# On self-similar sets with overlaps and inverse theorems for entropy in $\mathbb{R}^d$

Michael Hochman

**Abstract**

We study self-similar sets and measures on $\mathbb{R}^d$. Assuming that the defining iterated function system $\Phi$ does not preserve a proper affine subspace, we show that one of the following holds: (1) the dimension is equal to the trivial bound (the minimum of $d$ and the similarity dimension $s$); (2) for all large $n$ there are $n$-fold compositions of maps from $\Phi$ which are super-exponentially close in $n$; (3) there is a non-trivial linear subspace of $\mathbb{R}^d$ that is preserved by the linearization of $\Phi$ and whose translates typically meet the set or measure in full dimension. In particular, when the linearization of $\Phi$ acts irreducibly on $\mathbb{R}^d$, either the dimension is equal to $\min\{s, d\}$ or there are super-exponentially close $n$-fold compositions. We give a number of applications to algebraic systems, parametrized systems, and to some classical examples.

The main ingredient in the proof is an inverse theorem for the entropy growth of convolutions of measures on $\mathbb{R}^d$, and the growth of entropy for the convolution of a measure on the orthogonal group with a measure on $\mathbb{R}^d$. More generally, this part of the paper applies to smooth actions of Lie groups on manifolds.

## Contents



Supported by ERC grant 306494, partially supported by ISF grant 1409/11
*2010 Mathematics Subject Classification.* 28A80, 11K55, 11B30, 11P70

# 1 Introduction

Self similar sets and measures are among the simplest fractal objects: their defining property is that the whole is made up of finitely many objects similar to it, i.e. identical to the whole except for scaling, rotation and translation. When these smaller copies are sufficiently separated from each other the small-scale structure is relatively easy to understand and in particular the Hausdorff dimension can be computed precisely in terms of the defining similitudes. Without separation, however, things are significantly more complicated, and it is an open problem to compute the dimension. Many special cases of this problem have received attention, including the Erdös problem on Bernoulli convolutions, Furstenberg's projection problem for the 1-dimensional Sierpinski gasket (now settled), the Keane-Smorodinsky the $0, 1, 3$-problem, and "fat" Sierpinski gaskets (for more on these, see below).

For self-similar sets and measures in $\mathbb{R}$ there is a longstanding conjecture predicting that the dimension will be "as large as possible", subject to the combinatorial constraints, unless there are exact overlaps, i.e. unless some of the (iterated) small-scale copies of the original coincide. In recent work [12] we introduced methods from additive combinatorics to this problem and obtained a partial result towards the conjecture, showing that if the dimension is "too small" then there are super-exponentially close pairs of small-scale copies. In particular, for some important classes of self-similar sets, e.g. those defined by similarities with algebraic coefficients, this resolves the conjecture.

In the present paper we treat the general case of self-similar sets and measures in $\mathbb{R}^d$. Easy examples show that in the higher-dimensional setting the conjecture above is false as stated (Example 1.2). The main new feature of the problem is that the linear parts of the defining similarities may act reducibly on $\mathbb{R}^d$, and "excess dimension" may accumulate on non-trivial invariant subspaces and produce dimension loss. To correct this we propose here a modified version of the conjecture that takes this possibility into account (Conjecture 1.3), and prove a weak version of it (Theorem 1.5), analogous to the main result of [12]. We give various applications, in particular we show that the modified conjecture holds when the linear action is irreducible and the coefficients of the similarities are algebraic.

As in the 1-dimensional case, a central ingredient in the proof is an inverse theorem about the structure of probability measures on $\mathbb{R}^d$ whose convolutions have essentially the same entropy as the original (Theorem 2.8). In fact, what we really need is a result of this type for the convolution of a measure on $\mathbb{R}^d$ with a measure on the similarity group, or one on the isometry group (Theorem 2.12). These results are of independent interest, and provide a versatile tool for analyzing smooth images of product measures. We take the opportunity to develop these methods here, in particular stating results for convolutions in Lie groups and their actions (Theorems 2.12, 2.14 and the subsequent corollaries).

### 1.1 Setup: Self-similar sets and measures

Let $G$ denote the group of similarities of $\mathbb{R}^d$, namely maps $x \mapsto rUx + a$ for $r \in (0,\infty)$, $a \in \mathbb{R}^d$ and $U$ a $d \times d$ orthogonal matrix; we denote the map simply by $\varphi = rU + a$. In this paper an iterated function system means a finite family $\Phi = \{\varphi_i\}_{i\in\Lambda} \subseteq G$ consisting of contractions, so $\varphi_i = r_iU_i + a_i$ with $0 < r_i < 1$. A self similar set is the attractor of such a system, defined as the unique compact set $\emptyset \neq X \subseteq \mathbb{R}$ satisfying

$$X = \bigcup_{i\in\Lambda} \varphi_i X. \tag{1}$$

The self-similar measure determined by $\Phi$ and a positive probability vector $(p_i)_{i\in\Lambda}$ is the unique Borel probability measure $\mu$ on $\mathbb{R}^d$ satisfying

$$\mu = \sum_{i=1}^{k} p_i \cdot \varphi_i\mu.$$

Here and throughout, $\varphi\mu = \mu \circ \varphi^{-1}$ denotes the push-forward of $\mu$ by $\varphi$.

It is a classical problem to understand the small-scale structure of self-similar sets and measures, and especially their dimension. We shall write $\dim A$ for the

Hausdorff dimension of $A$ and define the dimension of a finite Borel measure $\theta$ by[1]

$$\dim \theta = \inf\{\dim E \,:\, \theta(E) > 0\}.$$

The textbook case of self-similar sets and measures occurs when the images $\varphi_i X$ are disjoint, or satisfy some weaker separation assumption (e.g. the open set condition). Then the dimension can be computed exactly: $\dim X$ is equal to the similarity dimension[2] s-dim $X$, i.e. the unique $s \geq 0$ solving the equation $\sum |r_i|^s = 1$, and $\dim \mu$ is equal to the similarity dimension of $\mu$, defined by

$$\text{s-dim}\, \mu = \frac{\sum p_i \log p_i}{\sum p_i \log r_i}.$$

It is when the images $\varphi_i X$ have more substantial overlap that the problem becomes very challenging. The similarity dimension, and the dimension $d$ of the ambient space $\mathbb{R}^d$, still constitute upper bounds. Thus one always has

$$\dim X \leq \min\{d, \text{s-dim}\, X\} \tag{2}$$
$$\dim \mu \leq \min\{d, \text{s-dim}\, \mu\}. \tag{3}$$

In general little more is known. In fact, we usually cannot even determine whether or not equality holds in (2) and (3). There is one exception to this, which arises from combinatorial coincidences of cylinder sets. For $i = i_1 \ldots i_n \in \Lambda^n$ write

$$\varphi_i = \varphi_{i_1} \circ \ldots \circ \varphi_{i_n}.$$

One says that exact overlaps occur if there is an $n$ and distinct $i, j \in \Lambda^n$ such that $\varphi_i = \varphi_j$ (in particular the images $\varphi_i X$ and $\varphi_j X$ coincide).[3] If this occurs then the attractor (or self-similar measure) can be expressed using an IFS $\Psi$ which is a proper subset of $\{\varphi_i\}_{i \in \Lambda^n}$, and a strict inequality in (2) and (3) may follow from the trivial bounds (2) and (3) applied to the IFS $\Psi$.

## 1.2 Main results

Define the distance between similarities $\psi = rU + a$ and $\psi' = r'U' + a'$ by[4]

$$d(\psi, \psi') = |\log r - \log r'| + \|U - U'\| + \|a - a'\|. \tag{4}$$

Here $\|\cdot\|$ denotes the Euclidean or operator norm as appropriate. Given an IFS $\Phi = \{\varphi_i\}_{i \in \Lambda}$, let

$$\Delta_n = \min\{d(\varphi_i, \varphi_j) \,:\, i, j \in \Lambda^n \,,\, i \neq j\}. \tag{5}$$

[1] This is the lower Hausdorff dimension. Many other notions of dimension exist but since self-similar measures are exact dimensional [8], for them all the major ones coincide.

[2] The similarity dimension depends on the IFS $\Phi$ rather than the attractor, but we prefer the shorter notation s-dim $X$ in which $\Phi$ is implicit. The meaning should always be clear from the context. A similar comment holds for the similarity dimension of measures.

[3] If $i \in \Lambda^k$, $j \in \Lambda^m$ and $\varphi_i = \varphi_j$, then $i$ cannot be a proper prefix of $j$ and vice versa, because the maps are all contractions. Thus $ij, ji \in \Lambda^{k+m}$ are distinct, and $\varphi_{ij} = \varphi_{ji}$. This shows that our definition is equivalent to one asking for coincidence of compositions of possibly different lengths. Stated differently, exact overlaps means that the semigroup generated by the $\varphi_i$, $i \in \Lambda$, is not freely generated by them.

[4] In [12] we used the stronger metric in which the term $|\log r - \log r'|$ is replaced by the discrete distance $\delta_{r,r'}$. One could do the same here but the metric above is better suited in some of the generalizations presented in Section 2.12 and is good enough for our applications, so we restrict ourselves to it.

Note that exact overlaps occur if and only if $\Delta_n = 0$ for all large $n$, and it is easy to see that $\Delta_n \to 0$ at least exponentially fast (this is an easy consequence of contraction). Convergence may or may not be faster than this, but we note that in some cases there is an exponential lower bound $\Delta_n \geq c^n > 0$.

The main result of [12] was a step towards the folklore conjecture that when $d = 1$, the occurrence of exact overlaps is the only mechanism which can lead to a strict inequality in (2) and (3). Specifically, we proved the following [12, Corollary 1.2]:

**Theorem 1.1.** *For a self-similar set $X \subseteq \mathbb{R}$, if $\dim X < \min\{1, \text{s-dim}\, X\}$ then $\Delta_n \to 0$ super-exponentially, i.e. $-\frac{1}{n}\log \Delta_n \to \infty$. The same conclusion holds if $\dim \mu < \min\{1, \text{s-dim}\, \mu\}$ for a self-similar measure $\mu$ on $X$.*

When $d \geq 2$, the analogous conjecture and analogous theorem are both false. A trivial class of counterexamples arise when the maps in $\Phi$ preserve a non-trivial affine subspace $V < \mathbb{R}^d$, which is equivalent to having $X \subseteq V$. In this case, if $\text{s-dim}\, X > \dim V$, then the trivial bound gives

$$\dim X \leq \min\{\dim V, \text{s-dim}\, X\} = \dim V < \min\{d, \text{s-dim}\, X\},$$

even though there may be no exact overlaps.

We say that $\Phi$ is affinely irreducible if the only trivial affine subspaces are simultaneously preserved by all $\varphi_i \in \Phi$. The following example shows that affine irreducibility is also not enough for an analog of Theorem 1.1 to hold.

**Example 1.2.** Begin with the IFS $\Phi = \{\varphi_\pm\}$ on $\mathbb{R}$ given by $\varphi_\pm(x) = \lambda^{-1}x \pm 1$, where $\lambda = 1.6956\ldots$ is the real root of $t^3 - t^2 - 2 = 0$. This example, due to Garsia [11], has the property that $\Delta_n \geq c \cdot 2^{-n}$, and the attractor is the interval $[-\frac{\lambda}{\lambda-1}, \frac{\lambda}{\lambda-1}]$. Let $\Phi^3 = \{\varphi_i\}_{i \in \{\pm\}^3}$ denote the IFS consisting of all three-fold compositions of the maps $\varphi_+, \varphi_-$. Then $\Phi^3$ has the same attractor and all the maps in $\Phi^3$ contract by the same ratio $\lambda^{-3} < 1/2$. Now let $\Psi = \{\varphi_-^3, \varphi_+^3\}$, where $\varphi^3 = \varphi \circ \varphi \circ \varphi$. Then $\Psi$ is an IFS with the same contraction ratio $\lambda^{-3}$ as $\Phi^3$ but it satisfies the strong separation condition (its attractor $Y$ is the disjoint union of $\varphi_+^3 Y$ and $\varphi_-^3 Y$), and hence $\dim Y = \log 2/\log \lambda^3 < 1$. Finally, take the product IFS $\Gamma = \Phi^3 \times \Psi$, consisting of all maps of the form $(x, y) \mapsto (\varphi x, \psi y)$ for $\varphi \in \Phi^3$, $\psi \in \Psi$. The attractor $Z$ of $\Gamma$ is just the product $Z = [-\frac{\lambda}{\lambda-1}, \frac{\lambda}{\lambda-1}] \times Y$ of the attractors of $\Phi^3$ and $\Psi$, and its dimension is $1 + \log 2/\log \lambda$. We can compute the similarity dimension of $Z$ using $\lambda^2 > 2$ and $\lambda^3 - \lambda^2 - 2 = 0$:

$$\text{s-dim}\, Z = \frac{\log |\Gamma|}{\log \lambda^3} = \frac{\log 16}{\log \lambda^3} = \frac{\log 16}{\log(2 + \lambda^2)} < 2.$$

We therefore have (using $\lambda < 2$):

$$\dim Z = 1 + \frac{\log 2}{\log \lambda} < \frac{\log 16}{\log \lambda^3} = \min\{2, \text{s-dim}\, Z\}.$$

On the other hand, since both $\Phi^3$ and $\Psi$ have exponential lower bounds on the distance between cylinders, there is also an exponential lower bound for $\Gamma$. Thus, the example shows that a strict inequality in (1) with neither exact overlaps or even super-exponential concentration of cylinders.

Two things stand out about this example. First, the foliation of $\mathbb{R}^2$ by horizontal lines is preserved by all maps in $\Gamma$, and, second, the excess similarity dimension is being "absorbed" in the intersection of the attractor of $\Gamma$ with these lines. Indeed, in these intersections we are seeing essentially the 1-dimensional IFS $\Phi$, and we are not getting all of the potential dimension out of it, since its similarity dimension is $> 1$ but attractor is "trapped" in a line. We do, however, have the maximal possible dimension for the intersection of $Z$ with those horizontal lines that intersect it.

For an IFS $\Phi = \{\varphi_i\}_{i\in\Lambda}$ on $\mathbb{R}^d$, we say that a linear subspace $V < \mathbb{R}^d$ is $D\Phi$-invariant if it is invariant under the orthogonal parts (i.e. differentials) $U_i = D\varphi_i$ of $\varphi_i \in \Phi$, and nontrivial if $0 < \dim V < d$. If every $D\Phi$-invariant subspace is trivial then $\Phi$ is said to be *linearly irreducible*. The discussion above suggests the following:

**Conjecture 1.3.** *Let $X \subseteq \mathbb{R}^d$ be the attractor of an affinely irreducible IFS $\Phi \subseteq G$. Then one of the following must hold:*

*(i)* $\dim X = \min\{d, \text{s-dim}\, X\}$.

*(ii) There are exact overlaps.*

*(iii) There is a non-trivial $D\Phi$-invariant linear subspace $V \leq \mathbb{R}^d$ and $x \in X$ such that*

$$\dim(X \cap (V + x)) = \dim V.$$

One might even conjecture a stronger form of (iii), e.g. that the set of points $x$ in question is of full dimension in $X$, or is large in some other sense.

The main result of this paper, Theorem 6.15, confirms a weakened version of Conjecture 1.3:

**Theorem 1.4.** *Let $X \subseteq \mathbb{R}^d$ be the attractor of an affinely irreducible IFS $\Phi \subseteq G$. Then one of the following must hold:*

*(i')* $\dim X = \min\{d, \text{s-dim}\, X\}$.

*(ii')* $\Delta_n \to 0$ *super-exponentially.*

*(iii') There exists a non-trivial $D\Phi$-invariant linear subspace $V \leq \mathbb{R}^d$ and $x \in X$ such that*

$$\dim(X \cap (V + x)) = \dim V.$$

The alternatives are not exclusive (all three may hold simultaneously).

The theorem follows, as in the one-dimensional case, from a more precise statement about the entropy of the measure at small scales. We require some notation. The level-$n$ dyadic partition $\mathcal{D}_n$ of $\mathbb{R}$ is the partition into intervals $[k/2^n, (k+1)/2^n)$, $k \in \mathbb{Z}$. The level-$n$ dyadic partition of $\mathbb{R}^d$ is given by

$$\mathcal{D}_n^d = \{I_1 \times \ldots \times I_d \,:\, I_i \in \mathcal{D}_n\}.$$

We omit the superscript $d$ when it is clear from the context.

For a probability measure $\nu$ and partitions $\mathcal{E}, \mathcal{F}$ of the underlying probability space we write $H(\nu, \mathcal{E}) = -\sum_{E\in\mathcal{E}} \nu(E)\log\nu(E)$ and $H(\nu, \mathcal{E}|\mathcal{F}) = H(\nu, \mathcal{E} \vee \mathcal{F}) - H(\nu, \mathcal{F})$ for the entropy and conditional entropy of $\nu$ with respect to $\mathcal{E}$ (conditioned on $\mathcal{F}$, respectively). Here $\mathcal{E} \vee \mathcal{F}$ is the common refinement of the partitions $\mathcal{E},\mathcal{F}$. We also write $H(\nu)$ for the entropy of an atomic measure $\nu$ with respect to the partition into points.

It is convenient to parametrize $G$ as a subset of $\mathbb{R}\times M_d(\mathbb{R})\times\mathbb{R}^d$, with $(t,U,a)$ corresponding to $2^{-t}U+a\in G$. Then the level-$n$ dyadic partition $\mathcal{D}_n^G$ of $G$ is defined as the partition induced from the corresponding level-$n$ partition of $\mathbb{R}\times M_d(\mathbb{R})\times\mathbb{R}^d\cong\mathbb{R}^{1+d^2+d}$. We also introduce the partitions $\mathcal{E}_n^G$ of $G$ induced by the dyadic partition according to the translation part of the similarities, which in the parametrization $G\subseteq\mathbb{R}\times M_d(\mathbb{R})\times\mathbb{R}^d$ is

$$\mathcal{E}_n^G=\{(\mathbb{R}\times M_d(\mathbb{R})\times D)\cap G\;:\;D\in\mathcal{D}_n^d\}.$$

Note that $\mathcal{D}_n^G$ refines $\mathcal{E}_n^G$.

Given a self-similar measure $\mu=\sum_{i\in\Lambda}p_i\cdot\varphi_i\mu$ and assuming all $p_i>0$, let

$$\nu^{(n)}=\sum_{i\in\Lambda^n}p_i\cdot\delta_{\varphi_i}.$$

This is a probability measure on $G$, but if we fix $\widetilde{x}$ in the attractor of $\Phi$ then the push-forward of $\nu^{(n)}$ via $g\mapsto g\widetilde{x}$ is the natural "$n$-th generation" approximation of $\mu$, given by[5]

$$\widetilde{\nu}^{(n)}=\sum_{i\in\Lambda^n}p_i\cdot\delta_{\varphi_i(\widetilde{x})}$$

(this measure depends on the choice of $\widetilde{x}$ but this is of little consequence). Let

$$r=\prod_{i\in\Lambda}r_i^{p_i}$$

denote the (geometric) average contraction and for $n\in\mathbb{N}$ let

$$n'=[n/\log(1/r)],$$

so that $2^{-n'}\sim r^n$.

Now, it is not hard to show $|H(\nu^{(n)},\mathcal{E}_{n'}^G)-H(\widetilde{\nu}^{(n)},\mathcal{D}_{n'})|=O(1)$ (in fact if we take $\widetilde{x}=0$ then two entropies are identical), and since it is easily seen that $\frac{1}{n'}H(\widetilde{\nu}^{(n)},\mathcal{D}_{n'})\to\dim\mu$, one concludes

$$\lim_{n\to\infty}\frac{1}{n'}H(\nu^{(n)},\mathcal{E}_{n'}^G)=\dim\mu.$$

Observe that when there are no exact overlaps, $\nu^{(n)}$ consists of $|\Phi|^n$ atoms whose masses are all the products $p_{i_1}\cdot\ldots\cdot p_{i_n}$, and hence $H(\nu^{(n)})=n\cdot(-\sum p_i\log p_i)$. Thus for fixed $n$,

$$\frac{1}{n'}H(\nu^{(n)},\mathcal{D}_k)\to\frac{-\sum p_i\log p_i}{\log r}=\text{s-dim}\,\mu\qquad\text{as }k\to\infty,$$

and if there is a strict inequality in (3) we would have

$$\begin{aligned}\frac{1}{n'}H(\nu^{(n)},\mathcal{D}_k^G|\mathcal{E}_{n'}^G) &= \frac{1}{n'}H(\nu^{(n)},\mathcal{E}_k^G)-\frac{1}{n'}H(\nu^{(n)},\mathcal{D}_k^G)\\ &\to \text{s-dim}\,\mu-\dim\mu\qquad\text{as }k\to\infty\\ &> 0\end{aligned}$$

[5]It is also common to approximated $\mu$ "at scale $\rho$" by putting the appropriate mass on the points $\varphi_{i_1\ldots i_m}(x)$, where $i_1\ldots i_m\in\Lambda^*$ are the sequences of minimal length such that $\varphi_{i_1\ldots i_m}$ contracts by at least $\rho$. We could use this approximation instead of $\nu^{(n)}$, but this would lead to messier notation and have little advantage.

Therefore it is possible to choose $k = k(n)$ such that the "excess" $\frac{1}{n'}H(\nu^{(n)}, \mathcal{D}^G_{k(n)}|\mathcal{E}^G_{n'})$ remains bounded away from 0 as $n \to \infty$. It is natural to ask at what rate this excess entropy emerges, that is, how fast $k(n)$ must grow for this to hold. The following theorem shows that it must grow at least super-linearly.

**Theorem 1.5.** *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$ be a self-similar measure for an affinely irreducible IFS $\Phi$. Then one of the following must hold:*

*(i")* $\dim \mu = \min\{d, \text{s-dim}\, \mu\}$.

*(ii")* $\lim_{n\to\infty} \frac{1}{n'}H(\nu^{(n)}, \mathcal{D}^G_{qn}|\mathcal{E}^G_{n'}) = 0$ *for all* $q > 1$.

*(iii") There is a non-trivial $D\Phi$-invariant linear subspace $V \le \mathbb{R}^d$ such that for $\mu$-a.e. $x$, the conditional measure $\mu_{V+x}$ on $V + x$ satisfies* $\dim \mu_{V+x} = \dim V$.

In fact, the second or third alternatives must hold irrespective of the validity of (i"). The usefulness of the theorem, however, lies in the fact that if (i") fails and (ii") holds then $\Delta_n \to 0$ super-exponentially.

Theorem 1.5 and Theorem 1.4 are usually applied by ruling out (iii') or (iii"), and then working out the implications for the dimension. One trivial way to rule it out is to just assume it:

**Corollary 1.6.** *If $\Phi$ is a linearly irreducible IFS $\Phi$, then its attractor $X$ satisfies (i') or (ii'), and every self-similar measure $\mu$ for $\Phi$ satisfies (i") or (ii").*

As there are no non-trivial linear subspaces of $\mathbb{R}$, every IFS acts linearly irreducibly, and we have recovered the main results of [12] (Theorem 1.1 above).

We say that $rU + a \in G$ is algebraic if $r$ and all the coordinates of $U$ and $a$ are algebraic numbers over $\mathbb{Q}$, and we say that an IFS $\Phi \subseteq G$ is algebraic if all of its elements are. If $\Phi$ is an algebraic IFS without exact overlaps, and we take $\widetilde{x} = 0$, then for each $n$, $\Delta_n$ is a polynomial in the algebraic parameters defining the maps on $\Phi$ and has degree $n$ and height at most exponential in $n$. This implies an exponential lower bound $\Delta_n \ge c^n$; this is a well known fact but we include a proof in Section 6.7. Thus we have ruled out (ii') ad (ii"), and obtained the following:

**Corollary 1.7.** *Let $\Phi$ be an algebraic IFS acting linearly irreducibly on $\mathbb{R}^d$ and without exact overlaps. Then $\dim \mu = \min\{d, \text{s-dim}\, \mu\}$ for every fully supported self-similar measure $\mu$ of $\Phi$, and $\dim X = \min\{d, \text{s-dim}\, X\}$.*

Our arguments are purely Euclidean and do not utilize any non-elementary properties of the orthogonal or similarity groups. However, the nature of these groups depends crucially on the dimension $d$. For $d \le 2$ the orthogonal group of $\mathbb{R}^d$ is abelian (and the similarity group is solvable). In particular, the set $\mathcal{U}_n = \{U_i\}_{i\in\Lambda^n}$ of the orthogonal parts of $\varphi_i$, $i \in \Lambda^n$, is of polynomial size in $n$, and does not contribute to the entropy $H(\nu^{(n)}, \mathcal{D}^G_{qn}|\mathcal{E}^G_{n'})$ (for the same reason, the contraction ratios do not contribute asymptotically to the entropy). For $d \ge 3$ the orthogonal group is a virtually simple Lie group with strong expansion properties, and typically $|\mathcal{U}_n|$ is exponential in $n$. Our methods do not make use of any special properties of the orthogonal group, but concurrently and independently with our work, Lindenstrauss and Varjú utilized the work of Bourgain and Gamburd [3] and of de Saxce [6] on spectral gap of random walks on the orthogonal group to prove the following result.

**Theorem 1.8** (Lindenstrauss-Varjú, [22]). *Let $U_1, \ldots, U_k \in SO(d)$ and $p = (p_1, \ldots, p_k)$ a probability vector. Suppose that the operator $f \mapsto \sum_{i=1}^k p_i f \circ U_i$ on $L^2(SO(d))$ has a spectral gap. Then there is a number $\widetilde{r} < 1$ such that for every choice $\widetilde{r} < r_1, \ldots, r_k < 1$, and for any $a_1, \ldots, a_k \in \mathbb{R}^d$, the self similar measure with weights $p$ for the IFS $\{r_i U_i + a_i\}_{i=1}^k$ is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^d$.*

The spectral gap hypothesis can currently be verified when the entries are algebraic and $U_i$ generate a dense subgroup of $O(d)$, but is conjectured to hold much more generally.

Compare this theorem to Corollary 1.7: The former ensures absolute continuity (which is a stronger property than full dimension), but only when the contraction of the IFS is uniformly close enough to 1, while the latter ensures that the dimension is $d$ as soon as there is no dimension obstruction (i.e. as soon as s-dim $\mu \geq d$), but does not give absolute continuity. It is probable that absolute continuity holds under the same assumptions but this remains open.

There are other cases in which possibility (iii') of Theorem 1.4 or (iii") of Theorem 1.5, can be ruled out. A trivial case is when the attractor $X$ of $\Phi$ satisfies $\dim X < k$, and all $D\Phi$-invariant subspaces have dimension $\geq k$. Another case is when $\Phi$ consists of homotheties (i.e. the orthogonal parts $U_i$ of the contractions are identities), and for every line $\ell$ in $\mathbb{R}^d$ we have

$$\sum_{i\,:\,\varphi_i(X)\cap\ell\neq\emptyset} r_i < 1$$

Then elementary covering considerations show that $\dim(X \cap \ell) < 1$ for every line $\ell \subseteq \mathbb{R}^d$, and consequently (iii') (and hence (iii")) fails for every subspace $V$. Similarly, if $\Phi$ consists of homotheties and $\mu = \sum p_i \cdot \varphi_i \mu$ is a self-similar measure such that for every line $\ell$,

$$\frac{\sum_{i\,:\,\varphi_i(X)\cap\ell\neq\emptyset} p_i \log p_i}{\sum_{i\,:\,\varphi_i(X)\cap\ell\neq\emptyset} p_i \log r_i} < 1$$

then one can deduce that $\dim \mu_{\ell+x} < 1$ for $\mu$-a.e. $x$, which by Marstrand's slice theorem rules out (iii"). Another alternative is to show that the linear images onto $(d-1)$-planes have dimension greater than $\dim \mu - 1$, in which case Dimension conservation [10] implies that in every dimension, the conditional measure on a.e. line has dimension $< 1$.

Unfortunately such arguments do not always apply, and we know of no general method to exclude (iii') and (iii"). See Theorem 1.16 and the discussion surrounding it.

## 1.3 Parametric families

Suppose that $I$ is a set of parameters and that for $t \in I$ we are given an IFS $\Phi_t = \{\varphi_{i,t}\}$, where $\varphi_{i,t}(x) = r_i(t)U_i(t)x + a_i(t)$ for functions $r_i, U_i, a_i$ defined on $I$. For $i, j \in \Lambda^n$ let

$$\Delta_{i,j}(t) = \varphi_{i,t}(0) - \varphi_{j,t}(0).$$

Then $\|\Delta_{i,j}(t)\|$ is the third term in the definition (4) of $d(\varphi_{i,t}, \varphi_{j,t})$, and hence, writing $\Delta_n(t)$ for the quantity defined as in (5) for the system $\Phi_t$, we have

$$\min\{\|\Delta_{i,j}(t)\| \,:\, i, j \in \Lambda^n \text{ distinct}\} \leq \Delta_n(t).$$

This gives the following formal consequence of Theorem 1.5:

**Theorem 1.9.** *Let $\{\Phi_t\}_{t\in I}$ be a parametric family of IFSs on $\mathbb{R}^d$. Let $E \subseteq I$ be the set*

$$E = \bigcap_{\varepsilon>0}\left(\bigcup_{N=1}^{\infty}\bigcap_{n>N}\left(\bigcup_{i,j\in\Lambda^n}\Delta_{i,j}^{-1}((-\varepsilon^n,\varepsilon^n)^d)\right)\right),$$

*and let $F \subseteq I$ be the set of parameters $t$ for which $\Phi_t$ is linearly reducible. Then for $t \in I \setminus (E \cup F)$, every self-similar measure $\mu$ for $\Phi_t$ satisfies* $\dim\mu = \min\{d, \text{s-dim}\,\mu\}$ *and similarly for the attractor of $\Phi_t$.*

The main case of interest is when $I \subseteq \mathbb{R}^m$. Then, under rather mild assumptions, the set $E$ of (potential) exceptions can be shown to be quite small. For $i,j \in \Lambda^{\mathbb{N}}$ let

$$\Delta_{i,j}(t) = \lim_{n\to\infty}\Delta_{i_1\ldots i_n, j_1\ldots j_n}(t).$$

**Theorem 1.10.** *Let $I \subseteq \mathbb{R}^m$ be connected and compact and let $\{\Phi_t\}_{t\in I}$ be a parametrized family of IFSs for which the associated functions $r_i(\cdot)$, $U_i(\cdot)$ and $a_i(\cdot)$ are real-analytic on a neighborhood of $I$. Suppose that*

$$\forall i,j \in \Lambda^{\mathbb{N}}\ \ (i \neq j \implies \Delta_{i,j} \not\equiv 0).$$

*Then the set $E$ of the previous theorem has Hausdorff and packing dimension $\leq m-1$. In particular if $\Phi_t$ is linearly irreducible for all $t \in I$, then outside a set of parameters $t$ of dimension $\leq m-1$ (and in particular for Lebesgue-a.e. parameter), the attractor and self-similar measures of $\Phi_t$ have the expected dimension (i.e. equality holds in equations* (2) *and* (3)*).*

The condition $\Delta_{i,j} \not\equiv 0$ rules out trivial cases. For instance the theorem cannot be expected to apply when $\Phi_t = \Phi$ does not depend on $t$ and the system $\Phi$ has exact overlaps, in which case there are indeed distinct $i,j \in \Lambda^{\mathbb{N}}$ with $\Delta_{i,j} \equiv 0$.

If $I \subseteq \mathbb{R}^m$ and the IFS is in $\mathbb{R}^d$, and $m \geq d$, then we expect that for each $i,j \in \Lambda^{\mathbb{N}}$ there typically will be a sub-manifold $I_{i,j} \subseteq I$ of dimension $m-d$ on which $\Delta_{i,j} = 0$ for $i \in I_{i,j}$. Thus, the dimension bound on $E$ that one expects is $m-d$ rather than the bound $m-1$ appearing in the theorem above. However, the hypothesis $\Delta_{i,j} \not\equiv 0$ in itself is certainly not enough to guarantee this bound. To see this, begin with any a 1-parameter family $\{\Phi_u\}_{u\in[0,1]}$ of linearly irreducible IFSs in $\mathbb{R}^2$, and define a two-parameter family by $\Phi_{(s,t)} = \Phi_{(s-t)^2}$, $(s,t) \in [0,1]^2$. One might expect, by the logic above, that $\dim E = m-d = 0$. But, evidently, on the 1-dimensional subspace $V = \{s = t\}$ we have $\Phi_{(s,t)} = \Phi_0$, and if the attractor of $\Phi_0$ happens to satisfy (2) with a strict inequality, then $\dim E \geq 1 \neq 0 = m-d$.

It is natural to suggest that, assuming linear irreducibility of the IFS, the "correct" bound for $E$ is

$$\dim E \leq m - \sup\left\{\dim \Delta_{i,j}^{-1}(0)\ :\ i,j \in \Lambda^{\mathbb{N}},\ i \neq j\right\}. \tag{6}$$

For $m = d = 1$, the bound proved in [12] coincides with this one. The difficulty in higher dimension is that the zero sets of real-analytic functions, and the behavior of the functions near them, are not so well understood (for real-analytic functions on the line things are simple: the zero set consists of isolated points, away from which the function grows polynomially in a well-understood manner). It seems likely that having effective bounds on the constants in Łojasiewicz's inequality [23] might advance the matter but this seems to a difficult question in itself. What we prove here is that the bound (6) holds if one makes an assumption analogous to the classical transversality assumption.

**Theorem 1.11.** *Let $I \subseteq \mathbb{R}^m$ be compact and let $\{\Phi_t\}_{t\in I}$ be a parametrized family of IFSs for which the associated functions $r_i(\cdot)$, $U_i(\cdot)$ and $a_i(\cdot)$ are real-analytic on a neighborhood of $I$. Suppose that there exists an $r \in \mathbb{N}$ such that for every distinct pair $i, j \in \Lambda^n$ and $t \in I$,*

$$\Delta_{i,j}(t) = 0 \qquad \Longrightarrow \qquad \mathrm{rank}\,(D\Delta_{i,j}(t)) \geq r.$$

*Then the set $E$ of Theorem 1.9 has Hausdorff and packing dimension $\leq m - r$.*

As noted above, it is likely that there is room for improvement in these results.

## 1.4 Applications

We demonstrate the use of Theorems 1.10 and 1.11 for families of self-similar measures in which one varies the translations, contractions, or the IFS. Proofs are given in Section 6.7.

Let $X_\Phi \subseteq \mathbb{R}^d$ denote the attractor of an IFS $\Phi$.

**Theorem 1.12.** *For a finite set $\Lambda$ and $d \in \mathbb{N}$ let $IFS_\Lambda \subseteq G(d)^\Lambda$ denote the set $|\Lambda|$-tuples of contracting similarities , which we identify with the set of IFSs indexed by $\Lambda$. Then*

$$\dim\{\Phi \in IFS_\Lambda \,:\, \dim X_\Phi < \min\{d, \text{s-dim}_\Phi X_\Phi\}\} \leq \dim IFS_\Lambda - 1.$$

*In particular,* $\dim X_\Phi = \dim\{1, \text{s-dim}\, X_\Phi\}$ *for a.e. IFS $\Phi \in IFS_\Lambda$.*

If one fixes the linear parts of the similarity maps and varies the translation part, one obtains a version of results by Simon and Solomyak [29]:

**Theorem 1.13.** *Let $\{U_i\}_{i\in\Lambda}$ be orthogonal maps acting irreducibly on $\mathbb{R}^d$ and fix $0 < r_i < 1$, $i \in \Lambda$, satisfying the condition*

$$i \neq j \qquad \Longrightarrow \qquad r_i + r_j < 1.$$

*Then there is a subset $A \subseteq (\mathbb{R}^d)^\Lambda$ with* $\dim(\mathbb{R}^d)^\Lambda \setminus A \leq d|\Lambda| - d$, *and such that for $a \in A$ the attractor of $\Phi = \{r_i U_i + a_i\}_{i\in\Lambda}$ satisfies* $\dim X_\Phi = \dim\{1, \text{s-dim}\, X_\Phi\}$*. In particular this is true for a.e. $a \in (\mathbb{R}^d)^\Lambda$.*

The condition on the contraction ratios plays a similar role in [29, Theorem 2.1(c)] and the forthcoming book [30], where it is used in conjunction with the transversality method. It is needed to control the rank of $D\Delta_{i,j}$, which in our setting is required in order to apply Theorem 1.11. It is not clear to what extent the restriction on the contractions in necessary, but without the irreducibility condition it certainly is, as follows from [29, Proposition 3.3].

Another variant of these results concerns projections of self-similar measures defined by homotheties. This is a variant of Marstrand's theorem and Furstenberg's projection problem [20, 12]:

**Theorem 1.14.** *Let $X \in \mathcal{P}(\mathbb{R}^d)$ be a self-similar set defined by an IFS consisting of homotheties and satisfying strong separation. Let $k < d$ and let $\Pi_{d,k}$ denote the set of orthogonal projections from $\mathbb{R}^d$ to $k$-dimensional subspaces. Then*

$$\dim\{\pi \in \Pi_{d,k} : \dim \pi X = \min\{k, \dim X\}\} \leq \dim \Pi_{d,k} - k.$$

A particularly interesting family are the Bernoulli convolutions with nonuniform contraction. Namely, for $0 < \beta, \gamma < 1$ let $\lambda_{\beta,\gamma}$ denote the self-similar measure of maximal dimension for the IFS $\{x \mapsto \beta x, x \mapsto \gamma x + 1\}$. Let $S \subseteq (0,1)^2$ be the set of $(\beta, \gamma)$ for which $\text{s-dim}\,\lambda_{\beta,\gamma} > 1$; it is expected that $\lambda_{\beta,\gamma}$ is absolutely continuous for a.e. $(\beta, \gamma) \in S$, but this has been established only in certain restricted parameter ranges, e.g. [27].

**Theorem 1.15.** $\dim \lambda_{\beta,\gamma} = \min\{1, \text{s-dim}\,\lambda_{\beta,\gamma}\}$ *outside a set of parameters* $(\beta, \gamma) \in (0,1)^2$ *of Hausdorff (and packing) dimension* 1. *In particular this holds for Lebesgue-a.e. pair* $(\beta, \gamma) \in (0,1)^2$.

Finally, our results can be applied to a higher-dimensional analogs of the Bernoulli convolutions problem, namely the "fat Sierpinski gasket", first studied by Simon and Solomyak [29]. For $\lambda \in (0,1)$ consider the system of contractions $\{\varphi_i\}_{i=a,b,c}$ where $a, b, c$ are the vertices of an equilateral triangle in $\mathbb{R}^2$ and $\varphi_u(x) = \lambda x + u$. The classical Sierpinski gasket arises from the choice $\lambda = 1/2$, and in general when $0 \leq \lambda \leq 1/2$ the open set condition is satisfied and the dimension of the attractor $S_\lambda$ is equal to the similarity dimension. When $\lambda > 2/3$ the attractor has non-empty interior, and this remains true for $\lambda \geq \lambda_*$, where $\lambda_* \approx 0.6478$ is the real root of $x^3 - x^2 + x = 1/2$; see Broomhead-Montaldi-Sidorov [5]. For $1/2 < \lambda \leq \lambda_*$, however, the the dimension is known only for certain special algebraic parameters and for Lebesgue-typical $\lambda$ in a certain sub-range, and similarly for absolute continuity of the appropriate self-similar measures. See Jordan [15] and Jordan-Pollicott [16].

**Theorem 1.16.** $\dim S_\lambda = \min\{2, \text{s-dim}\, S_\lambda\}$ *for* $\lambda \in (0,1)$ *outside a set of Hausdorff (and packing) dimension* 0.

The last result is an immediate consequence of Theorem 1.10 using the fact that $S_\lambda$ can be written also as the attractor of a linearly irreducible IFS (the one given above is reducible). The possibility of such a presentation of $S_\lambda$ comes from its rotational symmetries. Interestingly, our method do not give comparable results even for very slight variants of $S_\lambda$, e.g. the fat Sierpinski gaskets studied in [16].

## 1.5 Organization and notation

A key ingredient in our argument is played by on the growth of entropy of measures under convolution. This subject is developed in the next three sections: Section 2 introduces the statements and basic definitions, Section 3 contains preliminaries on entropy, saturation, concentration and convolutions, and Section 4 proves the main results on convolutions. In Section 5 we extend the results to convolutions of a measure on $\mathbb{R}^d$ with a measure on the isometry group. Finally, in Section 6 we state and prove our main theorem on self-similar sets and measures and their applications.

Some notation: $\mathbb{N} = \{1, 2, 3, \ldots\}$. All logarithms are to base 2. $\mathcal{P}(X)$ is the space of probability measures on $X$, endowed with the weak-* topology if appropriate. We follow standard "big $O$" notation: $O_\alpha(f(n))$ is an unspecified function bounded in absolute value by $C \cdot f(n)$ for some constant $C = C(\alpha)$ depending on $\alpha$. Similarly $o(1)$ is a quantity tending to 0 as the relevant parameter $\to \infty$. We implicitly suppress all dependence of constants on the dimension $d$ of $\mathbb{R}^d$. Thus $O(1)$ sometimes means $O_d(1)$. We sometimes write $-O(\cdot)$ instead of $+O(\cdot)$ to indicate that the error may be negative but formally the two notations are equivalent.

The statement "for all $s$ and $t > t(s), \ldots$" should be understood as saying "there exists a function $t(\cdot)$ such that for all $s$ and $t > t(s), \ldots$". The function $t(\cdot)$ will change between contexts, when we want a persistent name we will designate the function as $t_1(\cdot)$, $t_2(\cdot)$, $t_*(\cdot)$, etc.

For the reader's convenience we summarize our main notation in the table below.

| | |
|---|---|
| $d$ | Dimension of the ambient Euclidean space. |
| $B_r(x)$ | The open Euclidean ball of radius $r$ around $x$ |
| $\|x\|, \|A\|$ | Euclidean norm of $x \in \mathbb{R}^d$, operator norm of $A \in M_d(\mathbb{R})$ |
| $\dim$ | Hausdorff dimension of sets and measures |
| $\Phi = \{\varphi_i\}_{i\in\Lambda}$ | Iterated Function system, Section 1.1 |
| $X$ | Attractor of $\Phi$. Usually assume $0 \in X \subseteq [0, 1)$, Section 1.1 |
| $\mu$ | Self-similar measure (usually), Section 1.1 |
| $\varphi_{i_1\ldots i_n}$, $p_{i_1\ldots i_n}$ | $\varphi_{i_1} \circ \varphi_{i_2} \circ \ldots \varphi_{i_n}$ and $p_{i_1} \cdot p_{i_2} \cdot \ldots \cdot p_{i_n}$ |
| $\nu^{(n)}$ | $\sum_{i\in\Lambda^n} p_i \cdot \delta_{\varphi_i(0)}$, the $n$-th approximation of $\mu$ |
| $\mathcal{D}_n^k$ | $n$-th level dyadic partition of $\mathbb{R}^k$ ($k = d$ by default); Section 1.2 |
| $\mathcal{D}_n^G$ | Dyadic partition of $G \subseteq \mathbb{R}^+ \times M_d(\mathbb{R}) \times \mathbb{R}^d$, Section 1.2 |
| $\mathcal{E}_n^G$ | Dyadic partition of $G$ by translation part, Section 1.2 |
| $\mathcal{P}(X)$ | Space of probability measures on $X$. |
| $\mu_{x,n}, \mu^{x,n}$ | Component measures (raw and re-scaled), Section 2.3 |
| $S_t$ | Scaling map: $S_t(x) = 2^t x$ |
| $\tau_z$ | Translation map: $\tau_s(x) = x + s$ |
| $\mathbb{P}_{i\in I}$, $\mathbb{E}_{i\in I}$ | Distribution and expectation over components, Section 2.3 |
| $H(\mu, \mathcal{B})$ | Shannon entropy, Section 3.1 |
| $H(\mu, \mathcal{B}\|\mathcal{C})$ | Conditional entropy, Section 3.1 |
| $H_m(\mu)$ | $\frac{1}{m}H(\mu, \mathcal{D}_m)$, Section 3.1 |
| $G, G_0$ | The groups of similarities and isometries, respectively. |
| $\pi_V$ | Orthogonal projection to $V$ |
| $V^{(\varepsilon)}$ | $\varepsilon$-neighborhood of $V$ |
| $d(U, V)$ | Distance between linear subspaces of $U, V \leq \mathbb{R}^d$, Section 3.6 |
| $\sqsubseteq$ | Subset relation restricted to unit ball, Section 3.6 |
| $\angle(U, V)$ | (Modified) angle between linear subspaces, Section 3.6 |
| $\mu * \eta$ | Convolution of probability measure on $\mathbb{R}^d$. |
| $\nu \mathbin{\bullet} x$, $\nu \mathbin{\bullet} \mu$ | Action/convolution of $\nu \in \mathcal{P}(G)$ on $x \in \mathbb{R}^d$, $\mu \in \mathcal{P}(\mathbb{R}^d)$. |
| $m(\mu)$, $\Sigma(\mu)$ | Mean and covariance matrix of measure $\mu \in \mathcal{P}(\mathbb{R}^d)$, Section 4.2 |
| $\lambda_i(\mu), \lambda_i(\Sigma)$ | Eigenvalues of measure or covariance matrix, Section 4.2 |
| $\mathrm{eigen}_{1\ldots r}(\Sigma)$ | Span of top $r$ eigenvectors of $\Sigma$ (for measure, $\Sigma = \Sigma(\mu)$), Section 4.2 |
| $\mathrm{sat}(\eta, \varepsilon, n, m)$ | Set of $(V, \varepsilon, m)$-saturated subspaces at level $n$, Section 6.2 |

## Acknowledgment

I am grateful to Pablo Shmerkin and Boris Solomyak for their many helpful comments, and to Ariel Rapaport for his contribution to the argument in Section 6.4. Part of this work was done during a visit to Microsoft Research in Redmond, Washington, and I would like to thank Yuval Peres and the members of the theory group for their hospitality.


# 2 An inverse theorem for the entropy of convolutions

## 2.1 Entropy and additive combinatorics

A subject of independent interest and central to our work is an analysis of the growth of the entropy of measures under convolution, either with other measures or with measures on the group of isometries (or similarities). This topic will occupy us for a large part of the paper.

We begin with a discussion of convolutions on Euclidean space, leaving generalizations to later. It is convenient to introduce the normalized scale-$n$ entropy

$$H_n(\mu) = \frac{1}{n} H(\mu, \mathcal{D}_n).$$

This normalization makes $H_n(\mu)$ a finite-scale surrogate for the dimension of $\mu$. In particular, for $\mu \in \mathcal{P}([0,1)^d)$ we have

$$0 \leq H_n(\mu) \leq d,$$

with equality holding for all $n$ if and only if $\mu$ is Lebesgue measure on $[0,1)^d$, and in general for measures $\mu$ of bounded support,

$$0 \leq H_n(\mu) \leq d + O(\frac{1}{n}),$$

where the constant depends logarithmically on the diameter of the support.

Our aim is to obtain structural information about measures $\mu, \nu$ for which $\mu * \nu$ is small in the sense that

$$H_n(\mu * \nu) \leq H_n(\mu) + \delta, \tag{7}$$

where $\delta > 0$ is small but fixed, and $n$ is large. This problem is a relative of classical ones in additive combinatorics concerning the structure of sets $A, B$ whose sumset $A + B = \{a + b : a \in A, b \in B\}$ is appropriately small. The general principle is that when the sum is small, the sets should have some algebraic structure. Results to this effect are known as inverse theorems. For example the Freiman-Rusza theorem asserts that if $|A + B| \leq C|A|$ then $A, B$ are close, in a manner depending on $C$, to generalized arithmetic progressions[6] (the converse is immediate). See e.g [32].

[6] A generalized arithmetic progression is an injective affine image of a box in a higher-dimensional lattice.

The entropy of a discrete measure corresponds to the logarithm of the cardinality of a set, and convolution is the analog for measures of the sumset operation. Thus the analog of the condition $|A+A| \leq C|A|$ is

$$H_n(\mu * \mu) \leq H_n(\mu) + O(\frac{1}{n}). \tag{8}$$

An entropy version of Freiman's theorem was recently proved by Tao [31], who showed that if $\mu$ satisfies (8) then it is close, in an appropriate sense, to a uniform measures on a (generalized) arithmetic progression.

The condition (7), however, is significantly weaker than (8) even when $\nu = \mu$, and it is harder to draw conclusions from it about the global structure of $\mu$. Consider the following example. Start with an arithmetic progression of length $n_1$ and gap $\varepsilon_1$, and put the uniform measure on it. Now split each atom $x$ into an arithmetic progression of length $n_2$ and gap $\varepsilon_2 < \varepsilon_1/n_2$, starting at $x$ (so the entire gap fits in the space between $x$ and the next atom). Repeat this procedure $N$ times with parameters $n_i, \varepsilon_i$, and call the resulting measure $\mu$. Let $k$ be such that $\varepsilon_N$ is of order $2^{-k}$. It is not hard to verify that we can have $H_k(\mu) = 1/2$ but $|H_k(\mu) - H_k(\mu * \mu)|$ arbitrarily small. This example is actually the uniform measure on a (generalized) arithmetic progression, as predicted by Freiman-type theorems, but as we allow the rank $N$ to grow, the entropy growth can be made arbitrarily small. Furthermore, if one conditions $\mu$ on an exponentially small subset of its support one gets another example with the similar properties that is quite far from a generalized arithmetic progression.

Our main contribution to this matter is Theorem 2.8 below, which shows that constructions like the one above are, in a certain statistical sense, the only way that (7) can occur. We note that there is a substantial existing literature on the growth condition $|A+B| \leq |A|^{1+\delta}$, which is the sumset analog of (7). Such a condition appears in the sum-product theorems of Bourgain-Katz-Tao [4] and in the work of Katz-Tao [19], and in the Euclidean setting more explicitly in Bourgain's work on the Erdős-Volkmann conjecture [1] and Marstrand-like projection theorems [2]. However we have not found a result in the literature that meets our needs and, in any event, we believe that the formulation given here will find further applications.

## 2.2 Concentration and saturation on subspaces

We begin by discussing global properties of measures that lead to the inequality in (7), and formulate discrete analogs of them.

For a linear subspace $V \leq \mathbb{R}^d$ we say that a measure $\mu$ is absolutely continuous on a translate $V'$ of $V$ if it is absolutely continuous with respect to the $\dim V$-dimensional volume (Hausdorff measure) $\lambda_{V'}$ on $V'$. Suppose that $\mu \in \mathcal{P}(\mathbb{R}^d)$ is compactly supported on a translate $V_1$ of $V$, and is absolutely continuous there. Then the Lebesgue differentiation theorem implies that $\mu(B_r(x)) = c_x \cdot (r^{\dim V} + o(1))$ as $r \to 0$, and it follows that

$$H_n(\mu) = \dim V - o(1) \qquad \text{as } n \to \infty. \tag{9}$$

If $\nu \in \mathcal{P}(\mathbb{R}^d)$ is compactly supported on another translate $V_2$ of $V$, then $\nu * \mu$ is supported on $V_3 = V_1 + V_2$, which is a translate of $V$, and is absolutely continuous there. Thus it also satisfies (9), and consequently $H_n(\mu * \nu) = H_n(\mu) + o(1)$: i.e., at small scales there is negligible entropy growth, and (7) is satisfied.

More generally, let $W = V^{\perp}$ be the orthogonal complement of $V$ and write $\pi_W$ for the orthogonal projection to $W$. Suppose $\mu \in \mathcal{P}(\mathbb{R}^d)$ is compactly supported and its conditional measures on the translates of $V$ are absolutely continuous, that is, $\mu = \int \mu_w d\theta(w)$ where $\theta = \pi_W \mu$ and $\mu_w$ is $\theta$-a.s. supported and absolutely continuous on $\pi_W^{-1}(w) = V + w$. Then instead of (9), one can show that

$$H_n(\mu) = H_n(\pi_W(\mu)) + \dim V - o(1) \qquad \text{as } n \to \infty, \tag{10}$$

and, if $\nu$ is compactly supported on a translate of $V$, then $\mu * \nu$ again has absolutely continuous conditional measures on translates of $V$, and it projects to a translate of $\theta$, so it satisfies the same relation (10). Again, we have $H_n(\nu * \mu) = H_n(\mu) + o(1)$, and (7) is satisfied.

This discussion motivates the following finite-scale analogs. For $A \subseteq \mathbb{R}^d$ and $\varepsilon > 0$ denote the $\varepsilon$-neighborhood of $A$ by

$$A^{(\varepsilon)} = \{x \in \mathbb{R}^d \,:\, d(x, A) < \varepsilon\}.$$

**Definition 2.1.** Let $V \leq \mathbb{R}^d$ be a linear subspace and $\varepsilon > 0$. A measure $\mu \in \mathcal{P}(\mathbb{R}^d)$ is $(V, \varepsilon)$-concentrated if there is a translate $W$ of $V$ such that $\mu(W^{(\varepsilon)}) \geq 1 - \varepsilon$.

Note that $(V, \varepsilon)$-concentration does not imply that the measure is supported near $V$ itself, only near a translate of it. Next, discretizing (9) we have

**Definition 2.2.** Let $V \leq \mathbb{R}^d$ be a linear subspace, $\varepsilon > 0$ and $m \in \mathbb{N}$. A measure $\mu \in \mathcal{P}(\mathbb{R}^d)$ is $(V, \varepsilon)$-uniform at scale $m$, or $(V, \varepsilon, m)$-uniform, if it is $(V, 2^{-m})$-concentrated and $H_m(\mu) > \dim V - \varepsilon$.

Finally, discretizing (10), we have:

**Definition 2.3.** Let $V \leq \mathbb{R}^d$ be a linear subspace, $W = V^{\perp}$ its orthogonal complement, and $\varepsilon > 0$. A probability measure $\mu \in \mathcal{P}(\mathbb{R}^d)$ is $(V, \varepsilon)$-saturated at scale $m$, or $(V, \varepsilon, m)$-saturated, if

$$H_m(\mu) \geq H_m(\pi_W \mu) + \dim V - \varepsilon.$$

There are obvious relations between the notions above: being nearly uniform implies saturation, and saturation implies being essentially a convex combination of nearly uniform measures. Furthermore, as one would expect from the discussion above, if we convolve a measure which is highly concentrated on a subspace with another measure which is uniform or saturated on that subspace at some scale, there will be little entropy growth at that scale. For precise statements see Sections 3.3 and 3.7.

## 2.3 Component measures

Let $\mathcal{D}_n(x) \in \mathcal{D}_n$ denote the unique level-$n$ dyadic cell containing the point $x \in \mathbb{R}^d$. For $D \in \mathcal{D}_n$ let $T_D : \mathbb{R}^d \to \mathbb{R}^d$ be the unique homothety mapping $D$ to $[0,1)^d$. Recall that if $\mu \in \mathcal{P}(\mathbb{R})$ then $T_D \mu$ is the push-forward of $\mu$ through $T_D$ .

**Definition 2.4.** For $\mu \in \mathcal{P}(\mathbb{R}^d)$ and a dyadic cell $D$ with $\mu(D) > 0$, the (raw) $D$-component of $\mu$ is

$$\mu_D = \frac{1}{\mu(D)} \mu|_D,$$

and the (rescaled) $D$-component is

$$\mu^D = \frac{1}{\mu(D)} T_D(\mu|_D).$$

For $x \in \mathbb{R}^d$ with $\mu(\mathcal{D}_n(x)) > 0$ we write

$$\begin{aligned} \mu_{x,n} &= \mu_{\mathcal{D}_n(x)} \\ \mu^{x,n} &= \mu^{\mathcal{D}_n(x)}. \end{aligned}$$

These measures, as $x$ ranges over all possible values for which $\mu(\mathcal{D}_n(x)) > 0$, are called the level-$n$ components of $\mu$.

Our results on the multi-scale structure of $\mu \in \mathbb{R}^d$ are stated in terms of the behavior of random components of $\mu$, defined as follows.[7]

**Definition 2.5.** Let $\mu \in \mathcal{P}(\mathbb{R}^d)$.

1. A random level-$n$ component, raw or rescaled, is the random measure $\mu_D$ or $\mu^D$, respectively, obtained by choosing $D \in \mathcal{D}_n$ with probability $\mu(D)$; equivalently, this is the random measure $\mu_{x,n}$ or $\mu^{x,n}$, respectively, with $x$ chosen according to $\mu$.
2. For a finite set $I \subseteq \mathbb{N}$, a random level-$I$ component, raw or rescaled, is chosen by first choosing $n \in I$ uniformly, and then (conditionally independently on the choice of $n$) choosing a raw or rescaled level-$n$ component.

*Notation* 2.6. When the symbols $\mu^{x,i}$ and $\mu_{x,i}$ appear inside an expression $\mathbb{P}(\ldots)$ or $\mathbb{E}(\ldots)$, they will always denote random variables drawn according to the component distributions defined above. The range of $i$ will be specified as needed. When dealing with components of several measures $\mu, \nu$, we assume all choices of components are independent unless otherwise stated.

The definition is best understood with some examples. For $\mathcal{A}, \mathcal{B} \subseteq \mathcal{P}([0,1]^d)$, and writing $1_{\mathcal{A}}$ for the indicator function of $\mathcal{A}$, we have

$$\begin{aligned} \mathbb{P}_{i=n}\left(\mu^{x,i} \in \mathcal{A}\right) &= \int 1_{\mathcal{A}}(\mu^{x,n})\, d\mu(x) \\ \mathbb{P}_{0\le i\le n}\left(\mu^{x,i} \in \mathcal{A}\right) &= \frac{1}{n+1}\sum_{i=0}^{n} \int 1_{\mathcal{A}}(\mu^{x,i})\, d\mu(x) \\ \mathbb{P}_{i=n}\left(\mu^{x,i} \in \mathcal{A}\,,\, \nu^{y,i} \in \mathcal{B}\right) &= \int\int 1_{\mathcal{A}}(\mu^{x,n}) \cdot 1_{\mathcal{B}}(\nu^{y,n})\, d\mu(x)\, d\nu(y). \end{aligned}$$

This notation implicitly defines $x, i$ as random variables. Thus if $\mathcal{A}_0, \mathcal{A}_1, \ldots \subseteq \mathcal{P}([0,1]^d)$ and $D \subseteq [0,1]^d$ we could write

$$\mathbb{P}_{0\le i\le n}\left(\mu^{x,i} \in \mathcal{A}_i \text{ and } x \in D\right) = \frac{1}{n+1}\sum_{i=0}^{n} \mu\left(x \,:\, \mu^{x,i} \in \mathcal{A}_i \text{ and } x \in D\right).$$

[7]Definition 2.5 is motivated by Furstenberg's notion of a CP-distribution [9, 10, 13], which arise as limits as $N \to \infty$ of the distribution of components of level $1, \ldots, N$. These limits have a useful dynamical interpretation but in our finitary setting we do not require this technology.

Similarly, for $f : \mathcal{P}([0,1)^d) \to \mathbb{R}$ and $I \subseteq \mathbb{N}$,

$$\mathbb{E}_{i\in I}\left(f(\mu^{x,i})\right) = \frac{1}{|I|}\sum_{i\in I}\int f(\mu^{x,i})\,d\mu(x).$$

We use similar expectation notation to average a sequence $a_n, \ldots, a_{n+k} \in \mathbb{R}$:

$$\mathbb{E}_{n\le i\le n+k}\left(a_i\right) = \frac{1}{k+1}\sum_{i=n}^{n+k} a_i.$$

We note in particular one trivial identity that will be used repeatedly later on:

$$\mu = \mathbb{E}_{i=n}\left(\mu_{x,i}\right). \tag{11}$$

Component distributions have the convenient property that they are almost invariant under repeated sampling, i.e. choosing components of components. More precisely, for $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $m, n \in \mathbb{N}$, let $\mathbb{P}_n^\mu$ denote the distribution of components $\mu^{x,i}$, $0 \le i \le n$, as defined above; and let $\mathbb{Q}_{n,m}^\mu$ denote the distribution on components obtained by first choosing a random component $\mu^{x,i}$, $0 \le 1 \le n$, as above, and then, conditionally on $\theta = \mu^{x,i}$, choosing a component $\theta^{y,j}$, $i \le j \le i+m$ with the usual distribution (note that $\theta^{y,j} = \mu^{y,j}$ is indeed a component of $\mu$).

**Lemma 2.7.** *Given $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $m, n \in \mathbb{N}$, the total variation distance between $\mathbb{P}_n^\mu$ and $\mathbb{Q}_{n,m}^\mu$ satisfies*

$$\left\|\mathbb{P}_n^\mu - \mathbb{Q}_{n,m}^\mu\right\| = O(\frac{m}{n})$$

*In particular if $\mathcal{A}, \mathcal{B} \subseteq \mathcal{P}([0,1)^d)$ and $\varepsilon, \delta > 0$ are such that*

$$\begin{aligned} \mathbb{P}_{0\le i\le n}(\mu^{x,y} \in \mathcal{A}) &> 1-\varepsilon \\ \mathbb{P}_{i\le j\le i+m}(\theta^{y,i} \in \mathcal{B}) &> 1-\delta \qquad \textit{for every } \theta \in \mathcal{A} \end{aligned} \tag{12}$$

*Then*

$$\mathbb{P}_{0\le i\le n}(\mu^{x,i} \in \mathcal{B}) > 1-\varepsilon-\delta-O(\frac{m}{n})$$

*Proof.* Observe that both $\mathbb{P}_n^\mu$ and $\mathbb{Q}_{n,m}^\mu$ produce a component $\mu_{z,k}$ by choosing $z$ according to $\mu$, and independently choosing a level $k \in \mathbb{N}$. The difference is that $\mathbb{P}_n^\mu$ chooses $k$ uniformly in the range $0, \ldots, n$, whereas for $\mathbb{Q}_{n,m}^\mu$, an elementary calculation shows that with probability $1 - O(m/n)$ is choses $k$ uniformly in the range $m, m+1, \ldots, n$, and with probability $O(m/n)$ it is chooses $k \in \{0, 1, , \ldots, m-1\} \cup \{n+1, \ldots, n+m\}$ (one can easily determine the distribution in this case but it is not relevant here). This gives the first statement.

For the second statement, what we want to show is that $\mathbb{P}_n^\mu(\mathcal{B}) > 1-\varepsilon-\delta-O(m/n)$. This will follow from the first statement if we show that $\mathbb{Q}_{n,m}^\mu(\mathcal{B}) > 1-\varepsilon-\delta$. Let $\theta = \mu^{x,i}$ and $\theta^{y,j}$ be as in the previous paragraph, so $\theta^{y,j}$ is distributed according to $\mathbb{Q}_{n,m}^\mu$ . By the law of total probability and our hypotheses,

$$\begin{aligned} \mathbb{Q}_{n,m}^\mu(\mathcal{B}) &= \mathbb{P}(\theta^{y,j} \in \mathcal{B}) \\ &\ge \mathbb{P}(\theta^{y,j} \in \mathcal{B} | \mu^{x,i} \in \mathcal{A}) \cdot \mathbb{P}(\mu^{x,i} \in \mathcal{A}) \\ &> (1-\delta)(1-\varepsilon) \end{aligned}$$

and the claim follows. ∎

Similar statements hold for raw components and components of measures on the similarity group. We omit the proofs, which are the same.

## 2.4 An inverse theorem for convolutions on $\mathbb{R}^d$

Our main result on entropy growth is that the global obstructions described at the beginning of Section 2.2 are the only local obstructions.

**Theorem 2.8.** *For every $R, \varepsilon > 0$ and $m \in \mathbb{N}$ there is a $\delta = \delta(\varepsilon, R, m) > 0$ such that for every $n > n(\varepsilon, R, \delta, m)$, the following holds: if $\mu, \nu \in \mathcal{P}([-R, R]^d)$ and*

$$H_n(\mu * \nu) < H_n(\mu) + \delta,$$

*then there exists a sequence $V_0, \ldots, V_n \leq \mathbb{R}^d$ of subspaces such that*

$$\mathbb{P}_{0 \leq i \leq n}\left( \begin{array}{c} \mu^{x,i} \textit{ is } (V_i, \varepsilon, m)\textit{-saturated and} \\ \nu^{y,i} \textit{ is } (V_i, \varepsilon)\textit{-concentrated} \end{array} \right) > 1 - \varepsilon. \tag{13}$$

The proof of the theorem is given in Section 4.6.

*Remark* 2.9. 1. The dependence of $\delta$ on $\varepsilon, m$ is effective, but the bounds we obtain are certainly far from optimal, and we do not pursue this topic. Also note that the theorem is not a characterization (this is already the case in dimension 1, see discussion after [12, Theorem 2.7]).

2. We have assumed that $\mu, \nu \in \mathcal{P}([-R, R]^d])$ but the theorem can be extended to measures with unbounded support having finite entropy by an approximation argument, see also [12, Section 5.5].

3. An application of Markov's inequality shows that (up to replacing $\varepsilon$ by $\sqrt{\varepsilon}$) equation (13) is equivalent to

$$\mathbb{P}_{0 \leq i \leq n}\left(\mu^{x,i} \text{ is } (V_i, \varepsilon, m)\text{-saturated and}\right) > 1 - \varepsilon \tag{14}$$

$$\mathbb{P}_{0 \leq i \leq n}\left(\nu^{y,i} \text{ is } (V_i, \varepsilon)\text{-concentrated}\right) > 1 - \varepsilon. \tag{15}$$

4. There is no assumption in the theorem on the entropy of $\nu$, but if $H_n(\nu)$ is sufficiently close to 0 the conclusion will automatically hold with $V_i = \{0\}$ (indeed, a small value of $H_n(\nu)$ implies that with high probability $\nu^{y,i}$ will be highly concentrated on $\{0\}$, so (14) holds, and (15) is automatic, every measure is $(\{0\}, \varepsilon, m\})$-saturated).

5. The version of Theorem 2.8 given in [12] for the case $d = 1$ had a somewhat different, but equivalent, appearance. The statement there was that for small enough $\delta > 0$, if $H_n(\mu * \nu) \leq H_n(\mu) + \delta$, then there exist disjoint sets $I, J \subseteq \{0, \ldots, n\}$ with $|I \cup J| > (1 - \varepsilon)n$ such that (14) holds for $V_i = \mathbb{R}$ when the expectation is conditioned on $i \in I$, and (15) holds for $V_i = \{0\}$ when the expectation is conditioned on $i \in J$. Indeed, if such $I, J \subseteq \{0, \ldots, n\}$ are given, observe that by setting $V_i = \mathbb{R}$ for $i \in I$ and $V_i = \{0\}$ for $i \in J$, and defining $V_i$ arbitrarily on the at most $\varepsilon n$ remaining $i$, equations (14) and (15) will hold for slightly larger $\varepsilon$ also without conditioning on $I, J$, because every measure is $(\mathbb{R}, \varepsilon)$-concentrated and $(\{0\}, \varepsilon, m)$-saturated. Thus the version in [12] implies the $d = 1$ case of Theorem 2.8. Conversely, assuming subspaces $V_i$ as in Theorem 2.8, we recover the version from [12] by setting $I = \{i : V_i = \mathbb{R}\}$ and $J = \{j : V_j = \{0\}\}$ and adjusting $\varepsilon$.

Specializing to self-convolutions and using some of the basic relations between saturation, concentration and uniformity, one deduces a multi-scale Freiman-type result:

**Theorem 2.10.** *For every $\varepsilon > 0$ and $m \in \mathbb{N}$, there is a $\delta = \delta(\varepsilon, m) > 0$ such that for every $n > n(\varepsilon, \delta, m)$ and every $\mu \in \mathcal{P}([0,1)^d)$, if*

$$H_n(\mu * \mu) < H_n(\mu) + \delta,$$

*then there exists a sequence $V_0, \ldots, V_n < \mathbb{R}^d$ such that*

$$\mathbb{P}_{0\le i\le n}\left(\mu^{x,i} \text{ is } (V_i, \varepsilon, m)\text{-uniform}\right) \quad > \quad 1 - \varepsilon.$$

## 2.5 An inverse theorem for isometries acting on $\mathbb{R}^d$

Recall that $G = G(d)$ denotes the group of similarities of $\mathbb{R}^d$. For $g = rU + a$ we write $r_g = r, U_g = U$ and $a_g = a$. The dyadic partitions $\mathcal{D}_n^G$ and $\mathcal{E}_n^G$ of $G$ were defined in Section 1.2 using the identification of $G$ with a subset of $\mathbb{R} \times M_d(\mathbb{R}) \times \mathbb{R}^d$. For $\nu \in \mathcal{P}(G)$ and for $g \in G$, $n \in \mathbb{N}$, we define the raw component $\nu_{g,n}$ in terms of the partition $\mathcal{D}_n^G$,

$$\nu_{g,n} = c \cdot \nu|_{\mathcal{D}_n^G(g)},$$

where $c$ is a normalizing constant. We adopt the same notation and conventions for these components as laid out in Section 2.3.

It is not natural in this context to define "rescaled" components. When we need to rescale we shall do so explicitly using the maps $S_t \in G$,

$$S_t x = 2^t x.$$

For $\nu \in \mathcal{P}(G)$ and $\mu \in \mathcal{P}(\mathbb{R}^d)$ we write $\nu . \mu$ for the push-forward of $\nu \times \mu$ via $(\varphi, x) \mapsto \varphi(x)$, and similarly for $x \in \mathbb{R}^d$ write $\nu . x$ for the push-forward of $\nu$ by $g \mapsto gx$. Our aim is to understand when the entropy of $\nu . \mu$ is large relative to the entropy of $\mu$, for $\nu \in \mathcal{P}(G)$ and $\mu \in \mathcal{P}(\mathbb{R}^d)$.

While our methods are able to treat this setting, it is more transparent if we assume that $\nu$ is supported on the isometry group $G_0 < G$, and we shall mostly restrict our attention to this case.

The statement we would like to make is that, if $\nu \in \mathcal{P}(G_0)$ and $\mu \in \mathcal{P}(\mathbb{R}^d)$, and if $\nu$ is of large entropy, then $\mu . \nu$ will have substantially more entropy than $\mu$, at small enough scales, unless certain specific obstructions occur. In the present setting the obvious global obstruction is that $\mu$ may be close to uniform on an orbit of a subgroup $H < G_0$, and $\nu$ supported on $H$ or a left coset of $H$. However, locally, this situation is not very different from the one we have already seen, and it is more natural to study the concentration of $\mu$ on affine subspaces, as in the Euclidean case. This is because the orbit of a point $x \in \mathbb{R}^d$ under a closed subgroup $H < G_0$ is a finite union of smooth manifolds, and at small scales these look like affine subspaces of $\mathbb{R}^d$ (essentially, the tangent hyperplanes of the manifolds). Thus we continue to state our results in terms of the concentration on subspaces of (the components of) $\mu$ and (the components of) the image of $\nu$ under the action.

Even so, there are several complications related to the phenomenon above. The first is demonstrated by the following example. Let $d = 2$, let $\mu$ be the uniform measure on the circle $\{x \in \mathbb{R}^2 : \|x\|_2 = 1\}$, and let $\nu$ be the uniform measure on the group of rotations about the origin. Then $\nu . \mu = \mu$, so there is no entropy growth. In this case, as predicted in the previous paragraph, the components $\mu^{x,n}$ become saturated on lines when $n$ is large, but the line varies according to the point $x$ (the distribution of directions for $x \sim \mu$ is of course uniform). In contrast, recall from Theorem 2.8 that, for convolutions of measures on $\mathbb{R}^d$, at each scale there was a single subspace on which, with high probability, all components of $\mu$ at a given level became saturated, irrespective of their spatial positions.

Another complication is the possibility that at small scales $\mu$ indeed becomes saturated, and $\nu$ concentrated, on subspaces, but that these subspaces are trivial. In the Euclidean setting such an occurrence was possible only if $\nu$ had nearly vanishing entropy, since if $H_n(\nu)$ is substantial then the components of $\nu$ cannot with high probability be highly concentrated on points. In the current setting, however, this cannot be ruled out. To see this let $\mu = \delta_0$ and let $\nu$ be normalized Haar measure on the orthogonal group $O(d) = \mathrm{stab}_{G_0}(0)$. Then $\nu . \mu = \mu$, so there is no entropy growth, and $\nu$ has large entropy at all scales, but the components of $\mu$ are not saturated on any non-trivial subspace. Thus the theorem above applies, but $V_i = \{0\}$. This type of situation can be avoided, however, if no part of the measure $\mu$ is close to a proper affine subspace. To make this quantitative we introduce the following definition:

**Definition 2.11.** $\mu \in \mathcal{P}(\mathbb{R}^d)$ is $(\varepsilon, \sigma)$-non-affine if $\mu(V^{(\sigma)}) < \varepsilon$ for every proper affine subspace $V \leq \mathbb{R}^d$.

We can now state the inverse theorem. Informally, it says that if $\nu . \mu$ does not have substantially more entropy than $\mu$, then, to most components of $\mu$ and $\nu$ at a moderately small scale, we can associate a subspace (depending on the components in question) such that the sub-components of the components typically become concentrated or saturated on this subspace. Furthermore, these subspaces will frequently be non-trivial if $\mu$ is not too close to being supported on a proper affine subspace of $\mathbb{R}^d$. Here is the precise formulation:

**Theorem 2.12.** *For every $\varepsilon > 0$, $R > 0$ and $m \in \mathbb{N}$, there exists $\delta = \delta(\varepsilon, R, m) > 0$ such that for every $k > k(\varepsilon, R, m)$ and every $n > n(\varepsilon, R, m, k)$, the following holds. For every $\nu \in \mathcal{P}(G_0)$ and $\mu \in \mathcal{P}([-R, R]^d)$ that are supported on balls of radius $R$, either*

$$H_n(\nu . \mu) > H_n(\mu) + \delta,$$

*or else, to every pair of level-$k$ components $\widetilde{\nu}$ of $\nu$ and $\widetilde{\mu}$ of $\mu$ we can assign a sequence of subspaces $V_i = V_i(\widetilde{\nu}, \widetilde{\mu}) < \mathbb{R}^d$, $0 \leq i \leq n$, such that with probability at least $1 - \varepsilon$ over the choice of $\widetilde{\mu}, \widetilde{\nu}$,*

$$\mathbb{P}_{0 \leq i \leq n}\left( \begin{array}{c} \widetilde{\mu}^{x,i} \text{ is } (V_i, \varepsilon, m)\text{-saturated and} \\ S_i U_g^{-1}(\widetilde{\nu}_{g,i} . x) \text{ is } (V_i, \varepsilon)\text{-concentrated} \end{array} \right) > 1 - \varepsilon$$

*If in addition $\mu$ is $((\varepsilon/5d)^{2(d+1)}, \sigma)$-non-affine for some $\sigma > 0$, and the relation among parameters takes $\sigma$ into account, then for those $\widetilde{\nu}, \widetilde{\mu}$ in the set of good components above, then for those $\widetilde{\nu}, \widetilde{\mu}$ in the set of good components above,*

$$\frac{1}{n+1} \sum_{i=0}^{n} \dim V_i > \frac{1}{d+1} H_n(\widetilde{\nu}) - \varepsilon,$$

*and*

$$\mathbb{E}_{i=k}\left(\frac{1}{n+1}\sum_{j=0}^{n}\dim V_j(\nu_{g,i},\mu_{x,i})\right) > \frac{1}{d+1}H(\nu)-\varepsilon \tag{16}$$

*Remark* 2.13.

1. Given $\varepsilon$, the assumption that $\mu$ is $((\varepsilon/5d)^{2(d+1)},\sigma)$-non-affine is global, and imposes no restriction on the structure of $\mu$ below at scales smaller than $O(\varepsilon^{2(d+1)}))$. Indeed, if $\mu\in\mathcal{P}(\mathbb{R}^d)$ does not give mass to any affine subspace, then for any $\tau$ it is $(\tau,\sigma)$-non-concentrated for some $\sigma>0$. Thus, if we fix $\mu$ in advance, then for every $\varepsilon, m$ the conclusion of the theorem holds automatically for suitable parameters $\delta, k, n$, and all $\nu\in\mathcal{P}(G)$.
2. The average in (16) is over all pairs of components $\nu_{g,k},\mu_{x,k}$, not only those for which the first part of the conclusion holds. But the total mass of the exceptional components is at most $\varepsilon$, and $\dim V_i\leq d$, so the exceptional components contribute $O(\varepsilon)$ to the average, which is of the same order as the error term. Thus we get an equivalent statement if in (16) we average only over only the "good" components from the first part of the theorem.

The proof of the theorem is based on a linearization argument which allows us to apply Theorem 2.8 from the Euclidean setting. See Section 5.5.

## 2.6 Generalizations

It is possible to apply our methods also to convolutions in Lie groups, actions of Lie groups on manifolds, and more general settings. Let $I\subseteq\mathbb{R}^{d_1}$ and $J\subseteq\mathbb{R}^{d_2}$ be closed balls and $f:I\times J\to\mathbb{R}^d$ a $C^1$ map. For $z=(x,y)\in I\times J$ we can write the differential $Df(z):\mathbb{R}^{d_1+d_2}\to\mathbb{R}^d$ in matrix form, as

$$Df(z)=[A_z,B_z]:\mathbb{R}^{d_1+d_2}\to\mathbb{R}^d,$$

where $A_z\in M_{d\times d_1}$ and $B_z\in M_{d\times d_2}$.

**Theorem 2.14.** *Let $f:I\times J\to\mathbb{R}^d$ be as above. For every $\varepsilon>0$ and $m\in\mathbb{N}$ there exists $\delta=\delta(f,\varepsilon,m)>0$ such that for every $k>k(f,\varepsilon,m)$ and every $n>n(f,\varepsilon,m,k)$, the following holds. Let $\nu\in\mathcal{P}(I)$ and $\mu\in\mathcal{P}(J)$. Then either*

$$H_n(f(\mu\times\nu)) > \int H_n(f(\mu\times\delta_y))\,d\nu(y)+\delta \tag{17}$$

*or else, for independently chosen level-$k$ components $\widetilde{\mu},\widetilde{\nu}$ of $\mu,\nu$, respectively, with probability at least $1-\varepsilon$ there are subspaces $V_0,\ldots,V_n<\mathbb{R}^d$ such that*

$$\mathbb{P}_{0\leq i\leq n}\left(\begin{array}{c} A_{x,y}\widetilde{\mu}^{x,i}\text{ is }(V_i,\varepsilon,m)\text{-saturated and}\\ B_{x,y}\widetilde{\nu}^{y,i}\text{ is }(V_i,\varepsilon)\text{-concentrated}\end{array}\right)>1-\varepsilon$$

*and*

$$\frac{1}{n+1}\sum_{i=0}^{n}\dim V_i > c\int H_n(f(\delta_x\times\nu))\,d\widetilde{\mu}(x).$$

Note that since $I\times J$ is compact the norms of $A_{x,y}$ and $B_{x,y}$ are bounded over $(x,y)\in I\times J$, and since $\varepsilon$ may be small and $m$ large relative to these norms, we have not bothered to re-scale the measures $A_{x,y}\widetilde{\mu}^{x,i}, B_{x,y}\widetilde{\nu}^{y,i}$ to compensate for their contraction/expansion (the distortion caused by these matrices is also one reason for the dependence of the parameters on $f$, the other being the speed of linear approximation). The proof is given in Section 5.6.

We note two important special cases.

**Corollary 2.15.** *Let $G < GL_d(\mathbb{R}) \subseteq \mathbb{R}^{d^2}$ be a matrix group acting by left multiplication on $\mathbb{R}^d$. Let $\nu \in \mathcal{P}(G)$ and $\mu \in \mathcal{P}(\mathbb{R}^d)$ be measures of bounded support. Then for every $\varepsilon > 0$ and $m \in \mathbb{N}$ there is a $\delta = \delta(\nu, \mu, \varepsilon, m) > 0$, such that for $k > k(\nu, \mu, \varepsilon, m, \delta)$ and $n > n(\nu, \mu, \varepsilon, m, \delta, k)$, either*

$$H_n(\nu . \mu) > H_n(\mu) + \delta,$$

*or else, for independently chosen level-$k$ components $\widetilde{\mu}, \widetilde{\nu}$ of $\mu, \nu$, respectively, with probability at least $1 - \varepsilon$ there are subspaces $V_0, \ldots, V_n < \mathbb{R}^d$ such that*

$$\mathbb{P}_{0 \le i \le n}\left( \begin{array}{c} y . \widetilde{\mu}^{x,i} \text{ is } (V_i, \varepsilon, m)\text{-saturated and} \\ \widetilde{\nu}^{y,i} . x) \text{ is } (V_i, \varepsilon)\text{-concentrated} \end{array} \right) > 1 - \varepsilon$$

*(The dependence of $\delta, k, \nu$ on the measures depends only on their support and is uniform on compact sets).*

*If in addition $\mu$ is $((\varepsilon/3d)^{d+1}, \sigma)$-non-affine for some $\sigma > 0$, then for $\delta, k, n$ which also depend on $\sigma$, we also have*

$$\frac{1}{n+1} \sum_{i=0}^{n} \dim V_i > c \cdot H_n(\widetilde{\nu}) - \varepsilon.$$

*for a constant $c$ depending only on $d, \sigma$ and the support of $\nu$.*

**Corollary 2.16.** *Let $G < GL_d(\mathbb{R}) \subseteq \mathbb{R}^{d^2}$ be a matrix group and $\mu, \nu \in \mathcal{P}(G)$ measures of bounded support. Then for every $\varepsilon > 0$ and $m \in \mathbb{N}$ there is a $\delta > 0$ such that for every large enough $k$ and all suitably large enough $n$, either*

$$H_n(\mu * \nu) > H_n(\mu) + \delta,$$

*or else, for an independently chosen pair of raw level-$k$ components $\widetilde{\mu}, \widetilde{\nu}$ of $\mu, \nu$, respectively, with probability $> 1 - \varepsilon$, there are subspaces $V_0, \ldots, V_n < \mathbb{R}^{d^2}$ such that*

$$\mathbb{P}_{0 \le i \le n}\left( \begin{array}{c} y * \widetilde{\mu}^{x,i} \text{ is } (V_i, \varepsilon, m)\text{-saturated and} \\ \widetilde{\nu}^{y,i} * x \text{ is } (V_i, \varepsilon)\text{-concentrated} \end{array} \right) > 1 - \varepsilon$$

*and*

$$\frac{1}{n+1} \sum_{i=0}^{n} \dim V_i > c \cdot H_n(\nu) - \varepsilon.$$

*for a suitable constant $c$.*

Both corollaries follow from the the previous theorem by taking $f(x, y) = yx$ to be the appropriate action map; for the first corollary an additional argument is needed to produce the constant $c$. The dependence of the paramerets on the measures is only through their supports: If we fix a large ball in advance and assume the measures are supported on it, then the parameters depend only on the ball, not the measures.

*Remark* 2.17.

1. It is important to note the order of quantifiers in the theorem and corollaries: In the theorem all parameters depend on the function $f$, and in the corollaries the function $f$ is the action map restricted to the (compact) product of the supports of $\nu$ and $\mu$, which are fixed before the other parameters. The reason this works is that once the functions is fixed and the measures are fixed, and compactly supported, the speed with which the function $f$ approaches its linearzation is uniform, hence, at small enough scales, we are essentially dealing with linear convolutions rather than a non-linear image.

2. In some applications the order of quantifiers above is not sufficient and it is necessary to obtain statements that are uniform over many functions or independent of the support of the measures. Then a more quantitative analysis is needed. Such an example can be found in [14].

3. One can formulate the corollaries in abstract Lie groups using partitions introduced from local coordinates, or using general theorem on the existence of similar partitions in doubling metric spaces, see e.g. [17].

4. When dealing with more general group actions one would also like to relax the condition that the measures be compactly supported. But in doing so one must take into account how various properties of the action affect the dependence between parameters in the theorem. For example they are sensitive to the speed at which the action approaches its linearization (which may differ from point to point), how well the an element of the group is determined by its action on $k$-tuples, and how sensitive the latter procedure is to changes in the $k$-tuple. It turns out that the cleanest approach is to choose a left-invariant Riemmanian metric on the group and dyadic partition adapted to it. For a detailed development of this approach in one example we refer the reader to [14].

# 3 Entropy, concentration, uniformity and saturation

This section presents without proof some standard results about entropy, followed by a more detailed analysis of concentration, saturation and uniformity.

## 3.1 Preliminaries on entropy

The Shannon entropy of a probability measure $\mu$ with respect to a countable partition $\mathcal{E}$ is given by

$$H(\mu,\mathcal{E}) = -\sum_{E\in\mathcal{E}} \mu(E)\log\mu(E),$$

where the logarithm is in base 2 and $0\log 0 = 0$. The conditional entropy with respect to a countable partition $\mathcal{F}$ is

$$H(\mu,\mathcal{E}|\mathcal{F}) = \sum_{F\in\mathcal{F}} \mu(F)\cdot H(\mu_F,\mathcal{E}),$$

where $\mu_F = \frac{1}{\mu(F)}\mu|_F$ is the conditional measure on $F$. For a discrete probability measure $\mu$ we write $H(\mu)$ for the entropy with respect to the partition into points, and for a probability vector $\alpha = (\alpha_1, \ldots, \alpha_k)$ we write

$$H(\alpha) = -\sum \alpha_i \log \alpha_i.$$

and for $0 < \varepsilon < 1$ we abbreviate

$$H(\varepsilon) = H((\varepsilon, 1-\varepsilon))$$

Note that if $0 < \varepsilon < 1/2$ then $H(\varepsilon) = O(\varepsilon \log(1/\varepsilon))$.

We collect here some standard properties of entropy.

**Lemma 3.1.** *Let $\mu, \nu$ be probability measures on a common space, $\mathcal{E}, \mathcal{F}$ partitions of the underlying space and $\alpha \in [0,1]$.*

1. *$H(\mu, \mathcal{E}) \geq 0$, with equality if and only if $\mu$ is supported on a single atom of $\mathcal{E}$.*
2. *If $\mu$ is supported on $k$ atoms of $\mathcal{E}$ then $H(\mu, \mathcal{E}) \leq \log k$, with equality if and only if each of these atoms has mass $1/k$.*
3. *If $\mathcal{F}$ refines $\mathcal{E}$ (i.e. $\forall\, F \in \mathcal{F}\ \exists E \in \mathcal{E}$ s.t. $F \subseteq E$) then $H(\mu, \mathcal{F}) \geq H(\mu, \mathcal{E})$.*
4. *If $\mathcal{E} \vee \mathcal{F} = \{E \cap F \,:\, E \in \mathcal{E}\,,\, F \in \mathcal{F}\}$ denotes the join of $\mathcal{E}$ and $\mathcal{F}$, then*

$$H(\mu, \mathcal{E} \vee \mathcal{F}) = H(\mu, \mathcal{F}) + H(\mu, \mathcal{E}|\mathcal{F}),$$

*in particular*

$$H(\mu, \mathcal{E} \vee \mathcal{F}) \leq H(\mu, \mathcal{E}) + H(\mu, \mathcal{F}).$$

5. *$H(\cdot, \mathcal{E})$ and $H(\cdot, \mathcal{E}|\mathcal{F})$ are concave.*
6. *$H(\cdot, \mathcal{E})$ obeys the "convexity" bound*

$$H(\sum \alpha_i \mu_i, \mathcal{E}) \leq \sum \alpha_i H(\mu_i, \mathcal{E}) + H(\alpha).$$

*and similarly after conditioning on $\mathcal{F}$.*

In particular, we note that for $\mu \in \mathcal{P}([0,1)^d)$ we have the bounds $H(\mu, \mathcal{D}_m) \leq md$ and $H(\mu, \mathcal{D}_{n+m}|\mathcal{D}_n) \leq md$.

Although the function $(\mu, m) \mapsto H(\mu, \mathcal{D}_m)$ is not continuous in the weak-* topology on measures, the following estimates provide usable substitutes.

**Lemma 3.2.** *Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, let $\mathcal{E}, \mathcal{F}$ be partitions of $\mathbb{R}^d$, and $m, m' \in \mathbb{N}$.*

1. *Given a compact $K \subseteq \mathbb{R}^d$ and $\mu \in \mathcal{P}(K)$, there is a neighborhood $U \subseteq \mathcal{P}(K)$ of $\mu$ such that $|H(\nu, \mathcal{D}_m) - H(\mu, \mathcal{D}_m)| = O_d(1)$ for $\nu \in U$.*
2. *If each $E \in \mathcal{E}$ intersects at most $k$ elements of $\mathcal{F}$ and vice versa, then $|H(\mu, \mathcal{E}) - H(\mu, \mathcal{F})| = O(\log k)$.*
3. *If $f, g : \mathbb{R}^d \to \mathbb{R}^k$ and $\|f(x) - g(x)\| \leq C2^{-m}$ for $x \in \mathbb{R}^d$ then $|H(f\mu, \mathcal{D}_m) - H(g\mu, \mathcal{D}_m)| \leq O_{C,k}(1)$.*

4. *If* $\nu(\cdot)=\mu(\cdot+x_0)$ *then* $|H(\mu,\mathcal{D}_m)-H(\nu,\mathcal{D}_m)|=O_d(1)$.

5. *If* $|m'-m|\leq C$*, then* $|H(\mu,\mathcal{D}_m)-H(\mu,\mathcal{D}_{m'})|\leq O_{C,d}(1)$.

We will use some easy corollaries of Lemma 3.1 (5) and (6).

**Lemma 3.3.** *Let* $\mu,\nu\in\mathcal{P}([-r,r]^d)$*, let* $\delta>0$*, and let* $\theta=(1-\delta)\mu+\delta\nu$*. Then for partitions* $\mathcal{A},\mathcal{B}$ *of* $\mathbb{R}^d$ *we have*

$$\begin{aligned}|H(\theta,\mathcal{A})-H(\mu,\mathcal{A})| &\leq H(\delta)+\delta|H(\mu,\mathcal{A})-H(\nu,\mathcal{A})|,\\ |H(\theta,\mathcal{A}|\mathcal{B})-H(\mu,\mathcal{A}|\mathcal{B})| &\leq H(\delta)+\delta|H(\mu,\mathcal{A}|\mathcal{B})-H(\nu,\mathcal{A}|\mathcal{B})|.\end{aligned}$$

Recall that the total variation distance between $\mu,\nu\in\mathcal{P}(\mathbb{R}^d)$ is

$$\|\mu-\nu\|=\sup_A|\mu(A)-\nu(A)|,$$

where the supremum is over Borel sets $A$. This is a complete metric on $\mathcal{P}(\mathbb{R}^d)$. It follows from standard measure theory that given $\mu,\nu$ there are probability measures $\tau,\mu',\nu'$ such that $\mu=(1-\delta)\tau+\delta\mu'$ and $\nu=(1-\delta)\tau+\delta\nu'$, where $\delta=\frac{1}{2}\|\mu-\nu\|$. Combining this with Lemma 3.1 (5) and (6), we have

**Lemma 3.4.** *If* $\mathcal{A},\mathcal{B}$ *are partitions of* $\mathbb{R}^d$*, and if* $\mu,\nu\in\mathcal{P}(\mathbb{R}^d)$ *are supported on at most* $k$ *atoms of each partition and* $\|\mu-\nu\|<\varepsilon$*, then*

$$\begin{aligned}|H(\mu,\mathcal{A})-H(\nu,\mathcal{A})| &< 2k\varepsilon+2H(\frac{1}{2}\varepsilon).\\ |H(\mu,\mathcal{A}|\mathcal{B})-H(\nu,\mathcal{A}|\mathcal{B})| &< 2k\varepsilon+2H(\frac{1}{2}\varepsilon).\end{aligned}$$

*In particular, if* $\mu,\nu\in\mathcal{P}([0,1)^d)$*, then*

$$|H_m(\mu)-H_m(\nu)|<2d\varepsilon+\frac{2H(\frac{1}{2}\varepsilon)}{m}.$$

## 3.2 Global entropy from local entropy

Recall from Section 2.3 the definition of the raw and re-scaled components $\mu_{x,n}$, $\mu^{x,n}$, and note that

$$H(\mu^{x,n},\mathcal{D}_m)=H(\mu_{x,n},\mathcal{D}_{n+m}).$$

Also,

$$\begin{aligned}\mathbb{E}_{i=n}\left(H_m(\mu^{x,i})\right) &= \int\frac{1}{m}H(\mu^{x,n},\mathcal{D}_m)\,d\mu(x)\\ &= \frac{1}{m}\int H(\mu_{x,n},\mathcal{D}_{n+m})\,d\mu(x)\\ &= \frac{1}{m}H(\mu,\mathcal{D}_{n+m}\,|\,\mathcal{D}_n).\end{aligned}$$

The following basic lemmas enable us to get bounds on the scale-$n$ entropy of a measure, or a convolution of measures, in terms of the average scale-$m$ entropy of their components or convolution of their components, when $m\ll n$.

**Lemma 3.5.** *For $r \geq 1$, $\mu \in \mathcal{P}([-r,r]^d)$ and integers $m < n$,*

$$H_n(\mu) \quad = \quad \mathbb{E}_{0\leq i\leq n}\left(H_m(\mu^{x,i})\right) + O(\frac{m+\log r}{n}).$$

**Lemma 3.6.** *For $r > 0$, $\mu, \nu \in \mathcal{P}([-r,r]^d)$ and integers $m < n$,*

$$\begin{aligned} H_n(\mu * \nu) &\geq \mathbb{E}_{0\leq i\leq n}\left(\frac{1}{m}H(\mu_{x,i} * \nu_{y,i}, \mathcal{D}_{i+m}|\mathcal{D}_i)\right) + O(\frac{m+\log r}{n}) \\ &\geq \mathbb{E}_{0\leq i\leq n}\left(H_m(\mu^{x,i} * \nu^{y,i})\right) + O(\frac{1}{m} + \frac{m+\log r}{n}). \end{aligned}$$

For proofs see [12, Section 3.2], or the proof of the following variant, which is essentially the same as the Euclidean case.

**Lemma 3.7.** *Let $\nu \in \mathcal{P}(G_0)$ and $\mu \in \mathcal{P}(\mathbb{R}^d)$ be supported on sets of diameter $r$. Then for $m < n$,*

$$H_n(\nu . \mu) \geq \mathbb{E}_{0\leq i\leq n}\left(H_{i,m}(\nu_{g,i} . \mu)\right) - O(\frac{1}{m} + \frac{m+\log r}{n}).$$

*Proof.* We can assume that $n = n_0 m$, since replacing $n$ by the closest multiple of $m$ results in a change of $O(m/n)$ to $H_n(\nu . \mu)$, which is absorbed in the error term. Let us also introduce a parameter $0 \leq k < m$. Then

$$\begin{aligned} H_n(\nu . \mu) &= \frac{1}{n}H(\nu . \mu, \mathcal{D}_n) \\ &= \frac{1}{n}H(\nu . \mu, \mathcal{D}_{k+n}) + O(\frac{k}{n}) \\ &= \frac{1}{n}H(\nu . \mu, \mathcal{D}_k) + \frac{1}{n}H(\nu . \mu, \mathcal{D}_{k+n}|\mathcal{D}_k) + O(\frac{m}{n}) \end{aligned}$$

Since $\nu$ is supported on a set of diameter $O(1)$ and $\mu$ on a set of diameter $O(r)$, also $\nu . \mu$ is supported on a set of diameter $O(r)$, so the trivial entropy bound gives

$$\frac{1}{n}H(\nu . \mu, \mathcal{D}_k) = O(\frac{\log r + m}{n})$$

We next evaluate $\frac{1}{n}H(\nu . \mu, \mathcal{D}_{k+n}|\mathcal{D}_k)$. Recalling our assumption $n = n_0 m$ and the definition of conditional entropy, we have

$$\frac{1}{n}H(\nu . \mu, \mathcal{D}_{k+n}|\mathcal{D}_k) = \frac{1}{n}\sum_{j=0}^{n_0-1} H(\nu . \mu, \mathcal{D}_{k+(j+1)m}|\mathcal{D}_{k+jm})$$

For each $j$ we have the identities $\nu = \mathbb{E}_{i=j}(\nu_{g,i})$ and $\mu = \mathbb{E}_{i=j}(\mu_{x,i})$, which implies $\nu . \mu = \mathbb{E}_{i=j}(\nu_{g,i} . \mu)$. By concavity of entropy, we get

$$\begin{aligned} \frac{1}{n}\sum_{j=1}^{n_0} H(\nu . \mu, \mathcal{D}_{k+jm}|\mathcal{D}_{k+(j-1)m}) &= \frac{1}{n}\sum_{j=0}^{n_{0-1}} H(\mathbb{E}_{i=k+jm}(\nu_{g,i} . \mu), \mathcal{D}_{k+(j+1)m}|\mathcal{D}_{k+jm}) \\ &\geq \frac{1}{n}\sum_{j=1}^{n_0} \mathbb{E}_{i=k+jm}\left(H(\nu_{g,i} . \mu, \mathcal{D}_{k+(j+1)m}|\mathcal{D}_{k+jm})\right) \\ &= \frac{1}{n}\sum_{j=1}^{n_0} \mathbb{E}_{i=k+jm}\left(H(\nu_{g,i} . \mu, \mathcal{D}_{k+(j+1)m}) - H(\nu_{g,i} . \mu, \mathcal{D}_{k+jm})\right) \end{aligned}$$

Since $\nu_{g,i}\bm{.}\mu$ is supported on a set of diameter $O(2^{-i})$, for $i = k + jm$ we have $H(\nu_{g,i}\bm{.}\mu\mathcal{D}_{k+jm}) = O(1)$. Thus the total sum of error terms in the sum above is $O(n_0)$, which upon dividing by $n$ is $O(n_0/n) = O(1/m)$. The discussion so far shows that

$$\frac{1}{n}H(\nu\bm{.}\mu, \mathcal{D}_n) \geq \frac{1}{n}\sum_{j=1}^{n_0}\mathbb{E}_{i=k+jm}\left(H(\nu_{g,i}\bm{.}\mu, \mathcal{D}_{k+(j+1)m})\right) - O(\frac{1}{m} + \frac{m+\log r}{n})$$

Averaging now over $k = 0, \ldots, m$ gives

$$\begin{aligned}
\frac{1}{n}H(\nu\bm{.}\mu, \mathcal{D}_n) &= \frac{1}{m}\sum_{k=0}^{m-1}\frac{1}{n}H(\nu\bm{.}\mu, \mathcal{D}_{k+n}) - O(\frac{m}{n}) \\
&\geq \frac{1}{m}\sum_{k=0}^{m-1}\frac{1}{n}\sum_{j=1}^{n_0}\mathbb{E}_{i=k+jm}\left(H(\nu_{g,i}\bm{.}\mu, \mathcal{D}_{k+(j+1)m})\right) - O(\frac{1}{n} + \frac{m+\log r}{n}) \\
&= \frac{1}{n}\sum_{j=1}^{n}\frac{1}{m}\mathbb{E}_{i=j}\left(H(\nu_{g,i}\bm{.}\mu, \mathcal{D}_{i+m})\right) - O(\frac{1}{n} + \frac{m+\log r}{n}) \\
&= \mathbb{E}_{1\leq i\leq n}\left(\frac{1}{m}H(\nu_{g,i}\bm{.}\mu, \mathcal{D}_{i+m})\right) - O(\frac{1}{n} + \frac{n+\log r+k}{n})
\end{aligned}$$

as claimed. $\square$

We also need the following variant of Lemma 3.7:

**Lemma 3.8.** *Let $\nu \in \mathcal{P}(G_0)$ and $\mu \in \mathcal{P}(\mathbb{R}^d)$ be supported on balls of diameter $r$. Then for every $k, n \in \mathbb{N}$,*

$$H_n(\nu\bm{.}\mu) \geq \mathbb{E}_{i=k}(H_n(\nu_{g,i}\bm{.}\mu)) + O_{R,k}(\frac{1}{n})$$

*and in particular*

$$\mathbb{E}_{i=k}(H_n(\nu_{g,i}\bm{.}\mu) - H_n(\mu_{x,i})) \leq H_n(\nu\bm{.}\mu) - H_n(\mu) + O_{R,k}(\frac{1}{n})$$

*Proof.* Since $\mu, \nu$ are supported on balls of radius $R$, so is $\nu\bm{.}\mu$, so the scale-$k$ entropies of all these measures is $O_{R,k}(1)$. It follows that

$$H_n(\nu\bm{.}\mu) = \frac{1}{n}H(\nu\bm{.}\mu, \mathcal{D}_n|\mathcal{D}_k) + O_{R,k}(\frac{1}{n})$$

By concavity of conditional entropy,

$$\begin{aligned}
\frac{1}{n}H(\nu\bm{.}\mu, \mathcal{D}_n|\mathcal{D}_k) &= \frac{1}{n}H(\mathbb{E}_{i=k}(\nu_{g,i}\bm{.}\mu), \mathcal{D}_n|\mathcal{D}_k) \\
&\geq \mathbb{E}_{i=k}(\frac{1}{n}H(\nu_{g,i}\bm{.}\mu, \mathcal{D}_n|\mathcal{D}_k))
\end{aligned}$$

But $\nu_{g,i}\bm{.}\mu$ is supported on a set of diameter $O(2^{-i})$, so (taking $i = k$),

$$\begin{aligned}
\frac{1}{n}H(\nu_{g,k}\bm{.}\mu, \mathcal{D}_n|\mathcal{D}_k) &= \frac{1}{n}H(\nu_{g,k}\bm{.}\mu, \mathcal{D}_n) + O(\frac{1}{n}) \\
&= H_n(\nu_{g,k}\bm{.}\mu) + O(\frac{1}{n})
\end{aligned}$$

Combining the last three equations gives the first claim. For the second claim, note that we have

$$\begin{aligned} H_n(\mu) &= \frac{1}{n}H(\mu,\mathcal{D}_n|\mathcal{D}_k)+O_{R,k}(\frac{1}{n}) \\ &= \frac{1}{n}\mathbb{E}_{i=k}(H(\mu_{x,i},\mathcal{D}_n))+O_{R,k}(\frac{1}{n}) \\ &= \mathbb{E}_{i=k}(H_n(\mu_{x,i}))+O_{R,k}(\frac{1}{n}) \end{aligned}$$

(the first inequality again because $\mu$ is supported on a set of diameter $O(R)$). Subtracting this expression for $H_n(\mu)$ from the previous one for $H_n(\nu\boldsymbol{.}\mu)$ gives the claim. $\square$

### 3.3 First lemmas on concentration, uniformity, saturation

We consider here some basic connections between uniform, concentrated and saturated measures. We make the statements as general as possible, but in some cases, especially when dealing with uniform measures, it is necessary to assume that the support of the measures is bounded, and the constants in the error terms may depend on the diameter of the support. Since we are interested in the asymptotics as $m\to\infty$ we rarely make the dependence explicit, but it can be read off from the proofs.

Given partitions $\mathcal{E}$ and $\mathcal{F}$ of sets $X,Y$, respectively, write

$$\mathcal{E}\otimes\mathcal{F}=\{E\times F\,:\,E\in\mathcal{E}\,,\,F\in\mathcal{F}\}$$

for the product partition of $X\times Y$. We will also often identify $\mathcal{E}$ with the partition $\mathcal{E}\otimes\{Y\}$ of $X\times Y$, and similarly $\mathcal{F}$ with the partition $\{X\}\otimes\mathcal{F}$ of $X\times Y$.

For a linear subspace $V\leq\mathbb{R}^d$ we write $\mathcal{D}_n^V$ for the level-$n$ dyadic partition on $V$ with respect to some fixed (but arbitrary) orthogonal coordinate system in $V$, which we usually do not specify.

Let $V\leq\mathbb{R}^d$ be a linear subspace and $W=V^\perp$, and let $\mathcal{D}'_m=\mathcal{D}_m^V\otimes\mathcal{D}_m^W$ denote the product partition of $\mathbb{R}^d\cong V\times W$. Each element of $\mathcal{D}_m$ intersects at most $O(1)$ elements of $\mathcal{D}'_m$, and vice versa, so by Lemma 3.2 (2),

$$|H(\mu,\mathcal{D}_m)-H(\mu,\mathcal{D}'_m)|=O(1).$$

The same is true for the induced partitions on $W$, so, writing $\pi_W$ for the orthogonal projection to $W$,

$$|H(\pi_W\mu,\mathcal{D}_m)-H(\pi_W\mu,\mathcal{D}'_m)|=O(1)$$

and also

$$|H(\pi_W\mu,\mathcal{D}_m)-H(\pi_W\mu,\mathcal{D}_m^W)|=O(1).$$

Recall that we identify $\mathcal{D}_m^V,\mathcal{D}_m^W$ with the partitions $\pi_V^{-1}\mathcal{D}_m^V$, $\pi_W^{-1}\mathcal{D}_m^W$ of $\mathbb{R}^d$, respectively. With this identification we have $\mathcal{D}'_m=\mathcal{D}_m^V\vee\mathcal{D}_m^W$, and

$$H(\pi_W\mu,\mathcal{D}'_m)=H(\mu,\mathcal{D}_m^W).$$

From this discussion we have the following immediate consequence:

**Lemma 3.9.** *With the above notation, a measure $\mu \in \mathcal{P}(\mathbb{R}^d)$ is $(V, \varepsilon + O(1/m), m)$-saturated if and only if*

$$\frac{1}{m} H(\mu, \mathcal{D}_m^V | \mathcal{D}_m^{V^\perp}) \geq \dim V - (\varepsilon + O(\frac{1}{m})).$$

From similar considerations we have

**Lemma 3.10.** *If $\mu \in \mathcal{P}(\mathbb{R}^d)$ is $(V, \varepsilon, m)$-saturated and $g = 2^t U + a \in G$ is a similarity, then $g\mu$ is $(UV, \varepsilon + O(|t|/m), m)$-saturated; and similarly for uniformity.*

One way to get saturated measures is from uniform measures:

**Lemma 3.11.** *If $\mu \in \mathcal{P}([-r, r]^d)$ is $(V, \varepsilon, m)$-uniform then it is $(V, O_r(\varepsilon + 1/m), m)$-saturated.*

*Proof.* By uniformity, we can write $\mu = (1-\varepsilon)\mu' + \varepsilon\mu''$ , where $\mu'$ is supported on the $2^{-m}$-neighborhood of a translate of $V$. By concavity of conditional entropy,

$$\begin{aligned} H(\mu, \mathcal{D}_m | \mathcal{D}_m^{V^\perp}) &\geq (1-\varepsilon) H(\mu', \mathcal{D}_m | \mathcal{D}_m^{V^\perp}) \\ &\geq H(\mu', \mathcal{D}_m | \mathcal{D}_m^{V^\perp}) - \varepsilon H(\mu', \mathcal{D}_m). \end{aligned}$$

Since $\mu$, and hence $\mu'$, is supported on at most $O(r^d \cdot 2^m)$ atoms of $\mathcal{D}_m$, we have $H(\mu', \mathcal{D}_m) = O(m \log r)$, and the inequality above becomes

$$H(\mu, \mathcal{D}_m | \mathcal{D}_m^{V^\perp}) \geq H(\mu', \mathcal{D}_m | \mathcal{D}_m^{V^\perp}) - \varepsilon O(m \log r).$$

Since $\mu'$ is supported on a $2^{-m}$-neighborhood of a translate of $V$, it is supported on $O(1)$ atoms of $\mathcal{D}_m^{V^\perp}$, so $H(\mu', \mathcal{D}_m^{V^\perp}) = O(1)$, hence

$$\begin{aligned} H(\mu', \mathcal{D}_m | \mathcal{D}_m^{V^\perp}) &\geq H(\mu', \mathcal{D}_m) - H(\mu', \mathcal{D}_m^{V^\perp}) \\ &\geq H(\mu', \mathcal{D}_m) - O(1). \end{aligned}$$

Finally, by Lemma 3.3 applied to $\mu = (1-\varepsilon)\mu + \varepsilon\mu''$, and using the bound $O(r^d 2^m)$ on the number of $\mathcal{D}_m$-atoms supporting $\mu', \mu''$ and uniformity of $\mu$,

$$\begin{aligned} H(\mu', \mathcal{D}_m) &> H(\mu) - \varepsilon(m + O(\log r)) - H(\varepsilon) \\ &> m \dim V - \varepsilon(m + \log r) - H(\varepsilon) \end{aligned}$$

Putting it all together, using $H(\varepsilon) \leq 1$ and dividing by $m$ gives the claim. □

Another way to get saturated measures is to take convex combinations of saturated measures:

**Lemma 3.12.** *A convex combination of $(V, \varepsilon, m)$-saturated measures on $\mathbb{R}^d$ is $(V, \varepsilon + O(1/m), m)$-saturated.*

*Proof.* Immediate from Lemma 3.9 and concavity of conditional entropy (Lemma 3.1 (5)). □

Combining the two lemmas above gives the following:

**Corollary 3.13.** *A convex combination of $(V, \varepsilon, m)$-uniform measures on $[-r, r]^d$ is $(V, O_r(\varepsilon + 1/m), m)$-saturated.*

Saturation is also stable under small perturbations in the total variation metric:

**Lemma 3.14.** *Let $\mu,\nu \in \mathcal{P}([-r,r]^d)$. If $\mu$ is $(V,\varepsilon,m)$-saturated and $\|\mu-\nu\| < \delta$ then $\nu$ is $(V,\varepsilon',m)$-saturated for $\varepsilon' = \varepsilon + O(\delta\log r + 1/m)$.*

*Proof.* Take $\mathcal{A} = \mathcal{D}_m^V \vee \mathcal{D}_m^{V^\perp}$ and $\mathcal{B} = \mathcal{D}_m^{V^\perp}$ in Lemma 3.4, and use Lemma 3.9. ☐

Finally, we shall need an entropy bound for concentrated measures.

**Lemma 3.15.** *If $\mu \in \mathcal{P}([-r,r]^d)$ is $(V,2^{-m})$-concentrated then $H_m(\mu) \le \dim V + O_r(\frac{\log m}{m})$.*

*Proof.* Write $\mu = (1-2^{-m})\mu_1 + 2^{-m}\mu_2$ where $\mu_1 \in \mathcal{P}(W^{(2^{-m})})$ for some translate $W$ of $V$ and $\mu_2 \in \mathcal{P}([-r,r]^d)$. Since $H_m(\mu_i) = O_r(1)$ for $i=1,2$, by Lemma 3.3 it suffices for us to show that $H_m(\mu_1) \le \dim V + O_r(1/m)$. This again follows from the fact that $W^{(2^{-m})} \cap [-r,r]^d$ intersects $O(r^d 2^m)$ atoms of $\mathcal{D}_m$ and the trivial entropy bound. ☐

## 3.4 Concentration and saturation of components

In this section all measures are supported on $[0,1)^d$.

**Lemma 3.16.** *If $\mu \in \mathcal{P}([0,1)^d)$ is $(V,\varepsilon,n)$-saturated, then for every $1 \le m < n$,*

$$\mathbb{P}_{0\le i\le n}\left(\mu^{x,i} \text{ is } (V,\varepsilon',m)\text{-saturated}\right) > 1-\varepsilon',$$

*where $\varepsilon' = \sqrt{d\varepsilon + O(\frac{m}{n})}$.*

*Proof.* Without loss of generality, we may assume that $\mathcal{D}_n = \mathcal{D}_n^V \vee \mathcal{D}_n^W$ where $W = V^\perp$ (Lemma 3.9). By the fact that $\mu$ is $(V,\varepsilon,n)$-saturated and by Lemma 3.5, we have

$$\begin{aligned}
\dim V &+ H_n(\pi_W\mu) - \varepsilon \le \\
&\le H_n(\mu) \\
&= \mathbb{E}_{0\le i\le n}\left(H_m(\mu^{x,i})\right) + O(\frac{m}{n}) \\
&= \mathbb{E}_{0\le i\le n}\left(H_m(\mu^{x,i},\mathcal{D}_m^W)\right) + \mathbb{E}_{0\le i\le n}\left(\frac{1}{m}H\left(\mu^{x,i},\mathcal{D}_m|\mathcal{D}_m^W\right)\right) + O(\frac{m}{n}) \\
&= \mathbb{E}_{0\le i\le n}\left(H_m(\pi_W(\mu^{x,i}))\right) + \mathbb{E}_{0\le i\le n}\left(\frac{1}{m}H\left(\mu^{x,i},\mathcal{D}_m|\mathcal{D}_m^W\right)\right) + O(\frac{m}{n}).
\end{aligned}$$

Since $(\pi_W\mu)_{y,i}$ is the convex combination (using the natural weights) of $\pi_W(\mu_D)$ over those $D \in \mathcal{D}_i$ with $D \cap \pi_W^{-1}(y) \ne \emptyset$ (recall that we are assuming $\mathcal{D}_n = \mathcal{D}_n^V \vee \mathcal{D}_n^W$), concavity of entropy implies

$$\begin{aligned}
H_n(\pi_W\mu) &= \mathbb{E}_{0\le i\le n}\left(H_m((\pi_W\mu)^{y,i})\right) + O(\frac{m}{n}) \\
&\ge \mathbb{E}_{0\le i\le n}\left(H_m(\pi_W(\mu^{x,i})\right) + O(\frac{m}{n}).
\end{aligned}$$

Combining these we have

$$\mathbb{E}_{0\leq i\leq n}\left(\frac{1}{m}H\left(\mu^{x,i},\mathcal{D}_m|\mathcal{D}_m^W\right)\right)\geq \dim V-(\varepsilon+O(\frac{m}{n})).$$

But we also have the trivial bound $\frac{1}{m}H(\mu^{x,i},\mathcal{D}_m|\mathcal{D}_m^W)\leq \dim V\leq d$. Combining this and the last inequality, the lemma follows by Markov's inequality. □

The analogous statement for concentration is valid at individual scales (rather than for typical scales between 0 and $n$, as above):

**Lemma 3.17.** *If $\mu\in\mathcal{P}([0,1)^d)$ is $(V,\varepsilon)$-concentrated and $1\leq m\leq\log(1/\varepsilon)$, then*

$$\mathbb{P}_{i=m}\left(\mu^{x,i}\text{ is }(V,\sqrt{2^m\varepsilon})\text{-concentrated}\right)>1-\sqrt{2^{-m}\varepsilon}.$$

*Proof.* Let $W=V+v_0$ be such that $\mu(W^{(\varepsilon)})>1-\varepsilon$. For a dyadic cube $D$ write $T_D$ for the surjective homothety $D\to[0,1)^d$ and let $W^D=T_D(W)$. Clearly, for any $D\in\mathcal{D}_m$ we have $T_D(W^{(\varepsilon)})=(W^D)^{(2^m\varepsilon)}$. Take $\delta=\sqrt{2^m\varepsilon}\leq 1$ and let $\mathcal{E}\subseteq\mathcal{D}_m$ denote the family of cells $D$ such that

$$\mu_D(D\setminus W^{(\varepsilon)})=\mu^D([0,1]^d\setminus(W^D)^{(2^m\varepsilon)})>\delta.$$

It follows that

$$\varepsilon\geq\mu([0,1]^d\setminus W^{(\varepsilon)})\geq\sum_{D\in\mathcal{E}}\mu(D\setminus W^{(\varepsilon)})>\delta\cdot\mu(\cup\mathcal{E}),$$

so $\mu(\cup\mathcal{E})<\varepsilon/\delta=\sqrt{2^{-m}\varepsilon}$. Hence $\mu(\cup(\mathcal{D}_m\setminus\mathcal{E}))>1-\sqrt{2^{-m}\varepsilon}$, and the conclusion follows. □

We often will want to change the scale at which measures are saturated. Clearly if $\delta<\varepsilon$ and $\mu$ is $(V,\delta)$-concentrated, then it is also $(V,\varepsilon)$-concentrated. However for $\delta<\varepsilon$ and $k>m$ it is in general not true that if $\mu$ is $(V,\delta,k)$-saturated then $\mu$ is also $(V,\varepsilon,m)$-saturated (though of course it certainly is $(V,\varepsilon,k)$-saturated). The issue is that the first few scales do not greatly affect the entropy at a fine scale. In order to allow such change of parameters we will pass to components, using the lemmas above. We will also need a simple covering argument for intervals of $\mathbb{Z}$:

**Lemma 3.18.** *Let $I\subseteq\{0,\ldots,n\}$ and $m\in\mathbb{N}$ be given. Then there is a subset $I'\subseteq I$ such that $I\subseteq I'+[0,m]$ and $[i,i+m]\cap[j,j+m]=\emptyset$ for distinct $i,j\in I'$.*

*Proof.* Define $I'$ inductively. Begin with $I'=\emptyset$ and, at each successive stage, if $I\setminus\bigcup_{i\in I'}[i,i+m]\neq\emptyset$ then add its least element to $I'$. Stop when $I\subseteq\bigcup_{i\in I'}[i,i+m]$. □

**Proposition 3.19.** *For every $\varepsilon>0$ and $m\in\mathbb{N}$, if $k>k(\varepsilon,m)$ and $0<\delta<\delta(\varepsilon,m,k)$, then for all large enough $n>n(\varepsilon,m,k,\delta)$, the following holds. Let $\nu,\mu\in\mathcal{P}(\mathbb{R}^d)$ and let $V_0,V_1,\ldots,V_n\leq\mathbb{R}^d$ be linear subspaces such that*

$$\mathbb{P}_{0\leq i\leq n}\left(\begin{array}{c}\mu^{x,i}\text{ is }(V_i,\delta,k)\text{-saturated and}\\ \nu^{y,i}\text{ is }(V_i,\delta)\text{-concentrated}\end{array}\right)>1-\delta. \qquad (18)$$

*Then there are linear subspaces $V_0',\ldots,V_n' \leq \mathbb{R}^d$ such that*

$$\mathbb{P}_{0\leq i\leq n}\left(\begin{array}{c}\mu^{x,i} \text{ is } (V_i',\varepsilon,m)\text{-saturated and} \\ \nu^{y,i} \text{ is } (V_i',\varepsilon)\text{-concentrated}\end{array}\right) > 1-\varepsilon. \tag{19}$$

*Furthermore if $V_i = V$ is independent of $i$ then we can take $V_i' = V$.*

*Proof.* Fix $\delta,k$ and suppose that (18) holds for some $n$. Let $I\subseteq\{0,\ldots,n\}$ denote the set of indices $u$ such that

$$\mathbb{P}_{i=u}\left(\begin{array}{c}\mu^{x,i} \text{ is } (V_i,\delta,k)\text{-saturated and} \\ \nu^{y,i} \text{ is } (V_i,\delta)\text{-concentrated}\end{array}\right) > 1-\sqrt{\delta}.$$

By Markov's inequality,

$$|I| \geq (1-\sqrt{\delta})(n+1)$$

Let $I'\subseteq I$ be chosen as in the previous lemma with parameter $k$, so $I\subseteq I'+[0,k]$ and $[i,i+k]\cap[j,j+k]=\emptyset$ for distinct $i,j\in I'$. If $j=i+u$ for some $i\in I'$ and $0\leq u\leq k$, define $V_j'=V_i$. Define $V_j'$ arbitrarily for other $j$. Note that when $V_i=V$ is independent of $i$ then also $V_j'=V$ for $j$ as above, in which case we can set $V_i'=V$ for all $i$ and satisfy the last assertion in the statement.

To see that this choice works (assuming the parameters satisfy the proper relations), note that for any pair of components $\theta=\mu^{x,i},\eta=\nu^{y,i}$ such that $\theta$ is $(V_i,\delta,k)$-saturated and $\eta$ is $(V_i,\delta)$-concentrated, we have by Lemmas 3.16 and 3.17 that

$$\begin{aligned}\mathbb{P}_{i\leq j\leq i+k}(\theta^{w,j} \text{ is } (V_j',\sqrt{d\delta+O(\tfrac{m}{k})},m)\text{-saturated}) &> 1-\sqrt{d\delta+O(\tfrac{m}{k})}\\ \mathbb{P}_{i\leq j\leq i+k}(\eta^{z,j} \text{ is } (V_j',\sqrt{2^k\delta})\text{-concentrated}) &> 1-\sqrt{2^{-k}\delta}.\end{aligned}$$

so

$$\mathbb{P}_{i\leq j\leq i+k}\left(\begin{array}{c}\theta^{w,j} \text{ is } (V_j',\sqrt{d\delta+O(\frac{m}{k})},m)\text{-saturated and} \\ \eta^{z,j} \text{ is } (V_j',\sqrt{2^k\delta})\text{-concentrated}\end{array}\right) > 1-O(\sqrt{\delta+\frac{m}{k}}).$$

Write $U=\bigcup_{i\in I'}[i,i+k]$. The union is disjoint by assumption, so the bounds above combine to give

$$\mathbb{P}_{i\in U}\left(\begin{array}{c}\theta^{w,i} \text{ is } (V_j',\sqrt{d\delta+O(\frac{m}{k})},m)\text{-saturated and} \\ \eta^{zij} \text{ is } (V_j',\sqrt{2^k\delta})\text{-concentrated}\end{array}\right) > 1-O(\sqrt{\delta+\frac{m}{k}}).$$

Let $V=U\cap[0,n]$. Then we have the trivial inequalities

$$\begin{aligned}\mathbb{P}_{i\in V}(\ldots) &\geq \mathbb{P}_{i\in U}(\ldots)-\frac{|U\setminus V|}{|U|}\\ \mathbb{P}_{0\leq i\leq n}(\ldots) &\geq \frac{|V|}{n+1}\mathbb{P}_{i\in U}(\ldots).\end{aligned}$$

Since $I\subseteq U\subseteq[0,n+k]$ to we have $|U\setminus V|\leq k$ and $|U|\geq(1-\sqrt{\delta})(n+1)$, so combining the identities above with the previous inequality we get

$$\mathbb{P}_{0\leq i\leq n}\left(\begin{array}{c}\theta^{w,i} \text{ is } (V_j',\sqrt{d\delta+O(\frac{m}{k})},m)\text{-saturated and} \\ \eta^{zij} \text{ is } (V_j',\sqrt{2^k\delta})\text{-concentrated}\end{array}\right) > 1-O(\sqrt{\delta+\frac{m}{k}})-O(\frac{k}{n}).$$

Thus if $k$ is large enough relative to $\varepsilon,m$; $\delta$ is small enough relative to $\varepsilon,k$; and $n$ is large enough relative to $\varepsilon,k$, we obtain (19). ☐

We remark that the use of Lemma 3.18 and Markov's inequality in the proof is rather crude, and one might want to use Lemma 2.7 instead. This would have shown that one can associate to most components a subspace on which it is suitably concentrated and saturated, but the subspaces would generally depend on the component, and not just on the level it belongs to. The argument above gives the desired uniformity across each level.

## 3.5 The space of subspaces

Let $B_r(x)$ denote the open Euclidean ball of radius $r$ around $x \in \mathbb{R}^d$, and, as before, for $A \subseteq \mathbb{R}^d$ let $A^{(\varepsilon)} = \{x \in \mathbb{R}^d : d(x, A) < \varepsilon\}$. Define a metric on the space of linear subspaces $V, W \leq \mathbb{R}^d$ by

$$d(V, W) = \inf\{\varepsilon > 0 \,:\, V \cap B_1(0) \subseteq W^{(\varepsilon)} \text{ and } W \cap B_1(0) \subseteq V^{(\varepsilon)}\} \qquad (20)$$

This is just the Hausdorff metric on the intersections of $V, W$ with the closed unit ball, so the induced topology on the space of linear subspaces of $\mathbb{R}^d$ is compact (note that it decomposes into $d+1$ connected components, corresponding to the dimensions of the subspaces). It is also the same as the distance $\|\pi_V - \pi_W\|$, where $\|\cdot\|$ denotes the operator norm and $\pi_V, \pi_W$ the orthogonal projections.

It will be convenient to write

$$A \sqsubseteq A' \qquad \text{if} \qquad A \cap B_1(0) \subseteq A'.$$

This a transitive, reflexive relation. In this notation, the distance between subspaces $V, W \leq \mathbb{R}^d$ defined above is

$$d(V_1, V_2) = \inf\{\varepsilon > 0 \,:\, V_1 \sqsubseteq V_2^{(\varepsilon)} \text{ and } V_2 \sqsubseteq V_1^{(\varepsilon)}\}.$$

Define the "angle" between subspaces $V_1, V_2$ by $\angle(V_1, V_2) = 0$ if $V_1 \subseteq V_2$ or $V_2 \subseteq V_1$; otherwise set $W = V_1 \cap V_2$ and

$$\angle(V_1, V_2) = \inf\{\|v_1 - v_2\| \,:\, v_1 \in V_1 \cap W^{\perp}\,,\ v_2 \in V_2 \cap W^{\perp}\,,\ \|v_1\| = \|v_2\| = 1\}.$$

This is not the usual notion of angle, but it agrees with the usual definition up to a multiplicative constant, and is more convenient to work with.

The following properties are elementary and we omit their proof.

**Lemma 3.20.** *Let $V, W \leq \mathbb{R}^d$ be linear subspaces and $\varepsilon > 0$.*

1. *$d(V, W) \leq 1$ with equality if and only if $V \cap W^{\perp} \neq \{0\}$ or $W \cap V^{\perp} \neq \{0\}$. In particular if $\dim W > \dim V$ then $W \not\sqsubseteq V^{(1)}$ and $d(V, W) = 1$.*
2. *If $0 < \varepsilon < 1$ and $V \sqsubseteq W^{(\varepsilon)}$ then $\pi_W : V \to W$ is injective, $\dim V \leq \dim W$, and if $\dim V = \dim W$ then $W \sqsubseteq V^{(\varepsilon)}$ and $d(V, W) \leq \varepsilon$.*
3. *$\angle(V, W) \leq \sqrt{2} \cdot d(V, W)$.*
4. *If $V \not\sqsubseteq W^{(\varepsilon)}$ then there exists a vector $v \in V$ with $\angle(\mathbb{R}v, W) \geq \varepsilon$.*

We collect some elementary implications for concentration, uniformity and saturation:

**Lemma 3.21.** *Let $\mu \in \mathcal{P}([0,1)^d)$ and $V, W \leq \mathbb{R}^d$.*

1. *If $\mu$ is $(V,\varepsilon)$ concentrated and $d(W,V)<\delta$, then $\mu$ is $(W,\varepsilon+\sqrt{d}\delta)$-concentrated.*

2. *If $\mu$ is $(V,\varepsilon,m+1)$-uniform and $d(W,V)<\frac{1}{\sqrt{d}}2^{-(m+1)}$, then $\mu$ is $(W,\varepsilon,m)$-uniform.*

3. *If $\mu$ is $(V,\varepsilon,m)$-saturated and $d(W,V)<2^{-m}$, then $\mu$ is $(W,\varepsilon+O(1/m),m)$-saturated.*

4. *If $\mu$ is $(V,\varepsilon,m)$-saturated and $W\leq V$ is a subspace then $\mu$ is $(W,\varepsilon+O(1/m),m)$-saturated.*

5. *If $\mu\in\mathcal{P}([0,1)^d)$ is both $(V_1,\varepsilon,m)$, and $(V_2,\varepsilon,m)$-saturated, and $\angle(V_1,V_2)>\delta>0$, then $\mu$ is $(V_1+V_2,\varepsilon',m)$-saturated, where $\varepsilon'=2\varepsilon+O(\frac{1}{m}\log(\frac{1}{\delta}))$.*

*Proof.* If $d(W,V)<\delta$ then $V\cap B_1(0)\subseteq W^{(\delta)}\cap B_1(0)$, so $V\cap B_{\sqrt{d}}(0)\subseteq W^{(\sqrt{d}\delta)}\cap B_{\sqrt{d}}(0)$. It follows that if $(V+v)\cap[0,1)^d\neq\emptyset$ then $V+v\cap[0,1)^d\subseteq(W+v)^{(\sqrt{d}\cdot\delta)}$ (we use the fact that the diameter of $[0,1)^d$ is $\sqrt{d}$), so $(V^{(\varepsilon)}+v)\cap[0,1)^d\subseteq(W+v)^{(\varepsilon+\sqrt{d}\delta)}$. The first claim follows.

For (2), observe that if $d(W,V)<2^{-(m+1)}$ and $\mu$ is $(V,2^{-(m+1)}/\sqrt{d})$-concentrated, then by the first claim, $\mu$ is $(W,2^{-m})$-concentrated. Since by assumption $H_m(\mu,\mathcal{D}_m)>\dim V-\varepsilon$, and $d(V,W)<2^{-(m+1)}$ implies $\dim W=\dim V$, we have shown that $\mu$ is $(V,\varepsilon,m)$-uniform.

For (3), note that $d(V,W)<2^{-m}$ implies that $\|\pi_{V^\perp}-\pi_{W^\perp}\|<2^{-m}$, so $|H(\mu,\mathcal{D}_m|\mathcal{D}_m^{V^\perp})-H(\mu,\mathcal{D}_m|\mathcal{D}_m^{W^\perp})|=O(1)$, and the claim follows.

For (4), we may assume $W\neq V$. Let $W'<V$ denote the orthogonal complement of $W$ in $V$ and write $\mathbb{R}^d$ as the orthogonal direct sum $W\oplus W'\oplus V^\perp$. Without loss of generality we may assume $\mathcal{D}_m=\mathcal{D}_m^W\vee\mathcal{D}_m^{W'}\vee\mathcal{D}_m^{V^\perp}$ (Lemma 3.9); by doing so we implicitly increased $\varepsilon$ by $O(1/m)$. Since $\mu$ is $(V,\varepsilon,m)$-saturated,

$$\frac{1}{m}H(\mu,\mathcal{D}_m|\mathcal{D}_m^{V^\perp})\geq\dim V-\varepsilon.$$

Since $\mathcal{D}_m$ refines $\mathcal{D}_m^{W^\perp}=\mathcal{D}_m^{W'\oplus V^\perp}$ which in turn refines $\mathcal{D}_m^{V^\perp}$, we have

$$\begin{aligned}H(\mu,\mathcal{D}_m|D_m^{V^\perp}) &= H(\mu,\mathcal{D}_m\vee\mathcal{D}_m^{W'}|\mathcal{D}_m^{V^\perp})\\ &= H(\mu,\mathcal{D}_m^{W'}|\mathcal{D}_m^{V^\perp})+H(\mu,\mathcal{D}_m|\mathcal{D}_m^{W^\perp}).\end{aligned}$$

Inserting this into the inequality above gives

$$\frac{1}{m}H(\mu,\mathcal{D}_m|\mathcal{D}_m^{W^\perp})\geq\dim V-H(\mu,\mathcal{D}_m^{W'}|\mathcal{D}_m^{V^\perp})-\varepsilon.$$

Since $\frac{1}{m}H(\mu,\mathcal{D}_m^{W'}|\mathcal{D}_m^{V^\perp})\leq\dim W'+O(1/m)=\dim V-\dim W+O(1/m)$, this is precisely $(W,\varepsilon+O(1/m),m)$-saturation of $\mu$.

We turn to (5). Let $V_2'=V_2\cap(V_1\cap V_2)^\perp$, so that $V_2'<V_2$, $V_1\cap V_2'=\{0\}$, $\angle(V_1,V_2')=\angle(V_1,V_2)>\delta$ and $V_1+V_2'=V_1+V_2$. By (4) we can replace $V_2$ by $V_2'$ at the cost of increasing $\varepsilon$ by $O(1/m)$. Thus, we may assume from the start that $V_1\cap V_2=\{0\}$.

Write $V=V_1\oplus V_2$ (this is an algebraic, not an orthogonal, sum) and $W=V^\perp$. We can assume without loss of generality that $\mathcal{D}_m=\mathcal{D}_m^V\vee\mathcal{D}_m^W$. Also let $\mathcal{E}_m = \mathcal{D}_m^{V_1} \vee \mathcal{D}_m^{V_2} \vee \mathcal{D}_m^W$ be the partition corresponding to the direct sum $\mathbb{R}^d = V_1 \oplus V_2 \oplus W$.

By Lemma 3.9, we must show that

$$H_m(\mu, \mathcal{D}_m^V | \mathcal{D}_m^W) \geq \dim V - 2\varepsilon - O(\frac{\log(1/\delta)}{m}).$$

Because of the assumption $\angle(V_1, V_2) > \delta$, the partitions of $\mathcal{D}_m^{V_1} \vee \mathcal{D}_m^{V_2}$ and $\mathcal{D}_m^V$ of $V$, and also the corresponding partitions of $\mathbb{R}^d$, have the property that each atom of one intersects $O(1/\delta)$ atoms of the other. Thus

$$\left| H_m(\mu, \mathcal{D}_m^{V_1} \vee \mathcal{D}_m^{V_2} | \mathcal{D}_m^W) - H(\mu, \mathcal{D}_m^V | \mathcal{D}_m^W) \right| = O(\log(1/\delta)),$$

so it is sufficient for us to prove that

$$H_m(\mu, \mathcal{D}_m^{V_1} \vee \mathcal{D}_m^{V_2} | \mathcal{D}_m^W) \geq \dim V - 2\varepsilon - O(\frac{\log(1/\delta)}{m}). \tag{21}$$

Now,

$$\frac{1}{m} H(\mu, \mathcal{D}_m^{V_1} \vee \mathcal{D}_m^{V_2} | \mathcal{D}_m^W) = \frac{1}{m} H(\mu, \mathcal{D}_m^{V_1} | \mathcal{D}_m^W) + \frac{1}{m} H(\mu, \mathcal{D}_m^{V_2} | \mathcal{D}_m^{V_1} \vee \mathcal{D}_m^W). \tag{22}$$

Since $W \subseteq V_1^{\perp}$, we can assume that the partition $\mathcal{D}_m^{V_1^{\perp}}$ refines $\mathcal{D}_m^W$. Using the fact that $\mu$ is $(V_1, \varepsilon, m_1)$-saturated, we get a bound for the first term on the right hand side of the above identity:

$$\frac{1}{m} H(\mu, \mathcal{D}_m^{V_1} | \mathcal{D}_m^W) \geq \frac{1}{m} H(\mu, \mathcal{D}_m^{V_1} | \mathcal{D}_m^{V_1^{\perp}}) \geq \dim V_1 - \varepsilon.$$

As for the second term, again using the fact that each atom of $\mathcal{D}_m^{V_1} \vee \mathcal{D}_m^{V_2} \vee \mathcal{D}_m^W$ intersects $O(1/\delta)$ atoms of $\mathcal{D}_m^{V_2} \vee \mathcal{D}_m^{V_2^{\perp}}$ and vice versa, and similarly for $\mathcal{D}_m^{V_1} \vee \mathcal{D}_m^W$ and $\mathcal{D}_m^{V_2^{\perp}}$, we have

$$\begin{aligned} \frac{1}{m} H(\mu, \mathcal{D}_m^{V_2} | \mathcal{D}_m^{V_1} \vee \mathcal{D}_m^W) &= \frac{1}{m} H(\mu, \mathcal{D}_m^{V_2} | \mathcal{D}^{V_2^{\perp}}) - O(\frac{\log(1/\delta)}{m}) \\ &\geq \dim V_2 - \varepsilon - O(\frac{\log(1/\delta)}{m}). \end{aligned}$$

Combining the last two inequalities and (22) gives the desired inequality (21) $\square$

## 3.6 Geometry of thickened subspaces

In this section we develop some methods for understanding unions and intersections of thickened subspaces. We require some elementary linear algebra estimates.

**Lemma 3.22.** *Let $v_1, \ldots, v_k \in \mathbb{R}^d$ with $\|v_i\| \leq 1$, and suppose that*

$$d(v_i, \operatorname{span}\{v_1, \ldots, v_{i-1}\}) > \delta \qquad \text{for all } 1 \leq i \leq k.$$

*Then for any $v = \sum t_i v_i$ we have $\|(t_1, \ldots, t_k)\| \leq \sqrt{k} \cdot 2^k \|v\| / \delta^k$.*

*Proof.* We first claim for every $1 \leq i \leq k$ that

$$|t_i| \leq (1 + \frac{1}{\delta})^{k-i+1} \|v\|$$

This we show by induction on $k$. For $k = 1$ it is trivial. In general set $V_i = \text{span}\{v_1, \ldots, v_i\}$ and $W_i = V_i^{\perp}$. By hypothesis, $\left\|\pi_{W_{i-1}}(v_i)\right\| > \delta$ for all $1 \leq i \leq k$. Thus

$$\|v\| \geq \left\|\pi_{W_{k-1}}(v)\right\| = |t_k| \cdot \left\|\pi_{W_{k-1}}(v_k)\right\| > t_k \delta$$

so $|t_k| < \|v\| / \delta$, and the claim holds for $i = k$. Now,

$$\begin{aligned}
\left\|\sum_{i=1}^{k-1} t_i v_i\right\| &= \left\|\pi_{V_{k-1}}(v) - t_k \pi_{V_{k-1}}(v_k)\right\| \\
&\leq \|v\| + |t_k| \\
&\leq (1 + \frac{1}{\delta}) \|v\|
\end{aligned}$$

Thus, by the induction hypothesis, for $1 \leq i \leq k-1$,

$$\begin{aligned}
|t_i| &\leq (1 + \frac{1}{\delta})^{(k-1)-i+1} \left\|\sum_{i=1}^{k-1} t_i v_i\right\| \\
&\leq (1 + \frac{1}{\delta})^{k-i+1} \|v\|
\end{aligned}$$

as claimed. It remains to note that

$$\|t\| \leq \sqrt{k} \|t\|_\infty \leq \sqrt{k}(1 + \frac{1}{\delta})^k \|v\|$$

The claim follows (note that $\delta < 1$). □

It will be convenient to introduce notation for the constant

$$p_k = k \cdot 2^k$$

**Corollary 3.23.** *Let $V \leq \mathbb{R}^d$ be a subspace and $v_1, \ldots, v_k \in V^{(\varepsilon)}$ with $\|v_i\| \leq 1$. If $d(v_i, \text{span}\{v_1, \ldots, v_{i-1}\}) > \delta$ for all $1 \leq i \leq k$, then $\text{span}\{v_1, \ldots, v_k\} \sqsubseteq V^{(p_k \varepsilon / \delta^k)}$.*

*Proof.* Write $W = \text{span}\{v_i\}$ and let $w = \sum t_i v_i \in W$ be a unit vector. Write $t = (t_1, \ldots, t_k)$. Then by the last lemma, $\|t\|_2 \leq 2^k \sqrt{k} / \delta^k$. Thus

$$\begin{aligned}
d(w, V) &\leq \sum |t_i| \cdot d(v_i, V) \\
&< \varepsilon \cdot \|t\|_1 \\
&\leq \varepsilon \cdot \sqrt{k} \cdot \|t\|_2 \\
&\leq p_k \cdot \frac{\varepsilon}{\delta^k},
\end{aligned}$$

where we used the hypothesis and the general inequality $\|u\|_1 \leq \sqrt{k} \|u\|_2$. □

**Corollary 3.24.** *Suppose that $E, V \leq \mathbb{R}^d$ are subspaces such that $E \sqsubseteq V^{(\varepsilon)}$, and $e \in V^{(\varepsilon)} \cap B_1(0)$ is such that $d(e, E) > \delta > 0$. Then $E' = E \oplus \mathbb{R}e$ satisfies $E' \sqsubseteq V^{(8\varepsilon/\delta^2)}$.*

*Proof.* Every vector in $E'$ belongs to a subspace of the form $\mathbb{R}e' \oplus \mathbb{R}e$ for some $e' \in E$, so it is enough to show $\mathbb{R}e' \oplus \mathbb{R}e \sqsubseteq V^{(8\varepsilon/\delta^2)}$. But the pair $e', e$ satisfies the assumptions of the previous corollary with $k = 2$. Since $p_2 = 8$, the claim follows. $\square$

**Corollary 3.25.** *Suppose that $E, V, W \leq \mathbb{R}^d$ are subspaces such that $E \sqsubseteq V^{(\varepsilon)} \cap W^{(\varepsilon)}$, and $e \in (V^{(\varepsilon)} \cap W^{(\varepsilon)}) \cap B_1(0)$ is such that $d(e, E) > \delta > 0$. Let $E' = E \oplus \mathbb{R}e$. Then $E' \sqsubseteq V^{(8\varepsilon/\delta^2)} \cap W^{(8\varepsilon/\delta^2)}$.*

*Proof.* Immediate from the lemma. $\square$

Proposition 3.27 below takes a family $\mathcal{W}$ of subspaces and finds an essentially minimal subspace that almost-contains all $W \in \mathcal{W}$. The basic step in the proof is to do this for two subspaces, and this is given by the next corollary.

**Corollary 3.26.** *Given $\varepsilon > 0$ let $\varepsilon_k = 4\varepsilon^{1/3^k}$. Then for any $V, W \leq \mathbb{R}^d$, there is a $0 \leq k \leq d$ and a $k$-dimensional subspace $E \sqsubseteq V^{(\varepsilon_k)} \cap W^{(\varepsilon_k)}$ such that $V^{(\varepsilon_k)} \cap W^{(\varepsilon_k)} \sqsubseteq E^{(\varepsilon_{k+1})}$.*

*Proof.* Let $E$ be a subspace of maximal dimension satisfying $E \sqsubseteq V^{(\varepsilon_{\dim E})} \cap W^{(\varepsilon_{\dim E})}$ (such subspaces exist, e.g. $\{0\}$). Let $k = \dim E$. If $V^{(\varepsilon_k)} \cap W^{(\varepsilon_k)} \not\sqsubseteq E^{(\varepsilon_{k+1})}$ then by the previous corollary we can replace $E$ by $E' = E + \mathbb{R}e$ for some $e \in \partial B_1(0) \cap (V^{(\varepsilon_k)} \cap W^{(\varepsilon_k)} \setminus E^{(\varepsilon_{k+1})})$ and $E'$ will satisfy

$$E' \sqsubseteq V^{(8\varepsilon_k/\varepsilon_{k+1}^2)} \cap W^{(8\varepsilon_k/\varepsilon_{k+1}^2)} \subseteq V^{(\varepsilon_{k+1})} \cap W^{(\varepsilon_{k+1})}$$

where we have used

$$\frac{8\varepsilon_k}{\varepsilon_{k+1}^2} = \frac{8 \cdot 4\varepsilon^{1/3^k}}{4^2\varepsilon^{2/3^{k+1}}} = 2 \cdot \varepsilon^{1/3^{k+1}} = \frac{1}{2}\varepsilon_{k+1}$$

But $\dim E' = \dim E + 1$, which contradicts the maximality of $E$. $\square$

**Proposition 3.27.** *Let $\varepsilon > 0$ and $\varepsilon_k = 4\varepsilon^{1/3^k}$. Then for any family $\mathcal{W}$ of subspaces of $\mathbb{R}^d$, there is a subspace $V \leq \mathbb{R}^d$ such that $W \sqsubseteq V^{(\varepsilon_d)}$ for all $W \in \mathcal{W}$, and if $\widetilde{V}$ is a subspace such that $W \sqsubseteq \widetilde{V}^{(\varepsilon)}$ for all $W \in \mathcal{W}$, then $V \sqsubseteq \widetilde{V}^{(\varepsilon_d)}$.*

*Proof.* We may assume that $\varepsilon_d < 1$ since otherwise the statement is trivial (any subspace $V$ will do). Let $V$ be a subspace of minimal dimension such that $W \sqsubseteq V^{(\varepsilon_{d-\dim V})}$ for all $W \in \mathcal{W}$ (such subspaces exist, e.g. $V = \mathbb{R}^d$). Write $k = d - \dim V$. We can assume $k < d$ since the case $k = d$ corresponds to $V = \{0\}$, and then the conclusion is trivial.

We claim that $V$ is the desired subspace. First, $\varepsilon_k \leq \varepsilon_d$, so we have $W \sqsubseteq V^{(\varepsilon_k)} \sqsubseteq V^{(\varepsilon_d)}$ for all $W \in \mathcal{W}$, which is the first property.

For the second property of $V$, suppose that there is a subspace $\widetilde{V} \leq \mathbb{R}^d$ such that $W \sqsubseteq \widetilde{V}^{(\varepsilon)} \sqsubseteq \widetilde{V}^{(\varepsilon_k)}$ for $W \in \mathcal{W}$, but such that $V \not\sqsubseteq \widetilde{V}^{(\varepsilon_d)}$. Let $E$ be a subspace of maximal dimension satisfying $E \sqsubseteq V^{(\varepsilon_{k+1})} \cap \widetilde{V}^{(\varepsilon_{k+1})}$. Clearly $\dim E \leq \dim V$ (since $E \sqsubseteq V^{(\varepsilon_{k+1})}$ and $\varepsilon_{k+1} \leq \varepsilon_d < 1$), and we cannot have $\dim E = \dim V$ because then we would have $V \sqsubseteq E^{(\varepsilon_k)} \sqsubseteq \widetilde{V}^{(\varepsilon_k + \varepsilon_{k+1})} \sqsubseteq \widetilde{V}^{(\varepsilon_d)}$, contrary to assumption. So $\dim E < \dim V$. Thus, by the definition of $V$, there exists a $W \in \mathcal{W}$ with $W \not\sqsubseteq E^{(\varepsilon_{k+1})}$. Choose a vector $e \in (B_1(0) \cap W) \setminus E^{(\varepsilon_{k+1})}$, so that $d(e, E) \geq \varepsilon_{k+1}$, and note that since $W \sqsubseteq V^{(\varepsilon_k)} \cap \widetilde{V}^{(\varepsilon_k)}$ we also have $e \in V^{(\varepsilon_k)} \cap \widetilde{V}^{(\varepsilon_k)}$. Thus, by Corollary 3.25 (with $\varepsilon_k$ and $\varepsilon_{k+1}$ in the role of $\varepsilon, \delta$), the subspace $E' = E \oplus \mathbb{R}e$ satisfies $E' \sqsubseteq V^{(\varepsilon_{k+1})} \cap \widetilde{V}^{(\varepsilon_{k+1})}$. But $\dim E' > \dim E$, which contradicts the definition of $E$. We conclude that $V \subseteq \widetilde{V}^{(\varepsilon_d)}$, as desired. □

We note that the proof actually shows $V \sqsubseteq W^{(\varepsilon_{d-\dim V})}$ for all $W \in \mathcal{W}$ and that any $\widetilde{V}$ with this property satisfies $V \sqsubseteq \widetilde{V}^{(\varepsilon_{d-\dim V+1})}$.

From the last proposition we can derive a dual version: for any family $\mathcal{W}$ of subspaces and any $\varepsilon > 0$, there is a subspace $V$ such that $V \sqsubseteq W^{(\varepsilon_d)}$ for all $W \in \mathcal{W}$ and any other subspace $\widetilde{V}$ with this property satisfies $\widetilde{V} \sqsubseteq V^{(\varepsilon_d)}$. To see this, observe that $U_1 \sqsubseteq U_2^{(\varepsilon)}$ if and only if $U_2^{\perp} \sqsubseteq (U_2^{\perp})^{(\varepsilon)}$, and apply the previous proposition to $\mathcal{W}^{\perp} = \{W^{\perp} : W \in \mathcal{W}\}$. However, in a later application we will want to present the subspace $V$ as an intersection of a small number of (neighborhoods of) subspaces from $\mathcal{W}$. This is provided for in the following proposition.

**Proposition 3.28.** *Let $\varepsilon > 0$ and $\delta = 8^{d-1}\varepsilon^{1/3^{d^2}}$. Then for any family $\mathcal{W}$ of subspaces of $\mathbb{R}^d$, there is a subspace $V \leq \mathbb{R}^d$ such that $V \sqsubseteq W^{(\delta)}$ for every $W \in \mathcal{W}$, and subspaces $W_1, \ldots, W_k \in \mathcal{W}$ with $k \leq d - \dim V$ such that $\bigcap_{i=1}^{k} W_i^{(\varepsilon)} \sqsubseteq V^{(\delta)}$. In particular, if $V'$ is any other subspace satisfying $V' \sqsubseteq W^{(\varepsilon)}$ for every $W \in \mathcal{W}$, then $V' \sqsubseteq V^{(\delta)}$.*

*Furthermore, if we are given an increasing sequence $\mathcal{W}^1 \subseteq \mathcal{W}^2 \subseteq \ldots$ with each $\mathcal{W}^i$ a family of subspaces of $\mathbb{R}^d$, then we can assign $V^i$ to $\mathcal{W}^i$ as above in such a way that $V^{i+1} \sqsubseteq (V^i)^{(\delta)}$.*

*Proof.* Fix $\varepsilon, \delta, \mathcal{W}$ as in the statement. We shall recursively choose finite sequences of subspaces $W_1, W_2, \ldots \in \mathcal{W}$ and $V_0, V_1, \ldots \leq \mathbb{R}^d$, and of real numbers $\delta_0, \delta_1, \ldots > 0$, such that $\bigcap_{j=1}^{i} W_j^{(\varepsilon)} \sqsubseteq V_i^{(\delta_i)}$.

Begin with $V_0 = \mathbb{R}^d$ and $\delta_0 = \varepsilon$. Now for $j \geq 1$ suppose we have defined $V_i, W_i, \delta_i$ for $i < j$. Let $\delta_j^* = 8(\delta_{j-1})^{1/3^d}$. If $V_{j-1} \sqsubseteq W^{(\delta_j^*)}$ for all $W \in \mathcal{W}$, we terminate the construction. Otherwise, choose $W_j \in \mathcal{W}$ such that $V_{j-1} \not\sqsubseteq W_j^{(\delta_j^*)}$. Apply Corollary 3.26 to the subspaces $V_{j-1}$, $W_j$ with the parameter $\delta_{j-1}$. We obtain a subspace $V_j \leq \mathbb{R}^d$ and real numbers $\delta_{j-1} \leq \delta_j' \leq \delta_j \leq 4(\delta_{j-1})^{1/3^d}$ satisfying

$$V_j \sqsubseteq V_{j-1}^{(\delta_j')} \cap W_j^{(\delta_j')} \tag{23}$$

and

$$V_{j-1}^{(\delta_j')} \cap W_j^{(\delta_j')} \sqsubseteq V_j^{(\delta_j)}$$

(in the notation of the corollary, $\delta_j' = \varepsilon_k$ and $\delta_j = \varepsilon_{k+1}$, but if $k = d$ we can take $\delta_j' = \delta_j = \varepsilon_d$). Since $\varepsilon \leq \delta_{j-1} \leq \delta_j'$ and, by the induction hypothesis, $\bigcap_{i=0}^{j-1} W_j^{(\varepsilon)} \sqsubseteq V_{j-1}^{(\delta_{j-1})}$, the last equation implies that $\bigcap_{i=0}^{j} W_j^{(\varepsilon)} \sqsubseteq V_j^{(\delta_j)}$, and the conditions of the construction are satisfied.

We now claim that $\dim V_j < \dim V_{j-1}$ as long as they are defined. Indeed, suppose the construction completed the $j$-th step of the construction without terminating, so $V_{j-1} \not\sqsubseteq W_j^{(\delta_j^*)}$. In particular this means that $\delta_j \leq \delta_j^* < 1$. Now, we know that $V_j \sqsubseteq V_{j-1}^{(\delta_j)}$, which together with $\delta_j < 1$ implies $\dim V_j \leq \dim V_{j-1}$. Suppose that equality held. Then, again using $\delta_j < 1$, we would have the reverse containment $V_{j-1} \sqsubseteq V_j^{(\delta_j)}$. This, together with $V_j \sqsubseteq W_j^{(\delta'_j)}$ and $\delta'_j \leq \delta_j$, implies $V_{j-1} \sqsubseteq W_j^{(2\delta_j)}$. Since $2\delta_j \leq \delta^*_j$, this contradicts the assumption $V_{j-1} \not\sqsubseteq W_j^{(\delta^*_j)}$, so we must have $\dim V_j < \dim V_{j-1}$.

Since $\dim V_j$ is strictly decreasing, the procedure terminates after completing some $k \leq d$ iterations, which in our numbering means it completed step $k-1$ and terminated at step $k$. This means that $V_{k-1} \sqsubseteq W^{(\delta^*_k)}$ for all $W \in \mathcal{W}$ and $\bigcap_{i=0}^{k-1} W_i^{(\varepsilon)} \sqsubseteq V_k^{(\delta_{k-1})}$. Observe that

$$\delta_{k-1} \leq \delta^*_{k-1} < 8^{k-1}(\delta_0)^{1/(k-1)d} \leq \delta$$

(since $\delta_0 = \varepsilon$). Hence for $V = V_k$ we have $W \sqsubseteq V^{(\delta)}$ for all $W \in \mathcal{W}$ and $\bigcup_{i=1}^{k} W_i^{(\varepsilon)} \sqsubseteq V^{(\delta)}$, as desired.

The statement about $V'$ is immediate from the first statement of the lemma.

Finally, for the last part, we note that in the construction we may first exhaust the subspaces in $\mathcal{W}^1$, obtaining $V^1$, then move on to those in $\mathcal{W}^2$ obtaining possibly a different $V^2$, etc. The containment relation follows from (23). □

## 3.7 Minimally concentrated and maximally saturated subspaces

Our goal in this section is to identify, given a measure and associated parameters, a subspace $V$ on which it is in a sense most concentrated, and one on which it is most saturated, relative to the parameters. By this we mean that if $\widetilde{V}$ is another subspace for which the measure is concentrated or saturated, relative to comparable parameters, then $\widetilde{V}$ is, respectively, essentially contained in, or essentially contains, $V$.

The existence of a "minimal" subspace on which a given measure concentrates is proved by a variation on the argument in Proposition 3.27:

**Proposition 3.29.** *Let $\varepsilon > 0$ and $\varepsilon_k = 4\varepsilon^{1/3^k}$, and assume that $\varepsilon_d < 1/2$. Then for any $\eta \in \mathcal{P}([0,1)^d)$, there is a subspace $V \leq \mathbb{R}^d$ such that $\eta$ is $(V, \sqrt{d} \cdot \varepsilon_d)$-concentrated, and if $W$ is any subspace such that $\eta$ is $(W, \varepsilon)$-concentrated, then $V \sqsubseteq W^{(\varepsilon_d)}$.*

*Proof.* We can assume $\varepsilon_d < 1$. Choose a subspace $V \leq \mathbb{R}^d$ of minimal dimension such that $\eta$ is $(V, \sqrt{d} \cdot \varepsilon_{d-\dim V})$-concentrated (the family of such subspaces is non-empty, e.g. $V = \mathbb{R}^d$). Write $k = d - \dim V$ note that we can assume $k < d$ since otherwise $V = \{0\}$ and the claim is trivial.

We claim that $V$ is the desired subspace. Suppose that $\eta$ is $(W, \varepsilon)$-concentrated (and hence $(W, \varepsilon_k)$-concentrated) but that $V \not\sqsubseteq W^{(\varepsilon_d)}$. Let $E \sqsubseteq V^{(\varepsilon_{k+1})} \cap W^{(\varepsilon_{k+1})}$ be a subspace of maximal dimension. Then $\dim E < \dim V$ so by the definition of $V$ the measure $\eta$ is not $(E, \sqrt{d} \cdot \varepsilon_{k+1})$-concentrated. Now, consider translates of $V^{(\varepsilon_k)} + v$ and $W^{(\varepsilon_k)} + w$ which cover all but $\varepsilon_k$ and $\varepsilon$ of the mass of $\eta$, respectively. Choose $u \in (V^{(\varepsilon_k)} + v) \cap (W^{(\varepsilon_k)} + w)$ (the intersection is non-empty because it has $\eta$-mass at least $1 - 2\varepsilon_k > 0$), and observe that $V^{(\varepsilon_k)} + v \subseteq V^{(2\varepsilon_k)} + u$ and $W^{(\varepsilon_k)} + w \subseteq W^{(2\varepsilon_k)} + u$. Hence

$$\eta\left([0,1]^d \cap (V^{(2\varepsilon_k)} \cap W^{(2\varepsilon_k)} + u\right) > 1 - 2\varepsilon_k.$$

Now consider the translate $[0,1]^d \cap (E^{(\sqrt{d}\varepsilon_k)} + u)$. It covers at most $1 - \varepsilon_{k+1}$ of the mass of $\eta$, which, since since $2\varepsilon_k < \varepsilon_{k+1}$, is less than the mass of the previous intersection. Thus, translating back to the origin and scaling by $1/\sqrt{d}$ (so that $[0,1]^d + u$ is mapped into the unit ball), we find that there exists a point $e \in (B_1(0) \cap V^{(2\varepsilon_k/\sqrt{d})} \cap W^{(2\varepsilon_k/\sqrt{d})}) \setminus E^{(\varepsilon_k)}$. By Corollary 3.25 the subspace $E' = E + \mathbb{R}e$ satisfies $E' \sqsubseteq V^{(\varepsilon_{k+1})} \cap W^{(\varepsilon_{k+1})}$ (we have used that $8 \cdot 2\varepsilon_k/(\sqrt{d}\varepsilon_{k+1}^2) < \varepsilon_{k+1}$). But $\dim E' > \dim E$, contradicting the choice of $E$. We conclude therefore $V \sqsubseteq W^{(\varepsilon_d)}$, as desired. □

We turn to the analog of Proposition 3.29, which provides a "maximal" subspace on which a given measure is saturated to a certain degree. The argument is again similar to the measureless case.

**Proposition 3.30.** *Given $m \in \mathbb{N}$ and $\theta \in \mathcal{P}([0,1)^d d)$, there is a subspace $V \leq \mathbb{R}^d$ such that $\theta$ is $(V, O(\frac{\log m}{m}), m)$-saturated, and if $W$ is any subspace such that $\theta$ is $(W, \frac{1}{m}, m)$-saturated, then $W \sqsubseteq V^{(O((\log m)/m))}$.*

*Proof.* Write $\delta_k = C2^k k \log(m)/m$ where $C > 1$ is large enough to serve as the implicit a constant in the big-$O$ expressions we invoke below. Note that $\delta_k < \delta_{k+1}$ and $\delta_d = O_d(\frac{\log m}{m})$. Let $V$ be a subspace of maximal dimension such that $\theta$ is $(V, \delta_{\dim V}, m)$-saturated (such subspaces exist, e.g. $V = \{0\}$). Write $k = \dim V$ and suppose $\theta$ is $(W, 1/m, m)$-saturated for some $W$. If $W \not\sqsubseteq V^{(\delta_d)}$ then certainly $W \not\sqsubseteq V^{(\delta_k)}$, so by Lemma 3.20(4), there is a subspace $W' \subseteq W$ with $\angle(V, W') > \delta_k$. By Lemma 3.21 (3) $\theta$ is $(W', (1+C)/m, m)$-saturated, and since $(1+C)/m < \delta_k$ it is $(W', \delta_k, m)$-saturated. By Lemma 3.21 (5), $\theta$ is $(V + W', 2\delta_k + \frac{C}{m}\log(\frac{1}{\delta_k}), m)$-saturated. Since $2\delta_k + \frac{C}{m}\log(\frac{1}{\delta_k}) < \delta_{k+1}$ the measure $\theta$ is $(V + W', \delta_{k+1}, m)$-saturated. Since $V' = V + W'$ has dimension at least $1 + k$ and $\theta$ is $(V', \delta_{\dim V'}, m)$-saturated, this contradicts the definition of $V$. □

## 3.8 Measures with uniformly concentrated components

When a measure has the property that at each level the components are with high probability concentrated on a subspace, one may expect the subspace to vary slowly between levels. This is the content of the following proposition, which may be applied to the conclusion of Theorem 2.8, but is also needed in the theorem's proof.

**Proposition 3.31.** *Let $0 < \varepsilon < 1$ and set $\delta = 3 \cdot 8^{d-1}\varepsilon^{1/(4\cdot 3^{d^2})}$. Let $\eta \in \mathcal{P}([0,1)^d)$ and $n \in \mathbb{N}$, and suppose that for every $n \leq k \leq n + \frac{1}{2}\log(1/\varepsilon)$ there is given a linear subspace $W_k \leq \mathbb{R}^d$ satisfying*

$$\mathbb{P}_{i=k}(\eta^{x,i} \text{ is } (W_k, \varepsilon)\text{-concentrated}) > 1 - \varepsilon. \tag{24}$$

*Then there are subspaces $V_k \leq W_k$ such that for $n \leq k \leq \frac{1}{2}\log(1/\varepsilon)$,*

$$\mathbb{P}_{i=k}(\eta^{x,i} \text{ is } (V_k, \delta)\text{-concentrated}) > 1 - 2d\sqrt{\varepsilon}, \tag{25}$$

*and $V_j \sqsubseteq V_i^{(\delta)}$ for all $n \leq i \leq j \leq n + \frac{1}{2}\log(1/\varepsilon)$.*

*Proof.* Write $N = [\frac{1}{2}\log(1/\varepsilon)]$. For each $n \le i \le n+N$ set $\mathcal{W}^i = \{W_j : n \le j \le i\}$ and apply Proposition 3.28 the with parameter $\varepsilon^{1/4}$ to obtain a subspace $V_i$ satisfying $V_i \sqsubseteq W_j^{(\delta/3)}$ and $V_i \sqsubseteq V_j^{(\delta/3)}$ for $n \le j \le i$, and $r(i) \le d$ subspaces $W_{i,1}, \ldots, W_{i,r(i)} \in \mathcal{W}_i$ such that $\bigcap_{j=1}^{r(i)} W_{i,j}^{(\varepsilon^{1/4})} \sqsubseteq V_i^{(\delta/3)}$.

Now, given $i$ and $1 \le j \le r(i)$, there is by definition a $n \le k = k(i,j) \le i$ such that $W_{i,j} = W_{k(i,j)}$. For every component $\theta = \eta^{x,k}$ in the event in (24), we can apply Lemma 3.17 (using $i - k(i,j) \le \frac{1}{2}\log(1/\varepsilon)$) to get

$$\mathbb{P}_{u=i}(\theta^{x,u} \text{ is } (V_k, \varepsilon^{1/4})\text{-concentrated}) > 1 - \sqrt{\varepsilon}.$$

Thus by (24),

$$\mathbb{P}_{u=i}(\eta^{x,u} \text{ is } (V_k, \varepsilon^{1/4})\text{-concentrated}) > 1 - 2\sqrt{\varepsilon}.$$

Hence,

$$\begin{aligned} \mathbb{P}_{u=i}(\eta^{x,u} \text{ is } (V_{k(i,j)}, \varepsilon^{1/4})\text{-concentrated for all } 1 \le j \le r(i)) &> 1 - 2r(i)\sqrt{\varepsilon} \\ &\ge 1 - 2d\sqrt{\varepsilon}. \end{aligned}$$

Finally, if $\theta = \eta^{x,i}$ is in the event above then, using $\bigcap_{j=1}^{r(i)} W_{i,j}^{(\varepsilon^{1/4})} = \bigcap_{j=1}^{r(i)} W_{k(i,j)}^{(\varepsilon^{1/4})} \sqsubseteq V_i^{(\delta/3)}$ we have

$$\begin{aligned} \theta(V_i^{(\delta/3)}) &\ge 1 - \sum_{j=1}^{r(i)} (1 - \theta(W_{k(i,j)}^{(\varepsilon^{1/4})})) \\ &\ge 1 - r(i) \cdot \varepsilon^{1/4}. \end{aligned}$$

Since $r(i) \le d$ and $d\varepsilon^{1/4} \le \delta/3$, this means that $\theta$ is $(V_i, \delta/3)$-concentrated. Since this is true for components $\theta = \eta^{x,i}$ with probability $> 1 - 2d\sqrt{\varepsilon}$, we have established (25), in fact with $\delta/3$ instead of $\delta$.

Finally, we show that we can assume $V_k \le W_k$. If $\varepsilon$ is so large that $\delta \ge 1$ there is nothing to prove since we can take $V_k = W_k$ from the start, so assume $\delta < 1$. From this and the relation $V_i \sqsubseteq W_i^{(\delta/3)}$ it follows that $\pi_{W_i}$ is injective on$V_i$ and satisfies $d(V_i, \pi_{W_i} V_i) \le \delta/3$. Thus $V_i^{(\delta/3)} \sqsubseteq (\pi_{W_i} V_i)^{(\delta)}$, so if a measure $\theta$ is $(V_i, \delta/3)$-concentrated, it is also $(\pi_{W_i} V_i, \delta)$-concentrated. It follows that if we replace $V_i$ by $\pi_{W_i} V_i$, we still will have (25), as desired. Also, since $V_j \sqsubseteq V_i^{(\delta/3)}$ for $n \le i < j$ before the modification, and each subspace moves by at most $\delta/3$, after the change we have $V_j \sqsubseteq V_i^{(\delta)}$ for $n \le i < j$, as desired. □

**Corollary 3.32.** *For every $\ell \in \mathbb{N}$ and $0 < \varepsilon < 1$ the following holds with $\delta = 3 \cdot 8^{d-1} \varepsilon^{1/(4 \cdot 3^{d^2})}$. Let $\eta \in \mathcal{P}([0,1]^d)$ and $N > \frac{1}{2}\log(1/\varepsilon)$, and suppose that for each $0 \le q \le N$ there is given a subspace $W_q \le \mathbb{R}^d$ such that*

$$\mathbb{P}_{i=q}(\eta^{x,i} \textit{ is } (W_q, \varepsilon)\textit{-concentrated}) > 1 - \varepsilon.$$

*Then there are subspaces $V_q \le W_q$ such that*

$$\mathbb{P}_{i=q}(\eta^{x,i} \textit{ is } (V_q, \delta)\textit{-concentrated}) > 1 - 2d\sqrt{\varepsilon},$$

*and*

$$\frac{1}{N+1} \# \{0 \le i \le N : d(V_i, V_{i-\ell}) \le \delta\} \ge 1 - \frac{2(d+1)\ell}{\log(1/\varepsilon)}.$$

*(Note that the conclusion is of interest only when $\ell$ is small compared to* $\log(1/\varepsilon)$*).*

*Proof.* We may assume that $\delta < 1$, otherwise the statement is trivial.

Let $m = [\frac{1}{2}\log(1/\varepsilon)]$. For each $k < [N/m]$ write $I_k = \{mk, mk+1, \ldots, m(k+1)-1\}$ and for $k = [N/m]$ write $I_k = \{m[N/m], \ldots, N\}$. For each $k \le [N/m]$, apply the previous proposition with $n = km$ to find subspaces $V_q \le W_q$, $q \in I_k$, such that $V_j \sqsubseteq V_i^{(\delta)}$ for all $i < j$ in $I_k$. This defines $V_q$ for all $0 \le q \le N$.

Fix $k$. If $i < j$ are in $I_k$ then $V_j \sqsubseteq V_i^{(\delta)}$ (since $\delta < 1$), hence $\dim V_j \le \dim V_i$, and if $\dim V_i = \dim V_j$ then $d(V_i, V_j) \le \delta$ (since $V_j \sqsubseteq V_i^{(\delta)}$). Let $i_0 = mk$ and let $i_{u+1} \in I_k$ denote the least index such that $\dim V_{i_{u+1}} < \dim V_{i_u}$. There are at most $d$ such indices. It follows from the above that if $j + \ell \in I_k$ and $d(V_j, V_{j-\ell}) \ge \delta$ then $i_u \le j < i_u + \ell$ for some $u$. There are at most $(d+1)\ell$ such indices $j$, so

$$\#\{i \,:\, i+\ell \in I_k \text{and } d(V_i, V_{i-\ell}) \ge \delta\} \le (d+1)\ell.$$

As the sets $I_0, \ldots, I_{[N/m]}$ are disjoint and cover $\{0, \ldots, N+1\}$, the bound above applies to each of them, so

$$\#\{0 \le i < N \,:\, d(V_i, V_{i-\ell}) \ge \delta\} \le (\frac{N}{m} + 1)(d+1)\ell.$$

Dividing by $N+1$ and using $N > \frac{1}{2}\log(1/\varepsilon)$ gives the desired bound. □

# 4 The inverse theorem in $\mathbb{R}^d$

Our goal in this section is to prove Theorem 2.8.

## 4.1 Elementary properties of convolutions

We begin with the obvious.

**Lemma 4.1.** *For $m \in \mathbb{N}$ and $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ ,*

$$H_m(\mu * \nu) \ge H_m(\mu) - O(\frac{1}{m}).$$

*Also, if $\mu$ is $(V, \varepsilon, m)$-saturated then $\mu * \nu$ is $(V, \varepsilon', m)$-saturated, where $\varepsilon' = \varepsilon + O(1/m)$.*

*Proof.* Notice that $\mu * \delta_y(A) = \mu(A - y)$, so that $H(\mu * \delta_y, \mathcal{D}_m) = H(\mu, \mathcal{D}_m + y)$, where $\mathcal{D}_m + y = \{[a+y, b+y) \,:\, [a,b) \in \mathcal{D}_m\}$. Thus by Lemma 3.2 (4), we have $H_m(\mu * \delta_y) \ge H_m(\mu) - O(\frac{1}{m})$. Since $\mu * \nu = \int \mu * \delta_y d\nu(y)$, concavity of entropy implies $H_m(\mu * \nu) \ge H_m(\mu) - O(\frac{1}{m})$. The second part follows using the same relation and Lemma 3.12. □

**Corollary 4.2.** *Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, $m \in \mathbb{N}$, and let $V \le \mathbb{R}^d$ be a linear subspace. Suppose that $\mu$ is not $(V, 2\varepsilon, m)$-saturated, and that $\nu$ is $(V, \varepsilon, m)$-saturated. Then*

$$H_m(\mu * \nu) > H_m(\mu) + \varepsilon',$$

*where $\varepsilon' = \varepsilon - O(1/m)$.*

*Proof.* Write $W = V^{\perp}$. By the previous lemma (with the roles of $\mu, \nu$ reversed),

$$H_m(\mu * \nu) \geq H_m(\pi_W(\mu * \nu)) + \dim V - (\varepsilon + O(1/m)).$$

Since $\pi_W$ is linear, $\pi_W(\mu * \nu) = \pi_W \mu * \pi_W \nu$, by the previous lemma $H_m(\pi_W(\mu * \nu)) \geq H_m(\pi_W \mu) - O(1/m)$. Inserting this in the last inequality and using the assumption that $H_m(\mu) \leq H_m(\pi_W \mu) + \dim V - 2\varepsilon$, and absorbing another $O(1/m)$ into the error term, we have

$$\begin{aligned} H_m(\mu * \nu) &\geq H_m(\pi_W \mu) + \dim V - (\varepsilon + O(1/m)) \\ &\geq H_m(\mu) + (\varepsilon - O(1/m)), \end{aligned}$$

as claimed. □

**Lemma 4.3.** *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$ be $(V, \varepsilon)$-concentrated, $0 < \varepsilon < 1$. Then $\mu^{*k}$ is $(V, (1 - \varepsilon k))$-concentrated for all $k \in \mathbb{N}$ with $\varepsilon k < 1$.*

*Proof.* Let $\mu = \varepsilon\mu_1 + (1-\varepsilon)\mu_2$ with $\mu_1$,$\mu_2$ probability measures and $\mu_1$ supported on a translate of $V^{(\varepsilon)}$. Then we can write $\mu^{\times k} = (1 - \varepsilon)^k \mu_1^{\times k} + (1 - (1 - \varepsilon)^k)\nu_k$ for some probability measure $\nu_k$, so, writing $\pi_k(x_1 \ldots x_k) = \sum_{i=1}^{k} x_i$, we have

$$\mu^{*k} = \pi_k \mu^{\times k} = (1 - \varepsilon)^k \pi_k \mu_1^{\times k} + (1 - (1 - \varepsilon)^k)\pi_k \nu_k$$

Since $\mu_1$ is supported on a translate of $V^{(\varepsilon)}$, the measure $\mu_1^{*k} = \pi_k \mu^{\times k}$ is supported on a translate of $\sum_{i=1}^{k} V^{(\varepsilon)} = V^{(\varepsilon k)}$. So the splitting of $\mu^{*k}$ above shows that $(1 - \varepsilon)^k$ of the mass of $\mu^{*k}$ is supported on an $\varepsilon k$-neighborhood of a translate of $V$. Since $(1 - \varepsilon)^k \geq 1 - \varepsilon k$, the claim follows. □

## 4.2 Mean, covariance and concentration

A rough but convenient way to describe the distribution of a measure is via its mean and covariance matrix. In this section we develop some basic properties of these objects and their relation to concentration.

By a covariance matrix we shall mean a $d \times d$ real symmetric matrix with non-negative eigenvalues (we do not require them to be positive). We denote the eigenvalues of such a matrix $\Sigma$ by

$$\lambda_1(\Sigma) \geq \lambda_2(\Sigma) \geq \ldots \geq \lambda_d(\Sigma).$$

set $\lambda_k = 0$ for $k > d$, preserving monotonicity. Define $\text{eigen}_{1\ldots r}(\Sigma)$ to be the span in $\mathbb{R}^d$ of the eigenvectors corresponding to eigenvalues $\geq \lambda_r(\Sigma)$. Note that if $\lambda_r(\Sigma) = \lambda_{r+1}(\Sigma)$ then $\dim(\text{eigen}_{1\ldots r}(\Sigma)) > r$.

It is advantageous to think of a covariance matrix as the positive semi-definite bi-linear form which it determines. The correspondence between these objects is not one-to-one: The matrix determines the form but the form determines the matrix only given the standard basis. Nevertheless, given the inner product, the form determines the eigenvalues and eigenspaces, and we are primarily interested in these; since the inner product is always fixed in our discussion, we will not lose much by thinking in terms of linear forms, and use the same notation for both. One advantage of this approach is that a bi-linear form can be restricted to a linear subspace, giving another bi-linear form, which is positive semi-definite if the original one was.

**Lemma 4.4.** *Let $\Sigma$ be a positive semidefinite form on $\mathbb{R}^d$ and $U \leq \mathbb{R}^d$ as subspace. Suppose that $u_1, \ldots, u_k$ is an orthonormal basis for $U$ and that $\Sigma(u_i, u_i) < \varepsilon$ for $i = 1, \ldots, d$. Then $\lambda_1(\Sigma|_{U\times U}) \leq d\varepsilon$.*

*Proof.* Let $u = \sum a_i u_i \in U$ be a unit vector, write $a = (a_1, \ldots, a_{d-r+1})$, so that $\|a\|_2 = 1$. Using Cauchy-Schwartz inequality for the "semi-inner product" $\langle v, w\rangle \mapsto v^T \Sigma w$ (which may not be positive, but satisfies the requirements for the weak inequality), and again to get $\sum |a_i| \leq \sqrt{d}\,\|a\|_2$, we have

$$\begin{aligned} u^T \Sigma_\mu u &= \sum_{i,j} a_i a_j u_i^T \Sigma_\mu u_j \\ &\leq \sum_{i,j} a_i a_j \sqrt{(u_i^T \Sigma_\mu u_i)(u_j^T \Sigma_\mu u_j)} \\ &< \sum_{i,j} a_i a_j \varepsilon \\ &\leq \varepsilon \cdot (\sum |a_i|)(\sum |a_j|) \\ &\leq \varepsilon \cdot d \cdot \|a\|_2^2 \\ &= d \cdot \varepsilon \end{aligned}$$

This proves the claim. ☐

For $\mu \in \mathcal{P}(\mathbb{R}^d)$, the mean of $\mu$ is

$$m = m(\mu) = \int x \, d\mu(x),$$

and the covariance matrix of $\mu$ is

$$\Sigma(\mu) = \int (x - m)(x - m)^T \, d\mu(x).$$

In this case we abbreviate

$$\lambda_i(\mu) = \lambda_i(\Sigma(\mu)),$$

and similarly $\text{eigen}_{1\ldots r}(\mu)$. We note that scaling a measure by $r$ results in multiplying its covariance matrix by $r^2$, an operation which does not affect the eigenvalues or eigenspaces.

**Lemma 4.5.** *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$, write $\Sigma = \Sigma(\mu)$, and let $U \leq \mathbb{R}^d$ be a linear subspace. Then $\Sigma|_U = \Sigma(\pi_U \mu)$ (the equality is of bi-linear forms on $U$).*

*Proof.* Write $m = m(\mu)$ and $m_U = \pi_U m = m(\pi_U \mu)$ (the last equality is immediate). For vectors $u, v \in U$, we now have

$$\begin{aligned} u^T\Sigma v &= u^T\left(\int (x-m)(x-m)^T d\mu(x)\right) v \\ &= \int u^T(x-m)(x-m)^T v d\mu(x) \\ &= \int \langle u, x-m\rangle \langle v, x-m\rangle\, d\mu(x) \\ &= \int \langle u, \pi_U(x-m)\rangle \langle v, \pi_U(x-m)\rangle\, d\mu(x) \\ &= \int \langle u, x-m_U\rangle \langle v, x-m_U\rangle\, d\pi_U\mu(x) \\ &= u^T\Sigma(\pi_U\mu)v. \end{aligned}$$

This proves the claim. □

A measure $\mu$ is supported on an $r$-dimensional affine subspace of $\mathbb{R}^d$ if and only if $\lambda_i(\mu) = 0$ for $i > r$, in which case it is supported on a translate of $\mathrm{eigen}_{1\ldots r}\,\mu$. We will use a quantitative version of this fact:

**Lemma 4.6.** *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$ and write $\lambda_i = \lambda_i(\mu)$ and $V_r = \mathrm{eigen}_{1\ldots r}(\mu)$.*

1. *$\mu$ is $(V_r, O(\lambda_{r+1}^{1/3}))$-concentrated.*
2. *If $\mu \in \mathcal{P}([0,1])$ is $(V,\varepsilon)$-concentrated for some $r$-dimensional subspace $V$ and $\varepsilon > 0$, then $\lambda_{r+1} = O(\varepsilon)$ and $\mu$ is $(V_r, O(\varepsilon^{1/3}))$-concentrated.*
3. *Let $\mu = \mu_\omega \in \mathcal{P}([0,1]^d)$ be a random measure defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Set $A = \mathbb{E}(\Sigma(\mu))$. If $\lambda_{r+1}(A) < \varepsilon$, then, writing $V = \mathrm{eigen}_{1,\ldots,r}(A)$,*

$$\mathbb{P}\Big(\mu \textit{ is } (V, O(\varepsilon^{1/6}))\textit{-concentrated}\Big) > 1 - O(\sqrt{\varepsilon}).$$

*Proof.* Let $\xi$ be an $\mathbb{R}^d$-valued random variable distributed according to $\mu$ and let $m = m(\mu)$ and $\Sigma = \Sigma(\mu)$. Identifying column vectors with $d \times 1$ matrices and scalars with $1 \times 1$ matrices, for $u \in \mathbb{R}^d$ we have

$$\begin{aligned} \mathbb{E}\left(\langle u, \xi - m\rangle^2\right) &= \mathbb{E}\left(u^T(\xi-m)(\xi-m)^T u\right) \\ &= u^T\mathbb{E}\left((\xi-m)(\xi-m)^T\right) u \\ &= u^T\Sigma u. \end{aligned}$$

For a subspace $W$, let $\eta_W$ denote the rotation-invariant probability measure on the unit sphere in $W$. Then there exists a constant $c = c(r)$ such that

$$d(\xi, m + V_r)^2 = c \cdot \int \langle u, \xi - m\rangle^2\, d\eta_{V_r^\perp}(u).$$

Therefore,

$$\begin{aligned} \mathbb{E}\left(d(\xi, m+V_r)^2\right) &= \mathbb{E}\left(c \cdot \int \langle u, \xi - m\rangle^2\, d\eta_{V_r^\perp}(u).\right) \\ &= c \cdot \int \mathbb{E}\left(\langle u, \xi - m\rangle^2\right) d\eta_{V_r^\perp}(u). \\ &\le c \cdot \lambda_{r+1}. \end{aligned}$$

because $u^T\Sigma u \leq \lambda_{r+1}$ for every unit vector in $V_r^\perp$. Now (1) follows from Markov's inequality.

For (2), fix $V$ as in the statement. Since $r+1+\dim V^\perp > d$, we must have $\dim\left(\mathrm{eigen}_{1,\ldots,r+1}\cap V^\perp\right) \geq 1$. Fix a unit vector $w \in \mathrm{eigen}_{1,\ldots,r+1}\cap V^\perp$. Then

$$\begin{aligned}
\mathbb{E}\left(d(\xi, m+V)^2\right) &= \mathbb{E}\left(\sup_{u\in V^\perp}\frac{\langle u, \xi - m\rangle^2}{\|u\|^2}\right)\\
&\geq \sup_{u\in V^\perp}\mathbb{E}\left(\frac{\langle u, \xi - m\rangle^2}{\|u\|^2}\right)\\
&\geq \mathbb{E}\left(\langle w, \xi - m\rangle^2\right)\\
&= w^T\Sigma(\mu)w\\
&\geq \lambda_{r+1}. \qquad (26)
\end{aligned}$$

On the other hand, since $\mu \in \mathcal{P}([0,1]^d)$ we have $\|\xi\| \leq \sqrt{d}$ $\mu$-a.s., hence, writing $\delta = \varepsilon(1+2\sqrt{d})$,

$$\begin{aligned}
\mathbb{E}\left(d(\xi, m+V)^2\right) &\leq \delta^2\mathbb{P}(\xi \in (m+V)^{(\delta)}) + d\cdot\mathbb{P}(\xi \in [0,1]^d \setminus (m+V)^{(\delta)})\\
&\leq \delta^2 + d\cdot\mathbb{P}(\xi \in [0,1]^d \setminus (m+V)^{(\delta)}). \qquad (27)
\end{aligned}$$

Finally, since $\mu$ is $(V,\varepsilon)$-concentrated, there is a translate $U$ of $V$ such that $\mu(U^{(\varepsilon)}) > 1-\varepsilon$. Hence

$$\begin{aligned}
m &= \mathbb{E}(\xi)\\
&= \mu(U^{(\varepsilon)})\mathbb{E}(\xi|\xi \in U^{(\varepsilon)}) + (1-\mu(U^{(\varepsilon)})\mathbb{E}(\xi|\xi \in \mathbb{R}^d \setminus U^{(\varepsilon)}).
\end{aligned}$$

Since $U^{(\varepsilon)}$ is convex, $\mathbb{E}(\xi|\xi \in U^{(\varepsilon)}) \in U^{(\varepsilon)}$. Also, since $\|\xi\| \leq \sqrt{d}$, both expectations on the right hand side of the last equation have magnitude at most $\sqrt{d}$. Thus

$$d(m, U^{(\varepsilon)}) \leq \left\|m - \mathbb{E}(\xi|U^{(\varepsilon)})\right\| \leq 2\varepsilon\sqrt{d}.$$

Therefore $U^{(\varepsilon)} \subseteq m + V^{(\varepsilon+2\varepsilon\sqrt{d})} = m + V^{(\delta)}$, and consequently

$$\mathbb{P}(\xi \in [0,1]^d \setminus (m+V)^{(\delta)}) \leq \mathbb{E}(\xi \notin U^{(\varepsilon)}) < \varepsilon.$$

Combined with (26) and (27) this proves the first part of (2), the second part now follows from (1).

We turn to (3). Let $U = (\mathrm{eigen}_{1\ldots r} A)^\perp \leq \mathrm{eigen}_{r+1,\ldots,d} A$. Also for brevity write $\Sigma_\mu = \Sigma(\mu)$. For any unit vector $u \in U$, we have

$$\varepsilon > u^T Au = \mathbb{E}(u\Sigma_\mu u^T)$$

Since $u^T\Sigma_\mu u \geq 0$, by Markov's inequality,

$$\mathbb{P}(u^T\Sigma_\mu u > \sqrt{\varepsilon}) < \sqrt{\varepsilon}$$

Now fix an orthonormal basis $u_1 \ldots u_\ell$ of $U$ (so $\ell \leq d-(r+1)$). By the last inequality,

$$\mathbb{P}(u_i^T\Sigma_\mu u_i \leq \sqrt{\varepsilon} \text{ for all } i = 1,\ldots,d) \geq 1 - d\sqrt{\varepsilon}$$

By the Lemma 4.4, the condition in the event above implies that $\lambda_1(\Sigma_\mu|_{U\times U}) \leq d\sqrt{\varepsilon}$, where $\Sigma_\mu|_{U\times U}$ is the restriction of the quadratic for $\Sigma_\mu$ to $U \times U$, and by Lemma 4.5, $\Sigma_\mu|_{U\times U} = \Sigma(\pi_U\mu)$ (as linear forms on $U$). Combined with the previous probability estimate we get

$$\mathbb{P}\left(\lambda_1(\Sigma_{\pi_U\mu}) \leq d\sqrt{\varepsilon}\right) \geq 1 - d\sqrt{\varepsilon} \tag{28}$$

By the first part of this lemma, for $\mu$ in the event in (28), $\pi_U\mu$ is $(\{0\}, O(\varepsilon^{1/6}))$-concentrated, and this is the same as saying that $\mu$ is $(V, O(\varepsilon^{1/6}))$-concentrated, as claimed. □

Recall the definition of the distance between linear subspaces (20). We shall use the following basic fact, which we state without proof.

**Lemma 4.7.** *The maps $\Sigma \mapsto \lambda_i(\Sigma)$ are continuous on the set of positive semi-definite matrices. Furthermore, given $\tau > \sigma > 0$ and $1 \leq r \leq d$, the map $\Sigma \mapsto \mathrm{eigen}_{1\ldots r}\Sigma$ is continuous on the compact space of positive semi-definite matrices $\Sigma$ satisfying $\lambda_r(\Sigma) \geq \tau$ and $\lambda_{r+1}(\Sigma) \leq \sigma$.*

## 4.3 Gaussian measures and the Berry-Esseen-Rotar estimate

The standard $d$-dimensional Gaussian measure $\gamma = \gamma_d$ is given by $\gamma(A) = \int_A \varphi(x)dx$, where $\varphi = \varphi_d$ is $\varphi(x) = (2\pi)^{d/2}\exp(-\frac{1}{2}\|x\|^2)$. The mean and covariance are 0 and $I$ (the $d \times d$ identity matrix), respectively. Given a $d \times d$ covariance matrix $\Sigma$ and $m \in \mathbb{R}^d$, write $\Sigma = BB^T$. The Gaussian measure with mean $m \in \mathbb{R}^d$ and covariance $\Sigma$ is the push-forward of $\gamma$ by the map $x \mapsto Bx + m$ and is denoted $N(m, \Sigma)$. When $\Sigma$ is non-singular its density with respect to Lebesgue is

$$f(x) = \frac{1}{\sqrt{(2\pi)^d \det \Sigma}} \exp(-\frac{1}{2}(x-m)^T\Sigma^{-1}(x-m)).$$

When $\Sigma$ is singular and $r$ is such that $\lambda_r(\Sigma) > 0$, $\lambda_{r+1}(\Sigma) = 0$, one obtains a similar formula for the density on the affine space $V = \mathrm{eigen}_{1\ldots r}(\Sigma) + m$ with respect to the $r$-dimensional Hausdorff measure on $V$. In particular, if $\mu = N(m, \Sigma)$ and $\nu$ is the push-forward of $\mu$ through the map $x \mapsto rx$, then $\nu = N(rm, r^2\Sigma)$.

If $\mu_1, \ldots, \mu_k$ are measures then $\mu = \mu_1 * \ldots * \mu_k$ has mean $m(\mu) = \sum_{i=1}^k m(\mu_i)$ and covariance $\Sigma(\mu) = \sum_{i=1}^k \Sigma(\mu_i)$. If $\mu_i = N(m_i, \Sigma_i)$ then $\mu_1 * \ldots * \mu_k = N(\sum m_i, \sum \Sigma_i)$.

The central limit theorem asserts that, for $\mu_1, \mu_2, \ldots \in \mathcal{P}(\mathbb{R}^d)$ which are not too concentrated on subspaces, the convolutions $\mu_1 * \ldots * \mu_k$ can be re-scaled so that the resulting measure is close to a Gaussian measure. The Berry-Esseen estimate and its variants quantify the rate of this convergence. The following multi-dimensional variant is due to Rotar [28].

**Theorem 4.8.** *Let $\mu_1, \ldots, \mu_k$ be probability measures on $\mathbb{R}^d$ with finite third moments $\rho_i = \int \|x\|^3 \, d\mu_i(x)$. Let $\mu = \mu_1 * \ldots * \mu_k$ and let $\gamma$ be the Gaussian measure with the same mean and covariance matrix as $\mu$. Then for any convex Borel set $D \subseteq \mathbb{R}^d$,*

$$|\mu(D) - \gamma(D)| \leq C_1 \cdot \frac{\sum_{i=1}^k \rho_i}{\lambda_d(\mu)^{3/2}},$$

*where* $C_1 = C_1(d)$. *In particular, if* $\rho_i \leq C$ *and* $\lambda_d(\mu_i) \geq c$ *for constants* $c, C > 0$ *then*

$$|\mu(D) - \gamma(D)| = O_{c,C}(k^{-1/2}).$$

## 4.4 Multi-scale analysis of repeated self-convolutions

If $\mu \in \mathcal{P}(\mathbb{R}^d)$ is supported on a subspace $V \leq \mathbb{R}^d$ but not on a smaller subspace, and if the support is bounded, then $\mu^{*k}$ becomes increasingly smooth as a measure on $V$, in the sense that, by the central limit theorem, it converges (after suitable re-scaling) to a Gaussian on $V$. In this section we prove a localized version of this statement which applies with high probability to the components of the measure. Specifically, for $\mu \in \mathcal{P}(\mathbb{R}^d)$ of bounded support, for every $\delta > 0$ and integer scale $m$, there are subspaces $V_0, V_1, \ldots$ such that typical level-$i$ components of $\mu$ are $(\delta, 2^{-m})$-concentrated on $V_i$, and, when $k$ is large, a typical level-$i$ component of $\mu^{*k}$ is $(V_i, \delta, m)$-saturated.

For a linear subspace $V \leq \mathbb{R}^d$ let $\pi_V$ denote the orthogonal projection $\mathbb{R}^d \to V$. Recall our convention that $\lambda_i = 0$ for $i > d$, and in what follows define $\lambda_0(\Sigma) = d$, so that when $\mu \in \mathcal{P}([0,1)^d)$ and $\Sigma = \Sigma(\mu)$ the sequence $(\lambda_i(\mu))_{i=0}^{\infty}$ is monotone.

**Proposition 4.9.** *Let* $\sigma > 0$, $\delta > 0$, $R > 0$ *and*[8] $m > m(\delta, R)$. *Then there exists an integer* $p = p_0(\sigma, \delta, R, m)$ *such that for all* $k \geq k_0(\sigma, \delta, R, m)$ *and all* $0 \leq \rho < \rho_0(\sigma, \delta, R, m, k)$, *the following holds:*

*Let* $\mu_1, \ldots, \mu_k \in \mathcal{P}([-R, R]^d)$, *let* $\mu = \mu_1 * \ldots * \mu_k$ *and* $V = \mathrm{eigen}_{1,\ldots,r}\, \mu$ *for some* $0 \leq r \leq d$, *and suppose that* $\lambda_r(\mu) \geq \sigma k$ *and* $\lambda_{r+1}(\mu) \leq \rho$. *Then*

$$\mathbb{P}_{i=p-[\log \sqrt{k}]}\left(\mu^{x,i} \textit{ is } (V, \delta, m)\textit{-uniform}\right) > 1 - \delta. \tag{29}$$

*Remark* 4.10. Instead of $\lambda_{r+1}(\mu) < \rho$ we could require $\mu$ to be $(V, \rho)$-concentrated. This would give a formally equivalent statement (using Lemma 4.6 (2)).

*Proof.* It is a general fact that, for an absolutely continuous probability measure $\gamma$, for $\gamma$-a.e. $x$, as $p \to \infty$ the components $\gamma^{x,p}$ converge weak-* to Lebesgue measure on $[0,1]^d$, and in particular

$$\mathbb{E}_{i=p}(H_m(\gamma^{x,i})) \to d \qquad \text{as } p \to \infty \tag{30}$$

(this is a consequence of the martingale convergence theorem). There is no guaranteed rate of convergence, but if $\gamma$ has a continuous density function $f$, then convergence holds at every $x$ for which $f(x) > 0$, and the rate depends only on $f(x)$ and on the modulus of continuity of $f$ at $x$. In particular, when $f \in C^1$ has a smooth density $f$, the convergence rate at $x$ is controlled by $f(x)$ and the bounds on $\|\nabla f(x)\|$ near $x$. Thus, for any compact family $\mathcal{E} \subseteq M_d(\mathbb{R})$ of non-singular co-variance matrices and any compact $K \subseteq \mathbb{R}^d$, convergence in (30) is uniform as $\gamma$ ranges over the Gaussians $\gamma$ with mean 0 and co-variance matrix $\Sigma \in \mathcal{E}$, and $x$ ranges over $K$. Furthermore, given $\mathcal{E}$ we can choose a

[8] In the one-dimensional case in [12] there wan no requirement that $m$ be large. The reason this is necessary in the multi-dimensional case is that, even when $\mu \in \mathcal{P}([0,1]^d)$ is Lebesgue measure on $V \cap [0,1]^d$ for an affine subspace $V$, we do not generally have $H_m(\mu) = \dim V$, but rather only $H_m(\mu) = \dim V - o(1)$. One can change coordinates so that if $\mathcal{D}_m$ is defined in the new coordinates, $H_m(\mu) = \dim V$, but the coordinate change itself incurs an $O(1/m)$ loss for $H_m(\cdot)$.

compact $K_2 \subseteq \mathbb{R}^d$ so that it has arbitrarily large mass uniformly for such $\gamma$. Summarizing, given $0 < \sigma, \delta < 1$, there is a $p = p_0(\sigma, \delta, m)$ such that, for any Gaussian $\gamma$ with $\sigma \le \lambda_d(\gamma) \le 1/\sigma$,

$$\mathbb{P}_{i=p}\left(H_m(\gamma^{x,i}) > d - \delta\right) > 1 - \delta. \tag{31}$$

In addition, by Lemma 3.2 (1), there is a weakly open neighborhood $\mathcal{U}_\delta \subseteq \mathcal{P}(\mathbb{R}^d)$ of these Gaussians such that the inequality continues to be valid for all $\gamma \in \mathcal{U}_\delta$.

Next, let $\mu = \mu_1 * \ldots * \mu_k$ be as in the statement of the proposition and first assume $r = d$, so $\lambda_d(\mu) \ge \sigma k$. The third moments of the $\mu_i$ are bounded by $O_R(1)$, because $\mu_i \in \mathcal{P}([-R,R]^d)$. Thus by Theorem 4.8, if $k$ is large enough in a manner depending only on $\delta$, the scaling $\mu'$ of $\mu$ given by $\mu'(A) = \mu(2^{[\log\sqrt{k}]} \cdot A)$ will belong to $U_\delta$. Thus we obtain (31) for $\mu'$. Scaling everything back by a factor of $2^{[\log\sqrt{k}]}$ we obtain (29).

Now consider the case that $\Sigma(\mu)$ is singular, i.e. $\lambda_d(\mu) = 0$. Fix $r, \rho$ and $V = \mathrm{eigen}_{1\ldots r}\,\mu$ as in the statement of the proposition, and let $\pi = \pi_V$ denote the orthogonal projection to $V$. Then the argument in the last paragraph applies in $V$ to the measure $\pi\mu = \pi\mu_1 * \ldots * \pi\mu_k$ and ensures that

$$\mathbb{P}_{i=p-[\log\sqrt{k}]}\left(H_m(\pi\mu^{x,i}) > r - \delta/2 - O(1/m)\right) > 1 - \delta/2.$$

The $O(1/m)$ term arises because we have transferred the entropy bound from the dyadic partition $\mathcal{D}_m^V$ on $V$ to the dyadic partition $\mathcal{D}_m$ of $\mathbb{R}^d$. But, as we are assuming that $m$ is large relative to $\delta$, we can absorb this term in $\delta$ and assume that

$$\mathbb{P}_{i=p-[\log\sqrt{k}]}\left(H_m(\pi_V\mu^{x,i}) > r - \delta/2\right) > 1 - \delta/2. \tag{32}$$

Now, the hypothesis $\lambda_{r+1}(\mu) \le \rho$ means that $\mu$ is $(V, \sqrt{\rho})$-concentrated (Lemma 4.6) and so for $\rho$ small enough in a manner depending on the other parameters, a $(1 - \delta/2)$-fraction of the components $\mu^{x,p-[\log\sqrt{k}]}$ are $(V, 2^{-m})$-concentrated (Lemma (4.3)). In fact by taking $\rho$ small we can ensure an arbitrarily high degree of concentration. Furthermore, if enough of the mass of such a component $\mu^{x,p-[\log\sqrt{k}]}$ is concentrated on a small enough neighborhood of $V$, then on this neighborhood $\pi$ will be close enough to the identity map (in the supremum norm on continuous self-maps of $[0,1]^d$) that Lemma 3.2 (3) will imply $|H_m(\mu^{x,p-[\log\sqrt{k}]}) - H_m(\pi_V\mu^{x,p-[\log\sqrt{k}]})| < \delta/2$. Combined with (32), we obtain (29). □

We now specialize to convolutions of a single measure.

**Proposition 4.11.** *Let $\sigma, \delta > 0$ and $m > m(\delta)$. Then there exists $p = p_1(\sigma, \delta, m)$ such that for sufficiently large $k \ge k_1(\sigma, \delta, m)$ and sufficiently small $0 < \rho \le \rho_1(\sigma, \delta, m, k)$, the following holds.*

*Let $\mu \in \mathcal{P}(\mathbb{R}^d)$, fix an integer $i_0 \ge 0$, and write*

$$A = \mathbb{E}_{i=i_0}(\Sigma(\mu^{x,i})).$$

*If $\lambda_r(A) > \sigma$ and $\lambda_{r+1}(A) < \rho$ for some $1 \le r \le d$, then, setting $V = \mathrm{eigen}_{1\ldots r}(A)$, $\nu = \mu^{*k}$ and $j_0 = i_0 - [\log\sqrt{k}] + p$, we have*

$$\mathbb{P}_{j=j_0}\left(\nu^{x,j} \text{ is } (V,\delta,m)\text{-saturated}\right) > 1 - \delta.$$

*Proof.* Fix $\sigma$, $\delta$, $m$, $k$, $\rho$, $\mu$, $A$ , $V$, $i_0$ as in the statement, we will see that if the stated relationships hold and $p$ is defined as in the statement, then the conclusion holds.

Let $\widetilde{\mu}$ denote the $k$-fold self-product $\widetilde{\mu} = \mu \times \ldots \times \mu$ and $\pi : (\mathbb{R}^d)^k \to \mathbb{R}^d$ the map

$$\pi(x_1, \ldots, x_k) = \sum_{i=1}^{k} x_i.$$

Then $\nu = \pi\widetilde{\mu}$, and, since $\widetilde{\mu} = \mathbb{E}_{i=i_0}(\widetilde{\mu}_{x,i})$, we also have by linearity $\nu = \mathbb{E}_{i=i_0}(\pi(\widetilde{\mu}_{x,i}))$. Thus, by Corollary 3.13 and an application of Markov's inequality, there is a $\delta_1 > 0$, depending only on $\delta$ and $d$, such that if $m$ is large enough as a function of $\delta_1$ then the proposition will follow if we show that with probability $> 1 - \delta_1$ over the choice of the component $\widetilde{\mu}_{x,i_0}$ of $\widetilde{\mu}$, the measure $\tau = \pi(\widetilde{\mu}_{x,i_0})$ satisfies

$$\mathbb{P}_{j=j_0}\left(\tau^{y,j} \text{ is } (V, \delta_1, m)\text{-uniform}\right) > 1 - \delta_1.$$

If we manage to define a random subspace $W = W(\widetilde{\mu}_{x,i_0})$ such that

$$\mathbb{P}_{j=j_0}\left(d(W,V) < \frac{1}{\sqrt{d}}2^{-(m+1)} \text{ and } \tau^{y,j} \text{ is } (W, \delta_1, m+1)\text{-uniform}\right) > 1 - \delta_1,$$

then the previous inequality follows by applying Lemma 3.21 to each component $\eta^{y,i}$ in the last event (we use here the assumption that $m$ is large relative to $\delta_1$). We thus aim to define $W$ such that (33) holds.

Set $\eta = \pi(\widetilde{\mu}^{x,i_0})$ and notice that, with $\tau$ as before, the distribution of the components $\tau^{y,j_0}$ is the same as the distribution of the components of $\eta^{z,j_0-i_0}$. Thus what we really aim to prove is that we

$$\mathbb{P}_{j=j_0-i_0}\left(d(W,V) < \frac{1}{\sqrt{d}}2^{-(m+1)} \text{ and } \eta^{y,j} \text{ is } (W, \delta_1, m+1)\text{-uniform}\right) > 1 - \delta_1. \tag{33}$$

A random component $\widetilde{\mu}^{x,i_0}$ is itself a product measure $\widetilde{\mu}^{x,i_0} = \mu^{x_1,i_0} \times \ldots \times \mu^{x_k,i_0}$ (here $x = (x_1, \ldots, x_k)$), and the marginal measures $\mu^{x_j,i_0}$ of this product are distributed independently according to the distribution of the rescaled components of $\mu$ at level $i_0$. Recall that

$$\Sigma(\pi(\mu^{x_1,i_0} \times \ldots \times \mu^{x_k,i_0})) \quad = \quad \sum_{j=1}^{k} \Sigma(\mu^{x_j,i_0}) \tag{34}$$

Fixing a parameter $\delta_2$ which will depend on $\sigma, \delta_1$, by the weak law of large numbers, if $k$ is large enough in a manner depending on $\delta_2$, then with probability $> 1 - \delta_2$ over the choice of $\widetilde{\mu}_{x,i_0}$ we will have[9]

$$\left\| \frac{1}{k}\Sigma(\pi\widetilde{\mu}^{x,i_0}) - A \right\| < \delta_2. \tag{35}$$

[9] We use here the fact that we have a uniform bound for the rate of convergence in the weak law of large numbers for i.i.d. random variables $X_1, X_2, \ldots$. In fact, the rate can be bounded in terms of the mean and variance of $X_n$. Here $X_n$ are matrix-valued (they are distributed like the covariance matrix of the level-$i_0$ components of $\mu$), and therefore the mean and variance of the components of $X_n$ can be bounded independently of the measure $\mu \in \mathcal{P}([0,1)^d)$.

Using Lemma 4.7 and the fact that $\mu_r(A) > \sigma$ and $\lambda_{r+1}(A) < \rho$, and assuming as we may that $\rho < k/4$, we can choose $\delta_2$ in a manner depending on $\sigma, \delta_1$, in such a way that (35) implies

$$\begin{aligned}\lambda_r(\pi\widetilde{\mu}^{x,i_0}) &> \frac{k\sigma}{2}\\ \lambda_{r+1}(\pi\widetilde{\mu}^{x,i_0}) &< \frac{k\sigma}{4}\end{aligned}$$

and such that, if we write $W_{x,i_0} = \mathrm{eigen}_{1\ldots r}\,\Sigma(\pi\widetilde{\mu}_{x,i_0}) = \mathrm{eigen}_{1\ldots r}\,\Sigma(\pi\widetilde{\mu}^{x,i_0})$, then

$$d(W_{x,i_0}, V) < \frac{1}{\sqrt{d}}2^{-(m+1)}$$

Thus, assuming that $k$ is large enough, we have shown

$$\mathbb{P}_{i=i_0}\left(\lambda_r(\pi\widetilde{\mu}^{x,i}) > \frac{k\sigma}{2} \text{ and } d(W_{x,i}, V) < \frac{1}{\sqrt{d}}2^{-(m+1)}\right) > 1 - \delta_2 \tag{36}$$

Next, fix such a $k$. By hypothesis $\lambda_{r+1}(A) = \lambda_{r+1}(\mathbb{E}_{i=i_0}(\Sigma(\mu^{x,i}))) < \rho$, so by Lemma 4.6 (3),

$$\mathbb{P}_{i=i_0}\left(\mu^{x,i} \text{ is } (V, O(\rho^{1/6}))\text{-concentrated}\right) > 1 - O(\sqrt{\rho})$$

Using again the fact that $\widetilde{\mu}^{x,i_0}$ is a product of $k$ independent copies of level-$i_0$ components of $\mu$, the last inequality implies

$$\mathbb{P}_{i=i_0}\left(\text{all marginals of } \widetilde{\mu}^{x,i} \text{ are } (V, O(\rho^{1/6})\text{-concentrated}\right) > 1 - O(k\sqrt{\rho}) \tag{37}$$

If $\widetilde{\mu}^{x,i_0}$ is in the event above, then all its marginals are $(V, O(\rho^{1/6}))$-concentrated, so by Lemma 4.3, $\pi(\widetilde{\mu}^{x,i_0})$ is $(V, O(k\rho^{1/6}))$-concentrated. By Lemma 4.6 (2), $\lambda_{r+1}(\pi(\widetilde{\mu}^{x,i_0})) < O_k(\rho^{1/6})$, and we conclude that

$$\mathbb{P}_{i=i_0}\left(\lambda_{r+1}(\pi(\widetilde{\mu}^{x,i_0})) \leq O_k(\rho^{1/6})\right) > 1 - O(k\sqrt{\rho})$$

Combining this with (36) and assuming that $\rho$ is sufficiently small relative to $\delta_2$, we have

$$\mathbb{P}_{i=i_0}\left(\begin{array}{rl}\lambda_r(\pi\widetilde{\mu}^{x,i}) > & \frac{k\sigma}{2}\\ \lambda_{r+1}(\pi(\widetilde{\mu}^{x,i})) < & O_k(\rho^{1/6})\\ d(W_{x,i_0}, V) < & \frac{1}{\sqrt{d}}2^{-(m+1)}\end{array}\right) > 1 - 2\delta_2$$

Let $\widetilde{\mu}^{x,i_0}$ belong to the event above. Let us recall the dependences of the parameters: $\delta$ is given and determines $\delta_1$, then $m$ is large relative to $\delta_1$, then $\delta_2$ small depending on $\sigma, \delta_!$, then $k$ is correspondingly large, and $\rho$ correspondingly small. So we can assume that $k$ is large enough, and $\rho$ small enough, to apply Proposition 4.9 with parameters $\delta_1, m+1$ and $\sigma/2$, and conclude that there is a $p = p(\delta_1, m+1, k) = p(\delta, m, k)$ such that, writing $\eta = \pi\widetilde{\mu}^{x,i_0}$,

$$\mathbb{P}_{j=j_0-i_0}\left(\eta^{y,j} \text{ is } (W_{x,i_0}, \delta_1, m+1)\text{-uniform}\right) > 1 - \delta_1.$$

This and the estimate above on the probability that $d(W_{x,i_0}, V) < \frac{1}{\sqrt{d}}2^{-(m+1)}$ give (33), which is what we wanted. □

**Theorem 4.12.** *Let $\delta > 0$ and $m \in \mathbb{N}$. Then there exists*[10] *a $0 \le k \le k_2(\delta, m)$ such that for all sufficiently large $n \ge n_2(\delta, m, k)$, the following holds: For any $\mu \in \mathcal{P}(\mathbb{R}^d)$ there is a sequence $V_0, \dots, V_n$ of subspaces*[11] *of $\mathbb{R}^d$ such that, writing $\nu = \mu^{*k}$,*

$$\mathbb{P}_{0 \le i \le n}\left(\nu^{x,i} \text{ is } (V_i, \delta, m)\text{-saturated}\right) > 1 - \delta$$

*and*

$$\mathbb{P}_{0 \le i \le n}\left(\mu^{x,i} \text{ is } (V_i, \delta)\text{-concentrated}\right) > 1 - \delta.$$

*Proof.* It is a formal consequence of Proposition 3.19 that we may assume that $m$ is large in a manner depending on $\delta$. We also may assume that $\delta < 1/2$. Also, since we are free to take $R$ large relative to $n$, we can assume that $\mu$ is supported on $[0, 1)^d$.

Let $k_1(\cdot)$, $p_1(\cdot)$, $\rho_1(\cdot)$ be as in Proposition 4.11. We assume, without loss of generality, that these functions are monotone in each of their arguments.

Let $c > 1$ denote a constant good for all previous big-$O$ bounds.

The proof will depend on a function $\widetilde{\rho} : (0, d] \to (0, d]$ such that $\widetilde{\rho}(\sigma)$ is small in a manner depending on $\sigma, \delta, m$. Specifically, we require that $\widetilde{\rho}$ satisfy the following inequalities, where $\exp_2(y) = 2^y$ (for concreteness, on could define $\widetilde{\rho}(\sigma)$ to be one-half the minimum of the right-hand sides):

$$\widetilde{\rho}(\sigma) < \sigma, \tag{38}$$

$$\widetilde{\rho}(\sigma) < \rho_1(\sigma, \delta/2, m, k_1(\sigma, \delta/2, m)), \tag{39}$$

$$\widetilde{\rho}(\sigma) < \frac{\delta^{12}}{c^6 (2d)^{12}}, \tag{40}$$

$$\widetilde{\rho}(\sigma) < \frac{1}{c^6} \exp_2\left(-\frac{24(d+1) \cdot ([\log \sqrt{k_1(\sigma, \delta/2, m)}] - p_1(\sigma, \delta/2, m))}{\delta/2}\right), \tag{41}$$

$$\widetilde{\rho}(\sigma) < \frac{1}{c^4} \cdot \left(\frac{\delta}{(\sqrt{d} + 1) \cdot 3 \cdot 8^{d-1}}\right)^{24 \cdot 3^{d^2}}. \tag{42}$$

As before define $\lambda_0(\Sigma_i) = d$ and $\lambda_{d+1}(\Sigma_i) = 0$. Fix $n$ and $\mu$, we shall later see how large an $n$ is desirable. For $0 \le q \le n$ write

$$\Sigma_q = \mathbb{E}_{i=q}\left(\Sigma(\mu^{x,i})\right).$$

Define a sequence $\sigma_0 > \sigma_1 > \dots$ by $\sigma_0 = d$ and $\sigma_i = \widetilde{\rho}(\sigma_{i-1})$ (the sequence is decreasing because of (38)). For a covariance matrix $\Sigma$ and $s \in \mathbb{N}$, set

$$N_s(\Sigma) = \#\{1 \le j \le d \,:\, \lambda_j(\Sigma) \in (\sigma_s, \sigma_{s-1}]\}.$$

*Claim* 4.13. There is an $s \le \lceil 1 + 2d/\delta \rceil$ satisfying

$$\mathbb{P}_{0 \le q \le n}(N_s(\Sigma_q) = 0) > 1 - \frac{\delta}{2}.$$

[10] In [12] the corresponding statement holds for all large enough $k$. The reason the size of $k$ must be restricted is, roughly, that if $\mu$ is concentrated extremely near a subspace $V$ then it will remain so for a reasonable number of convolutions, but too many convolutions will make it drift away from $V$.

[11] The corresponding theorem in [12] is stated differently, in terms of disjoint subsets $I, J \subseteq \{1, \dots, n\}$. See remark after Theorem 2.8.

*Proof.* Note that $\sum_{r=1}^{\infty} N_r(\Sigma_q) = d$, so

$$\sum_{s=1}^{\lceil 1+2d/\delta \rceil} \mathbb{E}_{0\leq q\leq n}(N_i(\Sigma_q)) = \mathbb{E}_{0\leq q\leq n}(\sum_{s=1}^{\lceil 1+2d/\delta \rceil} N_i(\Sigma_q)) \leq d. \tag{43}$$

Thus there must exist an $s \leq \lceil 1 + 2d/\delta \rceil$ such that

$$\mathbb{E}_{0\leq q\leq n}(N_s(\Sigma_q)) \leq \frac{d}{\lceil 1+2d/\delta \rceil} < \frac{\delta}{2}.$$

Since $N_i(\cdot)$ is integer valued, we have

$$\mathbb{P}_{0\leq q\leq n}(N_s(\Sigma_q) \geq 1) \leq \mathbb{E}_{0\leq q\leq n}(N_s(\Sigma_q)),$$

so this is the desired $s$. □

Fix an $s$ that satisfies the conclusion of the lemma, write

$$\begin{aligned} \sigma &= \sigma_{s-1} \\ \rho &= \sigma_s \\ &= \widetilde{\rho}(\sigma), \end{aligned}$$

and set

$$k = k_1(\sigma, \frac{\delta}{2}, m).$$

Note that $k$ is bounded above by some expression $k_2(\delta, m)$ (also depending implicitly on the choice of the function $\widetilde{\rho}$), as in the statement, since its largest possible value occurs for $s = [1 + 2d/\delta]$, and once the function $\widetilde{\rho}$ is fixed, the magnitude $\sigma$, and hence $k$, is bounded.

Let

$$I = \{0 \leq q \leq n \,:\, N_s(\Sigma_q) = 0\}.$$

By our choice of $s$,

$$|I| \geq (1 - \frac{\delta}{2})(n+1).$$

For $q \in I$ let $1 \leq r_q \leq d$ denote the smallest integer such that

$$\lambda_{r_q}(\Sigma_q) \geq \sigma \qquad \text{and} \qquad \lambda_{r_q+1}(\Sigma_q) < \rho,$$

which exists by definition, and set

$$W_q = \text{eigen}_{1,\ldots,r_q}(\Sigma_q).$$

We define $W_q = \mathbb{R}^d$ for $q \notin I$. Finally, write

$$\ell = [\log \sqrt{k}] - p_1(\sigma, \frac{\delta}{2}, m).$$

*Claim* 4.14. For $q \in I$,

$$\mathbb{P}_{i=q}\left(\nu^{x,i-\ell} \text{ is } (W_i, \frac{\delta}{2}, m)\text{-saturated}\right) \;>\; 1 - \frac{\delta}{2} \tag{44}$$

$$\mathbb{P}_{i=q}\left(\mu^{x,i} \text{ is } (W_i, c\rho^{1/6})\text{-concentrated}\right) \;>\; 1 - c\rho^{1/6}. \tag{45}$$

*Proof.* The first inequality follows from Proposition 4.11 and our choice of parameters, specifically the definition of $\ell$ and assumption (39). The second follows from Lemma 4.6 (3) applied to the random component $\mu^{x,i}$, since $W_q = \text{eigen}_{1,\ldots,r_q}$, $\mathbb{E}_{i=q}\left(\lambda_{r_q+1}(\mu^{x,i})\right) < \rho$. $\square$

This is almost what we want, except that in (44) the level of the component is shifted by $\ell$ (that is, $\nu^{x,i-\ell}$ appears instead of $\nu^{x,i}$). To correct this we apply Corollary 3.32 to (45) with parameter $c\rho^{1/2}$ (we can do this since we are assuming that $n$ is large relative to $\rho$). Then, writing

$$\rho' = 3 \cdot 8^{d-1} c^{1/(4\cdot 3^{d^2})} \rho^{1/(24\cdot 3^{d^2})},$$

we conclude the there are subspaces $W_i' \leq W_i$ such that for all $0 \leq q \leq n$,

$$\begin{aligned} \mathbb{P}_{i=q}\left(\mu^{x,i} \text{ is } (W_i', \rho')\text{-concentrated}\right) &> 1 - 2d\sqrt{c\rho^{1/6}} \\ &> 1 - \delta \end{aligned} \tag{46}$$

(the last inequality by (40)), and

$$\begin{aligned} \frac{1}{n+1} \# \left\{0 \leq q \leq n \,:\, d(W_q', W_{q-\ell}') \leq \rho'\right\} &\geq 1 - \frac{2(d+1)\ell}{\log(1/c\rho^{1/6})} \\ &> 1 - \frac{\delta}{2} \end{aligned}$$

(the last inequality in by assumption (41)). Let

$$J = \{i \in I\,,\, d(W_i', W_{i-\ell}') \leq \rho'\}.$$

Since $\frac{1}{n+1}|I| > 1 - \delta/2$, the previous equation implies that

$$\frac{1}{n+1}|J| \geq 1 - \delta. \tag{47}$$

Now, for any $\ell \leq q \leq n$, applying (46) to $q - \ell$ we have

$$\mathbb{P}_{i=q}\left(\nu^{x,i-\ell} \text{ is } (W_{i-\ell}', \rho')\text{-concentrated}\right) > 1 - \delta.$$

Assuming also $q \in J$, we also have $d(W_{q-\ell}', W_q') \leq \rho'$, so by Lemma 3.21 (1) applied to each component $\nu^{x,i-\ell}$ in the event above,

$$\mathbb{P}_{i=q}\left(\nu^{x,i-\ell} \text{ is } (W_i', (\sqrt{d}+1)\rho')\text{-concentrated}\right) > 1 - \delta \qquad \text{for } q \in J.$$

Our assumption (42) implies that $(\sqrt{d}+1)\rho' < \delta$, and the last inequality yields

$$\mathbb{P}_{i=q}\left(\nu^{x,i-\ell} \text{ is } (W_i', \delta)\text{-concentrated}\right) > 1 - \delta \qquad \text{for } q \in J. \tag{48}$$

On the other hand for $q \in J$ we have $q \in I$ and so (44) holds. Since $W_q' \leq W_q$, by Lemma 3.21 (4), we have

$$\mathbb{P}_{i=q}\left(\mu^{x,i-\ell} \text{ is } (W_i', \frac{\delta}{2} + O(\frac{1}{m}), m)\text{-saturated}\right) > 1 - \delta \qquad \text{for } q \in J.$$

Since we are assuming $m$ large enough relative to $\delta$, this implies

$$\mathbb{P}_{i=q}\left(\mu^{x,i-\ell} \text{ is } (W_i', \delta, m)\text{-saturated}\right) > 1 - \delta \qquad \text{for } q \in J. \tag{49}$$

In conclusion, if we define $V_i = W_{i+\ell}'$ and replace $I$ by $(J - \ell) \cap [0, n]$, then equations (47), (48), and (49) give the desired conclusion, assuming that $m, \delta$ have the appropriate relationship to each other and to $\varepsilon$, and that $n$ and is large enough. $\square$

## 4.5 The Kaĭmanovich-Vershik lemma

The second ingredient in our proof of Theorem 2.8 is the following entropy analog of the Plünnecke-Rusza inequality:

**Lemma 4.15** (Kaĭmanovich-Vershik, [18]). *Let $\Gamma$ be a countable abelian group and let $\mu, \nu \in \mathcal{P}(\Gamma)$ be probability measures with $H(\mu) < \infty$, $H(\nu) < \infty$. Let*

$$\delta_k = H(\mu * (\nu^{*(k+1)})) - H(\mu * (\nu^{*k})).$$

*Then $\delta_k$ is non-increasing in $k$. In particular,*

$$H(\mu * (\nu^{*k})) \leq H(\mu) + k \cdot (H(\mu * \nu) - H(\nu)).$$

This lemma first appears in a study of random walks on groups by Kaĭmanovich and Vershik [18]. It was more recently rediscovered and applied in additive combinatorics by Madiman and co-authors [24, 25], and in a weaker form independently by Tao [31]. For a proof using our notation see [12].

For non-discrete measures in $\mathbb{R}^d$ we have the following analog:

**Corollary 4.16.** *Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ with $H_n(\mu), H_n(\nu) < \infty$. Then*

$$H_n(\mu * (\nu^{*k})) \leq H_n(\mu) + k \cdot (H_n(\mu * \nu) - H_n(\mu)) + O(\frac{k}{n}).$$

The error term arises in the same way as in Lemma 4.1. For the proof see [12] (the passage from $\mathbb{R}$ to $\mathbb{R}^d$ requires only notational changes).

## 4.6 Proof of the inverse theorem

We now prove Theorem 2.8, which we re-state for convenience.

**Theorem 4.17.** *For every $\varepsilon > 0$, $R > 0$ and $m \in \mathbb{N}$, there exists $\delta = \delta(\varepsilon, R, m) > 0$ such that for all $n > n(\varepsilon, R, m, \delta)$, the following holds: if $\nu, \mu \in \mathcal{P}([-R, R]^d)$ and*

$$H_n(\mu * \nu) < H_n(\mu) + \delta,$$

*then there exists a sequence $V_0, \ldots, V_n \leq \mathbb{R}^d$ of subspaces such that*

$$\mathbb{P}_{0 \leq i \leq n} \left( \begin{array}{c} \mu^{x,i} \textit{ is } (V_i, \varepsilon, m)\textit{-saturated and} \\ \nu^{x,i} \textit{ is } (V_i, \varepsilon)\textit{-concentrated} \end{array} \right) \quad > \quad 1 - \varepsilon$$

*Proof.* Fix $\varepsilon > 0$. It is a formal consequence of Proposition 3.19 that it suffices for us to prove the theorem with the assumption that $m$ is large in a manner depending on $\varepsilon$. We can also assume that $\varepsilon < 1/2$, and that $\varepsilon$ is small with respect to $d$. Also, as we are free to choose $n$ large relative to $R$, the distribution on components depends negligibly on dyadic scales greater than 0, and the scale-$n$ entropy of $\mu$ and $\mu * \nu$ differs negligibly from the same entropy conditioned on $\mathcal{D}_0$. Thus, without loss of generality, we can assume that the measures are supported on $[0, 1)^d$, and we omit mention of $R$ from now on.

Choose $k = k_2(\varepsilon, m)$ as in Theorem 4.12. We shall show that the conclusion holds if $n$ is large relative to the previous parameters.

Let $\mu, \nu \in \mathcal{P}([0, 1)^d)$. Denote

$$\tau = \nu^{*k}$$

Assuming $n$ is large enough, Theorem 4.12 provides us with subspaces $V_0,\ldots,V_n \subseteq \mathbb{R}^d$ such that

$$\mathbb{P}_{0\le i\le n}\left(\nu^{x,i} \text{ is } (V_i,\varepsilon)\text{-concentrated}\right) \ge 1-\varepsilon, \tag{50}$$

and

$$\mathbb{P}_{0\le i\le n}\left(\tau^{x,i} \text{ is } (V_i,\varepsilon,m)\text{-saturated}\right) > 1-\varepsilon.$$

If it holds that

$$\mathbb{P}_{0\le i\le n}\left(\mu^{x,i} \text{ is } (V_i,2\varepsilon,m)\text{-saturated}\right) > 1-2\varepsilon \tag{51}$$

then we are done, since (50) and (51) together are the second alternative of the theorem we want to prove (with a multiple of $\varepsilon$ instead of $\varepsilon$, but this is formally equivalent).

Otherwise, by Lemma 4.2 and the above we have

$$\begin{aligned}
&\mathbb{P}_{0\le i\le n}\left(H_m(\mu^{x,i}*\tau^{y,i}) > H_m(\mu^{x,i})+\varepsilon-O(\frac{1}{m})\right)\\
&\quad\ge\quad \mathbb{P}_{0\le i\le n}\left(\mu^{x,i} \text{ is not } (V_i,2\varepsilon,m)\text{-saturated and } \tau^{y,i} \text{ is } (V_i,\varepsilon,m)\text{-saturated}\right)\\
&\quad>\quad \mathbb{P}_{0\le i\le n}\left(\mu^{x,i} \text{ is not } (V_i,2\varepsilon,m)\text{-saturated}\right)\\
&\qquad\quad -\ \left(1-\mathbb{P}_{0\le i\le n}\left(\tau^{y,i} \text{ is } (V_i,\varepsilon,m)\text{-saturated}\right)\right)\\
&\quad>\quad 2\varepsilon-(1-(1-\varepsilon))\\
&\quad=\quad \varepsilon.
\end{aligned}$$

Let $\delta'(\mu^{x,i},\tau^{y,i}) = H_m(\mu^{x,i}*\tau^{y,i}) - H_m(\mu^{x,i})$. By the previous calculation, with probability at least $\varepsilon$ we have $\delta' \ge \varepsilon - O(1/m)$, and by Lemma 4.1 we always have $\delta' \ge -O(1/m)$. Thus

$$\mathbb{E}_{0\le i\le n}\left(\delta'(\mu^{x,i},\tau^{y,i})\right) \ge \varepsilon^2 - O(\frac{1}{m})$$

Thus, by Lemmas 3.5 and 3.6,

$$\begin{aligned}
H_n(\mu*\tau) &> \mathbb{E}_{0\le i<n}\left(H_m(\mu^{x,i}*\tau^{y,i})\right) - O(\frac{m}{n})\\
&\ge \mathbb{E}_{0\le i\le n}\left(H_m(\mu^{x,i}) + \delta'(\mu^{x,i},\tau^{y,i})\right) - O(\frac{m}{n})\\
&\ge \mathbb{E}_{0\le i\le n}\left(H_m(\mu^{x,i})\right) + \varepsilon^2 - O(\frac{1}{m}+\frac{m}{n})\\
&= H_n(\mu) + \varepsilon^2 - O(\frac{1}{m}+\frac{m}{n}).
\end{aligned}$$

So, assuming that $m$ is large and $n$ larger still, all in a manner depending on $\varepsilon, d$, we have

$$H_n(\mu*\tau) > H_n(\mu) + \frac{\varepsilon^2}{2}.$$

On the other hand, by Corollary 4.16 above,

$$H_n(\mu*\tau) \le H_n(\mu) + k\cdot(H_n(\mu*\nu) - H_n(\mu)) + O(\frac{k}{n}).$$

Assuming that $n$ is large enough in a manner depending on $d$, $\varepsilon$ and $k$, this and the previous inequality give

$$H_n(\mu * \nu) \geq H_n(\mu) + \frac{\varepsilon^2}{3k}.$$

This completes the proof of Theorem 2.8 with $\delta = \varepsilon^2/3k$. □

# 5 Inverse theorem for the action of the isometry group on $\mathbb{R}^d$

In this section we prove the inverse theorems for convolutions $\nu . \mu$ for $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $\nu \in \mathcal{P}(G_0)$, where $G_0$ is the group of isometries of $\mathbb{R}^d$. Our strategy is to linearize the action $G_0 \times \mathbb{R}^d \to \mathbb{R}^d$ and apply the Euclidean inverse theorem.

Recall that the elements of $G_0$ are denoted $g = U + a$, with $U$ an orthogonal matrix, $a \in \mathbb{R}^d$, and $gx = Ux + a$. Given $g \in G_0$ we denote the associated matrix and vector by $U_g$ and $a_g$. Also recall that $S_t$ is the scalar map $S_t(x) = 2^t x$, and introduce the translation map

$$\tau_s(x) = x + s.$$

## 5.1 Concentration and saturation on random subspaces

This section contains additional technical results on concentration and saturation of components of a measures. Our first goal is to show that if two measures $\eta, \theta \in \mathcal{P}(\mathbb{R}^d)$ are such that with high probability pairs of components $\eta^{x,i}$, $\theta^{y,i}$ are highly concentrated and saturated, respectively, on a subspace $V = V^{(i,x,y)}$, then we can assume that $V^{(i,x,y)}$ is essentially independent of $x, y$.

**Proposition 5.1.** *For every $\varepsilon > 0$ and $m \in \mathbb{N}$, there are $\varepsilon' = \varepsilon'(\varepsilon, m) \to 0$ and $m' = m'(\varepsilon, m) \to \infty$ as $\varepsilon \to 0$ and $m \to \infty$, such that the following holds. Suppose that $\theta, \eta$ are independent $\mathcal{P}(\mathbb{R}^d)$-valued random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and that $V = V(\theta, \eta) \leq \mathbb{R}^d$ is a linear subspace determined by the random measures $\theta, \eta$ (hence $V$ is random). If*

$$\mathbb{P}\left(\theta \text{ is } (V, \varepsilon', m')\text{-saturated and } \eta \text{ is } (V, \varepsilon')\text{-concentrated}\right) > 1 - \varepsilon'. \tag{52}$$

*Then there is a deterministic subspace $V_*$ such that*

$$\mathbb{P}_{0 \leq i \leq m'}\left(\theta^{x,i} \text{ is } (V_*, \varepsilon, m)\text{-saturated and } \eta^{y,i} \text{ is } (V_*, \varepsilon)\text{-concentrated}\right) > 1 - \varepsilon. \tag{53}$$

*(the probability in the last equation is over both the measures $\theta, \eta$ and $i, x, y$, independently).*

*Proof.* It is a formal consequence of Proposition 3.19 that it is enough to prove the statement under the assumption that $m$ is large relative to $\varepsilon$.

Fix $\varepsilon$. Assume $m, m'$ large relative to $\varepsilon$, and $\varepsilon'$ small relative to the other parameters. We shall show that (52) implies (53).

Apply Proposition 3.29 to $\eta$ with parameter $\varepsilon'$. We obtain a random subspace $V_c = V_c(\eta)$ and a constant $C_c \geq 1$ (which we will later assume is large compared to another constant $D$) such that, for

$$\delta_c = C_c \cdot (\varepsilon')^{1/3^d},$$

the measure $\eta$ is $(V_c, \delta_c)$-concentrated and if $\eta$ is $(W, \varepsilon')$-concentrated then $V_c \sqsubseteq W^{(\delta_c)}$.

Apply Proposition 3.30 to $\theta$ with parameter $m'$. We obtain a random subspace $V_s = V_s(\theta)$ such that, for a constant $C_s$ and

$$\delta_s = C_s \frac{\log(m')}{m'},$$

the measure $\theta$ is $(V_s, \delta_s, m')$-saturated and if $\theta$ is $(W, 1/m', m')$-saturated, then $W \sqsubseteq V_s^{(\delta_s)}$. We shall assume that $\varepsilon' < 1/m'$. Thus, if $\theta$ is $(W, \varepsilon', m')$-saturated then $W \sqsubseteq V_s^{(\delta_s)}$.

The random subspace $V$ satisfies (52), so we have

$$\mathbb{P}\left(V_c \sqsubseteq V^{(\delta_c)} \text{ and } V \sqsubseteq V_s^{(\delta_s)}\right) > 1 - \varepsilon'.$$

Thus, writing

$$\delta = \delta_c + \delta_s,$$

we have

$$\mathbb{P}(V_c \sqsubseteq V_s^{(\delta)}) > 1 - \varepsilon'. \tag{54}$$

Let $\mathcal{W} = \{W_1, \dots, W_N\}$ denote a minimal $\delta_c$-dense sequence of subspaces with respect to the metric (20). This metric is bi-Lipschitz equivalent to a smooth metric on the compact manifold of subspaces, so $N \leq D \cdot \delta_c^{-[d^2/2]}$ for some universal constant $D > 1$ (here $[d^2/2]$ is the dimension of the space of subspaces). Let

$$\mathcal{W}_0 = \{W \in \mathcal{W} : \mathbb{P}(d(V_c, W) < \delta_c) > \frac{\delta_c}{N}\}.$$

Apply Proposition 3.27 to $\mathcal{W}_0$ with parameter $2\delta$ to obtain the parameter

$$\delta' = 4 \cdot 2^{1/3^d} \cdot \delta^{1/3^d}$$

and a non-trivial subspace $V_*$ such that

a. $W \sqsubseteq V_*^{(\delta')}$ for all $W \in \mathcal{W}_0$,

b. If $\widetilde{V}_*$ is another subspace such that $W \sqsubseteq \widetilde{V}_*^{(2\delta)}$ for all $W \in \mathcal{W}_0$, then $V_* \sqsubseteq \widetilde{V}_*^{(2\delta')}$.

We claim that $V_*$ is the desired subspace. Writing $\mathcal{W}_1 = \mathcal{W} \setminus \mathcal{W}_0$,

$$\begin{aligned}
\mathbb{P}(d(V_c, W) \geq \delta_c \text{ for all } W \in \mathcal{W}_0) &= P(V_c \notin \bigcup_{W \in \mathcal{W}_0} B_{\delta_c}(W)) \\
&\leq \mathbb{P}(V_c \in \bigcup_{W \in \mathcal{W}_1} B_{\delta_c}(W)) \\
&\leq \sum_{W \in \mathcal{W}_1} \mathbb{P}(V_c \in B_{\delta_c}(W)) \\
&\leq |\mathcal{W}_1| \cdot \frac{\delta_c}{N} \\
&< \delta_c,
\end{aligned}$$

where in the first inequality we used the fact that $\bigcup_{W\in\mathcal{W}_0\cup\mathcal{W}_1} B_{\delta_c}(W)$ covers all subspaces, and in the last line we used $|\mathcal{W}_1| \leq |\mathcal{W}| = N$. Hence

$$\mathbb{P}(d(V_c, W) < \delta_c \text{ for some } W \in \mathcal{W}_0) > 1 - \delta_c.$$

Consequently, by property (a) of $V_*$ and the fact that $\delta_c \leq \delta'$,

$$\mathbb{P}(V_c \sqsubseteq V_*^{(2\delta')}) > 1 - \delta_c. \tag{55}$$

Since $V_c$ is a function of $\eta$ and $V_s$ is a function of $\theta$, and since $\eta, \theta$ are independent, also each of the pairs $V_c, \theta$ and $V_c, V_s$ is independent. Therefore, for a.e. value $\theta_0$ of $\theta$,

$$\mathbb{P}(d(V_c, W) < \delta_c | \theta = \theta_0) = \mathbb{P}(d(V_c, W) < \delta_c) > \frac{\delta_c}{N} \qquad \text{for all } W \in \mathcal{W}_0. \tag{56}$$

Observe that

$$\begin{aligned}
\frac{\delta_c}{N} &\geq \frac{1}{D}\delta_c^{1+[d^2/2]} \\
&= \frac{C_c^{1+[d^2/2]}}{D}(\varepsilon')^{(1+[d^2/2])/3^d} \\
&> (\varepsilon')^{1/2},
\end{aligned}$$

where, to justify the last inequality, we increase the constant $C_c$ if necessary to ensure $C_c^{1+[d^2/2]}/D \geq 1$, and note that $(1 + [d^2/2])/3^d < 1/2$. Thus if a fixed measure $\theta_0$ satisfies

$$\mathbb{P}(V_c \sqsubseteq V_s^{(\delta)} | \theta = \theta_0) > 1 - \sqrt{\varepsilon'} \tag{57}$$

then, by (56) and (57), for all $W \in \mathcal{W}_0$,

$$\begin{aligned}
\mathbb{P}(W \sqsubseteq V_s^{(2\delta)} | \theta = \theta_0) &\geq \mathbb{P}(d(V_c, W) < \delta_c \text{ and } V_c \sqsubseteq V_s^{(\delta)} | \theta = \theta_0) \\
&\geq \mathbb{P}(d(V_c \sqsubseteq V_s^{(\delta)} | \theta = \theta_0) - (1 - \mathbb{P}(d(V_c, W) < \delta_c | \theta = \theta_0)) \\
&> (1 - \sqrt{\varepsilon'}) - (1 - \sqrt{\varepsilon'}) \\
&= 0,
\end{aligned}$$

Since $V_s$ is a function of $\theta$, this says that for $\theta_0$ satisfying (57) we have $W \sqsubseteq V_s^{(2\delta)}$ for each $W \in \mathcal{W}_0$; consequently, by property (b) of the definition of $V_*$, for such $\theta_0$ we have that $V_* \sqsubseteq V_s^{(2\delta')}$.

By Markov's inequality and (54), the relation (57) holds with probability $1 - \sqrt{\varepsilon'}$ over the choice of $\theta_0$. Thus we conclude

$$\mathbb{P}(V_* \sqsubseteq V_s^{(2\delta')}) > 1 - \sqrt{\varepsilon'}. \tag{58}$$

Combining (55) and (58) and using $\sqrt{\varepsilon'} \leq \delta_c$, we find that

$$\mathbb{P}(V_c \sqsubseteq V_*^{(2\delta')} \text{ and } V_* \sqsubseteq V_s^{(2\delta')}) > 1 - 2\delta_c.$$

Finally, fix $\eta, \theta$ and associated to $V_s, V_c$ belonging to this event, we have that $\eta$ is $(V_c, \delta_c)$-concentrated. Therefore by Lemma 3.17,

$$\mathbb{P}_{0\leq i\leq m'}\left(\eta^{x,i} \text{ is } (V_c, \sqrt{2^i\delta_c})\text{-concentrated}\right) > 1 - \sqrt{\delta_c}$$

(here and below the randomness is over $i$, with $\eta, \theta$ fixed). Since $V_c \sqsubseteq V_*^{(2\delta')}$, and assuming as we may that $\varepsilon'$, and hence $\delta_c$, is small enough relative to $\varepsilon, m'$, this implies

$$\mathbb{P}_{0\le i\le m'}\left(\eta^{x,i} \text{ is } (V_*, \varepsilon)\text{-concentrated}\right) > 1 - \frac{\varepsilon}{3}. \tag{59}$$

Similarly, $\theta$ is $(V_s, \delta_s, m')$-saturated, so arguing in the same manner and using Lemma 3.16,

$$\mathbb{P}_{0\le i\le m'}\left(\theta^{y,i} \text{ is } (V_s, \sqrt{d\delta_s + O(\frac{m}{m'})}, m)\text{-saturated}\right) > 1 - O(\sqrt{\delta_s + \frac{m}{m'}}).$$

Assuming $\varepsilon'$ is small enough and $m'$ large enough relative to $\varepsilon, m$, the constant $\delta'$ can be assumed arbitrarily small compared to $\varepsilon, m$. Since $V_* \sqsubseteq V_s^{(2\delta')}$ we have $d(V_*, \pi_{V_s} V_*) < 2\delta'$, so by Lemma 3.21 and (4), and assuming the parameters satisfy the appropriate relationship, we have

$$\mathbb{P}_{0\le i\le m'}\left(\theta^{y,i} \text{ is } (V_*, \varepsilon, m)\text{-saturated}\right) > 1 - \frac{\varepsilon}{3}. \tag{60}$$

Thus, combining (59) and (60) for $\eta, \theta$ in the event in (58), we have

$$\mathbb{P}_{0\le i\le m'}\left(\theta^{y,i} \text{ is } (V_*, \varepsilon, m)\text{-saturated and } \eta^{x,i} \text{ is } (V_*, \varepsilon)\text{-concentrated}\right) > 1 - \frac{2}{3}\varepsilon.$$

Using (58) and assuming as we may that $\sqrt{\varepsilon'} < \varepsilon/3$, we obtain (53). ☐

**Corollary 5.2.** *Let $\varepsilon > 0$ and $m \in \mathbb{N}$. Then there exist $\varepsilon'' = \varepsilon''(\varepsilon, m) \to 0$ and $m'' = m''(\varepsilon, m) \to \infty$ as $\varepsilon \to 0$ and $m \to \infty$, such that for all large enough $n$, the following holds. Suppose that we are given subspaces $V^{(i,x,y)}$ for $0 \le i \le n$ and $x, y \in [0,1]^d$ and measures $\theta, \eta \in \mathcal{P}([0,1]^d)$ such that*

$$\mathbb{P}_{0\le i\le n}\left(\begin{array}{c}\theta^{x,i} \textit{ is } (V^{(i,x,y)}, \varepsilon'', m'')\textit{-saturated and}\\ \eta^{y,i} \textit{ is } (V^{(i,x,y)}, \varepsilon'')\textit{-concentrated}\end{array}\right) > 1 - \varepsilon'',$$

*then there are subspaces $V^i \le \mathbb{R}^d$ such that*

$$\mathbb{P}_{0\le i\le n}\left(\begin{array}{c}\theta^{x,i} \textit{ is } (V^i, \varepsilon, m)\textit{-saturated and}\\ \eta^{y,i} \textit{ is } (V^i, \varepsilon)\textit{-concentrated}\end{array}\right) > 1 - \varepsilon.$$

*Proof.* Apply the previous proposition to obtain $\varepsilon' = \varepsilon'(\frac{1}{2}\varepsilon, m)$ and $m' = m'(\frac{1}{2}\varepsilon, m)$, and set $m'' = m'$ and $\varepsilon'' = \min\{(\varepsilon')^2, 1/m'\}$.

For $0 \le k \le n$, let

$$p_k = \mathbb{P}_{i=k}\left(\begin{array}{c}\theta^{x,i} \text{ is } (V^{(i,x,y)}, \varepsilon'', m'')\text{-saturated and}\\ \eta^{y,i} \text{ is } (V^{(i,x,y)}, \varepsilon'')\text{-concentrated}\end{array}\right),$$

and assume as in the hypothesis that $\frac{1}{n+1}\sum_{k=0}^{n} p_k > 1 - \varepsilon''$. Let $I \subseteq \{0, \ldots, n\}$ denote the set of $k$ such that $p_k > 1 - \sqrt{\varepsilon''} = 1 - \varepsilon'$, so by Markov, $|I| > (1 - \varepsilon')(n+1)$.

For $i \in I$, consider the random and independently chosen components $\theta^{x,i}$, $\eta^{y,i}$ and the subspace $V_i = V^{i,x,y}$. Without loss of generality we may assume that $V^{i,x,y}$ depend only on $\theta^{x,i}$ and $\eta^{y,i}$, since the only stated property of $V^{i,x,y}$

involves these measures. From the previous proposition and our choice of $\varepsilon', m'$, we conclude that there exists a subspace $V^i$ such that

$$\mathbb{P}_{i\leq j\leq i+m'}\left(\begin{array}{c}\theta^{x,i} \text{ is } (V^i, \frac{1}{2}\varepsilon, m)\text{-saturated and} \\ \eta^{y,i} \text{ is } (V^i, \frac{1}{2}\varepsilon)\text{-concentrated}\end{array}\right) > 1 - \frac{\varepsilon}{2}.$$

The remainder of the argument involves choosing one of these subspaces $V^{i(j)}$, for every $j \in \bigcup_{u\in I}[u, u+m']$. The details are identical to the proof of Proposition 3.19. $\square$

We will actually need a more general version of the last corollary, but, as the proof is identical to the one above, we only give the statement.

**Corollary 5.3.** *Let $\varepsilon > 0$ and $m \in \mathbb{N}$. Then there exist $\varepsilon'' = \varepsilon''(\varepsilon, m) \to 0$ and $m'' = m''(\varepsilon, m) \to \infty$ as $\varepsilon \to 0$, such that for all large enough $n$, the following holds. Suppose that $\theta \in \mathcal{P}(G_0)$, $\eta \in \mathcal{P}(\mathbb{R}^d)$, and that for $0 \leq i \leq n$, $x \in \operatorname{supp}\eta$ and $g \in \operatorname{supp}\theta$ there are subspaces $V^{(i,x,g)}$ such that*

$$\mathbb{P}_{0\leq i\leq n}\left(\begin{array}{c}\eta^{x,i} \textit{ is } (V^{(i,x,y)}, \varepsilon'', m'')\textit{-saturated and} \\ S_i U_g^{-1}(\theta_{g,i}\boldsymbol{.}x) \textit{ is } (V^{(i,x,y)}, \varepsilon'')\textit{-concentrated}\end{array}\right) > 1 - \varepsilon''.$$

*Then there are subspaces $V^i \leq \mathbb{R}^d$ such that*

$$\mathbb{P}_{0\leq i\leq n}\left(\begin{array}{c}\eta^{x,i} \textit{ is } (V^i, \varepsilon, m)\textit{-saturated and} \\ S_i U_g^{-1}(\theta^{y,i}\boldsymbol{.}x) \textit{ is } (V^i, \varepsilon)\textit{-concentrated}\end{array}\right) > 1 - \varepsilon.$$

## 5.2 From concentration of Euclidean components to $G_0$-components

We turn our attention to measures $\eta \in \mathcal{P}(\mathbb{R}^d)$ of the form $\eta = \theta\boldsymbol{.}x$ for some $\theta \in \mathcal{P}(G_0)$ and $x_0 \in \mathbb{R}^d$. Our goal is to show that the concentration properties of typical components $\eta^{y,i}$ translates to similar properties of the "components" $\theta_{g,i}\boldsymbol{.}x$. The issue which we must overcome is that $\theta_{g,i}\boldsymbol{.}x$ is supported on $\mathcal{D}_i^G(g)\boldsymbol{.}x$, and this set generally intersects more than one dyadic cell of $\mathcal{D}_i^d$. Thus even if $\eta$ is highly concentrated on a translate of a subspace $W$ on each of these cells, taken together all one can say is that $\theta_{g,i}\boldsymbol{.}x$ is concentrated on the union of several translates of $W$.

For a linear subspace $W \leq \mathbb{R}^d$ we say that a measure $\eta \in \mathcal{P}(\mathbb{R}^d)$ is $(W,\delta)^m$-concentrated if for some $m' \leq m$ there are $m'$ translates $W_1, \ldots, W_{m'}$ of $W$ such that $\eta(\bigcup_{u=1}^{m'} W_u^{(\delta)}) \geq 1 - \delta$. Thus $(W,\delta)^1$-concentration is the same as $(W,\delta)$-concentration.

**Lemma 5.4.** *Let $R > 0$ , let $\theta \in \mathcal{P}(G_0)$ and $x \in [-R, R^d]$. Suppose that $\delta > 0$, $m \in \mathbb{N}$ and that $\theta\boldsymbol{.}x$ is $(W,\delta)^m$-concentrated. Then for* $n = [\frac{1}{2}\log(1/\delta)]$ and $\delta' = O_{R,m}(\frac{\log\log(1/\delta)}{\log(1/\delta)})$ *we have*

$$\mathbb{P}_{0\leq i\leq n}\left(S_i(\theta_{g,i}\boldsymbol{.}x) \textit{ is } (W,\delta')\textit{-concentrated}\right) > 1 - \delta'.$$

*Proof.* Although the "rescaled component" $\theta^{g,i}$ is not defined, it will be convenient to introduce the notation

$$\theta^{g,i}\boldsymbol{.}x = S_i(\theta_{g,i}\boldsymbol{.}x),$$

and define the distribution on these "components" in the usual manner. Note that there is a constant $C = C(R) \geq 1$ such that $\theta_{g,i}\mathbin{.}x$ is supported on a set of diameter $\leq C2^{-i}$, and $\theta^{g,i}\mathbin{.}x$ is supported on a set of diameter $\leq C$.

Let $W_1, \ldots, W_m$ be affine subspaces parallel to $W$ verifying that $\theta\mathbin{.}x$ is $(W,\delta)^m$-concentrated. We may assume the $W_u$ are distinct. For $u \neq v$ let

$$d_{u,v} = d(W_u, W_v) = \min\{d(x,y) \,:\, x \in W_u \,,\, y \in W_v\}.$$

Notice that for any $1 \leq k \leq n$,

- If $2^{-k} < \frac{1}{4C} d_{u,v}$ for some $u, v$, then for any $g$ the measure $\theta_{g,k}\mathbin{.}x$ is supported on a set of diameter at most $C2^{-k} < \frac{1}{4} d_{u,v}$, and on the other hand $\sqrt{\delta} \leq 2^{-n} \leq 2^{-k} \leq \frac{1}{4} d_{u,v}$, hence $\theta_{g,k}\mathbin{.}x$ gives positive mass to at most one the sets $W_u^{(\sqrt{\delta})}, W_v^{(\sqrt{\delta})}$.
- Let $I_{g,k} \subseteq \{1, \ldots, m\}$ be the set of indices $u$ such that $(\theta_{g,k}\mathbin{.}x)(W_u^{(\delta)}) > 0$. Given $\rho > 0$, if all distinct $u, v \in I_{g,k}$ satisfy $d_{u,v} \leq \rho 2^{-k}$, then there is a translate $W_{g,k}$ of $W$ such that $\bigcup_{u \in I_{g,k}} W_u^{(\delta)} \cap \mathrm{supp}(\theta_{g,k}\mathbin{.}x) \subseteq W_{g,k}^{(\rho 2^{-k} + 2\delta)}$, so $\theta^{g,k}\mathbin{.}x$ is $(W, \rho + 2^{k+1}\delta))$-concentrated.

Now, for $0 \leq k \leq n$ we have identity

$$\theta\mathbin{.}x = \mathbb{E}_{i=k}\left(\theta_{g,i}\mathbin{.}x\right).$$

Using the hypothesis that $(\theta\mathbin{.}x)(\bigcup_{u=1}^m W_u^{(\delta)}) > 1 - \delta$ and Markov's inequality we conclude that

$$\mathbb{P}_{i=k}\left((\theta_{g,i}\mathbin{.}x)(\bigcup_{u=1}^{m} W_u^{(\delta)}) > 1 - \sqrt{\delta}\right) > 1 - \sqrt{\delta}. \tag{61}$$

Fix a small parameter $\rho > 0$, and suppose that $k$ satisfies

$$\text{For each } 1 \leq u < v \leq m \text{ either } 2^{-k} < \frac{1}{4C} d_{u,v} \text{ or } d_{u,v} \leq \rho 2^{-k}, \tag{62}$$

or, equivalently, that $k$ does not belong to any of the intervals $J_{u,v} = [\log \frac{\rho}{d_{u,v}}, \log \frac{4C}{d_{u,v}})$. Then, setting

$$\sigma = \sigma(\rho) = \max\{\sqrt{\delta}, \rho + 2^{k+1}\delta\}$$

the two observations above and (61) imply

$$\mathbb{P}_{i=k}\left(\theta^{g,k}\mathbin{.}x \text{ is } (W,\sigma)\text{-concentrated}\right) > 1 - \sqrt{\delta}.$$

Note that $\delta 2^k \leq \delta 2^n \leq \sqrt{\delta}$, so in fact $\sigma \leq \rho + 2\sqrt{\delta}$.

Next, since the length of $J_{u,v}$ is $\log \frac{4C}{\rho}$ and there are at most $m(m-1)$ distinct values of $1 \leq u, v \leq m$, the fraction of $0 \leq k \leq n$ which satisfy (62) is at least $1 - m^2 \log(\frac{4C}{\rho})/n$. Averaging the last equation over $k = 0, \ldots, n$, we conclude that

$$\begin{aligned}\mathbb{P}_{0 \leq k \leq n}\left(\theta^{g,k}\mathbin{.}x \text{ is } (W, \rho + \sqrt{\delta})\text{-concentrated}\right) &> 1 - \sqrt{\delta} - \frac{m^2 \log(4C/\rho)}{n} \\ &= 1 - \sqrt{\delta} - O_m\left(\frac{\log(1/\rho)}{\log(1/\delta)}\right).\end{aligned}$$

Choosing $\rho = \frac{1}{\log(1/\delta)}$ gives the desired result. $\square$

**Proposition 5.5.** *For every $\varepsilon > 0$ and $R > 0$ there exists $n = n(\varepsilon, R)$ (with $n(\varepsilon, R) \to \infty$ as $\varepsilon \to 0$) and an $\delta = \delta(\varepsilon, R) > 0$ (with $\delta(\varepsilon, R) \to 0$ as $\varepsilon \to 0$) such that the following holds. Let $\nu \in \mathcal{P}(G_0)$ and $x_0 \in [-R, R]^d$, and write $\eta = \nu . x_0$. Let $V < \mathbb{R}^d$ be a linear subspace and $k \in \mathbb{N}$ such that*

$$\mathbb{P}_{j=k} \left( \eta^{x,j} \text{ is } (V, \delta)\text{-concentrated} \right) > 1 - \delta.$$

*Then*

$$\mathbb{P}_{k \le j \le k+n} \left( S_j(\nu_{g,j} . x) \text{ is } (V, \varepsilon)\text{-concentrated} \right) > 1 - \varepsilon.$$

*Proof.* Fix $\varepsilon, n, \delta$ for the moment and assume that the hypothesis holds. Consider the identities

$$\mathbb{E}_{j=k}(\eta_{x,j}) = \eta = \mathbb{E}_{j=k}(\nu_{g,j} . x).$$

This means that the measures $\nu_{g,k} . x$ are ($\nu$-almost-surely over choice of $g$) absolutely continuous with respect to the weighted average of the components $\eta_{x,k}$. In fact, since each $\nu_{g,k} . x$ is supported on a set that intersects $m = O_R(1)$ level-$k$ dyadic cells, each "component" $\nu_{g,k} . x$ is absolutely continuous with respect to the average of these $O(1)$ components $\eta_{x,k}$. Most of these components are $(V, \delta)$-concentrated, so a Markov-inequality argument (similar to the one in Lemma 6.5 below) shows that

$$\mathbb{P}_{j=k} \left( \nu_{g,j} . x \text{ is } (V, 2^{-k}\delta')^m\text{-concentrated} \right) > 1 - \delta',$$

where $\delta' \to 0$ as $\delta \to 0$. Equivalently,

$$\mathbb{P}_{j=k} \left( S_j(\nu_{g,j} . x) \text{ is } (V, \delta')^m\text{-concentrated} \right) > 1 - \delta'.$$

Apply the previous lemma to each component $\theta = \nu_{g,k}$ in the event above with parameter $\delta'$. Taking $\delta'' = O_{R,m}(\log\log(1/\delta')/\log(1/\delta'))$ and $n = [\frac{1}{2}\log(1/\delta')]$, the conclusion is

$$\mathbb{P}_{k \le i \le k+n} \left( S_i(\nu_{g,i} . x) \text{ is } (V, \delta'')\text{-concentrated} \right) > 1 - \delta'',$$

and $\delta''$ can be made arbitrarily small by taking $\delta$ small. This is what was claimed. ∎

**Proposition 5.6.** *For every $\delta > 0$, $R > 0$ and $m \in \mathbb{N}$, if $m' > m'(\delta, m, R)$, $0 < \delta' < \delta'(\delta, m, R)$, then for all large enough $n$ (depending on previous parameters), the following holds. Let $\mu \in \mathcal{P}(\mathbb{R}^d)$, $\nu \in \mathcal{P}(G_0)$ and $x_0 \in \mathbb{R}^d$, and write $\eta = \nu . x_0$. Let $V_0, V_1, \ldots, V_n \le \mathbb{R}^d$ be linear subspaces, and suppose that*

$$\mathbb{P}_{0 \le j \le n} \left( \begin{array}{c} \mu^{x,j} \text{ is } (V_j, \delta', m')\text{-saturated and} \\ \eta^{y,j} \text{ is } (V_j, \delta')\text{-concentrated} \end{array} \right) > 1 - \delta'.$$

*Then there are subspaces $V_0', V_1', \ldots, V_n'$ such that*

$$\mathbb{P}_{0 \le j \le n} \left( \begin{array}{c} \mu^{x,j} \text{ is } (V_j', \delta, m)\text{-saturated and} \\ S_j(\nu_{g,j} . x_0) \text{ is } (V_j', \delta)\text{-concentrated} \end{array} \right) > 1 - \delta.$$

*Proof.* Let $\delta, R, m$ be given. Fix a small auxiliary parameter $\delta_1$ which we will specify later and let $m_1$ be the number $n(\delta_1, R)$ from the previous proposition, in particular it can be made arbitrarily large by making $\delta_1$ small. Let $\varepsilon = \varepsilon(\delta_1)$ as in the previous proposition, and let $\delta_2 = \frac{1}{2}\varepsilon^2$. Then for all small enough $\delta'$

and all large enough $m'$, the hypothesis implies, by Proposition 3.19, that there are subspaces $V''_j$ such that

$$\mathbb{P}_{0\le j\le n}\left(\begin{array}{c}\mu^{x,j}\text{ is }(V''_j,\delta_2,m_1)\text{-saturated and}\\ \eta^{y,j}\text{ is }(V''_j,\delta_2)\text{-concentrated}\end{array}\right)>1-\delta_2.$$

By Markov's inequality, the set $I\subseteq\{0,\dots,n\}$ consisting of $k$ such that

$$\mathbb{P}_{j=k}\left(\begin{array}{c}\mu^{x,j}\text{ is }(V''_j,\delta_2,m_1)\text{-saturated and}\\ \eta^{y,j}\text{ is }(V''_j,\delta_2)\text{-concentrated}\end{array}\right)>1-\sqrt{\delta_2}$$

has size $|I|\ge(1-\sqrt{\delta_2})(n+1)$. Since $\delta_2<\varepsilon^2$, by our choice of $m_1$ and $\varepsilon$, for each $k\in I$ we have

$$\mathbb{P}_{k\le j\le k+m_1}\left(S_j(\nu_{g,j}\boldsymbol{.}x_0)\text{ is }(V''_k,\delta_1)\text{-concentrated}\right)>1-\delta_1.$$

Also, applying Lemma 3.16 to each $(V''_k,\delta_2,m_1)$-saturated component $\mu^{x,k}$ of $\mu$, we find that

$$\mathbb{P}_{k\le j\le k+m_1}\left(\mu^{x,j}\text{ is }(V''_k,\sqrt{d\sqrt{\delta_2}+O(\frac{m}{m_1})},m)\text{-saturated}\right)>1-\sqrt{d\sqrt{\delta_2}+O(\frac{m}{m_1})}.$$

Now, by choosing $\delta_1$ small enough we can ensure that $\varepsilon$ and $\delta_2$ is small, and $m_1$ is large, relative to $\delta,m$. With suitable choices, one now argues as in the proof of Proposition 3.19 to combine the last two equations over all $k\in I$ and define $V'_j$ with the desired properties. □

## 5.3 Entropy and the $G_0$-action on $\mathbb{R}^d$

For $g=U+a$ and $g'=U'+a'$ in $G_0$ and $x,x'\in\mathbb{R}^d$,

$$g'x'-gx=(U'-U)x+U'(x'-x)+(a'-a). \tag{63}$$

In particular

$$\|gx-g'x'\|\le\|U-U'\|\,\|x\|+\|U'\|\,\|x-x'\|+\|a-a'\|,$$

so if $g,g'$ are in a common level-$k$ dyadic cell and $x,x'\in[-R,R]^d$ are in a common level-$k$ dyadic cell, then $\|gx-g'x'\|=O_R(2^{-k})$. In particular if $\nu\in\mathcal{P}(G_0)$ and $\mu\in\mathcal{P}([-R,R]^d)$ are both supported on level-$k$ dyadic cells, then $\nu\boldsymbol{.}\mu$ is supported on a set of diameter $O_R(2^{-k})$.

For a probability measure $\theta$ on $\mathbb{R}^d$ or $G_0$ it will be convenient in this section to write

$$H_{i,n}(\theta)=\frac{1}{n}H(\theta,\mathcal{D}_{i+n}).$$

(This differs from $H_{i+n}(\theta)$ because we normalize by $1/n$ instead of $1/(i+n)$). In particular $H_n(\theta)=H_{0,n}(\theta)$. By the previous paragraph, if $\theta\in\mathcal{P}(\mathbb{R}^d)$ and $\nu\in\mathcal{P}(G_0)$ are supported on level-$i$ dyadic cells then $\nu\boldsymbol{.}\theta$ is supported on $O(1)$ level-$i$ dyadic cells, so

$$H_{i,n}(\nu\boldsymbol{.}\theta)=\frac{1}{n}H(\nu\boldsymbol{.}\theta,\mathcal{D}_{i+n}|\mathcal{D}_i)+O(\frac{1}{n}).$$

Also observe that for $\theta$ as above, $H_{i,n}(\theta)=H_n(S_i\theta)+O(1/n)$ (Lemma 3.1 (5)).

We now address the issue, described in Section 2.5, of pairs $\nu\in\mathcal{P}(G_0)$ and $x\in\mathbb{R}^d$ such that $\nu$ has substantial entropy but $\nu\boldsymbol{.}x$ does not (e.g. because $\nu$ is supported close to $\mathrm{stab}_{G_0}(x)$.

**Definition 5.7.** For $\sigma > 0$ we say that $x_1, \ldots, x_{d+1} \in \mathbb{R}^d$ are $\sigma$-independent if each $x_i$ is at distance at least $\sigma$ from the affine subspace spanned by the others.

The action of an element $g \in G$ is determined by its action on any $(d+1)$-tuple of affinely independent vectors in $\mathbb{R}^d$, in particular of any $\sigma$-independent $(d+1)$-tuple.

**Proposition 5.8.** *For every $\varepsilon, \sigma, R > 0$, and $k \in \mathbb{Z}$ and $m \in \mathbb{N}$, the following holds. For every $\sigma$-independent sequence $x_1, \ldots, x_{d+1} \in [-R, R]^d$ and every $\nu \in \mathcal{P}(G_0)$ that is supported on a level-$k$ dyadic cell, if*

$$H_{k,m}(\nu . x_i) < \varepsilon \quad \text{for all } i = 1, \ldots, d+1,$$

*then*

$$H_{k,m}(\nu) < (d+1)\varepsilon + O_{\sigma,R}(\frac{1}{m}).$$

*Proof.* Since $\nu$ is supported on a level-$k$ dyadic cell, each $\nu . x_i$ is supported on $O(1)$ level-$k$ dyadic cells, and therefore

$$H(\nu . x_i, \mathcal{D}_k) = O(1)$$

Thus the hypothesis is $\frac{1}{m} H(\nu . x_i, \mathcal{D}_{k+m}) < \varepsilon$ for $i = 1, \ldots, d+1$, and it is enough to prove that $\frac{1}{m} H(\nu, \mathcal{D}^G_{k+m}) < (d+1)\varepsilon + O_{\sigma,R}(1/m)$.

Define the map $f : G_0 \to (\mathbb{R}^d)^{(d+1)}$ by $g \to (gx_1, \ldots, gx_{d+1})$. Then $f$ is a diffeomorphism and one may easily verify that $f$ is uniformly bi-Lipschitz with its image,[12] with Lipschitz constants of $f$ and $f^{-1}$ depending only on $\sigma$ and $R$. Thus (e.g. by Lemma 3.2 (2) applied to $f^{-1}\mathcal{D}^{d(d+1)}_{k+m}$ and $\mathcal{D}^G_{k+m}$),

$$|\frac{1}{m} H(f\nu, \mathcal{D}^{d(d+1)}_{k+m}) - \frac{1}{m} H(\nu, \mathcal{D}^G_{k+m})| = O_{\sigma,R}(\frac{1}{m}).$$

Let $\pi_i : (\mathbb{R}^d)^{d+1} \to \mathbb{R}^d$ denote the projection to the $i$-th copy of $\mathbb{R}^d$. Then $\nu . x_i = \pi_i(f\nu)$. Therefore, if $\frac{1}{m} H(\nu . x_i, \mathcal{D}_{k+m}) < \varepsilon$ for all $i = 1, \ldots, d+1$, then $\frac{1}{m} H(\pi_i f\nu, \mathcal{D}_{k+m}) < \varepsilon$ for all $i = 1, \ldots, d+1$, and so $\frac{1}{m} H(f\nu, \mathcal{D}^{d(d+1)}_{k+m}) \leq (d+1)\varepsilon$ (because $\mathcal{D}^{d(d+1)}_{k+m} = \bigvee \pi_i^{-1} \mathcal{D}^d_{k+m}$, and using Lemma 3.1 (4)). The claim follows. □

Recall the definition of $(\varepsilon, \sigma)$-non-affine measures, Definition 2.11.

**Lemma 5.9.** *If $\mu \in \mathcal{P}(\mathbb{R}^d)$ is $(\varepsilon, \sigma)$-non-affine and $A \subseteq \mathbb{R}^d$ is a Borel set with $\mu(A) > ((d+1)\varepsilon)^{1/(d+1)}$, then there exists a $\sigma$-independent sequence $x_1, \ldots, x_{d+1} \in A$.*

*Proof.* Let $X_1, \ldots, X_{d+1}$ be independent $\mathbb{R}^d$-valued random variables, each distributed according to $\mu$. Let $V_i$ be the (random) affine subspace spanned by the $d$ vectors $\{X_j\}_{j \neq i}$. For each $i$ the vector $X_i$ is independent of $V_i$, and $X_i$ is distributed according to $\mu$, so, since $\mu$ is $(\varepsilon, \sigma)$-non-affine,

$$\mathbb{P}(X_i \notin V_i^{(\sigma)}) = \mu(\mathbb{R}^d \setminus V_i^{(\sigma)}) > 1 - \varepsilon.$$

[12] This fact depends of course on the metric with which we endowed $G_0$. In general when applying this type of argument to a non-compact group this is one point where the choice of metric must be carefully considered.

This implies

$$\mathbb{P}(X_i \notin V_i^{(\sigma)} \text{ for all } i = 1, \ldots, d+1) > 1 - (d+1)\varepsilon.$$

Therefore, if $\mu(A) > ((d+1)\varepsilon)^{1/(d+1)}$,

$$\begin{aligned}\mathbb{P}(X_i \notin V_i^{(\sigma)} \text{ and } X_i \in A \text{ for all } i = 1, \ldots, d+1) &\\ &\geq \mathbb{P}(X_i \in A \text{ for all } i) - (d+1)\varepsilon \\ &\geq \mu(A)^{d+1} - (d+1) \\ &> 0.\end{aligned}$$

Any realization $X_1, \ldots, X_{d+1}$ from the event above is $\sigma$-independent. □

**Corollary 5.10.** *Let $k \in \mathbb{Z}$ and let $\nu \in \mathcal{P}(G_0)$ be supported on a level-$k$ dyadic cell. Then for every $\varepsilon, \sigma, R > 0$, every $(\varepsilon, \sigma)$-non-affine measure $\mu \in \mathcal{P}([-R, R]^d)$, and for every $m \in \mathbb{N}$,*

$$\mu\left(x \in \mathbb{R}^d \,:\, H_{k,m}(\nu . x) > \frac{1}{d+1} H_{k,m}(\nu) - O_{\sigma,R}(\frac{1}{m})\right) > 1 - ((d+1)\varepsilon)^{1/(d+1)}.$$

*Proof.* Let $c = c(\sigma, R)$ denote the constant in the error term of Proposition 5.8. Let $A = \{x \in \mathbb{R}^d \,:\, H_{k,m}(\nu . x) \leq \frac{1}{d+1} H_{k,m}(\nu) - \frac{c}{m}\}$, we claim that $\mu(A) \leq ((d+1)\varepsilon)^{1/(d+1)}$. Otherwise, by the previous lemma, there is an $(\varepsilon, \sigma)$-non-affine tuple $x_1, \ldots, x_{d+1} \in A$. By the Proposition 5.8 applied to $x_1, \ldots, x_{d+1}$ and using the definition of $A$ we have

$$H_{k,m}(\nu) < (d+1)(\frac{1}{d+1} H_{k,m}(\nu) - \frac{c}{m}) + \frac{c}{m} < H_{k,m}(\nu),$$

which is a contradiction. □

## 5.4 Linearization of the $G_0$-action

Next we utilize the differentiability of the action of $G_0 \times \mathbb{R}^d \to \mathbb{R}^d$, which implies that at small scales, a convolution $\nu . \mu$ of $\nu \in \mathcal{P}(G)$ and $\mu \in \mathcal{P}(\mathbb{R}^d)$ can be well approximated by a Euclidean convolution. Since it is easy to give an elementary argument, we do so.

Let $g_0 = U_0 + a_0$ and $g = U + a$ be elements of $G_0$ and $x_0, x \in \mathbb{R}^d$. Then we have the identity

$$g . x = g . x_0 + U_0(x - x_0) + (U - U_0)(x - x_0) \tag{64}$$

Assuming further that $g, g_0$ belong to a common level-$k$ dyadic cell in $G_0$ and $x, x_0$ belong to a common level-$k$ dyadic cell in $\mathbb{R}^d$, we have $\|U - U_0\| = \|x - x_0\| = O(2^{-k})$, so

$$g . x \;=\; g . x_0 + U_0(x - x_0) + O(2^{-2k}). \tag{65}$$

Recall that $\tau_z(y) = y + z$ is the translation map. It follows from the above that if $\nu \in \mathcal{P}(G_0)$ and $\mu \in \mathcal{P}(\mathbb{R}^d)$ are supported on the level-$k$ dyadic cells containing

$g_0, x_0$ respectively, then for $f \in \mathrm{Lip}(\mathbb{R}^d)$ we have

$$\begin{aligned}
\int f\, d(\nu . \mu) &= \int\int f(g . x)\, d\nu(g)\, d\mu(x) \\
&= \int\int f(g . x_0 + U_0(x - x_0))\, d\nu(g)\, d\mu(x) + O(2^{-2k} \cdot \|f\|_{\mathrm{Lip}}) \\
&= \int f(y)\, d((\nu . x_0) * (U_0 \tau_{-x_0} \mu))(y) + O(2^{-2k} \cdot \|f\|_{\mathrm{Lip}}).
\end{aligned}$$

Let $\nu'$, $\mu'$ and $\theta$ be the measures obtained from $\nu . x$, $U_0 \tau_{x_0} \mu$ and $\nu . \mu$, respectively, by scaling them by a factor of $2^k$ and translating them so that they are supported on a closed ball $B$ of radius $O(1)$ at the origin. Define a metric on $\mathcal{P}(B)$ by

$$d(\alpha, \beta) = \sup_{\|f\|_{\mathrm{Lip}}=1} |\int f d\alpha - \int f d\beta|.$$

It is well known that $d(\cdot, \cdot)$ is compatible with the weak-* topology on $\mathcal{P}(B)$ (see e.g. [26, Chapter 14]), and the calculation above implies that $d(\nu' * \mu', \theta) = O(2^{-k})$. Thus when $k$ is large, by Lemma 3.2 (1), $|H_m(\nu' * \mu') - H_m(\theta)| = O(1/m)$. Restating this in terms of the original measure, we have shown:

**Lemma 5.11.** *For every $m \in \mathbb{N}$ and $k > k(m)$ the following holds. If $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $\nu \in \mathcal{P}(G_0)$ are supported on level-$k$ dyadic cubes, and $x_0 \in \operatorname{supp} \mu$ and $g_0 \in \operatorname{supp} \nu$, then*

$$\begin{aligned}
H_{k,m}(\nu . \mu) &= H_{k,m}\left((\nu . x_0) * U_0 \mu\right) + O(\frac{1}{m}) \\
&= H_{k,m}\left((U_0^{-1}(\nu . x_0)) * \mu\right) + O(\frac{1}{m})
\end{aligned}$$

We omitted the translation in the statement because it commutes with convolution, and does not affect entropy more than the error term. The second line follows from the first by applying $U_0^{-1}$ to the convolution.

Reasoning similarly, let $g, g_0 \in G_0$ belong to a common level-$\ell$ dyadic cube $D$, and $x, x_0 \in \mathbb{R}^d$ belong to a common level-$k$ dyadic cell. Then, using (64), and the fact that $\|(U - U_0)(y - x)\| = O(2^{-\ell-k})$, we have

$$gx = gx_0 + U_0(x_0 - x) + O(2^{-\ell-k})$$

Thus, for $\nu \in \mathcal{P}(D)$ and $f \in \mathrm{Lip}(\mathbb{R}^d)$, we have

$$\begin{aligned}
\int f\, d(\nu . x) &= \int f(gx), d\nu(g) \\
&= \int f(gx_0 + U_0(x_0 - x) + O(2^{-\ell-k})\, d\nu(g) \\
&= \int f(gx_0 + U_0(x_0 - x))\, d\nu(g) + O(2^{-\ell-k} \|f\|_{\mathrm{Lip}} \\
&= \int f(y)\, d(\tau_{U_0(x_0 - x)}(\nu . x))(y) + O(2^{-\ell-k} \|f\|_{\mathrm{Lip}}).
\end{aligned}$$

Now, $\nu . x$ and $\nu . x_0$ are measures supported on sets of diameter $O_{\|x\|, \|x_0\|}(2^{-\ell})$ (since $\nu$ is supported on a level-$\ell$ dyadic cell), so re-scaling by $2^\ell$ turns them into "macroscopic" measures. The equation above says that after this the resulting measures are, up to a translation, $2^{-k}$-close in the weak sense. Therefore,

**Lemma 5.12.** *For every $\varepsilon > 0$ and $k > k(\varepsilon)$, if $\nu \in \mathcal{P}(G_0)$ is supported on a level-$k$ dyadic cube, $x, y \in \mathbb{R}^d$ are in the same level-$k$ dyadic cube, and $V \leq \mathbb{R}^d$ is a linear subspace then*

$$S_k(\nu . x) \text{ is } (V, \varepsilon)\text{-concentrated} \quad \Longrightarrow \quad S_k(\nu . y) \text{ is } (V, 2\varepsilon)\text{-concentrated.}$$

## 5.5 Proof of the inverse theorem

We first prove a version of the inverse theorem 2.12 which assumes that $\nu, \mu$ are supported on small dyadic cubes. These cubes are introduced to ensure that the supports of the measures are small enough for the linearization machinery to kick in, and the proof focuses on this aspect of the argument. After the proof we explain how to get the stronger version, in which the measures have larger support, and give bounds on the dimensions of the subspaces produced by the theorem.

**Theorem 5.13.** *For every $\varepsilon > 0$, $R > 0$ and $m \in \mathbb{N}$ there is a $\delta = \delta(\varepsilon, R, m) > 0$, such that, for all $k > k(\varepsilon, R, m, \delta)$ and all $n > n(\varepsilon, R, m, \delta, k)$, the following holds: If $\nu \in \mathcal{P}(G_0)$ and $\mu \in \mathcal{P}([-R, R]^d)$ are supported on level-$k$ dyadic cells, then either*

$$H_n(\nu . \mu) > H_n(\mu) + \delta,$$

*or there is a sequence $V_k, \ldots, V_n$ of subspaces of $\mathbb{R}^d$ such that*

$$\mathbb{P}_{0 \leq i \leq n}\left(\mu^{x,i} \text{ is } (V_i, \varepsilon, m)\text{-saturated}\right) > 1 - \varepsilon$$

*and for all $x \in \operatorname{supp} \mu$,*

$$\mathbb{P}_{0 \leq i \leq n}\left(S_i(\nu_{g,i} . x) \text{ is } (U_g V_i, \varepsilon)\text{-concentrated}\right) > 1 - \varepsilon.$$

*Remark* 5.14.

1. Since $k$ is assumed large relative to $\varepsilon$, by Lemma 5.12 the last condition holds for all $x \in \operatorname{supp} \mu$ if and only if it holds for some $x \in \operatorname{supp} \mu$, up to a change of a factor of 2 in the degree of concentration.
2. The measures $\mu$, $\nu$ and $\mu . \nu$ are supported on sets of diameter $O_R(2^{-k})$, so when measuring their scale-$n$ entropy it might seem more natural to rescale them by $O_R(2^k)$. However, the statement of the theorem is formally unchanged if we do so, since we are taking $n$ large relative to $k$, and the average entropy over $n$ scales is negligibly affected by the first $k$ scales.

*Proof.* Let $\varepsilon, R$ and $m$ be given.

i. Apply Corollary 5.3 with parameters $\varepsilon$ and $m$ to obtain parameters $\varepsilon'$ and $m'$ and $n'$.

ii. Apply Proposition 5.6 with parameter $\frac{1}{4}\varepsilon'$, $R$ and $m'$ to obtain parameters $\varepsilon''$ and $m''$.

iii. Apply the Euclidean inverse theorem (Theorem 2.8) with parameters $\varepsilon'', R, m''$, obtaining $\delta'$ and $n''$. We are free to assume that $\delta'$ is arbitrarily small in a manner depending on the previous parameters, and that $n''$ is large with respect to previous parameters. In particular we assume $n''$ is large relative to

$$\delta = (\delta'/2)^2.$$

iv. Choose $k$ large enough that the conclusions of Lemma 5.12 hold for parameter $\varepsilon''$ and Lemma 5.11 holds for parameter $n'$ (instead of $m$ there). We also assume that for any $D \in \mathcal{D}_k^{G_0}$ and $g, h \in D$, the difference $\|U_g - U_{g'}\|$ is small enough that if $\theta \in \mathcal{P}([0,1]^d)$ is a $(U_g V, \frac{1}{4}\varepsilon')$-concentrated measure then it is also $(U_h V, \varepsilon')$-concentrated.

v. Let $n$ be very large in a manner depending on all previous parameters.

Now let $\nu \in \mathcal{P}(G), \mu \in \mathcal{P}(\mathbb{R}^d)$ be supported on level-$k$ dyadic cells, and suppose that

$$H_n(\nu \mathbin{.} \mu) \leq H_n(\mu) + \delta. \tag{66}$$

By Lemmas 3.5 and 3.7, assuming $n$ is large compared to $n''$, (66) implies

$$\mathbb{E}_{0 \leq i \leq n} \left( H_{i,n''}(\nu_{g,i} \mathbin{.} \mu) - H_{i,n''}(\mu_{x,i}) \right) < 2\delta.$$

By Lemma 5.11, our choice of $k$ and the fact that $n''$ are large in a manner depending on $\delta$,

$$\mathbb{E}_{0 \leq i \leq n} \left( H_{i,n''}((U_g^{-1}(\nu_{g,i} \mathbin{.} x)) * \mu_{x,i}) - H_{i,n''}(\mu_{x,i}) \right) < 3\delta.$$

Since $n''$ is large enough relative to $\delta$, the difference inside the expectation is essentially non-negative, the is, larger than $-\delta$ (Lemma 4.1). Since $\delta' = \sqrt{4\delta}$, by Markov's inequality we conclude that

$$\mathbb{P}_{0 \leq i \leq n} \left( H_{i,n''}((U_g^{-1}(\nu_{g,i} \mathbin{.} x)) * \mu_{x,i}) \leq H_{i,n''}(\mu_{x,i}) + \delta' \right) > 1 - \delta'.$$

Fix $g, x$ such that $\nu_{g,i}$ and $\mu_{x,i}$ are in the event above. Write $\eta = U_g^{-1}(\nu_{g,i} \mathbin{.} x)$ and $\theta = \mu_{x,i}$. Since

$$H_{i,n''}(\eta * \theta) \leq H_{i,n''}(\theta) + \delta',$$

and $\eta$ is supported on a set of diameter $O(R \cdot 2^{-i})$, we can, after implicitly rescaling by $2^i$, apply the Euclidean inverse theorem (Theorem 2.8) and conclude, by our choice of the parameters $n'', \delta'$, that there are subspaces $V_j = V_j^{(i,g,x)}$ for $i \leq j \leq i + n''$, such that

$$\mathbb{P}_{i \leq j \leq i+n''} \left( \begin{array}{c} \theta^{y,j} \text{ is } (V_j^{(i,g,x)}, \varepsilon'', m')\text{-saturated and} \\ \eta^{z,j} \text{ is } (V_j^{(i,g,x)}, \varepsilon'')\text{-concentrated} \end{array} \right) > 1 - \varepsilon''.$$

Since we can assume $n'' > n'$, by Proposition 5.6 and our choice of parameters, writing $\tau = \nu_{g,i}$,

$$\mathbb{P}_{i \leq j \leq i+n''} \left( \begin{array}{c} \theta^{y,j} \text{ is } (V_j^{(i,g,x)}, \frac{1}{2}\varepsilon', m')\text{-saturated and} \\ S_{-j} U_g^{-1}(\tau_{h,j} \mathbin{.} x) \text{ is } (V_j^{(i,g,x)}, \frac{1}{4}\varepsilon')\text{-concentrated} \end{array} \right) > 1 - \frac{1}{2}\varepsilon'$$

(in the last equation, $g, x$ are fixed, and the randomness is over $y, h$ and $j$). Recalling that $\mu, \nu$ are supported on level-$k$ dyadic cells and the definition of $k$, we can apply Lemma 5.12 in the event above to replace $\tau_{h,j} \mathbin{.} x$ by $\tau_{h,j} \mathbin{.} y$. As a result the degree of concentration degrades from $\varepsilon'/4$ to $\varepsilon'/2$. Then, since $h, g$ are in the same level $j$ (and hence level-$k$) component, we can exchange $U_g$ with $U_h$ in the event above with another $\varepsilon'/4$ degradation of the concentration. After these adjustments we have

$$\mathbb{P}_{i \leq j \leq i+n''} \left( \begin{array}{c} \theta^{y,j} \text{ is } (V_j^{(i,g,x)}, \varepsilon', m')\text{-saturated and} \\ S_{-j} U_h^{-1}(\tau_{h,j} \mathbin{.} y) \text{ is } (V_j^{(i,g,x)}, \varepsilon')\text{-concentrated} \end{array} \right) > 1 - \frac{1}{2}\varepsilon'.$$

So far we have seen that with high probability (at least $1-\delta'$) over choice of components $\theta = \mu_{x,i}$ and $\tau = \nu_{g,i}$, we can associate subspaces $V_j^{(i,g,x)}$ to a large fraction (at least $1-\varepsilon'/2$) of the components of $\theta, \tau$ at levels $i, \ldots, i+n''$. These components are also components of $\nu, \mu$, but each component of $\nu, \mu$ may arise in several ways as a component of a components. So we have not associated a subspace to (most) components of $\nu, \mu$, but rather to (most) components of $\nu, \mu$ we have associated several subspaces. To correct this we invoke Lemma 2.7, letting us select subspaces $V^{(i,g,x)}$ (no longer depending on $j$) such that

$$\mathbb{P}_{0\leq i\leq n}\left(\begin{array}{c} \mu^{x,i} \text{ is } (V^{(i,g,x)}, \varepsilon', m')\text{-saturated and} \\ S_{-i}U_g^{-1}\nu_{g,i}\boldsymbol{.}x \text{ is } (V^{(i,g,x)}, \varepsilon')\text{-concentrated} \end{array}\right) > 1 - \frac{1}{2}\varepsilon' - \delta' - O(\frac{n''}{n}).$$

The right hand side is $> 1-\varepsilon'$ assuming as we may that $\delta'$ is small compared to $\varepsilon'$ and $n$ large relative to $n''$. Applying Corollary 5.3, and by our choice of $\varepsilon'$, there are subspaces $V^i$, independent of $g, x$, such that

$$\mathbb{P}_{0\leq i\leq n}\left(\begin{array}{c} \mu^{x,i} \text{ is } (V^i, \varepsilon, m)\text{-saturated and} \\ S_iU_g^{-1}\nu_{g,i}\boldsymbol{.}x \text{ is } (V^i, \varepsilon)\text{-concentrated} \end{array}\right) > 1 - \varepsilon.$$

This implies the statement. □

We now prove Theorem 2.12, which we repeat for convenience:

**Theorem 5.15.** *For every $\varepsilon > 0$, $R > 0$ and $m \in \mathbb{N}$, there exists $\delta = \delta(\varepsilon, R, m) > 0$ such that for every $k > k(\varepsilon, R, m)$ and every $n > n(\varepsilon, R, m, k)$, the following holds. For every $\nu \in \mathcal{P}(G_0)$ and $\mu \in \mathcal{P}([-R, R]^d)$ that are supported on balls of radius $R$, either*

$$H_n(\nu\boldsymbol{.}\mu) > H_n(\mu) + \delta,$$

*or else, to every pair of level-$k$ components $\widetilde{\nu}$ of $\nu$ and $\widetilde{\mu}$ of $\mu$ we can assign a sequence of subspaces $V_i = V_i(\widetilde{\nu}, \widetilde{\mu}) < \mathbb{R}^d$, $0 \leq i \leq n$, such that with probability at least $1-\varepsilon$ over the choice of $\widetilde{\mu}, \widetilde{\nu}$,*

$$\mathbb{P}_{0\leq i\leq n}\left(\begin{array}{c} \widetilde{\mu}^{x,i} \textit{ is } (V_i, \varepsilon, m)\textit{-saturated and} \\ S_iU_g^{-1}(\widetilde{\nu}_{g,i}\boldsymbol{.}x) \textit{ is } (V_i, \varepsilon)\textit{-concentrated} \end{array}\right) > 1 - \varepsilon$$

*If in addition $\mu$ is $((\varepsilon/5d)^{2(d+1)}, \sigma)$-non-affine for some $\sigma > 0$, and the relation among parameters takes $\sigma$ into account, then for those $\widetilde{\nu}, \widetilde{\mu}$ in the set of good components above,*

$$\frac{1}{n+1}\sum_{i=0}^{n} \dim V_i > \frac{1}{d+1} H_n(\widetilde{\nu}) - \varepsilon \tag{67}$$

*and*

$$\mathbb{E}_{i=k}\left(\frac{1}{n+1}\sum_{j=0}^{n} \dim V_j(\nu_{g,i}, \mu_{x,i})\right) > \frac{1}{d+1} H(\nu) - \varepsilon \tag{68}$$

*Proof.* Fix $\varepsilon, R, m$ (the error terms below depend on them but we suppress it in the notation). Let also $\delta, k, n$ be parameters whose relations we will specify later, and suppose that

$$H_n(\nu\boldsymbol{.}\mu) < H_n(\mu) + \delta$$

By Lemma 3.8,

$$\mathbb{E}_{i=k}\left(H_n(\nu_{g,i}\boldsymbol{.}\mu)-H_n(\mu_{x,i})\right)<2\delta+O(\frac{1}{n})<3\delta.$$

By Markov's inequality, assuming $n$ large enough,

$$\mathbb{P}_{i=k}\left(H_n(\nu_{g,i}\boldsymbol{.}\mu)-H_n(\mu_{x,i})<\sqrt{3\delta}\right)>1-\sqrt{3\delta} \tag{69}$$

Assuming as we may that $\sqrt{3\delta}<\varepsilon$, the last probability is at least $1-\varepsilon$.

Now fix a pair of components $\widetilde{\nu},\widetilde{\mu}$ from the event in (69). Assuming that $\sqrt{3\delta}$ is small relative to $\varepsilon,R,m$ and that $k,n$ are large enough, we can apply the previous theorem to $\widetilde{\nu}$ and $\widetilde{\mu}$ (which are by definition supported on level-$k$ components) and obtain corresponding subspaces $V_i(\widetilde{\nu},\widetilde{\mu})$, $k\leq i\leq n$. This proves the first part of the present theorem.

For the second part (bounding the dimensions of the subspaces), suppose that $\mu$ is $(\sigma',\sigma)$-non-concentrated. Fix an auxiliary parameter $\varepsilon'$ depending in a manner we shall later determine on $\varepsilon,\sigma$ and $R$, and run first part using $\varepsilon'$ instead of $\varepsilon$, obtaining associated $\delta,k,n$ etc., and a set of level-$k$ components $\widetilde{\nu},\widetilde{\mu}$ of probability at least $1-\varepsilon'$ to which are associated subspaces $V_i(\widetilde{\mu},\widetilde{\nu})$ with the desired properties w.r.t. $\varepsilon'$. Define $V_i(\widetilde{\nu},\widetilde{\mu})=\mathbb{R}^d$ for any pair of level-$k$ components $\widetilde{\nu},\widetilde{\mu}$ for which is was not yet defined (i.e. pairs that are not in the event in (69)). For $i\geq k$ and components $\nu_{g,i}$ and $\mu_{x,i}$ set

$$V(\nu_{g,i},\mu_{x,i})=V_i(\nu_{g,k},\mu_{g,k}).$$

This is well defined because a level-$i$ component for $i\geq k$ determines uniquely the level-$k$ component it belongs to (on the other hand we are abusing notation slightly since, strictly speaking, $\nu_{g,i},\mu_{x,i}$ do not determine $g,x,i$; but as they are written explicitly, no confusion should occur).

Observe now that, by the first part of the proof,

$$\mathbb{P}_{0\leq i\leq n}\left(S_iU_g^{-1}\nu_{g,i}\boldsymbol{.}x\text{ is }(V(\nu_{g,i},\mu_{x,i}),\varepsilon)\text{-concentrated}\right)\,d\mu(x)>1-\varepsilon'.$$

Indeed, if we write $\widetilde{\nu},\widetilde{\mu}$ for the level-$k$ components to which $\nu_{g,i},\mu_{x,i}$ belong, respectively, then conditioned on $\widetilde{\nu},\widetilde{\mu}$ belonging to the event in (69), the probability of the event above is at least $1-\varepsilon'$; while conditioned on the complementary event, the probability is 1, since then $V(\widetilde{\nu},\widetilde{\mu})=\mathbb{R}^d$. Thus the unconditional probability above is at least $1-\varepsilon'$.

Set

$$\ell=[\log(1/\varepsilon')]$$

By Lemma 3.15, the previous inequality gives

$$\mathbb{P}_{0\leq i\leq n}\left(H_\ell(S_iU_g^{-1}\nu_{g,i}\boldsymbol{.}x)<\dim V(\nu_{g,i},\mu_{x,i})+O(\frac{\log\ell}{\ell})\right)\quad>\quad1-\varepsilon'.$$

Since by Lemma 3.2 (2),

$$\left|H_\ell(S_iU_g^{-1}\nu_{g,i}\boldsymbol{.}x)-H_{i,\ell}(\nu_{g,i}\boldsymbol{.}x)\right|=O(\frac{1}{\ell}),$$

we obtain

$$\begin{aligned}\mathbb{P}_{0\le i\le n}\left(H_{i,\ell}(\nu_{g,i}\mathbin{.}x)<\dim V(\nu_{g,i},\mu_{x,i})+O(\frac{\log\ell}{\ell})\right) &> 1-2\varepsilon'-O(\frac{1}{\ell})\\ &= 1-O(\frac{1}{\ell}).\end{aligned}\qquad(70)$$

We now use the assumption that $\mu$ is $((\varepsilon/5d)^{2(d+1)},\sigma)$-non-affine. By Corollary 5.10, for every component $\nu_{g,i}$ of $\nu$,

$$\mu\left(x\in\mathbb{R}^d\ :\ H_{i,\ell}(\nu_{g,i}\mathbin{.}x)>\frac{1}{d+1}H_{i,\ell}(\nu_{g,i})-O_{\sigma,R}(\frac{1}{\ell})\right)\quad>\quad 1-\frac{1}{5}\varepsilon^2,$$

Choosing the component $\nu_{g,i}$, $k\le i\le n$, at random, and then $x$ independently according to $\mu$, we conclude that $H_{i,\ell}(\nu_{g,i}\mathbin{.}x)>\frac{1}{d+1}H_{i,\ell}(\nu_{g,i})-O_{\sigma,R}(\frac{1}{\ell})$ with probability at least $1-\varepsilon^2/5$. Therefore, combined with (70), we have

$$\mathbb{P}_{0\le i\le n}\left(\begin{array}{c}H_{i,\ell}(\nu_{g,i}\mathbin{.}x)<\dim V(\nu_{g,i},\mu_{x,i})+O(\frac{\log\ell}{\ell})\\ \text{and } H_{i,\ell}(\nu_{g,i}\mathbin{.}x)>\frac{1}{d+1}H_{i,\ell}(\nu_{g,i})-O_{\sigma,R}(\frac{1}{\ell})\end{array}\right)\quad>\quad 1-O(\frac{1}{\ell})-\frac{1}{5}\varepsilon^2$$

Recalling that $\ell=\log(1/\varepsilon')$, by $\varepsilon'$ small we can assume the error term does not exceed $\varepsilon^2/4$. We obtain

$$\mathbb{P}_{0\le i\le n}\left(\frac{1}{d+1}H_{i,\ell}(\nu_{g,i})<\dim V(\nu_{g,i},\mu_{x,i})+O_{\sigma,R}(\frac{\log\ell}{\ell})\right)\quad>\quad 1-\frac{1}{4}\varepsilon^2$$

By Markov's inequality, there is a set $A\subseteq G_0\times\mathbb{R}^d$ with $\nu\times\mu(A)>1-\varepsilon/2$, such that for every $(g_0,x_0)\in A$, setting $\widetilde{\nu}=\nu_{g_0,k}$ and $\widetilde{\mu}=\mu_{x_0,k}$,

$$\mathbb{P}_{0\le i\le n}\left(\frac{1}{d+1}H_{i,\ell}(\widetilde{\nu}_{g,i})<\dim V(\widetilde{\nu}_{g,i},\widetilde{\mu}_{x,i})+O_{\sigma,R}(\frac{\log\ell}{\ell})\right)\quad>\quad 1-\frac{1}{2}\varepsilon$$

Outside of the event above we have the trivial bound $\frac{1}{d+1}H_{i,\ell}(\widetilde{\nu}_{g,i})\le\frac{1}{d+1}d<1$. On the other hand, by Lemma 3.5 (which holds also in $G$),

$$H_n(\widetilde{\nu})=\mathbb{E}_{0\le i\le n}(H_{i,\ell}(\widetilde{\nu}_{g,i}))+O(1/\ell+\ell/n)$$

Finally, since $V(\widetilde{\nu}_{g,i},\widetilde{\mu}_{x,i})=V_i(\widetilde{\nu},\widetilde{\mu})$), the last two equations give

$$\begin{aligned}\frac{1}{d+1}H_n(\widetilde{\nu}) &= \mathbb{E}_{0\le i\le n}\left(\frac{1}{d+1}H_{i,\ell}(\widetilde{\nu}_{g,i})\right)+O(\frac{1}{\ell}+\frac{\ell}{n})\\ &< \frac{1}{n+1}\sum_{i=0}^{n}\left(\dim V_i(\widetilde{\nu},\widetilde{\mu})+O_{\sigma,R}(\frac{\log\ell}{\ell})\right)+O(\frac{1}{\ell}+\frac{\ell}{n})+\frac{1}{2}\varepsilon\end{aligned}\qquad(71)$$

Taking $\varepsilon'$ small (and hence $\ell$ large) relative to $\varepsilon,R,\sigma$, and $n$ larger, and rearranging, we obtain (67).

Finally, recall that (71) holds on a set of pairs of level-$k$ components $\widetilde{\nu},\widetilde{\mu}$ of probability at least $1-\varepsilon/2$, and recall that

$$\mathbb{E}_{i=k}(H_n(\nu_{g,i}))=H_n(\nu)-O(\frac{k}{n})$$

The last statement of the theorem, (68), follows now by taking expectation of both sides in (71) and making $\varepsilon'$ small enough and $n$ large enough. $\square$

## 5.6 Generalizations

To derive Theorem 2.14 very few changes are needed to the convolution case. The $C^1$-assumption of $f$, and the compactness of its domain, easily imply analogs of Equations (63), (65) and (64) and their consequences (without quantitative control on the error, but one cannot expect it in the general setting). In particular, for large enough $k$ and suitably large $n$, with $\mu\times\nu$-probability at least $1-\delta$ over choice of $(x,y)$ we have

$$|H_n(f(\mu_{x,k}\times\delta_y)) - H_n(A_{x,y}\mu_{x,k})| < \frac{\delta}{10},$$

and

$$|H_n(f(\mu_{x,k}\times\nu_{y,k})) - H_n(A_{x,y}\mu_{x,k} * B_{x,y}\nu_{x,k})| < \frac{\delta}{10}$$

(note that since $n \gg k$, there is no advantage in scaling $1/\|A\|$, $1/\|B\|$ by $2^k$, as might seem natural).

By concavity and almost-convexity of entropy (Lemma 3.1 (5) and (6)), for $n \gg k$ we have

$$\left|H_n(f(\mu\times\nu)) - \int H_n(f(\mu_{x,k}\times\nu_{y,k}))\,d\mu\times\nu(x,y)\right| < \frac{\delta}{10},$$

and for every $y$, similarly,

$$\left|H_n(f(\mu\times\delta_y)) - \int H_n(f(\mu_{x,k}\times\delta_y))\,d\mu(x)\right| < \frac{\delta}{10}.$$

Thus the hypothesis (17) of Theorem 2.14 implies that for any $k$ and $n \gg k$,

$$\int H_n(f(\mu_{x,k}\times\nu_{y,k}))\,d\mu\times\nu(x,y) < \int H_n(f(\mu_{x,k}\times\delta_y))\,d\mu\times\nu(x,y) + \frac{8}{10}\delta.$$

By the above, for large $k$ this is

$$\int H_n(A_{x,y}\mu_{x,k} * B_{x,y}\nu_{x,k})\,d\mu\times\nu(x,y) < \int H_n(A_{x,y}\mu_{x,k})\,d\mu\times\nu(x,y) + \frac{6}{10}\delta.$$

Since for large $n$ we essentially have the reverse inequality between the integrands, we conclude that with high probability at least $1-\delta$ over the components $\widetilde{\mu} = \mu_{x,k}$ and $\widetilde{\nu} = \nu_{y,k}$, we have

$$H_n(A_{x,y}\widetilde{\mu} * B_{x,y}\widetilde{\nu}) < H_n(A_{x,y}\widetilde{\mu}) + \delta',$$

where $\delta'$ tends to zero with $\delta$. From here one can apply the Euclidean inverse theorem to the components $\widetilde{\nu},\widetilde{\mu}$ as we did in the proof of the convolution case, with very few changes other than notational ones. We omit the details.

In the special case of actions of matrix groups on $\mathbb{R}^d$ or on themselves, one has analogs of Corollary 5.10. In the first case essentially by the same lemma (using compactness of the domain of the action function in place of compactness of the orthogonal group). For a matrix group acting on itself, there are in fact trivial stabilizers, so there conclusion is automatic.

# 6 Self-similar sets and measures on $\mathbb{R}^d$

The derivation of our main result, Theorem 1.5, from the Theorem 2.12 (the inverse theorem for the $G_0$-action), follows lines similar to the argument in [12] for $\mathbb{R}$. One new ingredient is the explicit presence of the isometry group, but this is implicit in the original argument and the main change is notational. More significant is the appearance of invariant subspaces in the third alternative of the theorem. This will require some further analysis, and will occupy us in the first few subsections.

We remark that our analysis so far, and much of the analysis below, is of a finitary nature, involving entropies at fine (but finite) partitions. Certainly we must somewhere connect this to dimension, specifically to the dimension of the conditional measures of a given self-similar measure on the family of translates of a subspace (as in (iii") of Theorem 1.5). It is an unfortunate reality that such a connection seems to be available only when the subspace is invariant under the linearization of the IFS (see Section 6.4 below). If such results were available without invariance, much of the technical work of the next few sections could be avoided by passing to a limit at an earlier stage. However, understanding these "slice" measures for general self-similar measures remains an open problem.

## 6.1 Almost-invariance and invariance

We will obtain invariant subspaces from almost invariant ones:

**Definition 6.1.** A subspace $V \leq \mathbb{R}^d$ is $\varepsilon$-invariant under a subgroup $H < G_0$, or $(H, \varepsilon)$-invariant, if $d(hV, V) \leq \varepsilon$ for every $h \in H$.

Evidently, $(H, 0)$-invariance is $H$-invariance in the usual sense. Furthermore,

**Lemma 6.2.** *Let $H < G_0$ be a closed subgroup. For every $\varepsilon > 0$ there is a $\delta > 0$, such that if $V$ is $\delta$-invariant under $H$, then there is an $H$-invariant subspace $V'$ with $d(V, V') < \varepsilon$.*

*Proof.* Let $\mathcal{S}_H$ denote the space of $H$-invariant subspaces of $\mathbb{R}^d$. If the statement were false there would be some $\varepsilon > 0$ and a sequence $V_n \leq \mathbb{R}^d$ of subspaces such that $V_n$ is $1/n$-invariant for $H$, but $d(V_n, V') \geq \varepsilon$ for every $V' \in \mathcal{S}_H$. Using compactness of the space of subspaces, we can pass to a subsequence $V_{n_k}$ converging to some $V$. Since the linear action is continuous, $d(V, hV) = \lim d(V_{n_k}, hV_{n_k}) = 0$ for all $h \in H$, so $V \in \mathcal{S}_H$. But by hypothesis $d(V_{n_k}, V) \geq \varepsilon$ for all $k$, a contradiction. $\square$

In fact the choice $\delta = c \cdot \varepsilon^{d+1}$ works for an appropriate constant $c$ (or $c \cdot \varepsilon^{k+1}$ if one fixes the dimension $k$ of the subspace in question), but we will not use this.

Our second tool will be to construct almost-invariant subspaces from almost-invariant sets of vectors.

**Lemma 6.3.** *Let $0 < \varepsilon < 1$ and write $\varepsilon_n = \varepsilon^{n!}$. Let $H < G_0$ be a closed subgroup and let $E \subseteq B_1(0) \subseteq \mathbb{R}^d$ be a set such that $d(hv, E) < \varepsilon_n$ for all $v \in E$ and $h \in H$. Let $v_1, \ldots, v_k \in E$ be a maximal sequence of vectors satisfying $d(v_i, \text{span}\{v_1, \ldots, v_{i-1}\}) > \varepsilon_i$ for $1 < i \leq k$, and set $V = \text{span}\{v_1, \ldots, v_k\}$. Then $V$ is $(H, O(\varepsilon))$-invariant and $E \subseteq V^{(\varepsilon_{k+1})}$.*

*Proof.* We may assume $k < d$ since otherwise $V = \mathbb{R}^d$ and the statements is trivial. To see that $E \subseteq V^{(\varepsilon_{k+1})}$, note that if $v \in E \setminus V^{(\varepsilon_{k+1})}$ then the vector $v_{k+1} = v$ would extend the given sequence of vectors in a way that contradicts its maximality. For invariance, let $h \in H$ and set $w_i = hv_i$ and $W = hV = \text{span}\{w_i\}$. By assumption, for each $i$ there is a $w_i' \in E$ with $d(w_i, w_i') < \varepsilon_d \leq \varepsilon_{k+1}$, and we saw above that $w_i' \in V^{(\varepsilon_{k+1})}$, hence $w_i \in V^{(2\varepsilon_{k+1})}$. Also, $h$ is an isometry, so $d(w_i, \text{span}\{w_1, \ldots, w_{i-1}\}) > \varepsilon_i \geq \varepsilon_k$ for all $1 \leq i \leq k$, since the same is true for the $v_i$. Therefore, by Corollary 3.23, $\text{span}\{w_1, \ldots, w_k\} \sqsubseteq V^{(c \cdot \varepsilon_{k+1}/\varepsilon_k^k)}$, and using the fact that $\dim V = \dim W$, this implies

$$d(W, V) = O(\frac{2\varepsilon_{k+1}}{\varepsilon_k^k}) = O(\varepsilon^{(k+1)! - k! \cdot k}) = O(\varepsilon^{k!}) = O(\varepsilon),$$

as desired. □

## 6.2 Saturation at level $n$

We will be interested in the situation where the components of a measure at some scale typically are highly saturated on a subspace. More precisely,

**Definition 6.4.** For $V \leq \mathbb{R}^d$ , a measure $\mu \in \mathcal{P}(\mathbb{R}^d)$ is $(V, \varepsilon, m)$*-saturated at level* $n$ if

$$\mathbb{P}_{i=n}\left(\mu^{x,i} \text{ is } (V, \varepsilon, m)\text{-saturated}\right) > 1 - \varepsilon.$$

We write

$$\text{sat}(\mu, \varepsilon, m, n) = \{V \leq \mathbb{R}^d \,:\, \mu \text{ is } (V, \varepsilon, m)\text{-saturated at level } n\}.$$

Some technical properties related to this notion are summarized in the next lemma. In the formulation we write $\sum A$ for $\sum_{a \in A} a$.

**Lemma 6.5.** *Let* $\varepsilon, R > 0$ *and* $V \leq \mathbb{R}^d$. *Let* $\mu \in \mathcal{P}([-R, R]^d)$ *and suppose that* $\mu$ *is given as a convex combination of probability measures,* $\mu = \sum_{i=1}^k \alpha_i \mu_i$.

1. *If* $\mu$ *is* $(V, \varepsilon, m)$*-saturated, then*

$$\sum\{\alpha_i \,:\, \mu_i \textit{ is } (V, \varepsilon', m)\textit{-saturated}\} > 1 - \varepsilon',$$

   *where* $\varepsilon' = O(\sqrt{\varepsilon + (\log kR)/m})$.

2. *For* $n$ *sufficiently large in a manner depending on* $\mu, \alpha_i, \nu_i$, *if* $V \in \text{sat}(\mu, \varepsilon, m, n)$ *then*

$$\sum\{\alpha_i \,:\, V \in \text{sat}(\mu_i, \varepsilon', m, n)\} > 1 - \varepsilon',$$

   *where* $\varepsilon' = O(\sqrt{\varepsilon})$.

3. *If for some* $n$ *we have*

$$\sum\{\alpha_i \,:\, V \in \text{sat}(\mu_i, \varepsilon, m, n)\} > 1 - \varepsilon,$$

   *then* $V \in \text{sat}(\mu, \varepsilon', m, n)$, *where* $\varepsilon' = O(\sqrt{\varepsilon})$.

4. *Let $g = 2^{-t}U + a \in G$. If $V \in \text{sat}(\mu, \varepsilon, m, n)$ then $UV \in \text{sat}(g\mu, \varepsilon', m, [n-t])$ where $\varepsilon' \to 0$ as $(\varepsilon, \frac{1}{m}) \to 0$.*

5. *Under the same assumptions as in* (4), *$UV \in \text{sat}(g\mu, \varepsilon'', m, n)$ where $\varepsilon'' \to 0$ as $(\varepsilon, \frac{1}{m}) \to 0$.*

*Proof.* For (1), by absorbing an $O(1/m)$ error into $\varepsilon$ we can assume that $\mathcal{D}_m = \mathcal{D}_m^V \vee \mathcal{D}_m^{V^\perp}$ (Lemma 3.9). By Lemmas 3.1 (6) and the hypothesis, we have

$$\begin{aligned}\sum_{i=1}^{k} \alpha_i \cdot \frac{1}{m} H(\mu_i, \mathcal{D}_m | \mathcal{D}_m^{V^\perp}) &\geq \frac{1}{m} H(\mu, \mathcal{D}_m | \mathcal{D}_m^{V^\perp}) - \frac{\log k}{m} \\ &> \dim V - (\varepsilon + \frac{\log k}{m}).\end{aligned}$$

On the other hand, each $\mu_i$ is supported on $[-R, R]^d$ so each term in the average on the left hand side is bounded above by $\dim V + O(\frac{\log R}{m})$. Now (1) follows by Markov's inequality.

For (2), fix for convenience $\delta = \sqrt{\varepsilon}$. By standard differentiation theorems, for $\mu_i$-a.e. $x$, $\|\mu_{x,\ell} - (\mu_i)_{x,\ell}\| \to 0$ as $\ell \to \infty$. In particular for large $n$, for a set of $x$ of $\mu_i$-mass at least $1 - \delta$, we have $\mu^{x,n} = (1-\delta)\mu_i^{x,n} + \delta\theta$ for some $\theta \in \mathcal{P}([0,1]^d)$ (depending on $x, i$). For any such $n$ let

$$A = \{x \in [0,1]^d \,:\, \mu^{x,n} \text{ is } (V, m, \varepsilon)\text{-saturated}\}.$$

By hypothesis $\mu(A) > 1 - \varepsilon$. Since $\mu = \sum \alpha_i \mu_i$, by Markov's inequality we have

$$\sum\{\alpha_i \,:\, \mu_i(A) > 1 - \sqrt{\varepsilon}\} > 1 - \sqrt{\varepsilon}. \tag{72}$$

For $i$ satisfying $\mu_i(A) > 1 - \sqrt{\varepsilon}$, for a set $x$ of points having $\mu_i$-mass $1 - \delta - \sqrt{\varepsilon}$ we have that $\mu^{x,n}$ is $(V, \varepsilon, m)$-saturated and $\mu^{x,n} = (1-\delta)\mu_i^{x,n} + \delta\theta$ for some $\theta \in \mathcal{P}[0,1]^d)$. Now we can apply part (1) of this lemma to $\mu^{x,n}$, which is written as a combination of two measures ($k = 2$) and supported on $[0, 1)$ (so $R = 1$), and conclude that $\mu_i^{x,n}$ is $(V, O(\sqrt{\varepsilon}), m)$-saturated. This holds for at least a $(1 - \delta - \sqrt{\varepsilon})$-fraction of the components $\mu_i^{x,n}$. Since $\delta = \sqrt{\varepsilon}$ we find that $\mu_i$ is $(V, O(\sqrt{\varepsilon}), m, n)$-saturated. This together with (72) is what we wanted to prove.

For (3), observe that $\mu^{x,n}$ is a convex combination of components $\mu_i^{x,n}$ (the weights are proportional to $\alpha_i \mu_i(\mathcal{D}_n(x))$). By Lemmas 3.12 and 3.14, we will be done if we show with $\mu$-probability $> 1 - \sqrt{2\varepsilon}$ over the choice of $x$, the components $\mu_i^{x,n}$ which are $(V, \varepsilon, m)$-saturated constitute a $(1 - \sqrt{2\varepsilon})$-fraction of the mass of $\mu^{x,n}$.

To show this, let $I = \{1, \ldots, k\}$ and let $\alpha$ be the probability measure on $I$ arising from the weights $\alpha_i$. Consider the space $I \times \mathbb{R}^d$ with the probability measure $\theta$ given by $\theta(\{i\} \times A) = \alpha_i \mu_i(A)$. Define $f : I \times \mathbb{R}^d \to \mathbb{R}$ by

$$f(i, x) = \begin{cases} 1 & \text{if } \mu_i^{x,n} \text{ is } (V, \varepsilon, m)\text{-saturated} \\ 0 & \text{otherwise} \end{cases}.$$

Note that $f$ is $2^I \times \mathcal{D}_n$-measurable. Writing $I_0 = \{i \in I \,:\, \mu_i \text{ is } (V, \varepsilon, m, n)\text{-saturated}\}$, we have

$$\begin{aligned}
\int f d\theta &= \sum_{i\in I}\alpha_i \int f(i,x)d\mu_i(x) \\
&\geq \sum_{i\in I_0}\alpha_i \int f(i,x)d\mu_i(x) \\
&= \sum_{i\in I_0}\alpha_i\mu_i(x\,:\,\mu^{x,n}\text{ is }(V,\varepsilon,m)\text{-saturated}) \\
&> \sum_{i\in I_0}\alpha_i(1-\varepsilon) \\
&> (1-\varepsilon)^2 \\
&> 1-2\varepsilon
\end{aligned}$$

(the passage from the third to fourth equation is by the hypothesis). Let $\mathcal{B}$ be smallest the $\sigma$-algebra that makes the map $I\times\mathbb{R}^d\to\mathbb{R}^d$, $(i,x)\mapsto x$, measurable. The function $g=\mathbb{E}(f|\mathcal{B})$ also satisfies $g\leq 1$ and $\int g d\theta=\int f d\theta > 1-2\varepsilon$, so by Markov's inequality,

$$\theta((i,x)\,:\,g(x)>1-\sqrt{2\varepsilon})>1-\sqrt{2\varepsilon}.$$

But, writing $D=\mathcal{D}_n(x)$, the inequality $g(x)>1-\sqrt{2\varepsilon}$ just means that in the convex combination $\mu^{x,n}=\sum\frac{\alpha_i\mu_i(D)}{\sum\alpha_i\mu_i(D)}(\mu_i)^{x,n}$, at least $1-\sqrt{2\varepsilon}$ of the mass originates in terms for which $(\mu_i)^{x,n}$ is $(V,\varepsilon,m)$-saturated. Since the distribution on $x$ induced by $\theta$ is equal to $\mu$, this completes the proof.

For (4), consider $D\in\mathcal{D}_n$ such that $\mu^D$ is $(V,\varepsilon,m)$-saturated. Let $\nu=g(\mu_D)$. Then $\nu'=S_{[n-t]}\nu$ is the image of $\mu^D$ under a similarity that contracts by $O(1)$ and rotates by $U$, and so by Lemma 3.10, $\nu'$ is $(UV,\varepsilon+O(1/m),m)$-saturated. Writing $\nu'=\sum_{D\in\mathcal{D}_1}\nu'(D)\cdot\nu'_D$ we can apply (1) and conclude that, with $\varepsilon$ small and $m$ large, most mass in this convex combination comes from terms that are $(UV,\varepsilon',m)$-saturated. This means precisely that $\nu'$ is $(UV,\varepsilon',m,n)$-saturated. Now, since $g\mu$ is the convex combination of measures $\nu$ of which a $(1-\varepsilon)$-fraction are as above, (4) follows from (2).

(5) is proved in the same manner as (4), using $\nu'=S_n\nu$ instead of $S_{[n-t]}\nu$. □

## 6.3 Saturated subspaces of self-similar measures

From here until the end of the paper we again denote by $\mu$ a self-similar measure on $\mathbb{R}^d$ defined by an IFS $\Phi=\{\varphi_i\}_{i\in\Lambda}$ and a positive probability vector $p=(p_i)_{i\in\Lambda}$. As usual we write $\varphi_i=r_iU_i+a_i$, and for $i\in\Lambda^k$ we set $\varphi_i=\varphi_{i_1}\circ\ldots\circ\varphi_{i_k}$, $p_i=p_{i_1}\cdot\ldots\cdot p_{i_k}$, and define $r_i$, $U_i$, similarly. Denote by $G_\Phi\subseteq G_0$ the smallest closed group containing the orthogonal parts $U_i$, $i\in\Lambda$, of the maps $\varphi_i\in\Phi$.

In the next few results, all dependences between parameters and implicit constants depend on $\mu$ and $\Phi$.

The first lemma says that the set of subspaces that are $(\mu,\varepsilon,m)$-saturated at level $n$ is almost invariant under $G_\Phi$:

**Lemma 6.6.** *For every $\varepsilon>0$ there is a $\delta=\delta(\varepsilon)>0$ such that, if $m>m(\varepsilon)$ and $n>n(\varepsilon,m)$, the following holds. For any $V\in\mathrm{sat}(\mu,\delta,m,n)$ and $g\in G_\Phi$ there exists $W\in\mathrm{sat}(\mu,\varepsilon,m,n)$ such that $d(W,gV)<\varepsilon$.*

*Proof.* Let $\Lambda^{\leq n} = \bigcup_{i=1}^{n} \Lambda^i$. Then $S = \{U_i : i \in \bigcup_{n=1}^{\infty} \Lambda^n\}$ is a sub-semigroup of $G_\Phi$, and $\overline{S}$ is a closed subgroup of $G_\Phi$ (it is a general fact that a closed sub-semigroup of a compact group is a group). Since $\{U_i\}_{i\in\Lambda} \subseteq S$ in fact $\overline{S} = G_\Phi$. Since $S$ is dense in $\overline{S}$ and $S$ is the increasing union $S = \bigcup_{n=1}^{\infty}\{U_i : i \in \Lambda^{\leq n}\}$, we can choose $k_0$ large enough that for every $V$, $g \in G_\Phi$ there is a $i \in \Lambda^{\leq k_0}$ with $d(U_i^{-1}V, gV) < \varepsilon$.

Fix $0 \leq k \leq k_0$. Since $\mu = \sum_{i\in\Lambda^k} p_i \cdot \varphi_i\mu$ we can apply Lemma 6.5 (2) with a small parameter $\delta$. Writing $\varepsilon' = \sqrt{\delta}$, it follows that if $V \in \mathrm{sat}(\mu, \delta, m, n)$ for some $m$ and $n > n_0$, then $V \in \mathrm{sat}(\varphi_j\mu, \varepsilon', m, n)$ for all $j \in \Lambda^k$ outside a set $J \subseteq \Lambda^k$ with $\sum_{j\in J} p_j < \varepsilon'$. Choose $\delta$ small enough that $p_j > \varepsilon'$ for all $j \in \Lambda^k$ (this requires $\delta$ small in a manner depending only on $k_0$, and hence only on $\varepsilon$). Thus we have shown that if and $V \in \mathrm{sat}(\mu, \delta, m, n)$ for some $m$ and $n > n_0$, then $V \in \mathrm{sat}(\varphi_j\mu, \varepsilon', m, n)$ for all $j \in \Lambda^k$. By Lemma 6.5 (5), this in turn implies that is $U_j^{-1}V \in \mathrm{sat}(\mu, \varepsilon'', m, n)$, where $\varepsilon''$ can be made $< \varepsilon$ if $\varepsilon'$ and $k_0/m$ (and hence $k/m$) are small enough. This holds if $\delta$ is small and $m$ large relative to $\varepsilon$ (and hence $k_0$), and the claim follows from our choice of $k_0$. $\square$

The next proposition says, roughly, that there is an essentially maximal $(\varepsilon, m)$-saturated subspace at each small enough scale $n$, and that it is $(G_\Phi, \varepsilon)$-invariant.

**Proposition 6.7.** *For every $0 < \varepsilon < \frac{1}{10}$ there exists a $\delta = \delta(\varepsilon) > 0$ such that, for $m > m(\varepsilon)$ and $n > n(\varepsilon, m)$ there exists a $(G_\Phi, \varepsilon)$-invariant subspace $V_n^* \in \mathrm{sat}(\mu, \varepsilon, m, n)$ such that every $W \in \mathrm{sat}(\mu, \delta, m, n)$ satisfies $W \sqsubseteq (V_n^*)^{(\varepsilon)}$.*

*Proof.* Fix $\varepsilon > 0$ and apply the previous lemma to $\varepsilon' = \varepsilon^{d!}/2$ to obtain $\delta'$ and set $\delta = \delta'/2$. Suppose $m$ and $n$ are large enough to satisfy the conclusion of that lemma. Assume that $m$ is also large enough that, for a suitable parameter $\varepsilon'' < \varepsilon$, the following holds: if $V_1, V_2 \leq \mathbb{R}^d$ are subspaces with $\angle(V_1, V_2) > \varepsilon'$ and $\mu$ is $(V_i, \varepsilon'', m)$-saturated for $i = 1, 2$ then $\mu$ is $(V_1 + V_2, 3\varepsilon', m)$-saturated (such an $m$ and $\varepsilon''$ exists by Lemma 3.21 (5)). Also assume that $m$ is large enough that if $\mu$ is $(V, \delta, m)$-saturated and $V' \leq V$, then $\mu$ is $(V, \delta', m)$-saturated (such $m$ exist by Lemma 3.21 (4), using $\delta' = 2\delta$).

By the choice of $\delta'$, if $V \in \mathrm{sat}(\mu, \delta', m, n)$ and $g \in G_\Phi$, then there is a subspace $W \leq \mathbb{R}^d$ such that $d(W, gV) < \varepsilon'$ and $W \in \mathrm{sat}(\mu, \varepsilon', m, n)$. Let $\mathcal{W}$ denote the set of all one-dimensional subspaces $W$ that arise in this way, and write

$$E = \{w \in \mathbb{R}^d : \|w\| = 1 \text{ and } \mathbb{R}w \in \mathcal{W}\}.$$

Observe that if $w \in E$ then $W = \mathbb{R}w \in \mathcal{W}$ and there exists a $V \in \mathrm{sat}(\mu, \delta', m, n)$ and $g \in G_\Phi$ with $d(W, gV) < \varepsilon'$, hence for every $h \in G_\Phi$ we have $d(hW, hgV) < \varepsilon'$. By definition of $\mathcal{W}$ there is some $W' \in \mathcal{W}$ such that $d(W', hgV) < \varepsilon'$, so $d(hW, W') < 2\varepsilon' = \varepsilon^{d!}$. Thus there is $w' \in E$ with $W' = \mathbb{R}w'$ and $d(w, w') < \varepsilon^{d!}$.

It follows that the set $E$ satisfies the hypothesis of Lemma 6.3 for $\varepsilon$ and the group $G_\Phi$. Choosing a maximal sequence of unit vectors $v_1, \ldots, v_k \in E$ such that $d(v_i, \mathrm{span}\{v_1, \ldots, v_{i-1}\}) > \varepsilon^{i!}$ and setting $V = \mathrm{span}\{v_1, \ldots, v_k\}$, we conclude that $V$ is $(G_\Phi, O(\varepsilon))$-invariant and $E \subseteq V^{(\varepsilon)}$.

Since $V = \oplus_{i=1}^{k}\mathbb{R}v_i$ and $\angle(v_i, \mathrm{span}\{v_1, \ldots, v_{i-1}\}) > \varepsilon^{i!}$, and $\mathbb{R}v_i \in \mathrm{sat}(\mu, \varepsilon^{d!}/2, m, n)$ for all $i$, repeated application of Lemma 3.21 (5), assuming $m$ large enough relative to $\varepsilon$ (and hence $\varepsilon'$), gives that $V \in \mathrm{sat}(\mu, O(\varepsilon), m, n)$.

Finally, if $W \in \mathrm{sat}(\mu, \delta, m, n)$ then we can choose an orthonormal basis $\{w_i\}$ for $W$, so by choice of $m$, $\mathbb{R}w_i \in \mathrm{sat}(\mu, \delta', m, n)$, so $w_i \in E$. By Lemma 6.3,

$w_i \in V^{(\varepsilon)}$. The $w_i$ are orthonormal, so $d(w_i, \text{span}\{w_1, \ldots, w_{j-1}\} = 1$. Hence by Corollary 3.23, $W \sqsubseteq V^{(O(\varepsilon))}$.

We have proved the claim for $V_n^* = V$, up to some constant factors, to remove them begin with a small multiple of $\varepsilon$ instead of $\varepsilon$. □

The next proposition allows us to replace a saturated almost-invariant subspace with a truly invariant one, of some lesser saturation. It also shows that this new subspace is saturated at many levels, even though the original subspace a-priori was saturated at a single level.

**Proposition 6.8.** *For every $\varepsilon > 0$, $0 < \delta < \delta(\varepsilon)$, $m > m(\varepsilon, \delta)$ and every $n \in \mathbb{N}$, the following holds. If $W \in \text{sat}(\mu, \delta, m, n)$ is $(G_\Phi, \delta)$-invariant and $\widetilde{W}$ is a $G_\Phi$-invariant subspace with $d(W, \widetilde{W}) < \delta$, then for $m' = [\log(2/\delta)]$ and all large enough $n'$ we have $\widetilde{W} \in \text{sat}(\mu, \varepsilon, m', n')$.*

*Proof.* Fix $0 < \delta < \varepsilon$. Also fix $m$ large relative to $\delta$ (we shall see how large later). Let $n \in \mathbb{N}$, $W \leq \mathbb{R}^d$ and $m', n'$ be as in the statement, so our assumption is that

$$\mathbb{P}_{i=n}\left(\mu^{x,i} \text{ is } (W, \delta, m)\text{-saturated}\right) > 1 - \delta.$$

For each measure $\theta = \mu^{x,n}$ in the event above, writing $\delta_1 = \sqrt{d\delta + O(\frac{m'}{m})}$, Lemma 3.16 implies

$$\mathbb{P}_{0 \leq j \leq m}\left(\theta^{y,j} \text{ is } (W, \delta_1, m')\text{-saturated}\right) > 1 - \delta_1.$$

Assuming $m$ is large relative to $\delta$ (and hence $m'$), we can arrange $\delta_1 < 2\sqrt{d\delta}$. Combining the two inequalities above, we can find a $0 \leq k \leq m$ such that

$$\mathbb{P}_{i=n+k}\left(\mu^{x,i} \text{ is } (W, \delta_1, m')\text{-saturated}\right) > 1 - 2\delta_1.$$

Let $\mu^{x,n+k}$ be as in this last event. Since $d(W, \widetilde{W}) < \delta < 2^{-m'}$, by Lemma 3.21 (3), $\mu^{x,n+k}$ is also $(\widetilde{W}, \delta_2, m')$-saturated, where $\delta_2 = \delta_1 + O(1/m')$. Since this holds for a $1 - 2\delta_1 > 1 - \delta_2$ proportion of components $\mu^{x,n+k}$, (because, if $\delta$ is small, $\delta_2 \geq 2\delta_1$), we have $\widetilde{W} \in \text{sat}(\mu, \delta_2, m', n+k)$. Note that $\delta_2$ can be made arbitrarily small by choosing $\delta$ small enough.

Finally, let $n' > n + k$. Let $\Lambda(n') \subseteq \bigcup_{j=1}^\infty \Lambda^j$ denote the set of sequences $i = i_1 \ldots i_\ell$ such that $r_{i_1} \cdot \ldots \cdot r_{i_\ell} < 2^{-(n'-k)} \leq r_{i_1} \cdot \ldots \cdot r_{i_{\ell-1}}$. Then $\sum_{i \in \Lambda(n')} p_i = 1$ and $\mu = \sum_{i \in \Lambda(n')} p_i \varphi_i \mu$. By Lemma 6.5 (4), $\widetilde{W} = U_i \widetilde{W} \in (\varphi_i \mu, \delta_3, m', n')$ for all $i \in \Lambda(n')$, where $\delta_3 \to 0$ as $\delta \to 0$ and $m' \to \infty$. Since $\mu$ is a convex combination of the measures $\varphi_i \mu$, $i \in \Lambda(n')$, by Lemma 6.5 (3) we have $\widetilde{W} \in (\mu, \delta_4, m', n')$ for $\delta_4$ which can be made arbitrarily small (and in particular $< \varepsilon$) if $\delta$ is small and $m'$ large. This completes the proof. □

Finally, we show the existence of a "maximal" invariant subspace which is saturated to all degrees at sufficiently deep levels. Let us say that a $\mu$ is $V$-saturated if $\mu \in \text{sat}(V, \varepsilon, m, n)$ for all $\varepsilon > 0$, $m \geq m(\varepsilon)$ and all $n > n(\varepsilon, m)$.

**Proposition 6.9.** *There exists a unique subspace $\widetilde{V} \leq \mathbb{R}^d$ such that*

1. *$\mu$ is $\widetilde{V}$-saturated.*

*2.* $V \subseteq \widetilde{V}$ *whenever* $\mu$ *is* $V$*-saturated.*

*3.* $\widetilde{V}$ *is* $G_\Phi$*-invariant.*

*Proof.* A formal consequence of Lemma 3.21 (5) is that if $\mu$ is $\widetilde{V}_1$-saturated and $\widetilde{V}_2$-saturated then $\mu$ is $\widetilde{V}_1 + \widetilde{V}_2$-saturated. Thus we can take $\widetilde{V}$ to be the sum of all subspaces $V$ on which $\mu$ is saturated. (1) and (2) are then obvious, and (3) is a formal consequence of Lemma 6.2, Proposition 6.7 and Propositions 6.8, because taken together they show that if $\mu$ is $V$-saturated then $\mu$ is $V'$-saturated for a $G_\Phi$-invariant subspace $V'$, and $\dim V' \geq \dim V$ . Applying this to $V = \widetilde{V}$ we conclude $V' \subseteq \widetilde{V}$ and $\dim V' \geq \dim \widetilde{V}$ so $\widetilde{V} = V'$ is $G_\Phi$-invariant. □

We now need sufficient conditions for the subspace $\widetilde{V}$ from the last proposition to be of dimension $> 1$. To this end, we have the following.

**Proposition 6.10.** *If there exists a sequence* $V_i \in \text{sat}(\mu, \varepsilon_i, m_i, n_i)$ *with* $\varepsilon_i \to 0$, $m_i > m(\varepsilon_i)$ *and* $n_i > n(\varepsilon_i, m_i)$, *and if* $V_i \to V$, *then* $V \subseteq \widetilde{V}$, *where* $\widetilde{V}$ *is as in Proposition 6.9.*

*Proof.* In each of the three previous propositions, a $\delta = \delta(\varepsilon)$ was associated to an $\varepsilon$. We can assume that these functions $\delta$ are increasing (so decreasing $\varepsilon$ leads to no increase in $\delta(\varepsilon)$).

Let $V_i, \varepsilon_i, m_i, n_i$ be given. Assuming that $m_i, n_i$ are large enough relative to $\varepsilon_i$, by Proposition 6.7 there is a sequence $\varepsilon'_i = \varepsilon'_i(\varepsilon_i) \to 0$ depending monotonely on $\varepsilon_i$, such that for each $i$ there is a $(G_\Phi, \varepsilon'_i)$-invariant subspace $V_i^* \in \text{sat}(\mu, \varepsilon'_i, m_i, n_i)$ with $\angle(V_i, V_i^*) < \varepsilon'_i$ and $\dim V_i^* \geq \dim V_i$ (if $\delta(\cdot)$ is the function in that proposition than we choose $\varepsilon'_i = \delta^{-1}(\varepsilon_i)$).

We can henceforth assume that $m_i$ are large enough relative to $\varepsilon'_i$, and $n_i$ relative to $\varepsilon'_i, m_i$ (here we use that $\varepsilon'_i$ depends on $\varepsilon_i$ in a monotone way, so being large with respect to $\varepsilon'_i$ is the same as being large with respect to $\varepsilon_i$, which was assumed).

Passing to a subsequence we may assume that $V_i^*$ converge to some subspace $V^*$. Note that $V^*$ is $G_\Phi$-invariant, being the limit of $(G_\Phi, \varepsilon'_i)$-invariant subspaces.

By increasing $\varepsilon'_i$ if needed, we can assume that $m'_i = [\log(2/\varepsilon'_i)] \to \infty$ more slowly than linearly.

By Proposition 6.8 we can choose $\varepsilon''_i \to 0$, depending monotonely on $\varepsilon'_i$ such that if $W \in \text{sat}(\mu, \varepsilon'_i, m_i, n_i)$ is a $G_\Phi$-invariant subspace, and $d(V_i^*, W) < \varepsilon'_i$, then $W \in \text{sat}(\mu, \varepsilon''_i, m'_i, n')$, for all $n' > n_i + m'_i$ (recall that $m'_i = [\log(2/\varepsilon'_i)]$; if $\delta(\cdot)$ is the function in that proposition, choose $\varepsilon''_i = \delta^{-1}(\varepsilon'_i)$). Note that since we assumed that $m'_i \to \infty$ more slowly than linearly, every large enough integer occurs as $m'_i$ for some $i$.

Applying the previous paragraph to $W = V^*$, and since we have arranged that $\{m'_i\}$ includes all large enough integers, we see that $\mu$ if $V^*$-saturated. Thus $V^* \subseteq \widetilde{V}$.

Finally, combining $\angle(V_i, V_i^*) \to 0$ with $V_i^* \to V^*$ and $V_i \to V$, we conclude that $\angle(V^*, V) = 0$. Since $\dim V^* = \lim \dim V_i^* \geq \lim \dim V_i = \dim V$, we must have $V \subseteq V^*$. Since $V^* \subseteq \widetilde{V}$ we get $V \subseteq \widetilde{V}$, as claimed. □

## 6.4 Entropy and dimension for self-similar measures

If $\mu \in \mathcal{P}([0,1]^d)$ is exact dimensional, as self-similar measures are, the dimension of $\mu$ is given by the so-called entropy dimension:

$$\dim \mu = \lim_{n\to\infty} H_n(\mu). \tag{73}$$

We require a similar expression relating the dimension of conditional measures on affine subspaces to entropy. We parametrized affine subspaces as the set of fibers $\pi^{-1}(y)$ where $\pi$ is a linear map $\pi : \mathbb{R}^d \to \mathbb{R}^k$ and $y$ ranges over $\mathbb{R}^k$. The conditional measure $\mu_{\pi^{-1}(y)}$ of $\mu$ on $\pi^{-1}(y)$ is defined for $\pi\mu$-a.e. $y$ by the weak-* limit

$$\mu_{\pi^{-1}(y)} = \lim_{\ell\to\infty} \mu_{\pi^{-1}\mathcal{D}^k_\ell(y)},$$

which exists by the measure-valued version of the Martingale convergence theorem.

**Theorem 6.11.** *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$ be a self similar measure for the IFS $\Phi$ and let $\pi : \mathbb{R}^d \to \mathbb{R}^k$ be a linear map such that $\ker \pi$ is $D\Phi$-invariant. Then the conditional measure $\mu_{\pi^{-1}(y)}$ is exact dimensional for $\pi\mu$-a.e. $y$, and the dimension is given by*

$$\begin{aligned}\dim \mu_{\pi^{-1}(y)} &= \lim_{p\to\infty}\left(\liminf_{n\to\infty} \mathbb{E}_{0\le i\le n}\left(\frac{1}{p}H(\mu_{x,i}, \mathcal{D}_{i+p}|\pi^{-1}\mathcal{D}^k_{i+p})\right)\right)\\ &= \lim_{p\to\infty}\left(\limsup_{n\to\infty} \mathbb{E}_{0\le i\le n}\left(\frac{1}{p}H(\mu_{x,i}, \mathcal{D}_{i+p}|\pi^{-1}\mathcal{D}^k_{i+p})\right)\right)\end{aligned}$$

We will apply this theorem via the following corollary:

**Corollary 6.12.** *If $\mu$ is self-similar and $V$ is a saturated and $G_\Phi$-invariant subspace, then the conditional measures of $\mu$ on translates of $V$ are a.s. exact dimensional and of dimension $\dim V$. In particular this holds for the subspace described in Proposition 6.9.*

The proof we present for Theorem 6.11 has two ingredients. The first is exact dimensionality and dimension conservation:

**Theorem 6.13.** *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$ be a self similar measure for the IFS $\Phi$ and let $\pi : \mathbb{R}^d \to \mathbb{R}^k$ be a linear map such that $\ker \pi$ is $D\Phi$-invariant. Then $\pi\mu$ is exact dimensional, $\mu_{\pi^{-1}(y)}$ is exact dimensional for $\pi\mu$-a.e. $y$, its dimension is $\pi\mu$-a.s. independent of $y$, and*

$$\dim \pi\mu + \dim \mu_{\pi^{-1}(y)} = \dim \mu \qquad \text{for } \pi\mu\text{-a.e. } y.$$

This theorem follows from work of Falconer and Jin [7] (which in turn relies on methods of Feng and Hu [8]). Next, we require an expression for $\dim \pi\mu$ in terms of entropy of dyadic partitions. A special case of this result appears in [13] for the case that $G_\Phi$ is the full orthogonal group.

**Theorem 6.14.** *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$ be a self similar measure for the IFS $\Phi$ and let $\pi : \mathbb{R}^d \to \mathbb{R}^k$ be a linear map such that $\ker \pi$ is $D\Phi$-invariant. Then*

$$\begin{aligned}\dim \pi\mu &= \lim_{p\to\infty}\left(\liminf_{n\to\infty} \mathbb{E}_{0\le i\le n}\left(\frac{1}{p}H(\mu_{x,i}, \pi^{-1}\mathcal{D}^k_{i+p}\right)\right)\\ &= \lim_{p\to\infty}\left(\limsup_{n\to\infty} \mathbb{E}_{0\le i\le n}\left(\frac{1}{p}H(\mu_{x,i}, \pi^{-1}\mathcal{D}^k_{i+p}\right)\right).\end{aligned}$$

*Proof.* First, note that

$$\mathbb{E}_{0\leq i\leq n}\left(\frac{1}{p}H(\mu_{x,i},\pi^{-1}\mathcal{D}^k_{i+p})\right)=\mathbb{E}_{0\leq i\leq n}\left(\frac{1}{p}H(\mu,\pi^{-1}\mathcal{D}^k_{i+p}|\mathcal{D}_i)\right),$$

As we have seen, changing the dyadic partition to one adapted to a different coordinate system changes the right hand side of the last equation by $O(1/p)$, and in the statement of the theorem we consider the limit as $p\to\infty$. Thus, the statement is unaffected by changes to the coordinate system, and we may assume that $\pi$ is a coordinate projection. Therefore we can apply the local entropy averages lemma for projections [13]. The lemma is usually formulated for lower pointwise dimension, but the same proof exactly, replacing lim inf by lim sup, shows that

$$\limsup_{n\to\infty}-\frac{1}{n}\log(\mu((\pi^{-1}\mathcal{D}^k_n)(x)))\geq\limsup_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\frac{1}{p}H(\mu_{x,i},\pi^{-1}\mathcal{D}^k_{i+p})-O(\frac{1}{p})\qquad\mu\text{-a.e. }x.$$

Since $\pi\mu$ is exact dimensional, the left hand side is $\mu$-a.s. equal to $\dim\pi\mu$, and we have

$$\dim\pi\mu\geq\limsup_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\frac{1}{p}H(\mu_{x,i},\pi^{-1}\mathcal{D}^k_{i+p})-O(\frac{1}{p})\qquad\mu\text{-a.e. }x$$

Integrating this $d\mu$ and using Fatou's lemma, for all $p$,

$$\begin{aligned}\dim\pi\mu &\geq \limsup_{n\to\infty}\int\left(\frac{1}{n}\sum_{i=0}^{n-1}\frac{1}{p}H(\mu_{x,i},\pi^{-1}\mathcal{D}^k_{i+p})\right)d\mu(X)-O(\frac{1}{p})\\ &= \limsup_{n\to\infty}\mathbb{E}_{0\leq i\leq n}\left(\frac{1}{p}H(\mu_{x,i},\pi^{-1}\mathcal{D}^k_{i+p})\right)-O(\frac{1}{p}).\end{aligned}\tag{74}$$

Equation (74) is one half of the inequality we are after, and its proof only used exact dimensionality of $\mu$. For the reverse inequality we will use self-similarity. Fix $p$, and note that we have the identity

$$\mathbb{E}_{0\leq i\leq n}\left(\frac{1}{p}H(\mu_{x,i},\pi^{-1}\mathcal{D}^k_{i+p})\right)=\mathbb{E}_{0\leq i\leq n}\left(\frac{1}{p}H(\mu,\pi^{-1}\mathcal{D}^k_{i+p}|\mathcal{D}_i)\right)$$

where the expectation on the left is over $i$ and $x$, and on the right only over $i$. Let $r=\min\{r_i\,:\,i\in\Lambda\}$, and for each $i$ let $I_i\subseteq\Lambda^*$ denote the set of sequences $j_1\ldots j_k\in\Lambda^*$ such that $r\cdot 2^{-i}<r_{j_1}\ldots r_{j_k}<2^{-i}\leq r_{j_1}\ldots r_{j_{k-1}}$. It is a standard (and easy) fact that $\mu=\sum_{j\in I_i}p_j\cdot\varphi_j\mu$. By concavity of conditional entropy (Lemma 3.1(5)), for each $i$,

$$\begin{aligned}\frac{1}{p}H(\mu,\pi^{-1}\mathcal{D}^k_{i+p}|\mathcal{D}_i) &\geq \frac{1}{p}\sum_{i\in I_i}p_iH(\varphi_i\mu,\pi^{-1}\mathcal{D}_{i+p}|\mathcal{D}_i)+O(\frac{1}{p})\\ &= \frac{1}{p}\sum_{i\in I_i}p_iH(\varphi_i\mu,\pi^{-1}\mathcal{D}_{i+p})+O(\frac{1}{p}).\end{aligned}$$

where we used the fact that each $\varphi_i\mu$, $i\in I_i$, has diameter $O(2^{-i})$, and Lemma 3.2(2). Finally, since $\varphi_i$ contracts by $2^{-i}$ up to a constant factor, by changing scale, applying Lemma 3.2(5), and changing the coordinates system, we have

$$\frac{1}{p}\sum_{i\in I_i} p_i H(\varphi_i\mu, \pi^{-1}\mathcal{D}_{i+p}) = \frac{1}{p}H(\mu, \pi^{-1}\mathcal{D}_p) + O(\frac{1}{p}).$$

Note that we used here the fact that $\ker \pi$ is invariant under the linear part of $\varphi_i$.

Putting this all together, we have shown that for every $p$,

$$\mathbb{E}_{0\le i\le n}\left(\frac{1}{p}H(\mu_{x,i}, \pi^{-1}\mathcal{D}^k_{i+p})\right) \ge \frac{1}{p}H(\pi\mu, \mathcal{D}_p) + O(\frac{1}{p}).$$

Taking the lim inf as $n\to\infty$, we have

$$\liminf_{n\to\infty} \mathbb{E}_{0\le i\le n}\left(\frac{1}{p}H(\mu_{x,i}, \pi^{-1}\mathcal{D}^k_{i+p})\right) \ge \frac{1}{p}H(\pi\mu, \mathcal{D}_p) + O(\frac{1}{p}).$$

But, since $\pi\mu$ is exact dimensional, as $p\to\infty$ the right hand side tends to $\dim \pi\mu$. Combined with inequality (74), this proves the statement. □

We can now prove Theorem 6.11. Begin with the identity

$$\begin{aligned}\mathbb{E}_{0\le i\le n}\left(\frac{1}{p}H(\mu_{x,i}, \mathcal{D}_{i+p}|\pi^{-1}\mathcal{D}^k_{i+p})\right) &= \mathbb{E}_{0\le i\le n}\left(\frac{1}{p}H(\mu_{x,i}, \mathcal{D}_{i+p})\right) \\ &- \mathbb{E}_{0\le i\le n}\left(\frac{1}{p}H(\mu_{x,i}, \pi^{-1}\mathcal{D}^k_{i+p})\right)\end{aligned}$$

(this is just Lemma 3.1 (4) and linearity of expectation). Taking $n\to\infty$ and then $p\to\infty$, and using (73) and Theorem 6.14, the right hand side becomes $\dim\mu - \dim\pi\mu$, which by Theorem 6.13 is the a.s. dimension of fibers.

## 6.5 Proof of Theorem 1.5

Recall from the introduction that $r = \prod_{i\in\Lambda} r_i^{p_i}$, $n' = n\log(1/r)$ and $\nu^{(n)} = \sum_{i\in\Lambda^n} p_i\cdot\delta_{\varphi_i}$. Also recall the definition of the dyadic partition $\mathcal{D}_n = \mathcal{D}^G_n$, and the partition $\mathcal{E}_n = \mathcal{E}^G_n$ of $G$ according to the level-$n$ dyadic partition of the translation part of the maps. In this section we prove the following:

**Theorem 6.15.** *Let $\Phi = \{\varphi_i\}$ be an IFS on $\mathbb{R}^d$ that does not preserve a non-trivial affine subspace, and $\mu$ a self-similar measure for $\Phi$. Then either*

$$\lim_{n\to\infty}\frac{1}{n'}H(\nu^{(n)}, \mathcal{D}^G_{qn}|\mathcal{E}^G_{n'}) = 0 \quad \text{for all } q>1,$$

*or else there is a $D\Phi$-invariant subspace $V$ such that $\dim\mu_{V+x} = \dim V$ for $\mu$-a.e. $x$.*

This implies Theorem 1.5, see remark after its statement.

We begin the proof. First, note that $\mu(V) = 0$ for every proper affine subspace $V\subseteq\mathbb{R}^d$, since if $\mu(V)>0$ for some $V$ then it is easily shown that $\mu$ is supported on $V$, and hence $\Phi$ preserves $V$, contrary to hypothesis.

We now argue by contradiction: suppose that there is a $\delta_0>0$ and $q>1$ such that

$$\limsup_{n\to\infty}\frac{1}{qn'}H(\nu^{(n)}, \mathcal{D}_{(q+1)n'}|\mathcal{E}_{n'}) > \delta_0.$$

Let $\mathcal{F}$ denote the partition of $G$ according to the contraction ratio. This is an uncountable partition, but the possible contractions of $\varphi_i$, $i \in \Lambda^n$, are just all the $n$-fold products of the contractions $r_i$, $i \in \Lambda$. Thus only $O(n^{|\Lambda|+1})$ distinct contraction ratios occur in the support of $\nu^{(n)}$, so

$$\lim_{n\to\infty} \frac{1}{qn'} H(\nu^{(n)}, \mathcal{F}) = \lim_{n\to\infty} \frac{O(\log n)}{n'} = 0.$$

Using the identities $H(\cdot, \mathcal{D}|\mathcal{E} \vee \mathcal{F}) = H(\cdot, \mathcal{D}|\mathcal{E}) + H(\cdot, \mathcal{F}|\mathcal{E})$ and $H(\cdot, \mathcal{F}|\mathcal{E}) \leq H(\cdot, \mathcal{F})$, the two limits above imply

$$\limsup_{n\to\infty} \frac{1}{qn'} H(\nu^{(n)}, \mathcal{D}_{(q+1)n'}|\mathcal{E}_{n'} \vee \mathcal{F}) > \delta_0. \tag{75}$$

**Lemma 6.16.** $\lim_{n\to\infty} \int \frac{1}{qn'} H(g.\mu, \mathcal{D}_{(q+1)n'}|\mathcal{D}_{n'})\, d\nu^{(n)}(g) = \dim \mu$

*Proof.* If $g = 2^{-t}U + a$, then $g.\mu$ is supported on a set of diameter $O(2^{-t})$, hence $H(g.\mu, \mathcal{D}_{n'}) = O(|t - n'|)$. Similarly, by Lemma 3.2 (5), $H(g.\mu, \mathcal{D}_{(q+1)n'}) = H(\mu, \mathcal{D}_{qn'}) + O(|t - n'|)$.

If we choose $g = 2^{-t}U + a$ randomly according to $\nu^{(n)}$, then $t$ is distributed as the sum of $n$ independent random variables, each of which takes value $\log(1/r_i)$ with probability $p_i$ for $i \in \Lambda$, so by the law of large numbers, $t - n' = o(n')$ in probability. We also have a worst-case bound of $t \leq Cn$ (a.s. for $g \sim \nu^{(n)}$), because $\varphi_{i_1,\ldots,i_n}$ contracts by at least $(\min_{i\in\Lambda} r_i)^n$, and $\min_{i\in\Lambda} r_i < 1$. Hence the bound $t - n' = o(n')$ holds also in the mean sense. It follows from the first paragraph that

$$\begin{aligned}
\frac{1}{qn'} \int H(g.\mu, \mathcal{D}_{n'}) d\nu^{(n)}(g) &= o(1) \\
\frac{1}{qn'} \int H(g.\mu, \mathcal{D}_{(q+1)n'}) d\nu^{(n)}(g) &= \frac{1}{qn'} H(\mu, \mathcal{D}_{qn'}) + o(1) \\
&= \dim \mu + o(1)
\end{aligned}$$

Subtracting the first line from the second proves the claim. □

**Lemma 6.17.** $\lim_{n\to\infty} \frac{1}{qn'} H(\mu, \mathcal{D}_{(q+1)n'}|\mathcal{D}_{n'}) = \dim \mu$.

*Proof.* Using the conditional entropy formula,

$$\begin{aligned}
\frac{1}{(q+1)n'} H(\mu, \mathcal{D}_{(q+1)n'}) &= \frac{1}{(q+1)} \cdot \frac{1}{n'} H(\mu, \mathcal{D}_{n'}) + \\
&\quad + \frac{q}{(q+1)} \cdot \frac{1}{qn'} H(\mu, \mathcal{D}_{q(n'+1)}|\mathcal{D}_{n'}).
\end{aligned}$$

The lemma follows by taking $n \to \infty$ and using the fact that $\frac{1}{n} H(\mu, \mathcal{D}_n) \to \dim \mu$. □

Let $\nu_I^{(n)}$ denote, as usual the, conditional measure of $\nu^{(n)}$ on $I$.

**Lemma 6.18.** $\lim_{n\to\infty} \left( \sum_{I\in\mathcal{E}_{n'}\vee\mathcal{F}} \nu(I) \cdot \frac{1}{qn'} H(\nu_I^{(n)}.\mu, \mathcal{D}_{(q+1)n'}) \right) = \dim \mu$.

*Proof.* Write

$$\mu = \sum_{I\in\mathcal{E}_{n'}\vee\mathcal{F}} \nu(I)\cdot\left(\nu_I^{(n)}\mathbin{\bullet}\mu\right).$$

and note that

$$\nu_I^{(n)}\mathbin{\bullet}\mu = \int g\mathbin{\bullet}\mu \, d\nu_I^{(n)}(g)$$

Combining this with concavity of conditional entropy (Lemma 3.1 (5)) and the previous two lemmas,

$$\begin{aligned}
\dim\mu &= \lim_{n\to\infty}\frac{1}{qn'}H(\mu,\mathcal{D}_{(q+1)n'}|\mathcal{D}_{n'}) \\
&\geq \limsup_{n\to\infty}\sum_{I\in\mathcal{E}_{n'}\vee\mathcal{F}}\nu^{(n)}(I)\cdot\frac{1}{qn'}H(\nu_I^{(n)}\mathbin{\bullet}\mu,\mathcal{D}_{(q+1)n'}|\mathcal{D}_{n'}) \\
&\geq \liminf_{n\to\infty}\sum_{I\in\mathcal{E}_{n'}\vee\mathcal{F}}\nu^{(n)}(I)\cdot\frac{1}{qn'}H(\nu_I^{(n)}\mathbin{\bullet}\mu,\mathcal{D}_{(q+1)n'}|\mathcal{D}_{n'}) \\
&\geq \liminf_{n\to\infty}\sum_{I\in\mathcal{E}_{n'}\vee\mathcal{F}}\nu^{(n)}(I)\cdot\int\frac{1}{qn'}H(g\mathbin{\bullet}\mu,\mathcal{D}_{(q+1)n'}|\mathcal{D}_{n'})\,d\nu_I^{(n)}(g) \\
&= \lim_{n\to\infty}\int\frac{1}{qn'}H(g\mathbin{\bullet}\mu,\mathcal{D}_{(q+1)n'}|\mathcal{D}_{n'})\,d\nu^{(n)}(g) \\
&= \dim\mu,
\end{aligned}$$

as claimed. ☐

For $I\in\mathcal{E}_{n'}\vee\mathcal{F}$ consisting of similarities with contraction $2^{-t}$, define

$$\widetilde{\nu}_I^{(n)} = S_t\nu_I^{(n)}$$

This is a measure on the isometry group $G_0$.

**Lemma 6.19.** *For every $\delta>0$ and for arbitrarily large $n$ we can find $I\in\mathcal{E}_{n'}\vee\mathcal{F}$ with $\nu(I)>0$ and such that*

$$\frac{1}{qn'}H(\widetilde{\nu}_I^{(n)},\mathcal{D}_{qn'})>\delta_0$$

*and*

$$\frac{1}{qn'}H(\widetilde{\nu}_I^{(n)}\mathbin{\bullet}\mu,\mathcal{D}_{qn'})<\frac{1}{qn'}H(\mu,\mathcal{D}_{qn'})+\delta.$$

*Proof.* By (75), for infinitely many $n$ we have

$$\begin{aligned}
\frac{1}{qn'}\sum_{I\in\mathcal{E}_{n'}\vee\mathcal{F}}\nu^{(n)}(I)\cdot H(\nu_I^{(n)},\mathcal{D}_{(q+1)n'}) \\
= \frac{1}{qn'}H(\nu^{(n)},\mathcal{D}_{(q+1)n'}|\mathcal{E}_{n'}\vee\mathcal{F}) \\
> \delta_0
\end{aligned}\tag{76}$$

Suppose $I\in\mathcal{E}_{n'}\vee\mathcal{F}$ contains similitudes of contraction $t$. Since the action of $S_t$ on $G$ is just ordinary scaling in our coordinates on $G$, we have

$$\left|H(\nu_I^{(n)},\mathcal{D}_{(q+1)n'})-H(\widetilde{\nu}_I^{(n)},\mathcal{D}_{qn'})\right| = O(|t-n'|),$$

Using the fact that for $g = 2^{-t}U + a \sim \nu^{(n)}$ we have $t - n' = o(n')$ in probability as $n \to \infty$, and the pointwise bound $t = O(n')$, this and (76) imply that there are infinitely many $n$ such that

$$\sum_{I \in \mathcal{E}_{n'} \vee \mathcal{F}} \nu^{(n)}(I) \cdot \frac{1}{qn'} H(\widetilde{\nu}_I^{(n)}, \mathcal{D}_{qn'}) > \delta_0. \tag{77}$$

Similarly, we have $\widetilde{\nu}_I^{(n)} \boldsymbol{.} \mu = S_t(\nu_I^{(n)} \boldsymbol{.} \mu)$, so by Lemma 3.1 (5),

$$\left| H(\nu_I^{(n)} \boldsymbol{.} \mu, \mathcal{D}_{(q+1)n'}) - H(\widetilde{\nu}_I^{(n)} \boldsymbol{.} \mu, \mathcal{D}_{qn'}) \right| = O(|t - n'|).$$

Using the previous lemma and again the fact that $|t - n'| = o(n')$ in probability as $g = 2^{-t}U + a \sim \nu^{(n)}$,

$$\lim_{n \to \infty} \left( \frac{1}{qn'} \sum_{I \in \mathcal{E}_{n'} \vee \mathcal{F}} \nu^{(n)}(I) \cdot H(\widetilde{\nu}_I^{(n)} \boldsymbol{.} \mu, \mathcal{D}_{qn'}) \right) = \dim \mu.$$

On the other hand we know that also

$$\lim_{n \to \infty} \left( \frac{1}{qn'} H(\mu, \mathcal{D}_{qn'}) \right) = \dim \mu.$$

Therefore (using boundedness of the normalized entropy),

$$\lim_{n \to \infty} \sum_{I \in \mathcal{E}_{n'} \vee \mathcal{F}} \nu^{(n)}(I) \cdot \left| \frac{1}{qn'} H(\widetilde{\nu}_I^{(n)} \boldsymbol{.} \mu, \mathcal{D}_{qn'}) - \frac{1}{qn'} H(\mu, \mathcal{D}_{qn'}) \right| = 0. \tag{78}$$

Combining this with (77), for infinitely many $n$ we can find $I \in \mathcal{E}_{n'} \vee \mathcal{F}$ with the desired properties. $\square$

Now fix a parameter $\varepsilon > 0$, and let $\sigma > 0$ be such that $\mu$ is $((\varepsilon/5d)^{2(d+1)}, \sigma)$-non-affine (recall Definition 2.11). Such $\sigma$ exists because by assumption $\mu$ gives mass 0 to every proper affine subspace. Choose large $m \in \mathbb{N}$, and let $\delta > 0$ and $k \in \mathbb{N}$, be as in the conclusion of Theorem 2.12. Apply the theorem to the measures $\widetilde{\nu}_I^{(n)}$ for the set $I \in \mathcal{E}_{n'} \vee \mathcal{F}$ found in the previous lemma for the parameter $\delta$. We have arrived at the following conclusion:

> For every $\varepsilon > 0$, for arbitrarily large $n$, a $(1 - \varepsilon)$-fraction of the level-$k$ components $\theta = \mu_{x,k}$ of $\mu$ have associated to them a sequence of subspaces $V_1, \ldots, V_n$ of which at least a $c\delta_0$-fraction are of dimension $\geq 1$, and which satisfy
>
> $$\mathbb{P}_{0 \leq i \leq n} \left( \theta^{y,i} \text{ is } (V_i, \varepsilon, m)\text{-saturated} \right) > 1 - \varepsilon. \tag{79}$$

If the last equation held for $\mu$ instead of $\theta$ (possibly for a different sequence of subspaces), we would be in a position to apply Proposition 6.9 (4), which would give the second alternative of the present theorem. This "bootstrapping" from the component $\theta$ to $\mu$ is accomplished as follows. Let us say that a probability measure $\eta \in \mathcal{P}(\mathbb{R}^d)$ is fragmented at level $k$ if $\nu(D) > 0$ for at least two distinct $D \in \mathcal{D}_k$, otherwise it is unfragmented. We again abbreviate $\sum A = \sum_{a \in A} A$.

**Lemma 6.20.** *Given $k$, if $s \in \mathbb{N}$ is large enough, then*

$$\sum\{p_i \,:\, i \in \Lambda^s \text{ and } \varphi_i\mu \text{ is unfragmented at level } k\} > 1 - \varepsilon.$$

*Proof.* Let $E = \bigcup \partial D$, where the union is over $D \in \mathcal{D}_k$ such that $\operatorname{supp}\mu \cap \overline{D} \neq \emptyset$. Then $E$ is contained in the union of finitely many proper affine subspaces, so for a small enough $\rho > 0$ we will have $\mu(E^{(\rho)}) < \varepsilon$. Let $s$ be large enough that for $i \in \Lambda^s$ the measure $\varphi_i\mu$ is supported on a set of diameter $< \rho$. This means that if $\varphi_i\mu$ is fragmented then it is supported on $E^{(\rho)}$. Since $\mu = \sum_{i\in\Lambda^s} p_i \cdot \varphi_i\mu$, we conclude that

$$\begin{aligned}\sum\{p_i \,:\, i \in \Lambda^s \text{ and } \varphi_i\mu \text{ is fragmented at level } k\} &\leq \mu(E^{(s)}) \\ &< \varepsilon,\end{aligned}$$

as required. □

Let $s$ as in the lemma for the $k$ we found previously. Assuming $\varepsilon < 1/2$, by the lemma and our previous discussion we can find a level-$k$ component $\theta = \mu_D$, $D \in \mathcal{D}_k$, of $\mu$, for which (79) holds and, furthermore, $1 - \varepsilon$ of the mass of $\theta$ comes from components $\varphi_i\mu$, $i \in \Lambda^s$, supported entirely on $D$. We can now apply Lemma 6.5 (2) to conclude that there is an $i \in \Lambda^s$ such that for arbitrarily large $j$ there is a $V_n \in \operatorname{sat}(\varphi_i\mu, \varepsilon', m, j)$, where $\varepsilon' \to 0$ as $\varepsilon \to 0$. By Lemma 6.5 the same is true of $\mu$ for the subspace $U_i^{-1}V_j$ and some $\varepsilon''$ that also vanishes as $\varepsilon \to 0$. We can now invoke Proposition 6.10, which completes the proof.

## 6.6 Transversality and the dimension of exceptions

In this section we prove Theorems 1.10 and 1.11 on the dimension of exceptional parameters for parametric families of self-similar sets and measures. We adopt the notation from the introduction, so $\varphi_{i,t}(x) = r_i(t)U_i(t)x + a_i(t)$ are contracting similarities for $t$ in a compact connected set $I \subseteq \mathbb{R}^m$, for $i = i_1 \dots i_n \in \Lambda^n$ we define $\varphi_{i,t} = \varphi_{i_1,t} \circ \dots \circ \varphi_{i_n,t}$ and similarly $r_i(t)$ and $U_i(t)$. Recall that $\Delta_{i,j}(t) = \varphi_{i,t}(0) - \varphi_{j,t}(0)$ and define

$$\Delta'_n(t) = \min_{i\neq j\in\Lambda^n} \|\Delta_{i,j}(t)\|\,.$$

If, as in the introduction, we write $\Delta_n(t)$ for the minimum of $d(\varphi_{i,t}, \varphi_{j,t})$ over distinct $i, j \in \Lambda^n$, then we have $\Delta'_n \leq \Delta_n$ and hence $(\Delta'_n)^{-1}((-\varepsilon,\varepsilon)^d) \supseteq (\Delta_n)^{-1}((-\varepsilon,\varepsilon)^d)$. In particular, in order to prove Theorem 1.10, one may replace the set $E$ there with the set $E' = \bigcap_{\varepsilon>0} E'_\varepsilon$, where

$$E'_\varepsilon = \bigcup_{N=1}^{\infty} \bigcap_{n>N} \left( \bigcup_{i,j\in\Lambda^n} (\Delta_{i,j})^{-1}((-\varepsilon^n, \varepsilon^n)^d) \right).$$

Thus we wish to show that, under suitable hypotheses, $\dim_P E'_\varepsilon \to 0$ as $\varepsilon \to 0$.

We begin with the proof of Theorem 1.11. We require an elementary fact whose proof we include for completeness.

**Lemma 6.21.** *Let $V \subseteq \mathbb{R}^m$ be open and let $F : V \to \mathbb{R}^k$ be a $C^2$ map. Suppose that $K \subseteq V$ is compact and that $\operatorname{rank} DF \geq r$ everywhere in $K$. Then $K \cap F^{-1}((-\delta,\delta)^k)$ can be covered by at most $C \cdot 1/\delta^{m-r}$ balls of radius $\delta$, where $C$ depends only on the diameter of $K$ and the magnitude of the first and second partial derivatives of $F$ on $K$.*

*Proof.* We first reduce to the case that $k = r$. Assume this case is known. Consider the general case $k \geq r$. For each $r$-tuple of distinct coordinates $i = (i_1, \ldots, i_r) \in \{1, \ldots, k\}^r$, let $\pi_i : \mathbb{R}^k \to \mathbb{R}^r$ denote the projection to these coordinates. Now, if $\operatorname{rank} DF(x) \geq r$ then $\operatorname{rank} D(\pi_i \circ F)(x) \geq r$ for some $r$-tuple $i$, so we can find an open cover $V = \bigcup V_i$ indexed by tuples as above such that $D(\pi_i \circ F)$ has rank $r$ everywhere in $K \cap V_i$. Choose compact sets $K_i \subseteq K$ such that $K_i \subseteq V_i$ and $K = \bigcup K_i$. By our assumption, for each $i$ the set $K_i \cap (\pi_i \circ F)^{-1}((-\delta, \delta)^r)$ can be covered by $O(1/\delta^{m-r})$ balls of radius $\delta$. If $x \in F^{-1}((-\delta, \delta)^k)$ then certainly $x \in (\pi_i \circ F)^{-1}((-\delta, \delta)^r)$ for every tuple $i$, so the union of these $\binom{k}{r}$ covers is a cover of $K \cap F^{-1}((-\delta, \delta)^k)$ containing at most $\binom{k}{r} O(1/\delta^{m-r})$ balls of radius $\delta$, as required (note that restricting the function and composing with a projection can only decrease its $C^2$ norm, so the constant does not get worse).

Thus we may from the start assume that $k = r$ and that $\operatorname{rank} DF = r$ everywhere in $K$. Let $M$ denote the bound on the first and second derivatives of $F|_K$. Applying the constant rank theorem [21, Theorem 7.8], for each $x \in K$ there is a neighborhood $W_x \subseteq \mathbb{R}^m$ of $x$ and an open set $W'_x \subseteq \mathbb{R}^r$ such that $F|_{W_x} : W_x \to W'_x$ is a diffeomorphism and is $C^2$-conjugate to the projection $\pi_{1,\ldots,r} : \mathbb{R}^m \to \mathbb{R}^r$. The distortion of the conjugating maps is controlled by $M$. Since for $\pi_{1,\ldots,r}$ the statement is clear, the conclusion follows for $F|_{W_x}$. Finally, the neighborhoods $W_x$ contain balls centered at $x$ with radius again bounded in terms of $M$. Only $O((\operatorname{diam} K)^m)$ of these neighborhoods are needed to cover $K$, and the statement follows. □

Returning to our parametrized family of IFSs, assume that $D\Delta_{i,j}$ has rank at least $r$ at every point in $I$ and every distinct pair $i, j \in \Lambda^{\mathbb{N}}$.

**Lemma 6.22.** *For large enough $n$ and all $i, j \in \Lambda^n$ the rank of $\Delta_{i,j}$ is at least $r$ everywhere in $I$.*

*Proof.* It is easy to check that the power series for the functions $\Delta_{i,j}$ converge on a common neighborhood of $I$ in $\mathbb{C}^m$ (each function being defined by its complex power series), and since $\Delta_{i_1 \ldots i_n, j_1 \ldots j_n} \to \Delta_{i,j}$ uniformly on this neighborhood, we find that $D_v \Delta_{i_1,\ldots,i_n,j_1,\ldots,j_n} \to D_v \Delta_{i,j}$ as $n \to \infty$ for all $v$. The lemma now follows by a compactness argument. □

Finally, it is again clear that there is a uniform bound $M$ for the first and second derivatives of all the functions $\Delta_{i,j}$, $i, j \in \Lambda^n$. The proof of Theorem 1.11 is now concluded as follows. For large enough $n$, for each $i, j \in \Lambda^n$, the set

$$(\Delta_{i,j})^{-1}((-\varepsilon^n, \varepsilon^n)^d)$$

can be covered $O_M(1/\varepsilon^{n(m-r)})$ balls of radius $\varepsilon^n$. Thus the set

$$E'_{n,\varepsilon} = \bigcup_{i,j \in \Lambda^n} (\Delta_{i,j})^{-1}((-\varepsilon^n, \varepsilon^n)^d)$$

satisfies

$$N(E'_{n,\varepsilon}, \varepsilon^n) \leq |\Lambda|^n \cdot O_M(1/\varepsilon^{n(m-r)})$$

where $N(X, \delta)$ is the $\delta$-covering number of $X$. Thus for every $\varepsilon$ and $n$,

$$\begin{aligned} N(\bigcap_{k>n} E'_{k,\varepsilon}, \varepsilon^n) &\leq N(E'_{n,\varepsilon}.\varepsilon^n) \leq \\ &\leq |\Lambda|^n \cdot O_M(1/\varepsilon^{n(m-r)}), \end{aligned}$$

hence for each $\varepsilon > 0$,

$$\begin{aligned}\mathrm{bdim}\left(\bigcap_{k>n} E'_{k,\varepsilon}\right) &= \limsup_{n\to\infty} \frac{\log N\left(\bigcap_{k>n} E'_{k,\varepsilon}, \varepsilon^n\right)}{\log(1/\varepsilon^n)} \\ &\leq m - r + \frac{\log|\Lambda|}{\log(1/\varepsilon)},\end{aligned}$$

It follows that $\dim_{\mathrm{P}} E'_\varepsilon \leq m - r + \log|\Lambda|/\log(1/\varepsilon)$, and this tends to $m - r$ as $\varepsilon \to 0$, as required.

We now turn to the the proof of Theorem 1.10, which is very similar to the one-parameter case from [12]. Let $|i|$ denote the length of a sequence $i$ and for sequences $i, j$ let $i \wedge j$ denote the longest common initial segment of sequences (which may be 0). Let

$$\mathcal{B}_m = \{e_1, \ldots, e_m\}$$

denote the standard basis of $\mathbb{R}^m$ and let $D_v$ denote the directional derivative operator in direction $v$. Thus for $F = (F_1, \ldots, F_d) : I \to \mathbb{R}^d$ we have $D_v F = (D_v F_1, \ldots, D_v F_d) : I \to \mathbb{R}^d$. We also write $D$ for the differentiation operator for functions $\mathbb{R}^m \to \mathbb{R}^d$. It will be convenient for the rest of this section to use the supremum norm on vectors and matrices.

**Definition 6.23.** Let $I \subseteq \mathbb{R}^m$ be a connected compact set. A family $\{\Phi_t\}_{t\in I}$ of IFSs is transverse of order $k$ if the associated functions $r_i(\cdot)$, $a_i(\cdot)$, $U_i(\cdot)$ are $(k+1)$-times continuously differentiable in a neighborhood of $I$, and there is a constant $c > 0$ such that for all $n \in \mathbb{N}$ and all $i, j \in \Lambda^n$,

$$\begin{aligned}&\forall\, t_0 \in I \quad \exists\, p \in \{0, \ldots, k\} \quad \exists\, v_1, \ldots, v_p \in \mathcal{B}_m \\ &\text{such that} \quad \left\|(D_{v_p} \ldots D_{v_1} \Delta_{i,j})(t_0)\right\| > c \cdot |i \wedge j|^{-p} \cdot r_{i\wedge j}(t_0).\end{aligned}$$

A real-analytic function defined $F : I \to \mathbb{R}^d$ can be extended to a complex-analytic function on an open complex neighborhood of $I$. Such an $F$ is identically 0 if and only if at some point $t_0 \in U$ we have $D_{v_1} \ldots D_{v_n} F(t_0) = 0$ for every $n$ and $v_1 \ldots v_n \in \mathcal{B}_m$. For $i, j \in \Lambda^{\mathbb{N}}$ the functions $\Delta_{i,j}$ are real analytic if $r_i, a_i, U_i$ are, because on the common neighborhood of $I$ in which these functions are analytic, $\Delta_{i,j}$ is given as an absolutely convergent powers series in these functions. Thus the $\Delta_{i,j}$ extend to complex-analytic functions on a common neighborhoods of $I$.

We have the following analog of [12, Proposition 5.7]:

**Proposition 6.24.** *Let $I \subseteq \mathbb{R}^m$ be a connected compact set and $\{\Phi_t\}_{t\in I}$ a family of IFSs on $\mathbb{R}^d$ whose associated functions $r_i(\cdot), a_i(\cdot), U_i(\cdot)$ are real analytic on $I$. For $i, j \in \Lambda^{\mathbb{N}}$, suppose that $\Delta_{i,j} \equiv 0$ on $I$ if and only if $i = j$. Then $\{\Phi_t\}$ is transverse of order $k$ for some $k$.*

*Proof.* For $i, j \in \Lambda^n$, let $\ell = |i \wedge j|$ and let $u, v \in \Lambda^{n-\ell}$ denote the sequences obtained by deleting the first $\ell$ symbols of $i, j$. Define the function $\widetilde{\Delta}_{i,j}$ by

$$\widetilde{\Delta}_{i,j}(t) = \Delta_{u,v}(t).$$

We find that

$$\Delta_{i,j}(t) = r_{i\wedge j}(t) \cdot U_{i\wedge j}(t)(\widetilde{\Delta}_{i,j}(t)),$$

Let $n(u)$ denote the number of times that the symbol $u \in \Lambda$ appears in $i \wedge j$ and let $U^T$ be the transpose of $U$. Then (since $U^T_{i\wedge j} = U^{-1}_{i\wedge j}$),

$$\widetilde{\Delta}_{i,j}(t) = (\prod_{u\in\Lambda} r_u(t)^{n(u)}) \cdot U^T_{i\wedge j}(t)\Delta_{i,j}(t).$$

From here the analysis is entirely analogous to the proof of [12, Proposition 5.7], bounding iterated directional derivatives rather than the higher derivative $\widetilde{\Delta}^{(p)}_{i,j}$ from the original proof. We omit the details. □

Our next task is to show that transversality of order $k$ provides efficient coverings of pre-images $(\Delta_{i,j})^{-1}((-\varepsilon,\varepsilon)^d)$. The argument is again very similar to the one-dimensional case but with some additional technicalities. The key part of the argument in dimension 1 was the fact that if $F : [a,b] \to \mathbb{R}$ satisfies $|F'| > c$, then $F^{-1}(-\rho,\rho)$ is an interval of length $\leq 2\rho/c$. We now generalize this to higher dimensions.

Let $U \subseteq \mathbb{R}^{m-1}$, let $f : U \to \mathbb{R}$ be a Lipschitz function with Lipschitz constant $c$, and $E = \{(x, f(x)) \in \mathbb{R}^m : x \in U\}$ be its graph. Then we say that $E$ is a $c$-Lipschitz graph in $\mathbb{R}^m$ with domain $U$. More generally we apply this name to any isometric image of $E$ in $\mathbb{R}^m$.

**Lemma 6.25.** *Let $E \subseteq \mathbb{R}^m$ be a $c$-Lipschitz graph with domain $U = B_r(x) \subseteq \mathbb{R}^{m-1}$ and let $0 < \varepsilon < r$. Then the $\varepsilon$-neighborhood of $E$ can be covered by $O((r/\varepsilon)^{m-1})$ balls of radius $\varepsilon$ if $c < 1$, and by $O((cr/\varepsilon)^{m-1})$ such balls if $c \geq 1$.*

*Proof.* Assume that $c \leq 1$. Let $y = (u, f(u))$ be a point in the graph. Let $y^{\pm} = (u, f(u) \pm \varepsilon/2)$. Then the union $C = C(u) = B_\varepsilon(y^+) \cup B_\varepsilon(y^-)$ contains the cylinder $B_{\varepsilon/2}(u) \times [-3\varepsilon/2, 3\varepsilon/2]$. Since $f$ is $c$-Lipschitz, this implies that $C$ contains the $\varepsilon$-neighborhood of the graph over $B_{\varepsilon/2}(u)$. Now cover $B_r(x)$ by $O((r/\varepsilon)^{m-1})$ balls $B_{\varepsilon/2}(u_i)$. Then $\bigcup C(u_i)$ is covered by $O((r/\varepsilon)^{m-1})$ $\varepsilon$-balls, and contains the $\varepsilon$-neighborhood of the graph.

If $c \geq 1$, then $C(u)$ contains an $\varepsilon$-neighborhood of the graph over $B_{\varepsilon/2c}(u)$, and we obtain the desired bound by covering $B_r(x)$ by $O((cr/\varepsilon)^{m-1})$ balls of radius $\varepsilon/2c$. □

**Lemma 6.26.** *Let $I \subseteq \mathbb{R}^m$ be a compact set, let $0 < \delta < 1$ and let $I^{(\delta)}$ denote the $\delta$-neighborhood of $I$, let $F : I^{(\delta)} \to \mathbb{R}$ be twice continuously differentiable with $0 < c \leq \|DF\| \leq M$ and $\|D^2F\| \leq M$ on $I^{(\delta)}$. We assume $c \leq 1$. Then for $0 < \rho < \min\{\delta, c/M\}$, the set $I \cap F^{-1}(-\rho,\rho)$ can be covered by $O_{M,\mathrm{vol}(I^{(\delta)})}((c/\rho)^{m-1})$ balls of radius $\rho/c$.*

*Proof.* Let $t \in I$. Under our hypotheses, there is a ball $B_r(t) \subseteq I^{(\delta)}$, with radius $r$ less than $\min\{\delta, c/M\}$ and of this order, such that $\|DF(t) - DF(t')\| < \frac{1}{100}c$ for $t' \in B_r(t)$ (here we use the upper bound on the second derivative of $F$). It is then an easy fact from calculus, essentially, the implicit function theorem, that the level set $S = F^{-1}(0) \cap B_r(t)$ is the graph of a 1-Lipschitz function and that in the transverse direction to $S$ the function $F$ grows at a rate proportional to $c$ as long as we remain in $B_r(t)$. Thus, given $\rho > 0$, the set $F^{-1}((-\rho,\rho)) \cap B_r(t)$ is contained in the $O(\rho/c)$-neighborhood of the graph of a 1-Lipschitz function with domain $B_r(t)$ for $r = O_M(c)$, and by the previous lemma, if $\rho < \min\{\delta, c/M\}$, it can be covered by $O_M((r^{m-1}/(\rho/c)^{m-1}))$ balls of diameter $\rho/c$. Also, $I$ can be covered by $O(\mathrm{vol}(I^{(\delta)})/r^m)$ balls $B_r(t)$ as above, so it can be covered by $O_M((c/\rho)^{m-1})$ balls of diameter $\rho/c$. □

**Corollary 6.27.** *For $F : I^{(\delta)} \to \mathbb{R}^d$ and under the same assumptions as above, the same conclusion holds.*

*Proof.* We can write $I = I_1 \cup \ldots \cup I_d$ such that on each of the closed sets $I_i$ the assumption of the previous lemma holds for $F_i$ (the $i$-th component of $F$) with some degradation of $c$. Then $I \cap F^{-1}((-\rho,\rho)^d) \subseteq \bigcup_{i=1}^d I_i \cap F_i^{-1}(-\rho,\rho)$ and the lemma can be applied to each set in the union to obtain the desired result. □

**Proposition 6.28.** *Let $I \subseteq \mathbb{R}^m$ be a compact set, $I^{(\delta)}$ the $\delta$-neighborhood of $I$, and $F : I^{(\delta)} \to \mathbb{R}^d$ a $(k+1)$-times differentiable function. Suppose that there are constants $M > 0$ and $0 < b < 1$ such that*

1. *For every $t \in I$, $0 \le p \le k+1$ and $v_1, \ldots, v_p \in \mathcal{B}_m$ we have $|D_{v_1} \ldots D_{v_p} F(t)| \le M$ (for $p = 0$ this means $|F(t)| \le M$).*
2. *For every $t \in I$ there exist $p \in \{0, \ldots, k\}$ and $v_1, \ldots, v_p \in \mathcal{B}_m$ such that $\left\| D_{v_1} \ldots D_{v_p} F(t) \right\| > b$ (for $p = 0$ this means $F(t) > b$).*

*Then there exists $C = C(b, M, \mathrm{vol}\, I^{(\delta)}) \ge 1$ such that for every $0 < \rho < b \cdot b^{2^k}$, the set*

$$Z_\rho = I \cap F^{-1}((-\rho,\rho)^d)$$

*can be covered by $C^k (b/\rho)^{(m-1)/2^k}$ balls of radius $(\rho/b)^{1/2^k}$.*

*Proof.* Take $C$ large enough to play the role of the constant in the bound in the previous corollary, and large enough that $mC^{k-1} + C \le C^k$ for $k \ge 1$.

We argue by induction on $k$. The case $k = 0$ is trivial (because $|F(t)| > b$ and $\rho < b$ implies $Z_\rho = \emptyset$).

Now fix $k$ and suppose we have proved the claim for $k-1$. First, note that we can assume without loss of generality that $I \subseteq \overline{Z_b} = \{t \in I : \|F(t)\| \le b\}$, since clearly $Z_\rho \subseteq \overline{Z_b}$ and if we did not have $I \subseteq \overline{Z_b}$ we could simply replace $I$ by $I \cap \overline{Z_b}$, to make it hold.

Since $\|F(t)\| \le b$ on $I$, the hypothesis (2) necessarily holds at each point with $p \ge 1$. Thus we can write $I$ as a union of closed sets $I_v$, $v \in \mathcal{B}_m$, on each of which the induction hypothesis holds for one of the functions $G_v = D_v F$.

Fix $v \in \mathcal{B}_m$, take $\rho' = \sqrt{b\rho}$. Note that $0 < b < 1$ and $0 < \rho < b^{2^k}$, so $0 < \rho' < b^{2^{k-1}}$. Define

$$\begin{aligned} I'_v &= I_v \cap G_v^{-1}((-\rho',\rho')^d) \\ I''_v &= I_v \setminus I'_v \end{aligned}$$

We cover $Z_\rho$ in each of these sets separately.

First, we actually cover the entire set $I'_v$. Indeed, by the induction hypothesis, it can be covered by

$$C^{k-1}(\frac{b}{\rho'})^{(m-1)/2^{k-1}} = C^{k-1}(\frac{b}{\rho})^{(m-1)/2^k}$$

balls of radius $(\rho'/b)^{1/2^{k-1}} = (\rho/b)^{1/2^k}$.

On the other hand, on $I''_v$ we have $\|DF\| \ge \|G_v\| \ge \rho'$. By the previous corollary, $Z_\rho \cap I''_v = I''_v \cap F^{-1}((-\rho,\rho)^d)$ can be covered by

$$C(\rho'/\rho)^{m-1} = C(\frac{b}{\rho})^{(m-1)/2}$$

balls of diameter $\rho/\rho' = \sqrt{\rho/b}$, hence, since $\sqrt{\rho/b} < (\rho/b)^{1/2^k}$, we can cover $Z_\rho \cap I''$ by at most this many balls of radius $(\rho/b)^{1/2^k}$.

Taking the union of the covers we have found for $Z_\rho \cap I'_v$ and $Z_\rho \cap I''_v$, we obtain a cover of $Z_\rho \cap I_v$ by $(C^{k-1}+C)(b/\rho)^{(m-1)/2}$ balls of radius $(\rho/b)^{1/2^k}$. Summing over the $m$ elements $v \in \mathcal{B}_m$, we have covered $Z_\rho$ by

$$m(C^{k-1}+C)(\frac{1}{\rho})^{(m-1)/2}) \leq C^k(\frac{b}{\rho})^{(m-1)/2^k}$$

balls of radius $(\rho/b)^{1/2^k}$ (using our assumption $mC^{k-1}+C \leq C^k$). This is the desired cover. □

Theorem 1.10 now follows from Proposition 6.24 and the next result:

**Theorem 6.29.** *If $\{\Phi_t\}_{t\in I}$ satisfies transversality of order $k \geq 1$ on the compact set $I \subseteq \mathbb{R}^m$, then the set $E$ of "exceptional" parameters in Theorem 1.9 has packing (and hence Hausdorff) dimension at most $m-1$.*

*Proof.* Let $M$ be a uniform bound for $\|D_{v_1}\dots D_{v_{k+1}}\Delta_{i,j}(t)\|$ taken over $v_i \in \mathcal{B}_m$, $t \in I$ and $i, j \in \Lambda^*$. Such $M$ exists from $k$-fold continuous differentiability of $r_i(\cdot), a_i(\cdot)$ and the fact that $|r_i|$ are bounded away from 1 on $I$. By transversality there is a constant $c > 0$ such that for all $n \in \mathbb{N}$ and all $i, j \in \Lambda^n$,

$$\forall\, t_0 \in I \quad \exists\, p \in \{0,\dots,k\} \quad \exists\, v_1,\dots,v_p \in \mathcal{B}_m$$

$$\text{such that} \qquad \|(D_{v_p}\dots D_{v_1}\Delta_{i,j})(t_0)\| > c\cdot|i\wedge j|^{-p}\cdot r_{min}^{|i\wedge j|}(t_0),$$

where

$$r_{min} = \min\{r_i(t)\,:\, i\in\Lambda\,,\, t\in I\}.$$

We may assume that $c < 1$ and $k \geq 2$. In what follows we suppress the dependence on $k, M, c$ and $I$ in the $O(\cdot)$ notation: $O(\cdot) = O_{k,M,c,|I|}(\cdot)$.

Fix $n$ and distinct $i, j \in \Lambda^n$. Let $b = b_n = cn^{-k}r_{min}^n$, so that the hypothesis of the previous proposition is satisfied for the function $F = \Delta_{i,j}$ and this $b$. Therefore, for all $0 < \rho < n^{2^k}$, the set $\{t \in I\,:\, |\Delta_{i,j}| < \rho\}$ can be covered by at most $O((b/\rho)^{(m-1)/2^k})$ balls of radius $(\rho/b)^{1/2^k}$ each.

Now let $\varepsilon > 0$ be such that $\rho = \varepsilon^n$ satisfies $\rho < (b_n)^{2^k} = (cn^{-k}r_{min}^n)^{2^k}$ for all $n$ (this holds for all sufficiently small $\varepsilon > 0$). Fixing $n$ again, the discussion above applies to $(\Delta_{i,j})^{-1}(-\varepsilon^n, \varepsilon^n)$ for every distinct pair $i, j \in \Lambda^n$, so ranging over all such pairs we find that

$$E'_{\varepsilon,n} = \bigcup_{i,j\in\Lambda^n\,,\,i\neq j} (\Delta_{i,j})^{-1}(-\varepsilon^n, \varepsilon^n)$$

can be covered by $O(|\Lambda|^n(b_n/\varepsilon^n)^{(m-1)/2^k})$ balls of radius $(\varepsilon^n/b_n)^{1/2^k}$. Now,

$$E \subseteq E'_\varepsilon = \bigcup_{N=1}^{\infty}\bigcap_{n>N} E'_{\varepsilon,n}.$$

By the above, for each $\varepsilon$ and $N$ we have

$$\begin{aligned}\text{bdim}\left(\bigcap_{n>N} E'_{\varepsilon,n}\right) &\leq \lim_{n\to\infty} \frac{\log\left(|\Lambda|^n (b_n/\varepsilon^n)^{(m-1)/2^k}\right)}{\log\left((b_n/\varepsilon^n)^{1/2^k}\right)} \\ &= O\left(\frac{\log(|\Lambda|(r_{min}/\varepsilon)^{(m-1)/2^k})}{\log(r_{min}/\varepsilon)^{1/2^k}}\right).\end{aligned}$$

The last expression tends to $m-1$ as $\varepsilon \to 0$, uniformly in $N$. Thus the same is true of $E'_\varepsilon$, and $E \subseteq E'_\varepsilon$ for all $\varepsilon$, so $E$ has packing (and Hausdorff) dimension $m-1$. □

## 6.7 Applications and further comments

*Proof of Theorem 1.12.* Fix $\Lambda$. For $i, j \in \Lambda^{\mathbb{N}}$, given an IFS $\Phi = \{(\varphi_i\}_{i\in\Lambda} = (r_i U_i + a_i)_{i\in\Lambda}$, evidently

$$\Delta_{i,j}(\Phi) = \sum_{n=0}^{\infty} \left(r_{i_1\ldots i_{n-1}} U_{i_1\ldots i_{n-1}} a_{i_n} - r_{j_1\ldots j_{n-1}} U_{j_1\ldots j_{n-1}} a_{j_n}\right).$$

As a function of $(r_u, U_u, a_u)_{u\in\Lambda} \in (\mathbb{R}^+ \times \mathbb{R}^{d^2} \times \mathbb{R}^d)^\Lambda$ this is clearly a non-constant expression. The parametrization is trivially real-analytic, and the conclusion follows from Theorem (1.10). □

*Proof of Theorem 1.13.* Fix $\{U_i\}_{i\in\Lambda} \in G_0^\Lambda$ and $\{r_i\}_{i\in\Lambda} \in (0, 1/2)^\Lambda$. Given distinct $i, j \in \Lambda^{\mathbb{N}}$ let $k = k(i,j)$ be the first index where they differ. For $a = (a_u)_{u\in\Lambda} \in (\mathbb{R}^d)^\Lambda$, let $\Phi_a = \{r_u U_u + a_u\}_{u\in\Lambda}$, so

$$\Delta_{i,j}(a) = \sum_{n\geq k(i,j)} \left(r_{i_1\ldots i_{n-1}} U_{i_1\ldots i_{n-1}} a_{i_n} - r_{j_1\ldots j_{n-1}} U_{j_1\ldots j_{n-1}} a_{j_n}\right).$$

This is linear in the $a$ variables. Differentiating by the coordinates in $a_{i_k} = (a^1_{i_k}, \ldots, a^d_{i_k})$, we obtain a derivative matrix of the form

$$\left(\frac{\partial \Delta_{i,j}}{\partial a_{i_k}}\right) = r_{i_1\ldots i_{k-1}} U_{i_1\ldots i_{k-1}} + \sum_{n\in I} r_{i_1\ldots i_n} U_{i_1\ldots i_n} - \sum_{n\in J} r_{j_1\ldots j_n} U_{j_1\ldots j_n}, \qquad (80)$$

where $I = \{n > k : i_n = i_k\}$ and $J = \{n > k : j_n = i_k\}$. Similarly, setting $I' = \{n > k : i_n = j_k\}$ and $J' = \{n > k : j_n = j_k\}$ and differentiating $\Delta_{i,j}$ by the $a_{j_k}$ variable (and using $r_{i_1\ldots i_{k-1}} = r_{j_1\ldots j_{k-1}}$ and $U_{i_1\ldots i_{k-1}} = U_{j_1\ldots j_{k-1}}$),

$$\left(\frac{\partial \Delta_{i,j}}{\partial a_{j_k}}\right) = r_{j_1\ldots j_{k-1}} U_{j_1\ldots j_{k-1}} + \sum_{n\in I'} r_{i_1\ldots i_n} U_{i_1\ldots i_n} - \sum_{n\in J'} r_{j_1\ldots j_n} U_{j_1\ldots j_n}. \qquad (81)$$

In order for these matrices to be invertible, it is enough that on the right hand sides of equations (80) and (81), the norm of the sum of the last two terms is less than the norm of the first term. Let

$$\begin{aligned} R &= \sum_{n\in I} r_{i_1\ldots i_n} + \sum_{n\in J} r_{j_1\ldots j_n} \\ R' &= \sum_{n\in I'} r_{i_1\ldots i_n} + \sum_{n\in J'} r_{j_1\ldots j_n}. \end{aligned}$$

These are upper bounds for the norms in question. We have

$$\begin{aligned}
R + R' &= \left(\sum_{n\in I} r_{i_1\ldots i_n} + \sum_{n\in I'} r_{i_1\ldots i_n}\right) + \left(\sum_{n\in J} r_{j_1\ldots j_n} + \sum_{n\in J'} r_{j_1\ldots j_n}\right) \\
&\le r_{i_1\ldots i_{k-1}} \prod_{n\in I\cap I'} (r_{i_n} + r_{i_n}) + r_{j_1\ldots j_{k-1}} \prod_{n\in J\cap J'} (r_{j_n} + r_{j_n}) \\
&< 2r_{i_1\ldots i_{k-1}}.
\end{aligned}$$

(In the first inequality we used $r_i < 1$. In the second we used the fact that if $n \in I \cap I'$ then $i_n \neq j_n$ and hence $r_{i_n} + r_{j_n} < 1$, and similarly for $n \in J \cap J'$, and that $r_{i_1\ldots i_{k-1}} = r_{j_1\ldots j_{k-1}}$ by choice of $k$). Now, $R + R' < r_{i_1\ldots r_{k-1}}$ implies that either $R < r_{i_1\ldots i_{k-1}}$ or $R' < r_{i_1\ldots i_{k-1}}$. In the first case, the first term in (80) is a similarity with contraction $r_{i_1\ldots i_{k-1}}$, and the latter two terms together give a matrix whose norm is at most $R < r_{i_1\ldots i_{k-1}}$. Hence the sum is invertible, and $\operatorname{rank} D\Delta_{i,j} \ge d$. The same argument applies to (81) if $R' < r_{i_1\ldots i_{k-1}}$. The conclusion now follows from Theorem 1.11. □

*Proof of Theorem 1.14.* Let $(\varphi_i)_{i\in\Lambda}$ be given. For $i \in \Lambda^{\mathbb{N}}$ write $\varphi_i = \lim_{n\to\infty} \varphi_{i_1\ldots i_n}(0)$. Given distinct $i, j \in \Lambda^{\mathbb{N}}$ and $\pi \in \Pi_{d,k}$, evidently

$$\Delta_{i,j}(\pi) = \pi(\varphi_i) - \pi(\varphi_j) = \pi(\varphi_i - \varphi_j)$$

Now, it is easy to verify that for a fixed $0 \neq v \in \mathbb{R}^d$ the map $\pi \mapsto \pi(v)$, $\Pi_{d,k} \to \mathbb{R}^k$, has rank $k$ at every point. Taking $v = \varphi_i - \varphi_j$ this shows that $\Delta_{i,j}$ has rank $k$ at every point. An application of Theorem (1.11) completes the proof. □

*Proof of Theorem 1.15.* Writing $\Delta_{i,j}(\beta, \gamma)$ explicitly and noting that it is not constant and real-analytic, Theorem 1.15 is immediate from Theorem 1.10 (since the IFS in on the line, irreducibility is a non-issue). □

*Proof of Theorem 1.16.* We would again like to apply Theorem 1.16. Analyticity and non-triviality of $\Delta_{i,j}$ is again a simple matter, but the usual presentation of the fat Sierpinski gaskets uses an IFS consisting of homotheties, which act reducibly. However, the attractor of the fat Sierpinski gaskets are invariant under rotation by $2\pi/3$ about their center of mass, and hence they can be presented also as attractors of an IFS $x \mapsto \lambda U_i x + a_i$ where $a_i$ are the vertices of a triangle in $\mathbb{R}^2$ and the $U_i$ are rotations by $2\pi/3$. Unlike the usual presentation this IFS is irreducible. Theorem 1.16 now does the job. □

The argument in the last proof relied heavily on the possibility of presenting the attractor using an irreducible IFS. This is not always possible. For instance, if we take the fat Sierpinski gasket with the usual homothetic presentation, and augment it with an additional homothety, then the symmetry breaks down and there is no irreducible presentation. In this case Theorem 1.16 no longer gives information about the set of exceptional parameters, because the set of reducible parameters is large. Some additional argument is needed in this case.

Finally, the proof of Corollary 1.7 is based on the classical fact that polynomials of bounded height in a fixed set of algebraic numbers either vanishes or is exponentially large in the degree of the polynomial. For completeness we include a proof, noting that the version in [12, Lemma 5.10] erroneously omitted the height assumption:

**Lemma 6.30.** *Let $\mathcal{A} \subseteq \mathbb{R}$ be a finite set of algebraic numbers over $\mathbb{Q}$. If $x$ is a polynomial expression in the elements of $\mathcal{A}$ with coefficients of magnitude at most $h$, then either $x = 0$ or $|x| > s^n$.*

*Proof.* Let $\mathcal{A} = \{a_1, \dots, a_k\}$. Let $f(x_1, .., x_k)$ be an integer polynomial of degree $n$ and coefficients bounded by $h$ in absolute value. Assuming $x = f(a_1, ..., a_k)$ is not zero, it suffices to show that $|x| > c^n/h^u$ for some $c, u > 0$ depending only on $\mathcal{A}$.

Let $\mathbb{F} = \mathbb{Q}(a_1, \dots, a_k)$ be the field over $\mathbb{Q}$ generated by $\{a_i\}$.

We may assume that $a_i$ are algebraic integers. This is because we can choose positive integers $p_1, ..., p_k$ such that $b_i = p_i \cdot a_i$ is an algebraic integer. Let $p = p_1 \cdot \ldots \cdot p_k$ (note that this depends only on the $a_i$). Then

$$p^n \cdot f(a_1, ..., a_k) = g(b_1, ..., b_k),$$

and $g$ is an integer polynomial of degree $n$ with coefficients bounded by $h \cdot p^n$. So if we have $c = c(b_1, ..., b_k) > 0$ such that $g(b_1, ..., b_k) > c^n/(hp^n)^u$, then $f(a_1, ..., a_k) > c^n/(h^u \cdot p^{(u+1)n})$, which is what we wanted (using the constant $c/p^{u+1}$ instead of $c$).

Assuming now that $a_i$ are algebraic integers, let $\mathbb{F}'$ be the normal closure of $\mathbb{F} = \mathbb{Q}(a_1, ..., a_k)$ and $\Gamma = \mathrm{Gal}(\mathbb{F}'/\mathbb{Q})$, so the fixed field of $\Gamma$ is $\mathbb{Q}$. Note that $\mathbb{F}'$, hence $\Gamma$, depends only on the $a_i$, and $\Gamma$ is finite.

Now we do the usual thing: if $f(x_1, ..., a_k)$ is not zero then also $\prod_{s \in \Gamma} s(f(x))$ is non-zero, but it is both an algebraic integer and rational, so its absolute value is at least 1. Hence

$$1 \leq \prod_{s \in \Gamma} |f(sx)| = |f(x)| \cdot \prod_{s \in \Gamma \setminus \{\mathrm{id}\}} |f(sx)|.$$

The last product has $|\Gamma| - 1$ factors $|f(sx)|$, each of size at most $h \cdot \max\{|\Gamma\text{-conjugates of } a_i|\}^n$. Dividing gives the bound that we want. □

# References


[1] J. Bourgain. On the Erdős-Volkmann and Katz-Tao ring conjectures. *Geom. Funct. Anal.*, 13(2):334–365, 2003.

[2] J. Bourgain. The discretized sum-product and projection theorems. *J. Anal. Math.*, 112:193–236, 2010.

[3] J. Bourgain and A. Gamburd. On the spectral gap for finitely-generated subgroups of SU(2). *Invent. Math.*, 171(1):83–121, 2008.

[4] J. Bourgain, N. Katz, and T. Tao. A sum-product estimate in finite fields, and applications. *Geom. Funct. Anal.*, 14(1):27–57, 2004.

[5] Dave Broomhead, James Montaldi, and Nikita Sidorov. Golden gaskets: variations on the Sierpiński sieve. *Nonlinearity*, 17(4):1455–1480, 2004.

[6] Nicholas de Saxce. A product theorem in simple lie groups. *preprint*, 2014. http://arxiv.org/abs/1405.2003.

[7] Kenneth Falconer and Xiong Jin. Exact dimensionality and projections of random self-similar measures and sets. *preprint*, 2014. http://arxiv.org/abs/1212.1345.

[8] De-Jun Feng and Huyi Hu. Dimension theory of iterated function systems. *Comm. Pure Appl. Math.*, 62(11):1435–1500, 2009.

[9] Harry Furstenberg. Intersections of Cantor sets and transversality of semigroups. In *Problems in analysis (Sympos. Salomon Bochner, Princeton Univ., Princeton, N.J., 1969)*, pages 41–59. Princeton Univ. Press, Princeton, N.J., 1970.

[10] Hillel Furstenberg. Ergodic fractal measures and dimension conservation. *Ergodic Theory Dynam. Systems*, 28(2):405–422, 2008.

[11] Adriano M. Garsia. Arithmetic properties of Bernoulli convolutions. *Trans. Amer. Math. Soc.*, 102:409–432, 1962.

[12] Michael Hochman. On self-similar sets with overlaps and inverse theorems for entropy. *Ann. of Math. (2)*, 180(2):773–822, 2014.

[13] Michael Hochman and Pablo Shmerkin. Local entropy averages and projections of fractal measures. *Ann. of Math. (2)*, 175(3):1001–1059, 2012.

[14] Michael Hochman and Boris Solomyak. On the dimension of the furstenberg measure for $SL_2(\mathbb{R})$-random matrix products. *Inventiones Mathematicae*, 2016. to appear.

[15] Thomas Jordan. Dimension of fat Sierpiński gaskets. *Real Anal. Exchange*, 31(1):97–110, 2005/06.

[16] Thomas Jordan and Mark Pollicott. Properties of measures supported on fat Sierpinski carpets. *Ergodic Theory Dynam. Systems*, 26(3):739–754, 2006.

[17] Antti Käenmäki, Tapio Rajala, and Ville Suomala. Existence of doubling measures via generalised nested cubes. *Proc. Amer. Math. Soc.*, 140(9):3275–3281, 2012.

[18] V. A. Kaĭmanovich and A. M. Vershik. Random walks on discrete groups: boundary and entropy. *Ann. Probab.*, 11(3):457–490, 1983.

[19] Nets Hawk Katz and Terence Tao. Some connections between Falconer's distance set conjecture and sets of Furstenburg type. *New York J. Math.*, 7:149–187 (electronic), 2001.

[20] Richard Kenyon. Projecting the one-dimensional Sierpinski gasket. *Israel J. Math.*, 97:221–238, 1997.

[21] John M. Lee. *Introduction to smooth manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2003.

[22] Elon Lindenstrauss and Péter Varjú. Random walks in the group of euclidean isometries and self-similar measures. *preprint*, 2014. http://arxiv.org/find/all/1/all:+AND+varju+lindenstrauss/0/1/0/all/0/1.

[23] S. Łojasiewicz. Une propriété topologique des sous-ensembles analytiques réels. In *Les Équations aux Dérivées Partielles (Paris, 1962)*, pages 87–89. Éditions du Centre National de la Recherche Scientifique, Paris, 1963.

[24] M. Madiman. On the entropy of sums. In *Information Theory Workshop, 2008. ITW '08. IEEE*, pages 303–307, 2008.

[25] Mokshay Madiman, Adam W. Marcus, and Prasad Tetali. Entropy and set cardinality inequalities for partition-determined functions. *Random Structures Algorithms*, 40(4):399–424, 2012.

[26] Pertti Mattila. *Geometry of sets and measures in Euclidean spaces*, volume 44 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1995. Fractals and rectifiability.

[27] J. Neunhäuserer. Properties of some overlapping self-similar and some self-affine measures. *Acta Math. Hungar.*, 92(1-2):143–161, 2001.

[28] V. I. Rotar′. The rate of convergence in the multidimensional central limit theorem. *Teor. Verojatnost. i Primenen.*, 15:370–372, 1970.

[29] Károly Simon and Boris Solomyak. On the dimension of self-similar sets. *Fractals*, 10(1):59–65, 2002.

[30] Karoly Simon and Boris Solomyak. *Self-similar and self-affine sets and measures*. 2014. preliminary manuscript.

[31] Terence Tao. Sumset and inverse sumset theory for Shannon entropy. *Combin. Probab. Comput.*, 19(4):603–639, 2010.

[32] Terence Tao and Van Vu. *Additive combinatorics*, volume 105 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2006.

Email: mhochman@math.huji.ac.il
Address: Einstein Institute of Mathematics, Givat Ram, Jerusalem 91904, Israel