ON ABSOLUTE CONTINUITY OF INHOMOGENEOUS AND CONTRACTING ON AVERAGE SELF-SIMILAR MEASURES

SAMUEL KITTLE AND CONSTANTIN KOGLER

ABSTRACT. We give a condition for absolute continuity of self-similar measures in arbitrary dimensions. This allows us to construct the first explicit absolutely continuous examples of inhomogeneous self-similar measures in dimension one and two. In fact, for $d \geq 1$ and any given rotations in $O(d)$ acting irreducibly on $\mathbb{R}^d$ as well as any distinct translations, all having algebraic coefficients, we construct absolutely continuous self-similar measures with the given rotations and translations. We furthermore strengthen Varjú's result for Bernoulli convolutions, treat complex Bernoulli convolutions and in dimension ≥ 3 improve the condition on absolute continuity by Lindenstrauss-Varjú. Moreover, self-similar measures of contracting on average measures are studied, which may include expanding similarities in their support.

CONTENTS

1. INTRODUCTION

In the study of self-similar measures it is fundamental to determine their dimension and to find conditions for absolute continuity. For the former problem progress was made by Hochman ([Hoc14], [Hoc17]), relating the dimension of a self-similar measure to the entropy and Lyapunov exponent provided the generating measure satisfies a mild separation condition. While it was shown by Saglietti-Shmerkin-Solomyak [SSS18], building on methods pioneered by Solomyak [Sol95], that, under suitable assumptions, generic one-dimensional self-similar measures are absolutely continuous, finding explicit examples remains challenging. It was shown by Varjú [Var19] that Bernoulli convolution are absolutely continuous if their defining parameter is sufficiently close to 1 in terms of the Mahler measure. In dimension $d \geq 3$, assuming that the rotation part of the self-similar measure is fixed and has an L^2 spectral gap on $O(d)$, Lindenstrauss-Varjú [LV16] showed absolute continuity if all of the contraction rates are sufficiently close to 1. In this paper we strengthen and vastly generalise these two results. Moreover, we give the first explicit examples of absolutely continuous self-similar measures in dimension one and two with non-uniform contraction rates. For instance consider for $x \in \mathbb{R}$ the similarities

$$g_1(x) = \frac{q}{q+1}x + 1 \quad \text{and} \quad g_2(x) = \frac{q}{q+2}x + 2.$$

We then show that the self-similar measure of $\frac{1}{2}\delta_{g_1} + \frac{1}{2}\delta_{g_2}$ is absolutely continuous on $\mathbb{R}$ for any sufficiently large prime q. Furthermore, our methods allow to construct several classes of explicit absolutely continuous examples for $g_i(x) = \rho_i U_i x + b_i$ for $x \in \mathbb{R}^d$ in any dimension $d \geq 1$ as well as for every collection of orthogonal matrices U_i acting irreducibly on $\mathbb{R}^d$ and distinct vectors $b_i \in \mathbb{R}^d$, provided they all have algebraic entries.

Let $G = \mathrm{Sim}(\mathbb{R}^d)$ be the group of similarities on $\mathbb{R}^d$ and let $O(d)$ be the group of orthogonal $d \times d$ matrices. For each $g \in G$ there exists a scalar $\rho(g) > 0$, an orthogonal matrix $U(g) \in O(d)$ and a vector $b(g) \in \mathbb{R}^d$ such that $g(x) = \rho(g)U(g)x + b(g)$ for all $x \in \mathbb{R}^d$. A similarity is called contracting if $\rho(g) < 1$ and expanding when $\rho(g) > 1$.

Given a probability measure μ supported on finitely many contracting similarities of $\mathbb{R}^d$, there exists by Hutchinson's theorem [Hut81] a unique compactly supported μ-stationary probability measure ν on $\mathbb{R}^d$ called the self-similar measure of μ, i.e. a measure satisfying $\mu * \nu = \nu$ for $\mu * \nu$ the convolution of μ and ν as defined in (2.18). In this paper we study the larger class of self-similar measures that might also be supported on expanding similarities and are only contacting on average. The Lyapunov exponent of a probability measure μ on G is defined, whenever it exists, as

$$\chi_\mu = \mathbb{E}_{g\sim\mu}[\log \rho(g)].$$

Definition 1.1. *If $\chi_\mu < 0$, we call μ* ***contracting on average****. Moreover, if every $g \in \mathrm{supp}(\mu)$ is contracting, we say that μ is* ***contracting****. When $\chi_\mu < 0$ and there is $g \in \mathrm{supp}(\mu)$ such that $\rho(g) > 1$, then we call μ* ***only contracting on average****.*

As follows from [BE88], Hutchinson's theorem generalises to contracting on average measures. If μ is only contracting on average, the resulting self-similar measure is usually not compactly supported, yet we show that their mass is concentrated around the identity. Denote by $| \circ |$ the euclidean norm on $\mathbb{R}^d$. For the asymptotic notation used we refer to section 2.

Theorem 1.2. *(Generalisation of Hutchinson's theorem) Let μ be a finitely supported and contracting on average probability measure on $G = \mathrm{Sim}(\mathbb{R}^d)$. Then there exists a unique probability measure ν on $\mathbb{R}^d$ such that $\mu * \nu = \nu$. Moreover, there exists $\alpha = \alpha(\mu) > 0$ such that for $R > 0$,*

$$\nu(\{x \in \mathbb{R}^d \,:\, |x| \geq R\}) \ll_\mu R^{-\alpha}. \tag{1.1}$$

While fine estimates for $\nu(\{x \in \mathbb{R}^d \,:\, |x| \geq R\})$ have been established by various authors under a range of assumptions (cf. [Kes73], [Gol91], [GP15],[GP16],[Kev16], [Klo22]), the coarse bound (1.1) does not appear in the literature to the authors knowledge. We deduce (1.1) in section 4 from the large deviation principle.

When μ is only contracting on average and not supported on a set of similarities with a common fixed point, then we show in Lemma 4.3 that $\nu(\{x \in \mathbb{R}^d \,:\, |x| \geq$

$R\}) \gg_\mu R^{-\alpha_2}$ for some constant $\alpha_2 = \alpha_2(\mu) > 0$. Moreover, we note that when ν is absolutely continuous, it can be shown using (1.1) that ν has finite differential entropy (see the proof of Lemma 7.3). In similar vein, (1.1) will be used in section 7 for the proof of our main result, Theorem 2.4, to bound the entropy of a suitable stopped random walk.

Throughout this paper we denote by ν the self-similar measure associated to μ. If μ is (only) contracting on average, we say that ν is a (only) contracting on average self-similar measure. Moreover, μ or respectively ν is called homogeneous if there are $r \in \mathbb{R}_{>0}$ and $U \in O(d)$ such that $r = \rho(g)$ and $U = U(g)$ for all $g \in \mathrm{supp}(\mu)$. When this is not the case, we say that μ and ν are inhomogeneous. A particular goal of this paper is to give explicit examples of inhomogeneous as well as only contracting on average self-similar measures which are absolutely continuous.

To state our main result, we first discuss the Hausdorff dimension of ν. Recall that the Hausdorff dimension of ν is defined as

$$\dim \nu = \inf\{\dim E \,:\, E \subset \mathbb{R}^d \text{ measurable and } \nu(E) > 0\},$$

where $\dim E$ is the Hausdorff dimension of E. In order to state the landmark results by Hochman [Hoc14], [Hoc17], recall that the random walk entropy of a finitely supported measure μ is defined as

$$h_\mu = \lim_{n\to\infty} \frac{1}{n} H(\mu^{*n}) = \inf_{n\geq 1} \frac{1}{n} H(\mu^{*n}),$$

where $H(\cdot)$ is the Shannon entropy. Observe that if $\mathrm{supp}(\mu)$ has no exact overlaps, meaning that $\mathrm{supp}(\mu)$ generates a free semigroup, then $h_\mu = H(\mu) = -\sum_i p_i \log p_i$.

Moreover, as in [Hoc17], denote by $d(\cdot,\cdot)$ the metric on G defined for $g = \rho_1 U_1 + b_1$ and $h = \rho_2 U_2 + b_2$ as

$$d(g,h) = |\log \rho_1 - \log \rho_2| + ||U_1 - U_2|| + |b_1 - b_2|$$

for $||\cdot||$ the operator norm.

To distinguish between the results for dimension and absolute continuity, denote

$$\Delta_n = -\frac{1}{n} \log \min\{d(g,h) \text{ for } g,h \in \mathrm{supp}(\mu^{*n}) \text{ with } g \neq h\}$$

and

$$S_n = -\frac{1}{n} \log \min \left\{ d(g,h) \text{ for } g,h \in \bigcup_{i=1}^{n} \mathrm{supp}(\mu^{*i}) \text{ with } g \neq h \right\}.$$

The splitting rate of μ is then defined as

$$S_\mu = \limsup_{n\to\infty} S_n.$$

If $S_\mu < \infty$ we say that μ has exponential separation. For our purposes, it is necessary to work with S_μ, whereas Hochman's dimension result just require

$$\liminf_{n\to\infty} \Delta_n < \infty.$$

We call a subgroup H of $O(d)$ irreducible if H acts irreducibly on $\mathbb{R}^d$, i.e. the only H-invariant subspaces of $\mathbb{R}^d$ are $\{0\}$ and $\mathbb{R}^d$. Moreover we say that a measure $\mu = \sum_{i=1}^n p_i \delta_{g_i}$ on G or $O(d) \subset G$ irreducible if the group generated by $\{U(g_1), \ldots, U(g_n)\}$ is irreducible. When the elements in the support of μ have a common fixed point $x \in \mathbb{R}^d$, then δ_x is the self-similar measure of μ. To avoid the latter case, we say that μ has no common fixed point if the similarities in $\mathrm{supp}(\mu)$ do not.

It was shown by Hochman [Hoc17], generalizing [Hoc14], that if μ is a finitely supported, contracting and irreducible probability measure on G without a common fixed point such that $\liminf_{n\to\infty} \Delta_n < \infty$, then $\dim \nu = \min\{d, \frac{h_\mu}{|\chi_\mu|}\}$. In the accompaniment paper [KK24] we extend Hochman's result to include contracting on average measures.

Theorem 1.3. *(Generalisation of Hochman's theorem,* [KK24]*) Let μ be a finitely supported, contracting on average and irreducible probability measure on G without a common fixed point and satisfying* $\liminf_{n\to\infty} \Delta_n < \infty$*. Then*

$$\dim \nu = \min\left\{d, \frac{h_\mu}{|\chi_\mu|}\right\}.$$

Therefore ν can only be absolutely continuous if $h_\mu \geq d|\chi_\mu|$. Moreover the following general conjecture is expected to hold.

Conjecture 1.4. *Let μ be a finitely supported, contracting on average and irreducible probability measure on G without a common fixed point. Then ν is absolutely continuous if*

$$\frac{h_\mu}{|\chi_\mu|} > d.$$

Our main result establishes a weakening of the latter conjecture. Indeed, when the $O(d)$-part of our measure μ is fixed, we show conjecture 1.4 with the d being replaced by a constant depending on the $O(d)$-part as well as the logarithmic separation rate $\log S_\mu$. Given a measure μ on G we denote by $U(\mu)$ the pushforward of μ under the map $g \mapsto U(g)$. We first state a version of our main theorem for contracting measures.

Theorem 1.5. *Let $d \geq 1$ and $\varepsilon \in (0,1)$. Given an irreducible probability measure μ_U on $O(d)$ there exists a constant $C \geq 1$ and $\tilde{\rho} \in (0,1)$ depending on d, ε and μ_U such that the following holds. Let $\mu = \sum_{i=1}^k p_i \delta_{g_i}$ be a contracting and exponentially separated probability measure on G without a common fixed point satisfying $U(\mu) = \mu_U$ and $p_i \geq \varepsilon$ as well as $\rho(g_i) \in (\tilde{\rho}, 1)$ for all $1 \leq i \leq k$. Then the self-similar measure ν is absolutely continuous if*

$$\frac{h_\mu}{|\chi_\mu|} > C\left(\max\left\{1, \log \frac{S_\mu}{h_\mu}\right\}\right)^2.$$

Theorem 1.5 is a special case of the more general Theorem 2.4, which requires a few new definitions we state in section 2.1. When $d = 1$ we note that every probability measure on $O(1)$ is irreducible. We further observe that while Theorem 1.5 applies in the case when the spectral gap of μ_U is zero, the dependence of C and $\tilde{\rho}$ can be made more explicit in the presence of a spectral gap. To introduce notation, given a closed subgroup $H \subset G$ and assuming that μ_U is a probability measure on $O(d)$ with $\mathrm{supp}(\mu_U) \subset H$, we denote by $\mathrm{gap}_H(\mu_U)$ the L^2-spectral gap of μ_U in H as defined in (2.19).

Theorem 1.6. *Let d, ε, μ_U and μ be as in Theorem 1.5. Assume further that $\mathrm{gap}_H(\mu_U) \geq \varepsilon > 0$ for H the closure of the subgroup generated by the support of μ_U. Then there exists $C \geq 1$ and $\tilde{\rho} \in (0,1)$ only depending on d and ε such that the conclusion of Theorem 1.5 holds.*

We point out that in Theorem 1.6 the constants are independent of the subgroup H and the statement applies when H is a finite irreducible subgroup of $O(d)$ as well as when H is a positive dimensional irreducible Lie subgroup of $O(d)$. As is shown in section 9, this observation relies on uniform convergence of μ_U^{*n} towards the Haar probability measure m_H and on Schur's lemma implying that $\mathbb{E}_{h\sim m_H}[|x\cdot hy|^2] = d^{-1}$ for any unit vectors $x, y \in \mathbb{R}^d$ and any irreducible subgroup $H \subset O(d)$.

To construct explicit examples of absolutely continuous self-similar measures on $\mathbb{R}^d$, Theorem 1.5 requires us to estimate $h_\mu, |\chi_\mu|$ and S_μ. It is straightforward to deal with $|\chi_\mu|$ as it can be explicitly computed. Lower bounds on the random walk entropy follow in many cases (see section 10.1) by the ping-pong lemma or Breuillard's strong Tits alternative [Bre08]. It also holds that $h_{U(\mu)} \leq h_\mu$, so when $h_{U(\mu)} > 0$, we only need to control $|\chi_\mu|$ and S_μ. With current methods we can usually only bound S_μ if all of the coefficients of the elements in the support of μ are algebraic. In the latter case, as shown in section 10.2, when all of the coefficients of elements in the support of μ lie in a number field K and have logarithmic height at most L, then $S_\mu \ll_d L \cdot [K : \mathbb{Q}]$. We observe that $\log S_\mu$ is usually very small as it is double logarithmic in the arithmetic complexity of the coefficients. All this information makes it straightforward to find explicit examples of absolutely continuous self-similar measures. The constants C and $\tilde{\rho}$ in Theorem 1.5 can be computed from the involved terms, yet we do not make the dependence explicit in this work.

The proof of Theorem 1.5 and Theorem 2.4 builds on new techniques initiated by the first-named author in [Kit23] and further developed in this paper, while being inspired by ideas from [Hoc14], [Hoc17], [Var19] and [Kit21]. We give an outline of our proof in section 2.2 and note that the main novelties exploited are strong product bounds for detail at scale r (a notion introduced in [Kit21]) and a decomposition theory for stopped random walks to capture the amount of variance we can gain at a given scale. [Kit23] is concerned with constructing absolutely continuous Furstenberg measures of $\mathrm{SL}_2(\mathbb{R})$ on 1-dimensional projective space $\mathbb{P}^1(\mathbb{R}) = \mathbb{R}^2/\sim$ and an analogue of Theorem 2.4 is shown. However, we currently can't deduce a result similar to Theorem 1.5 for Furstenberg measures as the dynamics of the $\mathrm{SL}_2(\mathbb{R})$ action on $\mathbb{P}^1(\mathbb{R})$ are more difficult to control than the one of the $\mathrm{Sim}(\mathbb{R}^d)$ action on $\mathbb{R}^d$. Indeed, we exploit that one can rescale and translate self-similar measures without changing the Lyapunov exponent, the separation rate, the random walk entropy or the spectral gap of the generating measure.

To also treat contracting on average measures, we state the following version of Theorem 1.5. Our current methods require some control on the scaling rate of the expanding similarities.

Theorem 1.7. *Let d and μ_U be as in Theorem 1.5 and let $R > 1$ and $\varepsilon > 0$. Let $\mu = \sum_{i=1}^k p_i \delta_{g_i}$ be a contracting on average and exponentially separated probability measure on G without a common fixed point satisfying $U(\mu) = \mu_U$ and $p_i \geq \varepsilon$ as well as $\rho(g_i) \in [R^{-1}, R]$ for all $1 \leq i \leq k$. Then there is some $\tilde{\rho} \in (0,1)$ and $C > 1$ depending on d, R, ε and μ_U such that the conclusion of Theorem 1.5 holds provided that for some $\hat{\rho} \in (\tilde{\rho}, 1)$ we have*

$$\frac{\mathbb{E}_{\gamma\sim\mu}[|\hat{\rho} - \rho(\gamma)|]}{1 - \mathbb{E}_{\gamma\sim\mu}[\rho(\gamma)]} < 1 - \varepsilon.$$

In the presence of a spectral gap, the analogue of Theorem 1.6 also holds for Theorem 1.7. Using Theorem 1.5, Theorem 1.7 and Theorem 2.4 one can construct a versatile collection of explicit absolutely continuous self-similar measures. We give a few cases below and encourage the reader to find further examples. Indeed, as shown in Corollary 1.10 and Corollary 1.11, for any given irreducible probability measure μ_U on $O(d)$ supported on matrices with algebraic entries and algebraic vectors $b_1, \dots, b_k$ with $b_1 \neq b_2$, we can find explicit contracting as well as only contracting on average measures $\mu = \sum_{i=1}^k p_i \delta_{g_i}$ on G with $U(\mu) = \mu_U$ and $b(g_i) = b_i$ for $1 \leq i \leq k$ and having absolutely continuous self-similar measure.

Real and Complex Bernoulli Convolutions. While Theorem 1.5 applies to arbitrary self-similar measures, it gives new results for Bernoulli convolutions. Let $\lambda \in (1/2, 1)$ and denote by ν_λ the unbiased Bernoulli convolution of parameter λ, i.e. the law of the random variable $\sum_{n=0}^\infty \xi_n \lambda^n$ with $\xi_0, \xi_1, \dots$ independent Bernoulli random variables with $\mathbb{P}[\xi_i = 1] = \mathbb{P}[\xi_i = -1] = 1/2$. It was shown by Solomyak [Sol95] that for almost all $\lambda \in (1/2, 1)$ the Bernoulli convolution ν_λ has a density in $L^2(\mathbb{R})$, while Erdős [Erd39] proved that ν_λ is singular whenever λ^{-1} is a Pisot number.

The Mahler measure of an algebraic number λ is defined as

$$M_\lambda = |a| \prod_{|z_j|>1} |z_j|$$

with $a(x - z_1) \cdots (x - z_\ell)$ the minimal polynomial of λ over $\mathbb{Z}$. We note that as in Corollary 5.9 of [Kit23] it holds that

$$S_{\nu_\lambda} \leq \log M_\lambda. \tag{1.2}$$

Garsia [Gar62, Theorem 1.8] showed that ν_λ is absolutely continuous for algebraic λ with $M_\lambda = 2$, while the first-named author [Kit21] established that ν_λ is absolutely continuous if $M_\lambda \approx 2$. In landmark work, Varjú [Var19] proved for every $\varepsilon > 0$ there is a constant $C > 1$ such that that ν_λ is absolutely continuous if

$$\lambda > 1 - C^{-1} \min\{\log M_\lambda, (\log M_\lambda)^{-1-\varepsilon}\}. \tag{1.3}$$

When applying Theorem 1.5 to Bernoulli convolutions we deduce the following strengthening of (1.3), exploiting the comparison between the entropy and the Mahler measure for Bernoulli convolution due to [BV20].

Corollary 1.8. *There is an absolute constant $C > 1$ such that the following holds. Let $\lambda \in (1/2, 1)$ be a real algebraic number. Then the Bernoulli convolution ν_λ is absolutely continuous on $\mathbb{R}$ if*

$$\lambda > 1 - C^{-1} \min\{\log M_\lambda, (\log\log M_\lambda)^{-2}\}. \tag{1.4}$$

We estimate that a direct application of our method would lead to $C \approx 10^{10}$ in Corollary 1.8. It would be an interesting further direction to try to optimise C for Bernoulli convolutions and in particular for the case $\lambda = 1 - \frac{1}{n}$.

Our most general result, Theorem 2.4, also applies to complex Bernoulli convolutions, which are defined analogously for $\lambda \in \mathbb{D} = \{\lambda \in \mathbb{C} : |\lambda| < 1\}$. When $|\lambda| \in (0, 2^{-1/2})$, then $\dim \nu_\lambda \leq \frac{\log 2}{|\log \lambda|} < 2$ and ν_λ is singular to the Lebesgue measure on $\mathbb{C}$. It was shown by Shmerkin-Solomyak [SS16a] that the set of $\lambda \in \mathbb{C}$ with $|\lambda| \in (2^{-1/2}, 1)$ and ν_λ is singular has Hausdorff dimension zero, whereas Solomyak-Xu [SX03] showed that ν_λ is absolutely continuous on $\mathbb{C}$ for a non-real algebraic $\lambda \in \mathbb{D}$ with $M_\lambda = 2$ and [Kit21] applies as well. We extend Corollary 1.8 to complex parameters while assuming (1.5) in order to ensure that the rotation part of λ mixes fast enough and so that our measure is sufficiently non-degenerate (see section 2.1).

Corollary 1.9. *For every $\varepsilon > 0$ there is a constant $C \geq 1$ such that the following holds. Let $\lambda \in \mathbb{C}$ be a complex algebraic number such that $|\lambda| \in (2^{-1/2}, 1)$ and*

$$|\mathrm{Im}(\lambda)| \geq \varepsilon. \tag{1.5}$$

Then the Bernoulli convolution ν_λ is absolutely continuous on $\mathbb{C}$ if

$$|\lambda| > 1 - C^{-1} \min\{\log M_\lambda, (\log\log M_\lambda)^{-2}\}.$$

Self-similar measures on $\mathbb{R}^d$. With Theorem 1.5 and Theorem 1.7 numerous explicit examples of absolutely continuous self-similar measures in $\mathbb{R}^d$ can be constructed. In order to apply these results we need to estimate h_μ. In the following examples we have used the ping-pong lemma (see section 10) in two ways in order to establish lower bounds on h_μ. For the first class of examples we have applied p-adic ping-pong as in Lemma 10.4.

Corollary 1.10. *Let $d \geq 1$ and $\varepsilon > 0$, let $\mu_U = \sum_{i=1}^k p_i \delta_{U_i}$ be an irreducible probability measure on $O(d)$ with $p_i \geq \varepsilon$ and let $b_1, \dots, b_k \in \mathbb{R}^d$ with $b_1 \neq b_2$. Assume that $U_1, \dots, U_k$ and $b_1, \dots, b_k$ have algebraic coefficients. Let q be a prime number and for $1 \leq i \leq k$ consider*

$$g_i(x) = \frac{q}{q + a_{i,q}} U_i x + b_i \qquad \text{for any integer} \qquad a_{i,q} \in [1, q^{1-\varepsilon}].$$

Assume that $g_1, \dots, g_k$ do not have a common fixed point and consider $\mu = \sum_{i=1}^k p_i \delta_{g_i}$. Then the self-similar measure of μ is absolutely continuous for q sufficiently large depending on $d, \varepsilon, U_1, \dots, U_k$ and $b_1, \dots, b_k$.

We point out that any choice of integers $a_{i,q}$ works and that the necessary size of q to derive absolute continuity does not depend on this choice, leading to a vast number of examples. Moreover, we can adapt Corollary 1.10 to give only contracting on average examples. In order to satisfy the assumption from Theorem 1.7, we require that $\mu = \sum_{i=1}^k p_i \delta_{g_i}$ satisfies that $p_k \leq \frac{1}{4}$. This nonetheless leads to absolutely continuous examples with $U(\mu) = \mu_U$ for any given irreducible probability measure $\mu_U = \sum_{i=1}^k p_i \delta_{U_i}$ on $O(d)$ as we do not require that the U_i are distinct.

Corollary 1.11. *Let d, ε and $\mu_U = \sum_{i=1}^k p_i \delta_{U_i}$ as well as $b_1, \dots, b_k$ be as in Proposition 1.10. Let q be a prime number and consider for $1 \leq i \leq k-1$*

$$g_i(x) = \frac{q}{q+3} U_i x + b_i \qquad \text{and} \qquad g_k(x) = \frac{q}{q-1} U_k x + b_k.$$

Assume that $g_1, \dots, g_k$ do not have a common fixed point and further that

$$p_k \leq \frac{1}{3}.$$

Then the self-similar measure of $\mu = \sum_{i=1}^k p_i \delta_{g_i}$ is absolutely continuous for q sufficiently large depending on $d, \varepsilon, U_1, \dots, U_k$ and $b_1, \dots, b_k$.

We give a second class of examples that rely on Galois ping-pong in as Lemma 10.4.

Corollary 1.12. *Let $d \geq 1$ and $\varepsilon \in (0,1)$ and $\mu_U = \sum_{i=1}^k p_i \delta_{U_i}$ an irreducible probability measure on $O(d)$ with $p_i \geq \varepsilon$ for all $1 \leq i \leq k$. Assume furthermore that $U_1, \dots, U_k$ have algebraic entries. Let $\tilde{\rho} \in (0,1)$ be sufficiently close to 1 in terms of d, ε and μ_U and let $C > 1$ be sufficiently large depending on the same parameters.*

Suppose that $g_i(x) = \frac{a_i + b_i\sqrt{q}}{c_i} U_i x + d_i$ with $a_i, b_i, c_i \in \mathbb{Z}$ and $d_i \in \mathbb{Z}^d$ for $1 \leq i \leq k$ and a prime number q do not have a common fixed point. Then the self-similar measure associated to $\mu = \sum_{i=1}^k p_i \delta_{g_i}$ is absolutely continuous if the following properties are satisfied:

(i) $\frac{a_i + b_i\sqrt{q}}{c_i} \in (\tilde{\rho}, 1)$ *for* $1 \leq i \leq k$,

(ii) for $j = 1$ and for $j = 2$ we have

$$\left| \frac{a_j - b_j\sqrt{q}}{c_j} \right| < \frac{1}{3},$$

(iii) For $L = \max(\sqrt{q}, |a_i|, |b_i|, |c_i|, |d_i|_\infty)$ we have

$$C|\chi_\mu| \leq \frac{1}{(\log(\log L))^2}.$$

As a particular case of Corollary 1.12, we can consider as shown in Lemma 10.11 the maps

$$g_i(x) = \frac{\lceil \sqrt{q} \rceil - m_{i,q} + 2\sqrt{q}}{3\lceil \sqrt{q} \rceil} U_i x + d_i$$

for any $m_{i,q} \in \mathbb{Z}$ and $d_i \in \mathbb{Z}^d$ satisfying for some $\varepsilon > 0$ that

$$m_{i,q} \in [0, q^{1/2-\varepsilon}] \qquad \text{and} \qquad |d_i|_\infty \leq \exp(\exp(q^{\varepsilon/3})).$$

Then the self-similar measure of $\mu = \sum_{i=1}^n p_i \delta_{g_i}$ is absolutely continuous for sufficiently large primes q depending on d, μ_U and ε, provided that $g_1, \dots, g_k$ do not have a common fixed point. We note that since we have a double exponential range for d_i, we get abundantly many examples.

Dimension $d \geq 3$. Finally we discuss the case when $d \geq 3$. Under this assumption, $O(d)$ is a simple non-abelian Lie group and therefore instead of using the entropy and separation rate on G we can use the same quantities on $O(d)$.

We recall that Lindenstrauss-Varjú [LV16] proved the following. Given $d \geq 3$ and $\varepsilon \in (0,1)$ as well as a finitely supported probability measure μ_U on $SO(d)$, generating a dense subgroup of $SO(d)$ and with $\mathrm{gap}_{SO(d)}(\mu_U) \geq \varepsilon$, there exists a constant $\widetilde{\rho} \in (0,1)$ depending on d and ε such that every finitely supported contracting probability measure $\mu = \sum_{i=1}^k p_i \delta_{g_i}$ on G with $U(\mu) = \mu_U$ and

$$p_i \geq \varepsilon \quad \text{as well as} \quad \rho(g_i) \in (\tilde{\rho}, 1) \qquad \text{for all} \quad 1 \leq i \leq k \tag{1.6}$$

has absolutely continuous self-similar measure ν. Moreover, [LV16] show that ν has a C^k-density if the constant $\widetilde{\rho}$ is in addition sufficiently close to 1 in terms of k. The scalar $\widetilde{\rho}$ depends on the size of the spectral gap of $U(\mu)$. By current methods ([BG08], [BdS16]) spectral gap of $U(\mu)$ is only known when $\mathrm{supp}(U(\mu))$ generates a dense subgroup and all of the entries of elements in $\mathrm{supp}(U(\mu))$ are algebraic.

We note that $h_{U(\mu)} \leq h_\mu$ yet we do not have in general that $S_{U(\mu)} \geq S_\mu$. In the case when $S_{U(\mu)} \geq S_\mu$, which for example holds when the support of $U(\mu)$ generates a free group, (1.6) follows from Theorem 1.5. Moreover, our method can be adapted to work with $S_{U(\mu)}$ instead of S_μ and thereby we establish a generalisation of (1.6) that we state in Theorem 2.5. We note that our method does not require that

$\mathrm{supp}(\mu_U)$ generates a dense subgroup of $O(d)$ or $SO(d)$ and we can also treat contracting on average self-similar measures. Moreover, as shown in Corollary 1.10 and Corollary 1.12, we can also give examples when $\mathrm{supp}(\mu_U)$ generates a finite irreducible subgroup of $O(d)$.

Discussion of other work. In addition to the above discussed [Gar62], [SX03], [LV16], [Var19] and [Kit21] there is very little known about explicit examples of absolutely continuous self-similar measures. To the authors knowledge, the only further papers addressing this topic are [DFW07] and [Str24], which are concerned with homogeneous self-similar measures on $\mathbb{R}$ whose contraction rate λ satisfies that all of its Galois conjugates have absolute value < 1.

A related problem is to study the Furstenberg measure of $\mathrm{SL}_2(\mathbb{R})$ or of arbitrary simple non-compact Lie groups. The first examples of absolutely continuous Furstenberg measures were established by [Bou12], giving an intricate number theoretic construction and also providing examples with a C^k-density for any $k \geq 1$. Bourgain's methods were generalised and further used by [BISG17], [Leq22] and [Kog22]. Moreover, numerous new examples we recently given by [Kit23].

Returning to self-similar measures, we observe that the behavior of generic self-similar measures on $\mathbb{R}$ or $\mathbb{C}$ is better understood. [Shm14] showed, thereby improving the before mentioned [Sol95], that the set of $\lambda \in (1/2, 1)$ such that the Bernoulli convolution ν_λ is singular has Hausdorff dimension zero. In [SSS18] it was shown that when the translation part (with distinct translations) and the probability vector is fixed, then generic one-dimensional self-similar measures on $\mathbb{R}$ are almost surely absolutely continuous in the range where the similarity dimension > 1. This was generalised to $\mathbb{C}$ by [SS23]. A further line of research is to show that certain parametrized families of self-similar measures or other types of invariant function systems are generically absolutely continuous, see for example [SS16b] and [BSSŚ22].

We finally mention that Fourier decay of self-similar measures was studied by many authors recently. The interested reader is referred to [LS20], [Bré21], [LS22], [Rap22], [Sol22], [VY22] and [BKS24] and as well as [ARHW21] and [BS23] for self-conformal measures.

Acknowledgement. The first-named author gratefully acknowledges support from the Heilbronn Institute for Mathematical Research. This work is part of the second-named author's PhD thesis conducted at the University of Oxford. We furthermore thank Emmanuel Breuillard and Péter Varjú for comments on a preliminary draft.

2. Main Result and Outline

In this section we first state our main results and give an outline of the proof of the main theorem in section 2.2. Then we collect for the convenience of the reader some notation used throughout this paper in section 2.3 and comment on the organisation of the paper in section 2.4.

2.1. **Main Result.** Let μ be a probability measure on $G = \mathrm{Sim}(\mathbb{R}^d)$. To state our main results in full generality we introduce notions that capture how well $U(\mu)$ mixes on $O(d)$ and how degenerate ν is.

Denote by $\gamma_1, \gamma_2, \dots$ independent samples from μ, write $q_n := \gamma_1\gamma_2\dots\gamma_n$ and given $\kappa > 0$ let τ_κ be the stopping time defined by

$$\tau_\kappa := \inf\{n \geq 1 : \rho(q_n) \leq \kappa\}.$$

We then have the following definitions.

Definition 2.1. *Let μ be a probability measure on G generating a self-similar measure ν.*

*(i) We say that μ is (α_0, θ, A)-**non-degenerate** for $\alpha_0 \in (0,1)$ and $\theta, A > 0$ if for any proper subspace $W \subset \mathbb{R}^d$ and $y \in \mathbb{R}^d$,*

$$\nu(\{x \in \mathbb{R}^d \,:\, |x - (y + W)| < \theta \text{ or } |x| \geq A\}) \leq \alpha_0.$$

*(ii) We say that μ is (c, T)-**well-mixing** for $c \in (0,1)$ and $T \geq 0$ if there is some κ_0 such that for any $\kappa < \kappa_0$ and any unit vectors $x, y \in \mathbb{R}^d$ we have*

$$\mathbb{E}[|x \cdot U(q_{\tau_\kappa + F})y|^2] \geq c,$$

where F is a uniform random variable on $[0, T]$ which is independent of the γ_i.

For $d = 1$ our measure μ will always be $(1,1)$-well-mixing. As we show in section 9.1, when $U(\mu)$ is fixed there exists (c, T) depending only on $U(\mu)$ such that μ is (c, T)-well-mixing. This follows as $U(q_F) \to m_H$ in distribution as $T \to \infty$, where H is the closure of the subgroup generated by $\operatorname{supp}(U(\mu))$ and m_H the Haar probability measure on H. The latter would not be true if we would fix F to be a deterministic random variable and therefore we have introduced the above definition.

Dealing with non-degeneracy is more involved and uniform results for many classes of self-similar measures do not hold. However, instead of our given measure we can consider a conjugated measure to establish uniform non-degeneracy results. Indeed, for $\mu = \sum_{i=1}^k p_i \delta_{g_i}$ a measure on G and $h \in G$ we denote

$$\mu_h = \sum_{i=1}^k p_i \delta_{hg_ih^{-1}} \quad \text{and} \quad \mu_h' = \frac{1}{2}\delta_e + \frac{1}{2}\sum_{i=1}^k p_i \delta_{hg_ih^{-1}}.$$

Then as we show in Lemma 9.5, absolute continuity of any of the self-similar measures of μ, μ_h or μ_h' is equivalent and all of relevant quantities such as h_μ, S_μ and $|\chi_\mu|$ are the same or comparable.

Towards Theorem 1.5, Theorem 1.6 and Theorem 1.7, as we state in Proposition 2.2 and Proposition 2.3 we have essentially uniform (c, T)-mixing and uniform (α_0, θ, A)-non-degeneracy as long as we fix $U(\mu)$. We first state a uniform mixing result adapted for Theorem 1.5 and Theorem 1.6 in the contracting case.

Proposition 2.2. *Let $d \geq 1$, $\varepsilon \in (0,1)$ and let μ_U be an irreducible probability measure on $O(d)$. Then there exists $\tilde{\rho} \in (0,1)$, (c, T) and (α_0, θ, A) depending on d, ε and μ_U such that the following holds. Let $\mu = \sum_{i=1}^k p_i \delta_{g_i}$ be a contracting probability measure on G without a common fixed point and with $U(\mu) = \mu_U$ and*

$$p_i \geq \varepsilon \quad \text{as well as} \quad \rho(g_i) \in (\tilde{\rho}, 1) \qquad \text{for all} \quad 1 \leq i \leq k.$$

Then there is $h \in G$ such that $\mu_h' = \frac{1}{2}\delta_e + \frac{1}{2}\sum_{i=1}^k p_i \delta_{hg_ih^{-1}}$ is (c, T)-well-mixing and (α_0, θ, A)-non-degenerate.

Moreover, if $\mathrm{gap}_H(\mu_U) \geq \varepsilon > 0$ *for* H *the closure of the subgroup generated by the support of* μ_U*, then there exist* (c, T) *and* (α_0, θ, A) *depending only on* d *and* ε *such that the above conclusion holds.*

For Theorem 1.7 we state a similar result for contracting on average measures.

Proposition 2.3. *Let* d *and* μ_U *be as in Theorem 2.2 and let* $\varepsilon > 0$*. Let* $\mu = \sum_{i=1}^k p_i \delta_{g_i}$ *be a contracting on average probability measure on* G *without a common fixed point satisfying* $U(\mu) = \mu_U$ *and* $p_i \geq \varepsilon$ *for* $1 \leq i \leq k$*. Then there is some* $\tilde{\rho} \in (0, 1)$ *and* $C > 1$ *depending on* d, ε *and* μ_U *such that the following holds.*

The conclusion of Theorem 2.2 holds provided that for some $\hat{\rho} \in (\tilde{\rho}, 1)$ *we have*

$$\frac{\sum_{i=1}^k |\hat{\rho} - \rho(g_i)|}{k - \sum_{i=1}^k \rho(g_i)} < 1 - \varepsilon.$$

Proposition 2.2 and Proposition 2.3 are proved in section 9. We are now in a suitable position to state our main result. Theorem 1.5 and Theorem 1.6 and Theorem 1.7 follow from the main result Theorem 2.4 by applying Proposition 2.2 and Proposition 2.3 as well as Lemma 9.5.

Theorem 2.4. *For every* $d \in \mathbb{Z}_{\geq 1}$ *and* $R, c, T, \alpha_0, \theta, A > 0$ *with* $c, \alpha_0 \in (0, 1)$ *and* $T \geq 0$ *there is a constant* $C = C(d, R, c, T, \alpha_0, \theta, A)$ *depending on* $d, R, c, T, \alpha_0, \theta$ *and* A *such that the following holds. Let* μ *be a finitely supported, contracting on average, exponentially separated,* (c, T)*-well-mixing and* (α_0, θ, A)*-non-degenerate probability measure on* G *with* $\mathrm{supp}(\mu) \subset \{g \in G : \rho(g) \in [R^{-1}, R]\}$ *and satisfying*

$$\frac{h_\mu}{|\chi_\mu|} > C \left(\max\left\{1, \log \frac{S_\mu}{h_\mu}\right\}\right)^2.$$

Then the associated self-similar measure ν *is absolutely continuous.*

A similar result for Furstenberg measures of $\mathrm{SL}_2(\mathbb{R})$ was established by the first-named author [Kit23]. However in [Kit23] it is necessary to assume that $\alpha_0 \in (0, 1/3)$ and we currently can't prove an analogue of Proposition 2.2. Therefore in the case of self-similar measures we can deduce Theorem 1.5, and therefore the examples of absolutely continuous Furstenberg measures in [Kit23] are more intricate.

We next state a version of our main theorem for $d \geq 3$ that implies (1.6) by Proposition 2.2.

Theorem 2.5. *Let* $d \geq 3$ *and* $R, c, T, \alpha_0, \theta, A > 0$ *with* $c, \alpha_0 \in (0, 1)$ *and* $T \geq 1$*. Then there is a constant* $C = C(d, R, c, T, \alpha_0, \theta, A)$ *such that the following holds. Let* μ *be a finitely supported, contracting on average,* (c, T)*-well-mixing and* (α_0, θ, A)*-non-degenerate probability measure on* G *with* $\mathrm{supp}(\mu) \subset \{g \in G : \rho(g) \in [R^{-1}, R]\}$*. Moreover assume that all of the coefficients of the matrices in* $\mathrm{supp}(U(\mu))$ *lie in the number field* K *and have logarithmic height at most* $L \geq 1$*. Then* ν *is absolutely continuous if*

$$\frac{h_{U(\mu)}}{|\chi_\mu|} \geq C \max\left\{1, \log\left(\frac{L[K:\mathbb{Q}]}{h_{U(\mu)}}\right)\right\}^2.$$

As in (1.6) we do not assume in Theorem 2.5 that all the entries of $\mathrm{supp}(\mu)$ are algebraic and only require the latter for $U(\mu)$. By the Tits alternative $h_{U(\mu)} > 0$

as long as $\mathrm{supp}(U(\mu))$ generates a non virtually solvable semigroup and moreover, Breuillard's uniform Tits alternative [Bre08] results in uniform bounds for $h_{U(\mu)}$ under suitable assumptions. The advantage of Theorem 2.5 over (1.6) is that our result is particularly effective when $U(\mu)$ has high entropy (for example when $\mathrm{supp}(U(\mu))$ generates a free semigroup) and is explicit in terms of the dependence of the heights of the coefficients of $\mathrm{supp}(U(\mu))$. In addition, Theorem 2.5 applies to only contracting on average measures and does not require $\mathrm{supp}(U(\mu))$ to generate a dense subgroup of $SO(d)$.

2.2. **Outline.** We give a sketch for the proof of Theorem 2.4. Our proof extends the strategy of [Kit23] to self-similar measures and generalises it to higher dimensions, which in turn is inspired by ideas and techniques developed in [Hoc14], [Hoc17], [Var19] and [Kit21].

Let μ be a measure on $G = \mathrm{Sim}(\mathbb{R}^d)$ and let $\gamma_1, \gamma_2, \dots$ be independent μ-distributed random variables. For a stopping time τ write $q_\tau = \gamma_1 \cdots \gamma_\tau$. Note that if x is a sample of ν then so is $q_\tau x$. The basic idea of our proof is to decompose $q_\tau x$ as a sum

$$q_\tau x = X_1 + \cdots + X_n \tag{2.1}$$

with $X_1, \dots, X_n$ independent random variables. We aim to show that for each scale $r > 0$, a suitable stopping time τ and an appropriately chosen integer k we can find a decomposition (2.1) such that for all $i \in [n]$,

$$|X_i| \leq C^{-1} r \quad \text{and} \quad \sum_{i=1}^{n} \mathrm{Var}\, X_j \geq C k r^2 I \tag{2.2}$$

for a sufficiently large fixed constant $C = C(d) > 0$ only depending on d, where $\mathrm{Var}\, X_j$ is the covariance matrix of X_j and we denote by $\geq$ the partial order defined in (2.16). The proof of Theorem 2.4 comprises to establish (2.2) and to deduce from (2.2) that ν is absolutely continuous. For the former we use adequate entropy results and for the latter we work with the detail of a measure.

From Decomposition to Absolute Continuity. The notion of Detail $s_r(\nu)$ at scale $r > 0$ of a measure ν is a tool introduced in [Kit21] measuring how smooth ν is at scale r. Detail is an analogue of the entropy between scales $1 - H(\nu; r|2r)$ used by [Var19], yet with better properties. Our goal is to deduce from (2.2) that our self-similar measure ν satisfies for r sufficiently small,

$$s_r(\nu) \leq (\log r^{-1})^{-2}, \tag{2.3}$$

which implies that ν is absolutely continuous, as shown in [Kit21].

A novelty introduced in [Kit23] is a strong product bound for detail on $\mathbb{R}$, which we prove for $\mathbb{R}^d$ in this paper. Indeed, if $\lambda_1, \dots, \lambda_k$ are measures on $\mathbb{R}^d$, $a < b$ and $r > 0$ with $s_r(\lambda_i) \leq \alpha$ for some $\alpha > 0$ and all $r \in [a, b]$ and $1 \leq i \leq k$, then, as shown in Corollary 5.6,

$$s_{a\sqrt{k}}(\lambda_1 * \cdots * \lambda_k) \leq Q'(d)(\alpha^k + k! k a^2 b^{-2}) \tag{2.4}$$

for some constant $Q'(d)$ depending only on d. To prove (2.4), [Kit23] introduced k order detail, which we generalise to $\mathbb{R}^d$. We note that (2.4) is stronger than the product bounds [Kit21, Theorem 1.17] and [Var19, Theorem 3] and is required in our proof.

To convert (2.2) into (2.3), we first partition $[n]$ as $J_1 \sqcup \ldots \sqcup J_k$ such that the random variables $Y_j = \sum_{j \in J_i} X_j$ satisfy $\operatorname{Var} Y_j \gg_d C$. Then we apply a Berry-Essen type result to deduce that Y is well-approximated by a Gaussian random variable and therefore that $s_r(Y_j) \leq \alpha$ for some constant α depending on C, with α tending to zero as C tends to ∞. Finally we conclude by (2.4) that we roughly get $s_r(\nu) \leq Q'(d)^k \alpha^k = e^{k(\log Q'(d) + \log \alpha)}$. We choose $k \asymp \log\log r^{-1}$ and therefore show (2.3) provided that α is sufficiently small in terms of d or equivalently C is sufficiently large. This proves that ν is absolutely continuous.

From Decomposition on $\mathbb{R}^d$ to Decomposition on G. It remains to explain how to establish (2.2) for $k \asymp \log\log r^{-1}$, which we first translate into an analogous question on G. Indeed, we will make a decomposition of q_τ into

$$q_\tau = g_1 \exp(U_1) g_2 \exp(U_2) \cdots g_n \exp(U_n) \tag{2.5}$$

for random variables $g_1, \ldots, g_n$ on G and $U_1, \ldots, U_n$ on the Lie algebra $\mathfrak{g}$ of G. In order to express $q_\tau v$ as a sum of random variables using (2.5), we apply Taylor's theorem in Proposition 3.4 to deduce

$$q_\tau v \approx g_1 \cdots g_n v + \sum_{i=1}^{n} \zeta_i(U_i), \tag{2.6}$$

where

$$\zeta_i = D_u(g_1 g_2 \cdots g_i \exp(u) g_{i+1} g_{i+2} \cdots g_n v)|_{u=0}.$$

For notational convenience we write in this outline of proofs

$$g_i' = g_1 \cdots g_i \quad \text{and} \quad g_i'' = g_{i+1} \cdots g_n$$

and denote

$$\rho_x = D_u(\exp(u)x)|_{u=0}.$$

Then by the chain rule, as shown in Lemma 3.3,

$$\operatorname{Var}(\zeta_i(U_i)) = \rho(g_i')^2 \, U(g_i') \operatorname{Var}(\rho_{g_i'' x}(U_i)) U(g_i')^T.$$

We will use the (c,T)-well-mixing and (α_0, θ, A)-non-degeneracy condition to ensure that

$$\operatorname{Var}(\zeta_i(U_i)) \geq c_1 \rho(g_i')^2 \operatorname{tr}(U_i) I = c_1 \operatorname{tr}(\rho(g_i') U_i) I \tag{2.7}$$

for some constant $c_1 > 0$ depending on $d, c, T, \alpha_0, \theta$ and A and where $\operatorname{tr}(U_i)$ is the trace of the covariance matrix of U_i. This will be shown in Proposition 8.3 by ensuring that each of the g_i is a product of sufficiently many γ_j such that we can apply well-mixing and non-degeneracy as $g_i x$ is close in distribution to ν. In fact, we exploit suitable properties of the derivative of ρ_x and use a principal component decomposition.

So in order to achieve (2.2), we require that

$$|U_i| \leq \rho(g_i')^{-1} r \qquad \text{and} \qquad \sum_{i=1}^{n} \operatorname{tr}(\rho(g_i') U_i) \geq C^3 c_1^{-1} (\log\log r^{-1}) r^2 \tag{2.8}$$

for the constant C from (2.2). Note that to arrive at (2.2) we replace U_i by $C^{-1} U_i$ and use (2.7).

Entropy Gap and Trace Bounds for Stopped Random Walk. We prove (2.8) by establishing suitable entropy bounds on G and then translate them to the necessary trace bounds. We use the following notation. For a random variable g on G and $s > 0$, we define $\mathrm{tr}(g; s)$ to be the supremum of all $t \geq 0$ such that we can find some σ-algebra $\mathscr{A}$ and some $\mathscr{A}$-measurable random variable h taking values in G such that

$$|\log(h^{-1}g)| \leq s \quad \text{and} \quad \mathbb{E}[\mathrm{tr}(\log(h^{-1}g)|\mathscr{A})] \geq ts^2,$$

where $\log : G \to \mathfrak{g}$ is the Lie group logarithm and we assume that $h^{-1}g$ is supported on a small ball around the identity. The reason we need to work with the conditional trace is to use (2.12).

To establish (2.8) we therefore need to find a collection of scales $s_i = \rho(g_i')^{-1}r$ such that

$$\sum_{i=1}^{n} \mathrm{tr}(q_\tau; s_i) \geq Cc_1^{-1} \log\log r^{-1} \tag{2.9}$$

for C an absolute constant depending only on d.

To show (2.9) one converts entropy estimates for q_τ into trace estimates, using in essence that for an absolutely continuous random variable Z on $\mathbb{R}^\ell$ we have

$$H(Z) \leq \frac{\ell}{2} \log\left(\frac{2\pi e}{\ell} \cdot \mathrm{tr}(Z)\right), \tag{2.10}$$

where H is the differential entropy and $\mathrm{tr}(Z)$ is the trace of the covariance matrix of Z. Equality holds in (2.10) if and only if Z is a spherical Gaussian.

We will work with entropy between scales on G. Precise definitions are given in section 6. For the purposes of this outline consider the entropy between scales defined for a random variable g taking values in G, two scales $r_1, r_2 > 0$ and a parameter $a > 0$ as

$$H_a(g; r_1|r_2) = (H(gs_{r_1,a}) - H(s_{r_1,a})) - (H(gs_{r_2,a}) - H(s_{r_2,a})),$$

where $H(\cdot)$ is the differential entropy and $s_{r,a}$ is a smoothing function supported on a ball of radius ar and satisfying for $\ell = \dim \mathfrak{g}$ that

$$\mathrm{tr}(\log(s_{r,a})) \asymp \ell r^2 \quad \text{and} \quad H(s_{r,a}) = \frac{\ell}{2}\log 2\pi e r^2 + O_d(e^{-a^2/4}) - O_{d,a}(r). \tag{2.11}$$

The function $s_{r,a}$ is chosen such that $H(s_{r,a})$ is essentially maximal while being compactly supported, which is necessary towards establishing (2.9). The parameter $a > 0$ is useful as it gives us a uniform error bound in (2.11). By using moreover (2.10), we relate in Proposition 6.14 entropy between scales and the trace by

$$\mathrm{tr}(g; 2ar) \gg a^{-2}(H_a(g; r|2r) - O_d(e^{-a^2/4}) - O_{d,a}(r)). \tag{2.12}$$

For $\kappa > 0$ denote by

$$\tau_\kappa = \inf\{n \geq 1 \,:\, \rho(\gamma_1 \cdots \gamma_n) \leq \kappa\}.$$

It is then shown in Proposition 7.1 for $r_1 < r_2$ and with $r_1 \leq \kappa^{\frac{S_\mu}{|\chi_\mu|}}$ that as $\kappa \to 0$ the following entropy gap holds:

$$H_a(q_{\tau_\kappa}; r_1|r_2) \geq \left(\frac{h_\mu}{|\chi_\mu|} - d\right) \log \kappa^{-1} + \ell \cdot \log r_2 + o_{\mu,d,a}(\log \kappa^{-1}). \tag{2.13}$$

We will give a sketch of the proof of (2.13) in the beginning of section 7 and just note that the main point of (2.13) is that most of the elements in the support of

q_{τ_κ} are separated by $\kappa^{\frac{S_\mu}{|\chi_\mu|}}$, which by standard properties of entropy implies that $H(q_{\tau_\kappa} s_{r_1,a}) \approx H(q_{\tau_\kappa}) + H(s_{r_1,a})$. As we require to use a stopping time in (2.13), we will need to work with q_τ instead of a deterministic time throughout our proof.

By (2.13) it follows, assuming $h_\mu/|\chi_\mu|$ is sufficiently large and κ is sufficiently small, that

$$H_a(q_{\tau_\kappa}; \kappa^{\frac{S_\mu}{|\chi_\mu|}} | \kappa^{\frac{h_\mu}{2\ell|\chi_\mu|}}) \gg_d \frac{h_\mu}{|\chi_\mu|} \log \kappa^{-1}. \tag{2.14}$$

Using (2.14) and (2.12), we show in Proposition 7.5 with setting $S = 2\max\{S_\mu, h_\mu\}$ that for a collection of scales

$$s_i \in (\kappa^{\frac{S}{|\chi_\mu|}}, \kappa^{\frac{h_\mu}{2\ell|\chi_\mu|}}) \qquad \text{with} \qquad 1 \le i \le \hat{m}$$

and $\hat{m}$ being a fixed constant depending on S_μ and χ_μ that

$$\sum_{i=1}^{\hat{m}} \operatorname{tr}(q_{\tau_\kappa}; s_i) \gg_d \frac{h_\mu}{|\chi_\mu|} \max\left\{1, \log \frac{S_\mu}{h_\mu}\right\}^{-1}. \tag{2.15}$$

As we explain at the beginning of section 7, the error term $\max\left\{1, \log \frac{S_\mu}{h_\mu}\right\}^{-1}$ arises from the error $O_d(e^{-a^2/4})$ in (2.12).

Conclusion of Proof. The trace bound (2.15) is not sufficient to establish (2.9) as we require a lower bound depending on $\log\log r^{-1}$. To achieve such a bound and to conclude the proof, we concatenate several decompositions arising from (2.15) and therefore develop a suitable theory of such decompositions in section 8.

It therefore remains to find sufficiently many parameters $\kappa_1, \dots, \kappa_m$ such that the resulting intervals

$$(\kappa_1^{\frac{S}{|\chi_\mu|}}, \kappa_1^{\frac{h_\mu}{2\ell|\chi_\mu|}}), \quad (\kappa_2^{\frac{S}{|\chi_\mu|}}, \kappa_2^{\frac{h_\mu}{2\ell|\chi_\mu|}}), \quad \dots \quad (\kappa_m^{\frac{S}{|\chi_\mu|}}, \kappa_m^{\frac{h_\mu}{2\ell|\chi_\mu|}})$$

are disjoint. As we require that all of the scales are $\ge r$, we set $\kappa_1 = r^{\frac{|\chi_\mu|}{S}}$. On the other hand, we want all scales to be sufficiently small. We for example therefore require that $\kappa_m^{\frac{h_\mu}{2\ell|\chi_\mu|}} < e^{-10}$. Thus setting $\kappa_{i+1} = \kappa_i^{\frac{h_\mu}{3\ell S}}$, thereby ensuring that the resulting intervals are disjoint (provided h_μ/χ_μ is sufficiently large), a calculation shows that the maximal m we can take is

$$\max\left\{1, \log \frac{S_\mu}{h_\mu}\right\}^{-1} \log\log r^{-1} \ll m \ll_\mu \log\log r^{-1}.$$

Combining all of the above, it follows that when summing over all the scales

$$\sum_i \operatorname{tr}(q_{\tau_{\kappa_1}}; s_i) \gg_d \frac{h_\mu}{|\chi_\mu|} \max\left\{1, \log \frac{S_\mu}{h_\mu}\right\}^{-2} \log\log r^{-1}.$$

We therefore require in order to satisfy (2.9) that

$$\frac{h_\mu}{|\chi_\mu|} \max\left\{1, \log \frac{S_\mu}{h_\mu}\right\}^{-2} \ge C^3 c_1^{-1},$$

which leads us to the condition from Theorem 2.4 and concludes our sketch of proof.

2.3. **Notation.** We use the asymptotic notation $A \ll B$ or $A = O(B)$ to denote that $|A| \le CB$ for a constant $C > 0$. If the constant C depends on additional parameters we add subscripts. Moreover, $A \asymp B$ denotes $A \ll B$ and $B \ll A$.

For an integer $n \ge 1$ we abbreviate $[n] = \{1, 2, \dots, n\}$.

Given two positive semi-definite symmetric real $d \times d$ matrices M_1 and M_2 we write

$$M_1 \ge M_2 \qquad \text{if and only if} \qquad x^T M_1 x \ge x^T M_2 x \quad \text{for all } x \in \mathbb{R}^d. \tag{2.16}$$

For a random variable X on $\mathbb{R}^d$ we denote by $\mathrm{Var}(X)$ the covariance matrix of X and by $\mathrm{tr}(X) = \mathrm{tr}\,\mathrm{Var}(X)$ the trace of the covariance matrix.

Given a metric space (M, d), $p \in [1, \infty)$ and two probability measures λ_1 and λ_2 on M, we define

$$\mathcal{W}_p(\lambda_1, \lambda_2) = \inf_{\gamma \in \Gamma(\lambda_1, \lambda_2)} \left(\int_{M \times M} d(x, y)^p \, d\gamma(x, y) \right)^{\frac{1}{p}}, \tag{2.17}$$

where $\Gamma(\lambda_1, \lambda_2)$ is the set of couplings of λ_1 and λ_2, i.e. of probability measures γ on $M \times M$ whose projections to the first coordinate is λ_1 and to the second is λ_2.

Throughout this paper we fix $d \ge 1$ and write $G = \mathrm{Sim}(\mathbb{R}^d)$, except in section 6 where G is an arbitrary Lie group. The Lie algebra of G will be denoted $\mathfrak{g}$ and $\ell = \dim \mathfrak{g}$. We usually consider a fixed probability measure μ on G and independent samples $\gamma_1, \gamma_2, \dots$ of μ. We write for $\kappa > 0$

$$q_n = \gamma_1 \cdots \gamma_n \qquad \text{and} \qquad \tau_\kappa = \inf\{n \ge 1\, ;\, \rho(\gamma_n) \le \kappa\}.$$

When μ is a probability measure on $G = \mathrm{Sim}(\mathbb{R}^d)$ and ν is a probability measure $\mathbb{R}^d$ we denote by $\mu * \nu$ the probability measure uniquely characterized by

$$(\mu * \nu)(f) = \int \int f(gx) \, d\mu(g) d\nu(x)$$

for $f \in C_c(\mathbb{R}^d)$. When $\mu = \sum_i p_i \delta_{g_i}$ is finitely supported, then

$$\mu * \nu = \sum_i p_i g_i \nu, \tag{2.18}$$

where $g_i \nu$ is the pushforward of ν by g_i defined by $(g_i \nu)(B) = \nu(g_i^{-1} B)$ for all Borel sets $B \subset \mathbb{R}^d$.

Given a random variable g on G we denote, as defined in section 6, by $H(g)$ the Shannon entropy when g is discrete and the differential entropy when g is absolutely continuous. The various notions of entropy between scales as well as $\mathrm{tr}(g, r)$ are defined in section 6.

We will denote by m_G a normalised Haar measure on $\mathrm{Sim}(\mathbb{R}^d)$. Moreover if $H \subset O(d)$ is a closed subgroup, we will denote by m_H the Haar probability measure on H. For a probability measure μ_U on H, the L^2-spectral gap of μ_U in H is defined as

$$\mathrm{gap}_H(\mu_U) = 1 - ||T_{\mu_U}|_{L^2_0(G)}||, \tag{2.19}$$

where $(T_{\mu_U} f)(k) = \int f(hk) \, d\mu_U(h)$ for $f \in L^2(H)$ and $L^2_0(H) = \{f \in L^2(H) : m_H(f) = 0\}$ for $|| \circ ||$ the operator norm.

2.4. **Organisation.** In section 3 the Taylor expansion bound (2.6) is proved and we establish several probabilistic preliminaries. In section 4 we prove and generalise Theorem 1.2. We discuss order k detail in section 5, establish (5.2) as well as show how to convert (2.2) into suitable detail bounds. A theory of entropy on general Lie groups suitable for our purposes is developed in section 6 and (2.12) is shown. In section 7 we prove (2.13) and (2.15). Finally, we deduce Theorem 2.4 as well as Theorem 2.5 in section 8 by developing a decomposition theory for stopped random walks. We study (c, T)-well-mixing and (α_0, θ, A)-non-degeneracy in section 9 and prove Proposition 2.2 and Proposition 2.3. In section 10 we discuss explicit examples and in particular we prove Corollary 1.8, Corollary 1.9, Corollary 1.10, Corollary 1.11 and Corollary 1.12.

3. Preliminaries

In this section we first study the derivatives of the G action on $\mathbb{R}^d$ in section 3.1. As we want to work with conditional variance and entropy, we discuss regular conditional distributions in section 3.2 and then versions of the large deviation principle in section 3.3.

3.1. Derivative Bounds.

3.1.1. *Basic Properties.* Let $G = \mathrm{Sim}(\mathbb{R}^d)$ with Lie algebra $\mathfrak{g} = \mathrm{Lie}(G)$. For $x \in \mathbb{R}^d$ consider the map

$$w_x : \mathfrak{g} \to \mathbb{R}^d, \qquad u \mapsto \exp(u)x.$$

Denote by $\psi_x = D_0 w_x : \mathfrak{g} \to \mathbb{R}^d$ the differential at zero of w_x.

Note that we can embed $G = \mathrm{Sim}(\mathbb{R}^d)$ into $\mathrm{GL}_{d+1}(\mathbb{R})$ via the map

$$g \mapsto \begin{pmatrix} r(g)U(g) & b(g) \\ 0 & 1 \end{pmatrix}.$$

Therefore we can write $u \in \mathfrak{g}$ as $u = \left(\begin{smallmatrix} \alpha & \beta \\ 0 & 0 \end{smallmatrix}\right)$ with $\alpha \in \mathbb{R} \cdot \mathfrak{so}_d(\mathbb{R})$ and $\beta \in \mathbb{R}^d$. Thus it follows that $\psi_x(u) = u\left(\begin{smallmatrix} x \\ 1 \end{smallmatrix}\right) = \alpha x + \beta$. With this viewpoint we also use the following notation

$$ux = \psi_x(u) = \alpha x + \beta \tag{3.1}$$

The above embedding endows $\mathfrak{g}$ with a coordinate system, a natural inner product and denote by $|\circ|$ the associated norm. We collect some properties about the derivatives of w_x, ψ_x and the map g. For notational convenience we denote throughout this subsection by $\frac{\partial f}{\partial x}$ the derivative $D_x f$ of a function $f : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$ at a vector $x \in \mathbb{R}^{d_1}$. We furthermore write $\ell = \dim \mathfrak{g}$.

Lemma 3.1. *The following properties hold:*

(i) Let $g = \rho U + b \in G$. Then for all $x \in \mathbb{R}^d$, it holds that $\frac{\partial g}{\partial x} = \rho U$ and all of the second derivatives of g are zero.

(ii) Whenever $|u| \leq 1$ and $1 \leq i, j \leq \ell$,

$$\left| \frac{\partial w_x}{\partial u_i} \right| \ll_d |x| \quad \text{and} \quad \left| \frac{\partial w_x}{\partial u_i \partial u_j} \right| \ll_d |x|.$$

(iii) For any $x_1, x_2 \in \mathfrak{g}$ we have that

$$||\psi_{x_1} - \psi_{x_2}|| \ll_d |x_1 - x_2|.$$

(iv) Let $u \in \mathfrak{g} \backslash \{0\}$. Then there is a proper subspace $W_u \subset \mathbb{R}^d$ and a vector $u_0 \in \mathbb{R}^d$ such that if $\psi_x(u) = 0$ then $x \in u_0 + W_u$ for $x \in \mathbb{R}^d$.

(v) *For all $\theta, A > 0$ there is $\delta > 0$ such that the following is true. Let $v \in \mathfrak{g}$ be a unit vector. Then there is a proper subspace $W_v \subset \mathbb{R}^d$ and a vector $v_0 \in \mathbb{R}^d$ such that if*
$$x \in \mathbb{R}^d \backslash B_\theta(v_0 + W_v) \quad \text{and} \quad |x| \leq A$$
for $B_\theta(v_0 + W_v)$ the θ-ball around $v_0 + W_v$ then
$$|\psi_x(v)| \geq \delta.$$

Proof. (i) follows by definition and (ii) by compactness. For (iii) using notation (3.1) it holds for $u \in \mathfrak{g}$ with $|u| \leq 1$ that
$$\begin{aligned}|\psi_{x_1}(u) - \psi_{x_2}(u)| &= |\alpha x_1 - \alpha x_2| \leq ||\alpha|| \cdot |x_1 - x_2| \\ &\ll_d |\alpha| \cdot |x_1 - x_2| \leq |u| \cdot |x_1 - x_2|\end{aligned}$$
using that the operator norm $||\circ||$ is equivalent to the inner product norm on $\mathfrak{g}$. To show (iv), we may assume that $\beta \in \mathrm{Im}(\alpha)$ as otherwise there is nothing to show. Then set $W_u = \ker(\alpha)$ and $u_0 \in \mathbb{R}^d$ such that $\alpha u_0 = -\beta$, implying the claim. (v) follows from (iv) by continuity. □

For $u \in \mathfrak{g} \backslash \{0\}$ we define
$$E_\theta(u) = \mathbb{R}^d \backslash B_\theta(u_0 + W_u).$$

Given a random variable U taking values in $\mathfrak{g}$, we say that $u \in \mathfrak{g}$ is a first principal component if it is an eigenvector of its covariance matrix with maximal eigenvalue. Set
$$E_\theta(U) = \bigcup_{v \in P} E_\theta(v),$$
where P is the set of first principal components of U. Similarly if μ is a probability measure which is the law of a random variable U then we define $E_\theta(\mu) = E_\theta(U)$. Recall that given a random variable U in $\mathbb{R}^\ell$, we denote by $\mathrm{tr}(U)$ the trace of the covariance matrix of U.

Proposition 3.2. *For all theta, $A > 0$ there is some $\delta = \delta(d, \theta, A) > 0$ such that the following is true. Suppose that U is a random variable taking values in $\mathfrak{g}$ and that $x \in \mathbb{R}^d$ with $|x| \leq A$. Suppose that $x \in E_\theta(U)$. Then*
$$\mathrm{tr}(Ux) \geq \delta \cdot \mathrm{tr}(U).$$

Proof. We used here the notation (3.1) that $\psi_x(U) = Ux$. Write $\ell = \dim \mathfrak{g}$ and let $w_1, \ldots, w_\ell$ be an orthonormal basis of eigenvectors of the covariance matrix $\mathrm{Var}(U)$. We may assume that U has mean zero. Denote by $U_i = \langle U, w_i \rangle = U^T w_i$ for $1 \leq i \leq \ell$ and assume without loss of generality that $\mathrm{Var}(U_1) \geq \ldots \geq \mathrm{Var}(U_\ell)$ so that w_1 is a principal component. Then the $(U_i)_{1 \leq i \leq \ell}$ are uncorrelated since for $i \neq j$
$$\begin{aligned}\mathrm{cov}(U_i, U_j) = \mathbb{E}[U_i U_j] &= \mathbb{E}[\langle U^T w_i, U^T w_j \rangle] \\ &= \mathbb{E}[\langle U U^T w_i, w_j \rangle] = \langle \mathrm{Var}(U) w_i, w_j \rangle = 0\end{aligned}$$
and it holds that $U = \sum_{i=1}^{\ell} U_i w_i$ and that $\mathrm{Var}(U_1) \geq \frac{1}{\ell} \mathrm{tr}(U)$. Also by Proposition 3.1 (v) it holds that $|\psi_x(w_1)| \geq \delta$. We then compute
$$\mathrm{tr}(\rho_x(U)) = \mathbb{E}[|\rho_x(U)|^2] = \mathbb{E}\left[\sum_{i=1}^{\ell} U_i^2 |\rho_x(w_i)|^2\right] \geq \mathbb{E}[U_1^2 |\rho_x(w_1)|^2] \geq \frac{\delta}{\ell} \mathrm{tr}(U).$$

□

Lemma 3.3. *Let U be a random variable on $\mathfrak{g}$ and let $g \in G$ and $x \in \mathbb{R}^d$. Denote*

$$\zeta = D_u g \exp(u) x|_{u=0}.$$

Then

$$\operatorname{Var}(\zeta(U)) = \rho(g)^2 \cdot U(g)\psi_x \circ \operatorname{Var}(U) \circ \psi_x^T U(g)^T.$$

Proof. Note that by the chain rule $\zeta(U) = \rho(g)U(g)\psi_x(U)$ and therefore

$$\operatorname{Var} \zeta(U) = \rho(g)^2 U(g) \operatorname{Var}(\psi_x(U)) U(g)^T$$

Viewing $\psi_x : \mathfrak{g} \to \mathbb{R}^d$ as a matrix with our choice of coordinate system we write $\psi_x(U) = \psi_x \circ U$ and the claim follows. □

3.1.2. *Taylor Expansion Bound.* The aim of this subsection is to prove the following proposition, which crucially relies on the G action on $\mathbb{R}^d$ having no second derivatives.

Proposition 3.4. *For every $A > 0$ there exists $C = C(d, A) > 1$ such that the following holds. Let $n \geq 1$, $r \in (0, 1)$ and let $u^{(1)}, \ldots, u^{(n)} \in \mathfrak{g}$. Let $g_1, \ldots, g_n \in G$ with*

$$\rho(g_i) < 1, \qquad |b(g_i)| \leq A \quad \text{and} \quad |u^{(i)}| \leq \rho(g_1 \cdots g_i)^{-1} r < 1.$$

Let $v \in \mathbb{R}^d$ with $|v| \leq A$ and write

$$x = g_1 \exp(u^{(1)}) \cdots g_n \exp(u^{(n)}) v$$

and

$$\zeta_i = D_0(g_1 g_2 \cdots g_i \exp(u) g_{i+1} \cdots g_{n-1} g_n v)$$

and let

$$S = g_1 \cdots g_n v + \sum_{i=1}^{n} \zeta_i(u^{(i)}).$$

Then it holds that

$$|x - S| \leq C^n \rho(g_1 \cdots g_n)^{-1} r^2.$$

To prove Proposition 3.4 we use the following version of Taylor's theorem.

Theorem 3.5. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^2-function, let $R_1, \ldots, R_n > 0$ and write $U = [-R_1, R_1] \times \ldots \times [-R_n, R_n]$. For integers $i, j \in [1, n]$ let $K_{ij} = \sup_U |\frac{\partial^2 f}{\partial x_i \partial x_j}|$ and let $x \in U$. Then we have that*

$$\left| f(x) - f(0) - \sum_{i=1}^{n} x_i \frac{\partial f}{\partial x_i}\bigg|_{x=0} \right| \leq \frac{1}{2} \sum_{i,j=1}^{n} x_i K_{i,j} x_j.$$

Lemma 3.6. *Let*

$$w : \mathfrak{g} \times \mathfrak{g} \longrightarrow \mathbb{R}^d, \qquad (x, y) \longmapsto \exp(x) g \exp(y) v$$

for fixed g, v. Then if $|x|, |y| \leq 1$ it holds that

$$\left| \frac{\partial w(x, y)}{\partial x_i \partial y_i} \right| \ll_d \rho(g)|v|.$$

Proof. Let $\hat{v} = \exp(y)v$ and note that by compactness $|\frac{\partial \hat{v}}{\partial y_i}| \ll_d |v|$. Now let $\tilde{v} = g\hat{v}$. Therefore by Lemma 3.1 (i), $||\frac{\partial \tilde{v}}{\partial \hat{v}}|| \leq \rho(g)$ and by compactness $||\frac{\partial^2 w}{\partial x_i \partial \tilde{v}}|| \ll_d 1$. We conclude therefore by the chain rule

$$\left|\frac{\partial w}{\partial x_i \partial y_i}\right| = \left|\left|\frac{\partial w}{\partial x_i \partial \tilde{v}}\right|\right| \cdot \left|\left|\frac{\partial \tilde{v}}{\partial \hat{v}}\right|\right| \cdot \left|\frac{\partial \hat{v}}{\partial y_i}\right| \ll_d \rho(g)|v|.$$

□

Proposition 3.7. *There exists a constants $C = C(d) > 1$ such that the following holds. Suppose that $n \in \mathbb{Z}_{>0}$, $g_1, g_2, \dots, g_n \in G$ and let $u^{(1)}, \dots, u^{(n)} \in \mathfrak{g}$ be such that $|u^{(i)}| \leq 1$.*

Let $v \in \mathbb{R}^d$ and

$$x = g_1 \exp(u^{(1)}) g_2 \exp(u^{(2)}) \cdots g_n \exp(u^{(n)}) v.$$

Then for any $1 \leq i, j \leq \ell$ and any integers $k, \ell \in [1, n]$ with $k \leq \ell$ we have

$$\left|\frac{\partial^2 x}{\partial u_i^{(k)} \partial u_j^{(\ell)}}\right| \leq C^n \rho(g_1 \cdots g_\ell) |g_{\ell+1} \exp(u^{(\ell+1)}) \cdots g_n \exp(u^{(n)}) v|.$$

Proof. First we deal with the case $k = \ell$. Let

$$a = g_1 \exp(u^{(1)}) g_2 \exp(u^{(2)}) \cdots g_{k-1} \exp(u^{(k-1)}) g_k$$

and

$$b = g_{k+1} \exp(u^{(k+1)}) g_{k+2} \exp(u^{(k+2)}) \cdots g_n \exp(u^{(n)}) v$$

and let $\tilde{b} = \exp(u^{(k)})b$. We have

$$\frac{\partial x}{\partial u_i^{(k)}} = \frac{\partial x}{\partial \tilde{b}} \frac{\partial \tilde{b}}{\partial u_i^{(k)}}.$$

Note that by Lemma 3.1 (i) all of the second derivatives of x with respect to $\tilde{b}$ are zero and therefore

$$\left|\frac{\partial^2 x}{\partial u_i^{(k)} \partial u_j^{(k)}}\right| \leq \left|\left|\frac{\partial x}{\partial \tilde{b}}\right|\right| \cdot \left|\frac{\partial^2 \tilde{b}}{\partial u_i^{(k)} \partial u_j^{(k)}}\right|. \tag{3.2}$$

Thus by Lemma 3.1 (i) and (ii) we conclude that

$$\left|\frac{\partial^2 x}{\partial u_i^{(k)} \partial u_j^{(k)}}\right| \ll_d \rho(a)|b| \leq C^n \rho(g_1 \cdots g_\ell)|b|$$

for a suitable constant $C > 1$ using that $\rho(\exp(u^{(i)}))$ is bounded.

For the case $k < \ell$ we consider

$$a_1 = g_1 \exp(u^{(1)}) g_2 \exp(u^{(2)}) \cdots g_{k-1} \exp(u^{(k-1)}) g_k$$
$$a_2 = g_{k+1} \exp(u^{(k+1)}) g_{k+2} \exp(u^{(k+2)}) \cdots g_\ell$$
$$b = g_{\ell+1} \exp(u^{(\ell+1)}) g_{\ell+2} \exp(u^{(\ell+2)}) \cdots g_n \exp(u^{(n)}) v.$$

Then we consider $\tilde{b} = \exp(u^{(k)}) a_2 \exp(u^{(l)}) b$ and as before we conclude

$$\frac{\partial^2 x}{\partial u_i^{(k)} \partial u_j^{(k)}} = \frac{\partial x}{\partial \tilde{b}} \frac{\partial^2 \tilde{b}}{\partial u_i^{(k)} \partial u_j^{(k)}}.$$

We again arrive at (3.2) and deduce the claim as in the case $k = \ell$ using Lemma 3.6 instead of Lemma 3.1 (i). □

Proof. (of Proposition 3.4) We first show that there is a constant $C_1 = C_1(A, d)$ depending on A such that for all $1 \leq i \leq n$ we have that

$$|g_i \exp(u^{(i)}) \cdots g_n \exp(u^{(n)})v| \leq C_1^{n-i+1}. \tag{3.3}$$

Indeed, we note that for any $u \in \mathfrak{g}$ with $|u| \leq 1$ and $v_0 \in \mathbb{R}^d$ it holds that $|\exp(u)v_0 - v_0| \leq C_2(|v_0| + 1)$ for an absolute constant $C_2 = C_2(d)$. Without loss of generality we assume that $C_2(d) > 1$. Therefore $|\exp(u^{(n)})v| \leq C_2(2|v| + 1)$. Next note that as $\rho(g_n) < 1$,

$$\begin{aligned} |g_n \exp(u^{(n)})v| &\leq |g_n \exp(u^{(n)})v - g_n(0)| + |g_n(0)| \\ &\leq \rho(g_n)|\exp(u^{(n)})v| + |b(g_n)| \\ &\leq C_2(2|v| + |b(g_n)| + 1) \leq 4C_2(A + 1), \end{aligned}$$

using that $\rho(g_n) < 1$ and that $|v| \leq A$ and $|b(g_n)| \leq A$. Continuing this argument inductively, we may conclude that

$$|g_i \exp(u^{(i)}) \cdots g_n \exp(u^{(n)})v| \leq 4^{n-i+1} C_2^{n-i+1}(A + (n - i) + 1),$$

which implies (3.3).

By applying Theorem 3.5 together with Proposition 3.7 and (3.3) for a sufficiently large constant C depending on A and d in each of the coordinates of $\mathbb{R}^d$,

$$|x - S| \leq dn^2 C^n \rho(g_1 \cdots g_n)^{-1} r^2,$$

which implies the claim upon enlarging the constant C. □

3.2. **Regular Conditional Distributions.** In this section we review the definition of regular conditional distributions that will be used in section 6. On a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, we denote the conditional expectation by $\mathbb{E}[f|\mathscr{A}]$ for $f \in L^1(\Omega, \mathscr{F}, \mathbb{P})$ and a σ-algebra $\mathscr{A} \subset \mathscr{F}$. Given two measurable spaces $(\Omega_1, \mathscr{A}_1)$ and $(\Omega_2, \mathscr{A}_2)$, recall that a Markov kernel on $(\Omega_1, \mathscr{A}_1)$ and $(\Omega_2, \mathscr{A}_2)$ is a map $\kappa : \Omega_1 \times \mathscr{A}_2 \to [0, 1]$ if for any $A_2 \in \mathscr{A}$, the map $\kappa(\cdot, A_2)$ is $\mathscr{A}_1$-measurable and for any ω_1 the map $A_2 \to \kappa(\omega_1, A_2)$ is a probability measure.

Definition 3.8. *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $\mathscr{A} \subset \mathscr{F}$ be a σ-algebra. Let (E, ξ) be a measurable space and let $Y : (\Omega, \mathscr{F}) \to (E, \xi)$ be a random variable. Then we say that a Markov kernel*

$$(Y|\mathscr{A}) : \Omega \times \xi \to [0, 1]$$

*on $(\Omega, \mathscr{F})$ and (E, ξ) is a **regular conditional distribution** if*

$$(Y|\mathscr{A})(\omega, B) = \mathbb{P}[Y \in B \,|\, \mathscr{A}](\omega) = \mathbb{E}[1_{Y^{-1}(B)} \,|\, \mathscr{A}](\omega).$$

In other words,

$$\mathbb{E}[(Y|\mathscr{A})(\cdot, B)1_A] = \mathbb{P}[A \cap \{Y \in B\}]$$

for all $A \in \mathscr{A}$.

Regular conditional distributions exists whenever $(\Omega, \mathscr{F}, \mathbb{P})$ is a standard probability space. To give a construction, recall (c.f. section 3 of [EW11]) that there exist conditional measures $\mathbb{P}_\omega^{\mathscr{A}}$ uniquely characterized by

$$\mathbb{E}[f|\mathscr{A}](\omega) = \int f \, d\mathbb{P}_\omega^{\mathscr{A}}.$$

Then

$$(Y|\mathscr{A})(\omega, \cdot) = Y_* \mathbb{P}_\omega^{\mathscr{A}}$$

Indeed,

$$(Y|\mathscr{A})(\omega, B) = E[1_{Y^{-1}(B)}|\mathscr{A}](\omega) = \int 1_{Y^{-1}(B)}\, d\mathbb{P}_\omega^{\mathscr{A}} = \mathbb{P}_\omega^{\mathscr{A}}(Y^{-1}(B)) = Y_*\mathbb{P}_\omega^{\mathscr{A}}(B).$$

We denote by $[Y|\mathscr{A}]$ a random variable defined on a separate probability space with law $(Y|\mathscr{A})$.

We recall that given two further σ-algebras $\mathscr{G}_1, \mathscr{G}_2 \subset \mathscr{F}$, we say that they are independent given $\mathscr{A}$ if for all $U \in \mathscr{G}_1$ and $V \in \mathscr{G}_2$

$$\mathbb{P}[U \cap V|\mathscr{A}] = \mathbb{P}[U|\mathscr{A}]\mathbb{P}[V|\mathscr{A}]$$

almost surely. Similarly, two random variables Y_1 and Y_2 are independent given $\mathscr{A}$ if the σ-algebra they generate are. Note that if Y_1 is $\mathscr{A}$-measurable, then it is independent given $\mathscr{A}$ to every random variable Y_2.

Given a topological group G and two measures μ_1 and μ_2 we recall that the convolution $\mu_1 * \mu_2$ is defined as

$$(\mu_1 * \mu_2)(B) = \int\int 1_B(gh)\, d\mu_1(g) d\mu_2(g)$$

for any measurable set $B \subset G$.

Lemma 3.9. *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, G be a topological group and g, h be G-valued random variables. Let $\mathscr{A} \subset \mathscr{F}$ be a σ-algebra and assume that g and h are independent given $\mathscr{A}$. Then the following properties hold:*

(i) $(gh|\mathscr{A}) = (g|\mathscr{A}) * (h|\mathscr{A})$ *almost surely.*
(ii) $[gh|\mathscr{A}] = [g|\mathscr{A}] \cdot [h|\mathscr{A}]$ *almost surely.*

Proof. To show (i) we note that by assumption g and h are independent with respect to $\mathbb{P}_\omega^{\mathscr{A}}$ for almost all $\omega \in \Omega$. This implies that

$$\mathbb{E}_{\mathbb{P}_\omega^{\mathscr{A}}}[f(gh)] = \mathbb{E}_{\mathbb{P}_\omega^{\mathscr{A}}}[\mathbb{E}_{\mathbb{P}_\omega^{\mathscr{A}}}[f(gh)|h]] = \mathbb{E}_{(z_1,z_2)\sim\mathbb{P}_\omega^{\mathscr{A}}\times\mathbb{P}_\omega^{\mathscr{A}}}[f(g(z_1)h(z_1))],$$

proving (i). (ii) follows from (i) on a suitable separate probability space. □

3.3. **Large Deviation Principle.** In this subsection we review various versions of the large deviation principle. Applying the classical large deviation principle to ρ, we can state the following. Throughout this section we denote by μ a measure on G and by $\gamma_1, \gamma_2, \ldots$ independent samples from μ.

Lemma 3.10. *Let μ be a contracting on average probability measure on G. Then for every $\varepsilon > 0$ there is $\delta = \delta(\mu, \varepsilon) > 0$ such that for all sufficiently large n,*

$$\mathbb{P}\Big[\, |n\chi_\mu - \log \rho(\gamma_1)\cdots\rho(\gamma_n)| > \varepsilon n \Big] \leq e^{-\delta n}.$$

We generalise Lemma 3.10 to stopping times.

Lemma 3.11. *Let μ be a compactly supported contracting on average probability measure on G and let $\kappa > 0$ and denote*

$$\tau_\kappa = \inf\{n \geq 1 \,:\, \rho(\gamma_1 \ldots \gamma_n) \leq \kappa\}.$$

Then for every $\varepsilon > 0$ there is $\delta > 0$ such that for sufficiently small κ

$$\mathbb{P}\Big[\Big|\tau_\kappa - \frac{\log \kappa^{-1}}{|\chi_\mu|}\Big| > \varepsilon \log \kappa^{-1}\Big] \leq e^{-\delta \log \kappa^{-1}}$$

Proof. If $\tau_\kappa > \frac{\log \kappa^{-1}}{|\chi_\mu|} + \varepsilon \log \kappa^{-1}$ then

$$\rho(\gamma_1 \cdots \gamma_{\lfloor \frac{\log \kappa^{-1}}{|\chi_\mu|} + \varepsilon \log \kappa^{-1} \rfloor}) \geq \kappa,$$

which by Lemma 3.10 has probability at most $e^{-\delta \log \kappa^{-1}}$ for some $\delta > 0$ and sufficiently small κ.

Write $R = \inf\{\rho(g) \,:\, g \in \operatorname{supp}(\mu)\} \in (0,1)$, which is non-zero since μ is compactly supported. Therefore when $\tau_\kappa < \frac{\log \kappa^{-1}}{|\chi_\mu|} - \varepsilon \log \kappa^{-1}$ happens there must be some integer

$$k \in \left[\frac{\log \kappa^{-1}}{|\log R|}, \frac{\log \kappa^{-1}}{|\chi_\mu|} - \varepsilon \log \kappa^{-1}\right]$$

such that

$$\log \rho(\gamma_1 \cdots \gamma_k) \leq \log \kappa.$$

Note that for sufficiently small κ we have $k|\chi_\mu| \leq \log \kappa^{-1} - \varepsilon|\chi_\mu|\,|\log R|$ and therefore

$$\log \rho(\gamma_1 \cdots \gamma_k) \leq \log \kappa \leq k(\chi_\mu + \varepsilon|\log R|\chi_\mu). \tag{3.4}$$

By Lemma 3.10 the probability that (3.4) happens is $\leq e^{-\delta' k} = e^{-\delta' O_\mu(\log \kappa^{-1})}$ for some $\delta' > 0$. Since there are at most $O_\mu(\log \kappa^{-1})$ many possibilities for k, the claim follows by the union bound. □

From Lemma 3.10 and Theorem 1.2 we can deduce the following corollary.

Corollary 3.12. *Let μ be a contracting on average probability measure on G. Then for every $\varepsilon > 0$ there is $\delta = \delta(\mu, \varepsilon) > 0$ such that for all sufficiently large N*

$$\mathbb{P}\Big[\exists n \geq N : \rho(\gamma_1 \cdots \gamma_n) \geq \exp((\chi_\mu + \varepsilon)n)\Big] \leq e^{-\delta N} \tag{3.5}$$

and

$$\mathbb{P}\Big[\exists n, m \geq N : |b(\gamma_1 \cdots \gamma_n) - b(\gamma_1 \cdots \gamma_m)| \geq \exp((\chi_\mu + \varepsilon)\min(m,n))\Big] \leq e^{-\delta N}.$$

Proof. (3.5) follows from Lemma 3.10 and Borel-Cantelli. For (3.12) note that when $m \geq n+1$,

$$|b(\gamma_1 \cdots \gamma_n) - b(\gamma_1 \cdots \gamma_m)| \leq \rho(\gamma_1 \cdots \gamma_n)|b(\gamma_{n+1} \cdots \gamma_m)|.$$

Therefore by (3.5) it suffices to show that for sufficiently large N we have that

$$\mathbb{P}[\exists k \geq 1 \,:\, |b(\gamma_1 \cdots \gamma_k)| \geq e^{\varepsilon N}] \leq e^{-\delta N},$$

which readily follows from Theorem 1.2 and Borel-Cantelli as $b(\gamma_1 \cdots \gamma_k)$ converges exponentially fast in distribution to ν. □

The next lemma was proved in [Kit23].

Lemma 3.13. *(Corollary 7.9 of* [Kit23]*) There is a constant $c > 0$ such that the following is true for all $a \in [0,1)$ and $n \geq 1$. Let $X_1, \dots, X_n$ be random variables taking values in $[0,1]$ and let $m_1, \dots, m_n \geq 0$ be such that we have almost surely $\mathbb{E}[X_i | X_1, \dots, X_{i-1}] \geq m_i$ for $1 \leq i \leq n$. Suppose that $\sum_{i=1}^n m_i = an$. Then*

$$\log \mathbb{P}\left[X_1 + \ldots + X_n \leq \frac{1}{2}na\right] \leq -cna.$$

We generalise Lemma 3.13 to higher dimensions.

Lemma 3.14. *There is some absolute constant $c > 0$ such that the following is true. Suppose that $X_1, \dots, X_n$ are random $d \times d$ symmetric positive semi-definite matrices such that $X_i \leq bI$ for some $b > 0$ and*

$$\mathbb{E}[X_i | X_1, \dots, X_{i-1}] \geq m_i I.$$

Suppose that $\sum_{i=1}^n m_i = an$. Then there is some constant $C = C(a/b, d)$ depending only on a/b and d such that

$$\log \mathbb{P}\left[X_1 + \dots + X_n \leq \frac{na}{4} I\right] \leq -can + C$$

Here we are using the partial ordering (2.16).

Proof. For convenience write $Y_n = X_1 + \ldots + X_n$ and choose a set S of unit vectors in $\mathbb{R}^d$ such that if y is any unit vector in $\mathbb{R}^d$ then there exists $x \in S$ with $\|x - y\| \leq \frac{a}{8b}$. Note that the size of S depends only on d and a/b.

By Lemma 3.13 we know that for any $x \in S$,

$$\log \mathbb{P}\left[x^T Y_n x \leq \frac{na}{2}\right] \leq -can.$$

Let A be the event that there exists some $x \in S$ with $x^T Y_n x \leq \frac{na}{2}$. We have that $\log \mathbb{P}[A]$ is at most $-can + \log |S|$. It suffices therefore to show that on A^C we have $Y_n \geq \frac{na}{4} I$.

Indeed let $y \in \mathbb{R}^d$ be a unit vector. Choose some $x \in \mathbb{R}^d$ with $\|x - y\| \leq a/8b$. Suppose that A^C occurs. Note that we must have $Y_n \leq bnI$ and therefore $||Y_n|| \leq bn$. This means

$$\begin{aligned} y^T Y_n y &= x Y_n x + x^T Y_n (y - x) + (y - x)^T Y_n y \\ &> \frac{an}{2} - 2bn \cdot \frac{a}{8b} = \frac{an}{4}. \end{aligned}$$

and result follows. □

4. Polynomial Decay of Self-Similar Measures

In this section we prove Theorem 1.2, which we generalise to arbitrary complete metric spaces and also deal with measures that are not necessarily finitely supported. While existence and uniqueness of the self-similar measure ν is known, we include a short proof as the argument is needed to establish the polynomial decay of ν. The latter follows from first showing that ν is approximated very well by $\mu^{*n} * \delta_x$ and then applying the large deviation principle. For only contractive on average self-similar measures we show in section 4.2 a polynomial lower bound on $\nu(\{|x| \geq R\})$.

4.1. **Existence, Uniqueness and Polynomial Decay of Self-similar Measures.**

Let X be a complete metric space and for a map $g : X \to X$ denote by $\rho(g)$ the Lipschitz constant of ρ defined as

$$\rho(g) = \min\{\rho > 0 \,:\, d(gx, gy) \leq \rho \cdot d(x, y) \text{ for all } x, y \in X\}.$$

Consider the semi-group $S(X) = \{g : X \to X \text{ such that } \rho(g) < \infty\}$ endowed with the compact open topology. We note that ρ may not be continuous, for example when $X = \mathbb{R}$. Let μ be a probability measure on $S(X)$ such that $\log \circ \rho \in L^1(\mu)$. Then the Lyapunov exponent

$$\chi_\mu = \mathbb{E}_{g\sim\mu}[-\log \rho(\mu)] = \int \log \rho(g)\, d\mu(g)$$

exists and is finite. We denote for $x \in X$ and $R > 0$ by $B_R(x) = \{y \in X : d(x, y) < R\}$ the R-ball around x.

Theorem 4.1. *(Generalisation of Theorem 1.2) Let X be a complete metric space and let μ be a compactly supported probability measure on $S(X)$ with $\log \circ \rho \in L^1(\mu)$ and assume that $\chi_\mu < 0$. Then there exists a unique probability measure ν on X such that $\mu * \nu = \nu$. Moreover, if ρ is bounded on $\mathrm{supp}(\mu)$, there is $\alpha = \alpha(\mu) > 0$ such that for all $R > 0$ and $x \in X$,*

$$\nu(B_R(x)^c) \ll_{\mu,x} R^{-\alpha}. \tag{4.1}$$

For notational convenience write

$$A_x := \sup_{\gamma\in\mathrm{supp}(\mu)} d(x, \gamma x),$$

which is finite since μ is compactly supported. To show that there exists a unique stationary measure ν, we can drop the assumption of μ being compactly supported as long as A_x is finite for every $x \in X$.

Proof. (of existence and uniqueness in Theorem 4.1) To be explicit, assume that $\gamma_1, \gamma_2, \dots$ are sampled from the probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Observe that by the large deviation principle (Lemma 3.10) and Borel-Cantelli, there is $\lambda \in (0, 1)$ such that for almost all $\omega \in \Omega$ and n sufficiently large (depending on ω),

$$\rho(\gamma_1(\omega) \cdots \gamma_n(\omega)) \le \rho(\gamma_1(\omega)) \cdots \rho(\gamma_n(\omega)) \le \lambda^n.$$

Given $x \in X$ and $\omega \in \Omega$ write

$$z_n(x, \omega) = \gamma_1(\omega) \cdots \gamma_n(\omega)x.$$

Then almost surely, $z_n(x, \omega)$ is a Cauchy sequence. Indeed, almost surely for k sufficiently large $d(z_k(x, \omega), z_{k+1}(x, \omega)) \le A_x\lambda^k$ and thus for sufficiently large n and m,

$$\begin{aligned} d(z_n(x, \omega), z_m(x, \omega)) &\le \sum_{k=\min\{n,m\}}^{\max\{n,m\}-1} d(z_k(x, \omega), z_{k+1}(x, \omega)) \\ &\le A_x \sum_{k=\min\{n,m\}}^{\infty} \lambda^k = A_x \frac{\lambda^{\min\{n,m\}}}{1-\lambda}, \end{aligned} \tag{4.2}$$

which goes to zero as $\min\{n, m\} \to \infty$. Therefore, since X is complete, the limit $\lim_{n\to\infty} z_n(x, \omega)$ exists for almost all $\omega \in \Omega$. The latter limit does not depend on x as for almost all ω and sufficiently large n, $d(z_n(x, \omega), z_n(y, \omega)) \le \lambda^n d(x, y)$, which goes to zero. Thus there is a random variable $z : \Omega \to X$ such that for almost all ω,

$$z(\omega) = \lim_{n\to\infty} z_n(x, \omega) = \lim_{n\to\infty} \gamma_1(\omega) \cdots \gamma_n(\omega)x$$

for all $x \in X$.

Let ν be the distribution of z. The measure ν is stationary since for any continuous bounded function f on X, by dominated convergence,

$$\begin{aligned}(\mu * \nu)(f) &= \int\int f(gz)\, d\mu(g)d\nu(z)\\ &= \int\int f(gz(\omega))\, d\mu(g)d\mathbb{P}(\omega)\\ &= \lim_{n\to\infty}\int\int f(gz_n(x,\omega))\, d\mu(g)d\mathbb{P}(\omega)\\ &= \lim_{n\to\infty}\int f(z_{n+1}(x,\omega))\, d\mathbb{P}(\omega) = \nu(f)\end{aligned}$$

for any $x \in X$.

To show that the stationary measure is unique, let η be a further stationary measure. Then for all $n \geq 1$,

$$\int f(x)\, d\eta(x) = \int\int f(gx)\, d\mu^{*n}(g)d\eta(x) = \int\int f(z_n(x,\omega))\, d\mathbb{P}(\omega)d\eta(x).$$

Letting $n \to \infty$, the right hand side tends to $\nu(f)$ by dominated convergence. □

To complete the proof of Theorem 4.1 it remains to show the estimate (4.1). To do so, we establish that $\mu^{*n} * \delta_x$ converges to ν exponentially fast. For a function $f : X \to \mathbb{R}$, we denote by $||f||_\infty = \sup_{x\in X}|f(x)|$ and by $\mathrm{Lip}(f) = \sup_{x,y\in X}\frac{|f(x)-f(y)|}{d(x,y)}$ the Lipschitz constant of f.

Lemma 4.2. *Let X, μ, ν and A_x be as in Theorem 4.1. Then for a bounded Lipschitz function $f : X \to \mathbb{R}$ and $n \geq 1$,*

$$\left|\int f(z)\, d\nu(z) - \int f(gx)\, d\mu^{*n}(g)\right| \ll_\mu A_x \max\{||f||_\infty, \mathrm{Lip}(f)\}e^{-cn}.$$

Proof. We continue with the notation from the proof of Theorem 4.1. Denote by F_n the event that $\{\rho(\gamma_1)\cdots\rho(\gamma_n) \leq \lambda^n\}$ and by $E_n = \bigcap_{k=n}^{\infty} F_k$. Then by Lemma 3.10, $\mathbb{P}[E_n^c] \ll_\mu e^{-\delta n}$ for some $\delta > 0$. We have shown in (4.2) that for $\omega \in E_n$,

$$d(z(\omega), z_n(x,\omega)) \leq \frac{A_x}{1-\lambda}\lambda^n.$$

To conclude,

$$\begin{aligned}\left|\int f(z)\, d\nu(z) - \int f(gx)\, d\mu^{*n}\right| &\leq \int |f(z) - f(z_n(x,\omega))|\, d\mathbb{P}(\omega)\\ &= \int_{E_n} |f(z) - f(z_n(x,\omega))|\, d\mathbb{P}(\omega)\\ &\quad + \int_{E_n^c} |f(z) - f(z_n(x,\omega))|\, d\mathbb{P}(\omega)\\ &\ll_\mu A_x\lambda^n \mathrm{Lip}(f) + 2\mathbb{P}[E_n^c]||f||_\infty,\end{aligned}$$

showing the claim for a sufficiently small c such that $\max\{e^{-\delta}, \lambda\} \leq e^{-c}$. □

Proof. (of (4.1)) Let $x \in X$. We first apply Lemma 4.2 to a suitable function. Let F_R be the function from $\mathbb{R} \to \mathbb{R}$ that is 0 on $[-R/2, R/2]$, 1 on $[-R, R]^c$ and the linear interpolation between these intervals. Then we consider the function on X defined for $y \in X$ as

$$f_{R,x}(y) = F_R(d(y,x)).$$

Note that $f_{R,x}$ is Lipschitz with $\mathrm{Lip}(f_{R,x}) \leq 2R^{-1}$. Thus by applying Lemma 4.2 for $n \geq 1$,

$$\nu(B_R(x)^c) \leq \int f_{R,x}\, d\nu \leq (\mu^{*n} * \delta_x)(B_{R/2}(x)^c) + O_\mu(A_x R^{-1} e^{-cn}). \tag{4.3}$$

We next give a suitable bound for $(\mu^{*n} * \delta_x)(B_{R/2}(x)^c)$. If $\gamma_1, \dots, \gamma_n \in \mathrm{supp}(\mu)$ then

$$\begin{aligned} d(x, \gamma_1 \cdots \gamma_n x) &\leq \sum_{i=1}^{n} d(\gamma_1 \cdots \gamma_{i-1} x, \gamma_1 \cdots \gamma_i x) \\ &\leq A_x(1 + \rho(\gamma_1) + \ldots + \rho(\gamma_1 \cdots \gamma_{n-1})). \end{aligned}$$

Let $\varepsilon > 0$ be a fixed constant. Then note that

$$\begin{aligned} (\mu^{*n} * \delta_x)(B_{R/2}(x)^c) &= \int 1_{B_{R/2}(x)^c}(z_n(x,\omega))\, d\mathbb{P}(\omega) \\ &\leq \int_{E_{\varepsilon n}} 1_{B_{R/2}(x)^c}(z_n(x,\omega))\, d\mathbb{P}(\omega) + \mathbb{P}[E_{\varepsilon n}^c] \end{aligned}$$

Note that if $\omega \in E_{\varepsilon n}$ then for all $m \geq \varepsilon n$ it holds that $\rho(g_1 \cdots g_m) \leq \lambda^n$. Thus for such an ω and $\rho_{\sup} = \sup_{g \in \mathrm{supp}(\mu)} \rho(g)$,

$$\begin{aligned} d(x, z_n(x,\omega)) &\leq A_x(1 + \rho(\gamma_1) + \ldots + \rho(\gamma_1 \cdots \gamma_{\lfloor n\varepsilon \rfloor - 1}) + \lambda^{\lfloor n\varepsilon \rfloor} + \ldots \lambda^n) \\ &\leq A_x \left(\frac{1}{1-\lambda} + n\varepsilon \max\{1, \rho_{\sup}\}^{n\varepsilon} \right) \\ &\leq D_1 A_x (1 + D_2^{n\varepsilon}) \end{aligned}$$

for suitably large constants D_1 and D_2 depending on μ and sufficiently large n. Choosing n such that $4D_1 A_x(1 + D_2^{n\varepsilon}) \leq R \leq 4D_1 A_x(1 + D_2^{(n+1)\varepsilon})$ or equivalently $n \asymp_\mu \frac{1}{\varepsilon} \log R$ it therefore follows that

$$(\mu^{*n} * \delta_x)(B_{R/2}(x)^c) \leq \mathbb{P}[E_{\varepsilon n}^c] \ll_\mu e^{-\delta\varepsilon n}.$$

Combining the latter with (4.3), we conclude that

$$\nu(B_R(x)^c) \ll_\mu A_x R^{-1} e^{-cn} + e^{-\delta\varepsilon n} \ll_\mu A_x R^{-1} R^{-O(c/\varepsilon)} + R^{-O(\delta)} \ll_{\mu,x} R^{-\alpha}$$

for a suitable constant $\alpha > 0$. □

4.2. **Polynomial Lower Bound for Partially Expanding Self-Similar Measures.** Returning to self-similar measures on $\mathbb{R}^d$, we show a lower bound on $\nu(B_R^c)$ if μ is only contracting on average and not supported on contractions with a common fixed point. On the other hand, if μ is not only contracting on average, then the support of ν may be compact or non-compact. Indeed if for example $\mu = \frac{1}{2}(\delta_{g_1} + \delta_{g_2})$ with $g_1(x) = \frac{1}{2}x + 1$ and $g_2(x) = -x$, then ν has compact support. On the other hand, when we change g_2 to $g_2(x) = x + 1$, the support of ν is non-compact.

Lemma 4.3. *Let μ be a only contractive on average probability measure on $\mathrm{Sim}(\mathbb{R}^d)$ that is supported on finitely many similarities without a common fixed point. Then there is $\alpha_1 = \alpha_1(\mu) > 0$ such that for $x \in \mathbb{R}^d$,*

$$\nu(B_R(x)^c) \gg_{\mu,x} R^{-\alpha_1}. \tag{4.4}$$

Proof. We note that it suffices to prove the claim for a fixed $x \in \mathbb{R}^d$. Write $\mu = \sum_g p_g \delta_g$ and let $g_0 \in \mathrm{supp}(\mu)$ be a map with $\rho(g_0) > 1$ and $p_{g_0} > 0$. For convenience write $\rho_0 = \rho(g_0)$, $p_0 = p_{g_0}$ and denote by x_0 the unique fixed point of g_0, which exists since g_0^{-1} is contractive and has a unique fixed point. Since the support of ν contains at least two points, we may choose $r > 0$ such that $\nu(B_r(x_0)^c) > 0$ and note that $g_0^n B_r(x_0)^c \subset B_{\rho_0^n r}(x_0)^c$. Therefore

$$\begin{aligned}\nu(B_{\rho_0^n r}(x_0)^c) &= (\mu^{*n} * \nu)(B_{\rho_0^n r}(x_0)^c)) \\ &\geq (p_0^n \delta_{g_0^n} * \nu)(B_{\rho_0^n r}(x_0)^c) \\ &\geq p_0^n \nu(g_0^{-1} B_{\rho_0^n r}(x_0)^c) \\ &\geq p_0^n \nu(B_r(x_0)^c).\end{aligned}$$

To prove the claim, let $R \geq 1$ be such that $\rho_0^n r \leq R \leq \rho_0^{n+1} r$ for an integer $n \geq 0$. In particular $\log R \leq 2n \log \rho_0$ for sufficiently large R. Setting $\alpha_1 = -\frac{\log p_0}{2 \log \rho_0} > 0$,

$$\nu(B_R(x_0)^c) \geq \nu(B_{\rho_0^n r}(x_0)^c) \gg_\mu p_0^n = e^{(\log p_0)n} \gg_\mu e^{-\alpha_1 \cdot (2n \log \rho_0)} = R^{-\alpha_1}.$$

□

5. Order k Detail

The goal of this section is to prove the product bound (2.4) and to show how to convert (2.2) into suitable estimates for detail. We first recall in section 5.1 the definition of the detail $s_r(\lambda)$ of a measure λ on $\mathbb{R}^d$ at scale $r > 0$ that was first introduced by [Kit21]. We then expand the definition and results of order k detail $s_r^{(k)}(\lambda)$ of a measure from [Kit23] to measures of $\mathbb{R}^d$.

As mentioned in the outline of proofs, the advantage of using k-order detail over detail is that it leads to stronger product bounds. Indeed, we will show in Lemma 5.3 that

$$s_r^{(k)}(\lambda_1 * \cdots * \lambda_k) \leq s_r(\lambda_1) \cdots s_r(\lambda_k) \tag{5.1}$$

for measures $\lambda_1, \ldots, \lambda_k$ on $\mathbb{R}^d$ and $r > 0$. Moreover, if $s_r^{(k)}(\lambda) \leq \alpha$ for all $r \in [a, b]$ and some $k \geq 1$ then we show in Proposition 5.5 for a constant $Q'(d)$ depending only on d that

$$s_{a\sqrt{k}}(\lambda) \leq Q'(d)(\alpha + k!ka^2b^{-2}). \tag{5.2}$$

Combining (5.1) and (5.2), we deduce the strong product bound (Corollary 5.6) mentioned at (2.4) in the outline of proofs.

In section 5.3, we show that the difference in the detail of two measures is bounded in term of their Wasserstein distance. Finally, in section 5.4 we show how to convert the conditions from (2.2) into good estimates for detail. The latter requires Berry-Essen type results, the Wasserstein distance bounds from section 5.3, (5.1) and a suitable partition of $\sum_i X_i$.

All of these results will be used in section 8.

5.1. **Definitions.** Denote by η_y the standard Gaussian density on $\mathbb{R}^d$ with covariance matrix $y \cdot \mathrm{I}_d$, i.e.

$$\eta_y(x) = \frac{1}{(2\pi y)^{d/2}} \exp\left(-\frac{||x||^2}{2y}\right).$$

Moreover, we write

$$\eta_y^{(1)} = \frac{\partial}{\partial y}\eta_y.$$

Given a probability measure λ on $\mathbb{R}^d$ the detail of λ at scale $r > 0$ is defined as

$$s_r(\lambda) = r^2\, Q(d)\, ||\lambda * \eta_{r^2}^{(1)}||_1,$$

where $Q(d) = ||\eta_1^{(1)}||^{-1} = \frac{1}{2}\Gamma(\frac{d}{2})(\frac{d}{2e})^{-d/2}$ and note that by Stirling's approximation $d^{-1/2} \leq Q(d) \leq ed^{-1/2}$ for all $d \geq 1$. Moreover, $r^2 Q(d) = ||\eta_{r^2}^{(1)}||^{-1}$ and therefore $s_r(\lambda) \leq 1$ for every probability measure λ.

Proposition 5.1. [Kit21, section 2] *Let λ and μ be probability measures on $\mathbb{R}^d$. Then the following properties hold:*

(i) Suppose that there is $\beta > 1$ such that $s_r(\lambda) < (\log r^{-1})^{-\beta}$ for sufficiently small r. Then λ is absolutely continuous.
(ii) $s_r(\lambda * \mu) \leq s_r(\lambda)$.

Definition 5.2. *Given a probability measure λ on $\mathbb{R}^d$ and some $k \geq 1$ we define the **order** k **detail of** λ **at scale** r as*

$$s_r^{(k)}(\lambda) = r^{2k}\, Q(d)^k\, ||\lambda * \eta_{kr^2}^{(k)}||_1,$$

where $\eta_y^{(k)} = \frac{\partial^k}{\partial y^k}\eta_y$.

5.2. **Bounding Detail.** We have the following properties:

Lemma 5.3. *Let $k \geq 1$ and let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be probability measures on $\mathbb{R}^d$. Then*

$$s_r^{(k)}(\lambda_1 * \lambda_2 * \ldots * \lambda_k) \leq s_r(\lambda_1) s_2(\lambda_2) \cdots s_r(\lambda_k). \tag{5.3}$$

In particular, for any probability measure λ on $\mathbb{R}^d$ and $k \geq 1$,

$$s_r^{(k)}(\lambda) \leq 1. \tag{5.4}$$

Proof. Recall that by the Heat equation $\frac{\partial}{\partial y}\eta_y(x) = \frac{1}{2}\sum_{i=1}^d \frac{\partial^2}{\partial x_i^2}\eta_y(x)$ and therefore by standard properties of convolution

$$\begin{aligned}
\eta_{kr^2}^{(k)} &= \frac{1}{2^k}\sum_{i_1,\ldots,i_k=1}^{d} \frac{\partial^2}{\partial x_{i_1}^2}\cdots\frac{\partial^2}{\partial x_{i_k}^2}\eta_{kr^2} \\
&= \underbrace{\left(\frac{1}{2}\sum_{i=1}^d \frac{\partial^2}{\partial x_i^2}\eta_{r^2}\right) * \left(\frac{1}{2}\sum_{i=1}^d \frac{\partial^2}{\partial x_i^2}\eta_{r^2}\right) * \cdots * \left(\frac{1}{2}\sum_{i=1}^d \frac{\partial^2}{\partial x_i^2}\eta_{r^2}\right)}_{k \text{ times}} \\
&= \underbrace{\eta_{r^2}^{(1)} * \eta_{r^2}^{(1)} * \cdots * \eta_{r^2}^{(1)}}_{k \text{ times}}.
\end{aligned}$$

This concludes the proof of (5.3) as

$$\begin{aligned}
||\lambda_1 * \ldots * \lambda_k * \eta_{kr^2}^{(k)}||_1 &= ||\lambda_1 * \eta_{r^2}^{(1)} * \lambda_2 * \eta_{r^2}^{(1)} * \cdots * \lambda_k * \eta_{r^2}^{(1)}||_1 \\
&\leq ||\lambda_1 * \eta_{r^2}^{(1)}||_1 \cdot ||\lambda_2 * \eta_{r^2}^{(1)}||_1 \cdots ||\lambda_k * \eta_{r^2}^{(1)}||_1.
\end{aligned}$$

To show (5.4) we set $\lambda_1 = \lambda$ and $\lambda_2 = \ldots = \lambda_k = \delta_e$ and use that $s_r(\lambda_i) \leq 1$. $\square$

Lemma 5.4. *Let k be an integer greater than 1 and suppose that λ is a probability measure on $\mathbb{R}^d$. Suppose that $a, b, c > 0$ and $\alpha \in (0,1)$. Assume that $a < b$ and that for all $r \in [a, b]$ it holds that*

$$s_r^{(k)}(\lambda) \leq \alpha + cr^{2k}.$$

Then for all $r \in \left[a\sqrt{\frac{k}{k-1}}, b\sqrt{\frac{k}{k-1}}\right]$ we have

$$s_r^{(k-1)}(\lambda) \leq 2eQ(d)^{-1}\left(\alpha + (b^{-2(k-1)} + ckb^2)r^{2(k-1)}\right).$$

Proof. By the assumption and the definition of detail for $y \in [ka^2, kb^2]$ and writing $y = kr^2$,

$$||\lambda * \eta_y^{(k)}||_1 \leq r^{-2k}Q(d)^{-k}(\alpha + cr^{2k}) = \alpha y^{-k}k^kQ(d)^{-k} + cQ(d)^{-k}.$$

Therefore with $y \in [ka^2, kb^2]$,

$$\begin{aligned}
||\lambda * \eta_y^{(k-1)}||_1 &\leq ||\lambda * \eta_{kb^2}^{(k-1)}||_1 + \int_y^{kb^2} ||\lambda * \eta_u^{(k)}||_1 \, du \\
&\leq ||\eta_{kb^2}^{(k-1)}||_1 + \int_y^{kb^2} \alpha u^{-k}k^kQ(d)^{-k} + cQ(d)^{-k} \, du \\
&\leq (\tfrac{kb^2}{k-1})^{-(k-1)}Q(d)^{-(k-1)} + \alpha k^kQ(d)^{-k}\tfrac{y^{-(k-1)}}{k-1} + Q(d)^{-k}ckb^2,
\end{aligned}$$

where we bounded in the last inequality $||\eta_{kb^2}^{(k-1)}||_1$ by using that order $(k-1)$-detail is at most one, $\int_y^{kb^2} \alpha u^{-k}k^kQ(d)^{-k}\,du$ by $\int_y^{\infty} \alpha u^{-k}k^kQ(d)^{-k}\,du$ and $\int_y^{kb^2} cQ(d)^{-k}\,du$ by $\int_0^{kb^2} cQ(d)^{-k}\,du$. Using that $(\frac{k}{k-1})^{-(k-1)} < 1$ we therefore get

$$||\lambda * \eta_y^{(k-1)}||_1 \leq \alpha k^kQ(d)^{-k}\frac{y^{-(k-1)}}{k-1} + (b^{-(2k-2)} + Q(d)^{-1}ckb^2)Q(d)^{-(k-1)}.$$

Substituting the definition of order k detail gives for $y = (k-1)r^2 \in [ka^2, kb^2]$ or equivalently $r \in \left[a\sqrt{\frac{k}{k-1}}, b\sqrt{\frac{k}{k-1}}\right]$,

$$\begin{aligned}
s_r^{(k-1)}(\lambda) &= r^{2(k-1)}Q(d)^{k-1}||\lambda * \eta_{(k-1)r^2}^{(k-1)}||_1 \\
&\leq \alpha r^{2(k-1)}k^kQ(d)^{-1}\frac{((k-1)r^2)^{-(k-1)}}{k-1} + r^{2(k-1)}(b^{-2(k-1)} + Q(d)^{-1}ckb^2) \\
&\leq \alpha Q(d)^{-1}\left(1 + \frac{1}{k-1}\right)^k + (b^{-2(k-1)} + Q(d)^{-1}ckb^2)r^{2(k-1)}.
\end{aligned}$$

Finally using that $\left(1 + \frac{1}{k-1}\right)^k \leq 2e$ and that $2eQ(d)^{-1} \geq 1$ the proof is concluded. ∎

Proposition 5.5. *Let k be an integer greater than 1 and suppose that λ is a probability measure on $\mathbb{R}^d$. Suppose that $a, b > 0$ and $\alpha \in (0,1)$. Assume that $a < b$ and that for all $r \in [a, b]$ we have*

$$s_r^{(k)}(\lambda) \leq \alpha.$$

Then we have that

$$s_{a\sqrt{k}}(\lambda) \leq Q'(d)^{k-1}(\alpha + k! \cdot ka^2b^{-2})$$

for $Q'(d) = 4eQ(d)^{-1} \geq 1$.

Proof. We will show by induction for $j = k, k-1, \dots, 1$ that for all $r \in \left[a\sqrt{\frac{k}{j}}, b\sqrt{\frac{k}{j}}\right]$ we have

$$s_r^{(j)}(\lambda) \leq Q'(d)^{k-j}\left(\alpha + \frac{k!}{j!}b^{-2j}r^{2j}\right), \tag{5.5}$$

which implies the claim by setting $j = 1$ and $r = a\sqrt{k}$. The case $j = k$ follows from the conditions of the lemma. For the inductive step assume now that for all $r \in \left[a\sqrt{\frac{k}{j}}, b\sqrt{\frac{k}{j}}\right]$ we have that (5.5) holds. Then by Lemma 5.4 we have for all $r \in \left[a\sqrt{\frac{k}{j-1}}, b\sqrt{\frac{k}{j-1}}\right]$

$$\begin{aligned} s_r^{(j-1)}(\lambda) &\leq Q'(d)^{k-j} 2eQ(d)^{-1}\left(\alpha + \left(b^{-2(j-1)} + \frac{k!}{j!}b^{-2j}jb^2\right)r^{2(j-1)}\right) \\ &\leq Q'(d)^{k-j} 2eQ(d)^{-1}\left(\alpha + \left(1 + \frac{k!}{(j-1)!}\right)b^{-2(j-1)}r^{2(j-1)}\right) \\ &\leq Q'(d)^{k-(j-1)}\left(\alpha + \frac{k!}{(j-1)!}b^{-2(j-1)}r^{2(j-1)}\right). \end{aligned}$$

□

Combining Lemma 5.3 and Proposition 5.5, we arrive at the following corollary.

Corollary 5.6. *Let $k \geq 1$ and let $\lambda_1, \lambda_2, \dots, \lambda_k$ be probability measures on $\mathbb{R}^d$. Suppose that $a, b > 0$ and $\alpha \in (0, 1)$. Assume that $a < b$ and that for all $r \in [a, b]$ and $i \in [k]$ we have*

$$s_r(\lambda_i) \leq \alpha.$$

Then it holds that

$$s_{a\sqrt{k}}(\lambda) \leq Q'(d)^{k-1}(\alpha + k! \cdot ka^2b^{-2}).$$

5.3. **Wasserstein Distance.** Recall as in (2.17) that the Wasserstein 1-distance on $\mathbb{R}^d$ between λ_1 and λ_2 is defined as

$$\mathcal{W}_1(\lambda_1, \lambda_2) = \inf_{\gamma \in \Gamma(\lambda_1, \lambda_2)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y| \, d\gamma(x, y),$$

where $\Gamma(\lambda_1, \lambda_2)$ is the set of couplings between λ_1 and λ_2. We show that detail is comparable to the Wasserstein distance.

Lemma 5.7. *Let λ_1 and λ_2 be probability measures on $\mathbb{R}^d$. Then for $k \geq 1$ and $r > 0$,*

$$|s_r^{(k)}(\lambda_1) - s_r^{(k)}(\lambda_2)| \leq edr^{-1}\mathcal{W}_1(\lambda_1, \lambda_2),$$

where e is Euler's number.

Proof. Let X and Y be random variables with laws λ_1 and λ_2 respectively. Then

$$(\lambda_1 - \lambda_2) * \eta_{kr}^{(k)}(v) = \mathbb{E}\left[\eta_{kr}^{(k)}(v - X) - \eta_{kr}^{(k)}(v - Y)\right]$$

and therefore

$$\left|(\lambda_1 - \lambda_2) * \eta_{kr}^{(k)}(v)\right| \leq \mathbb{E}\left[\left|\eta_{kr}^{(k)}(v - X) - \eta_{kr}^{(k)}(v - Y)\right|\right].$$

Note that

$$\left|\eta_{kr}^{(k)}(v - X) - \eta_{kr}^{(k)}(v - Y)\right| \leq \int_X^Y \left|\nabla \eta_{kr}^{(k)}(v - u)\right| |du|,$$

where $\int_x^y \cdot |du|$ is understood to be the integral along the shortest path between x and y and ∇ is the gradient. Thus

$$\begin{aligned} ||(\lambda_1 - \lambda_2) * \eta_{kr}^{(k)}||_1 &\leq \int_{\mathbb{R}^d} \mathbb{E}\left[\int_X^Y \left|\nabla \eta_{kr}^{(k)}(v-u)\right| |du|\right] dv \\ &= \mathbb{E}\left[\int_X^Y \int_{\mathbb{R}^d} \left|\nabla \eta_{kr}^{(k)}(v-u)\right| dv \, |du|\right] \\ &= ||\nabla \eta_{kr}^{(k)}||_1 \mathbb{E}[|X-Y|] \\ &\leq \left(\sum_{i=1}^d \left|\left|\frac{\partial}{\partial x_i}\eta_{kr}^{(k)}\right|\right|_1\right) \mathbb{E}[|X-Y|] \end{aligned}$$

We next bound $||\frac{\partial}{\partial x_i}\eta_{kr}^{(k)}||_1$. As in the proof of Lemma 5.3, it follows that

$$\frac{\partial}{\partial x_i}\eta_{kr^2}^{(k)} = \left(\frac{\partial}{\partial x_i}\eta_{\frac{k}{k+1}r^2}\right) * \underbrace{\eta_{\frac{k}{k+1}r^2}^{(1)} * \dots * \eta_{\frac{k}{k+1}r^2}^{(1)}}_{k \text{ times}}.$$

Using standard properties of Gaussian integrals,

$$\left|\left|\frac{\partial}{\partial x_i}\eta_{\frac{k}{k+1}r^2}\right|\right|_1 = \sqrt{\frac{2(k+1)}{k\pi}} r^{-1} \leq \sqrt{\frac{k+1}{k}} r^{-1}$$

and therefore

$$\begin{aligned} \left|\left|\frac{\partial}{\partial x_i}\eta_{kr}^{(k)}\right|\right|_1 &\leq \left|\left|\frac{\partial}{\partial x_i}\eta_{\frac{k}{k+1}r^2}\right|\right|_1 \cdot ||\eta_{\frac{k}{k+1}r^2}^{(1)}||_1^k \\ &\leq \left(\frac{k+1}{k}\right)^{(k+1)/2} Q(d)^{-k} r^{-2k-1}. \end{aligned}$$

Using that $\left(\frac{k+1}{k}\right)^{(k+1)/2} \leq e$, we conclude

$$\begin{aligned} |s_r^{(k)}(\lambda_1) - s_r^{(k)}(\lambda_2)| &\leq r^{2k} Q(d)^k ||(\lambda_1 - \lambda_2) * \eta_{kr}^{(k)}||_1 \\ &\leq der^{-1}\mathbb{E}[|X-Y|]. \end{aligned}$$

Choosing a coupling for X and Y which minimizes $\mathbb{E}[|X-Y|]$ gives the required result. □

5.4. **Small Random Variables Bound in $\mathbb{R}^d$.** The aim of this subsection is to show that the sum of independent random variables in $\mathbb{R}^d$ have small detail whenever they are supported close to 0 and have sufficiently large variance. To state our result, we use the partial order (2.16) for positive semi-definite symmetric matrices.

Proposition 5.8. *For every positive integer $d \geq 1$ and every $\alpha > 0$ there exists some $C = C(\alpha, d) > 0$ such that the following is true for all $r > 0$ and positive integers k. Let $X_1, X_2, \dots, X_n$ be independent random variables taking values in $\mathbb{R}^d$ such that almost surely*

$$|X_i| \leq C^{-1} r \quad \textit{and} \quad \sum_{i=1}^n \operatorname{Var} X_i \geq Ckr^2 I.$$

Then

$$s_r^{(k)}(X_1 + \ldots + X_n) \le \alpha^k.$$

Proposition 5.8 relies on a higher dimensional Berry-Essen type result, which implies Proposition 5.8 for $k = 1$, as deduced in Lemma 5.11. To prove the higher dimensional Berry-Essen type result we first need the following.

Theorem 5.9. *Let $X_1, X_2, \ldots, X_n$ be independent random variables taking values in $\mathbb{R}$ with mean 0 and for each $i \in [n]$ let $\mathbb{E}[X_i^2] = \omega_i^2$ and $\mathbb{E}[|X_i|^3] = \gamma_i^3 < \infty$. Let $\omega^2 = \sum_{i=1}^n \omega_i^2$ and let $S = X_1 + \cdots + X_n$. Let N be a normal distribution with mean 0 and variance ω^2. Then for an absolute implied constant*

$$\mathcal{W}_1(S, N) \ll \frac{\sum_{i=1}^n \gamma_i^3}{\sum_{i=1}^n \omega_i^2}.$$

Proof. A proof of this result may be found in [Eri73]. □

From this we may deduce the following higher dimensional Berry-Essen type result.

Lemma 5.10. *Let $X_1, X_2, \ldots, X_n$ be independent random variables taking values in $\mathbb{R}^d$ with mean 0 and denote for each $i \in [n]$ write*

$$\Sigma_i = \operatorname{Var} X_i.$$

Suppose that $\delta > 0$ is such that for each $i \in [n]$ we have $|X_i| \le \delta$ almost surely. Let $\Sigma = \sum_{i=1}^n \Sigma_i$ and $S = X_1 + \ldots + X_n$. Let N be a multivariate normal distribution with mean 0 and covariance matrix Σ. Then

$$\mathcal{W}_1(S, N) \ll_d \delta.$$

Proof. First we will deduce this from Theorem 5.9 in the case $d = 1$. In this case simply note that

$$\sum_{i=1}^n \xi_i^3 = \sum_{i=1}^n \mathbb{E}[|X_i|^3] \le \sum_{i=1}^n \mathbb{E}[\delta |X_i|^2] = \delta \omega^2.$$

The result follows.

Now in the case $d \ge 1$ the result follows by looking at the projection of S and N onto each of the d coordinate axis. □

Lemma 5.11. *For every positive integer $d \ge 1$ and every $\alpha > 0$ there exists some $C = C(\alpha, d) > 0$ such that the following is true. Let $r > 0$ and let $X_1, X_2, \ldots, X_n$ be independent random variables taking values in $\mathbb{R}^d$ such that*

$$|X_i| \le C^{-1} r \quad \text{and} \quad \sum_{i=1}^n \operatorname{Var} X_i \ge C r^2 I.$$

Then

$$s_r(X_1 + \ldots + X_n) \le \alpha.$$

Proof. Denote for $1 \le i \le n$ by $X_i' = X_i - \mathbb{E}[X_i]$ and let $S' = \sum_{i=1}^n X_i'$. Note that $s_r(\sum_{i=1}^n X_i) = s_r(S')$. Write $\Sigma_i = \operatorname{Var} X_i$ and let $\Sigma = \sum_{i=1}^n \Sigma_i$. Let N be a multivariate normal distribution with mean 0 and covariance matrix Σ. Note that $|X_i'| \le 2C^{-1} r$ almost surely. Therefore by Lemma 5.10,

$$\mathcal{W}_1(S', N) \ll_d C^{-1} r.$$

Also

$$s_r(N) \leq s_r(\eta_{C^2r^2}) = \frac{\|\eta^{(1)}_{C^2r^2+r^2}\|}{\|\eta^{(1)}_{r^2}\|} = \frac{1}{C^2+1}.$$

Thus by Lemma 5.7,

$$s_r(X_1 + \ldots + X_n) = s_r(S') \ll_d C^{-1} + \frac{1}{1+C^2},$$

implying the claim. □

The proof of Proposition 5.8 in the case $d \geq 2$ is more involved than the proof in the case $d = 1$. In order to prove this proposition we also need the following lemma and a corollary of it.

Lemma 5.12. *Let V be a Euclidean vector space, let $v_1, \ldots, v_n \in V$ and write $S = v_1 + \cdots + v_n$. Let $c_1, c_2 > 0$ be such that for all $i \in [n]$ we have*

$$|v_i| \leq c_1 \quad \text{and} \quad v_i \cdot S \geq c_2 |v_i||S|.$$

Let k be a positive integer. Then we can partition $[n]$ as $J_1 \sqcup J_2 \sqcup \cdots \sqcup J_k$ such that for each $j \in [k]$ we have

$$|S_j - \tfrac{1}{k}S| < c_2^{-1}\sqrt{\tfrac{2c_1}{k}|S|} + 2c_2^{-2}c_1$$

where $S_j = \sum_{i \in J_j} v_i$.

Proof. Choose the J_j such that

$$\sum_{j=1}^{k} |S_j|^2 \tag{5.6}$$

is minimized. For each $i \in [n]$ let $j(i)$ denote the unique $j \in [k]$ such that $i \in J_j$. For each $i \in [n]$ and $j' \in [k]$ we know that moving i from $J_{j(i)}$ to $J_{j'}$ cannot decrease the sum in (5.6). Therefore

$$|S_{j(i)} - v_i|^2 + |S_{j'} + v_i|^2 \geq |S_{j(i)}|^2 + |S_{j'}|^2.$$

Expanding this out and cancelling gives

$$S_{j(i)} \cdot v_i - |v_i|^2 \leq S_{j'} \cdot v_i$$

and summing over all $i \in J_j$, we get

$$S_j \cdot S_j \leq S_j \cdot S_{j'} + \sum_{i \in J_j} |v_i|^2.$$

Let A_j denote $\sum_{i \in J_j} |v_i|^2$. Note that the above equation gives $|S_j - S_{j'}|^2 \leq A_j + A_{j'}$ and so

$$|S_j - \tfrac{1}{k}S| \leq \max_{j' \in [k]} |S_j - S_{j'}| \leq \sqrt{2 \max_{j' \in [k]} A_{j'}}. \tag{5.7}$$

Now let $\Lambda^2 = \max_{j' \in [k]} A_{j'}$. We compute

$$\begin{aligned}\sum_{i \in J_j} |v_i|^2 &\leq c_2^{-2}|S|^{-2} \sum_{i \in J_j} (v_i \cdot S)^2 \\ &\leq c_2^{-2}|S|^{-2} \sum_{i \in J_j} (v_i \cdot S) c_1 |S| \\ &= c_2^{-2} c_1 |S|^{-1} S \cdot S_j \leq c_2^{-2} c_1 |S_j| \leq c_2^{-2} c_1 (\tfrac{1}{k}|S| + \sqrt{2}\Lambda).\end{aligned}$$

Therefore $\Lambda^2 \leq c_2^{-2} c_1 (|S|/k + \sqrt{2}\Lambda)$, which gives

$$\left(\Lambda - c_2^{-2} c_1/\sqrt{2}\right)^2 \leq c_2^{-2} c_1 |S|/k + c_2^{-4} c_1^2/2$$

and so

$$\begin{aligned}\Lambda &\leq \sqrt{\frac{c_2^{-2} c_1 |S|}{k} + \frac{c_2^{-4} c_1^2}{2}} + \frac{c_2^{-2} c_1}{\sqrt{2}} \\ &\leq c_2^{-1} \sqrt{\tfrac{c_1}{k}|S|} + c_2^{-2} c_1 \sqrt{2},\end{aligned}$$

showing the required result by (5.7). □

Corollary 5.13. *Let $A_1, \ldots, A_n$ be symmetric positive semi-definite $d \times d$ matrices. Suppose that $\sum_{i=1}^n A_i \geq CkI$ and that for each $i \in [n]$ we have $\|A_i\| \leq c$. Then we can partition $[n]$ as $J_1 \sqcup J_2 \sqcup \cdots \sqcup J_k$ such that for each $j \in [k]$ we have*

$$\sum_{i \in J_j} A_i \geq \left(C - d\sqrt{2cC} - 2d^{3/2} c\right) I.$$

Proof. Let $M = \sum_{i=1}^n A_i$. We know that M is symmetric positive semi-definite and so it may be diagonalised as $M = P^{-1} D P$ for some orthogonal matrix P and a diagonal matrix D with non-zero real entries. Since $M \geq CkI$ all of the diagonal entries of D are at least Ck. Let $D' = \sqrt{CkD^{-1}}$ be a diagonal matrix and for each $i \in [n]$ let $A_i' = Q A_i Q$ where $Q = P^{-1} D' P$. Note that A_i' is symmetric positive semi-definite, $\|A_i'\| \leq c$ as $\|Q\| \leq 1$ and that $\sum_{i=1}^n A_i' = CkI$ since

$$QMQ = (P^{-1} D' P)(P^{-1} D P)(P^{-1} D' P) = P^{-1} D' D D' P = CkI.$$

We now apply Lemma 5.13 with V being the space of symmetric $d \times d$ matrices with inner product given by $A \cdot B = \sum_{x=1}^n \sum_{y=1}^n A_{xy} B_{xy} = \operatorname{tr} AB$ and with $v_1, \ldots, v_n$ being $A_1', \ldots, A_n'$. We will denote the norm induced by this inner product by $|\cdot|$. Note that given a symmetric matrix A we have that $|A|^2$ is equal to the sum of the squares of the eigenvalues of A and so in particular $\|\cdot\| \leq |\cdot| \leq \sqrt{d}\|\cdot\|$. This means that we can take $c_1 = \sqrt{d}c$ so that $|A_1'| \leq c_1$.

All that we need to do is find some lower bound on $A_i' \cdot CkI$ in terms of $|A_i'| \cdot |CkI|$. Note that $\operatorname{tr} A_i'$ is equal to the sum of the eigenvalues of A_i' and that $|A_i'|^2$ is equal to the sum of the squares of these eigenvalues. In particular since the eigenvalues are non-negative $\operatorname{tr} A_i' \geq |A_i'|$ and so

$$A_i' \cdot CkI = Ck \operatorname{tr} A_i' \geq Ck|A_i'| = |A_i'| \cdot |CkI|/\sqrt{d}.$$

This means that we can take $c_2 = 1/\sqrt{d}$.

We now apply Lemma 5.13 with $S = \sum_{i=1}^{n} A_i' = CkI$ to construct our partition $[n] = J_1 \sqcup J_2 \sqcup \dots \sqcup J_k$ such that for all $j \in [k]$,

$$\left\| \sum_{i \in J_j} A_i' - CI \right\| \leq \left| \sum_{i \in J_j} A_i' - CI \right| \leq d\sqrt{2cC} + 2d^{3/2}c.$$

Therefore

$$\left\| \sum_{i \in J_j} A_i - CIQ^{-2} \right\| \leq (d\sqrt{2cC} + 2d^{3/2}c)||Q^{-2}||$$

and hence,

$$\begin{aligned}\sum_{i \in J_j} A_i &\geq CIQ^{-2} - \left(d\sqrt{2cC} + 2d^{3/2}c\right)||Q^{-2}||I \\ &\left(C - d\sqrt{2cC} - 2d^{3/2}c\right)||Q^{-2}||I \\ &\geq \left(C - d\sqrt{2cC} - 2d^{3/2}c\right) I\end{aligned}$$

using that $||Q^{-1}|| \geq 1$ in the last line. □

Finally we can prove Proposition 5.8.

Proof of Proposition 5.8. Note that since $|X_i| \leq C^{-1}r$ almost surely we have $\| \operatorname{Var} X_i \| \leq C^{-2}r^2$. By Corollary 5.13 we can partition $[n]$ as $J_1 \sqcup J_2 \sqcup \dots \sqcup J_k$ such that for each $j \in [k]$ we have

$$\sum_{i \in J_j} \operatorname{Var} X_i \geq \left(C - d\sqrt{2C^{-1}} - 2d^{3/2}C^{-2}\right) r^2 I.$$

This means that by Lemma 5.11, provided that C is sufficiently large in terms of d, we know that

$$s_r\left(\sum_{i \in J_j} X_i\right) \leq \alpha.$$

The result now follows from Proposition 5.3. □

6. Entropy and Variance on General Lie groups

Throughout this section let G be an arbitrary Lie group of dimension ℓ with a fixed choice of Haar measure m_G and let $\mathfrak{g}$ be the Lie algebra of G. We fix an inner product on $\mathfrak{g}$, inducing an associated norm $| \circ |$. Also denote by

$$\log : G \to \mathfrak{g}$$

the logarithm on G, which is defined in a small neighbourhood around the identity.

We study entropy on arbitrary Lie groups. As exposed in the outline of proofs, we shall convert entropy estimates of a random variable Z to estimates of the variance of Z. Indeed, recall that if Z is an absolutely continuous random variable on $\mathbb{R}$ with variance σ^2 then

$$H(Z) \leq \frac{1}{2}\log(2\pi e\sigma^2), \tag{6.1}$$

where $H(Z)$ is the differential entropy of Z and equality holds in (6.1) if and only if Z is distributed like a Gaussian with variance σ^2. We will prove an analogue of this fact on Lie groups. To do so, for random variables g that are supported within small balls of a given point g_0 we consider the covariance matrix of the Lie group logarithm applied to $g_0^{-1}g$. This viewpoint allows us to apply a higher dimensional analogue of (6.1) to deduce an analogous result on G.

Indeed, we recall that for an ℓ-dimensional random variable X we denote by $\mathrm{tr}(X)$ the trace of the covariance matrix of X. In particular, we use the following definition. Given $g_0 \in G$ and a random variable g on G we define

$$\mathrm{tr}_{g_0}(g) = \mathrm{tr}(\log(g_0^{-1}g)),$$

whenever $\log(g_0^{-1}g)$ is defined. The analogue of (6.1), which will be proved in Proposition 6.8, then amounts to

$$H(g) \leq \frac{\ell}{2}\log\left(\frac{2\pi e}{\ell}\cdot \mathrm{tr}_{g_0}(g)\right) + O_G(\varepsilon) \tag{6.2}$$

for random variables supported on $B_\varepsilon(g_0)$ and $\varepsilon > 0$ sufficiently small.

A further goal of this section is to study entropy between scales on G. Indeed, we will define in section 6.5 an explicit family of smoothing distributions $s_{r,a}$ on G, which satisfy

$$\mathrm{tr}_e(s_{r,a}) \approx \ell r^2 \quad \text{and} \quad H(s_{r,a}) = \frac{\ell}{2}\log 2\pi e r^2 + O_\ell(e^{-a^2/4}) + O_{G,a}(r), \tag{6.3}$$

while being supported on $B_{ar}(e)$. The error $O_\ell(e^{-a^2/4})$ arises since $s_{r,a}$ is compactly supported while equality holds in (6.1) for Gaussians, which are non-compactly supported.

We then define the entropy at a scale $r > 0$ of a random variable as

$$H_a(g;r) = H(gs_{r,a}) - H(s_{r,a})$$

and the entropy between scales between two scales $r_1, r_2 > 0$ as

$$H_a(g;r_1|r_2) = H_a(g;r_1) - H_a(g;r_2).$$

Roughly speaking, $H_a(g;r_1|r_2)$ measures how much more information g has on scale ar_1 than it has on scale ar_2. We work with the parameter a as the uniform bounds (6.3) are useful for our purposes.

We next aim to relate the entropy between scales to the trace of a random variable. To do so we introduce the trace $\mathrm{tr}(g;r)$ for a random variable g at scale r, which we define as the supremum of all $t \geq 0$ such that we can find some σ-algebra $\mathscr{A}$ and some $\mathscr{A}$-measurable random variable h taking values in G such that

$$|\log(h^{-1}g)| \leq r \quad \text{and} \quad \mathbb{E}[\mathrm{tr}_h(g|\mathscr{A})] \geq tr^2.$$

Then we show in Proposition 6.14 that

$$\mathrm{tr}(g;2ar) \gg a^{-2}(H_a(g;r|2r) - O_\ell(e^{-a^2/4}) - O_{G,a}(r)). \tag{6.4}$$

In section 6.1 we give definitions and discuss basic properties of entropy on G, after which we discuss the Kullback-Leibler divergence on G in section 6.2. In section 6.3 we prove (6.2), after which we study conditional entropy in section 6.4. Finally we prove (6.4) in section 6.5.

6.1. **Entropy and Basic Properties.** For notational convenience, we denote

$$h(x) = -x\log(x)$$

for $x \in (0,\infty)$ and recall that h is concave. If $\lambda = \sum_i p_i \delta_{g_i}$ is a discrete probability measure on G, we define the Shannon entropy of λ as

$$H(\lambda) = \sum_i h(p_i).$$

On the other hand, given an absolutely continuous probability measure λ on G with density f_λ we define

$$H(\lambda) = \int h(f_\lambda)\, dm_G.$$

We extend the definition to finite measures λ that are either absolutely continuous or discrete by setting

$$H(\lambda) = ||\lambda||_1 H(\lambda/||\lambda||_1).$$

In this subsection we collect some useful basic properties of entropy.

Lemma 6.1. *Let $\lambda_1, \ldots, \lambda_n$ be absolutely continuous finite measures on G. Then*

$$H(\lambda_1 + \ldots + \lambda_n) \geq H(\lambda_1) + \ldots + H(\lambda_n).$$

Proof. It suffices to prove the claim for $n = 2$. Let f_1 and f_2 be the densities of λ_1 and λ_2. Then since h is concave

$$\begin{aligned} H(\lambda_1 + \lambda_2) &= (||\lambda_1||_1 + ||\lambda_2||_1) \int h\left(\frac{f_1 + f_2}{||\lambda_1||_1 + ||\lambda_2||_1}\right) dm_G \\ &\geq (||\lambda_1||_1 + ||\lambda_2||_1) \int \frac{||\lambda_1||_1}{||\lambda_1||_1 + ||\lambda_2||_1} h\left(\frac{f_1}{||\lambda_1||_1}\right) dm_G \\ &+ (||\lambda_1||_1 + ||\lambda_2||_1) \int \frac{||\lambda_2||_1}{||\lambda_1||_1 + ||\lambda_2||_1} h\left(\frac{f_2}{||\lambda_2||_1}\right) dm_G \\ &= H(\lambda_1) + H(\lambda_2). \end{aligned}$$

□

Lemma 6.2. *Let $p = (p_1, p_2, \ldots)$ be a probability vector and let $\lambda_1, \lambda_2, \ldots$ be probability measures on G either all absolutely continuous measures or all discrete measures with finite entropy such that $||\lambda_i|| = p_i$. Then*

$$H\left(\sum_{i=1}^{\infty} \lambda_i\right) \leq H(p) + \sum_{i=1}^{\infty} H(\lambda_i).$$

In particular, if $p_i = 0$ for all $i > k$ for some $k \geq 1$ then

$$H\left(\sum_{i=1}^{n} \lambda_i\right) \leq \log k + \sum_{i=1}^{n} H(\lambda_i).$$

Proof. Upon taking limits it suffices to prove the claim for n-dimensional probability vectors $p = (p_1, \ldots, p_n)$ and we only consider the case of absolutely continuous measures as the proof is analogous in the discrete case. We prove the first line in the case when the λ_i are absolutely continuous and denote their densities by f_i. Note that $h(\sum_{i=1}^n a_i) \le \sum_{i=1}^n h(a_i)$ for any $a_1, \dots, a_n \ge 0$. Therefore

$$\begin{aligned} H(\lambda_1 + \ldots + \lambda_n) &= \int h\left(\sum_{i=1}^n f_i\right) dm_G \\ &\le \sum_{i=1}^n \int h(f_i)\, dm_G \\ &= \sum_{i=1}^n \int (-f_i(x)\log(p_i^{-1} f_i) - f_i(x)\log(p_i))\, dm_G \\ &= \sum_{i=1}^n \int p_i h(p_i^{-1} f_i) dm_G + h(p_i) \\ &= \sum_{i=1}^n H(\lambda_i) + H(p). \end{aligned}$$

□

Lemma 6.3. *Let λ_1 be a discrete and and λ_2 be continuous probability measures on G. Then*

$$H(\lambda_1 * \lambda_2) \le H(\lambda_1) + H(\lambda_2)$$

Suppose further that λ_1 is supported on finitely many points with separation at least $2r$ and that the support of λ_2 is contained in a ball of radius r. Then

$$H(\lambda_1 * \lambda_2) = H(\lambda_1) + H(\lambda_2).$$

Proof. Write $\lambda_1 = \sum_{i=1}^n p_i \delta_{g_i}$ and let f be the density of λ_2. Then the density of $\lambda_1 * \lambda_2$ is given by $\sum_{i=1}^n p_i\, f \circ g_i^{-1}$. As $h(\sum_{i=1}^n a_i) \le \sum_{i=1}^n h(a_i)$ for any $a_1, \dots, a_n \ge 0$,

$$\begin{aligned} H(\lambda_1 * \lambda_2) &= \int h\left(\sum_{i=1}^n p_i\, f \circ g_i^{-1}\right) dm_G \\ &\le \sum_{i=1}^n \int h(p_i\, f \circ g_i^{-1})\, dm_G \\ &= \sum_{i=1}^n \int (p_i\, f \circ g_i^{-1})(\log(p_i) + \log(f \circ g_i^{-1}))\, dm_G \\ &= H(\lambda_1) + H(\lambda_2). \end{aligned}$$

If λ_1 is supported on finitely many points with separation at least $2r$ and that the support of λ_2 is contained in a ball of radius r, then the support of the functions $f \circ g_i^{-1}$ is disjoint and the inequality in the second line is an equality. □

6.2. **Kullback-Leibler Divergence.** If $\nu \ll \mu$ are measures on G, then we define the Kullback-Leibler divergence as

$$D_{\mathrm{KL}}(\nu \,||\, \mu) = -\int \log \frac{d\nu}{d\mu}\, d\nu.$$

Observe that if ν is absolutely continuous, then $H(\nu) = D_{\mathrm{KL}}(\nu \,||\, m_G)$. We collect some basic results on the Kullback-Leibler divergence on G.

Lemma 6.4. *Let $\nu \ll \mu$ be measures on G and assume that ν is supported on a set A of positive μ measure. Then*

$$D_{\mathrm{KL}}(\nu \,||\, \mu) \leq \log(\mu(A)).$$

Proof. For convenience write $\nu = f_\nu \, d\mu$. Then by Jensen's inequality,

$$D_{\mathrm{KL}}(\nu \,||\, \mu) = \int_A h\left(f_\nu \frac{\mu(A)}{\mu(A)}\right) d\mu = \int h(f_\nu \mu(A)) \frac{1_A}{\mu(A)} \, d\mu + \log(\mu(A)) \leq \log(\mu(A)).$$

□

Lemma 6.5. *Assume that we can write $G = X_1 \times \ldots \times X_m$ as a product of submanifolds $X_i \subset G$ and assume that $m_G = m_{X_1} \times \ldots \times m_{X_m}$ for measures m_{X_i} on X_i. Denote by π_i the projection from G to X_i and by $\pi_i\mu$ the pushforward of μ under π_i. Then for a probability measure $\mu \ll m_G$ it holds that*

$$D_{\mathrm{KL}}(\mu \,||\, m_G) \leq D_{\mathrm{KL}}(\pi_1\mu \,||\, m_{X_1}) + \ldots + D_{\mathrm{KL}}(\pi_m\mu \,||\, m_{X_m}).$$

Proof. It suffices to prove the claim for $m = 2$. Denote by f_μ the density of μ with respect to m_G and write

$$f_\mu^1(x_2) = \int f_\mu(x_1, x_2) \, dm_{G_1}(x_1) \quad \text{and} \quad f_\mu^2(x_1) = \int f_\mu(x_1, x_2) \, dm_{G_2}(x_2).$$

Therefore,

$$\begin{aligned}
D_{\mathrm{KL}}(\mu \,||\, m_G) &= \int\int h(f_\mu(x_1, x_2)) \, dm_{X_1}(x_1) dm_{X_2}(x_2) \\
&= \int\int h\left(\frac{f_\mu(x_1, x_2)}{f_\mu^2(x_1)} f_\mu^2(x_1)\right) dm_{X_1}(x_1) dm_{X_2}(x_2) \\
&= \int\int h\left(\frac{f_\mu(x_1, x_2)}{f_\mu^2(x_1)}\right) f_\mu^2(x_1) \, dm_{X_1}(x_1) dm_{X_2}(x_2) \\
&+ \int\int - \log(f_\mu^2(x_1)) f_\mu(x_1, x_2) \, dm_{X_1}(x_1) dm_{X_2}(x_2) \\
&\leq \int h(f_\mu^1(x_2)) \, dm_{X_2}(x_2) + \int h(f_\mu^2(x_1)) \, dm_{X_1}(x_1) \\
&= D_{\mathrm{KL}}(\pi_1\mu \,||\, m_{X_1}) + D_{\mathrm{KL}}(\pi_2\mu \,||\, m_{X_2}),
\end{aligned}$$

having used that h is concave and Jensen's inequality in the penultimate line. □

Lemma 6.6. *Let X be a manifold with a measure m_X and let $\Phi : G \to X$ be a diffeomorphism. Then for measures $\nu \ll \mu$ on G such that $\Phi_*\nu \ll m_X$ it holds that*

$$D_{\mathrm{KL}}(\Phi_*\nu || \Phi_*\mu) = D_{\mathrm{KL}}(\nu || \mu) - \int \log \frac{d\Phi_*\nu}{dm_X} \, dm_X$$

for $D_g\Phi$ the differential of Φ at $g \in G$.

Proof. Note that

$$\frac{d\Phi_*\nu}{d\Phi_*\mu}(x) = \frac{d\nu}{d\mu}(\Phi^{-1}(x)) \cdot \frac{d\Phi_*\nu}{dm_X}(x)$$

and therefore

$$\begin{aligned} D_{\mathrm{KL}}(\Phi_*\nu||\Phi_*\nu) &= -\int \log \frac{d\Phi_*\nu}{d\Phi_*\nu}\, d\Phi_*\nu \\ &= -\int \log\left(\frac{d\nu}{d\nu}(\Phi^{-1}(x)) \cdot \frac{d\Phi_*\nu}{dm_X}(x)\right) d\Phi_*\nu \\ &= D_{\mathrm{KL}}(\nu||\mu) - \int \log \frac{d\Phi_*\nu}{dm_X}\, dm_X. \end{aligned}$$

□

Lemma 6.7. *Let λ_1 be a probability measure on G and let λ_2 and λ_3 be measures on G such that $\lambda_1 \ll \lambda_2$, $\lambda_1 \ll \lambda_3$ and $\lambda_2 \ll \lambda_3$. Let $U \subset E$ and suppose that the support of λ_1 is contained in U. Then*

$$|D_{\mathrm{KL}}(\lambda_1 \,||\, \lambda_2) - D_{\mathrm{KL}}(\lambda_1 \,||\, \lambda_3)| \leq \sup_{x\in U}\left|\log \frac{d\lambda_2}{d\lambda_3}\right|.$$

Proof. We calculate

$$\begin{aligned} |D_{\mathrm{KL}}(\lambda_1 \,||\, \lambda_2) - D_{\mathrm{KL}}(\lambda_1 \,||\, \lambda_3)| &= \left|\int_U \log\frac{d\lambda_1}{d\lambda_2}\, d\lambda_1 - \int_U \log\frac{d\lambda_1}{d\lambda_3}\, d\lambda_1\right| \\ &\leq \int_U \left|\log\frac{d\lambda_1}{d\lambda_2} - \log\frac{d\lambda_1}{d\lambda_3}\right| d\lambda_1 \\ &= \int_U \left|\log\frac{d\lambda_2}{d\lambda_3}\right| d\lambda_1 \\ &\leq \sup_{x\in U}\left|\log\frac{d\lambda_2}{d\lambda_3}\right|. \end{aligned}$$

□

6.3. **Entropy and Trace.** In this subsection we prove (6.2). Recall that given $g_0 \in G$ and a random variable g on G we define

$$\mathrm{tr}_{g_0}(g) = \mathrm{tr}(\log(g_0^{-1}g)),$$

whenever $\log(g_0^{-1}g)$ is defined.

Proposition 6.8. *Let G be a Lie group of dimension ℓ. Let $\varepsilon > 0$ and suppose that g is a continuous random variable taking values in $B_\varepsilon(g_0)$ for some $g_0 \in G$. If ε is sufficiently small depending on G,*

$$H(g) \leq \frac{\ell}{2}\log\left(\frac{2\pi e}{\ell}\cdot \mathrm{tr}_{g_0}(g)\right) + O_G(\varepsilon).$$

Proof. We first note that if X is an ℓ-dimensional random vector, then

$$H(X) \leq \frac{\ell}{2}\log\left(\frac{2\pi e}{\ell}\cdot \mathrm{tr}(X)\right) \tag{6.5}$$

Indeed, it follows from the 1-dimensional case (6.1) that $H(X) \leq \frac{1}{2}\log((2\pi e)^\ell \cdot |\mathrm{Var}(X)|)$, where $|\mathrm{Var}(X)|$ is the determinant of the covariance matrix. Note that by the AM-GM inequality $|\mathrm{Var}(X)| \leq \mathrm{tr}(X)^\ell \ell^{-\ell}$, which implies (6.5).

Since $H(g_0^{-1}g) = H(g)$ and $\mathrm{tr}_{g_0}(g) = \mathrm{tr}_e(g_0^{-1}g)$, we may assume without loss of generality that $g_0 = e$. The density $\frac{dm_G}{d(m_{\mathfrak{g}}\circ\log)}$ is smooth and for $\varepsilon > 0$ sufficiently small is $1 + O_G(\varepsilon)$ on $B_\varepsilon(e)$ and therefore $\sup_{B_\varepsilon(e)} \left| \log \frac{dm_G}{d(m_{\mathfrak{g}} \circ \log)} \right| \ll_G \varepsilon$. Thus by Lemma 6.7,

$$|D_{\mathrm{KL}}(g \,||\, m_G) - D_{\mathrm{KL}}(g \,||\, m_{\mathfrak{g}} \circ \log)| \ll_G \varepsilon.$$

The claim follows since by (6.5)

$$D_{\mathrm{KL}}(g \,||\, m_{\mathfrak{g}} \circ \log) = D_{\mathrm{KL}}(\log(g) \,||\, m_{\mathfrak{g}}) = H(\log(g)) \leq \frac{\ell}{2} \log \left(\frac{2\pi e}{\ell} \mathrm{tr}_e(g) \right).$$

□

6.4. **Conditional Entropy and Conditional Trace.** The aim of this subsection is to prove an abstract result relating entropy between scales and the trace. To do so, we first discuss conditional entropy and conditional trace. Let Y be a random variable on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and $\mathscr{A} \subset \mathscr{F}$ be a σ-algebra. Denote by $(Y|\mathscr{A})$ the regular conditional distribution as defined in section 3.2. Assuming that $(Y|\mathscr{A})$ is almost surely absolutely continuous, we define

$$H((Y \mid \mathscr{A}))(\omega) = H((Y|\mathscr{A})(\omega)).$$

Recall that if X_1 and X_2 are two random variables then entropy of X_1 given X_2 is $H(X_1|X_2) = H(X_1, X_2) - H(X_2)$. If X_1 and X_2 have finite entropy and finite joint entropy, then by [Vig21],

$$H(X_1|X_2) = \mathbb{E}[H((X_1|X_2))]. \tag{6.6}$$

We next give an abstract definition of the entropy at a scale and for a smoothing functions s. Indeed, let g and s be random variables on G and assume that s is absolutely continuous. Then the entropy at scale s is defined as

$$H(g; s_1) = H(gs_1) - H(s_1)$$

Moreover, if s_1 and s_2 are absolutely continuous smoothing functions we define the entropy between scales s_1 and s_2 as

$$H(g; s_1|s_2) = H(g; s_1) - H(g; s_2).$$

The following basic result on the growth of conditional entropy holds.

Lemma 6.9. *Let g, s_1, s_2 be independent random variables taking values in G. Assume that s_1 and s_2 are absolutely continuous with finite differential entropy and assume that gs_1 and gs_2 also have finite differential entropy. Then*

$$H(gs_1|gs_2) \geq H(g; s_1|s_2) + H(s_1).$$

Proof. Note that

$$H(gs_2|gs_1) \geq H(gs_2|g, s_1) = H(gs_2|g) = H(s_2)$$

and so

$$H(gs_2, gs_1) = H(gs_2|gs_1) + H(gs_1) \geq H(gs_1) + H(s_2).$$

Therefore

$$\begin{aligned} H(gs_1|gs_2) &= H(gs_2, gs_1) - H(gs_2) \\ &\geq H(gs_1) - H(gs_2) + H(s_2) \\ &\geq H(g; s_1|s_2) + H(s_1). \end{aligned}$$

□

We next define the conditional trace of a random variable on G and relate it to the entropy between scales.

Definition 6.10. *Let g be a random variable defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and taking values in G. Let $\mathscr{A} \subset \mathscr{F}$ be a σ-algebra let g_0 be a $\mathscr{A}$-measurable random variable taking values on G. Then we denote by $\operatorname{tr}_{g_0}(g \,|\, \mathscr{A})$ the $\mathscr{A}$-measurable function given for $\omega \in \Omega$ by*

$$\operatorname{tr}_{g_0}(g \,|\, \mathscr{A})(\omega) = \operatorname{tr}_{g_0(\omega)}((g \,|\, \mathscr{A})(\omega)).$$

We note here that the variance of a measure μ is defined as the variance of a random variable with law μ. It follows from Proposition 6.8 that

$$H((g|\mathscr{A})) \leq \frac{\ell}{2} \log \left(\frac{2\pi e}{\ell} \cdot \operatorname{tr}_{g_0}(g|\mathscr{A}) \right) + O_G(\varepsilon). \tag{6.7}$$

Theorem 6.11. *Let g, s_1 and s_2 be independent absolutely continuous random variables taking values in G and suppose that that s_1 and s_2 are supported on B_ε for some sufficiently small $\varepsilon > 0$ and have finite differential entropy. Write $c = \frac{\ell}{2} \log \frac{2\pi e}{\ell} \operatorname{tr}_e(s_1) - H(s_1)$ and suppose that $\operatorname{tr}_e(s_1) \geq A\varepsilon^2$ for some positive constant A. Then*

$$\mathbb{E}[\operatorname{tr}_{gs_2}(g|gs_2)] \geq \frac{2}{\ell}(H(g; s_1|s_2) - c - C\varepsilon)\operatorname{tr}_e(s_1),$$

where C is some positive constant depending only on A and ℓ.

We first prove some basic result on the trace of the product of two random variables.

Lemma 6.12. *Let $\varepsilon > 0$ be sufficiently small and let a, b be random variables and $\mathscr{A}$ a σ-algebra. Suppose that b is independent from a and $\mathscr{A}$ and let g_0 be an $\mathscr{A}$-measurable random variable. Suppose that $g_0^{-1}a$ and b are almost surely contained in B_ε. Then*

$$\operatorname{tr}_{g_0}(ab|\mathscr{A}) = \operatorname{tr}_{g_0}(a|\mathscr{A}) + \operatorname{tr}_e(b) + O(\varepsilon^3).$$

Note that under the assumptions of Lemma 6.12 it holds by Lemma 3.9 that

$$[ab|\mathscr{A}] = [a|\mathscr{A}][b|\mathscr{A}] = [a|\mathscr{A}]b.$$

Therefore the claim follows from the following unconditional version.

Lemma 6.13. *Let $\varepsilon > 0$ be sufficiently small and let g and h be independent random variables taking values in G. Suppose that the image of g is contained in B_ε and the image of h is contained in $B_\varepsilon(h_0)$ for some $h_0 \in G$. Then*

$$\operatorname{tr}_{h_0}(hg) = \operatorname{tr}_{h_0}(h) + \operatorname{tr}_e(g) + O(\varepsilon^3).$$

Proof. Let $X = \log(h_0^{-1}h)$ and let $Y = \log(g)$. Then $|X|, |Y| \leq \varepsilon$ almost surely and by Taylor's theorem there is a random variable E with $|E| \ll \varepsilon^2$ almost surely such that

$$\log(\exp(X)\exp(Y)) = X + Y + E.$$

Therefore

$$\begin{aligned}
\operatorname{tr}_{h_0}(hg) &= \mathbb{E}[|X + Y + E|^2] - |\mathbb{E}[X + Y + E]|^2 \\
&= \mathbb{E}[|X + Y|^2] - |\mathbb{E}[X + Y]|^2 \\
&+ 2\mathbb{E}[(X + Y) \cdot E] + \mathbb{E}[|E|^2] - 2\mathbb{E}[X + Y]\mathbb{E}[E] - |\mathbb{E}[E]|^2 \\
&= \operatorname{Var}[X + Y] + O(\varepsilon^3) = \operatorname{Var}[X] + \operatorname{Var}[Y] + O(\varepsilon^3).
\end{aligned}$$

□

Proof. (of Theorem 6.11) We note that by 6.6 and Lemma 6.9, it holds that

$$\mathbb{E}[H((gs_1|gs_2))] \geq H(g;s_1|s_2) + H(s_1)$$

and so by (6.7),

$$\mathbb{E}\left[\frac{\ell}{2}\log\frac{2\pi e}{\ell}\mathrm{tr}_{gs_2}(gs_1|gs_2)\right] + O(\varepsilon) \geq H(g;s_1|s_2) + H(s_1).$$

Note that $(gs_2)^{-1}g = s_2^{-1}$, which is contained in $B_\varepsilon(e)$. Therefore by Lemma 6.12,

$$\mathrm{tr}_{gs_2}(gs_1|gs_2) \leq \mathrm{tr}_{gs_2}(g|gs_2) + \mathrm{tr}_e(s_1) + O(\varepsilon^3)$$

and so

$$H(g;s_1|s_2) + H(s_1) \leq \mathbb{E}\left[\frac{\ell}{2}\log\frac{2\pi e}{\ell}\left(\mathrm{tr}_{gs_2}(g|gs_2) + \mathrm{tr}_e(s_1) + O(\varepsilon^3)\right)\right] + O(\varepsilon).$$

Thus

$$\frac{2}{\ell}\left(H(g;s_1|s_2) - c\right) \leq \mathbb{E}\left[\log\left(1 + \frac{\mathrm{tr}_{gs_2}(g|gs_2)}{\mathrm{tr}_e(s_1)} + O_A(\varepsilon)\right)\right].$$

Using that $\log(1+x) \leq x$ for $x \geq 0$, we conclude the claim. □

6.5. **Entropy Between Scales.** In this subsection we prove an explicit result relating the entropy between scales and $\mathrm{tr}(g)$. To do so, we construct a suitable family of smoothing functions. Indeed for given $r > 0$ and $a \geq 1$, denote by $\eta_{r,a}$ a random variable on $\mathfrak{g}$ with density function $f_{r,a} : \mathfrak{g} \to \mathbb{R}$ given by

$$f_{r,a}(x) = \begin{cases} C_{r,a}e^{-\frac{|x|^2}{2r^2}} & \text{if } |x| \leq ar, \\ 0 & \text{otherwise,} \end{cases}$$

where $C_{r,a}$ is a normalizing constant to ensure that $f_{r,a}$ integrates to 1. We furthermore define

$$s_{r,a} = \exp(\eta_{r,a}).$$

We then define the entropy at scale r as

$$H_a(g;r) = H(g;s_{r,a}) = H(gs_{r,a}) - H(s_{r,a})$$

and the entropy between scales $r_1, r_2 > 0$ as

$$\begin{aligned} H_a(g;r_1|r_2) &= H(g;s_{r_1,a}|s_{r_2,a}) = H_a(g;r_1) - H_a(g;r_2) \\ &= (H(gs_{r_1,a}) - H(s_{r_1,a})) - (H(gs_{r_2,a}) - H(s_{r_2,a})). \end{aligned}$$

Recall that $\mathrm{tr}(g;r)$ is defined to be the supremum of all $t \geq 0$ such that we can find some σ-algebra $\mathscr{A}$ and some $\mathscr{A}$-measurable random variable h taking values in G such that

$$|\log(h^{-1}g)| \leq r \quad \text{and} \quad \mathbb{E}[\mathrm{tr}_h(g|\mathscr{A})] \geq tr^2.$$

Proposition 6.14. *Let g be a random variable taking values in G, let $a \geq 1$ and $r > 0$ be such that ar is sufficiently small in terms of G and assume that $g, s_{r,a}$ and $s_{2r,a}$ are independent random variables. Then*

$$\mathrm{tr}(g;2ar) \gg a^{-2}(H_a(g;r|2r) - O_\ell(e^{-a^2/4}) - O_{G,a}(r)),$$

for the implied constants depending on G.

Proposition 6.14 relies on the following lemma.

Lemma 6.15. *The following properties hold for $r > 0$ and $a \geq 1$:*

(i) $\ell r^2 \ll \operatorname{tr}(\eta_{r,a}) \leq \ell r^2$ *and*

$$H(\eta_{r,a}) = \frac{\ell}{2}\log 2\pi e r^2 + O_\ell(e^{-a^2/4}).$$

(ii) If ar *is sufficiently small,* $\ell r^2 \ll \operatorname{tr}_e(s_{r,a}) \leq \ell r^2$ *and*

$$H(s_{r,a}) = \frac{\ell}{2}\log 2\pi e r^2 + O_\ell(e^{-a^2/4}) + O_{G,a}(r).$$

Proof. We note that (ii) follows from (i) and the claim $\ell r^2 \ll \operatorname{tr}(\eta_{r,a}) \leq \ell r^2$ is obvious. To complete the proof of (i), we deal with $r = 1$ case first. Note first that

$$\int_{x\in\mathbb{R}^\ell,|x|\leq a} e^{-|x|^2/2}\,dx \leq \int_{x\in\mathbb{R}^\ell} e^{-|x|^2/2}\,dx = \prod_{i=1}^{\ell}\int_{\mathbb{R}} e^{-x_i^2/2}\,dx_i = (2\pi)^{\ell/2}$$

and by using spherical coordinates

$$\begin{aligned}\int_{x\in\mathbb{R}^\ell,|x|\geq a} e^{-|x|^2/2}\,dx &= c_\ell\int_a^\infty u^{\ell-1}e^{-u^2/2}\,du \\ &\ll_\ell \int_a^\infty e^{-u^2/3}\,du \leq \int_a^\infty e^{-au/3}\,du = \frac{3}{a}e^{-a^2/3} \ll_\ell e^{-a^2/4}.\end{aligned}$$

Thus we conclude

$$\int_{x\in\mathbb{R}^\ell,|x|\leq a} e^{-|x|^2/2}\,dx = (2\pi)^{\ell/2} - \int_{x\in\mathbb{R}^\ell,|x|\geq a} e^{-||x||^2/2}\,dx \geq (2\pi)^{\ell/2} - O_\ell(e^{-a^2/4})$$

and therefore $C_{1,a} = (2\pi)^{-\ell/2} + O_\ell(e^{-a^2/4})$. We are now in a suitable position to calculate $H(\eta_{1,a})$. Indeed,

$$\begin{aligned}H(\eta_{1,a}) &= \int_{|x|\leq a} -C_{1,a}e^{-|x|^2/2}\log\left(C_{1,a}e^{-|x|^2/2}\right)dx \\ &= \int_{|x|\leq a} C_{1,a}\left(\frac{|x|^2}{2} - \log C_{1,a}\right)e^{-|x|^2/2}\,dx\end{aligned}$$

We calculate

$$\begin{aligned}&\int_{x\in\mathbb{R}^\ell} C_{1,a}\left(\frac{|x|^2}{2} - \log C_{1,a}\right)e^{-|x|^2/2}\,dx \\ &= (2\pi)^{\ell/2}C_{1,a}\left(\frac{\ell}{2} - \log C_{1,a}\right) \\ &= \left(1 + O_\ell(e^{-a^2/4})\right)\left(\frac{\ell}{2}\log e + \frac{\ell}{2}\log 2\pi + O_\ell(e^{-a^2/4})\right) \\ &= \frac{\ell}{2}\log 2\pi e + O_\ell(e^{-a^2/4}).\end{aligned}$$

and again using spherical coordinates,

$$\begin{aligned}&\int_{|x|\geq a} C_{1,a}\left(\frac{|x|^2}{2} - \log C_{1,a}\right)e^{-|x|^2/2}\,dx \\ &= c_\ell\int_a^\infty C_{1,a}\left(\frac{u^2}{2} - \log C_{1,a}\right)u^{\ell-1}e^{-u^2/2}\,dx \\ &\ll_\ell O_\ell(e^{-a^2/4}).\end{aligned}$$

Thus the claimed bound on $H(\eta_{1,a})$ follows. Since $f_{r,a}(x) = r^\ell C_{1,a} f_{1,a}(x/r)$ it follows that $H(\eta_{r,a}) = \log(r^\ell) + H(\eta_{1,a})$ and hence the proof is complete. □

Proof. (of Proposition 6.14) We apply Theorem 6.11 to $s_1 = s_{r,a}$ and $s_2 = s_{2r,a}$ and we set $\varepsilon = \ell a r$. By Lemma 6.15 (ii) we have that $\mathrm{tr}_e(s_1) \gg \ell r^2 \gg_{a,\ell} \varepsilon^2$ and $c = \frac{\ell}{2}\log\frac{2\pi e}{\ell}\mathrm{tr}_e(s_1) - H(s_1) \leq O_\ell(e^{-a^2/4}) + O_{a,\ell}(r)$. Applying Theorem 6.11,

$$\mathbb{E}[\mathrm{tr}_{gs_2}(g|gs_2)] \geq cr^2(H(g; r|2r) - O_\ell(e^{-a^2/4}) - O_{G,a}(r))$$

for some absolute constant c depending on G. On the other hand, we have that $|\log((gs_2)^{-1}g)| = |\log s_2| \leq 2ar$ and therefore

$$\mathrm{tr}(g; 2ar) \geq (2ar)^{-2}\mathbb{E}[\mathrm{tr}_{gs_2}(g|gs_2)] \gg a^{-2}(H(g; r|2r) - O_\ell(e^{-a^2/4}) - O_{G,a}(r)).$$

□

7. Variance Growth on $\mathrm{Sim}(\mathbb{R}^d)$

In this section we return to $G = \mathrm{Sim}(\mathbb{R}^d)$ with dimension $\ell = \frac{d(d+1)}{2} + 1$. For μ a probability measure on G we denote by $\gamma_1, \gamma_2, \ldots$ independent μ-distributed samples of μ and write

$$q_n = \gamma_1 \cdots \gamma_n.$$

For $\kappa > 0$ be denote by τ_κ the stopping time

$$\tau_\kappa = \inf\{n : \rho(q_n) \leq \kappa\}.$$

The goal of this section is to give bounds for $\sum_{i=1}^N \mathrm{tr}(q_{\tau_\kappa}, s_i)$ for suitable scales s_i. Towards the proof of our main theorem as discussed in section 2.2, it would be ideal to give a bound roughly of the form

$$\sum_{i=1}^N \mathrm{tr}(q_{\tau_\kappa}, 2^i ar) \gg \frac{h_\mu}{|\chi_\mu|}\log\kappa^{-1} \qquad \text{with} \quad r \approx \kappa^{\frac{S_\mu}{|\chi_\mu|}} \quad \text{and} \quad 2^N r \approx \kappa^{\frac{h_\mu}{2\ell|\chi_\mu|}} \tag{7.1}$$

for sufficiently small κ. As we explain below, we can't quite achieve (7.1) and the bound we arrive at will also depend on the separation rate S_μ. To estimate the left hand side of (7.1) we apply Proposition 6.14 to each of the terms $\mathrm{tr}(q_{\tau_\kappa}, 2^i ar)$ which gives

$$\sum_{i=1}^N \mathrm{tr}(q_{\tau_\kappa}, 2^i ar) \gg a^{-2}(H_a(q_\tau; r|2^N r) + O_d(Ne^{-a^2/\ell}) + O_d(r)) \tag{7.2}$$

having used that by a telescoping sum

$$H_a(q_\tau; r|2^N r) = \sum_{i=1}^N H_a(q_\tau; 2^{i-1}r|2^i r).$$

The main contribution from (7.1) comes from suitable estimates for $H_a(q_\tau; r|2^N r)$. Indeed, we will show in Proposition 7.1 that, up to negligible error terms,

$$H_a(q_{\tau_\kappa}; r|2^N r) \gg \frac{h_\mu}{|\chi_\mu|}\log\kappa^{-1}. \tag{7.3}$$

To show this, we recall that

$$H_a(q_{\tau_\kappa}; r|2^N r) = H_a(q_{\tau_\kappa}; r) - H_a(q_{\tau_\kappa}; 2^N r)$$

and therefore we need to estimate the two arising terms $H_a(q_{\tau_\kappa}; r)$ and $H_a(q_{\tau_\kappa}; 2^N r)$. To bound the first term, as we explain after the statement of Lemma 7.2, we use that with high probability $\tau_\kappa \approx \log(\kappa^{-1})/|\chi_\mu|$ and so the points in the support of q_{τ_κ} are separated by distance $r \approx \kappa^{\frac{S_\mu}{|\chi_\mu|}} \approx \exp(-S_\mu \tau_\kappa)$. For the second term we use the large deviation principle and the polynomial decay of our self-similar measure.

Combining (7.2) with (7.3) would lead to (7.1) would it not be for the error term $O_d(Ne^{-a^2/\ell})$. Indeed, to not cancel out the lower bound from (7.3) we require that

$$Ne^{-a^2/\ell} \leq c\frac{h_\mu}{|\chi_\mu|} \log \kappa^{-1}$$

for a sufficiently small constant c. By our choice of N it holds that $N \approx \frac{S_\mu}{|\chi_\mu|} \log \kappa^{-1}$ and therefore

$$e^{-a^2/\ell} \leq c\frac{h_\mu}{S_\mu}.$$

So we have to set

$$a^2 = c \max\left\{1, \log \frac{S_\mu}{h_\mu}\right\}.$$

Applying then (7.2), since the error term $O_d(r)$ is negligible, we conclude that

$$\sum_{i=1}^{N} \mathrm{tr}(q_{\tau_\kappa}, 2^i ar) \gg \frac{h_\mu}{|\chi_\mu|} \log \kappa^{-1} \max\left\{1, \log \frac{S_\mu}{h_\mu}\right\}^{-1}. \tag{7.4}$$

We will give a precise proof of the latter bound in Proposition 7.5.

7.1. **Entropy Gap of Stopped Random Walk.** In this subsection we show that the entropy between scales is large for a suitable stopped random walk on $G = \mathrm{Sim}(\mathbb{R}^d)$. Indeed, we establish the following more precise version of (7.3).

Proposition 7.1. *Let μ be a finitely supported, contracting on average probability measure on G. Suppose that $S_\mu < \infty$ and that $h_\mu/|\chi_\mu|$ is sufficiently large. Let $S > S_\mu$, $\kappa > 0$ and $a \geq 1$ and suppose that $0 < r_1 < r_2 < a^{-1}$ with $r_1 < \exp(-S\log(\kappa^{-1})/|\chi_\mu|)$. Then as $\kappa \to 0$,*

$$H_a(q_{\tau_\kappa}; r_1|r_2) \geq \left(\frac{h_\mu}{|\chi_\mu|} - d\right) \log \kappa^{-1} + H(s_{r_2,a}) + o_{\mu,d,S,a}(\log \kappa^{-1}).$$

Proposition 7.1 directly follows from Lemma 7.2 and Lemma 7.3.

Lemma 7.2. *Under the assumptions of Proposition 7.1, as $\kappa \to 0$,*

$$H_a(q_{\tau_\kappa}; r_1) \geq \frac{h_\mu}{|\chi_\mu|} \log \kappa^{-1} + o_{\mu,d,S,a}(\log \kappa^{-1}).$$

Recall that $H_a(q_{\tau_\kappa}; r_1) = H(q_{\tau_\kappa} s_{r_1,a}) - H(s_{r_1,a})$. To give the proof idea, note that with high probability $\tau_\kappa \approx \log(\kappa^{-1})/|\chi_\mu|$. Also, by definition of h_μ, we have that $H(q_{\log(\kappa^{-1})/|\chi_\mu|}) \geq h_\mu \log(\kappa^{-1})/|\chi_\mu|$. On the other hand, $s_{r_1,a}$ is mostly contained in a ball around the identity with radius $O(\exp(-S\log(\kappa^{-1})/|\chi_\mu|))$, and therefore by Lemma 6.3 we have $H(q_{\log(\kappa^{-1})/|\chi_\mu|} \cdot s_{r_1,a}) = H(q_{\log(\kappa^{-1})/|\chi_\mu|}) + H(s_{r_1,a})$, which implies the claim. We proceed with a more rigorous proof.

Proof. For ease of notation we write in this proof $\tau = \tau_\kappa$. Fix some $\varepsilon > 0$ which is sufficiently small in terms of S and μ. Let $m = \lfloor \log(\kappa^{-1})/|\chi_\mu| \rfloor$ and define τ' as

$$\tau' = \begin{cases} \lceil (1+\varepsilon)m \rceil & \text{if } \tau > \lceil (1+\varepsilon)m \rceil, \\ \lfloor (1-\varepsilon)m \rfloor & \text{if } \tau < \lfloor (1-\varepsilon)m \rfloor, \\ \tau & \text{otherwise.} \end{cases}$$

For a random variable X denote by $\mathcal{L}(X)$ its law. Furthermore, given an event A, we will denote by $\mathcal{L}(X)|_A$ the measure given by the push forward of the restriction of $\mathbb{P}$ to A under the random variable X. Note that $||\mathcal{L}(X)|_A|| = \mathbb{P}[A]$.

By applying Lemma 6.1,

$$\begin{aligned} H(q_\tau s_{r_1,a}) &= H(\mathcal{L}(q_\tau) * \mathcal{L}(s_{r_1,a})) \\ &\geq H(\mathcal{L}(q_\tau)|_{\tau=\tau'} * \mathcal{L}(s_{r_1,a})) + H(\mathcal{L}(q_\tau)|_{\tau\neq\tau'} * \mathcal{L}(s_{r_1,a})) \\ &\geq H(\mathcal{L}(q_\tau)|_{\tau=\tau'} * \mathcal{L}(s_{r_1,a})) + \mathbb{P}[\tau \neq \tau'] H(\mathcal{L}(s_{r_1,a})), \end{aligned} \tag{7.5}$$

having used that

$$\begin{aligned} H(\mathcal{L}(q_\tau)|_{\tau\neq\tau'} * \mathcal{L}(s_{r_1,a})) &\geq H(\mathcal{L}(q_\tau)|_{\tau\neq\tau'} * \mathcal{L}(s_{r_1,a})|q_\tau) \\ &\geq \mathbb{P}[\tau \neq \tau'] H(\mathcal{L}(s_{r_1,a})). \end{aligned}$$

We next apply that $s_{r_1,a}$ has small support. Set $\delta = \frac{1}{4}(S - S_\mu)$. Write $D_m = \bigcup_{n=1}^m \operatorname{supp}(\mu^{*i})$ for all $m \geq 1$. Then for every N sufficiently large, $\exp(-(S_\mu + \delta)N) < d(x,y)$ for all $x, y \in D_N$. Therefore for ε and κ sufficiently small, $\exp(-(S_\mu + 2\delta)m) < d(x,y)$ for all $x, y \in D_{\lceil (1+\varepsilon)m \rceil}$. As $d(s_{r_1,a}, e) \ll_G r_1 a$ it follows that if κ is sufficiently small in terms of μ, a and S,

$$d(s_{r_1,a}, \mathrm{Id}) < O(a \exp(-Sm)) < \frac{1}{2} \min_{x,y \in \operatorname{supp}(q_{\tau'}), x \neq y} d(x,y).$$

In particular, by Lemma 6.3,

$$H(\mathcal{L}(q_\tau)|_{\tau=\tau'} * \mathcal{L}(s_{r_1,a})) = H(\mathcal{L}(q_\tau)|_{\tau=\tau'}) + \mathbb{P}[\tau = \tau'] H(\mathcal{L}(s_{r_1,a})). \tag{7.6}$$

Combining (7.6) with (7.5),

$$H(q_\tau s_{r_1,a}) \geq H(\mathcal{L}(q_\tau)|_{\tau=\tau}) + H(s_{r_1,a}).$$

It remains to estimate $H(\mathcal{L}(q_\tau)|_{\tau=\tau'})$. Consider the random variable

$$X' = (q_{\lfloor (1-\varepsilon)m \rfloor}, \gamma_{\lfloor (1-\varepsilon)m \rfloor} + 1, \gamma_{\lfloor (1-\varepsilon)m \rfloor + 2}, \dots, \gamma_{\lceil (1+\varepsilon)m \rceil + 1}).$$

As $q_{\tau'}$ is completely determined by X', we have $H(X'|q_{\tau'}) = H(X') - H(q_{\tau'})$.

Let K be the number of points in the support of μ. Note that if

$$\gamma_{\lfloor (1-\varepsilon)m \rfloor + 1}, \gamma_{\lfloor (1-\varepsilon)m \rfloor + 2}, \dots, \gamma_{\lceil (1+\varepsilon)m \rceil}$$

and τ' are fixed, then for any possible value of $q_{\tau'}$ there is at most one choice of $q_{\lfloor (1-\varepsilon)m \rfloor}$ which would lead to this value of $q_{\tau'}$. Therefore for each y in the image of $q_{\tau'}$ there are at most $(2\varepsilon m + 2)K^{2\varepsilon m + 2}$ elements x in the image of X' such that $\mathbb{P}[X' = x \cap q_{\tau'} = y] > 0$. Therefore $(X'|q_{\tau'})$ is almost surely supported on less than $(2\varepsilon m + 2)K^{2\varepsilon m + 2}$ points and hence by (6.6),

$$H(X'|q_{\tau'}) \leq \log\left((2\varepsilon m + 2)K^{2\varepsilon m + 2}\right) \leq \frac{2\varepsilon \log K}{|\chi_\mu|} \log \kappa^{-1} + o_{\mu,\varepsilon}(\log \kappa^{-1}).$$

On the other hand,

$$H(X') \geq H(q_m) \geq h_{\mathrm{RW}} \cdot m \geq \frac{h_{RW}}{|\chi_\mu|} \log \kappa^{-1} - o_\mu(\log \kappa^{-1}) \tag{7.7}$$

and therefore

$$H(q_{\tau'}) \geq \frac{h_{RW} - 2\varepsilon \log K}{|\chi_\mu|} \log \kappa^{-1} - o_{\mu,\varepsilon}(\log \kappa^{-1}).$$

To continue, we note that by Lemma 6.2,

$$H(q_{\tau'}) \leq H(\mathcal{L}(q_{\tau'})|_{\tau=\tau'}) + H(\mathcal{L}(q_{\tau'})|_{\tau\neq\tau'}) + \log 2. \tag{7.8}$$

We wish to bound $H(\mathcal{L}(q_{\tau'})|_{\tau=\tau'})$ from below. By the large deviation principle, $\mathbb{P}[\tau \neq \tau'] \leq \alpha^m$ for $\alpha \in (0,1)$ only depending on ε and μ. We also know that conditional on $\tau \neq \tau'$, there are at most $2K^{\lceil (1+\varepsilon)m \rceil}$ possible values for $q_{\tau'}$ and therefore

$$H(\mathcal{L}(q_{\tau'})|_{\tau\neq\tau'}) \leq \alpha^m \log\left(2K^{\lceil (1+\varepsilon)m \rceil}\right) = o_{\mu,\varepsilon}(\log \kappa^{-1}).$$

This implies

$$H(\mathcal{L}(q_{\tau'})|_{\tau=\tau'}) \geq \frac{h_{RW} - 2\varepsilon \log K}{|\chi_\mu|} \log \kappa^{-1} - o_{\mu,\varepsilon}(\log \kappa^{-1}).$$

Since ε can be made arbitrarily small, the claim follows. □

Lemma 7.3. *Under the assumptions of Proposition 7.1, as $\kappa \to 0$,*

$$H(q_{\tau_\kappa} s_{r_2,a}) \leq d \log \kappa^{-1} + o_{\mu,d,a}(\log \kappa^{-1}).$$

Proof. As in the proof of Proposition 7.2, write $\tau = \tau_\kappa$ and $K = |\mathrm{supp}(\mu)|$. We use the product structure on G combined with Lemma 6.5. Indeed, note that a choice of Haar measure on G is given as

$$\int f \, dm_G = \int f(\rho U + b)\, \rho^{-(d+1)} d\rho dU db,$$

for dr, db the Lebesgue measure and dU the Haar probability measure on $O(d)$. Therefore by Lemma 6.5, $H(q_\tau s_{r_2,a}) \leq$

$$D_{\mathrm{KL}}(\rho(q_\tau s_{r_2,a}) \,||\, \rho^{-(d+1)} d\rho) + D_{\mathrm{KL}}(U(q_\tau s_{r_2,a}) \,||\, dU) + D_{\mathrm{KL}}(b(q_\tau s_{r_2,a}) \,||\, db).$$

We give suitable bounds for each these terms. As dU is a probability measure $D_{\mathrm{KL}}(U(q_\tau s_{r_2,a}) \,||\, dU) \leq 0$ by Lemma 6.4.

We next deal with $D_{\mathrm{KL}}(b(q_\tau s_{r_2,a}) \,||\, db)$. Denote by ν_τ the distribution of $b(q_\tau s_{r_2,a})$. We claim that there is $\alpha = \alpha(\mu, d, a)$ such that

$$\nu_\tau(B_R^c) \leq R^{-\alpha} \tag{7.9}$$

for all sufficiently small κ and sufficiently large R. Note that

$$|b(q_\tau s_{r_2,a})| = |\rho(q_\tau) U(q_\tau) b(s_{r_2,a}) + b(q_\tau)| \leq \kappa |b(s_{r_2,a})| + |b(q_\tau)|$$

and therefore it suffices to show (7.9) for the distribution of $b(q_\tau)$, which we denote by ν'_τ. For $x \in \mathbb{R}^d$,

$$|b(q_\tau) - q_\tau(x)| \leq |q_\tau(0) - q_\tau(x)| \leq \rho(q_\tau)|x| \leq \kappa |x|$$

and so $|b(q_\tau)| \leq |q_\tau(x)| + \kappa|x|$. Therefore if $R \leq |b(q_\tau)|$ then either $R/2 \leq |q_\tau(x)|$ or $R/2 \leq \kappa|x|$. Also note that if x is sampled from ν independently from $\gamma_1, \gamma_2, \ldots$, so is $q_\tau(x)$. By Theorem 4.1 this implies that

$$\nu'_\tau(B_R^c) \leq \nu(B_{R/2}^c) + \nu(B_{R/2\kappa}^c) \leq R^{-\alpha_2} 2^{\alpha_2} (1 + \kappa^{-1})^{-\alpha_2},$$

showing (7.9).

To conclude we deduce from (7.9) that $D_{\mathrm{KL}}(\nu_\tau \,||\, db)$ is bounded by a constant depending on μ, d and a and therefore is $\leq o_{\mu,d,a}(\log \kappa^{-1})$. Indeed denote by f_τ the density of ν_τ such that

$$D_{\mathrm{KL}}(\nu_\tau \,||\, db) = \int -f_\tau \log f_\tau \, dm_{\mathbb{R}^d}.$$

Also let $L > 1$ be a constant and for $i = 0, 1, 2, \ldots$ write $p_i = \nu_\tau(B_{L^{i+1}} \backslash B_{L^i})$ such that $p_i \leq \nu_\tau(B_{L^i}^c) \leq L^{-i\alpha}$. Thus it holds by Jensen's inequality for $h(x) = -x \log x$,

$$\begin{aligned} D_{\mathrm{KL}}(\nu_\tau \,||\, db) &= \sum_{i \geq 0} \int_{B_{L^{i+1}} \backslash B_{L^i}} -f_\tau \log f_\tau \, dm_{\mathbb{R}^d} \\ &= \sum_{i \geq 0} \int_{B_{L^{i+1}} \backslash B_{L^i}} -f_\tau \log \left(\frac{f_\tau p_i}{p_i} \right) dm_{\mathbb{R}^d} \\ &= \sum_{i \geq 0} \left(\int h(f_\tau p_i) \frac{1_{B_{L^{i+1}} \backslash B_{L^i}}}{p_i} \, dm_{\mathbb{R}^d} + p_i \log(p_i) \right) \\ &\leq \sum_{i \geq 0} h(p_i) \leq \sum_{0 \leq i \leq I} h(p_i) + \sum_{i \geq I} h(L^{-i\alpha}) < \infty, \end{aligned}$$

having used in the last line that $\log(p_i) \leq 0$ and that $h(x)$ is monotonically decreasing for small x and therefore $h(p_i) \leq h(L^{-i\alpha})$ for $i \geq I$ with I sufficiently large.

Finally, we estimate $D_{\mathrm{KL}}(\rho(q_{\tau_\kappa} s_{r_2,a}) \,||\, \rho^{-(d+1)} d\rho)$. Fix $\varepsilon > 0$ and let A be the event that $\rho(q_\tau) \geq \kappa^{(1+\varepsilon)}$. By Lemma 3.11 there is $\delta > 0$ only depending on μ and ε such that $\mathbb{P}[A^c] \leq \kappa^\delta$. By Lemma 6.4,

$$\begin{aligned} D_{\mathrm{KL}}(\mathcal{L}(\rho(q_{\tau_\kappa} s_{r_2,a}))|_A \,||\, \rho^{-(d+1)} d\rho) &\leq \log \left(\int_{\kappa^{1+\varepsilon}}^{\infty} \rho^{-(d+1)} \, d\rho \right) \\ &= \log \left(d^{-1} \kappa^{-d(1+\varepsilon)} \right) \leq d(1+\varepsilon) \log \kappa^{-1}. \end{aligned}$$

To bound $H(\mathcal{L}(q_\tau s_{r_2,a})|_{A^c})$, we note that as in Lemma 6.3 it suffices to bound the Shannon entropy of $H(\mathcal{L}(q_\tau)|_{A^c})$. If $\tau \leq 2\frac{\log \kappa^{-1}}{|\chi_\mu|}$, the contribution can be bounded by $\kappa^\delta \frac{2 \log \kappa^{-1}}{|\chi_\mu|} \log K$. By the large deviation principle, when $n \geq 2\frac{\log \kappa^{-1}}{|\chi_\mu|}$ it holds that $\mathbb{P}[\tau = n] \leq \alpha^n$ for some $\alpha \in (0,1)$. Therefore the contribution in this case is $\leq \alpha^n n \log K$ where $\alpha \in (0,1)$ is some constant depending on μ. Summing over all $n \geq 2\frac{\log \kappa^{-1}}{|\chi_\mu|}$ and using Lemma 6.2, we conclude that $H(\mathcal{L}(q_\tau s_{r_2,a})|_{A^c})$ is bounded and therefore $o_{\mu,\varepsilon}(\log \kappa^{-1})$. As $\varepsilon > 0$ was arbitrary the claim follows. ☐

7.2. **Trace Bounds for Stopped Random Walk.** In this subsection we give a precise proof of (7.4) following the sketch given at the beginning of this section. We first convert Proposition 7.1 into an integral bound.

Proposition 7.4. *Let μ be a finitely supported, contracting on average probability measure on $G = \mathrm{Sim}(\mathbb{R}^d)$ and write $\ell = \dim G = \frac{d(d+1)}{2} + 1$. Suppose that $S_\mu < \infty$ and that $h_\mu/|\chi_\mu|$ is sufficiently large. Let $S > S_\mu$ and suppose that S is chosen* *sufficiently large such that* $h_\mu \le S$*. Then for sufficiently small* κ*,*

$$\int_{\kappa^{\frac{S}{|\chi_\mu|}}}^{\kappa^{\frac{h_\mu}{2\ell|\chi_\mu|}}} \frac{1}{u}\mathrm{tr}(q_{\tau_\kappa}; u)\, du \gg \left(\frac{h_\mu}{|\chi_\mu|}\right) \max\left\{1, \log\frac{S}{|\chi_\mu|}\right\}^{-1} \log\kappa^{-1}.$$

Proof. Let $\tau = \tau_\kappa$ and let $a \ge 1$ to be determined. Let

$$r_1 = a^{-1}\kappa^{\frac{S}{|\chi_\mu|}} = a^{-1}\exp\left(-\frac{S}{|\chi_\mu|}\log\kappa^{-1}\right)$$

and

$$N = \left\lfloor \left(\frac{S}{|\chi_\mu|} - \frac{h_\mu}{2\ell|\chi_\mu|}\right)\frac{\log\kappa^{-1}}{\log 2}\right\rfloor - 1.$$

Note that

$$\frac{1}{4}\frac{\kappa^{\frac{h_\mu}{2\ell|\chi_\mu|}}}{ar_1} = \frac{1}{4}\frac{\kappa^{-\frac{S}{|\chi_\mu|}}}{\kappa^{-\frac{h_\mu}{2\ell|\chi_\mu|}}} \le 2^N \le \frac{1}{2}\frac{\kappa^{-\frac{S}{|\chi_\mu|}}}{\kappa^{-\frac{h_\mu}{2\ell|\chi_\mu|}}} = \frac{1}{2}\frac{\kappa^{\frac{h_\mu}{2\ell|\chi_\mu|}}}{ar_1}.$$

Given $u \in [1, 2)$ and an integer $0 \le i \le N - 1$ let

$$k_i(u) = H_a(q_\tau; 2^{i-1}ur_1 | 2^i ur_1).$$

Then by Proposition 6.14, there is some constant $c = c(d) > 0$ depending only on d such that

$$\mathrm{tr}(q_\tau; a2^i ur_1) \ge ca^{-2}(k_i(u) - O_d(e^{-\frac{a^2}{4}}) - O_{d,a}(2^i r_1)). \tag{7.10}$$

Thus

$$\sum_{i=1}^{N}\mathrm{tr}(q_\tau; a2^i ur_1) \ge ca^{-2}\sum_{i=1}^{N} k_i(u) - O_d(Ne^{-\frac{a^2}{4}}a^{-2}) - O_{d,a}(N2^N r_1).$$

Note that for $u \in [1, 2)$ we have $a2^N ur_1 \le \kappa^{\frac{h_\mu}{2\ell|\chi_\mu|}}$ and $aur_1 \ge \kappa^{\frac{S}{|\chi_\mu|}}$. Therefore,

$$\begin{aligned}
&\int_{\kappa^{\frac{S}{|\chi_\mu|}}}^{\kappa^{\frac{h_\mu}{2\ell|\chi_\mu|}}} \frac{1}{u}\mathrm{tr}(q_\tau; u)\, du \\
\ge &\sum_{i=1}^{N}\int_{a2^i ur_1}^{a2^{i+1}ur_1} \frac{1}{u}\mathrm{tr}(q_\tau; u)\, du \\
\ge &\sum_{i=1}^{N}\int_{1}^{2} \frac{1}{u}\mathrm{tr}(q_\tau; a2^i ur_1)\, du \\
\ge &ca^{-2}\int_1^2 \frac{1}{u}\left(\sum_{i=1}^{N} k_i(u) - O_d(Ne^{-\frac{a^2}{4}}a^{-2}) - O_{d,a}(N2^N r_1)\right) du.
\end{aligned} \tag{7.11}$$

Observe that $\sum_{i=1}^N k_i(u) = H_a(q_{\tau_\kappa}; ur_1 | 2^N ur_1)$ and therefore by Proposition 7.1 and Lemma 6.15,

$$\begin{aligned}
\sum_{i=1}^{N} k_i(u) &\ge \left(\frac{h_\mu}{|\chi_\mu|} - d\right)\log\kappa^{-1} + \ell\cdot\log 2^N ur_1 + o_{\mu,d,S,a}(\log\kappa^{-1}) \\
&\ge \left(\frac{h_\mu}{|\chi_\mu|} - d - \frac{h_\mu}{2|\chi_\mu|}\right)\log\kappa^{-1} + o_{\mu,d,S,a}(\log\kappa^{-1}).
\end{aligned} \tag{7.12}$$

Let $C = C(d)$ be chosen such that the error term $O(Ne^{-\frac{a^2}{4}}a^{-2})$ in (7.11) can be bounded above by $CNe^{-\frac{a^2}{4}}a^{-2}$. Note that this is at most $C\frac{S}{|\chi_\mu|\log 2}e^{-\frac{a^2}{4}}a^{-2}\log\kappa^{-1}$. Let c be as in (7.10). We take our value of a to be

$$a = 2\sqrt{\log\left(\frac{4C}{c\log 2}\frac{S}{h_\mu}\right)}.$$

Then

$$CNe^{-\frac{a^2}{4}}a^{-2} \le ca^{-2}\frac{h_\mu}{4|\chi_\mu|}\log\kappa^{-1}.$$

We also note that $N2^N r_1 \le o_{\mu,d,S}(\log\kappa^{-1})$. Therefore combining (7.11) and (7.12),

$$\int_{\kappa^{\frac{S}{|\chi_\mu|}}}^{\kappa^{\frac{h_\mu}{2\ell|\chi_\mu|}}} \frac{1}{u}\operatorname{tr}(q_\tau;u)\,du \ge ca^{-2}\left(\frac{h_\mu}{|\chi_\mu|} - d - \frac{h_\mu}{2|\chi_\mu|} - \frac{h_\mu}{4|\chi_\mu|}\right)\log\kappa^{-1} + o_{\mu,d,S}(\log\kappa^{-1}).$$

Note further that $a^2 \ll_d \max\{1, \log\frac{S}{h_\mu}\}$. Thus we have for all sufficiently small κ (depending on μ and M),

$$\int_{\kappa^{\frac{S}{|\chi_\mu|}}}^{\kappa^{\frac{h_\mu}{2\ell|\chi_\mu|}}} \frac{1}{u}\operatorname{tr}(q_\tau;u)\,du \gg_d \left(\frac{h_\mu}{|\chi_\mu|}\right)\max\left\{1, \log\frac{S}{h_\mu}\right\}^{-1}\log\kappa^{-1}.$$

□

Finally we prove the following more precise version of (7.5). We show further that $s_{i+1} \ge \kappa^{-3}s_i$ in order to apply Proposition 6.15 to concatenate proper decompositions as defined and discussed in section 8.

Proposition 7.5. *Let μ be a finitely supported, contracting on average probability measure on $G = \operatorname{Sim}(\mathbb{R}^d)$ and write $\ell = \dim G = \frac{d(d+1)}{2} + 1$. Suppose that $S_\mu < \infty$ and that $h_\mu/|\chi_\mu|$ is sufficiently large. Let $S > S_\mu$ be chosen large enough that $S \ge h_\mu$. Suppose that κ is sufficiently small (depending on μ and S) and let $\widehat{m} = \lfloor\frac{S}{100|\chi_\mu|}\rfloor$.*

Then there exist $s_1, s_2, \dots, s_{\widehat{m}} > 0$ such that for each $i \in [\widehat{m}]$,

$$s_i \in (\kappa^{\frac{S}{|\chi_\mu|}}, \kappa^{\frac{h_\mu}{2\ell|\chi_\mu|}})$$

and for each $i \in [\widehat{m}-1]$ $s_{i+1} \ge \kappa^{-3}s_i$ and

$$\sum_{i=1}^{\widehat{m}} \operatorname{tr}(q_{\tau_\kappa}; s_i) \gg_d \left(\frac{h_\mu}{|\chi_\mu|}\right)\max\left\{1, \log\frac{S}{h_\mu}\right\}^{-1}.$$

Proof. Let $A = \kappa^{\frac{h_\mu}{4\widehat{m}\ell|\chi_\mu|} - \frac{S}{2\widehat{m}|\chi_\mu|}}$. Define $a_1, a_2, \dots, a_{2\widehat{m}+1}$ by $a_i = \kappa^{\frac{S}{|\chi_\mu|}}A^{i-1}$. Therefore $a_1 = \kappa^{\frac{S}{|\chi_\mu|}}$ and $a_{2\widehat{m}+1} = \kappa^{\frac{h_\mu}{2\ell|\chi_\mu|}}$. Furthermore, provided $h_\mu/|\chi_\mu|$ is sufficiently large, we have $\kappa^{-3} \le A \le \kappa^{-50}$. In particular $a_{i+1} \ge \kappa^{-3}a_i$.

Let U and V be defined by

$$U = \bigcup_{i=1}^{\widehat{m}}[a_{2i-1}, a_{2i}) \quad \text{and} \quad V = \bigcup_{i=1}^{\widehat{m}}[a_{2i}, a_{2i+1}).$$

Without loss of generality, upon replacing U with V, by Proposition 7.4

$$\int_U \frac{1}{u}\mathrm{tr}(q_{\tau_\kappa}; u)\, du \gg_d \left(\frac{h_\mu}{|\chi_\mu|}\right) \max\left\{1, \log\frac{S}{|\chi_\mu|}\right\}^{-1} \log\kappa^{-1}.$$

For $i \in [\widehat{m}]$ let $s_i \in (a_{2i-1}, a_{2i})$ be chosen such that

$$\mathrm{tr}(q_{\tau_\kappa}; s_i) \geq \frac{1}{2} \sup_{u \in (a_{2i-1}, a_{2i})} \mathrm{tr}(q_{\tau_\kappa}; u).$$

In particular,

$$\mathrm{tr}(q_{\tau_\kappa}; s_i) \geq \frac{1}{2\log A}\int_{a_{2i-1}}^{a_{2i}} \frac{1}{u}\mathrm{tr}(q_{\tau_\kappa}; u)\, du.$$

Summing over i gives

$$\begin{aligned}\sum_{i=1}^{\widehat{m}} \mathrm{tr}(q_{\tau_\kappa}; s_i) &\geq \frac{1}{2\log A}\int_U \frac{1}{u}\mathrm{tr}(q_{\tau_\kappa}; u)\, du \\ &\geq \frac{c}{2\log A}\left(\frac{h_\mu}{|\chi_\mu|}\right) \max\left\{1, \log\frac{S}{|\chi_\mu|}\right\}^{-1} \log\kappa^{-1}.\end{aligned}$$

As $\log A \asymp \log\kappa^{-1}$ it follows that, provided that κ is sufficiently small depending on μ, d, S,

$$\sum_{i=1}^{\widehat{m}} \mathrm{tr}(q_{\tau_\kappa}; s_i) \gg_d \left(\frac{h_\mu}{|\chi_\mu|}\right) \max\left\{1, \log\frac{S}{|\chi_\mu|}\right\}^{-1}.$$

Finally we note that as $A \geq \kappa^{-3}$ we have that $s_{i+1} \geq \kappa^{-3} s_i$. □

8. Decomposition of Stopped Random Walk

In this section Theorem 2.4 is proved. We construct samples from ν in a suitable way in order to bound the order k detail of ν. Given a probability measure μ on $G = \mathrm{Sim}(\mathbb{R}^d)$ we denote by $\gamma_1, \gamma_2, \ldots$ independent μ-distributed random variables and write $q_n = \gamma_1 \cdots \gamma_n$. Recall that if x distributed like ν and τ is a stopping time, then by Lemma 2.24 from [Kit23] the random variable $q_\tau x$ is distributed like ν.

As discussed in the outline of proofs, one uses Proposition 7.5 to make a decomposition

$$q_{\tau_\kappa} x = g_1 \exp(U_1) g_2 \exp(U_2) \cdots g_n \exp(U_n) x \tag{8.1}$$

with a suitable $\kappa > 0$ that satisfies

$$|U_i| \leq \rho(g_1 \cdots g_i)^{-1} r \quad \text{and} \quad \sum_{i=1}^{n} \mathrm{tr}(\rho(g_1 \cdots g_i) U_i) \geq C r^2 \tag{8.2}$$

for a sufficiently large constant C and a given scale $r > 0$. The definition of $\mathrm{tr}(q_{\tau_\kappa}, s_i)$ requires us to work with a σ-algebra $\mathscr{A}$ and with the conditional trace in (8.2). As stated in (2.9), we need to have (8.2) at $O(\log\log r^{-1})$ many suitable times κ_i.

Indeed, in order to deduce (8.2) from Proposition 7.5 we need to combine all the information at the scales $s_1, \ldots, s_{\widehat{m}}$. One also needs to ensure that the assumptions from the Taylor-approximation result Proposition 3.4 are satisfied for each scale s_i and that we can apply our (c,T)-well-mixing and (α_0, θ, A)-non-degeneracy conditions to deduce that

$$\mathrm{Var}(\zeta_i(U_i)) \geq c_1 \mathrm{tr}(\rho(g_1 \cdots g_i)U_i)I$$

for c_1 a constant depending on $d, c, T, \alpha_0, \theta$ and A. We will achieve the latter by ensuring that each g_i is a product of sufficiently many γ_i so that $g_i x$ is in distribution sufficiently close to ν.

To combine the trace bounds at the various scales while ensuring that the above conditions are satisfied, a theory of decompositions of the form (8.1) will be developed. We call decompositions (8.1) satisfying suitable properties *proper decompositions*. It is important for our purposes to track the amount of variance we can gain from a given proper decomposition, which is a quantity we will call the variance sum and denote by $V(\mu, n, K, \kappa, A; r)$ (see definition 8.2 for the various parameters).

In section 8.2 we will show that there exist proper decompositions that allow us to compare the variance sum V and tr. Proper decompositions can be concatenated in such a way that variance sum is additive, as is shown in section 8.3. We establish how to convert an estimate on the variance sum V into an estimate for detail in section 8.4. The proof of Theorem 2.4 culminates in section 8.5 combining the previous results. Finally, we establish Theorem 2.5 in section 8.6.

8.1. **Proper Decompositions.**

Definition 8.1. *Let μ be a probability measure on G, let $n, K \in \mathbb{Z}_{\geq 0}$ and let $A, r > 0$ and $r \in (0,1)$. Then a* ***proper decomposition*** *of (μ, n, K, A) at scale r consists of the following data*

- *(i) $f = (f_i)_{i=1}^n$ and $h = (h_i)_{i=1}^n$ random variables taking values in G,*
- *(ii) $U = (U_i)_{i=1}^n$ random variables taking values in $\mathfrak{g}$,*
- *(iii) $\mathscr{A}_0 \subset \mathscr{A}_1 \subset \ldots \subset \mathscr{A}_n$ a nested sequence of σ-algebras,*
- *(iv) $\gamma = (\gamma_i)_{i=1}^\infty$ be i.i.d. samples from μ and let $\mathscr{F} = (\mathscr{F}_i)_{i=1}^\infty$ be a filtration for γ with γ_{i+1} being independent from $\mathscr{F}_i$ for $i \geq 1$,*
- *(v) stopping times $S = (S_i)_{i=1}^n$ and $T = (T_i)_{i=1}^n$ for the filtration $\mathscr{F}$,*
- *(vi) $m = (m_i)_{i=1}^n$ non-negative real numbers,*

satisfying the following properties:

A1 *The stopping times satisfy*

$$S_1 \leq T_1 \leq S_2 \leq T_2 \leq \ldots \leq S_n \leq T_n,$$

$S_1 \geq K$ as well as $S_i \geq T_{i-1} + K$ and $T_i \geq S_i + K$ for $i \in [n]$,

A2 *We have $f_1 \exp(U_1) = \gamma_1 \ldots \gamma_{S_1}$ and for $2 \leq i \leq n$ we have $f_i \exp(U_i) = \gamma_{T_{i-1}+1} \cdots \gamma_{S_i}$. Furthermore for each i we have that f_i is $\mathscr{A}_i$-measurable,*

A3 *$h_i = \gamma_{S_i+1} \cdots \gamma_{T_i}$ and h_i is $\mathscr{A}_i$-measurable,*

A4 *$\rho(f_i) < 1$ for all $1 \leq i \leq n$,*

A5 *Whenever $|b(h_i)| > A$, we have $U_i = 0$,*

A6 *For each $1 \leq i \leq n$ we have*

$$|U_i| \leq \rho(f_1 h_1 f_2 h_2 \cdots h_{i-1} f_i)^{-1} r,$$

A7 *For each $1 \leq i \leq n$, we have that U_i is conditionally independent of $\mathscr{A}_n$ given $\mathscr{A}_i$,*

A8 *The U_i are conditionally independent given $\mathscr{A}_n$,*

A9 *For each $1 \leq i \leq n$, it holds*

$$\mathbb{E}\left[\frac{\mathrm{Var}(\rho(f_i)U(f_i)U_i b(h_i)|\mathscr{A}_i)}{\rho(f_1 h_1 f_2 h_2 \cdots f_{i-1} h_{i-1})^{-2} r^2} \mid \mathscr{A}_{i-1}\right] \geq m_i I.$$

Note that in **A9** by Var we mean the covariance matrix and we are using the ordering given by positive semi-definiteness (2.16) and we denote as in section 3.1 by $U_i b(h_i) = \psi_{b(h_i)}(U_i)$.

A proper decomposition as above gives us

$$\gamma_1 \cdots \gamma_{T_n} = f_1 \exp(U_1) h_1 f_2 \exp(U_2) h_2 \cdots h_{n-1} f_n \exp(U_n) h_n \tag{8.3}$$

We briefly comment on the various properties of proper decompositions. We use parameter K and **A1** to ensure that each of the $f_i x$ and $h_i x$ for $x \in \mathbb{R}^d$ are close in distribution to ν. Properties **A4**, **A5** and **A6** are needed in order to apply Proposition 3.4. We require **A7** so that we have $\mathrm{Var}(U_i|\mathscr{A}_n) = \mathrm{Var}(U_i|\mathscr{A}_i)$ and in particular the latter is a $\mathscr{A}_i$-measurable random variable. **A8** is needed so that $[U_1|\mathscr{A}_n], \dots, [U_n|\mathscr{A}_n]$ are independent random variables and therefore we can apply Proposition 5.8.

One works with two sequences of random variables f and h instead of one in order to be able to concatenate proper decompositions as in Proposition 8.4. Indeed, if we had proper decompositions of the form

$$\gamma_1 \cdots \gamma_{T_n} = g_1 \exp(U_1) g_2 \exp(U_2) g_3 \cdots g_n f \exp(U_n) g_{n+1}$$

we could show a variant of (8.6) and all other results on proper decompositions. However we could not prove anything like Proposition 8.4, whose flexible choice of the parameter M is necessary to apply Proposition 7.5.

We next define the V function mentioned above. The additional parameter $\kappa > 0$ is introduced in order to be able to concatenate the decompositions in a suitable way (Proposition 8.4).

Definition 8.2. *Given (μ, n, K, A) and $\kappa, r > 0$ we denote by*

$$V(\mu, n, K, \kappa, A; r)$$

*the **variance sum** defined as the supremum for $k = 0, 1, 2, \dots, n$ of all possible values of*

$$\sum_{i=1}^{k} m_i$$

for a proper decomposition of (μ, k, K, A) at scale r with $\rho(f_1 h_1 \cdots f_k h_k) \geq \kappa$ almost surely.

It is clear that for any $\kappa' > 0$ with $\kappa' \leq \kappa$ we have

$$V(\mu, n, K, \kappa', A; r) \geq V(\mu, n, K, \kappa, A; r). \tag{8.4}$$

8.2. **Existence of Proper Decompositions.** We show that for a suitable dependence of the involved parameters, we can construct proper decompositions comparing the variance sum and the trace.

Proposition 8.3. *Let $d \in \mathbb{Z}_{\geq 1}$ and $c, T, \alpha_0, \theta, A, R > 0$ with $c, \alpha_0 \in (0,1)$ and $T \geq 1$. Then there exists $c_1 = c_1(d, R, c, T, \alpha_0, \theta, A) > 0$ such that the following is true. Let μ be a contracting on average, (c,T)-well-mixing and (α_0, θ, A)-non-degenerate probability measure on G such that $\rho(g) \in [R^{-1}, R]$ for all $g \in \mathrm{supp}(\mu)$.*

Let $\kappa, s > 0$ with κ and s sufficiently small (in terms of μ and R). Let K be sufficiently large in terms of μ, R, and T. Then

$$V(\mu, 1, K, R^{-3K}\kappa, A; R^{-K}\kappa s) \geq c_1 \operatorname{tr}(q_{\tau_\kappa}; s).$$

Proof. We construct a proper decomposition with $n = 1$. Let F be uniform on $[0, T] \cap \mathbb{Z}$ and independent of γ. Let $\underline{S}$ be defined as

$$\underline{S} = \inf\{n : \rho(q_n) \leq R^{-K-1}\} + F$$

and let

$$S_1 := \inf\{n \geq \underline{S} : \rho(\gamma_{\underline{S}+1} \cdots \gamma_n) \leq \kappa\}.$$

Denote

$$\underline{f} = \gamma_1 \cdots \gamma_{\underline{S}} \qquad \text{and} \qquad g = \gamma_{\underline{S}+1}\gamma_{\underline{S}+2} \cdots \gamma_{S_1}.$$

By the definition of $\operatorname{tr}(q_{\tau_\kappa}, s)$ there is some σ-algebra $\mathscr{A}$, some random variable V taking values in $\mathfrak{g}$, some $\mathscr{A}$-measurable random variable $\overline{f}$ taking values in G such that $g = \overline{f}\exp(V)$ with $|V| \leq s$ and

$$\mathbb{E}[\operatorname{tr}(V|\mathscr{A})] \geq \frac{1}{2}s^2 \operatorname{tr}(q_{\tau_\kappa}, s). \tag{8.5}$$

We define $T_1 = S_1 + K$ and set

$$h_1 = \gamma_{S_1+1}\gamma_{S_1+2} \cdots \gamma_{T_1}.$$

Denote

$$U_1 = \begin{cases} V & \text{if } |b(h_1)| \leq A, \\ 0 & \text{otherwise} \end{cases} \qquad \text{and} \qquad f_1 = \begin{cases} \underline{f}\overline{f} & \text{if } |b(h_1)| \leq A, \\ \underline{f}g & \text{otherwise.} \end{cases}$$

Furthermore we set $\mathscr{A}_1 = \sigma(\underline{f}, f_1, h_1, \mathscr{A})$.

We have

$$R^{-K-2}R^{-T}\kappa \leq \rho(\underline{f}g) \leq R^{-K-1}\kappa.$$

In particular, we note that $|U_1| \leq s$ and so providing κ and s are sufficiently small in terms of R, we have $R^{-K-3}R^{-T}\kappa \leq \rho(f_1) \leq R^{-K}\kappa < 1$. This means that $|U_1| \leq s \leq \rho(f_1)^{-1}R^{-K}\kappa s$.

Now let $x \in \mathbb{R}^d$ be a unit vector. We wish to show that

$$\mathbb{E}\left[\operatorname{Var}(x \cdot \rho(f_1)U(f_1)U_1 b(h_1)|\mathscr{A}_1)\right] \geq c_1 \operatorname{tr}(q_{\tau_\kappa}; s) R^{-2K}\kappa^2 s^2.$$

Let $f' = \underline{f}^{-1}f_1$ and let $P_1, \ldots, P_d$ be orthogonal eigenvectors of the covariance matrix of $(U_1 b(h_1)|\mathscr{A})$ with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_d$. We have

$$\begin{aligned}
&\operatorname{Var}(x \cdot \rho(f_1)U(f_1)U_1 b(h_1)|\mathscr{A}_1) \\
&\quad \geq R^{-2K-6}R^{-2T}\kappa^2 \operatorname{Var}(x \cdot U(\underline{f})U(f')U_1 b(h_1)|\mathscr{A}_1) \\
&\quad = R^{-2K-6}R^{-2T}\kappa^2 \sum_{i=1}^{d} \left|x \cdot U(\underline{f})U(f')P_i\right|^2 \lambda_i \\
&\quad \geq R^{-2K-6}R^{-2T}\kappa^2 \left|x \cdot U(\underline{f})U(f')P_1\right|^2 \operatorname{tr}(U_1 b(h_1)|\mathscr{A}_1)/d.
\end{aligned}$$

By Proposition 3.2 we know that when $b(h_1) \in U_\theta(V)$ and $|b(h_1)| \leq A$ we have

$$\operatorname{tr}(U_1 b(h_1)|\mathscr{A}_1) \geq \delta \cdot \operatorname{tr}(U_1|\mathscr{A}_1) = \delta \cdot \operatorname{tr}(U_1|\mathscr{A}_1).$$

By our (α_0, θ, A)-non-degeneracy condition and Lemma 4.2 we know that providing K is sufficiently large this happens, conditional on $\mathscr{A}$, with probability at least $\frac{1}{2}(1-\alpha)$. Therefore by (8.5)

$$\mathbb{E}[\operatorname{tr}(U_1 b(h_1)|\mathscr{A}_1)] \geq \frac{1}{4}(1-\alpha)\delta \operatorname{tr}(q_{\tau_\kappa}; s)s^2.$$

By our (c, T)-well-mixing condition we have that providing K is sufficiently large in terms of μ,

$$\mathbb{E}\left[\left|x \cdot U(\underline{f})U(f')P_1\right|^2 |\mathscr{A}_1\right] \geq c.$$

Clearly $\operatorname{Var}(U_1 z(h_1)|\mathscr{A}_1)$ is $\sigma(h_1, \mathscr{A})$-measurable. Therefore

$$\begin{aligned}
&\mathbb{E}\left[\operatorname{Var}(x \cdot \rho(f_1)U(f_1)U_1 b(h_1)|\mathscr{A}_1)\right] \\
&\geq \mathbb{E}\left[R^{-2K-6}R^{-2T}\kappa^2 \left|x \cdot U(\underline{f})U(f')P_1\right|^2 \operatorname{tr}(U_1 b(h_1)|\mathscr{A}_1)/d\right] \\
&\geq R^{-2K-6}R^{-2T}d^{-1} \cdot \frac{1}{4}(1-\alpha)\delta \operatorname{tr}(q_{\tau_\kappa}; s)\kappa^2 s^2 \cdot c \\
&= c_1 \operatorname{tr}(q_{\tau_\kappa}; s)R^{-2K}\kappa^2 s^2
\end{aligned}$$

where $c_1 = R^{-6}R^{-2T}d^{-1}(1-\alpha)\delta c/4$. Since this is true for any unit vector $x \in \mathbb{R}^d$ we have

$$\mathbb{E}\left[\frac{\operatorname{Var}(\rho(f_1)U(f_1)U_1 b(h_1)|\mathscr{A}_1)}{R^{-2K}\kappa^2 s^2}\right] \geq c_1 \operatorname{tr}(q_{\tau_\kappa}; s)$$

as required. Finally note that

$$\rho(f_1 h_1) \geq R^{-1}\rho(\underline{f}gh_1) \geq R^{-1}R^{-T}R^{-K-1} \cdot \kappa R^{-1} \cdot R^{-K} = \kappa R^{-2K-3-T} \geq R^{-3K}\kappa$$

providing K is sufficiently large in terms of T and R. □

8.3. **Concatenating Decompositions.** We note that it is straightforward to show that for any measure μ and any admissible choice of coefficients, the variance sum is additive

$$\begin{aligned}
&V(\mu, n_1 + n_2, K, \kappa_1\kappa_2, A; r) \\
&\qquad \geq V(\mu, n_1, K, \kappa_1, A; r) + V(\mu, n_2, K, \kappa_2, A; \kappa_1^{-1}r).
\end{aligned} \tag{8.6}$$

However, in order to use Proposition 7.5 it is necessary to work with different scales r_1 and r_2 and therefore we show the following proposition.

Proposition 8.4. *Let μ be a probability measure on G. Let $R > 1$ be such that $\rho(g) \in [R^{-1}, R]$ for every $g \in \operatorname{supp}(\mu)$. Let $n_1, n_2, K \in \mathbb{Z}_{\geq 0}$ with $n_2, K > 0$ and let $\kappa_1, \kappa_2, r \in (0,1)$. Let $A > 0$ and let $M \geq R$. Then*

$$\begin{aligned}
&V(\mu, n_1 + n_2, K, R^{-1}M^{-1}\kappa_1\kappa_2, A; r) \\
&\qquad \geq V(\mu, n_1, K, \kappa_1, A; r) + V(\mu, n_2, K, \kappa_2, A; M\kappa_1^{-1}r).
\end{aligned}$$

Proof. For $j \in \{1, 2\}$ let $\gamma_1^{(j)}, \gamma_2^{(j)}, \ldots$ be a sequence of i.i.d. samples from μ defined on the probability space $\left(\Omega_{(j)}, \mathscr{F}_{(j)}, \mathbb{P}_{(j)}\right)$. Let $\hat{\gamma}_1, \hat{\gamma}_2, \ldots$ be a sequence of i.i.d. samples from μ defined on the probability space $\left(\hat{\Omega}, \hat{\mathscr{F}}, \hat{\mathbb{P}}\right)$. Consider the product probability space

$$(\Omega, \mathscr{F}, \mathbb{P}) = \left(\Omega_1 \times \hat{\Omega} \times \Omega_2, \mathscr{F}_1 \times \hat{\mathscr{F}} \times \mathscr{F}_2, \mathbb{P}_1 \times \hat{\mathbb{P}} \times \mathbb{P}_2\right).$$

Let $\left(\gamma_i^{(1)}, S_i^{(1)}, T_i^{(1)}, f_i^{(1)}, U_i^{(1)}, h_i^{(1)}, \mathscr{A}_i^{(1)}, m_i^{(1)}\right)$ be a proper decomposition for $(\mu, k_1, K, \kappa_1, A)$ at scale r defined on the probability space $\left(\Omega^{(1)}, \mathscr{F}^{(1)}, \mathbb{P}^{(1)}\right)$ such that $\sum_{i=1}^{k_1} m_i^{(1)}$ approaches $V(\mu, n_1, K, \kappa_1, A; r)$ and

$$\rho(f_1^{(1)} h_1^{(1)} \cdots f_{k_1}^{(1)} h_{k_1}^{(1)}) \geq \kappa_1.$$

Given $\omega_1 \in \Omega_1$ and $\hat{\omega} \in \hat{\Omega}$, let $\tau = \tau(\omega_1, \hat{\omega})$ be given by

$$\tau = \min\{k \in \mathbb{Z}_{\geq 0} : \rho(f_1^{(1)} h_1^{(1)} f_2^{(1)} h_2^{(1)} \ldots f_{k_1}^{(1)} h_{k_1}^{(1)} \hat{\gamma}_1 \hat{\gamma}_2 \ldots \hat{\gamma}_k) < M^{-1}\kappa_1\}$$

and let $\hat{\rho} = \rho(f_1^{(1)} h_1^{(1)} f_2^{(1)} h_2^{(1)} \ldots f_{k_1}^{(1)} h_{k_1}^{(1)} \hat{\gamma}_1 \hat{\gamma}_2 \ldots \hat{\gamma}_\tau)$ such that

$$\hat{\rho} \in [M^{-1}R^{-1}\kappa_1, M^{-1}\kappa_1].$$

Now given $\omega_1 \in \Omega_1$ and $\hat{\omega} \in \hat{\Omega}$, let $\left(\gamma_i^{(2)}, S_i^{(2)}, T_i^{(2)}, f_i^{(2)}, U_i^{(2)}, h_i^{(2)}, \mathscr{A}_i^{(2)}, m_i^{(2)}\right)$ be a proper decomposition for $(\mu, k_2, K, \kappa_2, A)$ at scale $M\kappa_1^{-1}r$ defined on the probability space $\left(\Omega^{(2)}, \mathscr{F}^{(2)}, \mathbb{P}^{(2)}\right)$ such that $\sum_{i=1}^{k_2} m_i^{(2)}$ approaches $V(\mu, n_2, K, \kappa_2, A; M\kappa_1^{-1}r)$ and

$$\rho(f_1^{(1)} h_1^{(1)} \cdots f_{k_2}^{(1)} h_{k_2}^{(1)}) \geq \kappa_2.$$

We now concatenate the two decompositions as follows. Let $\gamma_1, \gamma_2, \ldots$ be the sequence of random variables on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ defined by

$$\gamma_i = \begin{cases} \gamma_i^{(1)} & \text{if } i \leq T_{k_1}^{(1)} \\ \hat{\gamma}_{i - T_{k_1}^{(1)}} & \text{if } i > T_{k_1}^{(1)} \text{ and } i \leq T_{k_1}^{(1)} + \tau \\ \gamma_{i - T_{k_1}^{(1)} - \tau} & \text{if } i > T_{k_1}^{(1)} + \tau. \end{cases}$$

Clearly these are i.i.d. samples from μ. For $i = 1, 2, \ldots, k_1 + k_2$ we define S_i by

$$S_i = \begin{cases} S_i^{(1)} & \text{if } i \leq k_1 \\ S_{i-k_1}^{(2)} + T_{k_1}^{(1)} + \tau & \text{if } i > k_1 \end{cases}$$

and we define T_i analogously. We define f_i by

$$f_i = \begin{cases} f_i^{(1)} & \text{if } i \leq k_1 \\ \hat{\gamma}_1 \ldots \hat{\gamma}_\tau f_1^{(2)} & \text{if } i = k_1 + 1 \\ f_{i-k_1}^{(2)} & \text{if } i > k_1 + 1. \end{cases}$$

We define U_i by

$$U_i = \begin{cases} U_i^{(1)} & \text{if } i \leq k_1 \\ U_{i-k_1}^{(2)} & \text{if } i > k_1. \end{cases}$$

and define h_i and m_i analogously. Finally we define $\mathscr{A}_i$ by

$$\mathscr{A}_i = \begin{cases} \mathscr{A}_i^{(1)} \times \hat{\Omega} \times \Omega^{(2)} & \text{if } i \leq k_1 \\ \mathscr{A}_{k_1}^{(1)} \times \hat{\mathscr{F}} \times \mathscr{A}_{i-k_1}^{(2)} & \text{if } i > k_1. \end{cases}$$

It is easy to check that $(\gamma_i, S_i, T_i, f_i, U_i, h_i, \mathscr{A}_i, m_i)$ is a proper decomposition for $(\mu, R, k_1 + k_2, K, R^{-1}M^{-1}\kappa_1\kappa_2, A)$ at scale r and it holds that

$$\sum_{i=1}^{k_1+k_2} m_i = \sum_{i=1}^{k_1} m_i^{(1)} + \sum_{i=1}^{k_2} m_i^{(2)}.$$

Indeed, we note that for $i > k_2$ we have that since $M\kappa_1^{-1} \leq \hat{\rho}^{-1}$,

$$\begin{aligned}|U_i| = |U^{(2)}_{i-k_1}| &\leq \rho(f_1^{(2)} h_1^{(2)} f_2^{(2)} h_2^{(2)} \cdots h^{(2)}_{i-k_1-1} f^{(2)}_{i-k_1})^{-1} M\kappa_1^{-1} r \\ &\leq \hat{\rho}^{-1} \rho(f_1^{(2)} h_1^{(2)} f_2^{(2)} h_2^{(2)} \cdots h^{(2)}_{i-k_1-1} f^{(2)}_{i-k_1})^{-1} r \\ &= \rho(f_1 h_1 f_2 h_2 \cdots h_{i-1} f_i)^{-1} r.\end{aligned}$$

Similarly, for $i > k_2 + 1$ and using that $\hat{\rho}^2 M^2 \kappa_1^{-2} \leq 1$,

$$\begin{aligned}&\mathbb{E}\left[\frac{\mathrm{Var}(\rho(f_i)U(f_i)U_i b(h_i)|\mathscr{A}_i)}{\rho(f_1 h_1 f_2 h_2 \cdots f_{i-1} h_{i-1})^{-2} r^2} \mid \mathscr{A}_{i-1}\right] \\ &\quad = \mathbb{E}\left[\frac{\mathrm{Var}(\rho(f^{(2)}_{i-k_1})U(f^{(2)}_{i-k_1})U^{(2)}_{i-k_1} b(h^{(2)}_{i-k_1})|\mathscr{A}_i)}{\hat{\rho}^{-2}\rho(f_1^{(2)} h_1^{(2)} f_2^{(2)} h_2^{(2)} \cdots h^{(2)}_{i-k_1})^{-2} r^2} \mid \mathscr{A}_{i-1}\right] \\ &\quad \geq \mathbb{E}\left[\frac{\mathrm{Var}(\rho(f^{(2)}_{i-k_1})U(f^{(2)}_{i-k_1})U^{(2)}_{i-k_1} b(h^{(2)}_{i-k_1})|\mathscr{A}_i)}{\hat{\rho}^{-2}\rho(f_1^{(2)} h_1^{(2)} f_2^{(2)} h_2^{(2)} \cdots h^{(2)}_{i-k_1})^{-2} \hat{\rho}^2 M^2 \kappa_1^{-2} r^2} \mid \mathscr{A}_{i-1}\right] \\ &\quad \geq m^{(2)}_{i-k_1} I.\end{aligned}$$

The remainder of the properties are straightforward to check. □

Corollary 8.5. *Let μ be a probability measure on G. Let $R > 1$ be such that $\rho(g) \in [R^{-1}, R]$ for every $g \in \mathrm{supp}(\mu)$. Let $n, K \in \mathbb{Z}_{>0}$ and let $\kappa, r \in (0,1)$. Let $C, A > 0$ and let $M \geq R$. Then*

$$V(\mu, n, K, R^{-1}M^{-1}\kappa, A, C; M^{-1}r) \geq V(\mu, n, K, \kappa, A, C; r)$$

Proof. By Proposition 8.4 we have

$$\begin{aligned}&V(\mu, n, K, R^{-1}M^{-1}\kappa, A; M^{-1}r) \\ &\qquad \geq V(\mu, 0, K, 1, A; M^{-1}r) + V(\mu, n, K, \kappa, A; r).\end{aligned}$$

and simply note that $V(\mu, 0, K, 1, A, C; M^{-1}r) = 0$. □

8.4. **From Variance Sum to Bounding Detail.**

Proposition 8.6. *For every $d \geq 1$ and $A, \alpha > 0$ there is a constants $C = C(d, A, \alpha) > 0$ such that the following is true. Suppose that μ is a contracting on average probability measure on G. Then there is some $c = c(\mu) > 0$ such that whenever $\kappa \leq 1$ and $k, K, n \in \mathbb{Z}_{>0}$ with K and n sufficiently large (in terms of A, α and μ) and $r > 0$ is sufficiently small (in terms of A, α and μ) and*

$$V(\mu, R, n, K, \kappa, A; r) > Ck$$

we have

$$s_r^{(k)}(\nu) < \alpha^k + n\exp(-cK) + C^n \kappa^{-1} r.$$

Proof. Suppose that $(f, h, U, \mathscr{A}, \gamma, \mathscr{F}, S, T, m)$ is a proper decomposition of (μ, n, K, A) at scale r such that $\sum_{i=1}^n m_i \geq C/2$ and let v be an independent sample from ν. Let

$$I = \{i \in [1, n] \cap \mathbb{Z} : |b(h_i)| \leq A\}$$

and let $m = |I|$. Enumerate I as $i_1 < i_2 < \cdots < i_m$ and define $g_1, \ldots, g_m$ by $g_1 = f_1 h_1 \ldots f_{i_1}$ and $g_j = h_{i_{j-1}} f_{i_{j-1}+1} \ldots f_{i_j}$ for $2 \leq j \leq m$. Define $\overline{v}$ by $\overline{v} = h_{i_m} f_{i_m+1} \ldots h_n v$ and let $V_j = U_{i_j}$. Let x be defined by

$$x = g_1 \exp(V_1) \ldots g_m \exp(V_m) \overline{v}.$$

Note that x is a sample from ν. Let $\hat{\mathscr{A}}$ be the σ-algebra generated by $\mathscr{A}_n$ and v. Note that the g_j and $\overline{v}$ are $\hat{\mathscr{A}}$-measurable.

We will bound the order k detail of x by showing that with high probability we can apply Proposition 3.4 to $g_1, \dots, g_m$, $V_1, \dots, V_m$, and $\overline{v}$ and then bound the order k detail of this using Proposition 5.8.

Let E be the event that $|\overline{v}| \leq 2A$ and that for each $j = 1, \dots, m$ we have $|b(g_j)| \leq 2A$, $\rho(g_j) < 1$ and $|V_j| \leq \rho(g_1 \dots g_j)^{-1} r$. By Corollary 3.12 we know that $\mathbb{P}[E^C] \leq \exp(-c_1 K)$ for some $c_1 = c_1(\mu, A) > 0$.

For $j = 1, \dots, m$ define ζ_j by

$$\zeta_j = D_u(g_1 \cdots g_j \exp(u) g_{j+1} \cdots g_m \overline{v})|_{u=0}.$$

By Proposition 3.4 on E we have

$$\left| x - g_1 \dots g_m \overline{v} - \sum_{j=1}^{m} \zeta_j(V_j) \right| \leq C_1^m \rho(g_1 \dots g_m)^{-1} r^2$$

for some $C_1 = C_1(A) > 0$. Clearly the right hand side is at most $C_1^n \kappa^{-1} r^2$. By Lemma 5.7 this means that on E we have

$$s_r^{(k)}(x|\hat{\mathscr{A}}) \leq s_r^{(k)} \left(\sum_{j=1}^{m} \zeta_j(V_j) | \hat{\mathscr{A}} \right) + C_1^n e d \kappa^{-1} r$$

where e is Euler's number.

Let $C_3 = C_3(\alpha, d)$ be the constant C from Proposition 5.8 with the same values of α and d and let F be the event that

$$\sum_{j=1}^{m} \operatorname{Var} \zeta_j(V_j | \hat{\mathscr{A}}) \geq k C_3 I.$$

By Proposition 5.8, using that by **A8** the $[V_1|\hat{\mathscr{A}}], \dots, [V_m|\hat{\mathscr{A}}]$ are independent almost surely, we have that on F

$$s_r^{(k)} \left(\sum_{j=1}^{m} \zeta_j(V_j) | \hat{\mathscr{A}} \right) \leq \alpha^k.$$

Therefore

$$s_r^{(k)}(x|\hat{\mathscr{A}}) \leq \alpha^k + C_1^n e d \kappa^{-1} r + \mathbb{I}_{E^C \cup F^C}$$

and so by the convexity of order k detail we have

$$s_r^{(k)}(x) \leq \alpha^k + C_1^n e d \kappa^{-1} r^2 + \mathbb{P}[E^C] + \mathbb{P}[F^C].$$

We already have that $\mathbb{P}[E^C] \leq \exp(-c_1 K)$ so it only remains to bound $\mathbb{P}[F^C]$.

For $i = 1, \dots, n$ define

$$\hat{\zeta}_i = D_u(f_1 h_1 \cdots h_{i-1} f_i \exp(u) b(h_i))|_{u=0}$$

and let $\underline{F}$ be the event that

$$\left\| \sum_{i=1}^{n} \operatorname{Var} \hat{\zeta}_i(U_i|\hat{\mathscr{A}}) - \sum_{j=1}^{m} \operatorname{Var} \zeta_j(V_j|\hat{\mathscr{A}}) \right\| < 1.$$

Let $C_3 = C_3(\alpha, d)$ be the constant C from Proposition 5.8 with the same values of α and d and let $\overline{F}$ be the $\hat{\mathscr{A}}$-measurable event that $\sum_{i=1}^{n} \operatorname{Var}(\hat{\zeta}_i(U_i)|\hat{\mathscr{A}}) \geq (C_3+1)kIr^2$. Clearly $\underline{F} \cup \overline{F} \subset F$ so it suffices to bound $\mathbb{P}[\underline{F}^C]$ and $\mathbb{P}[\overline{F}^C]$.

Since $g_1, \dots, g_m$ and $\overline{v}$ are $\hat{\mathscr{A}}$ measurable, by Lemma 3.3 we have for $j = 1, \dots, m$ that $\operatorname{Var}(\zeta_j(V_j)|\hat{\mathscr{A}})$ is equal to

$$\rho(g_1 \dots g_j)^2 \cdot U(g_1 \dots g_j)\psi_{g_{j+1}\cdots g_m \overline{v}} \circ \operatorname{Var}(V_j|\hat{\mathscr{A}}) \circ \psi^T_{g_{j+1}\cdots g_m} U(g_1 \dots g_j)^T$$

and that

$$\operatorname{Var}(\hat{\zeta}_{i_j}(U_{i_j})|\hat{\mathscr{A}}) = \rho(g_1 \cdots g_j)^2 \cdot U(g_1 \dots g_j)\psi_{b(h_{i_j})} \circ \operatorname{Var}(V_j|\hat{\mathscr{A}}) \circ \psi^T_{b(h_{i_j})} U(g_1 \dots g_j)^T.$$

We also have that $|V_j| \leq \rho(g_1 \cdots g_j)^{-1} r$ almost surely and so consequently $\| \operatorname{Var} V_j \| \leq \rho(g_1 \cdots g_j)^{-2} r^2$. Therefore by Lemma 3.1 (iii),

$$\| \operatorname{Var} \zeta_j(V_i|\hat{\mathscr{A}}) - \operatorname{Var} \hat{\zeta}_{i_j}(U_{i_j}|\hat{\mathscr{A}})\| \ll_d |b(h_j) - g_{j+1} \dots g_m \overline{v}|^2 r^2.$$

Furthermore we have that whenever $i \notin I$ that $\operatorname{Var}(\hat{\zeta}_i(U_i)|\hat{\mathscr{A}}) = 0$. We may assume without loss of generality that $n \exp(-K\chi_\mu/10) < 1$. This means that, providing K is sufficiently large (in terms of d), in order for $\underline{F}$ to occur it is sufficient that for each $j = 1, \dots, m$ we have

$$|b(h_j) - g_{j+1} \dots g_m \overline{v}| < \exp(-K\chi_\mu/10) < 1/n.$$

By Corollary 3.12 this occurs with probability at least $1 - m\exp(-c_2 K)$ for some $c_2 = c_2(\mu) > 0$ and therefore $\mathbb{P}[\underline{F}^C] \leq m \exp(-c_2 K) \leq n \exp(-c_2 K)$.

Finally we wish to bound $\mathbb{P}[\overline{F}^C]$. Let

$$\begin{aligned} \Sigma_i &= r^{-2} \operatorname{Var}(\hat{\zeta}_i(U_i)|\hat{\mathscr{A}}) = r^{-2} \operatorname{Var}(\hat{\zeta}_i(U_i)|\mathscr{A}_i) \\ &= r^{-2} \operatorname{Var}(\rho(f_1 h_1 \cdots h_{i-1} f_i) U(f_1 h_1 \cdots h_{i-1} f_i) U_i b(h_i)|\mathscr{A}_i)) \end{aligned}$$

By construction we know that

$$\mathbb{E}[\Sigma_i | \Sigma_1, \dots, \Sigma_{i-1}] \geq m_i I.$$

We also know that $\|\Sigma_i\| \leq A^2$ since $||\psi_{b(h_i)}|| \leq |b(h_i)| \leq A$. This means that we can apply Lemma 3.14. By Lemma 3.14 we know that providing C is sufficiently large we have

$$\mathbb{P}\left[\sum_{i=1}^{n} \Sigma_i \geq (C_3+1)kI\right] \geq 1 - \exp\left(-c_3 k \sum_{i=1}^{n} m_i\right)$$

for some absolute $c_3 > 0$. Providing we choose C to be sufficiently large, we therefore have $\mathbb{P}[\overline{F}^C] \leq \exp(-c_3 k C) \leq \alpha^k$ this is less than α^k.

Putting everything together we have

$$s_r^{(k)}(x) \leq 2\alpha^k + n \exp(-c_3 K) + edC_1^n \kappa^{-1} r.$$

Replacing α be a slightly smaller value gives the required result. □

8.5. **Conclusion of Proof of Theorem 2.4.** We finally show a decay in detail under the assumption of Theorem 2.4. What follows is a rather intricate calculation and we refer the reader to the outline of proofs in section 2.2 for intuition and a sketch of the argument.

Proposition 8.7. *Let $d \in \mathbb{Z}_{\geq 1}$ and $c, T, \alpha_0, \theta, A, R > 0$ with $c, \alpha \in (0,1)$ and $T \geq 1$. Then there exists $C = C(d, R, c, T, \alpha_0, \theta, A) > 0$ such that the following is true. Let μ be a contracting on average, (c,T)-well-mixing and (α_0, θ, A)-non-degenerate probability measure on G with $\rho(g) \in [R^{-1}, R]$ for all $g \in \operatorname{supp}(\mu)$ and assume that*

$$\frac{h_\mu}{|\chi_\mu|} > C \max\left\{1, \left(\log \frac{S_\mu}{h_\mu}\right)^2\right\}.$$

Then for all sufficiently small $r > 0$ and all integers $k \in [\log\log r^{-1}, 2\log\log r^{-1}]$ we have that

$$s_r^{(k)}(\nu) < (\log r^{-1})^{-10d}.$$

Proof. We prove this by repeatedly applying Proposition 8.3 and Proposition 8.4 and then applying Proposition 8.6. First let C be as in Proposition 8.6 with $\alpha = \exp(-20d)$.

Now let $r > 0$ be sufficiently small and let $K = \exp(\sqrt{\log\log r^{-1}})$. This value of K is chosen so that K grows more slowly than $(\log r^{-1})^{\varepsilon}$ but faster than any polynomial in $\log\log r^{-1}$ as $r \to 0$. Let $S = 2\max\{h_\mu, S_\mu\}$.

Note that $\frac{h_\mu}{2\ell S} < 1$ and for $i = 1, 2, \dots$ let

$$\kappa_i = \exp\left(-\frac{|\chi_\mu| \log r^{-1}}{2S}\left(\frac{h_\mu}{3\ell S}\right)^{i-1}\right) = r^{\frac{|\chi_\mu|}{2S}(\frac{h_\mu}{3\ell S})^{i-1}}$$

with $\ell = \dim G$. Then

$$\kappa_1 = r^{\frac{|\chi_\mu|}{2S}} \quad \text{and} \quad \kappa_{i+1} = \kappa_i^{\frac{h_\mu}{3\ell S}}$$

and let m be chosen as large as possible such that

$$\kappa_m < \min\{R^{-10K}, 2^{-10K}\}.$$

We require $\kappa_m < R^{-10K}$ later in the proof and assume $\kappa_m < 2^{-10K}$ so that κ_m is surely sufficiently small when r is small enough so that we can apply Proposition 7.5. Note that this gives

$$\log\log R + \sqrt{\log\log r^{-1}} \ll \log\log r^{-1} + m \log \frac{h_\mu}{2\ell S} + \log \frac{\chi_\mu}{2S}$$

which is equivalent to

$$m \log\left(4\ell \max\left\{1, \frac{S_\mu}{h_\mu}\right\}\right) = m \log \frac{2\ell S}{h_\mu} \ll_d \log\log r^{-1}$$

and therefore it follows that

$$\left(\max\left\{1, \log\frac{S_\mu}{h_\mu}\right\}\right)^{-1} \log\log r^{-1} \ll_d m \ll_d \log\log r^{-1}.$$

Now as in Proposition 7.5 let $\hat{m} = \lfloor \frac{S}{100|\chi_\mu|} \rfloor$. For each $i = 1, 2, \dots, m$ let $s_1^{(i)}, s_2^{(i)}, \dots, s_{\hat{m}}^{(i)} > 0$ be the s_i from Proposition 7.5 with κ_i in the role of κ. So

$s_j^{(i)} \in (\kappa_i^{\frac{S}{|\chi_\mu|}}, \kappa_i^{\frac{h_\mu}{2\ell|\chi_\mu|}})$. By Proposition 8.3 we have for each $j \in [\hat{m}]$,

$$V(\mu, 1, K, R^{-3K}\kappa_i, A; R^{-K}\kappa_i s_j^{(i)}) \geq c_1 \mathrm{tr}(q_{\tau_{\kappa_i}}; s_j^{(i)})$$

for some constant $c_1 = c_1(c, T, \alpha_0, \theta, A, R, d) > 0$. Therefore by Proposition 8.4 with $M = R^{-1_{\{\geq 2\}}(j)} R^{-3K}\kappa_i s_{j+1}^{(i)}/s_j^{(i)}$, where we denote $1_{\{\geq 2\}}(j) = 1$ whenever $j \geq 2$, we can prove inductively for $j = 2, 3, \ldots, \hat{m}$ that

$$V(\mu, j, K, R^{-1}R^{-3K}\kappa_i s_1^{(i)}/s_j^{(i)}, A; R^{-K}\kappa_i s_1^{(i)}) \geq c_1 \sum_{a=1}^{j} \mathrm{tr}(q_{\tau_{\kappa_i}}; s_j^{(i)}).$$

We have used here that $s_{j+1}^{(i)}/s_j^{(i)} \geq \kappa_i^{-3}$ and so $M \geq R^{-6K}\kappa_i^{-2} \geq R^{10K} \geq R$ since $\kappa_i < R^{-10K}$. By Proposition 7.5 and (8.4) we conclude that

$$V(\mu, \hat{m}, K, R^{-4K}\kappa_i s_1^{(i)}/s_{\hat{m}}^{(i)}, A; R^{-K}\kappa_i s_1^{(i)}) \geq c_2 \frac{h_\mu}{|\chi_\mu|} \max\left\{1, \log \frac{S_\mu}{h_\mu}\right\}^{-1}$$

for some constant $c_2 > 0$ depending on all of the parameters.

Note that for $i = 1, 2, \ldots, m-1$ when $h_\mu/|\chi_\mu|$ is sufficiently large we have

$$\begin{aligned} R^{-4K}\kappa_{i+1}s_1^{(i+1)}/s_{\hat{m}}^{(i)} &\geq R^{-4K}\kappa_{i+1}^{\frac{S}{|\chi_\mu|}+1}\kappa_i^{-\frac{h_\mu}{2\ell|\chi_\mu|}} \\ &\geq R^{-4K}\kappa_i^{\frac{h_\mu}{3\ell|\chi_\mu|}-\frac{h_\mu}{2\ell|\chi_\mu|}+\frac{h_\mu}{3\ell S}} \\ &\geq R^{-4K}\kappa_i^{-1} \geq R^{6K} \geq R. \end{aligned}$$

as $\kappa_{i+1} = \kappa_i^{\frac{h_\mu}{3\ell S}}$ and $\kappa_i < R^{-10K}$ and so we may repeatedly apply Proposition 8.4 with

$$M = R^{-1_{\{\geq 2\}}(i)} R^{-4K}\kappa_{i+1}s_1^{(i+1)}/s_{\hat{m}}^{(i)},$$

where we denote $1_{\{\geq 2\}}(i) = 1$ whenever $i \geq 2$, to inductively show for $i = 2, 3, \ldots, m$ that

$$\begin{aligned} &V(\mu, R, i\hat{m}, K, R^{-1}R^{-4K}\kappa_1 s_1^{(1)}/s_{\hat{m}}^{(i)}, A; R^{-K}\kappa_1 s_1^{(1)}) \\ &\quad \geq c_2 i \frac{h_\mu}{|\chi_\mu|} \max\left\{1, \log \frac{S_\mu}{h_\mu}\right\}^{-1}. \end{aligned}$$

This means using (8.4)

$$\begin{aligned} &V(\mu, R, m\hat{m}, K, R^{-5K}\kappa_1 s_1^{(1)}/s_{\hat{m}}^{(m)}, A; R^{-K}\kappa_1 s_1^{(1)}) \\ &\geq c_3 \frac{h_\mu}{|\chi_\mu|} \max\left\{1, \log \frac{S_\mu}{h_\mu}\right\}^{-2} \log\log r^{-1} \end{aligned}$$

for some constant $c_3 > 0$ depending on all of the parameters. Since

$$R^{-K}\kappa_1 s_1^{(1)} \geq R^{-K}\kappa_1^{\frac{S}{|\chi_\mu|}+1} = R^{-K} r^{\frac{1}{2}+\frac{|\chi_\mu|}{2S}} \geq R^{-K} r^{\frac{1}{2}+\frac{1}{4d}} \geq r$$

for r sufficiently small by Corollary 8.5 with $M = R^{-K}\kappa_1 s_1^{(1)} r^{-1} \geq R$

$$V(\mu, R, m\hat{m}, K, R^{-5K} r/s_{\hat{m}}^{(m)}, A; r) \geq c_3 \frac{h_\mu}{|\chi_\mu|} \max\left\{1, \log \frac{S_\mu}{h_\mu}\right\}^{-2} \log\log r^{-1}.$$

Note that $1/s_{\hat{m}}^{(m)} \geq \kappa_m^{-\frac{h_\mu}{2\ell|\chi_\mu|}}$ and so in particular providing $h_\mu/|\chi_\mu|$ is sufficiently large we have $R^{-5K} r/s_{\hat{m}}^{(m)} \geq R^K r$. By Proposition 8.6 provided

$$\frac{h_\mu}{|\chi_\mu|} \max\left\{1, \log \frac{S_\mu}{h_\mu}\right\}^{-2} \geq 2c_3^{-1} C$$

we deduce

$$s_r^{(k)}(\nu) \leq \exp(-20dk) + m\hat{m}\exp(-c_4 K) + R^{-K} C^{m\hat{m}}$$

for some constant $c_4 = c_4(\mu) > 0$ and $k \in [\log\log r^{-1}, 2\log\log r^{-1}]$. Since $m\hat{m} \ll_\mu \log\log r^{-1}$ it is easy to see that

$$m\hat{m}\exp(-c_4 K) + R^{-K} C^{m\hat{m}} < \left(\log r^{-1}\right)^{-20d}$$

whenever $r > 0$ is sufficiently small (in terms of μ). Since $k \geq \log\log r^{-1}$ we have that $\exp(-20dk) \leq \left(\log r^{-1}\right)^{-20d}$. Overall this means that provided $r > 0$ is sufficiently small (in terms of μ) we have

$$s_r^{(k)}(\nu) < \left(\log r^{-1}\right)^{-10d}. \qquad \square$$

We deduce the main theorem from Proposition 8.7.

Proof. (of Theorem 2.4) We combine Proposition 8.7 with Lemma 5.5. Given $r > 0$ sufficiently small, let $k = \frac{3}{2}\log\log r^{-1}$, $a = r/\sqrt{k}$ and $b = rk^k$.

Suppose that $s \in [a, b]$ and note that then $k \in [\log\log s^{-1}, 2\log\log s^{-1}]$ and $\frac{1}{2}\log r^{-1} < \log s^{-1}$ for r sufficiently small and therefore by Proposition 8.7

$$s_s^{(k)}(\nu) < (\log s^{-1})^{-10d} < 2^{10d}(\log r^{-1})^{-10d}.$$

By Lemma 5.5 it follows that

$$s_r(\nu) \leq Q'(d)^{k-1}(2^{10d}(\log r^{-1})^{-10d} + k^{-k}),$$

which is easily shown to be $\leq (\log r^{-1})^{-2}$ for r sufficiently small. Indeed, recall that $Q'(d) \leq ed^{-1/2} \leq e$ for all $d \geq 1$ and therefore $Q'(d)^k \leq (\log(r^{-1}))^e$.

This concludes the proof of the main theorem of this paper. $\square$

8.6. **Proof of Theorem 2.5.** In this section we show how to work with the entropy and separation rate on $O(d)$ instead of the one on G. Recall that for a measure μ on G the measure $U(\mu)$ on $O(d)$ is the pushforward of μ under the map $g \mapsto U(g)$. We then denote for a finitely supported μ by $h_{U(\mu)}$ and $S_{U(\mu)}$ the analogously defined Shannon entropy and separation rate of $U(\mu)$. As we show in section 10.2, when all of the coefficients of the matrices in $\mathrm{supp}(U(\mu))$ lie in the number field K and have logarithmic height at most $L \geq 1$, then

$$S_{U(\mu)} \ll_d L[K : \mathbb{Q}].$$

Therefore Theorem 2.5 follows from Theorem 8.8.

Theorem 8.8. *Let $d \geq 3$ and $R, c, T, \alpha_0, \theta, A > 0$ with $c, \alpha_0 \in (0,1)$ and $T \geq 1$. Then there is a constant $C = C(d, R, c, T, \alpha_0, \theta, A)$ such that the following holds. Let μ be a finitely supported, contracting on average, (c, T)-well-mixing and (α_0, θ, A)-non-degenerate probability measure on G with* $\mathrm{supp}(\mu) \subset \{g \in G : \rho(g) \in [R^{-1}, R]\}$*. Then ν is absolutely continuous if*

$$\frac{h_{U(\mu)}}{|\chi_\mu|} \geq C \max\left\{1, \log\left(\frac{S_{U(\mu)}}{h_{U(\mu)}}\right)\right\}^2.$$

The proof of Theorem 8.8 is analogous to the proof of Theorem 2.4. The only point where a slightly different argument is needed is the following version of Proposition 7.1. The remainder of the proof is verbatim to the proof of Theorem 2.4 with only changing the notation of h_μ to $h_{U(\mu)}$ and S_μ to $S_{U(\mu)}$.

Proposition 8.9. *Let μ be a finitely supported, contracting on average probability measure on G. Suppose that $S_{U(\mu)} < \infty$ and that $h_{U(\mu)}/|\chi_\mu|$ is sufficiently large. Let $S > S_{U(\mu)}$, $\kappa > 0$ and $a \geq 1$ and suppose that $0 < r_1 < r_2 < a^{-1}$ with $r_1 < \exp(-S\log(\kappa^{-1})/|\chi_\mu|)$. Then as $\kappa \to 0$,*

$$H_a(q_{\tau_\kappa}; r_1|r_2) \geq \left(\frac{h_{U(\mu)}}{|\chi_\mu|} - d - 1\right)\log\kappa^{-1} + H(s_{r_2,a}) + o_{\mu,d,S,a}(\log\kappa^{-1}).$$

Proof. The proof is similar to the one of Proposition 7.1 thus we only provide a sketch. Lemma 7.3 still holds and therefore we only need to show that

$$H_a(q_{\tau_\kappa}; r_1) \geq \left(\frac{h_{U(\mu)}}{|\chi_\mu|} - 1\right)\log\kappa^{-1} + o_{\mu,d,S,a}(\log\kappa^{-1}), \tag{8.7}$$

where $H_a(q_{\tau_\kappa}; r_1) = H(q_{\tau_\kappa}s_{r_1,a}) - H(s_{r_1,a})$. To show (8.7) we apply Lemma 6.6 with $X = G \to \mathbb{R}_{>0} \times O(d) \times \mathbb{R}^d$ and $\Phi : G \to X, g \mapsto (\rho(g), U(g), b(g))$ and m_X the product measure on X as used in Lemma 7.3. Then we note that $\frac{d\Phi_* m_G}{dm_X} = 1$. Thus by Lemma 6.6,

$$\begin{aligned} H(q_{\tau_\kappa}s_{r_1,a}) &= D_{\mathrm{KL}}(U(q_{\tau_\kappa}s_{r_1,a}) \,||\, dU) + D_{\mathrm{KL}}(\rho(q_{\tau_\kappa}s_{r_1,a}) \,||\, \rho^{-(d+1)}d\rho) \\ &\quad + D_{\mathrm{KL}}(b(q_{\tau_\kappa}s_{r_1,a}) \,||\, db). \end{aligned}$$

As in Proposition 7.1 one shows that

$$D_{\mathrm{KL}}(U(q_{\tau_\kappa}s_{r_1,a}) \,||\, dU) \geq \frac{h_{U(\mu)}}{|\chi_\mu|}\log\kappa^{-1} + D_{\mathrm{KL}}(U(s_{r_1,a}) \,||\, dU) + o_{\mu,d,S,a}(\log\kappa^{-1}).$$

On the other hand,

$$D_{\mathrm{KL}}(\rho(q_{\tau_\kappa}s_{r_1,a}) \,||\, \rho^{-(d+1)}d\rho) \gg D_{\mathrm{KL}}(\rho(s_{r_1,a}) \,||\, \rho^{-(d+1)}d\rho)$$

and

$$D_{\mathrm{KL}}(b(q_{\tau_\kappa}s_{r_1,a}) \,||\, db) \gg D_{\mathrm{KL}}(b(s_{r_1,a}) \,||\, db)$$

and note that by Lemma 6.5,

$$D_{\mathrm{KL}}(U(s_{r_1,a}) \,||\, dU) + D_{\mathrm{KL}}(\rho(s_{r_1,a}) \,||\, \rho^{-(d+1)}d\rho) + D_{\mathrm{KL}}(b(s_{r_1,a}) \,||\, db) \geq H(s_{r_1,a}).$$

All these estimates combined imply the claim. ☐

9. Well-Mixing and Non-Degeneracy

In this section we study (c,T)-well mixing as well as (α_0,θ,A)-non-degeneracy. The goal of this section is prove Proposition 2.2 and Proposition 2.3. We treat (c,T)-well-mixing in section 9.1 and show that we have uniform results as long as $U(\mu)$ is fixed. In section 9.2 we conclude the proofs of Proposition 2.2 and Proposition 2.3 by proving strong results on non-degeneracy.

9.1. (c, T)**-well-mixing.** In this subsection we establish in Lemma 9.2 that we have uniform (c, T)-well-mixing whenever $U(\mu)$ is fixed and show that (c, T) can taken to be uniform when we know a lower bound on the spectral gap of $U(\mu)$. We start with a preliminary lemma that will also be used in section 9.2. Throughout this section and next we denote by m_H the Haar probability measure on H and by $I \in O(d)$ the identity matrix.

Lemma 9.1. *(Schur-type Lemma) Suppose that $d \geq 1$ and that H is an irreducible subgroup of $O(d)$ and let V be a uniform random variable on H. Let B be a random variable independent from V taking values in $\mathbb{R}^d$. Then VB has mean zero and covariance matrix of the form λI for some $\lambda \geq 0$.*

Proof. For $h \in H$ the random variables hVB and VB have the same law. This means that the mean of VB is invariant under H and so since H is irreducible it must be zero. Moreover the covariance matrix M of VB is invariant under conjugation by elements of H. Since M is symmetric positive definite, it has an eigenvector v and therefore $Mv = \lambda v$ and $hMv = Mhv = \lambda v$ for some $\lambda \geq 0$ and all $h \in H$. Since H is irreducible it therefore follows that $M = \lambda I$ as claimed. □

Lemma 9.2. *Let μ_U be a finitely supported probability measure on $O(d)$ such that $\mathrm{supp}(\mu_U)$ acts irreducibly on $\mathbb{R}^d$. Then there exists $T = T(\mu_U)$ only depending on μ_U such that every finitely supported probability measure μ on G with $U(\mu)$ is $(\frac{1}{2d}, T)$-well-mixing.*

Proof. Let $H \subset O(d)$ be the closure of the group generated by $\mathrm{supp}(\mu_U)$. Then H is compact and let m_H the Haar probablility measure on G and denote by V a uniform random variable on H. We first claim that for all unit vectors x and y in $\mathbb{R}^d$ we have

$$\mathbb{E}[|x \cdot Vy|^2] = \frac{1}{d}. \tag{9.1}$$

Indeed, we can view y as a random variable independent from V and therefore Vy has mean zero and covariance matrix λI. Moreover, since $\mathbb{E}[|Vy|^2] = d\lambda = 1$ it follows that $\lambda = \frac{1}{d}$ and therefore (9.1) holds.

Let F be a uniform random variable on $[0, T]$. Then F is distributed as

$$\frac{1}{T+1} \sum_{i=0}^{T} \mu^{*i}. \tag{9.2}$$

We claim that (9.2) converges as $T \to \infty$ to m_H in the weak*-topology. Indeed, we note that any weak*-limit m of (9.2) is μ_U-stationary and, upon performing an ergodic decomposition, we may assume without loss of generality that m is in addition ergodic. As this is equivalent to the measure being extremal, we conclude that m is invariant under the group generated by $\mathrm{supp}(\mu_U)$ and therefore also by H, implying that $m = m_H$.

Finally we just choose $c = \frac{1}{2d}$ and T sufficiently large depending on μ_U such that (9.2) is sufficiently close in distribution to m_H and therefore $\mathbb{E}[|x \cdot U(q_F)y|^2] \geq \frac{1}{2d}$ for all unit vectors $x, y \in \mathbb{R}^d$, implying the claim. □

For a closed subgroup $H \subset O(d)$ and a probability measure μ_U supported on H we denote as defined in (2.19) by $\mathrm{gap}_H(\mu_U)$ the L^2-spectral gap of μ_U on $L^2(H)$. We aim to show uniform well-mixing as long as $\mathrm{gap}_H(\mu_U) \geq \varepsilon$ independent of the subgroup H. To do so, we first show that we have uniform convergence in the Wasserstein distance with a rate only depending on ε and d.

Lemma 9.3. *Let $d \geq 1, \varepsilon \in (0,1)$ and let μ_U be a probability measure on $O(d)$. Assume that $\mathrm{gap}_H(\mu_U) \geq \varepsilon$ for H the subgroup generated by the support of μ_U. Then for $n \geq 1$*

$$\mathcal{W}_1(\mu_U^{*n}, m_H) \ll_d (1-\varepsilon)^{\alpha n}$$

for $\alpha = (1 + \frac{1}{2}\dim O(d))^{-1}$.

Proof. We consider the metric $d(g_1, g_2) = ||g_1 - g_2||$ on $O(d)$ for $||\circ||$ the operator norm and note that it is bi-invariant and restricts to H. Denote by $B_\delta^H(h)$ for $h \in H$ and $\delta > 0$ the δ-ball around H and denote

$$P_\delta = \frac{1_{B_\delta^H(e)}}{m_H(B_\delta^H(e))}.$$

For $\delta \in (0,1)$ we note that $m_H(B_\delta^H(e)) \gg_d \delta^{\dim O(d)}$ for an implied constant depending only on d and therefore $||P_\delta||_2 \ll_d \delta^{-(\dim O(d))/2}$. Also we note that for $h \in H$ we have $(\mu^{*n} * P_\delta)(h) = \frac{\mu^{*n}(B_\delta^H(h))}{m_H(B_\delta^H(e))}$. By the triangle inequality,

$$\mathcal{W}_1(\mu^{*n}, m_H) \leq \mathcal{W}_1(\mu^{*n}, \mu^{*n} * P_\delta) + \mathcal{W}_1(\mu^{*n} * P_\delta, m_H).$$

Note $\mathcal{W}_1(\mu^{*n}, \mu^{*n} * P_\delta) \ll_d \delta$ and since H is compact,

$$\begin{aligned}\mathcal{W}_1(\mu^{*n} * P_\delta, m_H) &\leq ||\mu^{*n} * P_\delta - 1||_1 \\ &\leq ||\mu^{*n} * P_\delta - 1||_2 \\ &\leq (1-\varepsilon)^n ||P_\delta||_2 \ll_d (1-\varepsilon)^n \delta^{-(\dim O(d))/2}.\end{aligned}$$

To conclude, if follows

$$\mathcal{W}_1(\mu^{*n}, m_H) \ll_d \delta + (1-\varepsilon)^n \delta^{-(\dim O(d))/2}.$$

Therefore setting $\delta = (1-\varepsilon)^{\alpha n}$ for $\alpha = (1 + \frac{1}{2}\dim O(d))^{-1}$ implies the claim. $\square$

Lemma 9.4. *Let $d \geq 1, \varepsilon \in (0,1)$ and let μ_U be a probability measure on $O(d)$. Assume that $\mathrm{gap}_H(\mu_U) \geq \varepsilon$ for H the subgroup generated by the support of μ_U. Then there exists $T = T(d, \varepsilon)$ only depending on d and ε such every probability measure μ on G with $U(\mu) = \mu_U$ is $(\frac{1}{2d}, T)$-well-mixing.*

Proof. The proof is similar to the one of Lemma 9.2 and recall the notation used in it. Consider a list of tuples of unit vectors $(x_1, y_1), \ldots, (x_m, y_m)$ such that for every two unit vectors x and y in $\mathbb{R}^d$ there is some $i \in [m]$ such that

$$\sup_{U \in O(d)} \Big| |x \cdot Uy|^2 - |x_i \cdot Uy_i|^2 \Big| < \frac{1}{4d}.$$

Such a list of tuples exists as the action of $O(d)$ on $\mathbb{S}^{d-1} \subset \mathbb{R}^d$ is uniformly continuous. We claim that for T large enough depending only on ε we have for all $i \in [m]$ that

$$\mathbb{E}[|x_i \cdot U(q_F)y_i|^2] \geq \frac{3}{4d}.$$

Indeed, we note that for $h_1, h_2 \in H$ we have

$$\Big| |x_i \cdot h_1 y_i|^2 - |x_i \cdot h_1 y_i|^2 \Big| \leq \Big| |x_i \cdot h_1 y_i| + |x_i \cdot h_1 y_i| \Big| \cdot \Big| |x_i \cdot h_1 y_i| - |x_i \cdot h_1 y_i| \Big| \leq 2||h_1 - h_2||.$$

Thus it follows that

$$\mathbb{E}[|x_i \cdot Vy_i|^2 - |x_i \cdot U(q_n)y_i|^2] \leq 2\mathcal{W}_1(\mu^{*n}, m_H)$$

and the claim follows by Lemma 9.3. This concludes the proof as for all x and y we have

$$\mathbb{E}[|x \cdot U(q_F)y|^2] \geq \sup_{i\in[m]} \mathbb{E}[|x_i \cdot U(q_F)y_i|^2] - \frac{1}{4d} \geq \frac{1}{2d}.$$

□

Another direction to show uniform well-mixing would be to study the stopped random walk $U(q_{\tau_\kappa})$ and to show that $U(q_{\tau_\kappa}) \to m_H$. We do not pursue this direction further and just note that the results by Kesten [Kes74] can be applied to this problem.

9.2. (α_0, θ, A)**-non-degeneracy.** In order to state our results on (α_0, θ, A)-non-degeneracy it is useful to understand that we can translate and rescale our generating measures, without changing any of the fundamental properties. It is also beneficial to replace μ by $\frac{1}{2}\delta_e + \frac{1}{2}\mu$ and we show in the following lemma that these changes do not change our self-similar measure or any of the relevant constants in a fundamental way.

Lemma 9.5. *Let $\mu = \sum_i p_i\delta_{g_i}$ be a contracting on average probability measure on G with self-measure ν. Let $h \in G$ and consider the measures*

$$\mu_h = \sum_i p_i \delta_{hg_ih^{-1}} \quad \text{and} \quad \mu'_h = \frac{1}{2}\delta_e + \frac{1}{2}\mu_h.$$

Then the following properties hold:

(i) $h_\mu = h_{\mu_h} = 2h_{\mu'_h}$,
(ii) $\chi_\mu = \chi_{\mu_h} = 2\chi_{\mu'_h}$,
(iii) $S_\mu = S_{\mu_h} = S_{\mu'_h}$,
(iv) $\mathrm{gap}_H(\mu) = \mathrm{gap}_{U(h)HU(h)^{-1}}(\mu_h) = 2\mathrm{gap}_{U(h)HU(h)^{-1}}(\mu'_h)$,
(v) *μ_h and $\mu_{h'}$ have $h\nu$ as self-similar measure.*

Proof. As conjugation is a bijection on G and by using [HS17, Lemma 6.8], (i) follows. Moreover, (ii) follows since $\rho(hg_ih^{-1}) = \rho(g_i)$ and (iv) follows similarly. To show (iii) note $S_{\mu_h} = S_{\mu'_h}$ since by the triangle inequality $d(g, h) \leq d(g, e) + d(e, h)$ for all $g, h \in G$. To show that $S_\mu = S_{\mu_h}$, set

$$A = \min_{g_1,g_2\in\mathrm{supp}(\mu), g_1\neq g_2} d(g_1, g_2)$$

and note that there is a constant C_h depending on h such that $d(hg_1h^{-1}, hg_2h^{-1}) \leq C_hd(g_1, g_2)$ for $d(g_1, g_2) \leq A$. Thus it holds that

$$\begin{aligned} S_{\mu_h} &= \limsup_{g_1,g_2\in S_n, g_1\neq g_2} -\frac{1}{n} \log d(hg_1h^{-1}, hg_2h^{-1}) \\ &\leq \limsup_{g_1,g_2\in S_n, g_1\neq g_2} -\frac{1}{n} \log C_hd(g_1, g_2) = S_\mu \end{aligned}$$

Applying the same argument to conjugation by h^{-1} implies the claim. Finally, we note that μ_h and μ'_h have the same self-similar measure and it holds that

$$h\nu = h\sum_i p_ig_i\nu = \sum_i p_ihg_ih^{-1}h\nu$$

and therefore $h\nu$ is the self-similar measure of μ_h and μ'_h. □

In particular, it follows that the self-similar measure of μ is absolutely continuous if and only if the one of μ_h or $\mu_{h'}$ is and all of the relevant quantities are the same up to a factor of 2.

To give an idea of the proof of the main results in this subsection, we first discuss the case of real Bernoulli convolutions ν_λ with $\lambda \in (1/2, 1)$. Indeed, we distinguish between $\lambda \geq \lambda_0$ and $\lambda \leq \lambda_0$ for some λ_0 sufficiently close to 1. Note that ν_λ is supported on $[-(1-\lambda)^{-1}, (1-\lambda)]$ and thus when $\lambda \leq \lambda_0$ one easily checks uniform non-degeneracy depending only on λ_0 using for example that Bernoulli convolutions are symmetric around 0. In the case $\lambda \geq \lambda_0$ we rescale ν_λ by $\sqrt{1-\lambda^2}$ to ν'_λ so that ν'_λ has variance one. Then we can deduce from the Berry-Essen type Lemma 5.10 that $\mathcal{W}_1(\nu'_\lambda, \mathcal{N}(0,1)) \ll \sqrt{1-\lambda^2} \ll \sqrt{1-\lambda_0^2}$. We then deduce from the latter as in Lemma 9.9 that ν'_λ is uniformly non-degenerate for λ_0 sufficiently close to 1.

Our results will be deduced from suitable results in the case when μ has a uniform contraction ratio and then in the general case from comparing our given measure with a self-similar measure with uniform contraction ratio. We now state the main proposition of this section.

Proposition 9.6. *Let $d \geq 1$, $\varepsilon > 0$ and let μ_U be an irreducible probability measure on $O(d)$. Then there is $\tilde{\rho} \in (0,1)$ and some (α_0, θ, A) depending on d, ε and μ_U such that the following is true. Let $\mu = \sum_{i=1}^k p_i \delta_{g_i}$ be a contracting on average probability measure on G satisfying $U(\mu) = \mu_U$ and*

$$p_i \geq \varepsilon \quad \text{as well as} \quad \rho(g_i) \in (\tilde{\rho}, 1) \qquad \text{for all} \quad 1 \leq i \leq k.$$

Suppose further that there is some $\hat{\rho} \in (\tilde{\rho}, 1)$ such that

$$\frac{\mathbb{E}_{\gamma\sim\mu}|\hat{\rho} - \rho(\gamma)|}{1 - \mathbb{E}_{\gamma\sim\mu}[\rho(\gamma)]} < 1 - \varepsilon.$$

Then there is some $h \in G$ with $U(h) = I$ such that the conjugate measure $\mu'_h = \frac{1}{2}\delta_e + \frac{1}{2}\sum_i p_i \delta_{hg_ih^{-1}}$ is (α_0, θ, A)-non-degenerate.

Moreover, if in addition $\mathrm{gap}_H(\mu) \geq \varepsilon$, for H the closure of the subgroup generated by $\mathrm{supp}(\mu)$, then $\tilde{\rho}$ and (α_0, θ, A) can be made uniform in d and ε.

We first show how to deduce from Proposition 9.6 the two propositions 2.2 and 2.3 from section 2.1. To do so we first state the following lemma.

Lemma 9.7. *Suppose $x_1 < x_2$ and let X be a real-valued random variable such that $X \leq x_2$ almost surely and $\mathbb{P}[X \leq x_1] \geq 1/2 + p$ for some $p > 0$. Then*

$$\mathbb{E}[|X - x_1|] \leq \mathbb{E}[|X - x_2|] - 2p(x_2 - x_1).$$

Proof. Let X_1 and X_2 have the same law as X and be coupled such that at least one of them is at most x_1 almost surely. Let A be the event that both X_1 and X_2 are at most x_1. Noting that A has probability at least $2p$ we compute

$$\begin{aligned}
\mathbb{E}[|X_1 - x_1| + |X_2 - x_1|] &= \mathbb{E}[(|X_1 - x_1| + |X_2 - x_1|)\mathbb{I}_{A^C}] \\
&\quad + \mathbb{E}[(|X_1 - x_1| + |X_2 - x_1|)\mathbb{I}_A] \\
&\leq \mathbb{E}[(|X_1 - x_2| + |X_2 - x_2|)\mathbb{I}_{A^C}] \\
&\quad + \mathbb{E}[(|X_1 - x_2| + |X_2 - x_2| - 2(x_2 - x_1))\mathbb{I}_A] \\
&\leq \mathbb{E}[|X_1 - x_2| + |X_2 - x_2|] - 4p(x_2 - x_1).
\end{aligned}$$

The result follows. □

We now prove Proposition 2.2 and Proposition 2.3.

Proof of Proposition 2.2. Let $\gamma_1, \gamma_2, \dots$ be i.i.d. samples from μ. Let $p_{\min}$ be the smallest of the $p_1, \dots, p_k$ and let $\rho_{\min}$ be the smallest of the $\rho(g_1), \dots, \rho(g_k)$. Clearly

$$\mathbb{P}[\rho(\gamma_1 \dots \gamma_n) \leq \rho_{\min}] \geq 1 - (1 - p_{\min})^n.$$

In particular there is some n depending only on ε such that this is at least $3/4$. Note that by Lemma 9.7 with $x_1 = \rho_{\min}$ and $x_2 = 1$ and $p = \frac{1}{4}$ we have

$$\begin{aligned}\frac{\mathbb{E}[|\rho(\gamma_1 \cdots \gamma_n) - \rho_{\min}|]}{1 - \mathbb{E}[\rho(\gamma_1 \dots \gamma_n)]} &\leq \frac{1 - \mathbb{E}[\rho(\gamma_1 \dots \gamma_n)] - (1 - \rho_{\min})/2}{1 - \mathbb{E}[\rho(\gamma_1 \dots \gamma_n)]} \\ &= 1 - \frac{1 - \rho_{\min}}{2(1 - \mathbb{E}[\rho(\gamma_1 \dots \gamma_n)])} \\ &\leq 1 - \frac{1 - \rho_{\min}}{2(1 - \rho_{\min}^n)} \\ &\leq 1 - \frac{1}{2n}.\end{aligned}$$

The result now follows by applying Proposition 9.6, Lemma 9.2 and Lemma 9.4 to μ^{*n}. □

Proof of Proposition 2.3. This follows directly by Proposition 9.6 and Lemma 9.4. □

Now we prove Proposition 9.6. We use the following definition.

Definition 9.8. *Given two measures λ_1, λ_2 on $\mathbb{R}^d$ we define*

$$\mathcal{PW}_1(\lambda_1, \lambda_2) := \inf_{\gamma \in \Gamma(\lambda_1, \lambda_2)} \sup_{p \in P(d)} \int |px - py| \, d\gamma(x, y)$$

where $P(d)$ is the set of orthogonal projections onto one dimensional subspaces of $\mathbb{R}^d$ and $\Gamma(\lambda_1, \lambda_2)$ is the set of couplings between λ_1 and λ_2.

We use this to show that if a measure is sufficiently close to a spherical normal distribution then it is (α_0, θ, A)-non-degenerate.

Lemma 9.9. *Let I be the $d \times d$ identity matrix. Then given any $p \in P(d)$ we have*

$$\mathbb{E}_{x \sim N(0,1)}[|px|] = \sqrt{\frac{2}{\pi}}.$$

Moreover, for any $\varepsilon > 0$ there exists $\alpha_0 \in (0, 1)$ and $\theta, A > 0$ such that if ν is a measure on $\mathbb{R}^d$ and

$$\mathcal{PW}_1(\nu, N(0, I)) < \sqrt{\frac{2}{\pi}} - \varepsilon$$

then ν is (α_0, θ, A)-non-degenerate.

Proof. The first part follows since if $X \sim \mathcal{N}(0, I)$ and $u \in \mathbb{R}^d$ is a unit vector, then $\langle X, u \rangle$ is distributed as $\mathcal{N}(0, 1)$. The second part follows from the first part, the fact that the $y \in \mathbb{R}$ such that $\mathbb{E}_{x \sim N(0,1)}|x - y|$ is minimal is $y = 0$ and Markov's inequality.

More precisely, we aim to estimate for all $y_0 \in \mathbb{R}^d$ and all proper subspaces $W \subset \mathbb{R}^d$

$$\nu(\{x \in \mathbb{R}^d : |x - (y_0 + W)| < \theta \text{ or } |x| \geq A\}),$$

which is bounded by $\nu(\{x \in \mathbb{R}^d : |x-(y_0+W)| < \theta\}) + \nu(\{x \in \mathbb{R}^d : |x| \geq A\})$. To deal with the second term we note that by Markov's inequality for a coupling γ between ν and $\mathcal{N}(0,1)$ we have

$$\begin{aligned}\nu(\{x \in \mathbb{R}^d : |x| \geq A\}) &\leq A^{-1} \int |x| \, d\nu(x) \\ &\leq A^{-1} \left(\int |y| \, d\mathcal{N}(0,I)(y) + \int |x-y| \, d\gamma(x,y) \right).\end{aligned}$$

In order to apply our bound for $\mathcal{PW}_1(\nu, N(0,I))$ we consider the projections $p_1, \dots, p_d$ to the coordinate axes. Then $|x-y| \leq \sum_{i=1}^d |p_i x - p_i y|$ and therefore by choosing a suitable coupling, it follows that for A sufficiently large only depending on d and ε we have that $\nu(\{x \in \mathbb{R}^d : |x| \geq A\}) \leq \varepsilon/16$.

To deal with the first term $\nu(\{x \in \mathbb{R}^d : |x-(y_0+W)| < \theta\})$, we assume without loss of generality that W has dimension $d-1$ and we let p be the orthogonal projection to the orthogonal complement of W. Then it holds that $|x-(y_0+W)| = |px - py_0|$ and therefore

$$\nu(\{x \in \mathbb{R}^d : |x-(y_0+W)| < \theta\}) = \nu(\{x \in \mathbb{R}^d : |px-py_0| < \theta\}).$$

In the following we identify $p\mathbb{R}^d$ as the real line. Let γ be any coupling between ν and $\mathcal{N}(0,I)$. Then it holds that

$$\begin{aligned}\int |px-py| \, d\gamma(x,y) &\geq \int |px-py| 1_{|px-py_0|<\theta}(x,y) \, d\gamma(x,y) \\ &\geq \nu(\{x \in \mathbb{R}^d : (|px-py_0| < \theta\}) \int |py-py_0| - \theta) \, d\mathcal{N}(0,I)(y) \\ &\geq \nu(\{x \in \mathbb{R}^d : (|px-py_0| < \theta\}) \left(\sqrt{\frac{2}{\pi}} - \theta \right),\end{aligned}$$

having used in the last line that $y \in \mathbb{R}$ such that $\mathbb{E}_{x \sim N(0,1)} |x-y|$ is minimal is $y=0$. By choosing a suitable coupling and setting $\theta = \varepsilon/4$ it therefore follows for ε sufficiently small that

$$\nu(\{x \in \mathbb{R}^d : (|px-py_0| < \theta\}) \leq \frac{\sqrt{\frac{2}{\pi}} - \varepsilon/2}{\sqrt{\frac{2}{\pi}} - \varepsilon/4} \leq 1 - \varepsilon/8.$$

The claim follows by combining the above two estimates. □

To make this useful we need to show that our self-similar measures are close to spherical normal distributions. We prove this in the case where all of the ρ_i are equal with the following proposition.

Proposition 9.10. *Given any $\varepsilon > 0$ and any irreducible probability measure $\mu_U = \sum_{i=1}^k p_i \delta_{U_i}$ on $O(d)$ there is some $\tilde{\rho} \in (0,1)$ depending on ε and μ_U such that the following is true. Let $\mu = \sum_{i=1}^k p_i \delta_{g_i}$ be a probability measure on G without a common fixed point and with $U(\mu) = \mu_U$ as well as $p_i \geq \varepsilon$ for all $1 \leq i \leq k$. Assume there is $\rho \in (\tilde{\rho}, 1)$ such that $\rho(g_i) = \rho$ for all $1 \leq i \leq k$. Then there exists some $h \in G$ with $U(h) = I$ such that the self similar measure ν'_h generated by the conjugate measure $\mu'_h = \frac{1}{2}\delta_e + \frac{1}{2}\sum_i p_i \delta_{h g_i h^{-1}}$ satisfies*

$$\mathcal{W}_3(\nu'_h, N(0,I)) < \varepsilon.$$

If moreover $\mathrm{gap}_H(\mu_U) \geq \varepsilon$ *then* $\tilde{\rho}$ *is uniform in* d *and* ε*.*

We then extend to the general case using the following lemma.

Lemma 9.11. *Let* γ *and* $\tilde{\gamma}$ *be contracting on average random variables taking values in* G *such that* $U(\gamma) = U(\tilde{\gamma})$ *and* $z(\gamma) = z(\tilde{\gamma})$ *almost surely. Let* ν *and* $\tilde{\nu}$ *be the self similar measures generated by the laws of* γ *and* $\tilde{\gamma}$ *respectively. Then*

$$\mathcal{PW}_1(\nu, \tilde{\nu}) \leq \frac{\mathbb{E}[|\rho(\gamma) - \rho(\tilde{\gamma})|]}{1 - \mathbb{E}[\rho(\gamma)]} \sup_{p \in P(d)} \mathbb{E}_{x \sim \tilde{\nu}} |px|.$$

We now have all the ingredients needed to prove Proposition 9.6.

Proof of Proposition 9.6. Without loss of generality we replace μ by $\frac{1}{2}\delta_e + \frac{1}{2}\mu$. Let $\tilde{g}_i : x \mapsto \hat{\rho} U_i x + b_i$ and let $\tilde{\mu} = \sum_{i=1}^n \delta_{\tilde{g}_i}$ with self-similar measure $\tilde{\nu}$. Then by Proposition 9.10 there is some $h \in G$ with $U(h) = I$ such that

$$\mathcal{W}_3(\tilde{\nu}_h, N(0, I)) < \varepsilon/10.$$

Clearly this implies $\mathcal{W}_1(\tilde{\nu}_h, N(0, I)) < \varepsilon/10$ and therefore $\mathcal{PW}_1(\tilde{\nu}_h, N(0, I)) < \varepsilon/10$ and so by Lemma 9.11 if we define $\mu_h = \sum_{i=1}^k p_i \delta_{h g_i h^{-1}}$ and let ν_h be the self similar measure generated by μ_h we have $\mathcal{PW}_1(\nu_h, N(0, I)) < \sqrt{\frac{\pi}{2}} - \varepsilon/2$. The result follows by Lemma 9.9. □

Now we just need to prove Lemma 9.11 and Proposition 9.10. We start with Lemma 9.11.

Proof of Lemma 9.11. Let x be a sample from ν and $\tilde{x}$ be a sample from $\tilde{\nu}$ such that $(x, \tilde{x})$ is independent from $(\gamma, \tilde{\gamma})$. Note that this means that γx is a sample from ν and $\tilde{\gamma}\tilde{x}$ is a sample from $\tilde{\nu}$. Let $p \in P(d)$. We have

$$\begin{aligned} \mathbb{E}[|p\gamma x - p\tilde{\gamma}\tilde{x}|] &\leq \mathbb{E}[|p\gamma(x - \tilde{x})|] + \mathbb{E}[|p(\gamma - \tilde{\gamma})\tilde{x}|] \\ &= \mathbb{E}[\rho(\gamma)]\mathbb{E}[|pU(\gamma)(x - \tilde{x})|] + \mathbb{E}[|\rho(\gamma) - \rho(\tilde{\gamma})|]\mathbb{E}[|pU(\gamma)(\tilde{x})|]. \end{aligned}$$

Therefore by taking a series of couplings such that $\sup_{p \in P(d)} \mathbb{E}[|px - p\tilde{x}|] \to \mathcal{PW}_1(\nu, \tilde{\nu})$ we get

$$\mathcal{PW}_1(\nu, \tilde{\nu}) \leq \mathbb{E}[\rho(\gamma)]\mathcal{PW}_1(\nu, \tilde{\nu}) + \mathbb{E}[|\rho(\gamma) - \rho(\tilde{\gamma})|]\mathbb{E}_{x \sim \tilde{\nu}}[|p(x)|].$$

□

Now we wish to prove Proposition 9.10. First we need the following result.

Lemma 9.12. *Let* μ_U *be a probability measure on* $O(d)$ *and let* H *be the closure of the group generated by the support of* μ_U *and let* V *be a uniform random variable on* H*. Let* $\gamma_1, \gamma_2, \dots$ *be independent samples from* $\frac{1}{2}\delta_e + \frac{1}{2}\mu_U$*. Then for every* $\varepsilon > 0$ *there exists* $N \in \mathbb{Z}_{>0}$ *such that whenever* $n \geq N$ *we have*

$$\mathcal{W}_3(\gamma_1 \dots \gamma_n, V) < \varepsilon.$$

Furthermore if $\mathrm{gap}_H(\mu_U) \geq \varepsilon$*, then* N *can be made uniform* d *and* ε*.*

Proof. This follows similar to the arguments given in section 9.1 since the measure $\mu'_U = \frac{1}{2}\delta_e + \frac{1}{2}\mu_U$ satisfies that $(\mu'_U)^{*n} \to m_H$ as $n \to \infty$. In the presence of a spectral gap we apply Lemma 9.3 and use that by compactness of H the L^3-Wasserstein distance is comparable with the L^1-Wasserstein distance. □

It is convenient to work with measures which are appropriately translated.

Definition 9.13. *We say that a probability measure μ on G is* centred at zero *if* $\mathbb{E}_{\gamma\sim\mu}[\gamma(0)] = 0$.

Lemma 9.14. *Suppose that μ is a probability measure on G which is centred at zero and has uniform contraction ratio $\rho \in (0,1)$. Then if $\gamma_1, \gamma_2, \dots$ are i.i.d. samples from μ and $n \in \mathbb{Z}_{>0}$ we have*

$$\mathbb{E}[\gamma_1 \dots \gamma_n(0)] = 0$$

and

$$\mathbb{E}[|\gamma_1 \dots \gamma_n(0)|^2] = \frac{1-\rho^{2n}}{1-\rho^2}\mathbb{E}[|\gamma_1(0)|^2].$$

Proof. Both of these follow by an induction argument left to the reader. □

In order to prove this we need the following theorem of Sakhanenko from [Sak85].

Theorem 9.15. *For every $p, d \geq 1$ there is some constant $c = c(p,d) > 0$ such that the following holds. Suppose that $X_1, \dots, X_n$ are independent random variables taking values in $\mathbb{R}^d$ with mean 0. Let $\Sigma_i = \operatorname{Var} X_i$, suppose that $\sum_{i=1}^n \Sigma_i^2 = I$ and let $L_p = \sum_{i=1}^p \mathbb{E}[|X_i|^p]$. Then*

$$\mathcal{W}_p\left(\sum_{i=1}^n X_i, N(0,I)\right) \leq CL_p.$$

This is enough to deduce the following estimate.

Lemma 9.16. *Let $(p_1, \dots, p_k)$ be a probability vector, $U_1, \dots, U_k \in O(d)$ generate an irreducible subgroup, $b_1, \dots, b_k \in \mathbb{R}^d$ and let $\rho \in (0,1)$. Let μ be the probability measure on G given by $\mu = \sum_{i=1}^k p_i \delta_{g_i}$ where $g_i : x \mapsto \rho U_i x + b_i$. Suppose that μ is centred at zero and that all of the b_i have modulus at most 1. Let $\gamma_1, \gamma_2, \dots$ be i.i.d. samples from μ. Let $\varepsilon \in (0,1)$.*

Given $\ell \in \mathbb{Z}_{>0}$ we define $S_\ell := \mathbb{E}[|\gamma_1 \dots \gamma_\ell(0)|^2]$ and

$$W_\ell := \mathcal{W}_3\left(d^{1/2} S_\ell^{-1/2} \gamma_1 \dots \gamma_\ell(0), N(0,I)\right).$$

Suppose that there exist $m, n \in \mathbb{Z}_{>0}$ such that for V a uniform random variable on the closure of the subgroup generated by the $U_1, \dots, U_k$ we have

$$\mathcal{W}_3(U(\gamma_1 \dots \gamma_m), V) < \varepsilon \qquad \text{and} \qquad \frac{m}{S_n^{1/2}} < \varepsilon.$$

Then for $n' \in \mathbb{Z}_{>0}$,

$$W_{(m+n)n'} \ll_d (T^{-1/6} + T^{1/2}\varepsilon)(W_n + 1) \tag{9.3}$$

where $T := \sum_{i=0}^{n'-1} \rho^{(m+n)i}$. In particular if $\rho^{(m+n)n'} > 1/2$ then $n'/2 \leq T \leq n$ and therefore

$$W_{(m+n)n'} \ll_d ((n')^{-1/6} + (n')^{1/2}\varepsilon)(W_n + 1) \tag{9.4}$$

Proof. For $i = 1, \dots, n'$ let

$$X_i := \gamma_{(i-1)(n+m)+1} \dots \gamma_{(i-1)(n+m)+m}$$

and

$$Y_i := \gamma_{(i-1)(n+m)+m+1} \dots \gamma_{i(n+m)}$$

such that

$$Z_i = X_i Y_i = \gamma_{(i-1)(n+m)+1} \dots \gamma_{i(n+m)}.$$

Furthermore consider $V_1, \dots, V_k$ independent random variables which are uniform on H (the closure of the subgroup generated by the U_i), independent of the Y_i and are such that

$$\mathbb{E}[\|U(X_i) - V_i\|^3] < \varepsilon^3.$$

Note that

$$Z_1 \dots Z_{n'}(0) = Z_1(0) + \rho^{(m+n)} U(Z_1) Z_2(0) + \\ \dots + \rho^{(m+n)(n'-1)} U(Z_1 \dots Z_{n'-1}) Z_{n'}(0).$$

Also note that

$$\begin{aligned} &\mathcal{W}_3\left(\rho^{(m+n)(i-1)} U(Z_1 \dots Z_{i-1}) Z_i(0), \rho^{(m+n)(i-1)+m} V_i Y_i(0)\right) \\ &\quad = \rho^{(m+n)(i-1)} \mathcal{W}_3\left(U(Z_1 \dots Z_{i-1})(\rho^m U(X_i) Y_i(0) + X_i(0)), \rho^m V_i Y_i(0)\right) \\ &\quad \leq \rho^{(m+n)(i-1)} (m + \varepsilon \rho^m (\mathbb{E}\left[|Y_i(0)|^3\right])^{1/3}) \\ &\quad \ll_d \varepsilon \rho^{(m+n)(i-1)} S_n^{1/2} (W_n + 1), \end{aligned}$$

having used the triangle inequality in the second line and that $|X_i(0)| \leq m$ as $\sup_i |b_i| \leq 1$ as well as that

$$\begin{aligned} &\mathcal{W}_3\left(U(Z_1 \dots Z_{i-1}) U(X_i) Y_i(0), V_i Y_i(0)\right) \\ &\quad = \mathcal{W}_3\left(U(Z_1 \dots Z_{i-1}) U(X_i) Y_i(0), U(Z_1 \dots Z_{i-1}) V_i Y_i(0)\right) \end{aligned}$$

as V_i is distributed like the Haar measure on H.

Note that by Lemma 9.1 the covariance matrix of $V_i Y_i(0)$ is $d^{-1} S_n I$. Therefore by Theorem 9.15 letting $A = d^{-1/2} \left(\frac{1-\rho^{2n'(m+n)}}{1-\rho^{2(m+n)}}\right)^{1/2} S_n^{1/2}$ we have that

$$\begin{aligned} &\mathcal{W}_3\left(A^{-1}\left(\sum_{i=1}^{n'} \rho^{(m+n)(i-1)} V_i Y_i(0)\right), N(0, I)\right) \\ &\quad \ll \left(\sum_{i=1}^{n'} \mathbb{E}\left[|A^{-1} \rho^{(m+n)(i-1)} Y_i(0)|^3\right]\right)^{1/3} \\ &\quad \ll_d A^{-1} \left(\frac{1 - \rho^{3(m+n)n'}}{1 - \rho^{3(m+n)}}\right)^{1/3} (W_n + 1) \\ &\quad \ll_d T^{-1/6} (W_n + 1), \end{aligned}$$

where we exploited that

$$\frac{1 - \rho^{2n'(m+n)}}{1 - \rho^{2(m+n)}} = \frac{1 - \rho^{n'(m+n)}}{1 - \rho^{(m+n)}} \frac{1 + \rho^{n'(m+n)}}{1 + \rho^{(m+n)}} \in [T/2, T]$$

and a similar estimate for $\left(\frac{1-\rho^{3(m+n)n'}}{1-\rho^{3(m+n)}}\right)^{1/3}$.

Therefore we may deduce that

$$\mathcal{W}_3\left(A^{-1} \gamma_1 \dots \gamma_{(m+n)n'}(0), N(0, I)\right) \ll_d T^{-1/6}(W_n + 1) + \varepsilon T^{1/2}(W_n + 1)$$

By Lemma 9.14 we have that

$$\frac{d^{-1/2} S_{n'}^{-1/2}}{A} = 1 + O(\frac{m}{n}) = 1 + O(\varepsilon).$$

We conclude

$$\begin{aligned} W_{(m+n)n'} &\ll_d T^{-1/6}(W_n + 1) + \varepsilon T^{1/2}(W_n + 1) + \varepsilon \\ &\ll_d T^{-1/6}(W_n + 1) + \varepsilon T^{1/2}(W_n + 1) \end{aligned}$$

as required. □

From this we can deduce the following.

Corollary 9.17. *For every $\varepsilon > 0$ and every irreducible probability measure μ_U on $O(d)$ there is $C > 0$ and $\tilde{\rho} \in (0, 1)$ such that the following is true. Let $\mu = \sum_{i=1}^{k} p_i \delta_{g_i}$ be a probability measure on G such that $U(\mu) = \mu_U$ and $p_i \geq \varepsilon$ for all $1 \leq i \leq k$. Assume further that $\max_{1 \leq i \leq k} |b(g_i)| = 1$ and for some $\rho \in (\tilde{\rho}, 1)$ we have $\rho(g_i) = \rho$ for all $1 \leq i \leq k$. Suppose that μ is centred at zero and let $\gamma_1, \gamma_2, \dots$ be i.i.d. samples from μ. Then for every $k \in \mathbb{Z}_{>0}$ such that $C^{k+1} < \frac{\rho^C}{1-\rho^C}$ there is some $n \in \mathbb{Z}_{>0}$ such that*

$$\frac{1}{1 - \rho^n} \in [C^k, C^{k+1}]$$

and

$$\mathcal{W}_3(d^{1/2} S_n^{-1/2} \gamma_1 \dots \gamma_n(0), N(0, I)) < C.$$

Moreover, if $\mathrm{gap}_H(\mu) \geq \varepsilon$, then C and $\tilde{\rho}$ can be made uniform d and ε.

Proof. Let $\varepsilon' > 0$ be sufficiently small. Choose $m = m(\mu_U, \varepsilon')$ such that

$$\mathcal{W}_3(U(\gamma_1 \dots \gamma_m), V) < \varepsilon'$$

and choose $n_0 = n_0(\varepsilon, \varepsilon', \tilde{\rho})$ such that

$$\frac{m}{S_{n_0}^{1/2}} < \varepsilon'.$$

Note that this is possible by Lemma 9.14 as $\varepsilon \leq \mathbb{E}[|\gamma_1(0)|^3] \leq 1$ and providing we choose $\tilde{\rho}$ to be sufficiently close to 1 in terms of ε'. Now inductively chose n'_k such that $\sum_{i=0}^{n'_k - 1} \rho^{(m+n_k)i} \in [\varepsilon'^{-3/2}, 2\varepsilon'^{-3/2}]$ and define $n_{k+1} := n'_k(n_k + m)$. Repeat this process until we find some k such that $\sum_{i=0}^{\infty} \rho^{(m+n_k)i} < \varepsilon'^{-3/2}$ and let k^* denote this value of k. By Lemma 9.16 this means that for $i = 1, \dots, k^*$ we have

$$W_i \ll_d \varepsilon'^{1/4}(W_{i-1} + 1).$$

Providing we take $\tilde{\rho}$ to be sufficiently close to 1 we can bound n_0 and W_{n_0} from above purely in terms of ε and ε'. This means that, providing we choose ε' to be sufficiently small, there is some $C_1 = C_1(\varepsilon, \varepsilon')$ such that for each $i = 1, \dots, k^*$ we have

$$W_{n_i} < C_1.$$

We also have that

$$\frac{1 - \rho^{n_{i+1}}}{1 - \rho^{m+n_i}} \in [\varepsilon'^{-3/2}, 2\varepsilon^{-3/2}]$$

and so providing we choose $\tilde{\rho}$ to be sufficiently large we have

$$\frac{1 - \rho^{n_{i+1}}}{1 - \rho^{n_i}} \leq 4\varepsilon'^{-3/2}.$$

The result follows. When we have a spectral gap, all of these constants can be chosen to be uniform. □

Now we have enough to prove Proposition 9.10.

Proof of Proposition 9.10. Without loss of generality we may assume that μ is centred at zero and that $\max_{i=1}^{k} |b_i| = 1$.

Let $\varepsilon' > 0$. By Lemma 9.12 there is some $m \in \mathbb{Z}_{>0}$ depending only on ε and ε' such that

$$\mathcal{W}_3(U(\gamma_1 \dots \gamma_m(0)), V) < \varepsilon'.$$

By Lemma 9.14 there is some N depending only on μ_U and ε' such that for any $n \geq N$ we have

$$\frac{m}{S_n^{1/2}} < \varepsilon'.$$

Let C be as in Corollary 9.17 and choose n such that

$$\frac{1}{1-\rho^{m+n}} \in [C^{-1}\varepsilon'^{-3/2}, C\varepsilon'^{-3/2}].$$

Providing we choose $\tilde{\rho}$ sufficiently close to 1 we will also have $n \geq N$. By letting $n' \to \infty$ in Lemma 9.16 we deduce that

$$\mathcal{W}_3(A^{-1}\nu, N(0, I)) \ll_d C\varepsilon'^{1/4}$$

where $A = d^{1/2}(1-\rho^2)^{1/2} = \lim_{\ell\to\infty} d^{1/2} S_\ell^{-1/2}$. In the presence of a spectral gap, all of these bounds are easily seen to be uniform. □

10. Construction of Examples

Throughout this section we denote as usual by $G = \mathrm{Sim}(\mathbb{R}^d)$. We first study random walk entropy in section 10.1 and then the separation rate in section 10.2. We prove Corollary 1.8 on real Bernoulli convolutions in section 10.3 as well as treat complex Bernoulli convolutions in section 10.4 proving Corollary 1.9. Finally, we discuss examples in $\mathbb{R}^d$ in section 10.5 and show Corollary 1.10, Corollary 1.11 and Corollary 1.12.

10.1. **Bounding Random Walk Entropy.** The techniques from [HS17, Section 6.3] or [Kit23, Section 9.2] follow through to our setting. In particular we have the following using Breuillard's strong Tits alternative.

Proposition 10.1. *(*[HS17, Section 6.3]*) Let $d \geq 1$. Then for every $p_0 > 0$ there exists $\rho = \rho(p_0, d)$ such that if $\mu = \sum_{i=1}^{k} p_i \delta_{g_i}$ is a finitely supported probability measure on G with $p_i \geq p_0$ and* $\mathrm{supp}(\mu)$ *generates a non-virtually solvable subgroup, then $h_\mu \geq \rho$.*

We will also use the following version of the ping-pong lemma for which we provide a full proof for the convenience of the reader.

Lemma 10.2. *(Ping-Pong) Let G be a group acting on a set X and let $g_1, g_2 \in G$. Assume there exist disjoint non-empty sets $A_1, A_2 \subset X$ such*

$$g_1(A_1 \cup A_2) \subset A_1 \quad \text{and} \quad g_2(A_1 \cup A_2) \subset A_2.$$

Then g_1 and g_2 generate a free semigroup.

When this happens we say that g_1 and g_2 *play ping pong.*

Proof. Let $w_1 = h_1 h_2 \cdots h_{\ell_1}$ and $w_2 = f_1 f_2 \cdots f_{\ell_2}$ with distinct sequences $h_i, f_j \in \{g_1, g_2\}$. Assume without loss of generality that $\ell_1 \leq \ell_2$. First assume that there is some $1 \leq k \leq \ell_1$ such that $h_k \neq f_k$. Choose the smallest such k and note that it suffices to show that $h_k \cdots h_{\ell_1} \neq f_k \cdots f_{\ell_2}$, which follows by applying the resulting maps to any $x \in A_1 \cup A_2$ and noting that $h_k \cdots h_{\ell_1} x \neq f_k \cdots f_{\ell_2} x$. On the other hand assume that $h_i = f_i$ for all $1 \leq i \leq \ell_1$, in which case we need to show that $w' = f_{\ell_1+1} \cdots f_{\ell_2}$ is not the identity. Without loss of generality assume that $f_{\ell_1+1} = g_1$. Then for $x \in A_2$ we have that $w'x \in A_1$ and thus w' is not the identity. We note that in particular it follows by the assumptions that g_1 and g_2 have infinite order. □

Lemma 10.3. *Let μ be a finitely supported probability measure on G such that $g_1, g_2 \in \mathrm{supp}(\mu)$ generate a free semigroup. Then*

$$h_\mu \gg \min\{\mu(g_1), \mu(g_2)\}.$$

Proof. Denote $\mu' = \frac{1}{2}\delta_e + \frac{1}{2}\mu$. Then by [HS17, Lemma 6.8] we have $h_{\mu'} = h_\mu/2$. Thus the claim follows from Proposition [Kit23, Section 9.7] (generalised to G and applied to $K = \min\{\mu(g_1), \mu(g_2)\}/2$). □

10.1.1. *p-adic Ping-Pong.* We first use ping-pong in a p-adic setting. For a number field K with ring of integers O_K. Let $\mathfrak{p} \subset O_K$ be a prime ideal and we denote by $R_\mathfrak{p}$ the localization of O_K at P defined as

$$R_\mathfrak{p} = \left\{ \frac{a}{b} : a \in O_K, b \in O_K \backslash \mathfrak{p} \right\}.$$

Lemma 10.4. *(p-adic Ping-Pong) Let K be a number field and let O_K be its ring of integers. Let $\mathfrak{p} \subset O_K$ be a prime ideal and let $M_\mathfrak{p}$ be the ideal of $R_\mathfrak{p}$ defined by*

$$M_\mathfrak{p} = \left\{ \frac{a}{b} : a \in \mathfrak{p}, b \in O_K \backslash \mathfrak{p} \right\}.$$

Let $g_1, g_2 \in G$ be such that all of the entries of $\rho(g_1)U(g_1)$ and $\rho(g_2)U(g_2)$ are in $M_\mathfrak{p}$ and all components of b_1 and b_2 are in $R_\mathfrak{p}$. Suppose that

$$M_\mathfrak{p} \times \cdots \times M_\mathfrak{p} + b_1 \neq M_\mathfrak{p} \times \cdots \times M_\mathfrak{p} + b_2.$$

Then g_1 and g_2 generate a free semigroup.

Proof. This follows immediately from Lemma 10.2 with $X = R_\mathfrak{p} \times \cdots \times R_\mathfrak{p}$ and $A_i = M_\mathfrak{p} \times \cdots \times M_\mathfrak{p} + b_i$ for $i = 1, 2$. □

10.1.2. *Ping-Pong under a Galois transform.* We can also apply the ping-pong lemma using field automorphisms. Recall that given a number field K, the automorphism group $\mathrm{Aut}(K/\mathbb{Q})$ consists of field automorphisms that fix $\mathbb{Q}$.

Lemma 10.5. *(Galois Ping-Pong) Let g_1 and g_2 be two elements in G whose coefficients lie in a real number field K and without a common fixed point. Let $\Phi \in \mathrm{Aut}(K/\mathbb{Q})$ be such that for $i = 1, 2$ we have*

$$|\rho(\Phi(g_i))| < 1/3.$$

Then g_1 and g_2 generate a free semigroup.

Proof. For $i = 1, 2$ write $h_i = \Phi(g_i)$ and let p_i be the fixed point of h_i, which has coefficients in K since it arises from a linear equation over K. Then $h_1 \neq h_2$ as g_1 and g_2 have no common fixed point. Consider $A_i = B_{d(h_1,h_2)/2}(h_i)$ (the open ball around h_i of radius $d(h_1, h_2)/2$) and note further that $h_1(A_1 \cup A_2) \subset A_1$ and $h_2(A_1 \cup A_2) \subset A_2$. So the claim follows by Lemma 10.2. □

10.2. **Heights and Separation.** In this subsection we will review some techniques for bounding S_μ using heights. We recall the following definition.

Definition 10.6 (Height)**.** *Let $\alpha_1 \in \mathbb{C}$ be algebraic with algebraic conjugates*

$$\alpha_2, \alpha_3, \dots, \alpha_d.$$

Suppose that the minimal polynomial for α_1 over $\mathbb{Z}$ has positive leading coefficient a_0. Then we define the height *of α_1 by*

$$\mathcal{H}(\alpha_1) := \left(a_0 \prod_{i=1}^{n} \max\{1, |\alpha_i|\} \right)^{1/d}$$

and we define the logarithmic height *of α_1 to be $h(\alpha_1) = \log \mathcal{H}(\alpha_1)$.*

We wish to use this to bound the size of polynomials of algebraic numbers. To do this we need the following way of measuring the complexity of a polynomial.

Definition 10.7. *Given some polynomial $P \in \mathbb{Z}[X_1, X_2, \dots, X_n]$ we define the* length *of P, which we denote by $\mathcal{L}(P)$, to be the sum of the absolute values of the coefficients of P.*

We recall the following basic facts about heights.

Lemma 10.8. *The following properties hold:*

(i) $\mathcal{H}(\alpha^{-1}) = \mathcal{H}(\alpha)$ *for any non-zero algebraic number α.*
(ii) If α is a non-zero algebraic number of degree d,
$$\mathcal{H}(\alpha)^{-d} \leq |\alpha| \leq \mathcal{H}(\alpha)^d.$$
(iii) Given $P \in \mathbb{Z}[X_1, X_2, \dots, X_n]$ of degree at most $L_1 \geq 0$ in X_1, ..., $L_n \geq 0$ in X_n and algebraic numbers $\xi_1, \xi_2, \dots, \xi_n$ we have
$$\mathcal{H}(P(\xi_1, \xi_2, \dots, \xi_n)) \leq \mathcal{L}(P)\mathcal{H}(\xi_1)^{L_1} \dots \mathcal{H}(\xi_n)^{L_n}$$

Proof. (i) and (ii) are well-known and (iii) is [Mas16, Proposition 14.7]. □

Proposition 10.9. *Suppose that μ is a finitely supported measure on $G = \mathrm{Sim}(\mathbb{R}^d)$. Let S be the set of coefficients of $\rho(g), U(g)$ and $b(g)$ with $g \in \mathrm{supp}(\mu)$ supported on a finite set of points. Suppose that all of the elements of S are algebraic and let K be the number field generated by S. Then*

$$S_\mu \ll_d [K : \mathbb{Q}] \max(\{h(y) : y \in S\} \cup \{1\}).$$

Proof. We let $m, n \in \mathbb{Z}_{>0}$ and we consider an expression of the from

$$a_1^{-1} a_2^{-1} \dots a_n^{-1} b_1 b_2 \dots b_m.$$

We wish to show that this is either the identity or at least some distance away from the identity. Let $C := \max\{\mathcal{H}(y) : y \in S\}$. First note that

$$\rho(a_1^{-1} a_2^{-1} \dots a_n^{-1} b_1 b_2 \dots b_m) - 1$$

is a polynomial in elements of S and their inverses with length 2 and total degree at most $n + m$. Therefore by Lemma 10.8

$$H(\rho(a_1^{-1} a_2^{-1} \dots a_n^{-1} b_1 b_2 \dots b_m) - 1) \leq 2C^{m+n}$$

and so either $\rho(a_1^{-1} a_2^{-1} \dots a_n^{-1} b_1 b_2 \dots b_m) = 1$ or

$$|\rho(a_1^{-1} a_2^{-1} \dots a_n^{-1} b_1 b_2 \dots b_m) - 1| \geq 2^{-[K:\mathbb{Q}]} C^{-(m+n)[K:\mathbb{Q}]}.$$

By a similar argument we see that either $U(a_1^{-1}a_2^{-1}\dots a_n^{-1}b_1b_2\dots b_m) = I$ or

$$||U(a_1^{-1}a_2^{-1}\dots a_n^{-1}b_1b_2\dots b_m) - I|| \geq (d^{m+n}+1)^{-[K:\mathbb{Q}]}C^{-(m+n)[K:\mathbb{Q}]}$$

and that either $b(a_1^{-1}a_2^{-1}\dots a_n^{-1}b_1b_2\dots b_m) = 0$ or

$$|b(a_1^{-1}a_2^{-1}\dots a_n^{-1}b_1b_2\dots b_m)| \geq (d^{m+n}+1)^{-[K:\mathbb{Q}]}C^{-(m+n)[K:\mathbb{Q}]}.$$

Overall this means that either $a_1^{-1}a_2^{-1}\dots a_n^{-1}b_1b_2\dots b_m = \mathrm{Id}$ or

$$\log d(a_1^{-1}a_2^{-1}\dots a_n^{-1}b_1b_2\dots b_m, \mathrm{Id}) \gg_d -(m+n)(\log C+1)[K:\mathbb{Q}].$$

The result follows. □

10.3. **Real Bernoulli Convolutions.** In this section we prove Corollary 1.8, stating that there is $C \geq 1$ such that if λ is algebraic with

$$\lambda > 1 - C^{-1}\min\{\log M_\lambda, (\log\log M_\lambda)^{-2}\},$$

then the Bernoulli convolution ν_λ is absolutely continuous.

Proof of Corollary 1.8. As in the paragraph before Proposition 9.6, Bernoulli convolutions are uniformly non-degenerate. Since we are in $d = 1$ they are $(1,0)$-well-mixing and therefore Theorem 2.4 applies. For convenience write $\eta = \log M_\lambda$ and $h_\lambda = h_{\nu_\lambda}$. We don't keep track of possible enlargements of C. That Bernoulli convolutions are uniformly non-degenerate follows from Proposition 2.2. Then Theorem 2.4 implies that if

$$(1-\lambda)^{-1}h_\lambda > C\left(\max\{1, \log\eta/h_\lambda\}\right)^2, \tag{10.1}$$

then ν_λ is absolutely continuous. Recall that by [BV20, Theorem 5] (which is stated with logarithms base 2) there is an absolute $c_0 \in (0,1)$ such that $c_0\min(\log 2, \eta) \leq h_\lambda \leq \min(\log 2, \eta)$.

We proceed with a case distinction. First assume that $\eta \leq \log 2$. Then $c_0^{-1} \geq \eta/h_\lambda \geq 1$ and therefore by (10.1) the condition $(1-\lambda)^{-1}c_0\eta > C$ is sufficient for absolute continuity, which is equivalent to

$$\lambda > 1 - C^{-1}\eta. \tag{10.2}$$

Next assume that $\eta \geq \log 2$. Then $c_0\log 2 \leq h_\lambda \leq \log 2$ and so (10.1) gives

$$(1-\lambda)\max\{1, \log\eta + \log(c_0\log 2)^{-1}\}^2 < C^{-1}.$$

Note that $\max\{1, \log\eta + \log(c_0\log 2)^{-1}\} \leq 2\log(c_0\log 2)^{-1}\max\{1, \log\eta\}$. Therefore we get the condition

$$\lambda > 1 - C^{-1}\max\{1, \log\eta\}^{-2} = 1 - C^{-1}\min\{1, (\log\eta)^{-2}\}. \tag{10.3}$$

To deduce (1.4), we note that there is a unique $\eta' > 0$ with $\eta' = (\log\eta')^{-2}$ and this η' satisfies $2 \leq \eta' \leq 5/2$. Moreover $\log\eta < (\log\eta)^{-2}$ for $0 < \eta < \eta'$ and $\log\eta > (\log\eta)^{-2}$ for $\eta > \eta'$. Therefore (1.4) holds for $\eta < \log(2)$ and $\eta > 2\eta'$ by (10.2) and (10.3). In the range $\log(2) < \eta < 2\eta'$, we enlarge C to ensure that (1.4) holds. □

We note that if λ is algebraic and not the root of any non-zero polynomial with coefficients $0, \pm 1$, then $h_\lambda = 2$ and also as mentioned in Remark 5.10 of [Kit21], $M_\lambda \geq 2$. Therefore for such a λ, ν_λ is absolutely continuous if

$$\lambda > 1 - C^{-1}\min\{1, (\log\log M_\lambda)^{-2}\}. \tag{10.4}$$

10.4. **Complex Bernoulli Convolutions.**

Proof of Corollary 1.9. We can't directly apply Proposition 2.2 so we give a direct proof of mixing and non-degeneracy. First note that (1.5) ensures that there is some $c > 0$ and $T \geq 1$ depending only on ε such that the (c, T)-well-mixing property is satisfied.

To deal with non-degeneracy, we distinguish the case when $|\lambda| \leq \lambda_0$ and $|\lambda| \geq \lambda_0$ for some λ_0 sufficiently close to 1. As in the case of real Bernoulli convolution, for any given λ_0, the family of Bernoulli convolutions with $|\lambda| \leq \lambda_0$ are easily seen to be uniformly non-degenerate depending on λ_0. To deal with the case $\lambda \geq \lambda_0$, we rescale our measure to the one given by the law of $B_\lambda = \sqrt{1-|\lambda|^2} \sum_{i=0}^{\infty} \pm\lambda^i$ and denote the resulting measure by ν_λ'. Now let Σ be the covariance matrix of ν_λ' under the natural identification of $\mathbb{C}$ with $\mathbb{R}^2$. Note that the trace of Σ is 1 and we claim that the smallest eigenvalue of Σ is $\gg_\varepsilon 1$. Indeed, for a unit vector $x \in \mathbb{R}^2$ we want to estimate $x^T \Sigma x$, which is by identifying $\mathbb{C}$ with $\mathbb{R}^2$ equal to

$$\mathbb{E}[|B_\lambda \cdot x|^2] = (1 - |\lambda|^2) \sum_{i=0}^{\infty} |\lambda^i \cdot x|^2 \gg_\varepsilon 1,$$

which follows as $|\lambda^i \cdot x|^2 \gg |\lambda|^2$ unless λ^i and x are almost colinear, which is only the case for a very small proportion of i's. It follows that

$$\inf_{p \in P(2)} \mathbb{E}_{x \sim \mathcal{N}(0,\Sigma)}[|px|] \gg_\varepsilon 1$$

for p ranging in the orthogonal projections of $\mathbb{R}^2$ as in section 9.2. By for example Lemma 5.10 we know that $\mathcal{W}_1(\nu_\lambda', N(0,\Sigma)) \ll \sqrt{1-|\lambda|^2}$. Therefore for λ_0 sufficiently close to 1 in terms of ε, uniform non-degeneracy follows as in Lemma 9.9. Having establish uniform well-mixing and non-degeneracy, Corollary 1.9 is established by the same argument as the proof of Corollary 1.8. □

10.5. **Examples in $\mathbb{R}^d$.** In this section we prove Corollary 1.10, Corollary 1.11 and Corollary 1.12 on general examples with absolutely continuous self-similar measures.

Proof of Corollary 1.10. We first show that g_1 and g_2 generate a free semigroup for sufficiently large q by using Lemma 10.4. For simplicity we first treat the case when all of the entries are rational. Then consider the q-adic numbers $\mathbb{Q}_q$ and the q-adic integers $\mathbb{Z}_q$. As the $U_1, \dots, U_k$ and the $b_1, \dots, b_k$ are fixed, for a sufficiently large prime q all of their entries are in $\mathbb{Z}_q \backslash q\mathbb{Z}_q$. On the other hand, by construction $\rho(g_i) \in q\mathbb{Z}_q$ for $1 \leq i \leq k$ and as $q\mathbb{Z}_q$ is an ideal therefore also all of the entries of $\rho(g_i)U_i$ are in $q\mathbb{Z}_q$. By Lemma 10.4 it therefore suffices to check that $(q\mathbb{Z}_q)^d + b_1 \neq (q\mathbb{Z}_q)^d + b_2$ or equivalently $b_1 - b_2 \notin (q\mathbb{Z}_q)^d$, which is clearly the case for sufficiently large q. Thus g_1 and g_2 generate a free semigroup. The same argument applies in the general case for K the number field generated by the coefficients of the entries of g_i and by choosing any prime ideal that factors (q).

Thus it follows by Lemma 10.3 that $h_\mu \gg \varepsilon$ and note that by Lemma 10.8 it holds that $S_\mu \ll_{K,d} \log q$. Hence there exists a constant C depending on all the relevant parameters such that the self-similar measure of μ is absolutely continuous if

$$C|\chi_\mu| \leq \frac{1}{(\log \log q)^2}.$$

Therefore it remains to estimate the Lyapunov exponent. Indeed, note that

$$\log\left(\frac{q}{q+a_{i,q}}\right) = \log\left(1 - \frac{a_{i,q}}{q+a_{i,q}}\right) \geq \log\left(1 - \frac{q^{1-\varepsilon}}{q}\right) \gg -q^{-\varepsilon}.$$

Therefore $|\chi_\mu| \ll q^{-\varepsilon}$ and the claim follows for sufficiently large q. □

Proof of Corollary 1.11. As in the proof of Corollary 1.10, g_1 and g_2 generate a free semigroup for sufficiently large q and therefore $h_\mu \gg \varepsilon$. Write $\alpha_1 = p_1 + \ldots + p_{k-1}$ and $\alpha_2 = p_k$. Then we have as $\alpha_1 + \alpha_2 = 1$,

$$\mathbb{E}_{\gamma\sim\mu}[\rho(\gamma)] = \alpha_1 \frac{q}{q+3} + \alpha_2 \frac{q}{q-1} = \frac{q^2 + (4\alpha_2 - 1)q}{(q+3)(q-1)}$$

and thus

$$1 - \mathbb{E}_{\gamma\sim\mu}[\rho(\gamma)] = \frac{(q+3)(q-1) - (q^2 + (4\alpha_2 - 1)q)}{(q+3)(q-1)} = \frac{(3 - 4\alpha_2)q - 3}{(q+3)(q-1)}.$$

On the other hand, choosing $\hat{\rho} = \frac{q}{q+3}$ we have

$$\mathbb{E}_{\gamma\sim\mu}[|\hat{\rho} - \rho(\gamma)|] = \alpha_2 \left(\frac{q}{q-1} - \frac{q}{q+3}\right) = \frac{4q\alpha_2}{k(q+3)(q-1)}.$$

Thus it follows that

$$\lim_{q\to\infty} \frac{\mathbb{E}_{\gamma\sim\mu}[|\hat{\rho} - \rho(\gamma)|]}{1 - \mathbb{E}_{\gamma\sim\mu}[\rho(\gamma)]} = \frac{4\alpha_2}{3 - 4\alpha_2} < 1 \tag{10.5}$$

provided that $\alpha_2 = p_k < \frac{3}{8}$. If we assume that $p_k \leq \frac{1}{3}$ then we have that the limit in (10.5) is uniformly away from 1. As in Corollary 1.10, we have that $S_\mu \ll_{K,d} \log q$. Therefore by Theorem 1.7 there exists a constant C depending on all of the parameters such that μ is absolutely continuous if

$$C|\chi_\mu| \leq \frac{1}{(\log\log q)^2}.$$

As in Corollary 1.10 it follows that $|\chi_\mu| \ll q^{-1}$ and hence the claim follows. □

We next prove Corollary 1.12 and first show the following basic lemma.

Lemma 10.10. *Let K be a real algebraic number field satisfying $\mathbb{Q}(\sqrt{q}) \subset K$ for a prime q. Then there exists a field automorphism $\Phi \in \mathrm{Aut}(K/\mathbb{Q})$ such that $\Phi(\sqrt{q}) = -\sqrt{q}$.*

Proof. Write $K_0 = \mathbb{Q}(\sqrt{q})$ and assume that $K = K_0(\alpha_1, \ldots, \alpha_\ell)$ for some $\alpha_1, \ldots, \alpha_\ell \in K$. Denote by $\Theta \in \mathrm{Aut}(K_0/\mathbb{Q})$ the automorphism with $\Theta(\sqrt{q}) = -\sqrt{q}$. When $\ell = 1$ we consider the surjective map $K_0[X] \to K_0(\alpha)$ with $P \mapsto \Theta(P)(\alpha_1)$ for $\Theta(P)$ the polynomial to which all coefficients we have applied Θ. This map induces a field automorphism of $K_0(\alpha)$ with the required properties and our proof is concluded by an induction on ℓ with the same argument. □

Proof of Corollary 1.12. By Theorem 1.5 there exists $\tilde{\rho} \in (0,1)$ and $C \geq 1$ depending on d, ε and μ_U such that μ is absolutely continuous if $p_i \geq \varepsilon$ as well as $\frac{a_i + b_i\sqrt{q}}{c_i} \in (\tilde{\rho}, 1)$ for all $1 \leq i \leq k$ as well as

$$\frac{h_\mu}{|\chi_\mu|} \geq C\left(\max\left\{1, \log\frac{S_\mu}{h_\mu}\right\}\right)^2.$$

Let K be the number field generated by all the coefficients of elements in $\mathrm{supp}(\mu)$. Then by Lemma 10.10 there is a field automorphism $\Phi \in \mathrm{Aut}(K/\mathbb{Q})$ such that $\Phi(\sqrt{q}) = -\sqrt{q}$ and therefore we have that $|\rho(\Phi(g_j))| < \frac{1}{3}$ for $j = 1, 2$. Thus by Lemma 10.5 and Lemma 10.3 we have that $h_\mu \gg \varepsilon$. We also have $h_\mu \leq \log \varepsilon^{-1}$. On the other hand, it follows by Lemma 10.8 (iii) and Proposition 10.9 that $S_\mu \ll_{d,\mu_U} \log L$, which readily implied the claim upon changing the constant C. □

Lemma 10.11. *In the setting of Corollary 1.12, for $\varepsilon > 0$ choose*

$$a_i = \lceil \sqrt{q} \rceil - m_{i,q}, \qquad b_i = 2 \qquad c_i = 3\lceil \sqrt{q} \rceil$$

for $m_{i,q}$ an integer satisfying $m_{i,q} \in [0, q^{1/2-\varepsilon}]$ and any $d_i \in \mathbb{Z}^d$ with $|d_i|_\infty \leq \exp(\exp(q^{\varepsilon/3}))$. Then μ is absolutely continuous for sufficiently large q depending on d, p_0, ε and $U_1, \ldots, U_k$, provided $g_1, \ldots, g_k$ does not have a common fixed point.

Proof. It holds that $\frac{a_i + b_i\sqrt{q}}{c_i} \in (0,1)$ converges to 1 as $q \to \infty$ and that $|\frac{a_i - b_i\sqrt{q}}{c_i}| < \frac{1}{3}$. We next estimate the Lyapunov exponent of μ. Indeed, note that for q large enough,

$$\begin{aligned} \log\left(\frac{a_i + b_i\sqrt{q}}{c_i}\right) &\geq \log\left(\frac{\lceil \sqrt{q} \rceil - q^{1/2-\varepsilon} + 2\sqrt{q}}{3\lceil \sqrt{q} \rceil}\right) \\ &\geq \log\left(1 - \frac{2(\lceil \sqrt{q} \rceil - \sqrt{q}) + q^{1/2-\varepsilon}}{3\lceil \sqrt{q} \rceil}\right) \gg -q^{-\varepsilon} \end{aligned}$$

and therefore $|\chi_\mu| \ll q^{-\varepsilon}$. In our case, for large q we have $L = |d_i|_\infty = \exp(\exp(q^{\varepsilon/3}))$ and therefore $\log(\log L) = q^{\varepsilon/3}$. Thus for sufficiently large q we have that $C|\chi_\mu| \leq (\log\log L)^{-2} = q^{-2\varepsilon/3}$ and the claim follows. □

References

[ARHW21] A. Algom, F. Rodriguez Hertz, and Z. Wang, *Pointwise normality and Fourier decay for self-conformal measures*, Adv. Math. **393** (2021).

[BdS16] Y. Benoist and N. de Saxcé, *A spectral gap theorem in simple Lie groups*, Invent. Math. **202** (2016), no. 2, 337–361.

[BE88] M. F. Barnsley and J. H. Elton, *A new class of markov processes for image encoding*, Adv. in Appl. Probab **20** (1988), no. 1, 14–32.

[BG08] J. Bourgain and A. Gamburd, *On the spectral gap for finitely-generated subgroups of* SU(2), Invent. Math. **171** (2008), no. 1, 83–121.

[BISG17] R. Boutonnet, A. Ioana, and A. Salehi Golsefidy, *Local spectral gap in simple Lie groups and applications*, Invent. Math. **208** (2017), no. 3, 715–802.

[BKS24] S. Baker, O. Khalil, and T. Sahlsten, *Fourier decay from l2-flattening* (2024). `https://arxiv.org/abs/2407.16699`.

[Bou12] J. Bourgain, *Finitely supported measures on $SL_2(\mathbb{R})$ which are absolutely continuous at infinity*, Geometric aspects of functional analysis, 2012, pp. 133–141.

[Bré21] J. Brémont, *Self-similar measures and the Rajchman property*, Ann. H. Lebesgue **4** (2021), 973–1004.

[Bre08] E. Breuillard, *A strong tits alternative* (2008). `https://arxiv.org/abs/0804.1395`.

[BS23] S. Baker and T. Sahlsten, *Spectral gaps and fourier dimension for self-conformal sets with overlaps* (2023). `https://arxiv.org/abs/2306.01389`.

[BSSŚ22] B. Bárány, K. Simon, B. Solomyak, and A. Śpiewak, *Typical absolute continuity for classes of dynamically defined measures*, Adv. Math. **399** (2022).

[BV20] E. Breuillard and P. P. Varjú, *Entropy of Bernoulli convolutions and uniform exponential growth for linear groups*, J. Anal. Math. **140** (2020), no. 2, 443–481.

[DFW07] X.-R. Dai, D.-J. Feng, and Y. Wang, *Refinable functions with non-integer dilations*, J. Funct. Anal. **250** (2007), no. 1, 1–20.

[Erd39] P. Erdös, *On a family of symmetric Bernoulli convolutions*, Amer. J. Math. **61** (1939), 974–976.

[Eri73] R. V. Erickson, *On an L_p version of the Berry-Esseen theorem for independent and m-dependent variables*, Ann. Probability **1** (1973), 497–503. MR383502

[EW11] M. Einsiedler and T. Ward, *Ergodic theory with a view towards number theory*, Graduate Texts in Mathematics, vol. 259, Springer-Verlag London, Ltd., London, 2011. MR2723325

[Gar62] A. M. Garsia, *Arithmetic properties of Bernoulli convolutions*, Trans. Amer. Math. Soc. **102** (1962), 409–432.

[Gol91] C. M Goldie, *Implicit renewal theory and tails of solutions of random equa- tions*, Ann. Appl. Probab. **1** (1991), no. 1, 126–166.

[GP15] Y. Guivarc'h and E. L. Page, *On the homogeneity at infinity of the stationary probability for an affine random walk*, Recent trends in ergodic theory and dynamical systems, 2015, pp. 119–130.

[GP16] ———, *Spectral gap properties for linear random walks and pareto's asymptotics for affine stochastic recursions*, Ann. Inst. H. Poincaré Probab. Statist. **52** (2016), no. 2, 503–574.

[Hoc14] M. Hochman, *On self-similar sets with overlaps and inverse theorems for entropy*, Ann. of Math. (2) **180** (2014), no. 2, 773–822.

[Hoc17] ———, *On self-similar sets with overlaps and inverse theorems for entropy in R^d* (2017). `https://arxiv.org/pdf/1503.09043.pdf`.

[HS17] M. Hochman and B. Solomyak, *On the dimension of Furstenberg measure for $SL_2(\mathbb{R})$ random matrix products*, Invent. Math. **210** (2017), no. 3, 815–875.

[Hut81] J. E. Hutchinson, *Fractals and self-similarity*, Indiana Univ. Math. J. **30** (1981), no. 5, 713–747.

[Kes73] H. Kesten, *Random difference equations and renewal theory for products of random matrices*, Acta Math. **131** (1973), 207–248.

[Kes74] ———, *Renewal theory for functionals of a Markov chain with general state space*, Ann. Probability **2** (1974), 355–386.

[Kev16] P. Kevei, *A note on the kesten–grincevičius–goldie theorem*, Electron. Commun. Probab. **21** (2016), 1–12.

[Kit21] S. Kittle, *Absolutely continuous self-similar measures with exponential separation* (2021). `https://arxiv.org/abs/2103.12684`, to appear at Annales scientifiques de l'Ecole normale supérieure.

[Kit23] ———, *Absolutely continuous Furstenberg measures* (2023). `https://arxiv.org/pdf/2305.05757.pdf`.

[KK24] S. Kittle and C. Kogler, *Dimension theory of contracting on average self-similar measures* (2024). In preparation.

[Klo22] B. R. Kloeckner, *Optimal transportation and stationary measures for iterated function systems*, Math. Proc. Cambridge Philos. Soc. **172** (2022), no. 1, 163–187.

[Kog22] C. Kogler, *Local limit theorem for random walks on symmetric spaces* (2022). `https://arxiv.org/abs/2211.11128`.

[Leq22] F. Lequen, *Absolutely continuous Furstenberg measures for finitely-supported random walks* (2022). `https://arxiv.org/abs/2205.11138.pdf`.

[LS20] J. Li and T. Sahlsten, *Fourier transform of self-affine measures*, Adv. Math. **374** (2020).

[LS22] ———, *Trigonometric series and self-similar sets*, J. Eur. Math. Soc. (JEMS) **24** (2022), no. 1, 341–368.

[LV16] E. Lindenstrauss and P. P. Varjú, *Random walks in the group of Euclidean isometries and self-similar measures*, Duke Math. J. **165** (2016), no. 6, 1061–1127.

[Mas16] D. Masser, *Auxiliary polynomials in number theory*, Cambridge Tracts in Mathematics, vol. 207, Cambridge University Press, Cambridge, 2016.

[Rap22] A. Rapaport, *On the Rajchman property for self-similar measures on $\mathbb{R}^d$*, Adv. Math. **403** (2022), Paper No. 108375, 53. MR4405371

[Sak85] A. I. Sakhanenko, *Estimates in an invariance principle*, Matematicheskie Trudy [In Russian] **5** (1985), 27–44.

[Shm14] P. Shmerkin, *On the exceptional set for absolute continuity of Bernoulli convolutions*, Geom. Funct. Anal. **24** (2014), no. 3, 946–958.

[Sol22] B. Solomyak, *Fourier decay for homogeneous self-affine measures*, J. Fractal Geom. **9** (2022), no. 1-2, 193–206.

[Sol95] ———, *On the random series $\sum \pm\lambda^n$ (an Erdos problem)*, Ann. of Math. (2) **142** (1995), no. 3, 611–625.

[SS16a] P. Shmerkin and B. Solomyak, *Absolute continuity of complex bernoulli convolutions*, Math. Proc. Cambridge Philos. Soc. **161** (2016), 435–453.

[SS16b] ———, *Absolute continuity of self-similar measures, their projections and convolutions*, Trans. Amer. Math. Soc. **368** (2016), no. 7, 5125–5151.

[SS23] B. Solomyak and A. Spiewak, *Absolute continuity of self-similar measures on the plane* (2023). https://arxiv.org/abs/2301.10620, to appear at Indiana Univ. Math. J.

[SSS18] S. Saglietti, P. Shmerkin, and B. Solomyak, *Absolute continuity of non-homogeneous self-similar measures*, Adv. Math. **335** (2018), 60–110. MR3836658

[Str24] L. Streck, *On absolute continuity and maximal garsia entropy for self-similar measures with algebraic contraction ratio* (2024). https://arxiv.org/abs/2303.07785.

[SX03] B. Solomyak and H. Xu, *On the 'mandelbrot set' for a pair of linear maps and complex bernoulli convolutions*, Nonlinearity **16** (2003), no. 5, 1733–1749.

[Var19] P. P. Varjú, *Absolute continuity of Bernoulli convolutions for algebraic parameters*, J. Amer. Math. Soc. **32** (2019), no. 2, 351–397.

[Vig21] J. P. Vigneaux, *Entropy under disintegrations*, Geometric science of information, 2021, pp. 340–349.

[VY22] P. P. Varjú and H. Yu, *Fourier decay of self-similar measures and self-similar sets of uniqueness*, Anal. PDE **15** (2022), no. 3, 843–858.

Email address: s.kittle@ucl.ac.uk, kogler@maths.ox.ac.uk

Samuel Kittle, Department of Mathematics, University College London, 25 Gordon Street, London WC1H 0AY, United Kingdom

Constantin Kogler, Mathematical Institute, University of Oxford, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, United Kingdom