UNIVERSITY OF CALIFORNIA,
IRVINE

Nonstationary random dynamical systems

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Mathematics

by

Grigorii Monakov

Dissertation Committee:
Professor Anton Gorodetski, Chair
Professor Hamid Hezari
Professor Svetlana Jitomirskaya

2026

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisor, Anton Gorodetski. His support began even before I entered the PhD program and has been steadfast ever since. I am also incredibly grateful to my collaborator, Victor Kleptsyn, for his invaluable insights and contributions. The work in the foundation of this thesis is either joint with or inspired by conversations with Anton Gorodetski and Victor Kleptsyn.

I am grateful to the faculty members from whom I took courses and learned a great deal. In particular, I would like to thank Lana Jitomirskaya, Katya Krupchik, Paata Ivanishvili, Roman Vershinin, Connor Mooney, Hamid Hezari, Kenneth Ascher, and Knut Solna.

Finally, I want to thank all of my colleagues, friends, and family who supported and helped me along the way.

The work comprising this thesis was supported in part by NSF Grants No. DMS–1855541 and DMS–2247966.

VITA

Grigorii Monakov

EDUCATION

Doctor of Philosophy in Mathematics **2026**
University of California *Irvine, California*

Master of Science in Mathematics **2021**
Saint-Petersburg State University *Saint-Petersburg, Russia*

Bachelor of Science in Mathematics **2019**
Saint-Petersburg State University *Saint-Petersburg, Russia*

PUBLICATIONS AND PREPRINTS

Central Limit Theorem for non-stationary random products of $\mathrm{SL}(2,\mathbb{R})$ matrices **2026**
(joint with A. Gorodetski, V. Kleptsyn), arxiv:2411.12003

Log-Hölder regularity of stationary measures **2025**
Communications in Mathematical Physics

Non-stationary Itô-Kawada and Ergodic Theorems for random isometries **2025**
Groups, Geometry, and Dynamics

Generalized Bounded Distortion Property **2025**
(joint with G. Borissov), Journal of Dynamics and Differential Equations

Ergodic Theorem for nonstationary random walks on compact abelian groups **2024**
Proceedings of the American Mathematical Society

Hölder regularity of stationary measures **2025**
(joint with A. Gorodetski, V. Kleptsyn), Inventiones mathematicae

Probabilistic shadowing in linear skew products **2025**
(joint with S. Tikhomirov), Discrete and Continuous Dynamical Systems

A model of voting dynamics under bounded confidence with nonstandard norming **2022**
(joint with S. Pilyugin; M. Tarasova, A. Tarasov), Networks and Heterogeneous Media

Displacement of viscous fluids in a set of parallel pipes **2020**
(joint with S. Tikhomirov, A Yakovlev), Computational Mathematics and Mathematical Physics

ABSTRACT OF THE DISSERTATION

Nonstationary random dynamical systems

By

Grigorii Monakov

Doctor of Philosophy in Mathematics

University of California, Irvine, 2026

Professor Anton Gorodetski, Chair

This thesis is dedicated to studying the quantitative regularity of measures produced by actions of random dynamical systems and to applying these geometric estimates to products of random matrices. First, we consider smooth random dynamical systems defined by a distribution with a finite moment of the norm of the differential, and prove that under suitable non-degeneracy conditions any stationary measure must be Hölder continuous. This result is a vast generalization of the classical statement on Hölder continuity of stationary measures of random walks on linear groups. Second, we extend this framework to rougher settings by considering Lipschitz and Hölder continuous random dynamical systems defined by a distribution with a finite logarithmic moment, proving that under suitable non-degeneracy conditions every stationary measure must be log-Hölder continuous. Finally, we leverage these regularization properties to prove a Central Limit Theorem for non-stationary random products of $\mathrm{SL}(2,\mathbb{R})$ matrices, generalizing the classical results by Benoist and Quint that was obtained in the case of iid random matrix products.

The content of this thesis is based on three papers, two of which ([24, 23]) are joint work with Anton Gorodetski and Victor Kleptsyn, and the last of which ([34]) was prepared with their continuous support and advice.

Chapter 1

Introduction

1.1 Regularity of stationary measures

1.1.1 Hölder regularity of stationary measures

Let M be a smooth closed Riemannian manifold and $f : M \to M$ be a homeomorphism. One of the most extensively studied objects in this setting is the *invariant measure*.

Definition 1.1.1. Borel measure ν on M is called invariant if $f_*\nu = \nu$.

This thesis is dedicated to studying the random analog of invariant measures called *stationary measures*. Before we can define this concept we need to introduce random dynamical systems. Let μ be a Borel probability measure on $\mathrm{Diff}^1(M)$, the set of C^1-diffeomorphisms of M. Consider the corresponding random dynamical system, given by the compositions

$$T_n := f_n \circ \cdots \circ f_1,$$

where $f_i \in \mathrm{Diff}^1(M)$ are chosen randomly and independently, with respect to the distribution

μ.

If an initial point $x \in M$ is distributed with respect to a probability measure ν, one can consider the distribution $\mu*\nu$ of its random image $f(x)$. In other words, $\mu*\nu$ is the μ-averaged push-forward image of the measure ν:

$$\mu * \nu := \int_{\mathrm{Diff}^1(M)} (f_*\nu)\, d\mu(f).$$

Definition 1.1.2. The measure ν is called *μ-stationary* if $\mu * \nu = \nu$.

As it was mentioned above, stationary measures of random dynamical systems are natural analogues of invariant measures of deterministic maps. Their properties are of crucial importance for many results in random dynamics, see [2, 12, 6, 10, 14, 15, 29, 28, 31, 32] and references therein.

One of the important properties is regularity (or singularity) with respect to a reference measure. In our setting the reference measure is the Lebesgue measure on the manifold M determined by the Riemannian structure. We define two following types of measure regularity:

Definition 1.1.3. We say that the measure ν on M is (C, α)-Hölder, if

$$\forall x \in M \quad \forall r > 0 \quad \nu(B_r(x)) < Cr^{\alpha}.$$

The measure ν on M is (C, α)-log-*Hölder* if

$$\forall x \in M \quad \forall r > 0 \quad \nu(B_r(x)) \leq C|\log(r)|^{-\alpha}.$$

For a diffeomorphism $f \in \mathrm{Diff}^1(M)$ denote

$$\mathfrak{L}(f) = \max(\mathrm{Lip}(f), \mathrm{Lip}(f^{-1})). \tag{1.1.1}$$

The first one of the main results of this work (discussed in more details and proved in Chapter 2) can be formulated as follows:

Theorem 1.1.4. *Suppose that μ is a probability distribution on $\mathrm{Diff}^1(M)$ such that*

$$\int \mathfrak{L}(f)^\gamma d\mu(f) < \infty \quad \textit{for some } \gamma > 0.$$

Suppose also that there is no probability measure on the manifold M invariant under every map $f \in \mathrm{supp}\,\mu$. Then every stationary measure of a random dynamical system defined by the distribution μ is Hölder continuous.

Remark 1.1.5. While we formulate the results for C^1-diffeomorphisms, only smoothness of the phase space is important for the proof, not the smoothness of the maps. We only need the maps to be bi-Lipschitz. Hence, Theorem 1.1.4 holds for random dynamical systems defined by a probability measure μ on the space of bi-Lipschitz homeomorphisms: $\{f \in \mathrm{Homeo}(M) \mid \mathfrak{L}(f) < \infty\}$. For the precise statement see Theorem 2.1.3.

Remark 1.1.6. An assumption that the maps $\{f_i\}$ are identically distributed can be essentially relaxed, see Theorem 2.1.8 below. An example showing the necessity of the finite moment in Theorem 1.1.4 is provided in Section 2.6. At the same time, weakening the assumption on finite moments and/or lowering regularity of maps to Hölder continuous leads to lowering regularity of the stationary measures to log-Hölder, see Section 1.1.2 and Chapter 3. For the discussion of necessity of other conditions see remarks 2.1.5, 2.1.9, and 2.1.10.

In Section 2.1 below we provide formal definitions and more general and stronger versions of this result. Specifically, in Theorem 2.1.6 we show that averaged images of any initial measure

become Hölder regular on any scale larger than some threshold, which decays exponentially in the number of iterates. Theorem 2.1.8 is a similar result for a non-stationary case, namely, it claims that, informally speaking, by averaging with respect to a different distribution on each step, one "regularizes" a probability measure on a manifold, bringing it closer to the subspace of Hölder measures.

1.1.2 log-Hölder regularity of stationary measures

The Theorem 1.1.4 gives rise to a series of natural questions. Namely, one can ask

- What do we get if we impose a weaker condition on the tails of distribution μ? For instance, if we only assume that $\int (\log \|f\|_{\mathrm{Diff}^1})^\gamma d\mu(f) < \infty$.
- Is it possible to weaken the regularity assumption on the dynamical system? For example, what if we only assume that homeomorphisms $f \in \mathrm{supp}(\mu)$ are Hölder continuous and not differentiable?

Chapter 3 addresses both of these questions. Our main results in this direction are the following two theorems (for rigorous formulations and further discussion see Section 3.1):

Theorem 1.1.7. *Let μ be a probability measure on $\mathrm{Lip}(M)$ – space of bi-Lipschitz homeomorphisms of M, satisfying the following assumptions:*

- *Lipschitz constant has finite α-logarithmic moment with respect to μ.*
- *There is no measure m on M, such that $f_*m = m$ for μ-a.e. f.*

Then every μ-stationary measure is $\alpha/2$-log-Hölder.

Theorem 1.1.8. *Let μ be a probability measure on $\mathrm{Hol}(M)$ – space of bi-Hölder homeomorphisms of M, satisfying the following assumptions:*

- *Hölder constant has a finite logarithmic moment with respect to μ.*
- *There is no measure m on M, such that $f_*m = m$ for μ-a.e. f.*

Then every μ-stationary measure is log*-Hölder.*

Remark 1.1.9. Both Theorem 1.1.7 and Theorem 1.1.8 have non-stationary versions that deal with finite number of random iterations, see Theorem 3.1.4 and Theorem 3.1.9.

Remark 1.1.10. Moreover, at the end of Chapter 3 (see Section 3.5) we provide an even more precise (and, inevitably, more technical) result in the spirit of log-Hölder continuity about interaction of two measures generated by non-stationary random dynamical systems. That statement allows us to prove a non-stationary Central Limit Theorem for products of random $\mathrm{SL}(2,\mathbb{R})$ matrices with optimal assumption on the tails (finite $2+\varepsilon$ moment for any $\varepsilon > 0$, see Section 1.2 and Chapter 4).

1.2 Application: Central Limit Theorem for products of random $\mathrm{SL}(2,\mathbb{R})$ matrices

1.2.1 Historical background

The two most fundamental results in probability that are present in almost every textbook are the (strong) Law of Large Numbers (LLN) and the Central Limit Theorem (CLT). In the most basic form, if $\{\xi_n\}$ is an iid sequence of random variables with finite expectation a and finite variance σ^2, the LLN claims that almost surely $\frac{1}{n}\sum_{i=1}^{n}\xi_i \to a$, and the CLT claims that $\frac{\sum_{i=1}^{n}\xi_i - na}{\sqrt{n}\sigma}$ converges in distribution to a normal distribution $\mathcal{N}(0,1)$ with mean 0 and variance 1.

There are many ways to relax the assumptions in both cases. In particular, the random variables do not have to be identically distributed. For example, a non-stationary version of the LLN known as *Kolmogorov's Law* [30] claims that if $\{\xi_i\}$ is a sequence of independent random variables with $a_i = \mathbb{E}\xi_i$, $\sigma_i^2 = \mathrm{Var}\,(\xi_i)$, and $\sum_{i=1}^{\infty} \frac{\sigma_i^2}{i^2} < \infty$, then $\frac{\sum_{i=1}^{n}(\xi_i - a_i)}{n} \to 0$ almost surely. On the other hand, if for some $\delta > 0$ the sequence $\mathbb{E}|\xi_i|^{2+\delta}$ is uniformly bounded, then the sequence of random variables $\left(\sum_{i=1}^{n}(\xi_i - a_i)\right) / \sqrt{\sum_{i=1}^{n}\sigma_i^2}$ converges in distribution to $\mathcal{N}(0,1)$, provided[1] $\sum_{i=1}^{n}\sigma_i^2 \to \infty$.

There are plenty of different generalizations and forms of these statements. For example, for some of the analogs of the LLN and CLT for the sums of iid random variables in the context of random walks on groups see the survey [13] and monograph [9], and references therein. Here we discuss random matrix products. In this case, a multiplicative version of the LLN is given by Furstenberg and Kesten [17]. A stronger result is the famous Furstenberg Theorem, which also guarantees positivity of the Lyapunov exponent. In order to formulate it we will need some algebraic conditions on the distribution of our random matrices.

Let $\{X_k, k \geq 1\}$ be independent and identically distributed random variables, taking values in $\mathrm{SL}(d,\mathbb{R})$, the $d \times d$ matrices with determinant one, let G_X be the smallest closed subgroup of $\mathrm{SL}(d,\mathbb{R})$ containing the support of the distribution of X_1.

Definition 1.2.1. G_X is called *strongly irreducible* if there is no finite family of proper non-zero subspaces $V_1, V_2, \ldots, V_N \subset \mathbb{R}^d$ such that

$$X(\cup_{i=1}^{N} V_i) = \cup_{i=1}^{N} V_i \text{ for all } X \in G_X,$$

and *proximal* (or *contracting*) if there exists a sequence $X_1, X_2, \ldots, \in G_X$ and a sequence of real numbers $x_1, x_2, \ldots$ such that the sequence of operators $\{x_i X_i\}$ converges in norm to a

[1] There are stronger versions of CLT for sums of independent not identically distributed random variables, such as Lyapunov or Lindeberg–Feller Theorems; here we provide the statement that is a direct additive analog of the main result of this thesis.

linear endomorphism of $\mathbb{R}^d$ of rank one.

See [10] for a detailed discussion of these notions.

Theorem 1.2.2 (H. Furstenberg [16]). *Assume that*

$$\mathbb{E}[\log \|X_1\|] < \infty, \text{ where } \|X_1\| = \sup_{v\in\mathbb{R}^d, |v|=1} |X_1 v|.$$

Also, assume that G_X is not compact and is strongly irreducible.Then there exists a positive constant λ_F (Lyapunov exponent) such that with probability one

$$\lim_{n\to\infty} \frac{1}{n} \log \|X_n \dots X_2 X_1\| = \lambda_F > 0.$$

Remark 1.2.3. In the case of random products of $\mathrm{SL}(2,\mathbb{R})$ matrices, the assumption that G_X is not compact and is strongly irreducible is equivalent to the assumption that there exists no measure on $\mathbb{RP}^1$ invariant under the action of every map from G_X, see [4, Lemma 3.6].

The CLT for the products of iid random matrices is also available. The initial results were obtained for matrices with positive coefficients [5], [17]. In the case of absolutely continuous distributions it was obtained by Tutubalin [38, 39]. The requirements on regularity of distributions was relaxed by Le Page, who proved the CLT for random matrix products under the assumption of finite exponential moments [35], see also [10], [19], [26], [20], [27]. Finally, the assumption on the moments of the distribution was optimized by Benoist and Quint [7]:

Theorem 1.2.4 (Benoist, Quint, [7]). *Let $\{X_k, k \geq 1\}$ be independent and identically distributed random matrices in $\mathrm{SL}(d,\mathbb{R})$. Assume that G_X is non-compact and strongly irreducible and*

$$\mathbb{E}\left[(\log \|X_1\|)^2\right] < \infty. \tag{1.2.1}$$

Then there exists $\sigma > 0$ such that the random variables

$$\frac{\log \|X_n \dots X_1\| - n\lambda_F}{\sqrt{n}},$$

where $\lambda_F > 0$ is the Lyapunov exponent, converge in distribution to $\mathcal{N}(0, \sigma^2)$.

Notice that both Theorems 1.2.2 and 1.2.4 require the random matrices to be identically distributed. That requirement allows to consider a stationary measure for the random dynamics on the projective space, which is a key notion used in the proofs of both results. For example, the proofs of the earlier versions of CLT relied on the following result by Guivarc'h:

Theorem 1.2.5 (Guivarc'h, [25])**.** *Suppose that, in the setting above, G_X is strongly irreducible and proximal, and*

$$\mathbb{E}\|X\|^\gamma < \infty$$

for some $\gamma > 0$. Then the random dynamical system given by the random projective maps f_{X_j} has unique stationary measure ν on $\mathbb{RP}^{d-1}$, and ν is Hölder continuous (here f_X is the projectivisation of the matrix X (see Eq. (4.5.1) *below)).*

An alternative proof of Theorem 1.2.5 (inspired by [11]) can be found in [6]. A proof of Hölder continuity of stationary measures under somewhat weaker assumptions was obtained in [1].

Notice that Theorem 1.2.5 can be considered a partial case of Theorem 1.1.4. Indeed, strong irreducibility and proximality conditions imply absence of common invariant measure for the corresponding projective maps (this follows, for example, from Propositions 1.6 and 2.3 in [10, Chapter III]), and finite moment condition implies the corresponding condition in Theorem 1.1.4.

Later on Benoist and Quint proved a more precise version of Theorem 1.2.5 that allowed logarithmic tails (and established an analog of log-Hölder regularity for the stationary measure),

and used that result to establish Theorem 1.2.4.

At the same time, the classical LLN and CLT hold for sums of independent and not necessarily identically distributed real valued random variables, and it is natural to expect that non-stationary versions of the LLN and CLT for random matrix products should hold as well. Indeed, the non-stationary version of the Furstenberg Theorem was recently provided in [22], and it already found interesting applications in spectral theory [21, 41]. The non-stationary version of the CLT for random products of $\mathrm{SL}(2,\mathbb{R})$ matrices (see Theorem 1.2.6) is the final main result of this thesis (and Chapter 4 of this thesis), and Theorem 3.5.1 in Chapter 3 is specifically designed to play the same role in its proof as Theorem 1.2.5 and its generalizations play in the proof of 1.2.4.

1.2.2 Definitions and statement of main results

Let us now provide the setting needed to state CLT for nonstationary products of $\mathrm{SL}(2,\mathbb{R})$ random matrices.

Let $\mathcal{K}$ be a compact subset (with respect to the weak-* topology) in the set of probability measures on the group $\mathrm{SL}(2,\mathbb{R})$. We will say that *the measures condition* is satisfied if for every measure $\mu \in \mathcal{K}$ there are no Borel probability measures ν_1, ν_2 on $\mathbb{RP}^1$ such that $(f_A)_*\nu_1 = \nu_2$ for μ-almost every $A \in \mathrm{SL}(2,\mathbb{R})$, where f_A is the projectivisation of the matrix A (see Eq. (4.5.1) below).

Let us fix some sequence $\{\mu_i\}_{i\in\mathbb{N}}$, $\mu_i \in \mathcal{K}$, and let $A_i \in \mathrm{SL}(2,\mathbb{R})$ be independent matrix-valued random variables, with A_i being distributed w.r.t. μ_i. Set

$$T_n = A_n A_{n-1} \dots A_1,$$

and denote

$$L_n = \mathbb{E} \log \|T_n\|. \tag{1.2.2}$$

If the measures condition is satisfied, then for any $\{\mu_i\}_{i\in\mathbb{N}}$, $\mu_i \in \mathcal{K}$, the sequence $\{L_n\}$ must grow at least linearly, i.e. the norms of the random products must grow exponentially on average, see [22, Theorem 1.5]. A related statement on exponential growth of the norms in the case of non-stationary linear cocycles over Markov chains was established by Goldsheid [18]. Moreover, if additionally a uniform bound on some exponential moment exists for distributions from $\mathcal{K}$, the non-random sequence $\{L_n\}$ describes the behavior of almost every random product, and in this sense serves as a non-stationary analog of Lyapunov exponent. Namely, almost surely one has $\lim_{n\to\infty} \frac{1}{n} (\log \|T_n\| - L_n) = 0$, see [22, Theorem 1.1]. This provides a direct analog of the LLN for non-stationary random matrix products.

That compels the question whether an analog of CLT for non-identically distributed random variables must hold in this setting. The main result of Chapter 4 provides a positive answer in dimension two:

Theorem 1.2.6. *Let $\mathcal{K}$ be a compact subset in the set of probability measures on the group $\mathrm{SL}(2,\mathbb{R})$ that satisfies the measures condition, and assume that there exists $\gamma > 2$ and $M > 0$ such that for any $\mu \in \mathcal{K}$ one has*

$$\mathbb{E}_\mu (\log \|A\|)^\gamma < M. \tag{1.2.3}$$

Then the random variables

$$\frac{\log \|T_n\| - L_n}{\sqrt{\mathrm{Var}(\log \|T_n\|)}}$$

converge in distribution to $\mathcal{N}(0,1)$, with the convergence that is uniform with respect to the choice of the sequence $\mu_1, \mu_2, \dots \in \mathcal{K}$.

Also, there are constants $C_1, C_2 > 0$ and an index n_0 such that for all $n \geq n_0$ and all $\mu_1, \ldots, \mu_n \in \mathcal{K}$ one has

$$C_1^2 n \leq \mathrm{Var}(\log \|T_n\|) \leq C_2^2 n. \tag{1.2.4}$$

Remark 1.2.7. (a) The condition (1.2.3) with the assumption $\gamma > 2$ is optimal, in a sense that it cannot be strengthened to $\gamma = 2$. This is in contrast with the iid case, compare with Theorem 1.2.4. We discuss this below, see Example 4.9.2 in Section 4.9.

(b) One should expect that, under suitable conditions, Theorem 1.2.6 should hold for random $\mathrm{SL}(d, \mathbb{R})$ matrix products for every $d \geq 2$. To prove such a statement, it would be helpful to have a non-stationary analog of simplicity of the Lyapunov spectrum, see [26], [20] for the case of iid random matrix products. In the case of some specific regular distributions in $\mathrm{SL}(d, \mathbb{R})$ such an analog was recently established [3], but a statement for a general sequence of distributions is currently not available, even if certainly expected.

The conclusion of Theorem 1.2.6 also applies to the distribution of log-lengths of individual vectors, $\log |T_n v|$, as well as to the matrix elements $(T_n)_{i,j}$.

Theorem 1.2.8. *Under the assumptions of Theorem 1.2.6, for any nonzero $v \in \mathbb{R}^2$ and any $i, j = 1, 2$ the random variables*

$$\frac{\log |T_n v| - L_n}{\sqrt{\mathrm{Var}(\log \|T_n\|)}}, \quad \frac{\log |(T_n)_{ij}| - L_n}{\sqrt{\mathrm{Var}(\log \|T_n\|)}}$$

converge in distribution to $\mathcal{N}(0, 1)$, where $(T_n)_{ij}$ is the (i, j)-th element of the matrix T_n.

Chapter 2

Hölder regularity of stationary measures

2.1 Definitions and main results

Definition 2.1.1. We say that the measure ν on M is (α, C)-Hölder, if

$$\forall x \in M \quad \forall r > 0 \quad \nu(B_r(x)) < Cr^{\alpha}. \tag{2.1.1}$$

The measures that satisfy (2.1.1) appear in geometric measure theory literature under the names "upper Ahlfors regular", or "Frostman measure".

Definition 2.1.2. For a diffeomorphism $f \in \mathrm{Diff}^1(M)$ we define

$$\mathrm{Lip}(f) = \sup_{x \neq y} \left(\frac{d(f(x), f(y))}{d(x, y)} \right)$$

and

$$\mathfrak{L}(f) = \max(\mathrm{Lip}(f), \mathrm{Lip}(f^{-1})).$$

Our first main result claims that Hölder regularity of stationary measures is completely abundant. The statement below is a more formal and detailed version of Theorem 1.1.4.

Theorem 2.1.3. *Let μ be a probability measure on $\mathrm{Diff}^1(M)$, satisfying the following assumptions:*

- ***(finite moment condition)*** *There exists $\gamma > 0$ and $C_0 > 0$ such that*

$$\int \mathfrak{L}(f)^\gamma \, d\mu(f) < C_0; \tag{2.1.2}$$

- ***(no common invariant measure)*** *There is no probability measure m on M such that $f_* m = m$ for μ-a.e. f.*

Then there exist $\alpha > 0$ and C such that any μ-stationary probability measure ν on M is (α, C)-Hölder.

Remark 2.1.4. Denote by $\mathcal{M}$ the space of all probability measures on $\mathrm{Diff}^1(M)$ equipped with weak-$*$ topology. If $\mathcal{K} \subset \mathcal{M}$ is a compact set, such that the assumptions of Theorem 2.1.3 (with uniform $\gamma > 0$ and $C_0 > 0$) hold for all $\mu \in \mathcal{K}$, then α and C in the conclusion can be chosen uniformly in $\mu \in \mathcal{K}$.

Remark 2.1.5. Notice that the condition on absence of common invariant measures cannot be dropped without adding some other assumptions on a model. Indeed, otherwise it could happen that all maps from $\operatorname{supp} \mu$ preserve the same non-Hölder measure, e.g. some atomic measure.

Theorem 2.1.3 states that (under suitable moment condition) all the stationary measures for a random dynamical system with no common invariant measure are uniformly Hölder. Thus, it is natural to ask if (averaged) iterations $\mu^{*n} * \nu_0$ of a given non-stationary initial measure ν_0 become "more and more Hölder" as the number n of iterations increases.

However, if both initial measure ν_0 and the measure μ are atomic (for instance, if ν_0 is a Dirac measure, and the measure μ is supported on a finite number of diffeomorphisms), then any of these iterated images has atoms, and thus cannot be Hölder. Hence, one can only hope (and expect) the Hölder property on not-too-small scales. And indeed, it turns out to be the case: after n iterations one has Hölder property on all scales above an exponentially small one. Namely, we have the following theorem, our second main result:

Theorem 2.1.6. *Assume that μ satisfies the assumptions of Theorem 2.1.3. Then there exist $\alpha > 0$, C and $\kappa < 1$ such that for any initial measure ν_0, any number of iterations $n \in \mathbb{N}$, and any $x \in M$ one has:*

$$\text{if } \; r > \kappa^n \quad \text{then } \; (\mu^{*n} * \nu_0)(B_r(x)) < Cr^\alpha. \tag{2.1.3}$$

Remark 2.1.7. Equivalently, the conclusion (2.1.3) in Theorem 2.1.6 can be replaced by the following one (perhaps, with different values of the constants):

$$\text{for any } \; r > 0 \quad \text{we have } \; (\mu^{*n} * \nu_0)(B_r(x)) < C(r^\alpha + \kappa^n).$$

Notice that Theorem 2.1.3 is an immediate consequence of Theorem 2.1.6. Indeed, if ν_0 is a stationary measure, then $\mu^{*n} * \nu_0 = \nu_0$, and since $\kappa^n \to 0$ as $n \to \infty$, applying Theorem 2.1.6 to the measure ν_0 shows that ν_0 is Hölder continuous, and hence Theorem 2.1.3 follows.

Finally, in some situations one has to consider non-stationary random dynamical systems, where the maps f_i applied on different steps are chosen with respect to different distributions μ_i. One example of such situation is non-stationary version of the Furstenberg theorem [22, 18], related to the Anderson Localisation for random Schrödinger operators in $l^2(\mathbb{Z})$, usually referred to as *Anderson Model*, in presence of a non-constant background potential.

Our third theorem states that in the non-stationary setting, under natural genericity as-

sumptions the random (averaged) image of any given probability measure on M after n iterations satisfies the Hölder property on not-too-small scales.

Theorem 2.1.8. *Let $\mathcal{K} \subset \mathcal{M}$ be a compact, satisfying the following conditions:*

- ***(uniform finite moment condition)*** *There exists $\gamma > 0$, C_0 such that for any $\mu \in \mathcal{K}$ one has*

$$\int \mathfrak{L}(f)^\gamma \, d\mu(f) < C_0; \tag{2.1.4}$$

- ***(no deterministic images)*** *For any $\mu \in \mathcal{K}$ there are no probability measures ν, ν' on M such that $f_*\nu = \nu'$ for μ-almost all $f \in \mathrm{Diff}^1(M)$.*

Then there exist $\alpha > 0$, C, $\kappa < 1$ such that for any initial measure ν_0, any n, and any distributions $\mu_1, \dots, \mu_n \in \mathcal{K}$ the n-th image of ν_0 satisfies (α, C)-Hölder property on the scales larger than κ^n:

$$\forall x \in M \quad \forall r > \kappa^n \quad (\mu_n * \cdots * \mu_1 * \nu_0)(B_r(x)) < Cr^\alpha.$$

Remark 2.1.9. Theorems 2.1.3, 2.1.6 and 2.1.8 also hold for the case of M being a compact manifold with a boundary. In this case we consider μ that is a probability measure on the space of all diffeomorphisms of M on the image.

Remark 2.1.10. It is certainly reasonable to ask whether one can relax any other assumptions in Theorem 1.1.4. For example, what if the random dynamical systems is defined by diffeomorphisms that preserve the Lebesgue measure on the phase space? Can one claim in this case that under suitable assumptions all the stationary measures must be Hölder? Another natural question is whether in the case the independence assumption is dropped, the property (2.1.3) still holds? The following simple example shows that in general one should

not expect it. Consider an iid sequence of random diffeomorphisms $f_1, f_2, \ldots$, and construct another sequence of random maps $g_i = f_{i+1}^{-1} \circ f_i$. Then $\{g_i\}$ is a sequence of random maps such that g_i and g_k are independent if $|i-k| > 2$, but it is clear that one should not expect (2.1.3) to hold.

2.1.1 "No invariant measure" vs "no deterministic images" conditions

The "no deterministic images" assumption in Theorem 2.1.8 is a natural generalisation to the non-stationary setting of the "no common invariant measures" one. Indeed, in the non-stationary setting it does not make sense to directly compare measures before and after applying a random map, since by taking any (bounded) family of maps g_i and considering a family of conjugations $f_i \mapsto \tilde{f}_i := g_i \circ f_i \circ g_{i-1}^{-1}$ one can turn random orbits $x_i = f_i(x_{i-1})$ into random orbits $y_i = \tilde{f}_i(y_{i-1})$ given by $y_i = g_i(x_i)$, not changing any essential properties of the non-stationary random dynamical system, but destroying a common invariant measure.

Nevertheless, notice that in the stationary case, when both of these conditions make sense, "no common invariant measure" and "no measure with deterministic images" are essentially different: absence of one-step deterministic images is a strictly stronger assumption than absence of a common invariant measure. Indeed, for instance, one can take two diffeomorphisms f, g of the circle, where f is a North-South map with two fixed points (so the only invariant measures are concentrated at these points), and g is a sufficiently small irrational rotation (thus having only one invariant measure, namely Lebesgue measure on the circle). Thus, f and g have no common invariant measure. However, the composition $g^{-1} \circ f$ (that is close to f) has a fixed point x_0 (near a fixed point of the map f). The relation $g^{-1} \circ f(x_0) = x_0$ implies that $f(x_0) = g(x_0)$, and thus the Dirac measure at x_0 has a deterministic image.

However, being able to compress several steps into a single one makes these two assumptions

equivalent for their usage in the above theorems:

Proposition 2.1.11. *Let $\mu \in \mathcal{M}$ be a probability measure on $\mathrm{Diff}^1(M)$ such that there is no common invariant measure for all $f \in \operatorname{supp}\mu$. Then there exists $k \in \mathbb{N}$ such that μ^{*k} satisfies "no deterministic images" condition (i.e. there are no probability measures ν, ν' on M such that $f_*\nu = \nu'$ with $f = f_k \circ \ldots \circ f_1$ for $\mu \times \mu \times \ldots \times \mu$-almost all $(f_1, f_2, \ldots, f_k) \in \left(\mathrm{Diff}^1(M)\right)^k$).*

Proof of Proposition 2.1.11. Assume that for arbitrarily large k there exist measures ν_k, ν_k' such that $f_*\nu_k = \nu_k'$ for μ^{*k}-a.e. $f \in \mathrm{Diff}^1(M)$. For any such k consider the intermediate images (note that intermediate images must be deterministic, too):

$$\nu_{k,0} := \nu_k, \quad \nu_{k,j} = f_*\nu_{k,j-1}, \quad j = 1, \ldots, k, \quad \text{for } \mu\text{-a.e. } f \in \mathrm{Diff}^1(M).$$

Let $\bar{\nu}_k$ be the time averages of these measures:

$$\bar{\nu}_k := \frac{1}{k}\sum_{j=0}^{k-1} \nu_{k,j}.$$

Then, due to the standard time-averaging (Krylov-Bogolyubov) argument any accumulation point $\bar{\nu}$ of the measures $\bar{\nu}_k$ is a common invariant measure for μ-a.e. $f \in \mathrm{Diff}^1(M)$. Namely, for μ-a.e. $f \in \mathrm{Diff}^1(M)$ one has

$$f_*\bar{\nu}_k = \bar{\nu}_k + \frac{1}{k}(f_*\nu_{k,k-1} - \nu_{k,0}),$$

providing $f_*\bar{\nu} = \bar{\nu}$ after passing to a limit point $\bar{\nu}$. This leads to a contradiction with the assumption of the proposition. □

Notice that Proposition 2.1.11 implies that it suffices to prove Theorems 2.1.3 and 2.1.6 under a more restrictive "no deterministic images" assumption instead of the "no common invariant measures" condition. In particular, Theorem 2.1.6 can be considered a corollary of

Theorem 2.1.8.

2.2 Ideas of the proof

In this section we informally present the ideas and motivations behind the proofs of our main results, before passing to the formal proofs in Section 2.4.

We start with slightly weaker, but less technically complicated arguments, that allow us to establish the *existence* of a Hölder stationary measure.

As the first step, for any $\alpha > 0$ and a measure ν on M define its *energy* as

$$\mathcal{E}_\alpha(\nu) := \iint_M d(x,y)^{-\alpha} d\nu(x) d\nu(y) \tag{2.2.1}$$

An immediate application of Markov's inequality shows that when $\mathcal{E}_\alpha(\nu)$ is finite, the measure is $\frac{\alpha}{2}$-Hölder, and the Hölder constant can be estimated in terms of the energy. Indeed, for any $x \in M$ and $r > 0$ one has

$$\mathcal{E}_\alpha(\nu) \geq (2r)^{-\alpha} \nu(B_r(x))^2, \tag{2.2.2}$$

as pairwise distances between the points of $B_r(x)$ do not exceed $2r$. Hence, to establish the Hölder property of a measure it would suffice to show that its energy is finite for some $\alpha > 0$.

The key step here is to show that for a sufficiently small α, convolution with μ makes sufficiently high energies decrease. Namely, there exists $\alpha_0 > 0$ and constants $C_0 > 0$, $0 < \lambda < 1$ such that for any $\alpha < \alpha_0$ one has

$$\mathcal{E}_\alpha(\mu * \nu) \leq \lambda \mathcal{E}_\alpha(\nu) + C_0. \tag{2.2.3}$$

To do so, we re-interpret the energy (2.2.1) in terms similar to L_2-norm of a convenient function. Namely, for a given measure ν, consider the function

$$\rho_\alpha[\nu](y) := \int_M \varphi_\alpha(d(x,y))\, d\nu(x), \tag{2.2.4}$$

where $\varphi_\alpha(r) := r^{-\frac{k+\alpha}{2}}$ and k is the dimension of the manifold M. Take the square of the L_2-norm (with respect to the Lebesgue measure on M) of this function,

$$\widetilde{\mathcal{E}}_\alpha(\nu) := \int_M \rho_\alpha[\nu](y)^2\, d\mathrm{Leb}_M(y). \tag{2.2.5}$$

It turns out that there exists constant c_α (that depends only on α and the dimension of the manifold M) such that for any ν

$$\widetilde{\mathcal{E}}_\alpha(\nu) = (c_\alpha + o(1)) \cdot \mathcal{E}_\alpha(\nu), \tag{2.2.6}$$

where $o(1)$ tends to zero as either of energies $\mathcal{E}_\alpha$, $\widetilde{\mathcal{E}}_\alpha$ tend to infinity. Indeed, the right hand side of (2.2.5) can be rewritten as

$$\begin{aligned}\widetilde{\mathcal{E}}_\alpha(\nu) &= \int_M \rho_\alpha^2[\nu](y)\, d\mathrm{Leb}_M(y)\\ &= \iiint_{M\times M\times M} \varphi_\alpha(d(x,y))\varphi_\alpha(d(z,y))\, d\nu(x)d\nu(z)\, d\mathrm{Leb}_M(y)\\ &= \iint_{M\times M} K_\alpha(x,z)d\nu(x)d\nu(z),\end{aligned} \tag{2.2.7}$$

where

$$K_\alpha(x,z) := \int_M \varphi_\alpha(d(x,y))\varphi_\alpha(d(z,y))\, d\mathrm{Leb}_M(y). \tag{2.2.8}$$

If M was replaced by the Euclidean space $\mathbb{R}^k$, the kernel (2.2.8) would have exactly the form $c_\alpha|x-z|^{-\alpha}$, where c_α is a constant. Indeed, an isometric change of coordinates preserves

the integral (2.2.8), and the isometries of $\mathbb{R}^k$ act transitively on pairs of points of the same distance, hence (2.2.8) depends only on $|x-z|$. At the same time, rescaling of $\mathbb{R}^k$ with any factor $\lambda > 0$ multiplies both $\varphi_\alpha(d(x,y))$ and $\varphi_\alpha(d(z,y))$ by $\lambda^{-\frac{k+\alpha}{2}}$, and has the Jacobian of λ^k, hence $K_\alpha(x,z)$ scales as $|x-z|^{-\alpha}$, thus providing the above form.

Hence, on $\mathbb{R}^k$ one would have exact proportionality $c_\alpha \mathcal{E}_\alpha = \widetilde{\mathcal{E}}_\alpha$. Meanwhile, for a general compact manifold M high values of energy $\mathcal{E}$ can come only from points close to each other, and on a small scale a Riemannian manifold is almost Euclidean.

Finally, the key argument goes by contradiction. Namely, for a small α the energy $\mathcal{E}_\alpha$ is almost unchanged by applying a given diffeomorphism f, as the distances are changed at most $\mathfrak{L}(f)$ times, and their α's powers are changed by a factor close to 1. Due to (2.2.6), this implies that for a large value of energy the same applies to $\widetilde{\mathcal{E}}_\alpha$.

However, the latter energy admits an L_2-norm interpretation: the definition (2.2.5) is the squared length $\langle \rho, \rho \rangle_{L_2(M,\mathrm{Leb}_M)}$ of a vector (function) $\rho \in L_2(M)$, linearly associated to the measure ν. In particular, if the L_2-norm of the average of images $f_*\nu$ (where f is distributed w.r.t. μ) does not decrease, it means that (most of) these vectors are essentially (almost) aligned. Considering the non-probability measure

$$\Theta_\alpha[\nu] := \rho_\alpha[\nu]^2(y)\, dLebM(y), \tag{2.2.9}$$

we see that the (almost) alignment of the vectors implies that the corresponding normalized (probability) measure $\theta_\alpha[\nu] := \frac{1}{\widetilde{\mathcal{E}}_\alpha(\nu)}\Theta_\alpha[\nu]$ on M is a measure with (almost) deterministic image. Passing to the limit as α tends to 0, we find a measure with the deterministic image, thus obtaining a contradiction.

The relation (2.2.3) that we have (non-rigorously) established implies that when the energy $\mathcal{E}_\alpha(\nu)$ is large, the passage to the averaged image $\mu * \nu$ decreases it. This already suffices to

establish the *existence* of a Hölder stationary measure. Indeed, take any measure ν_0 of finite α-energy (for instance, the Lebesgue measure on M). Consider the sequence of its averaged images,

$$\nu_n := \mu * \nu_{n-1}, \quad n = 1, 2, \dots;$$

their $\mathcal{E}_\alpha$ energies then stay uniformly bounded. Hence, the same bound for the energy applies to their Cesaro averages $\bar{\nu}_n := \frac{1}{n}\sum_{j=0}^{n-1} \nu_j$. Finally, any accumulation point of this latter sequence is necessarily a stationary measure, is of finite energy, and, hence, Hölder regular. This completes the (informal sketch of the) proof of the existence of a Hölder stationary measure.

Now, for a stationary measure ν the inequality (2.2.3) reads as

$$\mathcal{E}_\alpha(\nu) \le \lambda \mathcal{E}_\alpha(\nu) + C_0,$$

thus implying that either the energy $\mathcal{E}_\alpha(\nu) \le \frac{C_0}{1-\lambda}$, or that this energy is infinite. Unfortunately, the above arguments do not exclude the latter possibility, though suggest that it should not take place: large energies "decrease at infinity". Also, if one takes a Dirac measure as an initial one, an averaging with respect to a distribution that has atomic component will always have some atoms, and hence will never be Hölder regular. So in order to prove Theorems 2.1.3–2.1.8, definitions of energy in (2.2.1) and (2.2.5) are to be modified to exclude the infinity from the list of possibilities.

Such a modification is done by replacing the interaction kernel $d(x, y)^{-\alpha}$ in the definition of the energy $\mathcal{E}_\alpha$ by a bounded function, choosing a distance ε at which the interaction in the energy is cut-off. To do so, we take (see Definition 2.3.5 below)

$$\mathcal{E}_{\alpha,\varepsilon}(\nu) := \iint_M U_{\alpha,\varepsilon}(d(x, y)) d\nu(x) d\nu(y).$$

for some bounded interaction potential $U_{\alpha,\varepsilon}(r)$. Unfortunately, bluntly taking $[\max(r,\varepsilon)]^{-\alpha}$ in the role of $U_{\alpha,\varepsilon}(r)$ makes it harder to work with the L_2-interpretation. Thus we first modify the L_2-definition, applying a cut-off in the construction of the function ρ, and then use it to define the energy.

2.2.1 Structure of the paper

We start by providing rigorous definitions for the cut-off energies $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}$ and $\mathcal{E}_{\alpha,\varepsilon}$ in Sections 2.3.1 and 2.3.2 (see Definitions 2.3.2 and 2.3.5 below). We then state some of their properties in Section 2.3.3. In particular, we show that when the energies are large, they are comparable to each other (see Proposition 2.3.10, which is analogous to (2.2.6) above).

Next, in Section 2.3.4 we introduce measures $\theta_{\alpha,\varepsilon}[\nu]$ associated with the measure ν, and describe the effect of an action of "not too distorting" diffeomorphisms (see Proposition 2.3.21).

We provide the formal proof of Theorems 2.1.3–2.1.8 in Section 2.4. The main intermediate result is Proposition 2.4.1 on exponential decrease of large energies; it is an analog of (2.2.3) and is established in Section 2.4.1. The final step in the proof of Theorems 2.1.3–2.1.8 is made in Section 2.4.2.

To make an exposition more reader-friendly, the detailed proofs of the most technical statements are postponed till Section 2.5. Namely, in Section 2.5.1 we discuss the properties of the functions $\varphi_{\alpha,\varepsilon}(r)$ and $U_{\alpha,\varepsilon}(r)$, and prove Proposition 2.3.6. In Section 2.5.2 we provide the proof of Proposition 2.3.10 that claims that the energies $\mathcal{E}_{\alpha,\varepsilon}(\nu)$ and $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)$ cannot differ too much. The proof of Proposition 2.3.21 is given in Section 2.5.4, Proposition 2.3.16 is proven in Section 2.5.3, and in Section 2.5.5 the proof of Lemma 2.4.3 is given.

Finally, in Section 2.6 we give an example that shows that the finite moment condition (2.1.2) cannot be replaced by the condition $\int \log \mathfrak{L}(f)\, d\mu(f) < \infty$ that would be an analog of the

condition on distribution in the classical Furstenberg Theorem on random matrix products.

2.3 Tools for the proof

In this section, we introduce the ε-cut-off energies $\mathcal{E}_{\alpha,\varepsilon}$ and $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}$, as well as the corresponding measures $\Theta_{\alpha,\varepsilon}(\nu)$ and $\theta_{\alpha,\varepsilon}(\nu)$ associated to a given measure ν. We then study their properties and how these objects are changed under an action of a diffeomorphism.

2.3.1 L_2-type energy $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}$

To implement the strategy described in Section 2.2 in a way applicable to any initial probability measure on the manifold, let us start by modifying the definitions of the energy given by (2.2.1) and (2.2.5). As L_2-point of view is the essential component of the proof, we start by modifying the function used to realise it.

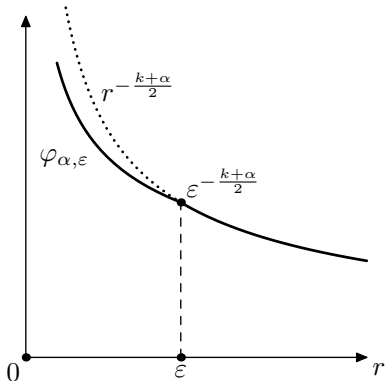

Figure 2.1: Function $\varphi_{\alpha,\varepsilon}(r)$: it coincides with $r^{-\frac{k+\alpha}{2}}$ (graph drawn by a dotted line) for $r > \varepsilon$, and scales as a slower-growing power of r for $r < \varepsilon$

Definition 2.3.1. We define (see Fig. 2.1)

$$\varphi_{\alpha,\varepsilon}(r) := \begin{cases} r^{-\frac{k+\alpha}{2}}, & r \geq \varepsilon \\ \frac{1}{\varepsilon^{\alpha}} \cdot r^{-\frac{k-\alpha}{2}}, & r < \varepsilon. \end{cases} \tag{2.3.1}$$

The following definition should be considered an analog of (2.2.5) and (2.2.9) from Section 2.2:

Definition 2.3.2. For a measure ν on M, we define a function

$$\rho_{\alpha,\varepsilon}[\nu](y) := \int_M \varphi_{\alpha,\varepsilon}(d(x,y))d\nu(x). \tag{2.3.2}$$

and a (finite, not necessarily a probability) measure

$$\Theta_{\alpha,\varepsilon}[\nu] := \rho^2_{\alpha,\varepsilon}[\nu](y)\, d\mathrm{Leb}_M(y).$$

We also define

$$\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) := \Theta_{\alpha,\varepsilon}[\nu](M) = \int_M \rho^2_{\alpha,\varepsilon}[\nu](y)\, d\mathrm{Leb}_M(y). \tag{2.3.3}$$

In the same way as in Section 2.2, we rewrite the definition (2.3.3) as a pairwise interaction with some kernel $K_{\alpha,\varepsilon}$:

Lemma 2.3.3.

$$\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) = \iint_{M\times M} K_{\alpha,\varepsilon}(x,z)d\nu(x)d\nu(z),$$

where

$$K_{\alpha,\varepsilon}(x,z) := \int_M \varphi_{\alpha,\varepsilon}(d(x,y))\varphi_{\alpha,\varepsilon}(d(z,y))\, d\mathrm{Leb}_M(y). \tag{2.3.4}$$

Proof. In the same way as in (2.2.7), it suffices to substitute (2.3.2) that defines $\rho^2_{\alpha,\varepsilon}[\nu](y)$, into (2.3.3), obtaining a triple integral

$$\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) = \iiint_{M\times M\times M} \varphi_{\alpha,\varepsilon}(d(x,y))\varphi_{\alpha,\varepsilon}(d(z,y))\, d\nu(x)d\nu(z)\, d\mathrm{Leb}_M(y), \tag{2.3.5}$$

and then change the order of integration. □

2.3.2 Interaction-type energy $\mathcal{E}_{\alpha,\varepsilon}$

Now let us define the energy $\mathcal{E}_{\alpha,\varepsilon}$. As the reader will see, we actually need both of these notions (as well as the relations between them): while $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}$ is L_2-based and thus adapted for the final steps of the proof, it is the energy $\mathcal{E}_{\alpha,\varepsilon}$ that behaves well under an action of a diffeomorphism.

In the same way as in Section 2.2, we first replace the manifold M with the Euclidean space $\mathbb{R}^k$ in (2.3.4):

Definition 2.3.4. For any two points $x, z \in \mathbb{R}^k$ consider the integral

$$I_{\alpha,\varepsilon}(x,z) := \int_{\mathbb{R}^k} \varphi_{\alpha,\varepsilon}(|y-x|)\varphi_{\alpha,\varepsilon}(|y-z|)\, d\mathrm{Leb}_{\mathbb{R}^k}(y). \tag{2.3.6}$$

For any isometry $f : \mathbb{R}^k \to \mathbb{R}^k$ it is easy to see that

$$I_{\alpha,\varepsilon}(f(x), f(z)) := \int_{\mathbb{R}^k} \varphi_{\alpha,\varepsilon}(|y - f(x)|)\varphi_{\alpha,\varepsilon}(|y - f(z)|)\, d\mathrm{Leb}_{\mathbb{R}^k}(y) =$$
$$= \int_{\mathbb{R}^k} \varphi_{\alpha,\varepsilon}(|f(y) - f(x)|)\varphi_{\alpha,\varepsilon}(|f(y) - f(z)|)\, d\mathrm{Leb}_{\mathbb{R}^k}(f(y)) = I_{\alpha,\varepsilon}(x, z).$$

Since $I_{\alpha,\varepsilon}(x, z)$ depends only on the distance $r = |x-z|$ we can define a function $U_{\alpha,\varepsilon} : \mathbb{R} \to \mathbb{R}$ as follows:

$$U_{\alpha,\varepsilon}(r) := I_{\alpha,\varepsilon}(x, z), \tag{2.3.7}$$

where x and z are any points in $\mathbb{R}^k$, such that $r = |x - z|$. We will show that $U_{\alpha,\varepsilon}(r)$ is finite for every $\varepsilon > 0$, $\alpha > 0$, and $r \geq 0$ in Lemma 2.5.4

Now we are ready to give a definition of energy $\mathcal{E}_{\alpha,\varepsilon}(\nu)$ that is a modified version of (2.2.1):

Definition 2.3.5. We define

$$\mathcal{E}_{\alpha,\varepsilon}(\nu) := \iint_{M\times M} U_{\alpha,\varepsilon}(d(x, z))\, d\nu(x)d\nu(z).$$

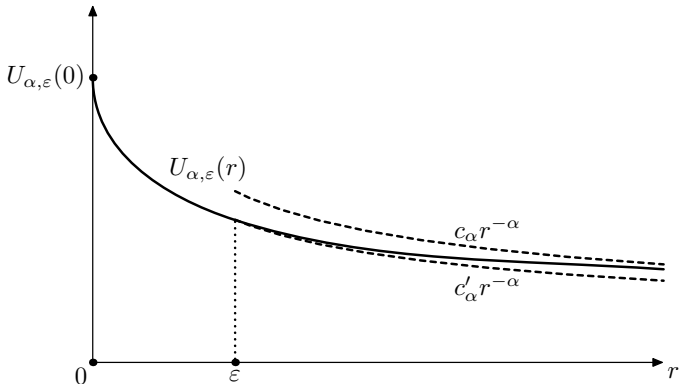

Figure 2.2: The function $U_{\alpha,\varepsilon}(r)$

Note that the function $U_{\alpha,\varepsilon}(r)$ that we defined can indeed be seen as a cut-off version of $r^{-\alpha}$. Namely, it has the following properties (see Fig. 2.2).

Proposition 2.3.6. *For any $\alpha \in (0, \frac{1}{2})$ the following hold:*

1. ***(Finiteness and monotonicity)*** *For any $\varepsilon > 0$ the function $U_{\alpha,\varepsilon}(r)$ is non-increasing and finite everywhere on $[0, +\infty)$; in particular, for any $r > 0$ one has*

$$U_{\alpha,\varepsilon}(r) \leq U_{\alpha,\varepsilon}(0). \tag{2.3.8}$$

2. ***(Outside cut-off: power law)*** *For some constants c_α, c'_α for any $r \geq \varepsilon > 0$ one has*

$$c'_\alpha r^{-\alpha} \leq U_{\alpha,\varepsilon}(r) \leq c_\alpha r^{-\alpha}; \tag{2.3.9}$$

also, for any $r > 0$

$$\frac{U_{\alpha,\varepsilon}(r)}{c_\alpha r^{-\alpha}} \to 1, \qquad \text{as } \varepsilon \to 0. \tag{2.3.10}$$

3. ***(α-slow changing)*** *For any ε and any $r' > r > 0$, one has*

$$\left(\frac{r'}{r}\right)^{-\alpha} \leq \frac{U_{\alpha,\varepsilon}(r')}{U_{\alpha,\varepsilon}(r)} \leq 1 \tag{2.3.11}$$

We postpone the proof of these properties until Section 2.5.1. Specifically, conclusion (1) will be proven in Lemma 2.5.4 and Lemma 2.5.6, conclusion (2) in Corollary 2.5.2 and Lemma 2.5.5, and conclusion (3) in Proposition 2.5.3.

Meanwhile, we will be using Proposition 2.3.6 to prove our main results. We will assume that for any given $\alpha \in (0, \frac{1}{2})$ the constants c_α, c'_α are as in the conclusions of this proposition.

Note that the value of $U_{\alpha,\varepsilon}(0)$, appearing in conclusion (1) of Proposition 2.3.6, can be calculated explicitly:

Lemma 2.3.7. *$U_{\alpha,\varepsilon}(0) = \frac{L}{\alpha}\,\varepsilon^{-\alpha}$, where $L = 2k\,\mathrm{Leb}_{\mathbb{R}^k}(B_1(0))$ is a constant.*

Proof.

$$
\begin{aligned}
U_{\alpha,\varepsilon}(0) &= \int_{\mathbb{R}^k} \varphi_{\alpha,\varepsilon}(|x|)^2 \, d\mathrm{Leb}_{\mathbb{R}^k}(x) \\
&= \int_{|x|<\varepsilon} \left(\frac{|x|^{\frac{\alpha}{2}-\frac{k}{2}}}{\varepsilon^{\alpha}} \right)^2 d\mathrm{Leb}_{\mathbb{R}^k}(x) + \int_{|x|>\varepsilon} \left(|x|^{-\frac{\alpha}{2}-\frac{k}{2}} \right)^2 d\mathrm{Leb}_{\mathbb{R}^k}(x) \\
&= \mathrm{Leb}_{\mathbb{R}^k}(B_1(0)) \left(\int_0^{\varepsilon} \frac{r^{\alpha-k}}{\varepsilon^{2\alpha}} \, kr^{k-1} dr + \int_{\varepsilon}^{\infty} r^{-\alpha-k} \, kr^{k-1} dr \right) \\
&= k \, \mathrm{Leb}_{\mathbb{R}^k}(B_1(0)) \left(\frac{r^{\alpha}}{\alpha\varepsilon^{2\alpha}} \Big|_0^{\varepsilon} - \frac{r^{-\alpha}}{\alpha} \Big|_{\varepsilon}^{\infty} \right) = \frac{L}{\alpha} \, \varepsilon^{-\alpha},
\end{aligned}
$$

where we have used the standard formula for integrals of radial functions,

$$
\int_{\mathbb{R}^k} f(|x|) \, d\mathrm{Leb}_{\mathbb{R}^k}(x) = \mathrm{Leb}_{\mathbb{R}^k}(B_1(0)) \int_0^{\infty} f(r) \, kr^{k-1} dr.
$$

□

Due to the monotonicity conclusion (1) of Proposition 2.3.6 one gets the following uniform upper bound:

Corollary 2.3.8. *For any $\alpha \in (0, \frac{1}{2})$ there exists C''_α such that for any measure ν on M and any $\varepsilon > 0$ one has*

$$
\mathcal{E}_{\alpha,\varepsilon}(\nu) \leq C''_\alpha \varepsilon^{-\alpha}.
$$

Proof. The function $U_{\alpha,\varepsilon}(r)$ is monotonously decreasing, so

$$
\mathcal{E}_{\alpha,\varepsilon}(\nu) = \iint_{M\times M} U_{\alpha,\varepsilon}(d(x,z)) \, d\nu(x) d\nu(z) \leq U_{\alpha,\varepsilon}(0) = C''_\alpha \varepsilon^{-\alpha},
$$

where $C''_\alpha = \frac{L}{\alpha}$. □

Proposition 2.3.6 also allows us to have Hölder type bounds in terms of the energy, the

statement that is an analogue to (2.2.2) above:

Lemma 2.3.9. *For any $\alpha, \varepsilon > 0$ and any probability measure ν on M, one has*

$$\forall y \in M \quad \forall r > \varepsilon \quad \nu(B_r(y)) \leq \sqrt{\frac{\mathcal{E}_{\alpha,\varepsilon}(\nu) 2^\alpha}{c'_\alpha}} \cdot r^{\frac{\alpha}{2}}, \tag{2.3.12}$$

where c'_α is given by (2.3.9) *from Proposition 2.3.6.*

Proof. One has

$$\mathcal{E}_{\alpha,\varepsilon}(\nu) \geq U_{\alpha,\varepsilon}(2r) \cdot \nu(B_r(y))^2 \geq c'_\alpha (2r)^{-\alpha} \cdot \nu(B_r(y))^2,$$

where the first inequality is the Markov lower bound for the integral (for any $x, z \in B_r(y)$ one has $U_{\alpha,\varepsilon}(d(x,z)) \geq U_{\alpha,\varepsilon}(2r)$), and the second is due to (2.3.9). Dividing by $c'_\alpha (2r)^{-\alpha}$ and taking square root, we obtain the desired (2.3.12). □

2.3.3 Properties of the energies $\mathcal{E}_{\alpha,\varepsilon}$ and $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}$

In the same way as for (2.2.6), the two energies, $\mathcal{E}_{\alpha,\varepsilon}$ and $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}$, are comparable. Namely, we have the following statement.

Proposition 2.3.10. *For any $\alpha \in (0, 1/2)$ and any $\delta > 0$ there exists $C > 0$ such that for any $\varepsilon > 0$ and any measure ν on M such that $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) > C$ or $\mathcal{E}_{\alpha,\varepsilon}(\nu) > C$ one has*

$$\frac{\mathcal{E}_{\alpha,\varepsilon}(\nu)}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)} \in (1-\delta, 1+\delta).$$

The idea of the proof is the same as for (2.2.6): large values of energy can come only from points that are close to each other, and locally any manifold looks like a Euclidean space. Again, we postpone the formal proof of this proposition until Section 2.5.2.

It is convenient to re-formulate Proposition 2.3.10 using the following notation:

Definition 2.3.11. We say that two positive numbers, A and A', are (δ, C)-*close*, if

$$A < (1+\delta)A' + C \quad \text{and} \quad A' < (1+\delta)A + C;$$

this can be equivalently rewritten as

$$\frac{1}{1+\delta}A - \frac{1}{1+\delta}C < A' < (1+\delta)A + C.$$

We denote it $A \approx_{(\delta,C)} A'$.

Remark 2.3.12. Proposition 2.3.10 can be equivalently reformulated in the following way, that includes all possible probability measures ν, not necessarily the high energy ones. For any $\alpha \in (0, 1/2)$ and any $\delta > 0$ there exists $C > 0$ such that

$$\forall \varepsilon > 0 \quad \forall \nu \quad \mathcal{E}_{\alpha,\varepsilon}(\nu) \approx_{(\delta,C)} \widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu). \tag{2.3.13}$$

The next proposition and its corollary give a bound on how the energy $\mathcal{E}_{\alpha,\varepsilon}$ can change under an action of a diffeomorphism. Namely, it is changed by a factor that is at most $\mathfrak{L}(f)^\alpha$, where $\mathfrak{L}(f)$ is as in Definition 2.1.2. Thus once α is close to 0, the factor is close to 1.

Proposition 2.3.13. *For any $\alpha \in (0, 1/2)$, any $\varepsilon > 0$ and any $f \in \mathrm{Diff}^1(M)$ one has*

$$\mathfrak{L}(f)^{-\alpha} \leq \frac{\mathcal{E}_{\alpha,\varepsilon}(f_*\nu)}{\mathcal{E}_{\alpha,\varepsilon}(\nu)} \leq \mathfrak{L}(f)^\alpha.$$

Proof. Indeed, by definition of $\mathfrak{L}(f)$ we know that $\mathfrak{L}(f) \geq 1$ and

$$d(f(x), f(z)) \leq \mathfrak{L}(f) d(x, z) \quad \text{for any } x, z \in M.$$

Proposition 2.3.6 thus implies (see (2.3.11)) that

$$\mathfrak{L}(f)^{-\alpha} U_{\alpha,\varepsilon}(d(x,z)) \leq U_{\alpha,\varepsilon}(d(f(x),f(z)))$$

Integrating over $d\nu(x)\, d\nu(z)$, one gets the desired

$$\mathfrak{L}(f)^{-\alpha} \mathcal{E}_{\alpha,\varepsilon}(\nu) \leq \mathcal{E}_{\alpha,\varepsilon}(f_*\nu).$$

The second inequality follows in the same way from

$$\mathfrak{L}(f)^{-1} d(x,z) \leq d(f(x),f(z))$$

□

Corollary 2.3.14. *For any $\delta > 0$ and any $R < \infty$ there exists $\alpha_0 > 0$ such that for any $\alpha \in (0, \alpha_0)$, any $\varepsilon > 0$ and any $f \in \mathrm{Diff}^1(M)$ with $\mathfrak{L}(f) < R$ one has*

$$\frac{\mathcal{E}_{\alpha,\varepsilon}(f_*\nu)}{\mathcal{E}_{\alpha,\varepsilon}(\nu)} \in (1-\delta, 1+\delta).$$

Joining this statement with Proposition 2.3.10, we get the same statement for the energy $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}$:

Corollary 2.3.15. *For any $\delta > 0$ and any $R < \infty$ there exists $\alpha_0 > 0$ such that for any $\alpha \in (0, \alpha_0)$, there exists $C > 0$ such that for any $f \in \mathrm{Diff}^1(M)$ with $\mathfrak{L}(f) < R$, any $\varepsilon > 0$ and any ν such that $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) > C$ or $\mathcal{E}_{\alpha,\varepsilon}(\nu) > C$ one has*

$$\frac{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(f_*\nu)}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)} \in (1-\delta, 1+\delta). \tag{2.3.14}$$

A final concluding remark in this direction is that the statement of Corollary 2.3.15 sur-

vives if one takes an expectation of energy of a random image, provided that the moment condition (2.1.2) is satisfied.

Proposition 2.3.16. *If a probability measure $\mu \in \mathcal{M}$ satisfies assumption* (2.1.2) *then for any $\delta > 0$ there exists $\alpha_0 > 0$ such that for any $\alpha \in (0, \alpha_0)$ the following holds. There exists $C > 0$ such that for any $\varepsilon > 0$ and any measure ν on the manifold M with $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) > C$ one has:*

$$\frac{\mathbb{E}_\mu \left[\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(f_*\nu)\right]}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)} \in (1-\delta, 1+\delta). \tag{2.3.15}$$

For any given individual diffeomorphism f and any measure ν, the quotient $\mathcal{E}_{\alpha,\varepsilon}(f_*\nu)/\mathcal{E}_{\alpha,\varepsilon}(\nu)$ belongs to $[\mathfrak{L}(f)^{-\alpha}, \mathfrak{L}(f)^{\alpha}]$ and hence converges to 1 as α tends to 0 uniformly in ε. It is not difficult to justify that one can bring the expectation under the limit, obtaining the convergence to 1 of the quotient $\mathbb{E}_\mu \mathcal{E}_{\alpha,\varepsilon}(f_*\nu)/\mathcal{E}_{\alpha,\varepsilon}(\nu)$. Conclusion (2.3.15) states the same for the energies $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}$; we provide a formal proof of Proposition 2.3.16 in Section 2.5.3.

2.3.4 Normalizations $\theta_{\alpha,\varepsilon}$, Wasserstein distance

In the same way as in the sketch of the proof, we consider the normalizations $\theta_{\alpha,\varepsilon}$ of measures $\Theta_{\alpha,\varepsilon}$ given by Definition 2.3.2. It is these measures that would turn out to be closer and closer to having a deterministic image assuming that the conclusions of our main results do not hold.

Definition 2.3.17.

$$\theta_{\alpha,\varepsilon}(\nu) = \frac{1}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)}\, \Theta_{\alpha,\varepsilon}[\nu]. \tag{2.3.16}$$

In what follows it will be convenient to use Wasserstein metric in the space of probability

measures on M. Let us recall its definition, as well as definition of the total variation distance between measures:

Definition 2.3.18. Let ν_1, ν_2 be two probability measures on a measure space $(M, \mathcal{B})$. Then the *Wasserstein distance* (or, to be completely precise, *1-Wasserstein distance*) between them is defined as

$$W(\nu_1, \nu_2) = \inf_{\gamma} \iint_{M \times M} d(x, y)\, d\gamma(x, y),$$

where the infimum is taken over all probability measures γ on $(M \times M, \mathcal{B} \times \mathcal{B})$ with the marginals (projections on the x and y coordinates) $P_x(\gamma) = \nu_1$ and $P_y(\gamma) = \nu_2$.

Definition 2.3.19. Let ν_1, ν_2 be two probability measures on a measure space $(M, \mathcal{B})$. Then the *total variation* distance between them is

$$\mathrm{TV}(\nu_1, \nu_2) = \sup_{B \in \mathcal{B}} |\nu_1(B) - \nu_2(B)|.$$

Also, we will need the following statement that can be found, for example, in [40, Theorem 6.15]: the Wasserstein metric is bounded from above by the diameter times the total variation distance.

Lemma 2.3.20. *For any two probability measures ν_1, ν_2 on a manifold M we have*

$$W(\nu_1, \nu_2) \leq \mathrm{diam}(M) \cdot \mathrm{TV}(\nu_1, \nu_2).$$

We are going to use the following statement. Assuming that α is sufficiently small, the energy of a measure ν is sufficiently high, and a diffeomorphism f is "not too distorting", not only are the energies of ν and of $f_*\nu$ close to each other, but also the measure $\theta_{\alpha,\varepsilon}[f_*\nu]$ is close to the push-forward $f_*\theta_{\alpha,\varepsilon}[\nu]$.

Proposition 2.3.21. *For any $\delta > 0$ and any $R > 0$ there exists $\alpha_0 > 0$ such that for any $\alpha \in (0, \alpha_0)$ there exists $C > 0$ such that for any $f \in \mathrm{Diff}^1(M)$ with $\mathfrak{L}(f) < R$, any $\varepsilon > 0$, and any ν such that $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) > C$ or $\mathcal{E}_{\alpha,\varepsilon}(\nu) > C$ one has*

$$W(f_*\theta_{\alpha,\varepsilon}(\nu), \theta_{\alpha,\varepsilon}(f_*\nu)) < \delta. \tag{2.3.17}$$

The idea of the proof of this proposition is close to the one for Proposition 2.3.10: large energy means that most of it comes from points close to each other. Moreover, considering the energy as a triple integral (2.3.5), one sees that most of it comes out from all the three points being close to each other. Now, when a diffeomorphism f is applied, these local contributions are changed *roughly* in the same way as the kernel $U_{\alpha,\varepsilon}(d(x,z))$. Finally, the latter almost does not change if $\mathfrak{L}(f)^\alpha$ is close to 1, thus implying the desired closeness of two measures.

Once again, we postpone the formal proof of this proposition until Section 2.5.4.

2.4 Proofs of main results

In this section we prove Theorems 2.1.3, 2.1.6, and 2.1.8, except for the technically complicated parts, that we moved to Section 2.5 (specifically, the proofs of Propositions 2.3.6, 2.3.10, 2.3.16, 2.3.21, together with the proof of Lemma 2.4.3).

2.4.1 Exponential decrease of large energies

The key step of the proof of our main results is the following proposition:

Proposition 2.4.1. *If the measure μ satisfies the **finite moment condition** and the **no***

deterministic images *condition, then there exist* $\alpha > 0, C < \infty, \lambda < 1$, *such that for any* $\varepsilon > 0$ *if* $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) > C$ *then*

$$\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\mu * \nu) < \lambda \widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu). \tag{2.4.1}$$

Proof. Fix some small number $\delta > 0$, take $\lambda = 1 - \delta$, and assume the contrary: for any $\alpha > 0$ and $C < \infty$ there exists an $\varepsilon > 0$ and a probability measure ν with energy $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) > C$, such that

$$\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\mu * \nu) \geq (1 - \delta)\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu). \tag{2.4.2}$$

Recall the following standard statement:

Lemma 2.4.2. *Assume that in some Hilbert space* $\mathcal{H}$ *a probability measure is given; in other words, one is given a random variable* v *taking values in this space. Assume that this measure has finite second moment, and let* $\overline{v} := \mathbb{E}v$ *be its expectation. Then*

$$\mathbb{E}\langle v - \overline{v}, v - \overline{v}\rangle = \mathbb{E}\langle v, v\rangle - \langle \overline{v}, \overline{v}\rangle \tag{2.4.3}$$

In dimension one, it is exactly the equivalence between the two definitions of variance of a random variable; and in general, it is straightforward to check (2.4.3): it suffices to substitute $v = (v - \overline{v}) + \overline{v}$ to the expectation in the right hand side and use the linearity of the scalar product.

Now, let us apply this statement to $L_2(M, \mathrm{Leb}_M)$. Namely, for a given measure ν its averaged image $\mu * \nu$ is the expectation of $f_*\nu$, where the diffeomorphism f of M is taken randomly w.r.t. the measure μ. Hence, the same applies for the density $\rho_{\alpha,\varepsilon}[\nu]$, given by (2.3.2) (see

Definition 2.3.2), since it depends linearly on ν:

$$\rho_{\alpha,\varepsilon}[\mu * \nu] = \mathbb{E}_\mu \, \rho_{\alpha,\varepsilon}[f_*\nu]. \tag{2.4.4}$$

Substituting this into (2.4.3) (with $v = \rho_{\alpha,\varepsilon}[f_*\nu]$), and taking into account the definition (2.3.3), we get

$$\mathbb{E}_\mu \left[\int_M \left(\rho_{\alpha,\varepsilon}[f_*\nu](y) - \rho_{\alpha,\varepsilon}[\mu * \nu](y)\right)^2 d\mathrm{Leb}_M(y) \right] \\ = \mathbb{E}_\mu \left[\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(f_*\nu) \right] - \widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\mu * \nu); \tag{2.4.5}$$

in particular,

$$\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\mu * \nu) \le \mathbb{E}_\mu \left[\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(f_*\nu) \right]. \tag{2.4.6}$$

By Proposition 2.3.16 we can find sufficiently small $\alpha > 0$ and sufficiently large $C < \infty$, such that for any $\varepsilon > 0$ and any measure ν on M with $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) > C$ the inequality (2.3.15) holds, and thus (taking only the upper bound)

$$\mathbb{E}_\mu \left[\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(f_*\nu) \right] \le (1+\delta) \widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu). \tag{2.4.7}$$

Denote by ν a probability measure with $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) > C$, for which the inequality (2.4.2) holds. Substituting (2.4.7) and (2.4.2) in the right hand side of (2.4.5), we get

$$\mathbb{E}_\mu \left[\int_M \left(\rho_{\alpha,\varepsilon}(f_*\nu)(y) - \rho_{\alpha,\varepsilon}(\mu * \nu)(y)\right)^2 d\mathrm{Leb}_M(y) \right] < 2\delta \, \widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) \tag{2.4.8}$$

and thus

$$\mathbb{E}_\mu \left[\int_M \frac{\left(\rho_{\alpha,\varepsilon}(f_*\nu)(y) - \rho_{\alpha,\varepsilon}(\mu * \nu)(y)\right)^2}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)} d\mathrm{Leb}_M(y) \right] < 2\delta. \tag{2.4.9}$$

The final step is to show that measures

$$\theta_{\alpha,\varepsilon}(f_*\nu) = \frac{\rho_{\alpha,\varepsilon}(f_*\nu)(y)^2}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(f_*\nu)} d\mathrm{Leb}_M(y) \quad \text{and} \quad \theta_{\alpha,\varepsilon}(\mu * \nu) = \frac{\rho_{\alpha,\varepsilon}(\mu * \nu)(y)^2}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\mu * \nu)} d\mathrm{Leb}_M(y)$$

are close with high probability. And indeed, we have the following statement, estimating (even) the total variations distance:

Lemma 2.4.3. *Under the assumptions above*

$$\mathbb{P}_\mu \left[\mathrm{TV}(\theta_{\alpha,\varepsilon}(f_*\nu), \theta_{\alpha,\varepsilon}(\mu * \nu)) > 2\sqrt[4]{\delta} \right] < 4\sqrt{\delta}, \tag{2.4.10}$$

and hence

$$\mathbb{P}_\mu \left[W(\theta_{\alpha,\varepsilon}(f_*\nu), \theta_{\alpha,\varepsilon}(\mu * \nu)) > 2\,\mathrm{diam}(M)\sqrt[4]{\delta} \right] < 4\sqrt{\delta}. \tag{2.4.11}$$

Proof of Lemma 2.4.3 is the most technical part of the proof of Proposition 2.4.1, and we postpone it till Section 2.5.5.

We are now ready to conclude the proof of Proposition 2.4.1. Namely, the conclusion (2.4.11) states that with high probability the measure $\theta_{\alpha,\varepsilon}(f_*\nu)$ is close to the deterministic one, $\theta_{\alpha,\varepsilon}(\mu * \nu)$. The next step is to show that with high probability the first measure is close to the f_*-image of a given measure, $\theta_{\alpha,\varepsilon}(\nu)$.

To do so, choose $R < \infty$, such that

$$\mathbb{P}_\mu \left[\mathfrak{L}(f) < R \right] > 1 - \sqrt[4]{\delta}$$

By Proposition 2.3.21 we can choose C big enough, so that for any $f \in \mathrm{Diff}^1(M)$, such that $\mathfrak{L}(f) < R$ inequality (2.3.17) holds, and using triangle inequality we arrive to

$$\mathbb{P}_\mu \left[W(f_* \theta_{\alpha,\varepsilon}(\nu), \theta_{\alpha,\varepsilon}(\mu * \nu)) > 2 \operatorname{diam}(M) \sqrt[4]{\delta} + \delta \right] < 5 \sqrt[4]{\delta}. \tag{2.4.12}$$

Now consider the sequence $\delta_n = \frac{1}{n}$ and denote by ν_n the corresponding measure and corresponding parameters by α_n, ε_n, for which (2.4.2) holds. Then the measures $\theta_{\alpha_n,\varepsilon_n}(\nu_n)$ have a weakly convergent subsequence. Futhermore, we can take another subsequence, such that the measures $\theta_{\alpha_{n_k},\varepsilon_{n_k}}(\mu * \nu_{n_k})$ also converge. It remains to notice that due to inequality (2.4.12) the limit measure

$$m = \lim_{k \to \infty} \theta_{\alpha_{n_k},\varepsilon_{n_k}}(\nu_{n_k})$$

has an almost surely constant image under the action of $f \in \mathrm{supp}(\mu)$, that is equal to

$$\tilde{m} = \lim_{k \to \infty} \theta_{\alpha_{n_k},\varepsilon_{n_k}}(\mu * \nu_{n_k}),$$

which contradicts our assumption. □

Finally, Proposition 2.4.1 can be generalized by the following statement.

Proposition 2.4.4. *Let $\mathcal{K} \subset \mathcal{M}$ be a compact subset that satisfies the following conditions:*

- ***(uniform finite moment condition)*** *There exists $C_0, \gamma > 0$ such that for any $\mu \in \mathcal{K}$*

one has

$$\int \mathfrak{L}(f)^\gamma \, d\mu(f) < C_0;$$

- ***(no deterministic images)*** *For any $\mu \in \mathcal{K}$ there are no probability measures ν, ν' on M such that $f_*\nu = \nu'$ for μ-almost all $f \in \mathrm{Diff}^1(M)$.*

Then the constants α, λ and C in Proposition 2.4.1 can be chosen uniformly in $\mu \in \mathcal{K}$.

Proof. Indeed, assume the contrary. It is easy to see that in the statements in Section 2.3 the constants can be chosen uniformly for all $\mu \in \mathcal{K}$. Now, repeating the proof of Proposition 2.4.1, we see that the only part to be modified is passage to the limit. Again taking $\delta_n = \frac{1}{n}$ one finds α_n, ε_n and a measure $\mu_n \in \mathcal{K}$ such that (2.4.2) holds. Passing to a subsequence (n_k), one can ensure the existence of three limits:

$$m = \lim_{k\to\infty} \theta_{\alpha_{n_k},\varepsilon_{n_k}}(\nu_{n_k})$$

$$\tilde{m} = \lim_{k\to\infty} \theta_{\alpha_{n_k},\varepsilon_{n_k}}(\mu_{n_k} * \nu_{n_k}),$$

$$\tilde{\mu} = \lim_{k\to\infty} \mu_{n_k}$$

In turn, this implies that for $\tilde{\mu}$-a.e. diffeomorphism f one has $f_*m = \tilde{m}$, thus obtaining a contradiction with the "no deterministic images" assumption at $\tilde{\mu} \in \mathcal{K}$. □

Finally, we note that the conclusion of Proposition 2.4.1 can be modified to include all possible measures ν (without assuming the energy higher than C). To do so, we need to adjust the upper bound (2.4.1) in order to include low-energy measures ν:

Lemma 2.4.5. *Assume that the measure μ satisfies the finite moment condition (2.1.2) with some γ, and that $\alpha < \gamma$ and C are given. Then there exists C' such that for any measure ν*

on M with $\mathcal{E}_{\alpha,\varepsilon}(\nu) \leq C$ one has

$$\mathcal{E}_{\alpha,\varepsilon}(\mu * \nu) \leq C'. \tag{2.4.13}$$

Corollary 2.4.6. *In the assumptions of Proposition 2.4.1 one can conclude that there exist $\alpha > 0, \widetilde{C} < \infty, \lambda < 1$, such that for any $\varepsilon > 0$ and any measure ν on M*

$$\mathcal{E}_{\alpha,\varepsilon}(\mu * \nu) < \max(\lambda \mathcal{E}_{\alpha,\varepsilon}(\nu), \widetilde{C}) \tag{2.4.14}$$

Proof of Lemma 2.4.5. Note first that due to Proposition 2.3.10, for any α and C_1 there exists C_2 such that for any ε and any measure ν one has

$$\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) < C_1 \Rightarrow \mathcal{E}_{\alpha,\varepsilon}(\nu) < C_2$$

and vice versa,

$$\mathcal{E}_{\alpha,\varepsilon}(\nu) < C_1 \Rightarrow \widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) < C_2. \tag{2.4.15}$$

In other words, having an upper bound on one of the energies $\mathcal{E}_{\alpha,\varepsilon}(\nu)$, $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)$ implies a bound for the other one.

Next, the finite moment condition implies the finiteness of the expectation

$$C_I := \mathbb{E}_\mu \mathfrak{L}(f)^\alpha < \infty. \tag{2.4.16}$$

Now, (2.4.5) implies that for any measure ν we have

$$\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\mu * \nu) \leq \mathbb{E}_\mu \widetilde{\mathcal{E}}_{\alpha,\varepsilon}(f_* \nu);$$

At the same time Proposition 2.3.10 for $\delta = \frac{1}{2}$ implies that for some constant C_1 one has for

any measure ν'

$$\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu') \leq \max(2\mathcal{E}_{\alpha,\varepsilon}(\nu'), C_1).$$

Applying this for $\nu' = f_*\nu$ and joining it with Proposition 2.3.13, we get

$$\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(f_*\nu) \leq \max(2\mathcal{E}_{\alpha,\varepsilon}(f_*\nu), C_1) \leq 2\mathfrak{L}(f)^\alpha \mathcal{E}_{\alpha,\varepsilon}(\nu) + C_1.$$

Taking the expectation w.r.t. μ and using (2.4.16), we finally get a uniform bound

$$\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\mu * \nu) \leq 2C_I C + C_1 =: C'.$$

□

Proof of Corollary 2.4.6. In the proof of Proposition 2.4.1 we can take α arbitrarily small, so we can assume that $\alpha < \gamma$. Applying Proposition 2.3.10 to both sides of (2.4.1), where we take δ sufficiently small so that $\frac{1-\delta}{1+\delta} > \lambda$, allows to conclude that there exists some constant C_3 such that for all ε and ν

$$\mathcal{E}_{\alpha,\varepsilon}(\mu * \nu) < \lambda' \mathcal{E}_{\alpha,\varepsilon}(\nu) \quad \text{if } \mathcal{E}_{\alpha,\varepsilon}(\nu) > C_3,$$

where $\lambda' = \frac{1+\delta}{1-\delta}\lambda < 1$. Meanwhile, Lemma 2.4.5 implies that there exists some C_4 such that

$$\mathcal{E}_{\alpha,\varepsilon}(\mu * \nu) < C_4 \quad \text{if } \mathcal{E}_{\alpha,\varepsilon}(\nu) \leq C_3.$$

Taking $\widetilde{C} := C_4$, we get the desired (2.4.14). □

2.4.2 Hölder bounds: proofs of Theorems 2.1.3–2.1.8

Now we are ready to complete the proof of Theorem 2.1.6 (and thus of Theorem 2.1.3).

Proof of Theorem 2.1.6. Let $\alpha, \widetilde{C}$ be as in Corollary 2.4.6: for any ε and for any measure ν one has

$$\mathcal{E}_{\alpha,\varepsilon}(\mu * \nu) < \max(\lambda \mathcal{E}_{\alpha,\varepsilon}(\nu), \widetilde{C});$$

applying this n times, we get

$$\mathcal{E}_{\alpha,\varepsilon}(\mu^n * \nu) < \max(\lambda^n \mathcal{E}_{\alpha,\varepsilon}(\nu), \widetilde{C}).$$

Recall that Lemma 2.3.7 gives a uniform upper bound $\mathcal{E}_{\alpha,\varepsilon}(\nu) \leq C''_\alpha \varepsilon^{-\alpha}$, where $C''_\alpha = \frac{L}{\alpha}$; choose $\varepsilon := \lambda^{\frac{n}{\alpha}}$, then we have

$$\mathcal{E}_{\alpha,\varepsilon}(\mu^n * \nu) < \max(C''_\alpha, \widetilde{C}).$$

Now, applying Lemma 2.3.9 for any $r > \varepsilon = \kappa^n$, where $\kappa := \lambda^{\frac{1}{\alpha}}$, we deduce that

$$(\mu^n * \nu)(B_r(x)) < \sqrt{\frac{2^\alpha \max(C''_\alpha, \widetilde{C})}{c'_\alpha}} \cdot r^{\frac{\alpha}{2}},$$

which, after renaming $\alpha := \alpha/2$ and choosing appropriate C completes the proof of Theorem 2.1.6.

□

Arguing in the same way as in the proof of Theorem 2.1.6, Theorem 2.1.8 can be deduced from Proposition 2.4.4.

Remark 2.4.7. The same proof works for the setting mentioned in Remark 2.1.9. One has to consider an embedding $M \hookrightarrow \tilde{M}$ into a closed manifold $\tilde{M}$, to define energies $\mathcal{E}_{\alpha,\varepsilon}$ and $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}$ with respect to the Lebesgue measure on $\tilde{M}$ and to repeat the original proof verbatim.

2.5 Proofs of used properties

Here we provide the proofs of the technical statements that were formulated and used above, as described in Section 2.2.1.

2.5.1 Properties of $\varphi_{\alpha,\varepsilon}(r)$ and $U_{\alpha,\varepsilon}(r)$

Let us prove Proposition 2.3.6. We start with some scaling properties of $\varphi_{\alpha,\varepsilon}$ and of $U_{\alpha,\varepsilon}$:

Lemma 2.5.1. *For any $\lambda > 0$ and any $\alpha, \varepsilon > 0$ we have*

$$\forall r > 0 \quad \varphi_{\alpha,\lambda\varepsilon}(\lambda r) = \lambda^{-\frac{k}{2}-\frac{\alpha}{2}} \varphi_{\alpha,\varepsilon}(r) \tag{2.5.1}$$

and

$$\forall r > 0 \quad U_{\alpha,\lambda\varepsilon}(\lambda r) = \lambda^{-\alpha} U_{\alpha,\varepsilon}(r). \tag{2.5.2}$$

In particular, one has

$$U_{\alpha,\varepsilon}(r) = r^{-\alpha} U_{\alpha,\frac{\varepsilon}{r}}(1) = \varepsilon^{-\alpha} U_{\alpha,1}(r/\varepsilon). \tag{2.5.3}$$

Proof. Relation (2.5.1) follows directly from (2.3.1) in the definition of $\varphi_{\alpha,\varepsilon}$. For (2.5.2), it suffices to make a change of variable $y = \lambda y'$ in the integral (2.3.6). Finally, (2.5.3) immediately follows from (2.5.2) □

Corollary 2.5.2. *For any $\alpha > 0$ and any $r > \varepsilon > 0$ one has*

$$U_{\alpha,\varepsilon}(r) \geq c'_\alpha r^{-\alpha},$$

where $c'_\alpha := U_{\alpha,1}(1)$.

Proof. As a function of ε, the function $\varphi_{\alpha,\varepsilon}(r)$ and hence $U_{\alpha,\varepsilon}(r)$ is decreasing. The conclusion thus directly follows from (2.5.3). □

Proposition 2.5.3.

$$\forall \alpha, \varepsilon > 0 \quad \forall r' > r > 0 \quad (r'/r)^{-\frac{k}{2}-\frac{\alpha}{2}} \leq \frac{\varphi_{\alpha,\varepsilon}(r')}{\varphi_{\alpha,\varepsilon}(r)} \leq (r'/r)^{-\frac{k}{2}+\frac{\alpha}{2}}. \tag{2.5.4}$$

$$\forall \alpha, \varepsilon > 0 \quad \forall r' > r > 0 \quad (r'/r)^{-\alpha} \leq \frac{U_{\alpha,\varepsilon}(r')}{U_{\alpha,\varepsilon}(r)} \leq (r'/r)^{\alpha}. \tag{2.5.5}$$

Proof. Note first that, on each of the intervals $(0, \varepsilon]$ and $[\varepsilon, \infty)$ the function $\varphi_{\alpha,\varepsilon}$ takes the form $\text{const} \cdot r^{-\frac{k}{2}} r^{\pm\frac{\alpha}{2}}$, thus implying the first statement on each of these intervals. Now, to handle the case $r' > \varepsilon > r$, it suffices to multiply the obtained estimates for (r, ε) and for (ε, r').

To establish the second conclusion, set $\lambda := \frac{r}{r'}$, $\lambda < 1$. Take two points x', z' at the distance r' from each other, then the points $x = \lambda x'$, $z = \lambda z'$ are at distance r. Let us make a change of variable $y = \lambda y'$ in (2.3.6) for $I_{\alpha,\varepsilon}(x, z)$:

$$\begin{aligned} U_{\alpha,\varepsilon}(r) = I_{\alpha,\varepsilon}(x,z) &= \int_{\mathbb{R}^k} \varphi_{\alpha,\varepsilon}(|y-x|)\varphi_{\alpha,\varepsilon}(|y-z|)\, d\mathrm{Leb}_{\mathbb{R}^k}(y) \\ &= \int_{\mathbb{R}^k} \varphi_{\alpha,\varepsilon}(|\lambda y' - \lambda x'|)\varphi_{\alpha,\varepsilon}(|\lambda y' - \lambda z'|)\, d\mathrm{Leb}_{\mathbb{R}^k}(\lambda y') \\ &= \lambda^k \int_{\mathbb{R}^k} \varphi_{\alpha,\varepsilon}(\lambda|y' - x'|)\varphi_{\alpha,\varepsilon}(\lambda|y' - z'|)\, d\mathrm{Leb}_{\mathbb{R}^k}(y') \end{aligned}$$

Now, due to (2.5.4),

$$\lambda^{-(k-\alpha)} \le \frac{\varphi_{\alpha,\varepsilon}(\lambda|y'-x'|)\varphi_{\alpha,\varepsilon}(\lambda|y'-z'|)}{\varphi_{\alpha,\varepsilon}(|y'-x'|)\varphi_{\alpha,\varepsilon}(|y'-z'|)} \le \lambda^{-(k+\alpha)}$$

Hence, the ratio of $U_{\alpha,\varepsilon}(r)$ and

$$\lambda^k \int_{\mathbb{R}^k} \varphi_{\alpha,\varepsilon}(|y'-x'|)\varphi_{\alpha,\varepsilon}(|y'-z'|)\, d\mathrm{Leb}_{\mathbb{R}^k}(y') = \lambda^k U_{\alpha,\varepsilon}(r')$$

must be between $\lambda^{-(k-\alpha)}$ and $\lambda^{-(k+\alpha)}$. As λ^k cancels out, we get the desired

$$\lambda^\alpha U_{\alpha,\varepsilon}(r') \le U_{\alpha,\varepsilon}(r) \le \lambda^{-\alpha} U_{\alpha,\varepsilon}(r').$$

□

Let us now show that the values of $U_{\alpha,\varepsilon}(r)$ are always finite.

Lemma 2.5.4. *For any $\alpha \in (0, \frac{1}{2})$, $\varepsilon > 0$, and any $r \ge 0$ the value $U_{\alpha,\varepsilon}(r)$ is finite, and*

$$\forall r \ge 0 \quad U_{\alpha,\varepsilon}(r) \le \langle \varphi_{\alpha,\varepsilon}, \varphi_{\alpha,\varepsilon} \rangle_{L_2(\mathbb{R}^k)} = U_{\alpha,\varepsilon}(0). \tag{2.5.6}$$

Proof. Note that the $\varphi_{\alpha,\varepsilon} \in L_2(\mathbb{R}^k)$. Indeed, the inequality $2 \cdot \frac{k-\alpha}{2} < k$ ensures squared integrability at the origin, and as $2 \cdot \frac{k+\alpha}{2} > k$, the integral converges at infinity, too.

Now, $U_{\alpha,\varepsilon}(r)$ is a scalar product between two r-shifted copies of $\varphi_{\alpha,\varepsilon}$. Hence, it is finite for all $r \ge 0$, and we get the desired (2.3.8). □

The next lemma explains that the interaction potential $U_{\alpha,\varepsilon}(r)$ can be considered as a cut-off of $r^{-\alpha}$ (or, more precisely, of $c_\alpha r^{-\alpha}$):

Lemma 2.5.5. *For any $\alpha \in (0, \frac{1}{2})$ the function $U_{\alpha,\varepsilon}(r)$ is decreasing in ε, and one has*

$$\forall r > 0 \quad \lim_{\varepsilon \to 0} U_{\alpha,\varepsilon}(r) = c_\alpha r^{-\alpha} \tag{2.5.7}$$

where

$$c_\alpha := \int_{\mathbb{R}^k} |y|^{-\frac{k+\alpha}{2}} |y - (1, 0, \dots, 0)|^{-\frac{k+\alpha}{2}} \, d\mathrm{Leb}_{\mathbb{R}^k}(y)$$

is a constant depending only on α.

Proof. The monotonicity part immediately comes from the monotonicity of $\varphi_{\alpha,\varepsilon}$ in ε, as we have already seen in the proof of Corollary 2.5.2. As the functions under the integral in (2.3.6) are positive and monotonous in ε, one can pass to the limit as $\varepsilon \to 0$, obtaining the convolution square of $r^{-\frac{k+\alpha}{2}}$ in $\mathbb{R}^k$. The latter equals $c_\alpha r^{-\alpha}$ due to the scaling reasons, and the constant c_α comes from substituting $r = 1$. □

Lemma 2.5.6. *For any $\alpha, \varepsilon > 0$ function $U_{\alpha,\varepsilon}(r)$ is non-increasing in r.*

Proof. Actually, the convolution of any positive, spherically symmetric, radially non-increasing and square integrable functions $\psi_1(r)$, $\psi_2(r)$ in $\mathbb{R}^k$ is again spherically symmetric radially non-increasing function. Indeed, any such function can be approximated from below by a linear combination of indicator functions $\mathbb{1}_{B_R(0)}(x) = \mathbb{1}_{|x|<R}$ of balls centred at the origin, taken with positive coefficients. It now suffices to establish the statement for such indicator functions, as one can then pass to the (monotonous) limit.

Therefore it is enough to show that the volume of the intersection of two such balls $B_{R_1}(0)$ and $B_{R_2}(x)$ is a non-increasing function of $|x|$ (see Fig. 2.3). But that is a partial case of the following simple statement:

Lemma 2.5.7. *Suppose $A \subset \mathbb{R}^k$ is convex, $B \subset A$ is measurable, and $\mathrm{Leb}_{\mathbb{R}^k}(B) < \infty$. Then*

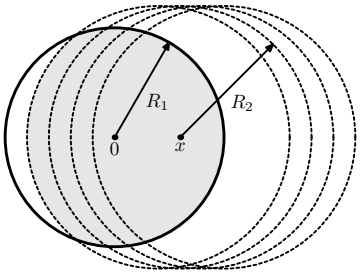

Figure 2.3: Intersections of $B_{R_1}(0)$ with $B_{R_2}(x)$

for any non-zero vector $\bar{v} \in \mathbb{R}^k$ *the measure* $\mathrm{Leb}_{\mathbb{R}^k}(A \cap (B + t\bar{v}))$ *as a function of* $t \geq 0$ *is non-increasing.*

Indeed, monotonicity in Lemma 2.5.7 follows from the fact that due to convexity of A, for any point $b \in B$, as soon as $b + t_1\bar{v} \notin A$, we must have $b + t\bar{v} \notin A$ for all $t > t_1$. □

Proof of Proposition 2.3.6. The finiteness part of conclusion (1) is proven in Lemma 2.5.4; at the same time, Lemma 2.5.6 ensures the monotonicity in r of $U_{\alpha,\varepsilon}(r)$, thus completing the proof of conclusion (1).

Next, the lower bound in (2.3.9) in conclusion (2) is established in Corollary 2.5.2. The limit (2.5.7) from the conclusion of Lemma 2.5.5 is exactly the limit part (2.3.10) of conclusion (2); due to Lemma 2.5.5 the function $U_{\alpha,\varepsilon}(r)$ is decreasing in ε, hence for any $\varepsilon > 0$ one has

$$U_{\alpha,\varepsilon}(r) \leq \lim_{\varepsilon' \to 0} U_{\alpha,\varepsilon'}(r) = c_\alpha r^{-\alpha}$$

thus establishing the upper bound part of (2.3.9), and completing the proof of conclusion (2).

Finally, the lower bound part of (2.3.11) in conclusion (3) follows from Proposition 2.5.3, while the upper bound is implied by the inequality $U_{\alpha,\varepsilon}(r') \leq U_{\alpha,\varepsilon}(r)$. The latter follows from the monotonicity in r of the function $U_{\alpha,\varepsilon}(r)$, established in Lemma 2.5.6. □

2.5.2 Comparing $\mathcal{E}_{\alpha,\varepsilon}(\nu)$ and $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)$: Proof of Proposition 2.3.10

Let us recall that in (2.2.8) we defined

$$K_\alpha(x,z) = \int_M \varphi_\alpha(d(x,y))\varphi_\alpha(d(z,y))\, d\mathrm{Leb}_M(y).$$

We will need the following statement:

Lemma 2.5.8. *For any $\alpha > 0$, the interaction potentials $K_{\alpha,\varepsilon}(x,z)$ and $U_{\alpha,\varepsilon}(d(x,z))$ are comparable in the following sense:*

$$\forall \delta > 0\, \exists C: \quad \forall \varepsilon > 0,\, \forall x, z \in M$$

$$K_{\alpha,\varepsilon}(x,z) \approx_{(\delta,C)} U_{\alpha,\varepsilon}(d(x,z)),\ \textit{i.e.}$$

$$\frac{1}{1+\delta}(U_{\alpha,\varepsilon}(d(x,z)) - C) < K_{\alpha,\varepsilon}(x,z) < (1+\delta)U_{\alpha,\varepsilon}(d(x,z)) + C. \quad (2.5.8)$$

Postponing its proof for the moment, note that it implies Proposition 2.3.10.

Proof of Proposition 2.3.10. It suffices to integrate (2.5.8) w.r.t. $\nu\times\nu\,(x,z)$. As the constant C does not depend on these points nor on the measure ν, one gets

$$\frac{1}{1+\delta}(\mathcal{E}(\nu) - C) < \widetilde{\mathcal{E}}(\nu) < (1+\delta)\mathcal{E}(\nu) + C$$

with the same constant C. As it was noticed in Remark 2.3.12, this is an equivalent form of Proposition 2.3.10. □

Proof of Lemma 2.5.8. Let $\alpha > 0$ be given. Take any $r_0 > 0$ and divide the integral (2.3.4)

defining $K_{\alpha,\varepsilon}(x,z)$ into two parts, depending on whether the distance $d(x,y)$ exceeds r_0:

$$K_{\alpha,\varepsilon}(x,z) = \int_{B_{r_0}(x)} \varphi_{\alpha,\varepsilon}(d(x,y))\varphi_{\alpha,\varepsilon}(d(z,y))\, d\mathrm{Leb}_M(y) + \\ + \int_{M\setminus B_{r_0}(x)} \varphi_{\alpha,\varepsilon}(d(x,y))\varphi_{\alpha,\varepsilon}(d(z,y))\, d\mathrm{Leb}_M(y).$$

Denote the first and the second summands as $K^{(r_0)}_{\alpha,\varepsilon}(x,z)$ and $\mathbf{K}^{(r_0)}_{\alpha,\varepsilon}(x,z)$.

Note that the second summand is uniformly bounded: indeed, the factor $\varphi_{\alpha,\varepsilon}(d(x,y))$ does not exceed a constant $\varphi_{\alpha,\varepsilon}(r_0) \le \varphi_\alpha(r_0)$, while the second factor $\varphi_{\alpha,\varepsilon}(d(y,z))$ is a function with the integral on M that is bounded uniformly in $z \in M$. The latter uniform bound can be seen by again decomposing the integral in two:

$$\int_M \varphi_{\alpha,\varepsilon}(d(y,z))\, d\mathrm{Leb}_M(y) = \int_{B_{r_0}(z)} \varphi_{\alpha,\varepsilon}(d(y,z))\, d\mathrm{Leb}_M(y) + \\ + \int_{M\setminus B_{r_0}(z)} \varphi_{\alpha,\varepsilon}(d(y,z))\, d\mathrm{Leb}_M(y). \quad (2.5.9)$$

The second summand in (2.5.9) does not exceed $\mathrm{vol}(M)\cdot\varphi_\alpha(r_0)$, as

$$\varphi_\alpha(r) = \lim_{\varepsilon\to+0} \varphi_{\alpha,\varepsilon}(r) = r^{-\frac{k+\alpha}{2}}$$

is an upper bound for $\varphi_{\alpha,\varepsilon}(r)$ for all $\varepsilon > 0$. The first one can be estimated uniformly in $z \in M$ by passage to the geodesic coordinates centred at z, comparing it to the same integral in a ball in $\mathbb{R}^k$: the Jacobian of the change of variables is (for r_0 smaller than the injectivity radius in M) uniformly bounded, and the integral of $r^{-\frac{k+\alpha}{2}}$ on $B_{r_0}(0) \subset \mathbb{R}^k$ converges.

We thus have

$$\mathbf{K}^{(r_0)}_{\alpha,\varepsilon}(x,z) \le C^K_\alpha(r_0) \quad (2.5.10)$$

for some constant $C_\alpha^K(r_0)$.

Now, let us transform the first integral, $K_{\alpha,\varepsilon}^{(r_0)}(x,z)$. For sufficiently small $r_0 > 0$ (smaller than the injectivity radius at every point) one can take the geodesic coordinates at any point $x \in M$ in a ball of radius r_0. Thus, we can take a point $x' \in \mathbb{R}^k$, let $\psi : B_{r_0}(x') \to B_{r_0}(x)$ be geodesic coordinates, and denote $z' := \psi^{-1}(z)$.

Fix any $\delta_1 > 0$. Due to the compactness of M, for a sufficiently small r_0 the Jacobian and the bi-Lipschitz constants of ψ are δ_1-close to 1:

$$\mathfrak{L}(\psi) < 1 + \delta_1, \quad \mathrm{Jac}(\psi) \in \left(\frac{1}{1+\delta_1}, 1+\delta_1\right). \tag{2.5.11}$$

Making a change of variables $y = \psi(y')$ in the integral for $K_{\alpha,\varepsilon}^{(r_0)}(x,z)$, we get

$$\begin{aligned} K_{\alpha,\varepsilon}^{(r_0)}(x,z) &= \int_{B_{r_0}(x)} \varphi_{\alpha,\varepsilon}(d(x,y))\varphi_{\alpha,\varepsilon}(d(z,y))\, d\mathrm{Leb}_M(y) \\ &= \int_{B_{r_0}(x')} \varphi_{\alpha,\varepsilon}(d(\psi(x'),\psi(y')))\varphi_{\alpha,\varepsilon}(d(\psi(z'),\psi(y')))\, d\mathrm{Leb}_M(\psi(y')); \end{aligned}$$

all the three quotients

$$\frac{\varphi_{\alpha,\varepsilon}(d(\psi(x'),\psi(y')))}{\varphi_{\alpha,\varepsilon}(|x'-y'|)}, \quad \frac{\varphi_{\alpha,\varepsilon}(d(\psi(z'),\psi(y')))}{\varphi_{\alpha,\varepsilon}(|z'-y'|)}, \quad \left.\frac{d\,\mathrm{Leb}_M}{d\,\psi_*\mathrm{Leb}_{\mathbb{R}^k}}\right|_y$$

are close to 1: the first two quotients, as well as their reciprocals, do not exceed $(1+\delta_1)^{\frac{k+\alpha}{2}}$ due to Proposition 2.5.3, and the last one and its reciprocal are bounded from above by $(1+\delta_1)$. Thus, $K_{\alpha,\varepsilon}^{(r_0)}(x,z)$ differs from

$$I_{\alpha,\varepsilon}^{(r_0)}(x',z') := \int_{B_{r_0}(x')} \varphi_{\alpha,\varepsilon}(|x'-y'|)\varphi_{\alpha,\varepsilon}(|z'-y'|)\, d\mathrm{Leb}_{\mathbb{R}^k}(y') \tag{2.5.12}$$

by the factor at most $(1+\delta_1)^{k+\alpha+1}$:

$$\frac{1}{(1+\delta_1)^{k+\alpha+1}} \le \frac{K^{(r_0)}_{\alpha,\varepsilon}(x,z)}{I^{(r_0)}_{\alpha,\varepsilon}(x',z')} \le (1+\delta_1)^{k+\alpha+1}$$

Now, $I^{(r_0)}_{\alpha,\varepsilon}(x',z')$ is also a part of $I_{\alpha,\varepsilon}(x',z')$, that differs from it by

$$\begin{aligned}\overline{I}^{(r_0)}_{\alpha,\varepsilon}(x',z') &:= I_{\alpha,\varepsilon}(x',z') - I^{(r_0)}_{\alpha,\varepsilon}(x',z') \\ &= \int_{\mathbb{R}^k\setminus B_{r_0}(x')} \varphi_{\alpha,\varepsilon}(|x'-y'|)\varphi_{\alpha,\varepsilon}(|z'-y'|)\, d\mathrm{Leb}_{\mathbb{R}^k}(y') \quad (2.5.13)\end{aligned}$$

Note that this difference is bounded uniformly in ε, x', z'. Indeed, on one hand, for $|x'-z'| < \frac{r_0}{2}$ this is due to the convergence of the integral of $r^{-k-\alpha}$ at infinity, as in the domain of integration in (2.5.13) both distances $|x'-y'|, |z'-y'|$ are bounded from below by $\frac{r_0}{2} > \varepsilon$, and hence the integrated function does not exceed

$$\varphi_{\alpha,\varepsilon}(|x'-y'|)\varphi_{\alpha,\varepsilon}(|z'-y'|) = |x'-y'|^{-\frac{k+\alpha}{2}}|z'-y'|^{-\frac{k+\alpha}{2}} \le 2^{\frac{k+\alpha}{2}}|x'-y'|^{-(k+\alpha)}.$$

On the other hand, for $|x'-z'| \ge \frac{r_0}{2}$ such a bound follows from Lemma 2.5.5, as

$$\overline{I}^{(r_0)}_{\alpha,\varepsilon}(x',z') \le U_{\alpha,\varepsilon}(|x'-z'|) \le c_\alpha \left(\frac{r_0}{2}\right)^{-\alpha}.$$

Thus, first choosing δ_1 so that $(1+\delta_1)^{k+\alpha+2} < 1+\delta$ and then accordingly choosing r_0 (so that (2.5.11) holds), we obtain the desired estimate (2.5.8).

□

2.5.3 $E_\mu\left[\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(f_*\nu)\right]$ vs $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)$: Proof of Proposition 2.3.16

Proof of Proposition 2.3.16. Let us start by establishing that the estimate (2.3.15) holds with $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}$ replaced by $\mathcal{E}_{\alpha,\varepsilon}$. Indeed, from Proposition 2.3.13 we know that for any $f \in \mathrm{Diff}^1(M)$ one has

$$\mathfrak{L}(f)^{-\alpha} \leq \frac{\mathcal{E}_{\alpha,\varepsilon}(f_*\nu)}{\mathcal{E}_{\alpha,\varepsilon}(\nu)} \leq \mathfrak{L}(f)^{\alpha}.$$

Taking the expectation with respect to measure μ we get

$$\frac{\mathbb{E}_\mu\left[\mathcal{E}_{\alpha,\varepsilon}(f_*\nu)\right]}{\mathcal{E}_{\alpha,\varepsilon}(\nu)} \in \left(\mathbb{E}_\mu\left[\mathfrak{L}(f)^{-\alpha}\right], \mathbb{E}_\mu\left[\mathfrak{L}(f)^{\alpha}\right]\right).$$

It remains to notice that by Hölder's inequality we have

$$\mathbb{E}_\mu\left[\mathfrak{L}(f)^{\alpha}\right] \leq \left(\mathbb{E}_\mu\left[\mathfrak{L}(f)^{\gamma}\right]\right)^{\alpha/\gamma}$$

and by Jensen's inequality we also have

$$\left(\mathbb{E}_\mu\left[\mathfrak{L}(f)^{\alpha}\right]\right)^{-1} \leq \mathbb{E}_\mu\left[\mathfrak{L}(f)^{-\alpha}\right].$$

Hence, taking C_0 such that $\mathbb{E}_\mu\left[\mathfrak{L}(f)^{\gamma}\right] < C_0$, one gets that for any $\alpha < \gamma$,

$$\frac{\mathbb{E}_\mu\left[\mathcal{E}_{\alpha,\varepsilon}(f_*\nu)\right]}{\mathcal{E}_{\alpha,\varepsilon}(\nu)} \in \left(C_0^{-\frac{\alpha}{\gamma}}, C_0^{\frac{\alpha}{\gamma}}\right).$$

Since $C_0^{\frac{\alpha}{\gamma}}$ tends to 1 as α tends to 0, for any $\delta_E > 0$ there exists α_0 such that for any $\alpha \in (0, \alpha_0)$ one has

$$\frac{\mathbb{E}_\mu\left[\mathcal{E}_{\alpha,\varepsilon}(f_*\nu)\right]}{\mathcal{E}_{\alpha,\varepsilon}(\nu)} \in \left(\frac{1}{1+\delta_E}, 1+\delta_E\right). \tag{2.5.14}$$

Let us now return back to the original (2.3.15). Namely, let $\delta > 0$ be given; choose and fix $\delta_E > 0$ such that $(1+\delta_E)^3 < 1+\delta$, and let α_0 be chosen w.r.t. δ_E so that (2.5.14) holds.

By Proposition 2.3.10, there exists $C_E > 0$ such that for any measure ν' on M and any $\varepsilon > 0$ one has

$$\frac{1}{1+\delta_E}\,\mathcal{E}_{\alpha,\varepsilon}(\nu') - \frac{C_E}{1+\delta_E} < \widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu') < (1+\delta_E)\mathcal{E}_{\alpha,\varepsilon}(\nu') + C_E \tag{2.5.15}$$

Applying this for $\nu' = f_*\nu$ and taking the expectation w.r.t. μ provides

$$\frac{1}{1+\delta_E}\,\mathbb{E}_\mu\left[\mathcal{E}_{\alpha,\varepsilon}(f_*\nu)\right] - \frac{C_E}{1+\delta_E} < \mathbb{E}_\mu\left[\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(f_*\nu)\right] < (1+\delta_E)\mathbb{E}_\mu\left[\mathcal{E}_{\alpha,\varepsilon}(f_*\nu)\right] + C_E;$$

using (2.5.14), we thus get

$$\frac{1}{(1+\delta_E)^2}\,\mathcal{E}_{\alpha,\varepsilon}(\nu) - \frac{C_E}{1+\delta_E} < \mathbb{E}_\mu\left[\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(f_*\nu)\right] < (1+\delta_E)^2\mathcal{E}_{\alpha,\varepsilon}(\nu) + C_E.$$

Finally, using (2.5.15) with $\nu' = \nu$ to estimate $\mathcal{E}_{\alpha,\varepsilon}$ via $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}$, we get

$$\frac{1}{(1+\delta_E)^2}\cdot\frac{1}{(1+\delta_E)}\left(\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) - C_E\right) - \frac{C_E}{1+\delta_E} < \mathbb{E}_\mu\left[\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(f_*\nu)\right] < (1+\delta_E)^2\cdot\left((1+\delta_E)\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) + C_E\right) + C_E;$$

as $(1+\delta_E)^3 < 1+\delta$, we get the desired

$$\frac{1}{(1+\delta)}(\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) - C') < \mathbb{E}_\mu\left[\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(f_*\nu)\right] < (1+\delta)\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) + C'$$

for some constant C'. As $\delta > 0$ was arbitrary, and taking (2.5.15) into account again, (2.3.15) follows. □

2.5.4 Estimating $W(f_*\theta_{\alpha,\varepsilon}(\nu), \theta_{\alpha,\varepsilon}(f_*\nu))$: Proof of Proposition 2.3.21

Proof of Proposition 2.3.21. For given $\alpha, \varepsilon > 0$ and measure ν, consider a (non-probability) measure $m_{\alpha,\varepsilon}(\nu)$ on $M \times M \times M$, given by

$$m_{\alpha,\varepsilon}(\nu) = \varphi_{\alpha,\varepsilon}(d(x,y))\varphi_{\alpha,\varepsilon}(d(z,y))\,\nu(dx)\,\mathrm{Leb}_M(dy)\,\nu(dz),$$

where we use (x, y, z) for the coordinates on $M \times M \times M$.

Denote by π_1, π_2 the projections of $M \times M \times M$ on the first and second coordinate respectively, and by $\pi_{1,3}$ the projection on $M \times M$ corresponding to the first and third coordinates. Then directly by definition

$$(\pi_2)_*\, m_{\alpha,\varepsilon}(\nu) = \Theta_{\alpha,\varepsilon}(\nu);$$

also, define

$$\widehat{\Theta}_{\alpha,\varepsilon}(\nu) := (\pi_{1,3})_*\, m_{\alpha,\varepsilon}(\nu) = K_{\alpha,\varepsilon}(x,z)\,\nu(dx)\,\nu(dz); \tag{2.5.16}$$

the second equality is due to (2.3.4). Finally, consider the measure

$$\Theta'_{\alpha,\varepsilon}(\nu) := (\pi_1)_*\, m_{\alpha,\varepsilon}(\nu) \tag{2.5.17}$$

as well as the normalizations of these measures,

$$\widehat{\theta}_{\alpha,\varepsilon}(\nu) := \frac{1}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)}\widehat{\Theta}_{\alpha,\varepsilon}(\nu), \quad \theta'_{\alpha,\varepsilon}(\nu) := \frac{1}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)}\Theta'_{\alpha,\varepsilon}(\nu). \tag{2.5.18}$$

For a high-energy measure ν most of the measure $m_{\alpha,\varepsilon}(\nu)$ is concentrated near the diagonal, and hence the projections on the first and on the second coordinates are close to each other.

The following lemma formalizes this argument:

Lemma 2.5.9. *For any δ_1, α there exists C' such that for any $\varepsilon > 0$ and ν with $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) > C'$ one has*

$$W(\theta_{\alpha,\varepsilon}(\nu), \theta'_{\alpha,\varepsilon}(\nu)) < \delta_1.$$

Proof. Take $r_0 := \frac{\delta_1}{2}$ and let

$$A_{r_0} := \{(x, y, z) \in M^3 \mid d(x, y) < r_0\}, \quad \overline{A}_{r_0} := M^3 \setminus A_{r_0}.$$

Using the notations as in the proof of Lemma 2.5.8, we have

$$m_{\alpha,\varepsilon}(\nu)(\overline{A}_{r_0}) = \iint \mathbf{K}^{(r_0)}_{\alpha,\varepsilon}(x, z)\, d\nu(x)\, d\nu(z) \le C^K_\alpha(r_0),$$

where $C^K_\alpha(r_0)$ is an upper bound for $\mathbf{K}^{(r_0)}_{\alpha,\varepsilon}(x, z)$, see (2.5.10). Thus, the non-normalized measure $m_{\alpha,\varepsilon}(\nu)(\overline{A}_{r_0})$ does not exceed a constant $C^K_\alpha(r_0)$.

Now, we can couple the normalized measures $\theta_{\alpha,\varepsilon}(\nu)$ and $\theta'_{\alpha,\varepsilon}(\nu)$ using the projection of $\frac{1}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)} m_{\alpha,\varepsilon}(\nu)$ on the first two coordinates; this coupling leads to the upper bound for the Wasserstein distance

$$\begin{aligned} W(\theta_{\alpha,\varepsilon}(\nu), \theta'_{\alpha,\varepsilon}(\nu)) &\le \iiint d(x, y)\, \frac{dm_{\alpha,\varepsilon}(\nu)(x, y, z)}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)} \\ &\le r_0 \cdot \frac{m_{\alpha,\varepsilon}(\nu)(A_{r_0})}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)} + \operatorname{diam}(M) \cdot \frac{m_{\alpha,\varepsilon}(\nu)(\overline{A}_{r_0})}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)} \le r_0 + \frac{C^K_\alpha(r_0)}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)}, \end{aligned} \tag{2.5.19}$$

where the first summand corresponds to the points with $d(x, y) < r_0$, and the second to the points with $d(x, y) \ge r_0$. As we chose $r_0 = \frac{\delta_1}{2}$, it suffices to require that

$$\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) > \frac{2\, C^K_\alpha(r_0)}{\delta_1} =: C'$$

to ensure that the total Wasserstein distance does not exceed δ_1. $\square$

On the other hand, for the measures $\theta'_{\alpha,\varepsilon}(\nu)$ the analogue of Proposition 2.3.21 can be established directly, and one can even estimate the total variations distance:

Lemma 2.5.10. *For any $\delta_2 > 0$ and any $R > 0$ there exists $\alpha_2 > 0$ such that for any $\alpha \in (0, \alpha_2)$ there exists $C > 0$ such that for any $f \in \mathrm{Diff}^1(M)$ with $\mathfrak{L}(f) < R$, any $\varepsilon > 0$, and any ν such that $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) > C$ or $\mathcal{E}_{\alpha,\varepsilon}(\nu) > C$ one has*

$$\mathrm{TV}(f_*\theta'_{\alpha,\varepsilon}(\nu), \theta'_{\alpha,\varepsilon}(f_*\nu)) < \delta_2, \tag{2.5.20}$$

and, hence,

$$W(f_*\theta'_{\alpha,\varepsilon}(\nu), \theta'_{\alpha,\varepsilon}(f_*\nu)) < \delta_2 \cdot \mathrm{diam}(M). \tag{2.5.21}$$

Proof. We will first use Lemma 2.5.8 together with Proposition 2.5.3 to compare $K_{\alpha,\varepsilon}(x, z)$ with $K_{\alpha,\varepsilon}(f(x), f(z))$. Namely, choose α_1 and δ_3 so small that

$$(1 + \delta_3)^2 \cdot R^{\alpha_1} < 1 + \frac{\delta_2}{10}.$$

Then from Lemma 2.5.8 we know that for every $\alpha < \alpha_1$ there exists $C > 0$ such that for any x, z, ε we have

$$K_{\alpha,\varepsilon}(x, z) \approx_{(\delta_3, C)} U_{\alpha,\varepsilon}(d(x, z)), \quad K_{\alpha,\varepsilon}(f(x), f(z)) \approx_{(\delta_3, C)} U_{\alpha,\varepsilon}(d(f(x), f(z))),$$

while from (2.5.5) in Proposition 2.5.3 we get

$$\frac{U_{\alpha,\varepsilon}(d(f(x), f(z)))}{U_{\alpha,\varepsilon}(d(x, z))} \in (\mathfrak{L}(f)^{-\alpha}, \mathfrak{L}(f)^{\alpha}) \subset (R^{-\alpha_1}, R^{\alpha_1}).$$

Joining these three estimates together, we obtain

$$K_{\alpha,\varepsilon}(x,z) \approx_{(\frac{\delta_2}{10},C'')} K_{\alpha,\varepsilon}(f(x),f(z)) \tag{2.5.22}$$

for some explicit constant C''.

Now, applying f_*^{-1} does not change the total variations distance, so instead of (2.5.20) we can show the equivalent statement

$$\mathrm{TV}(\theta'_{\alpha,\varepsilon}(\nu), f_*^{-1}\theta'_{\alpha,\varepsilon}(f_*\nu)) < \delta_2. \tag{2.5.23}$$

The measures here can be obtained as projections of measures on $M \times M$:

$$\theta'_{\alpha,\varepsilon}(\nu) = (\pi_1)_* \widehat{\theta}_{\alpha,\varepsilon}(\nu),$$

$$f_*^{-1}\theta'_{\alpha,\varepsilon}(f_*\nu) = (\pi_1)_* \left(f_*^{-1}\widehat{\theta}_{\alpha,\varepsilon}(f_*\nu)\right)$$

where π_1 is the projection on the first coordinate. To obtain the desired (2.5.23), we will actually show that for sufficiently high energy $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)$ even before projection one has

$$\mathrm{TV}(\widehat{\theta}_{\alpha,\varepsilon}(\nu), f_*^{-1}\widehat{\theta}_{\alpha,\varepsilon}(f_*\nu)) < \delta_2. \tag{2.5.24}$$

To do so, note that both these measures are absolutely continuous with respect to $\nu \times \nu$:

$$\widehat{\theta}_{\alpha,\varepsilon}(\nu) = \frac{1}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)} K_{\alpha,\varepsilon}(x,z)\,\nu(dx)\,\nu(dz),$$

$$f_*^{-1}\widehat{\theta}_{\alpha,\varepsilon}(f_*\nu) = \frac{1}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(f_*\nu)} K_{\alpha,\varepsilon}(f(x),f(z))\,\nu(dx)\,\nu(dz).$$

Now, (2.5.22) implies that once $K_{\alpha,\varepsilon}(x,z) > C_2 =: \frac{20C''}{\delta_2}$, the $K_{\alpha,\varepsilon}$ parts of these densities are

close to each other:

$$\frac{1}{1+\frac{\delta_2}{6}} < \frac{K_{\alpha,\varepsilon}(f(x),f(z))}{K_{\alpha,\varepsilon}(x,z)} < 1+\frac{\delta_2}{6}. \tag{2.5.25}$$

On the other hand, due to Corollary 2.3.15, there exists α_0 such that for all $\alpha < \alpha_0$, once the energy $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)$ is sufficiently large, the normalization constants are also close to each other,

$$\frac{1}{1+\frac{\delta_2}{6}}\,\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) < \widetilde{\mathcal{E}}_{\alpha,\varepsilon}(f_*\nu) < (1+\frac{\delta_2}{6})\,\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu). \tag{2.5.26}$$

Take now $\alpha_2 = \min(\alpha_1, \alpha_0)$, so that for any $\alpha < \alpha_2$ both inequalities (2.5.25) and (2.5.26) hold for sufficiently large $K_{\alpha,\varepsilon}(x,z)$ and $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)$ respectively. Multiplying these inequalities, we get that on the set

$$A := \{(x,z) \mid K_{\alpha,\varepsilon}(x,z) > C_2\}$$

the quotient of densities w.r.t. $\nu\times\nu$ of the two measures $\widehat{\theta}_{\alpha,\varepsilon}(\nu)$ and $f_*^{-1}\widehat{\theta}_{\alpha,\varepsilon}(f_*\nu)$ is in the interval $\left(\frac{1}{(1+\frac{\delta_2}{6})^2}, (1+\frac{\delta_2}{6})^2\right)$.

Finally, the $\widehat{\theta}_{\alpha,\varepsilon}(\nu)$-measure of its complement does not exceed $\frac{C_2}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)}$. Hence, if the energy $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)$ is sufficiently high to make sure that $\frac{C_2}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)} < \frac{\delta_2}{6}$, the part that the normalized measures have in common is at least

$$\frac{1-\frac{\delta_2}{6}}{(1+\frac{\delta_2}{6})^2} > 1-\frac{\delta_2}{2},$$

and hence the total variation (2.5.24) indeed does not exceed δ_2.

An application of Lemma 2.3.20 concludes the proof of the upper bound (2.5.21) for the Wasserstein distance.

□

Lemma 2.5.9 and Lemma 2.5.10 together imply Proposition 2.3.21. Indeed, we have:

$$W(f_*\theta_{\alpha,\varepsilon}(\nu), \theta_{\alpha,\varepsilon}(f_*\nu)) \leq W(f_*\theta_{\alpha,\varepsilon}(\nu), f_*\theta'_{\alpha,\varepsilon}(\nu)) + \\ + W(f_*\theta'_{\alpha,\varepsilon}(\nu), \theta'_{\alpha,\varepsilon}(f_*\nu)) + W(\theta'_{\alpha,\varepsilon}(f_*\nu), \theta_{\alpha,\varepsilon}(f_*\nu))$$

The first and the third summands can be estimated directly using Lemma 2.5.9. Indeed, for the first summand, it suffices to note that the application of f increases the Wasserstein distance at most $\mathfrak{L}(f) < R$ times; for the last summand, to apply this lemma, we note that an upper bound on $\mathfrak{L}(f)$ implies that $f_*\nu$ is of sufficiently high energy provided that ν is of high energy.

Finally, the second summand is estimated directly by Lemma 2.5.10. □

2.5.5 Estimating $W(\theta_{\alpha,\varepsilon}(f_*\nu), \theta_{\alpha,\varepsilon}(\mu * \nu))$: Proof of Lemma 2.4.3

Proof of Lemma 2.4.3. We start by working with the non-normalized measures $\Theta_{\alpha,\varepsilon}(f_*\nu)$, $\Theta_{\alpha,\varepsilon}(\mu * \nu)$. To simplify the notation, denote

$$g_f(y) := \rho_{\alpha,\varepsilon}[f_*\nu](y), \quad \overline{g}(y) := \rho_{\alpha,\varepsilon}[\mu * \nu](y).$$

Therefore, we have

$$\Theta_{\alpha,\varepsilon}[f_*\nu] = (g_f(y))^2 d\mathrm{Leb}_M(y) \text{ and } \Theta_{\alpha,\varepsilon}[\mu * \nu] = (\overline{g}(y))^2 d\mathrm{Leb}_M(y). \tag{2.5.27}$$

From (2.4.2) we have

$$\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) \leq \frac{1}{1-\delta}\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\mu * \nu).$$

Joining with (2.4.8), we have

$$\mathbb{E}_f\left[\int_M |g_f(y)-\overline{g}(y)|^2\, d\mathrm{Leb}_M(y)\right] \le \frac{2\delta}{1-\delta}\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\mu*\nu). \tag{2.5.28}$$

Next, there is a following general inequality in normed spaces: for any normed space V and any nonzero $u, v \in V$, one has

$$\left\|\frac{u}{\|u\|}-\frac{v}{\|v\|}\right\| = \left\|\frac{u}{\|u\|}\cdot\left(1-\frac{\|u\|}{\|v\|}\right)+\frac{u-v}{\|v\|}\right\| \le 2\frac{\|u-v\|}{\|v\|}. \tag{2.5.29}$$

Let us apply it to the functions $g_f, \overline{g} \in L_2(M)$; denote

$$h_f := \frac{g_f}{\|g_f\|_{L_2}}, \quad \bar{h} := \frac{\overline{g}}{\|\overline{g}\|_{L_2}},$$

so that normalisation to probability measures of (2.5.27) yields

$$\theta_{\alpha,\varepsilon}[f_*\nu] = h_f^2(y)\, d\mathrm{Leb}_M(y), \quad \theta_{\alpha,\varepsilon}[\mu*\nu] = \bar{h}^2(y)\, d\mathrm{Leb}_M(y).$$

Then, (2.5.29) implies that

$$\|h_f-\bar{h}\|_{L_2} \le 2\frac{\|g_f-\overline{g}\|_{L_2}}{\|\overline{g}\|_{L_2}} \tag{2.5.30}$$

On the other hand, the total variation distance between the normalised measures can be written as

$$\mathrm{TV}(\theta_{\alpha,\varepsilon}[f_*\nu], \theta_{\alpha,\varepsilon}[\mu*\nu]) = \frac{1}{2}\|h_f^2-\bar{h}^2\|_{L_1} = \frac{1}{2}\|(h_f-\bar{h})(h_f+\bar{h})\|_{L_1},$$

and hence by Cauchy inequality

$$\mathrm{TV}(\theta_{\alpha,\varepsilon}[f_*\nu], \theta_{\alpha,\varepsilon}[\mu * \nu])^2 \leq \|h_f - \bar{h}\|_{L_2}^2 \cdot \left\| \frac{h_f + \bar{h}}{2} \right\|_{L_2}^2 \leq \|h_f - \bar{h}\|_{L_2}^2 \leq 4 \frac{\|g_f - \overline{g}\|_{L_2}^2}{\|\overline{g}\|_{L_2}^2};$$

the second inequality is due to $\left\| \frac{h_f + \bar{h}}{2} \right\|_{L_2} \leq 1$, and the last one is due to (2.5.30). The right hand side of the inequality is equal to

$$\frac{4}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\mu * \nu)} \int_M |g_f - \overline{g}|^2 \, d\mathrm{Leb}_M;$$

taking the expectation with respect to f and applying (2.5.28), we get

$$\mathbb{E}_f \mathrm{TV}(\theta_{\alpha,\varepsilon}[f_*\nu], \theta_{\alpha,\varepsilon}[\mu * \nu])^2 \leq 4 \cdot \frac{2\delta}{1 - \delta} \leq 16\delta.$$

In the last inequality we assumed that $\delta \leq \frac{1}{2}$, as the statement holds trivially otherwise. Applying the Markov inequality, we see that

$$\mathbb{P}_\mu \left[\mathrm{TV}(\theta_{\alpha,\varepsilon}[f_*\nu], \theta_{\alpha,\varepsilon}[\mu * \nu]) > 2\sqrt[4]{\delta} \right] < 4\sqrt{\delta},$$

thus establishing conclusion (2.4.10). Finally, the latter immediately implies conclusion (2.4.11) due to Lemma 2.3.20.

□

2.6 Example: necessity of exponential tails

Notice that our main result is applicable to the case when all the maps we consider are projective maps of $\mathbb{RP}^d$. Hölder continuity of a stationary measure in that case is a well

known result by Guivarc'h [25], as we discussed in Section 1.2.1 above. The condition on finiteness of moments (used both in our main result and in [25]) is more restrictive than the condition in the Furstenberg Theorem on random matrix products that ensures positivity of Lyapunov exponent, i.e. $\mathbb{E}\log\|A\| < \infty$, see [16]. Let us give an example that shows that this is not an artifact of the proof, and that the original Furstenberg condition is not sufficient to guarantee Hölder continuity of stationary measures.

Proposition 2.6.1. *There exists a probability distribution μ on $\mathrm{SL}(2,\mathbb{R})$ such that*

1) The smallest subgroup generated by $\operatorname{supp}\mu$ is proximal (in particular, not compact) and strongly irreducible;

2) $\mathbb{E}_\mu \log\|A\| < \infty$;

3) For any $\gamma > 0$ we have $\mathbb{E}_\mu\|A\|^\gamma = \infty$;

4) Any stationary measure on $\mathbb{RP}^1$ is not Hölder continuous.

Proof. For each $n \in \mathbb{N}$ set

$$A_n = R_{\frac{1}{5n}} \cdot \begin{pmatrix} 2^{n^2} & 0 \\ 0 & 2^{-n^2} \end{pmatrix} \cdot R_{-\frac{1}{5n}},$$

where $R_\alpha = \begin{pmatrix} \cos 2\pi\alpha & \sin 2\pi\alpha \\ -\sin 2\pi\alpha & \cos 2\pi\alpha \end{pmatrix}$, and define the distribution μ on $\mathrm{SL}(2,\mathbb{R})$ by

$$\mu(A_n) = p_n \equiv \frac{1}{2^n}.$$

It is straightforward to check that the conditions 1), 2), and 3) of the Proposition 2.6.1 are satisfied. Let us show that 4) also holds. Let us denote by $f_n : \mathbb{RP}^1 \to \mathbb{RP}^1$ a projectivization

of $A_n : \mathbb{R}^2 \to \mathbb{R}^2$. Slightly abusing the notation, let us use the same symbol μ also for the distribution on the space of projective maps with $\mu(f_n) = p_n = \frac{1}{2^n}$. Each map f_n has an attracting fixed point that we denote by s_n. Let ν be a stationary probability measure on $\mathbb{RP}^1$. Due to the explicit form of the matrices A_n we have $\operatorname{supp}\nu \subseteq [0, 2/5]$ (here we identify $\mathbb{RP}^1$ with $[0,1]/_{\{0\sim 1\}}$). If $x \in [0, 2/5] \subset \mathbb{RP}^1$ and $v \in \mathbb{R}^2$ is the corresponding unit vector, then

$$|A_n v| \geq 2^{n^2} \cos\frac{2\pi}{5} > \frac{2^{n^2}}{10}, \text{ and hence } |f_n'(x)| = \frac{1}{|A_n v|^2} < \frac{100}{2^{2n^2}},$$

Notice that with probability p_n^k we can apply the matrix A_n in the random product k times in a row. Since

$$f_n^k(\operatorname{supp}\nu) \subset f_n^k([0,2/5]) \subset \left[s_n - \left(\frac{100}{2^{2n^2}}\right)^k, s_n + \left(\frac{100}{2^{2n^2}}\right)^k \right] \equiv I_{n,k},$$

we have $\nu(I_{n,k}) \geq p_n^k = \frac{1}{2^{kn}}$. This implies that

$$\frac{\log \nu(I_{n,k})}{\log |I_{n,k}|} \leq \frac{-n\log 2}{\frac{\log 2}{k} + \log 100 - 2n^2 \log 2} \to 0 \quad \text{as} \quad n \to \infty,$$

hence the measure ν cannot be Hölder continuous. □

Chapter 3

Log-Hölder regularity of stationary measures

This chapter is structured as follows: in Section 3.1 we formulate the main results, Section 3.2 contains a proof of Theorem 3.1.4, Section 3.3 outlines the proof of Theorem 3.1.9, Section 3.4 collects computational lemmata necessary for the proofs, and Section 3.5 contains the theorem about joint regularity of two measures, which is used in Chapter 4 to prove Central Limit Theorem for nonstationary products of random $\mathrm{SL}(2,\mathbb{R})$ matrices with optimal moments assumption.

3.1 Main results

Let M be a smooth closed Riemannian manifold of dimension k and μ be a Borel probability measure on $\mathrm{Homeo}(M)$, the set of continuous homeomorphisms of M. Consider the

corresponding random dynamical system, given by the compositions

$$T_n := f_n \circ \cdots \circ f_1,$$

where $f_i \in \mathrm{Homeo}(M)$ are chosen randomly and independently, with respect to the distribution μ.

If an initial point $x \in M$ is distributed with respect to a probability measure ν, one can consider the distribution $\mu * \nu$ of its random image $f(x)$. In other words, $\mu * \nu$ is the μ-averaged push-forward image of the measure ν:

$$\mu * \nu := \int_{\mathrm{Homeo}(M)} (f_* \nu) \, \mathrm{d}\mu(f).$$

The measure ν is called *μ-stationary* if $\mu * \nu = \nu$. Let $\mathcal{M}$ denote the space of all Borel probability measures on M equipped with the weak-* topology. Recall the following:

Definition 3.1.1. A measure $\nu \in \mathcal{M}$ is (C, α)-*Hölder* if

$$\forall x \in M \quad \forall r > 0 \quad \nu(B_r(x)) \leq C r^{\alpha}.$$

A measure $\nu \in \mathcal{M}$ is (C, α)-log-*Hölder* if

$$\forall x \in M \quad \forall r > 0 \quad \nu(B_r(x)) \leq C |\log(r)|^{-\alpha}.$$

Let $\mathrm{Hol}(M)$ denote the space of all bi-Hölder continuous homeomorphisms of M and $\mathrm{Lip}(M)$ denote the space of all bi-Lipschitz continuous homeomorphisms of M. We define Hölder and Lipschitz constants by the following formulae:

Definition 3.1.2. For a map $f : M \to M$ and $\gamma > 0$, let

$$H_\gamma(f) = \sup_{\substack{x,y\in M\\ x\neq y}} \frac{d(f(x), f(y))}{d(x, y)^\gamma}$$

be the (possibly infinite) γ-Hölder norm of this map, and if f is a homeomorphism, let

$$H'_\gamma(f) = \max\left(H_\gamma(f), H_\gamma\left(f^{-1}\right)\right).$$

Next, for $f \in \mathrm{Hol}(M)$ let

$$\gamma(f) = \sup\{\gamma > 0 \ : \ H'_\gamma(f) < \infty\}$$

be the critical Hölder regularity for f (that might not be attained), and define

$$L(f) = H'_{\gamma(f)/2}(f)$$

be the Hölder constant corresponding to the regularity $\gamma(f)/2$, that is automatically finite. We also introduce the quantity

$$\Lambda(f) = 2 \cdot \frac{1 + \log(L(f))}{\gamma(f)}$$

as a (convenient for the future computations) measure of (ir)regularity of a bi-Hölder homeomorphism. Finally, for a bi-Lipschitz $f \in \mathrm{Lip}(M)$ we denote the corresponding bi-Lipschitz constant by

$$\mathfrak{L}(f) = H'_1(f).$$

Now we are ready to formulate our main results.

Theorem 3.1.3. *Let μ be a probability measure on* $\mathrm{Hol}(M)$*, satisfying the following assumptions:*

- *There exists $\beta > 0$ such that*

$$\int_{\mathrm{Hol}(M)} \Lambda(f)^{\beta} \ d\mu(f) < \infty.$$

- ***(no invariant measure)*** *There is no measure $m \in \mathcal{M}$, such that $f_* m = m$ for μ-a.e. f.*

Then there exist $\alpha > 0$ and C such that every μ-stationary probability measure ν on M is (C, α)-log-*Hölder.*

Moreover, we prove the following regularity estimate for a nonstationary random dynamical system after finitely many iterations:

Theorem 3.1.4. *Let K be a compact set (with respect to weak-* topology) in the space of Borel probability measures on* $\mathrm{Homeo}(M)$*, satisfying the following assumptions:*

- *For every $\mu \in K$ we have* $\mathrm{supp}(\mu) \subset \mathrm{Hol}(M)$.
- *There exists $\beta > 0$ and C_0 such that for every $\mu \in K$*

$$\int_{\mathrm{Hol}(M)} \Lambda(f)^{\beta} \ d\mu(f) < C_0. \tag{3.1.1}$$

- ***(no deterministic image)*** *For every $\mu \in K$ there are no measures $m_1, m_2 \in \mathcal{M}$, such that $f_* m_1 = m_2$ for μ-a.e. f.*

Then there exist $\alpha > 0$, C and $\kappa > 1$ such that for every initial measure ν_0, every sequence

of measures $\mu_1, \mu_2, \ldots \in K$, every number of iterations $n \in \mathbb{N}$, and every $x \in M$ one has:

$$\text{if } \; r > e^{-\kappa^n} \quad \text{then } \; [\mu_n * \mu_{n-1} * \ldots * \mu_1 * \nu_0](B_r(x)) < C|\log(r)|^{-\alpha}.$$

We would like to reiterate the following connection between the **no invariant measure** and the **no deterministic image** assumptions.

Proposition 3.1.5. *Let μ be a probability measure on Homeo(M) that satisfies **no invariant measure** condition (there is no common invariant measure for all $f \in \operatorname{supp}\mu$). Then there exists $k \in \mathbb{N}$ such that μ^{*k} satisfies **no deterministic images** condition (i.e. there are no probability measures ν, ν' on M such that $f_*\nu = \nu'$ with $f = f_k \circ f_{k-1} \circ \ldots \circ f_1$ for $\mu \times \mu \times \ldots \times \mu$-almost all $(f_1, f_2, \ldots, f_k) \in (\text{Homeo}(M))^k$).*

Indeed, if for any $k \in \mathbb{N}$ there exists a measure with the same image under every $f \in \operatorname{supp}(\mu^k)$ then any accumulation point of Krylov-Bogolyubov time averages of these measures has to be an invariant measure (see Section 2.1.1 for more details).

Applying Theorem 3.1.4 to a one-point set $K = \{\mu\}$ and taking into account Proposition 3.1.5 we obtain the following

Corollary 3.1.6. *Assume that μ satisfies the assumptions of Theorem 3.1.3. Then there exist $\alpha > 0$, C and $\kappa > 1$ such that for every initial measure ν_0, every number of iterations $n \in \mathbb{N}$, and every $x \in M$ one has:*

$$\text{if } \; r > e^{-\kappa^n} \quad \text{then } \; [\mu^{*n} * \nu_0](B_r(x)) < C|\log(r)|^{-\alpha}. \tag{3.1.2}$$

Remark 3.1.7. It is also easy to see that Theorem 3.1.3 follows from Corollary 3.1.6, hence all we need to prove is Theorem 3.1.4.

Using similar techniques we are also able to provide more refined estimates for the case of

bi-Lipschitz homeomorphisms.

Theorem 3.1.8. *Let μ be a probability measure on* $\mathrm{Lip}(M)$*, satisfying the following assumptions:*

- *There exists $\alpha > 0$ such that*

$$\int_{\mathrm{Lip}(M)} (\log(\mathfrak{L}(f)))^{\alpha} \, d\mu(f) < \infty. \tag{3.1.3}$$

- *There is no measure $m \in \mathcal{M}$, such that $f_* m = m$ for μ-a.e. f.*

*Then there exists C, such that every stationary measure $\nu \in \mathcal{M}$ is $(C, \alpha/2)$-*log*-Hölder.*

Similar to the case of Hölder continuous homeomorphisms, a nonstationary version of Theorem 3.1.8 is available.

Theorem 3.1.9. *Let K be a compact set (with respect to weak-* topology) in the space of Borel probability measures on Homeo(M), satisfying the following assumptions:*

- *For every $\mu \in K$ we have* $\mathrm{supp}(\mu) \subset \mathrm{Lip}(M)$.
- *There exists $\alpha > 0$ and C_0 such that for every $\mu \in K$*

$$\int_{\mathrm{Lip}(M)} (\log(\mathfrak{L}(f)))^{\alpha} \; d\mu(f) < C_0. \tag{3.1.4}$$

- *For every $\mu \in K$ there are no measures $m_1, m_2 \in \mathcal{M}$, such that $f_* m_1 = m_2$ for μ-a.e. f.*

Then there exist C and $\kappa > 1$ such that for every initial measure ν_0, every sequence of measures $\mu_1, \mu_2, \ldots \in K$, every number of iterations $n \in \mathbb{N}$, and every $x \in M$ one has:

$$\text{if } \; r > e^{-\kappa^n} \quad \text{then } \; [\mu_n * \mu_{n-1} * \ldots * \mu_1 * \nu_0](B_r(x)) < C |\log(r)|^{-\frac{\alpha}{2}}.$$

The following corollary also holds:

Corollary 3.1.10. *Assume that μ satisfies the assumptions of Theorem 3.1.8. Then there exist C and $\kappa > 1$ such that for every initial measure ν_0, every number of iterations $n \in \mathbb{N}$, and every $x \in M$ one has:*

$$\text{if } \; r > e^{-\kappa^n} \quad \text{then } \; [\mu^{*n} * \nu_0](B_r(x)) < C|\log(r)|^{-\frac{\alpha}{2}}.$$

3.2 Hölder continuous homeomorphisms

3.2.1 Outline of the proof

This section is devoted to the proof of Theorem 3.1.4. As it was already mentioned in Corollary 3.1.6 and Remark 3.1.7, it will immediately imply Theorem 3.1.3. The regularity of a measure ν on M can be established by considering the integral

$$\iint_{M^2} |\log(d(x,z))|^{\alpha} \, \mathrm{d}\nu(x) \, \mathrm{d}\nu(z). \tag{3.2.1}$$

Namely, if it is finite, then the measure of any ball $B_r(x)$ by the Markov inequality doesn't exceed

$$\nu(B_r(x)) \leq c|\log(r)|^{-\frac{\alpha}{2}}.$$

Moreover, a similar estimate can be obtained given finiteness of any integral

$$\mathcal{E}_U(\nu) = \iint_{M \times M} U(d(x,z)) \, \mathrm{d}\nu(x) \, \mathrm{d}\nu(z), \tag{3.2.2}$$

where the function $U(r)$ has the same singularity as $|\log(r)|^\alpha$ as $r \to 0$. Indeed, if a lower bound $U(r) > c|\log(r)|^\alpha$ holds for some $r > \varepsilon$, then for the same r one will have the bound for the measures $\nu(B_r(x))$.

To estimate integrals of the type (3.2.1), we will apply the machinery introduced in Chapter 2. Namely, we will consider the quantity $\mathcal{E}_{\alpha,\varepsilon}(\nu) = \mathcal{E}_{U_{\alpha,\varepsilon}}(\nu)$, associated to some function $U_{\alpha,\varepsilon}$, and satisfying the following properties:

(i) For a fixed $\alpha > 0$ it admits a uniform upper bound, depending only on the choice of the minimal radius ε:

$$\mathcal{E}_{\alpha,\varepsilon}(\nu) \leq U_{\alpha,\varepsilon}(0) \tag{3.2.3}$$

(see Lemma 3.2.6).

(ii) Outside the radius ε it satisfies the lower bound

$$U_{\alpha,\varepsilon}(r) \geq c|\log(r)|^\alpha \tag{3.2.4}$$

(see Proposition 3.2.4 part **(III)**).

(iii) A convolution with any $\mu \in K$ reduces the quantity $\mathcal{E}_{\alpha,\varepsilon}$ by a linear factor: for some $\lambda < 1$, $\tilde{C}$ for any $\varepsilon > 0$, any ν and any $\mu \in K$ one has

$$\mathcal{E}_{\alpha,\varepsilon}(\mu * \nu) < \max(\lambda \mathcal{E}_{\alpha,\varepsilon}(\nu), \tilde{C}) \tag{3.2.5}$$

(see Proposition 3.2.27 and Corollary 3.2.31).

Having constructed such functions, for every n we choose ε so that the upper bound

$$\mathcal{E}_{\alpha,\varepsilon}(\mu_n * \mu_{n-1} * \ldots * \mu_1 * \nu) < \max(\lambda^n U_{\alpha,\varepsilon}(0), \tilde{C}) \tag{3.2.6}$$

obtained from joining (i) and (iii), would be equal to $\tilde{C}$. Then, the application of the Markov inequality to (ii) concludes the proof.

The most technically involved part of the proof is establishing inequality (3.2.5). The key idea is to replace the energy $\mathcal{E}_{\alpha,\varepsilon}(\nu)$ with an equivalent one $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)$ (see Section 3.2.5), where $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}$ can be represented as a square of the norm of an $L^2(M, \mathrm{Leb})$ function. Let us call this function $\rho_{\alpha,\varepsilon}[\nu](y)$. It will follow from the construction that

$$\rho_{\alpha,\varepsilon}[\mu * \nu](y) = \mathbb{E}_{\mu}\rho_{\alpha,\varepsilon}[f_*\nu](y),$$

and that functions $\rho_{\alpha,\varepsilon}[f_*\nu](y)$ are comparable in L^2 norm with $\rho_{\alpha,\varepsilon}[\nu](y)$ for bi-Hölder f. Now the idea of the proof can be formulated as follows: the energy $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\mu * \nu)$ is a square of norm of a convex combination of $L^2(M, \mathrm{Leb})$ vectors of norm comparable to the one of $\rho_{\alpha,\varepsilon}[\nu]$:

$$\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\mu * \nu) = \|\mathbb{E}_{\mu}\rho_{\alpha,\varepsilon}[f_*\nu]\|^2_{L^2}, \quad \text{and} \quad \|\rho_{\alpha,\varepsilon}[f_*\nu]\|_{L^2} \sim \|\rho_{\alpha,\varepsilon}[\nu]\|_{L^2}.$$

If the contraction doesn't happen (inequality (3.2.5) doesn't hold), then the convex combination has a norm that is comparable to that of its components. This fact implies that the components are aligned with each other. In other words, the functions $\rho_{\alpha,\varepsilon}[f_*\nu](y)$ are close to one another for μ-almost all f. Hence, so are the measures given by

$$\theta_{\alpha,\varepsilon}[f_*\nu] = \frac{\rho^2_{\alpha,\varepsilon}[f_*\nu](y)\,\mathrm{dLeb}(y)}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(f_*\nu)}$$

(notice that due to our choice of $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)$ and $\rho_{\alpha,\varepsilon}[\nu]$ the measure $\theta_{\alpha,\varepsilon}[\nu]$ always has total mass equal to 1). Finally, we show that measures $\theta_{\alpha,\varepsilon}[f_*\nu]$ and $f_*\theta_{\alpha,\varepsilon}[\nu]$ are also close for any bi-Hölder f. Now it remains to notice that for μ-almost every f the measures $f_*\theta_{\alpha,\varepsilon}[\nu]$ have to be close to one another, which turns the measure $\theta_{\alpha,\varepsilon}[\nu]$ into a measure with an "almost deterministic image". After passing to a limit, we obtain a true measure with deterministic

image and hence arrive to a contradiction.

The plan for the rest of this section is as follows:

- In Section 3.2.2 we define functions $U_{\alpha,\varepsilon}$ and $\varphi_{\alpha,\varepsilon}$ (which are needed to construct $\mathcal{E}_{\alpha,\varepsilon}$ and $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}$ mentioned above).
- In Section 3.2.3 we collect some useful properties of $U_{\alpha,\varepsilon}$, including the lower bound (3.2.4) (Proposition 3.2.4 part **(III)**).
- In Section 3.2.4 we introduce $\mathcal{E}_{\alpha,\varepsilon}$. We prove the upper bound (3.2.3) (Lemma 3.2.6), estimates on how $\mathcal{E}_{\alpha,\varepsilon}$ behaves if we replace ν with $f_*\nu$ for a bi-Hölder f (Proposition 3.2.8), and a Markov's type inequality (Lemma 3.2.7) to be used in the end of the proof of Theorem 3.1.4.
- In Section 3.2.5 we introduce $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}$, $\rho_{\alpha,\varepsilon}$, and $\theta_{\alpha,\varepsilon}$ mentioned above. We prove that energies $\mathcal{E}_{\alpha,\varepsilon}$ and $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}$ are equivalent (Proposition 3.2.14).
- In Section 3.2.6 we prove that measures $\theta_{\alpha,\varepsilon}[f_*\nu]$ and $f_*\theta_{\alpha,\varepsilon}[\nu]$ are close to each other for any bi-Hölder f (Proposition 3.2.22). We also show that the average (with respect to μ) value of $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(f_*\nu)$ is close to $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)$ (see Proposition 3.2.25). This is where the tail estimate (3.1.1) comes into play.
- In Section 3.2.7 we prove the contraction property (3.2.5) (Proposition 3.2.27).
- In Section 3.2.8 we establish the upper bound (3.2.6) and finish the proof of Theorem 3.1.4.

3.2.2 Choice of the functions $U_{\alpha,\varepsilon}$ and $\varphi_{\alpha,\varepsilon}$

As mentioned above, we would like to find a function $\varphi_\alpha(r)$, such that $\varphi_\alpha * \varphi_\alpha(r)$ has a singularity of order $|\log(r)|^\alpha$ at the origin. A natural candidate is given by a square root of

the derivative of $|\log(r)|^\alpha$. As we will be working on a Riemannian manifold of dimension k we adjust φ_α accordingly:

Definition 3.2.1. Define

$$\varphi_\alpha(x) = \begin{cases} \frac{(-\log(x))^{\frac{\alpha-1}{2}}}{x^{\frac{k}{2}}}, & \text{if } 0 < x < \frac{1}{e}, \\ 0, & \text{otherwise,} \end{cases}$$

and for $r > 0$

$$U_\alpha(r) = \int_{\mathbb{R}^k} \varphi_\alpha(|\mathbf{x}|)\varphi_\alpha(|\bar{r} - \mathbf{x}|)\,\mathrm{d}\mathbf{x} = \\ = \int_{\{|\mathbf{x}|<\frac{1}{e}\}\cap\{|\bar{r}-\mathbf{x}|<\frac{1}{e}\}} \varphi_\alpha(|\mathbf{x}|)\varphi_\alpha(|\bar{r} - \mathbf{x}|)\,\mathrm{d}\mathbf{x},$$

where $\mathbf{x} \in \mathbb{R}^k$ and $\bar{r} = (r, 0, \ldots, 0) \in \mathbb{R}^k$.

The function $U_\alpha(r)$ has the same singularity at the origin as $|\log(r)|^\alpha$ (see Lemma 3.4.1), but still needs a little adjustment, because the integral

$$\iint_{M^2} U_\alpha(d(x, z))\,\mathrm{d}\nu(x)\,\mathrm{d}\nu(z)$$

might be infinite and the estimate mentioned in Part (i) of the outline will be meaningless. To fix that problem we define the following family of cut-offs:

Definition 3.2.2. Define

$$\varphi_{\alpha,\varepsilon}(x) = \begin{cases} \varphi_\alpha(x), & \text{if } x \geq \varepsilon; \\ \varphi_\alpha(\varepsilon), & \text{if } 0 < x < \varepsilon \end{cases}$$

and

$$U_{\alpha,\varepsilon}(r) = \int_{\mathbb{R}^k} \varphi_{\alpha,\varepsilon}(|\mathbf{x}|)\varphi_{\alpha,\varepsilon}(|\bar{r}-\mathbf{x}|)\,\mathrm{d}\mathbf{x} = \\ = \int_{\{|\mathbf{x}|<\frac{1}{e}\}\cap\{|\bar{r}-\mathbf{x}|<\frac{1}{e}\}} \varphi_{\alpha,\varepsilon}(|\mathbf{x}|)\varphi_{\alpha,\varepsilon}(|\bar{r}-\mathbf{x}|)\,\mathrm{d}\mathbf{x}, \quad (3.2.7)$$

where $\mathbf{x} \in \mathbb{R}^k$ and $\bar{r} = (r, 0, \ldots, 0) \in \mathbb{R}^k$.

3.2.3 Properties of $U_{\alpha,\varepsilon}$

In this section we collect some important properties of $U_{\alpha,\varepsilon}$.

An important symmetry that the function $U_{\alpha,\varepsilon}$ possesses is described in the following

Remark 3.2.3. Note that the definition of $U_{\alpha,\varepsilon}$ can be reformulated as follows: for any two points $\mathbf{x}, \bar{z} \in \mathbb{R}^k$ consider the integral

$$I_{\alpha,\varepsilon}(\mathbf{x}, \bar{z}) := \int_{\mathbb{R}^k} \varphi_{\alpha,\varepsilon}(|\bar{y}-\mathbf{x}|)\varphi_{\alpha,\varepsilon}(|\bar{y}-\bar{z}|)\ \mathrm{dLeb}(\bar{y}). \quad (3.2.8)$$

Due to the spherical symmetry, this integral depends only on the distance $r = |\mathbf{x}-\bar{z}|$ between these two points:

$$I_{\alpha,\varepsilon}(\mathbf{x}, \bar{z}) = U_{\alpha,\varepsilon}(|\mathbf{x}-\bar{z}|)$$

for some function $U_{\alpha,\varepsilon}$ of $r = |\mathbf{x}-\bar{z}|$ (which of course has to coincide with $U_{\alpha,\varepsilon}$ defined by (3.2.7)). We can take this as a definition of $U_{\alpha,\varepsilon}(r)$.

Let us denote by $c_{\alpha,k}$ the following constant:

$$c_{\alpha,k} = \frac{k\omega_k}{\alpha},$$

where ω_k is the volume of a k-dimensional unit ball. The following Proposition establishes the properties of $U_{\alpha,\varepsilon}$ outlined in parts (i) and (ii) of the plan, together with some other estimates that will be useful later.

Proposition 3.2.4. *For every dimension $k \geq 1$ the following holds:*

(I) *For every $\alpha > 0$ and $\varepsilon > 0$ the function $U_{\alpha,\varepsilon}$ is non-increasing.*

(II) *For every $\alpha_0 > 0$ there exist $r_0 > 0$ and $\varepsilon_0 > 0$, such that for every $0 < \alpha < \alpha_0$, every $0 < \varepsilon < \varepsilon_0$, and every $0 < r < r_0$ we have:*

$$U_{\alpha,\varepsilon}(r) < 2c_{\alpha,k}(-\log(r))^{\alpha}. \tag{3.2.9}$$

(III) *For every $\alpha_0 > 0$ there exist $\varepsilon_0 > 0$, such that for every $0 < \alpha < \alpha_0$, and every r, ε, such that $0 < \varepsilon < r < \varepsilon_0$ we have:*

$$U_{\alpha,\varepsilon}(r) > \frac{c_{\alpha,k}}{2}(-\log(r))^{\alpha}. \tag{3.2.10}$$

(IV) *For every $\alpha_0 > 0$ and every $\delta > 0$ there exist $\varepsilon_0 > 0$ and $r_0 > 0$ such that for every $0 < \alpha < \alpha_0$, every $0 < \varepsilon < \varepsilon_0$, and every $0 < r_1 \leq r_2 \leq r_0$ we have*

$$U_{\alpha,\varepsilon}(r_1) \leq (1+\delta)\left(\frac{\log(r_1)}{\log(r_2)}\right)^{\alpha} U_{\alpha,\varepsilon}(r_2). \tag{3.2.11}$$

We will postpone the proof of Proposition 3.2.4 until Section 3.4.

3.2.4 Definition and properties of $\mathcal{E}_{\alpha,\varepsilon}$: behaviour under images and convolutions

We are ready to define the energy mentioned in (3.2.2).

Definition 3.2.5.

$$\mathcal{E}_{\alpha,\varepsilon}(\nu) = \iint_{M^2} U_{\alpha,\varepsilon}(d(x,z))d\nu(x)d\nu(z).$$

Let us point out a few useful properties of $\mathcal{E}_{\alpha,\varepsilon}$. We start with a precise estimate for the energy of an arbitrary measure ν for given α and ε (as mentioned in part (i) of the outline):

Lemma 3.2.6. *For every $\alpha > 0$, every $\nu \in \mathcal{M}$ and every $0 < \varepsilon < 1/e$ we have the following upper bound:*

$$\mathcal{E}_{\alpha,\varepsilon}(\nu) \leq (\omega_k + c_{\alpha,k})(-\log(\varepsilon))^\alpha \tag{3.2.12}$$

Proof. A straightforward computation shows that

$$\mathcal{E}_{\alpha,\varepsilon}(\nu) \leq U_{\alpha,\varepsilon}(0) = \int_{0<|\bar{x}|<\varepsilon} \varphi_\alpha^2(\varepsilon)\,\mathrm{d}\bar{x} + \int_{\varepsilon<|\bar{x}|<\frac{1}{e}} \varphi_\alpha^2(|\bar{x}|)\,\mathrm{d}\bar{x}.$$

Both summands can be computed exactly. For the first one we substitute the definition of $\varphi_\alpha(\varepsilon)$:

$$\int_{0<|\bar{x}|<\varepsilon} \varphi_\alpha^2(\varepsilon)\,\mathrm{d}\bar{x} = \frac{\omega_k \varepsilon^k (-\log(\varepsilon))^{\alpha-1}}{\varepsilon^k} = \omega_k(-\log(\varepsilon))^{\alpha-1}.$$

For the second one we use spherical coordinates (see Lemma 3.4.1 for details):

$$\int_{\varepsilon<|\bar{x}|<\frac{1}{e}} \varphi_\alpha^2(|\bar{x}|)\,\mathrm{d}\bar{x} = c_{\alpha,k}((-\log(\varepsilon))^\alpha - 1).$$

Adding together the two inequalities above we arrive to (3.2.12).

□

Next, we show a Markov type inequality that estimates $\nu(B_r(x))$ through $\mathcal{E}_{\varepsilon,\alpha}(\nu)$:

Lemma 3.2.7. *For every $\alpha > 0$ there exists $C_{\alpha,k} < \infty$ and $\varepsilon_0 > 0$, such that for every $0 < \varepsilon < \varepsilon_0$, every $\nu \in \mathcal{M}$, every $0 < \varepsilon < r < \frac{1}{e}$ and every ball $B_r(x) \subset M$ the following holds:*

$$\nu(B_r(x)) \leq C_{\alpha,k}\sqrt{\mathcal{E}_{\alpha,\varepsilon}(\nu)}(-\log(r))^{-\frac{\alpha}{2}}. \tag{3.2.13}$$

Proof. Let us choose ε_0 such that Part **(III)** of Proposition 3.2.4 is applicable. Then for small enough r applying Markov inequality to the definition of $\mathcal{E}_{\alpha,\varepsilon}$ we obtain:

$$\mathcal{E}_{\alpha,\varepsilon}(\nu) \geq U_{\alpha,\varepsilon}(2r) \cdot \nu(B_r(x))^2 \geq \frac{C_{\alpha,k}}{2}(-\log(2r))^{\alpha} \cdot \nu(B_r(x))^2.$$

Choosing $C_{\alpha,k}$ big enough we can ensure that (3.2.13) holds for all r such that $\varepsilon < r < 1/e$. □

Now we need some preparations in order to describe the energy $\mathcal{E}_{\alpha,\varepsilon}(\mu * \nu)$ that we get after one step of our random dynamics. The following estimates will be useful for proving the contraction mentioned in part (iii) of the outline. We start with the following series of statements describes the change of the energy $\mathcal{E}_{\alpha,\varepsilon}(\nu)$ after applying a Hölder continuous homeomorphism.

Proposition 3.2.8. *For every $\delta > 0$ there exists $\varepsilon_0 > 0$, such that for every $0 < \alpha < 1$ there exist $A = A(\alpha, \delta)$, such that for every $0 < \varepsilon < \varepsilon_0$, every $f \in \mathrm{Hol}(M)$ and every $\nu \in \mathcal{M}$ the following holds:*

$$\mathcal{E}_{\alpha,\varepsilon}(f_*\nu) \leq \Lambda(f)^{\alpha}\left((1+\delta)\mathcal{E}_{\alpha,\varepsilon}(\nu) + A\right). \tag{3.2.14}$$

Proof. According to the Part **(IV)** of Proposition 3.2.4 we can choose $\varepsilon_0 > 0$ and $r_0 > 0$,

such that for every $0 < \alpha < 1$, every $0 < \varepsilon < \varepsilon_0$ and every $0 < r_1 \leq r_2 \leq r_0$ we have

$$U_{\alpha,\varepsilon}(r_1) \leq (1+\delta)\left(\frac{\log(r_1)}{\log(r_2)}\right)^{\alpha} U_{\alpha,\varepsilon}(r_2).$$

Let us split the integral from the definition of energy the $\mathcal{E}_{\alpha,\varepsilon}(f_*\nu)$ in the following way:

$$\begin{aligned}
\mathcal{E}_{\alpha,\varepsilon}(f_*\nu) &= \iint_{M\times M} U_{\alpha,\varepsilon}(d(x,z))\,\mathrm{d}f_*\nu(x)\,\mathrm{d}f_*\nu(z) = \\
&= \iint_{M\times M} U_{\alpha,\varepsilon}(d(f(x),f(z)))\,\mathrm{d}\nu(x)\,\mathrm{d}\nu(z) = \\
&= \iint_{d(x,z)<r_0} U_{\alpha,\varepsilon}(d(f(x),f(z)))\,\mathrm{d}\nu(x)\,\mathrm{d}\nu(z) + \\
&\qquad + \iint_{d(x,z)\geq r_0} U_{\alpha,\varepsilon}(d(f(x),f(z)))\,\mathrm{d}\nu(x)\,\mathrm{d}\nu(z).
\end{aligned}$$

To estimate the first summand we take $r_2 = d(x,z)$ and $r_1 = \min(d(f(x),f(z)), r_2)$ and we deduce that

$$\begin{aligned}
U_{\alpha,\varepsilon}(d(f(x),f(z))) \leq U_{\alpha,\varepsilon}(r_1) &\leq (1+\delta)\left(\frac{\log(r_1)}{\log(r_2)}\right)^{\alpha} U_{\alpha,\varepsilon}(r_2) \leq \\
&\leq (1+\delta)\max\left(1, \left(\frac{\log(d(f(x),f(z)))}{\log(d(x,z))}\right)^{\alpha}\right) U_{\alpha,\varepsilon}(d(x,z))
\end{aligned}$$

for any $x, z \in M$, such that $d(x,z) \leq r_0$. By definition of $\gamma(f)$ and $L(f)$ we know that

$$d(f(x),f(z)) \geq \left(\frac{d(x,z)}{L(f)}\right)^{\frac{2}{\gamma(f)}}$$

for any $x, z \in M$. After substituting that into previous inequality we arrive to

$$\begin{aligned}
&U_{\alpha,\varepsilon}(d(f(x),f(z))) \leq \\
&\leq (1+\delta)\left(\frac{2(\log(d(x,z)) - \log(L(f)))}{\gamma(f)\log(d(x,z))}\right)^{\alpha} U_{\alpha,\varepsilon}(d(x,z)) \leq \\
&\qquad\qquad \leq (1+\delta)\Lambda(f)^{\alpha} U_{\alpha,\varepsilon}(d(x,z)).
\end{aligned}$$

For the first summand the last inequality gives us the following:

$$\iint_{d(x,z)<r_0} U_{\alpha,\varepsilon}(d(f(x),f(z)))\,\mathrm{d}\nu(x)\,\mathrm{d}\nu(z) \leq$$
$$\leq (1+\delta)\iint_{M\times M} \Lambda(f)^{\alpha} U_{\alpha,\varepsilon}(d(x,z))\,\mathrm{d}\nu(x)\,\mathrm{d}\nu(z) \leq$$
$$\leq (1+\delta)\Lambda(f)^{\alpha}\mathcal{E}_{\alpha,\varepsilon}(\nu).$$

Let us take r_0 and ε_0 small enough to ensure that the conclusion of Part **(II)** of Proposition 3.2.4 also holds. That allows us to estimate the second summand as follows:

$$\iint_{d(x,z)\geq r_0} U_{\alpha,\varepsilon}(d(f(x),f(z)))\,\mathrm{d}\nu(x)\,\mathrm{d}\nu(z) \leq$$
$$\leq \iint_{M\times M} U_{\alpha,\varepsilon}\left(\left(\frac{r_0}{L(f)}\right)^{\frac{2}{\gamma(f)}}\right)\mathrm{d}\nu(x)\,\mathrm{d}\nu(z) \leq$$
$$\leq 2c_{\alpha,k}\left(-\log\left(\left(\frac{r_0}{L(f)}\right)^{\frac{2}{\gamma(f)}}\right)\right)^{\alpha} \leq$$
$$\leq 2c_{\alpha,k}(-\log(r_0))^{\alpha}\Lambda(f)^{\alpha}.$$

Taking $A = 2c_{\alpha,k}(-\log(r_0))^{\alpha}$ finishes the proof. □

An immediate corollary can be formulated as follows:

Corollary 3.2.9. *For every $\delta > 0$ there exists $\varepsilon_0 > 0$, such that for every $0 < \alpha < 1$ there exists $C < \infty$, such that for every $0 < \varepsilon < \varepsilon_0$, every $f \in \mathrm{Hol}(M)$ and every $\nu \in \mathcal{M}$, such that $\mathcal{E}_{\alpha,\varepsilon}(\nu) > C$ the following holds:*

$$\Lambda(f)^{-\alpha}\frac{1}{1+\delta} \leq \frac{\mathcal{E}_{\alpha,\varepsilon}(f_*\nu)}{\mathcal{E}_{\alpha,\varepsilon}(\nu)} \leq \Lambda(f)^{\alpha}(1+\delta). \tag{3.2.15}$$

Another corollary of this statement that would be convenient to use is

Corollary 3.2.10. *There exists $\varepsilon_0 > 0$, such that for every $0 < \alpha < 1$ there exists $C < \infty$,*

such that for every $0 < \varepsilon < \varepsilon_0$*, every* $f \in \mathrm{Hol}(M)$ *and every* $\nu \in \mathcal{M}$ *the following holds:*

$$\mathcal{E}_{\alpha,\varepsilon}(f_*\nu) \leq \max\left(2\Lambda(f)^\alpha \mathcal{E}_{\alpha,\varepsilon}(\nu), C\right). \tag{3.2.16}$$

Proof. Applying Corollary 3.2.9 to $\delta = \frac{1}{2}$, $\tilde{\nu} := f_*\nu$ and $\tilde{f} = f^{-1}$ we conclude that there exists $C < \infty$, such that if $\mathcal{E}_{\alpha,\varepsilon}(\tilde{\nu}) = \mathcal{E}_{\alpha,\varepsilon}(f_*\nu) > C$ then

$$\mathcal{E}_{\alpha,\varepsilon}(f_*^{-1}f_*\nu) \geq \frac{1}{1+\frac{1}{2}}\Lambda\left(f^{-1}\right)^{-\alpha}\mathcal{E}_{\alpha,\varepsilon}(f_*\nu).$$

It remains to recall that $\Lambda(f) = \Lambda(f^{-1})$ and the inequality (3.2.16) follows. □

3.2.5 Definition of $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}$ and comparison to $\mathcal{E}_{\alpha,\varepsilon}$

Following the technique from Chapter 2 we introduce another energy $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)$. The intuition behind it can be described as follows: on one hand, for singular enough ν (such that $\mathcal{E}_{\alpha,\varepsilon}(\nu)$ is big) the values $\mathcal{E}_{\alpha,\varepsilon}(\nu)$ and $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)$ will be close to each other, so we can use $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)$ to estimate $\mathcal{E}_{\alpha,\varepsilon}(\nu)$. On the other hand, the energy $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}$ can be represented as a square of the norm of an $L^2(M, \mathrm{Leb})$ vector, which allows us to use geometry of this Hilbert space when working with $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}$. This makes $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}$ a crucial tool for proving the contraction property described in part (iii) of the outline.

We start by defining the corresponding $L^2(M, \mathrm{Leb})$ vector mentioned above:

Definition 3.2.11. For $\nu \in \mathcal{M}$ define

$$\rho_{\alpha,\varepsilon}[\nu](y) = \int_M \varphi_{\alpha,\varepsilon}(d(x,y))\,\mathrm{d}\nu(x). \tag{3.2.17}$$

It will be convenient to have a notation for a measure that has density with respect to Leb,

which is equal to $\rho_{\alpha,\varepsilon}[\nu]$:

$$\Theta_{\alpha,\varepsilon}[\nu] = \rho^2_{\alpha,\varepsilon}[\nu](y)\,\mathrm{dLeb}(y).$$

Next, we define $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}$ and a normalized version of the measure $\Theta_{\alpha,\varepsilon}$:

Definition 3.2.12. For $\nu \in \mathcal{M}$ define:

$$\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) = \Theta_{\alpha,\varepsilon}[\nu](M); \tag{3.2.18}$$

and

$$\theta_{\alpha,\varepsilon}[\nu] = \frac{\Theta_{\alpha,\varepsilon}[\nu]}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)}.$$

In order to compare $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)$ to $\mathcal{E}_{\alpha,\varepsilon}(\nu)$ we show that the former can be represented as an integral of a certain kernel over $\nu \times \nu$. Later we will show that this kernel is close to $U_{\alpha,\varepsilon}(d(x,z))$.

Lemma 3.2.13.

$$\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) = \iint_{M\times M} K_{\alpha,\varepsilon}(x,z)\; d\nu(x)\; d\nu(z),$$

where

$$K_{\alpha,\varepsilon}(x,z) := \int_M \varphi_{\alpha,\varepsilon}(d(x,y))\varphi_{\alpha,\varepsilon}(d(z,y))\;\; d\mathrm{Leb}(y). \tag{3.2.19}$$

Proof. It suffices to substitute (3.2.17) that defines $\rho^2_{\alpha,\varepsilon}[\nu](y)$, into (3.2.18), obtaining a triple

integral

$$\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) = \iiint_{M\times M\times M} \varphi_{\alpha,\varepsilon}(d(x,y))\varphi_{\alpha,\varepsilon}(d(z,y))\ \mathrm{d}\nu(x)\,\mathrm{d}\nu(z)\ \mathrm{dLeb}(y),$$

and then change the order of integration. □

The following Proposition formalizes the statement that for singular enough ν the values $\mathcal{E}_{\alpha,\varepsilon}(\nu)$ and $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)$ are close to each other.

Proposition 3.2.14. *For every $\alpha > 0$ and every $\delta > 0$ there exists $C > 0$ such that for every $\varepsilon > 0$ if $\mathcal{E}_{\alpha,\varepsilon}(\nu) > C$ or $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) > C$ then one has*

$$\frac{\mathcal{E}_{\alpha,\varepsilon}(\nu)}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)} \in (1-\delta, 1+\delta).$$

In what follows it will be convenient to use the following notation:

Definition 3.2.15. We say that two positive numbers, A and A', are (δ, C)-*close*, if

$$A < (1+\delta)A' + C \quad \text{and} \quad A' < (1+\delta)A + C;$$

this can be equivalently rewritten as

$$\frac{1}{1+\delta}A - \frac{1}{1+\delta}C < A' < (1+\delta)A + C.$$

We denote it $A \approx_{(\delta,C)} A'$.

Remark 3.2.16. We can reformulate Proposition 3.2.14 using notations introduced above: for every $\alpha > 0$ and every $\delta > 0$ there exists $C < \infty$, such that for every $\varepsilon > 0$ and every $\nu \in \mathcal{M}$ one has

$$\mathcal{E}_{\alpha,\varepsilon}(\nu) \approx_{(\delta,C)} \widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu).$$

Joining Proposition 3.2.14 and Corollary 3.2.9 we get

Corollary 3.2.17. *For every $\delta > 0$ there exists $\varepsilon_0 > 0$, such that for every $R < \infty$ there exists $\alpha_0 > 0$, such that for every $0 < \alpha < \alpha_0$ there exists $C > 0$ such that for every $f \in \mathrm{Hol}(M)$ with $L(f) < R$ and $\gamma(f)^{-1} < R$, every $0 < \varepsilon < \varepsilon_0$ and every ν, such that $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) > C$ or $\mathcal{E}_{\alpha,\varepsilon}(\nu) > C$ one has*

$$\frac{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(f_*\nu)}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)} \in (1-\delta, 1+\delta). \tag{3.2.20}$$

To prove Proposition 3.2.14 we will need the following:

Lemma 3.2.18. *For every $\alpha > 0$, the interaction potentials $K_{\alpha,\varepsilon}(x,z)$ and $U_{\alpha,\varepsilon}(d(x,z))$ are comparable in the following sense:*

$$\forall \delta > 0\, \exists C: \quad \forall \varepsilon > 0,\, \forall x, z \in M$$

$$K_{\alpha,\varepsilon}(x,z) \approx_{(\delta,C)} U_{\alpha,\varepsilon}(d(x,z)),\ \textit{i.e.}$$

$$\frac{1}{1+\delta}(U_{\alpha,\varepsilon}(d(x,z)) - C) < K_{\alpha,\varepsilon}(x,z) < (1+\delta)U_{\alpha,\varepsilon}(d(x,z)) + C. \tag{3.2.21}$$

Let us postpone the proof of Lemma 3.2.18 for a moment and deduce Proposition 3.2.14.

Proof of Proposition 3.2.14. It suffices to integrate (3.2.21) w.r.t. $\nu \times \nu\,(x,z)$. As the constant C does not depend on these points nor on the measure ν, one gets

$$\frac{1}{1+\delta}(\mathcal{E}_{\alpha,\varepsilon}(\nu) - C) < \widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) < (1+\delta)\mathcal{E}_{\alpha,\varepsilon}(\nu) + C$$

with the same constant C. As it was noticed in Remark 3.2.16, this is an equivalent form of Proposition 3.2.14. □

Proof of Lemma 3.2.18. Let $\alpha > 0$ be given. Take any $r_0 > 0$ and divide the integral (3.2.19)

defining $K_{\alpha,\varepsilon}(x,z)$ into two parts, depending on whether the distance $d(x,y)$ exceeds r_0:

$$K_{\alpha,\varepsilon}(x,z) = \int_{B_{r_0}(x)} \varphi_{\alpha,\varepsilon}(d(x,y))\varphi_{\alpha,\varepsilon}(d(z,y))\ \mathrm{dLeb}(y) + \\ + \int_{M\setminus B_{r_0}(x)} \varphi_{\alpha,\varepsilon}(d(x,y))\varphi_{\alpha,\varepsilon}(d(z,y))\ \mathrm{dLeb}(y).$$

Denote the first and the second summands as $K^{(r_0)}_{\alpha,\varepsilon}(x,z)$ and $\mathbf{K}^{(r_0)}_{\alpha,\varepsilon}(x,z)$. Note that the second summand is uniformly bounded: indeed, the factor $\varphi_{\alpha,\varepsilon}(d(x,y))$ doesn't exceed a constant $\varphi_{\alpha,\varepsilon}(r_0) \leq \varphi_\alpha(r_0)$, while the second factor $\varphi_{\alpha,\varepsilon}(d(y,z))$ is a function with the integral on M that is bounded uniformly in $z \in M$. The latter uniform bound can be seen by again decomposing the integral in two:

$$\int_M \varphi_{\alpha,\varepsilon}(d(y,z))\ \mathrm{dLeb}(y) = \int_{B_{r_0}(z)} \varphi_{\alpha,\varepsilon}(d(y,z))\ \mathrm{dLeb}(y) + \\ + \int_{M\setminus B_{r_0}(z)} \varphi_{\alpha,\varepsilon}(d(y,z))\ \mathrm{dLeb}(y). \quad (3.2.22)$$

The second summand in (3.2.22) does not exceed $\mathrm{vol}(M)\cdot\varphi_\alpha(r_0)$.The first one can be estimated uniformly in $z \in M$ by passage to the geodesic coordinates centred at z, comparing it to the same integral in a ball in $\mathbb{R}^k$: the Jacobian of the change of variables is (for r_0 smaller than the injectivity radius in M) uniformly bounded, and the integral of $\frac{(-\log(r))^{\frac{\alpha-1}{2}}}{r^{\frac{k}{2}}}$ on $B_{r_0}(0) \subset \mathbb{R}^k$ converges.

We thus have

$$\mathbf{K}^{(r_0)}_{\alpha,\varepsilon}(x,z) \leq C^K_\alpha(r_0) \quad (3.2.23)$$

for some constant $C^K_\alpha(r_0)$.

Now, let us transform the first integral, $K^{(r_0)}_{\alpha,\varepsilon}(x,z)$. For sufficiently small $r_0 > 0$ (smaller than the injectivity radius at every point) one can take the geodesic coordinates at any point

$x \in M$ in a ball of radius r_0. Thus, we can take a point $x' \in \mathbb{R}^k$, let $\psi : B_{r_0}(x') \to B_{r_0}(x)$ be geodesic coordinates, and denote $z' := \psi^{-1}(z)$.

Fix any $\delta_1 > 0$. Due to the compactness of M, for a sufficiently small r_0 the Jacobian and the bi-Lipschitz constants of ψ are δ_1-close to 1:

$$\mathfrak{L}(\psi) < 1 + \delta_1, \quad \operatorname{Jac}(\psi) \in \left(\frac{1}{1+\delta_1}, 1 + \delta_1 \right).$$

Making a change of variables $y = \psi(y')$ in the integral for $K^{(r_0)}_{\alpha,\varepsilon}(x, z)$, we get

$$K^{(r_0)}_{\alpha,\varepsilon}(x,z) = \int_{B_{r_0}(x)} \varphi_{\alpha,\varepsilon}(d(x,y))\varphi_{\alpha,\varepsilon}(d(z,y)) \operatorname{dLeb}_M(y) = \\ = \int_{B_{r_0}(x')} \varphi_{\alpha,\varepsilon}(d(\psi(x'),\psi(y')))\varphi_{\alpha,\varepsilon}(d(\psi(z'),\psi(y'))) \operatorname{dLeb}_M(\psi(y'));$$

all the three quotients

$$\frac{\varphi_{\alpha,\varepsilon}(d(\psi(x'),\psi(y')))}{\varphi_{\alpha,\varepsilon}(|x'-y'|)}, \quad \frac{\varphi_{\alpha,\varepsilon}(d(\psi(z'),\psi(y')))}{\varphi_{\alpha,\varepsilon}(|z'-y'|)}, \quad \left.\frac{\mathrm{d}\operatorname{Leb}_M}{\mathrm{d}\,\psi_*\operatorname{Leb}_{\mathbb{R}^k}}\right|_y$$

are close to 1: the first two by a factor $(1+\delta_1)^{\frac{k+\alpha}{2}}$ since

$$1 \le \frac{\varphi_{\alpha,\varepsilon}(r_1)}{\varphi_{\alpha,\varepsilon}(r_2)} \le \left(\frac{\log(r_1)}{\log(r_2)}\right)^{\frac{\alpha-1}{2}} \left(\frac{r_2}{r_1}\right)^{\frac{k}{2}} < \left(\frac{r_2}{r_1}\right)^{\frac{k+\alpha}{2}} \quad \text{for any } 0 < r_1 \le r_2 < \frac{1}{e},$$

and the last one by the factor $(1+\delta_1)$. Thus, $K^{(r_0)}_{\alpha,\varepsilon}(x,z)$ differs from

$$I^{(r_0)}_{\alpha,\varepsilon}(x',z') := \int_{B_{r_0}(x')} \varphi_{\alpha,\varepsilon}(|x'-y'|)\varphi_{\alpha,\varepsilon}(|z'-y'|) \operatorname{dLeb}_{\mathbb{R}^k}(y') \tag{3.2.24}$$

by the factor at most $(1+\delta_1)^{k+\alpha+1}$. Now, $I^{(r_0)}_{\alpha,\varepsilon}(x',z')$ is also a part of $I_{\alpha,\varepsilon}(x',z')$ (define by

(3.2.8)), that differs from it by

$$\overline{I}^{(r_0)}_{\alpha,\varepsilon}(x',z') := \int_{\mathbb{R}^k\setminus B_{r_0}(x')} \varphi_{\alpha,\varepsilon}(|x'-y'|)\varphi_{\alpha,\varepsilon}(|z'-y'|)\ \mathrm{dLeb}_{\mathbb{R}^k}(y'),$$

that is bounded uniformly in ε, x', z' (for $|x'-z'| < \frac{r_0}{2}$ both functions outside of $B_{r_0}(x')$ are bounded and have compact supports, for $|x'-z'| \geq \frac{r_0}{2}$ we have $\overline{I}^{(r_0)}_{\alpha,\varepsilon}(x',z') \leq U_{\alpha,\varepsilon}\left(\frac{r_0}{2}\right)$). Thus, first choosing δ_1 so that $(1+\delta_1)^{k+\alpha+1} < 1+\delta$ and then accordingly choosing r_0, we obtain the desired estimate (3.2.21).

□

3.2.6 Properties of $\theta_{\alpha,\varepsilon}(\nu)$ and $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)$ under Hölder homeomorphisms

First, in this section we show that computing the measure $\theta_{\alpha,\varepsilon}$ "almost commutes" with a Hölder continuous homeomorphisms, i.e. $\theta_{\alpha,\varepsilon}(f_*\nu) \simeq f_*\theta_{\alpha,\varepsilon}(\nu)$ (see Proposition 3.2.22). Second, we show that the energy $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)$ doesn't change much on average if we apply a random Hölder homeomorphism (see Proposition 3.2.25). Both facts are used to prove the contraction described in part (iii) of the outline of the proof.

In what follows it will be convenient to use Wasserstein metric in the space of probability measures on M. Let us recall its definition, as well as definition of the total variation distance between measures:

Definition 3.2.19. Let ν_1, ν_2 be two probability measures on a measure space $(M, \mathcal{B})$. Then the *Wasserstein distance* between them is defined as

$$W(\nu_1,\nu_2) = \inf_{\gamma} \iint_{M\times M} d(x,y)\ \mathrm{d}\gamma(x,y),$$

where the infimum is taken over all probability measures γ on $(M\times M, \mathcal{B}\times\mathcal{B})$ with the

marginals (projections on the x and y coordinates) $P_x(\gamma) = \nu_1$ and $P_y(\gamma) = \nu_2$.

Definition 3.2.20. Let ν_1, ν_2 be two probability measures on a measure space $(M, \mathcal{B})$. Then the *total variation* distance between them is

$$\mathrm{TV}(\nu_1, \nu_2) = \sup_{B \in \mathcal{B}} |\nu_1(B) - \nu_2(B)|.$$

Also, we will need the following statement that can be found, for example, in [40, Theorem 6.15]: the Wasserstein metric is bounded from above by the diameter times the total variation distance.

Lemma 3.2.21. *For every two probability measures ν_1, ν_2 on a manifold M we have*

$$W(\nu_1, \nu_2) \leq \mathrm{diam}(M) \cdot \mathrm{TV}(\nu_1, \nu_2).$$

Equipped with these notions we would like to prove the following

Proposition 3.2.22. *For every $\delta > 0$ and every $R < \infty$ there exists $\alpha_0 > 0$, such that for every $0 < \alpha < \alpha_0$ there exists $C > 0$, such that for every $\varepsilon > 0$, every $f \in \mathrm{Hol}(M)$, such that $\gamma(f)^{-1}, L(f) < R$ and every $\nu \in \mathcal{M}$ such that $\mathcal{E}_{\alpha,\varepsilon}(\nu) > C$ or $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) > C$ one has*

$$W(f_* \theta_{\alpha,\varepsilon}[\nu], \theta_{\alpha,\varepsilon}[f_* \nu]) < \delta. \tag{3.2.25}$$

Proof. For given $\alpha, \varepsilon > 0$ and measure ν, consider a (non-probability) measure $m_{\alpha,\varepsilon}(\nu)$ on $M \times M \times M$, given by

$$m_{\alpha,\varepsilon}(\nu) = \varphi_{\alpha,\varepsilon}(d(x,y)) \varphi_{\alpha,\varepsilon}(d(z,y)) \, \mathrm{d}\nu(x) \, \mathrm{dLeb}(y) \, \mathrm{d}\nu(z).$$

Denote by π_1, π_2 the projections of $M \times M \times M$ on the first and second coordinates respectively, and by $\pi_{1,3}$ the projection on $M \times M$ corresponding to the first and third coordinates.

Then directly from definition

$$(\pi_2)_* m_{\alpha,\varepsilon}(\nu) = \Theta_{\alpha,\varepsilon}(\nu);$$

also, define

$$\widehat{\Theta}_{\alpha,\varepsilon}(\nu) := (\pi_{1,3})_* m_{\alpha,\varepsilon}(\nu) = K_{\alpha,\varepsilon}(x,z)\,\nu(dx)\,\nu(dz); \tag{3.2.26}$$

the second equality is due to (3.2.19). Finally, consider the measure

$$\Theta'_{\alpha,\varepsilon}(\nu) := (\pi_1)_* m_{\alpha,\varepsilon}(\nu) \tag{3.2.27}$$

as well as the normalizations of these measures,

$$\widehat{\theta}_{\alpha,\varepsilon}(\nu) := \frac{1}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)}\widehat{\Theta}_{\alpha,\varepsilon}(\nu), \quad \theta'_{\alpha,\varepsilon}(\nu) := \frac{1}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)}\Theta'_{\alpha,\varepsilon}(\nu). \tag{3.2.28}$$

For a high-energy measure ν most of the measure $m_{\alpha,\varepsilon}(\nu)$ is concentrated near the diagonal, and hence the projections on the first and on the second coordinates are close to each other. The following lemma formalizes this argument:

Lemma 3.2.23. *For every δ_1, α there exists C' such that for every $\varepsilon > 0$ and ν with $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) > C'$ one has*

$$W(\theta_{\alpha,\varepsilon}(\nu), \theta'_{\alpha,\varepsilon}(\nu)) < \delta_1.$$

Proof. Take $r_0 := \frac{\delta_1}{2}$ and let

$$A_{r_0} := \{(x,y,z) \in M^3 \mid d(x,y) < r_0\}, \quad \overline{A}_{r_0} := M^3 \setminus A_{r_0}.$$

From the proof of Lemma 3.2.18, we have

$$m_{\alpha,\varepsilon}(\nu)(\overline{A}_{r_0}) = \int \mathbf{K}^{(r_0)}_{\alpha,\varepsilon}(x,z)\, d\nu(x)\, d\nu(z) \le C^K_\alpha(r_0),$$

where the second inequality is due to (3.2.23). Thus, the non-normalized measure $m_{\alpha,\varepsilon}(\nu)(\overline{A}_{r_0})$ does not exceed a constant $C^K_\alpha(r_0)$.

Now, we can couple the normalized measures $\theta_{\alpha,\varepsilon}(\nu)$ and $\theta'_{\alpha,\varepsilon}(\nu)$ using the projection of $\frac{1}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)} m_{\alpha,\varepsilon}(\nu)$ on the first two coordinates; this coupling leads to the upper bound for the Wasserstein distance:

$$W(\theta_{\alpha,\varepsilon}(\nu), \theta'_{\alpha,\varepsilon}(\nu)) \le r_0 \cdot \frac{m_{\alpha,\varepsilon}(\nu)(A_{r_0})}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)} + \operatorname{diam}(M) \cdot \frac{m_{\alpha,\varepsilon}(\nu)(\overline{A}_{r_0})}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)} \le r_0 + \frac{C^K_\alpha(r_0)}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)}, \quad (3.2.29)$$

where the first summand corresponds to the points with $d(x,y) < r_0$, and the second to the points with $d(x,y) \ge r_0$. As we chose $r_0 = \frac{\delta_1}{2}$, it suffices to require that

$$\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) > \frac{2\, C^K_\alpha(r_0)}{\delta_1} =: C'$$

to ensure that the total Wasserstein distance does not exceed δ_1. □

On the other hand, for the measures $\theta'_{\alpha,\varepsilon}(\nu)$ the analogue of Proposition 3.2.22 can be established directly, and one can even estimate the total variations distance:

Lemma 3.2.24. *For every $\delta_2 > 0$ and every $R > 0$ there exists $\alpha_2 > 0$ such that for every $0 < \alpha < \alpha_2$ there exists $C > 0$ such that for every $\varepsilon > 0$, every $f \in \operatorname{Hol}(M)$ with $\gamma(f)^{-1}, L(f) < R$ and every ν such that $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) > C$ or $\mathcal{E}_{\alpha,\varepsilon}(\nu) > C$ one has*

$$\operatorname{TV}(f_*\theta'_{\alpha,\varepsilon}(\nu), \theta'_{\alpha,\varepsilon}(f_*\nu)) < \delta_2, \quad (3.2.30)$$

and, hence,

$$W(f_*\theta'_{\alpha,\varepsilon}(\nu), \theta'_{\alpha,\varepsilon}(f_*\nu)) < \delta_2 \cdot \mathrm{diam}(M). \tag{3.2.31}$$

Proof. We will first use Lemma 3.2.18 together with Part **(IV)** of Proposition 3.2.4 to compare $K_{\alpha,\varepsilon}(x,z)$ with $K_{\alpha,\varepsilon}(f(x),f(z))$. Fix α_0 and δ_3 so small that

$$(1+\delta_3)^2 \cdot [2R(1+\log(R))]^{\alpha_0} < 1 + \frac{\delta_2}{10}.$$

Then from Lemma 3.2.18 we know that there exists $C > 0$ such that for any x, z, ε we have

$$K_{\alpha,\varepsilon}(x,z) \approx_{(\delta_3,C)} U_{\alpha,\varepsilon}(d(x,z)), \quad K_{\alpha,\varepsilon}(f(x),f(z)) \approx_{(\delta_3,C)} U_{\alpha,\varepsilon}(d(f(x),f(z))). \tag{3.2.32}$$

Without loss of generality we can assume that $d(x,z) < d(f(x),f(z))$. Then by Part **(IV)** of Proposition 3.2.4 we can choose a particular $r_0 = r_0(\alpha_0,\delta_3)$, such that if $d(f(x),f(z)) < r_0$ then

$$\frac{U_{\alpha,\varepsilon}(d(x,z))}{U_{\alpha,\varepsilon}(d(f(x),f(z)))} < (1+\delta_3)\left(\frac{\log(d(x,z))}{\log(d(f(x),f(z)))}\right)^{\alpha}. \tag{3.2.33}$$

Notice that inequality (3.2.33) only holds for $\varepsilon < \varepsilon_0(\alpha_0,\delta_3)$. That can be guaranteed by taking C big enough, because, according to Lemma 3.2.6, the energy $\mathcal{E}_{\varepsilon,\alpha}(\nu)$ for $\varepsilon \geq \varepsilon_0$ has an explicit upper bound.

Using the fact that $L(f), \gamma(f)^{-1} < R$ we conclude that

$$\left(\frac{\log(d(x,z))}{\log(d(f(x),f(z)))}\right)^{\alpha} \leq$$
$$\leq \left(\frac{\frac{2}{\gamma(f)}\log\left[d(f(x),f(z))L(f)^{-1}\right]}{\log(d(f(x),f(z)))}\right)^{\alpha} <$$
$$< [2R(1+\log(R))]^{\alpha_0}$$

and

$$U_{\alpha,\varepsilon}(d(x,z)) < (1+\delta_3)\left[2R(1+\log(R))\right]^{\alpha_0} U_{\alpha,\varepsilon}(d(f(x),f(z))). \tag{3.2.34}$$

On the other hand, if $d(f(x),f(z)) > r_0$ then

$$d(x,z) \geq \left(\frac{r_0}{L(f)}\right)^{\frac{2}{\gamma(f)}} > \left(\frac{r_0}{R}\right)^{\frac{2}{R}}$$

and hence

$$U_{\alpha,\varepsilon}(d(x,z)) < U_{\alpha_0}\left(\left(\frac{r_0}{R}\right)^{\frac{2}{R}}\right). \tag{3.2.35}$$

Combining estimates (3.2.32), (3.2.34) and (3.2.35) we conclude that

$$K_{\alpha,\varepsilon}(x,z) \approx_{(\frac{\delta_2}{10},C'')} K_{\alpha,\varepsilon}(f(x),f(z)) \tag{3.2.36}$$

for some explicit constant C''.

Now, applying f_*^{-1} does not change the total variations distance, so instead of (3.2.30) we can show the equivalent statement

$$\mathrm{TV}(\theta'_{\alpha,\varepsilon}(\nu), f_*^{-1}\theta'_{\alpha,\varepsilon}(f_*\nu)) < \delta_2. \tag{3.2.37}$$

The measures here can be obtained as projections of measures on $M \times M$:

$$\theta'_{\alpha,\varepsilon}(\nu) = (\pi_1)_* \widehat{\theta}_{\alpha,\varepsilon}(\nu),$$

$$f_*^{-1}\theta'_{\alpha,\varepsilon}(f_*\nu) = (\pi_1)_* \left(f_*^{-1}\widehat{\theta}_{\alpha,\varepsilon}(f_*\nu)\right)$$

where π_1 is the projection on the first coordinate. To obtain the desired (3.2.37), we will

actually show that for sufficiently high energy $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)$ even before projection one has

$$\mathrm{TV}(\widehat{\theta}_{\alpha,\varepsilon}(\nu), f_*^{-1}\widehat{\theta}_{\alpha,\varepsilon}(f_*\nu)) < \delta_2. \tag{3.2.38}$$

To do so, note that both these measures are absolutely continuous with respect to $\nu \times \nu$:

$$\widehat{\theta}_{\alpha,\varepsilon}(\nu) = \frac{1}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)} K_{\alpha,\varepsilon}(x,z)\,\nu(dx)\,\nu(dz),$$

$$f_*^{-1}\widehat{\theta}_{\alpha,\varepsilon}(f_*\nu) = \frac{1}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(f_*\nu)} K_{\alpha,\varepsilon}(f(x),f(z))\,\nu(dx)\,\nu(dz).$$

Now, (3.2.36) implies that once $K_{\alpha,\varepsilon}(x,z) > C_2 := \frac{20C''}{\delta_2}$, the $K_{\alpha,\varepsilon}$ parts of these densities are close to each other:

$$\frac{1}{1+\frac{\delta_2}{6}} < \frac{K_{\alpha,\varepsilon}(f(x),f(z))}{K_{\alpha,\varepsilon}(x,z)} < 1 + \frac{\delta_2}{6}.$$

On the other hand, due to Corollary 3.2.17, once the energy $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)$ is sufficiently large, the normalization constants are also close to each other:

$$\frac{1}{1+\frac{\delta_2}{6}}\,\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) < \widetilde{\mathcal{E}}_{\alpha,\varepsilon}(f_*\nu) < (1+\frac{\delta_2}{6})\,\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu).$$

Multiplying the two inequalities, we get that on the set

$$A := \{(x,z) \mid K_{\alpha,\varepsilon}(x,z) > C_2\}$$

the quotient of densities w.r.t. $\nu \times \nu$ of the two measures is in the interval

$$\left(\frac{1}{\left(1+\frac{\delta_2}{6}\right)^2}, \left(1+\frac{\delta_2}{6}\right)^2\right).$$

Finally, the $\widehat{\theta}_{\alpha,\varepsilon}(\nu)$-measure of its complement does not exceed $\frac{C_2}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)}$. Hence, if the energy $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)$ is sufficiently high to make sure that $\frac{C_2}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)} < \frac{\delta_2}{6}$, the part that the normalized measures have in common is at least

$$\frac{1-\frac{\delta_2}{6}}{\left(1+\frac{\delta_2}{6}\right)^2} > 1-\frac{\delta_2}{2},$$

and hence the total variation (3.2.38) indeed does not exceed δ_2.

An application of Lemma 3.2.21 concludes the proof of the upper bound (3.2.31) for the Wasserstein distance.

□

Lemma 3.2.23 and Lemma 3.2.24 together imply Proposition 3.2.22. Indeed, we have:

$$W(f_*\theta_{\alpha,\varepsilon}(\nu), \theta_{\alpha,\varepsilon}(f_*\nu)) \leq W(f_*\theta_{\alpha,\varepsilon}(\nu), f_*\theta'_{\alpha,\varepsilon}(\nu)) + \\ + W(f_*\theta'_{\alpha,\varepsilon}(\nu), \theta'_{\alpha,\varepsilon}(f_*\nu)) + W(\theta'_{\alpha,\varepsilon}(f_*\nu), \theta_{\alpha,\varepsilon}(f_*\nu))$$

The first and the third summands can be estimated directly using Lemma 3.2.23. It suffices to note that since $L(f), \gamma(f)^{-1} < R$ we have control over the increase of the Wasserstein distance after application of f. Indeed, a simple application of Jensen's inequality gives us

$$W(f_*\theta_{\alpha,\varepsilon}(\nu), f_*\theta'_{\alpha,\varepsilon}(\nu)) < R \cdot W(\theta_{\alpha,\varepsilon}(\nu), \theta'_{\alpha,\varepsilon}(\nu))^{\frac{2}{R}}.$$

For the last summand, to apply Lemma, we note that an upper bound on $L(f)$ and $\gamma(f)^{-1}$ implies that $f_*\nu$ is of sufficiently high energy provided that ν is of high energy.

Finally, the second summand is estimated directly by Lemma 3.2.24. □

Proposition 3.2.25. *If a probability measure μ on* $\mathrm{Hol}(M)$ *satisfies assumption (3.1.1) then*

for every $\delta > 0$ there exists $\alpha_0 > 0$ such that for every $0 < \alpha < \alpha_0$ there exists $C > 0$ such that for every $\varepsilon > 0$ and every measure $\nu \in \mathcal{M}$ such that $\mathcal{E}_{\alpha,\varepsilon}(\nu) > C$ the following formula holds:

$$\frac{\mathbb{E}_\mu\left[\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(f_*\nu)\right]}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)} \in (1-\delta, 1+\delta). \tag{3.2.39}$$

Proof. Let us fix some $\delta_E > 0$ and pick $\tilde{\delta} > 0$, such that $(1+\tilde{\delta})^2 < 1+\delta_E$. It follows directly from condition (3.1.1) and Hölder inequality that there exists $\alpha_0 > 0$, such that for every $0 < \alpha < \alpha_0$ one has

$$\int_{\mathrm{Hol}(M)} \Lambda(f)^\alpha \ \mathrm{d}\mu(f) < 1+\tilde{\delta}. \tag{3.2.40}$$

According to Corollary 3.2.9 we can choose C such that inequality (3.2.15) holds for $\tilde{\delta}$. Let $f \in \mathrm{Hol}(M)$ be a random homeomorphism distributed with respect to the measure μ. Then integrating inequality (3.2.15) with respect to the measure μ and combining it with inequality (3.2.40) we get:

$$\frac{1}{1+\delta_E} < \frac{1}{(1+\tilde{\delta})^2} \leq \frac{\mathbb{E}_\mu\left[\mathcal{E}_{\alpha,\varepsilon}(f_*\nu)\right]}{\mathcal{E}_{\alpha,\varepsilon}(\nu)} \leq (1+\tilde{\delta})^2 < 1+\delta_E. \tag{3.2.41}$$

Let us now return back to the original (3.2.39). Namely, let $\delta > 0$ be given; choose and fix $\delta_E > 0$ such that $(1+\delta_E)^3 < 1+\delta$, and let α_0 be chosen w.r.t. δ_E so that (3.2.41) holds.

By Proposition 3.2.14, there exists $C_E > 0$ such that for any measure ν' on M and any $\varepsilon > 0$ one has

$$\frac{1}{1+\delta_E}\,\mathcal{E}_{\alpha,\varepsilon}(\nu') - \frac{C_E}{1+\delta_E} < \widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu') < (1+\delta_E)\mathcal{E}_{\alpha,\varepsilon}(\nu') + C_E \tag{3.2.42}$$

Applying this for $\nu' = f_*\nu$ and taking the expectation w.r.t. μ provides

$$\frac{1}{1+\delta_E}\,\mathbb{E}_\mu\left[\mathcal{E}_{\alpha,\varepsilon}(f_*\nu)\right] - \frac{C_E}{1+\delta_E} < \mathbb{E}_\mu\left[\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(f_*\nu)\right] < (1+\delta_E)\mathbb{E}_\mu\left[\mathcal{E}_{\alpha,\varepsilon}(f_*\nu)\right] + C_E;$$

using (3.2.41), we thus get

$$\frac{1}{(1+\delta_E)^2}\,\mathcal{E}_{\alpha,\varepsilon}(\nu) - \frac{C_E}{1+\delta_E} < \mathbb{E}_\mu\left[\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(f_*\nu)\right] < (1+\delta_E)^2\mathcal{E}_{\alpha,\varepsilon}(\nu) + C_E.$$

Finally, using (3.2.42) with $\nu' = \nu$ to estimate $\mathcal{E}_{\alpha,\varepsilon}$ via $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}$, we get

$$\frac{1}{(1+\delta_E)^2}\cdot\frac{1}{(1+\delta_E)}\left(\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) - C_E\right) - \frac{C_E}{1+\delta_E} < \mathbb{E}_\mu\left[\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(f_*\nu)\right] < (1+\delta_E)^2\cdot\left((1+\delta_E)\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) + C_E\right) + C_E;$$

as $(1+\delta_E)^3 < 1+\delta$, we obtain the desired

$$\frac{1}{1+\delta}(\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) - C') < \mathbb{E}_\mu\left[\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(f_*\nu)\right] < (1+\delta)\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) + C'$$

for some constant C'. As $\delta > 0$ was arbitrary (3.2.39) follows.

□

Remark 3.2.26. Notice that if instead of one measure μ on $\mathrm{Hol}(M)$ we consider a compact K of such measures, such that there exists $\beta > 0$ and C_0 such that for every $\mu \in K$

$$\int_{\mathrm{Hol}(M)} \Lambda(f)^\beta \ \mathrm{d}\mu(f) < C_0$$

then constants α_0 and C in Proposition 3.2.25 can be chosen uniformly for all measures in K.

3.2.7 Contraction of $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}$ under random dynamics

In this section we prove a proposition that formalizes (3.2.5). Namely, we show that if $\mathcal{E}_{\alpha,\varepsilon}(\nu)$ is big enough then $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\mu * \nu) < \lambda \widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)$ for some $\lambda < 1$. This will allow us to control the energy after n iterations of random dynamics and finish the proof of Theorem 3.1.4.

Proposition 3.2.27. *Under the assumptions of Theorem 3.1.4 there exists $\alpha > 0$, $\lambda < 1$, and $C < \infty$, such that for every $\varepsilon > 0$, every $\nu \in \mathcal{M}$ and every $\mu \in K$ if $\mathcal{E}_{\alpha,\varepsilon}(\nu) > C$ then*

$$\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\mu * \nu) < \lambda \widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu). \tag{3.2.43}$$

Proof. Fix some small number $\delta > 0$, take $\lambda = 1 - \delta$, and assume the contrary: for any $\alpha > 0$ and $C < \infty$ there exists a measure $\mu \in K$, an $\varepsilon > 0$ and a probability measure ν with energy $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) > C$, such that

$$\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\mu * \nu) \geq (1 - \delta) \widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu). \tag{3.2.44}$$

Recall the following standard statement:

Lemma 3.2.28. *Assume that in some Hilbert space $\mathcal{H}$ a probability measure is given; in other words, one is given a random variable v taking values in this space. Assume that this measure has finite second moment, and let $\overline{v} := \mathbb{E}v$ be its expectation. Then*

$$\mathbb{E}\langle v - \overline{v}, v - \overline{v} \rangle = \mathbb{E}\langle v, v \rangle - \langle \overline{v}, \overline{v} \rangle \tag{3.2.45}$$

Let us apply this statement to $L^2(M, \mathrm{Leb})$. Namely, for a given measure ν its averaged image $\mu * \nu$ is the expectation of $f_*\nu$, where the homeomorphism f of M is taken randomly w.r.t. the measure μ. Hence, the same applies for the density $\rho_{\alpha,\varepsilon}[\nu]$, given by (3.2.17) (see

Definition 3.2.11):

$$\rho_{\alpha,\varepsilon}[\mu * \nu] = \mathbb{E}_\mu \, \rho_{\alpha,\varepsilon}[f_*\nu]. \tag{3.2.46}$$

Substituting this into (3.2.45) (with $v = \rho_{\alpha,\varepsilon}[f_*\nu]$), and taking into account the definition (3.2.18), we get

$$\mathbb{E}_\mu \left[\int_M \left(\rho_{\alpha,\varepsilon}[f_*\nu](y) - \rho_{\alpha,\varepsilon}[\mu * \nu](y)\right)^2 \, \mathrm{dLeb}(y) \right] = \\ = \mathbb{E}_\mu \left[\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(f_*\nu) \right] - \widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\mu * \nu); \tag{3.2.47}$$

in particular,

$$\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\mu * \nu) \le \mathbb{E}_\mu \left[\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(f_*\nu) \right]. \tag{3.2.48}$$

By Proposition 3.2.25 we can find sufficiently small $\alpha > 0$ and sufficiently large $C < \infty$ (uniformly in μ), such that for any $\varepsilon > 0$ and any measure ν on M with $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) > C$ the inequality (3.2.39) holds, and thus (taking only the upper bound)

$$\mathbb{E}_\mu \left[\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(f_*\nu) \right] \le (1+\delta) \widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu). \tag{3.2.49}$$

Denote by ν a probability measure with $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) > C$, for which the inequality (3.2.44) holds. Substituting (3.2.49) and (3.2.44) in the right hand side of (3.2.47), we get

$$\mathbb{E}_\mu \left[\int_M \left(\rho_{\alpha,\varepsilon}(f_*\nu)(y) - \rho_{\alpha,\varepsilon}(\mu * \nu)(y)\right)^2 \, d\mathrm{Leb}(y) \right] < 2\delta \, \widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu). \tag{3.2.50}$$

and thus

$$\mathbb{E}_\mu\left[\int_M \frac{\left(\rho_{\alpha,\varepsilon}(f_*\nu)(y)-\rho_{\alpha,\varepsilon}(\mu*\nu)(y)\right)^2}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)}d\mathrm{Leb}(y)\right]<2\delta. \tag{3.2.51}$$

The final step is to show that measures

$$\theta_{\alpha,\varepsilon}(f_*\nu)=\frac{\rho_{\alpha,\varepsilon}(f_*\nu)(y)^2}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(f_*\nu)}d\mathrm{Leb}(y) \quad \text{and} \quad \theta_{\alpha,\varepsilon}(\mu*\nu)=\frac{\rho_{\alpha,\varepsilon}(\mu*\nu)(y)^2}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\mu*\nu)}d\mathrm{Leb}(y)$$

are close with high probability. And indeed, we have the following statement, estimating (even) the total variations distance:

Lemma 3.2.29. *Under the assumptions above*

$$\mathbb{P}_\mu\left[\mathrm{TV}(\theta_{\alpha,\varepsilon}(f_*\nu),\theta_{\alpha,\varepsilon}(\mu*\nu))>10\sqrt[8]{\delta}\right]<4\sqrt{\delta}, \tag{3.2.52}$$

and hence

$$\mathbb{P}_\mu\left[W(\theta_{\alpha,\varepsilon}(f_*\nu),\theta_{\alpha,\varepsilon}(\mu*\nu))>10\mathrm{diam}(M)\sqrt[8]{\delta}\right]<4\sqrt{\delta}. \tag{3.2.53}$$

Proof of Lemma 3.2.29. We start by working with the non-normalized measures $\Theta_{\alpha,\varepsilon}(f_*\nu)$ and $\Theta_{\alpha,\varepsilon}(\mu*\nu)$. To simplify the notation, denote

$$g_f(y):=\rho_{\alpha,\varepsilon}[f_*\nu](y), \quad \overline{g}(y):=\rho_{\alpha,\varepsilon}[\mu*\nu](y).$$

Therefore, we have

$$\Theta_{\alpha,\varepsilon}[f_*\nu]=(g_f(y))^2d\mathrm{Leb}(y) \quad \text{and} \quad \Theta_{\alpha,\varepsilon}[\mu*\nu]=(\overline{g}(y))^2d\mathrm{Leb}(y).$$

From (3.2.44) we have

$$\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) \leq \frac{1}{1-\delta}\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\mu * \nu).$$

Joining with (3.2.50), we have

$$\mathbb{E}_f\left[\int_M |g_f(y) - \overline{g}(y)|^2 \, d\mathrm{Leb}(y)\right] \leq \frac{2\delta}{1-\delta}\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\mu * \nu). \tag{3.2.54}$$

By Markov inequality, with the probability at least $1 - \frac{2\sqrt{\delta}}{1-\delta}$, one has

$$\begin{aligned}\int_M |g_f(y) - \overline{g}(y)|^2 \, d\mathrm{Leb}(y) &\leq \sqrt{\delta}\,\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\mu * \nu) \\ &= \sqrt{\delta}\,\Theta_{\alpha,\varepsilon}[\mu * \nu](M) = \sqrt{\delta}\int_M \overline{g}^2(y)\, d\mathrm{Leb}(y). \end{aligned} \tag{3.2.55}$$

Now, for any f for which (3.2.55) holds, consider the set

$$X_f := \{y \in M \,:\, |g_f(y) - \overline{g}(y)| \leq \sqrt[8]{\delta} \cdot \overline{g}(y)\}.$$

Again from Markov inequality type argument, one has

$$\Theta_{\alpha,\varepsilon}[\mu * \nu](X_f) \geq \left(1 - \sqrt[4]{\delta}\right)\Theta_{\alpha,\varepsilon}[\mu * \nu](M). \tag{3.2.56}$$

Indeed, on the complement $M \setminus X_f$ one has $|g_f(y) - \overline{g}(y)|^2 > \sqrt[4]{\delta} \cdot \overline{g}^2(y)$; integrating, one gets

$$\int_{M\setminus X_f} |g_f(y) - \overline{g}(y)|^2 \, d\mathrm{Leb}(y) > \sqrt[4]{\delta} \cdot \Theta_{\alpha,\varepsilon}[\mu * \nu](M \setminus X_f).$$

Thus, if (3.2.56) did not hold, it would imply

$$\Theta_{\alpha,\varepsilon}[\mu * \nu](M \setminus X_f) > \sqrt[4]{\delta} \cdot \Theta_{\alpha,\varepsilon}[\mu * \nu](M),$$

hence providing a contradiction with (3.2.55).

Now, (3.2.56) implies that the non-normalized measures $\Theta_{\alpha,\varepsilon}[f_*\nu]$ and $\Theta_{\alpha,\varepsilon}[\mu * \nu]$ share a common part

$$\Xi_f := (1 - \sqrt[8]{\delta})^2\, \overline{g}^2(y) \cdot \mathbb{1}_{X_f}(y)\, d\mathrm{Leb}(y)$$

of measure at least

$$\Xi_f(M) \geq (1 - \sqrt[8]{\delta})^2(1 - \sqrt[4]{\delta}) \cdot \Theta_{\alpha,\varepsilon}[\mu * \nu](M). \tag{3.2.57}$$

Finally, note that the normalization constants are also close with high probability, namely, for f such that the inequality (3.2.55) holds. Indeed, the inequality (3.2.55) can be rewritten as

$$\|g_f - \overline{g}\|^2_{L_2(M)} \leq \sqrt{\delta} \cdot \|\overline{g}\|^2_{L_2(M)};$$

hence, for any f for which it holds, the triangle inequality implies

$$\Theta_{\alpha,\varepsilon}[f_*\nu](M) = \|g_f\|^2_{L_2(M)} \leq (1 + \sqrt[4]{\delta})^2 \Theta_{\alpha,\varepsilon}[\mu * \nu](M). \tag{3.2.58}$$

Now, the normalized measures $\theta_{\alpha,\varepsilon}[f_*\nu]$ and $\theta_{\alpha,\varepsilon}[\mu * \nu]$ share the common part $\frac{1}{C_f}\Xi_f$, where

$$C_f = \max(\Theta_{\alpha,\varepsilon}[f_*\nu](M), \Theta_{\alpha,\varepsilon}[\mu * \nu](M))$$

is the maximum of two normalization constants. Using (3.2.57) and (3.2.58), we see that this part has measure at least

$$\frac{1}{C_f}\Xi_f(M) \geq \frac{(1 - \sqrt[8]{\delta})^2(1 - \sqrt[4]{\delta})}{(1 + \sqrt[4]{\delta})^2} \geq 1 - 10\sqrt[8]{\delta}.$$

Hence, for any f satisfying (3.2.55) the total variation distance between the normalized measures does not exceed

$$\mathrm{TV}(\theta_{\alpha,\varepsilon}(f_*\nu), \theta_{\alpha,\varepsilon}(\mu * \nu)) \leq 10\sqrt[8]{\delta}.$$

As (3.2.55) holds with the probability at least $1 - \frac{2\sqrt{\delta}}{1-\delta} > 1 - 4\sqrt{\delta}$, this establishes conclusion (3.2.52). Finally, this immediately implies conclusion (3.2.53) due to Lemma 3.2.21. $\square$

We are now ready to conclude the proof of Proposition 3.2.27. Namely, the conclusion (3.2.53) states that with high probability the measure $\theta_{\alpha,\varepsilon}(f_*\nu)$ is close to the deterministic one, $\theta_{\alpha,\varepsilon}(\mu * \nu)$. The next step is to show that with high probability the first measure is close to f_*-image of a given measure, $\theta_{\alpha,\varepsilon}(\nu)$.

It easily follows from inequality (3.1.1) that we can choose $R < \infty$ (again, uniformly in μ), such that

$$\mathbb{P}_\mu \left[L(f) < R \text{ and } \gamma(f)^{-1} < R \right] > 1 - \sqrt[4]{\delta}$$

By Proposition 3.2.22 we can choose C big enough, so that for any $f \in \mathrm{Hol}(M)$, such that $L(f), \gamma(f)^{-1} < R$ inequality (3.2.25) holds, and using triangle inequality we arrive to

$$\mathbb{P}_\mu \left[W(f_*\theta_{\alpha,\varepsilon}(\nu), \theta_{\alpha,\varepsilon}(\mu * \nu)) > 10\mathrm{diam}(M)\sqrt[8]{\delta} + \delta \right] < 5\sqrt[4]{\delta}. \tag{3.2.59}$$

Now consider the sequence $\delta_n = \frac{1}{n}$ and denote by μ_n and ν_n the corresponding measures and corresponding parameters by α_n, ε_n, for which (3.2.44) holds. Then the measures $\theta_{\alpha_n,\varepsilon_n}(\nu_n)$ have a weakly convergent subsequence. Furthermore, we can take another subsequence, such that the measures $\theta_{\alpha_{n_k},\varepsilon_{n_k}}(\mu_{n_k} * \nu_{n_k})$ also converge. Passing to a subsequence one more time

we can guarantee (due to compactness of K) that

$$\lim_{k\to\infty} \mu_{n_k} = \mu$$

for some $\mu \in K$. It remains to notice that due to inequality (3.2.59) the limit measure

$$m = \lim_{k\to\infty} \theta_{\alpha_{n_k},\varepsilon_{n_k}}(\nu_{n_k})$$

has an almost surely constant image under the action of $f \in \mathrm{supp}(\mu)$, that is equal to

$$\tilde{m} = \lim_{k\to\infty} \theta_{\alpha_{n_k},\varepsilon_{n_k}}(\mu_{n_k} * \nu_{n_k}) = \lim_{k\to\infty} \theta_{\alpha_{n_k},\varepsilon_{n_k}}(\mu * \nu_{n_k}),$$

which contradicts our assumption. □

3.2.8 Proof of Theorem 3.1.4

Now we have all the ingredients to implement the plan outlined in Section 3.2.1. First we modify Proposition 3.2.27 (contraction of energy) so that we have control over the image of any probability measure, not just those with high energies. Then we apply Markov's inequality (Lemma 3.2.7) to finish the proof of Theorem 3.1.4.

We start by expanding Proposition 3.2.27 so that it includes all measures on M. To do so, we need to adjust the upper bound (3.2.43) in order to include low-energy measures ν:

Lemma 3.2.30. *Assume that the measure μ satisfies the condition (3.1.1) with some β, and that $\alpha < \beta$ and C are given. Then there exists C' such that for any $\varepsilon > 0$ and any measure ν on M with $\mathcal{E}_{\alpha,\varepsilon}(\nu) \le C$ one has*

$$\mathcal{E}_{\alpha,\varepsilon}(\mu * \nu) \le C'.$$

Corollary 3.2.31. *In the assumptions of Proposition 3.2.27 one can conclude that there exist* $\alpha > 0, \widetilde{C} < \infty, \lambda < 1$*, such that for every* $\mu \in K$*, every* $\varepsilon > 0$ *and every measure* ν *on* M

$$\mathcal{E}_{\alpha,\varepsilon}(\mu * \nu) < \max(\lambda \mathcal{E}_{\alpha,\varepsilon}(\nu), \widetilde{C}). \tag{3.2.60}$$

Proof of Lemma 3.2.30. Condition (3.1.1) implies the finiteness of the expectation

$$C_I := \mathbb{E}_\mu \Lambda(f)^\alpha < \infty. \tag{3.2.61}$$

Now, (3.2.48) implies that for every measure ν we have

$$\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\mu * \nu) \leq \mathbb{E}_\mu \widetilde{\mathcal{E}}_{\alpha,\varepsilon}(f_* \nu).$$

At the same time Proposition 3.2.14 for $\delta = \frac{1}{2}$ implies that for some constant C_1 one has for any measure ν'

$$\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu') \leq \max(2\mathcal{E}_{\alpha,\varepsilon}(\nu'), C_1).$$

Applying this for $\nu' = f_*\nu$ and joining it with Corollary 3.2.10, we get

$$\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(f_*\nu) \leq \max(2\mathcal{E}_{\alpha,\varepsilon}(f_*\nu), C_1) \leq 2\max\left(2\Lambda(f)^\alpha \mathcal{E}_{\alpha,\varepsilon}(\nu), C_2\right) + C_1,$$

where C_2 is equal to C from Corollary (3.2.10). Taking the expectation w.r.t. μ and using (3.2.61), we finally get a uniform bound

$$\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\mu * \nu) \leq 4C_I C + 2C_2 + C_1 =: C'.$$

□

Proof of Corollary 3.2.31. In the proof of Proposition 3.2.27 we can take α arbitrarily small, so we can assume that $\alpha < \beta$. Applying Proposition 3.2.14 to both sides of (3.2.43), where we take δ sufficiently small so that $\frac{1-\delta}{1+\delta} > \lambda$, allows to conclude that there exists some constant C_3 such that for all ε and ν

$$\mathcal{E}_{\alpha,\varepsilon}(\mu * \nu) < \lambda' \mathcal{E}_{\alpha,\varepsilon}(\nu) \quad \text{if } \mathcal{E}_{\alpha,\varepsilon}(\nu) > C_3,$$

where $\lambda' = \frac{1+\delta}{1-\delta}\lambda < 1$. Meanwhile, Lemma 3.2.30 implies that there exists some C_4 such that

$$\mathcal{E}_{\alpha,\varepsilon}(\mu * \nu) < C_4 \quad \text{if } \mathcal{E}_{\alpha,\varepsilon}(\nu) \leq C_3.$$

Taking $\widetilde{C} := C_4$, we get the desired (3.2.60). □

Now we are ready to complete the proof of Theorem 3.1.4 (and hence of Theorem 3.1.3).

Proof of Theorem 3.1.4. Let $\alpha, \widetilde{C}$ be as in Corollary 3.2.31: for any $\mu \in K$, any $\varepsilon > 0$ and for any measure ν one has

$$\mathcal{E}_{\alpha,\varepsilon}(\mu * \nu) < \max(\lambda \mathcal{E}_{\alpha,\varepsilon}(\nu), \widetilde{C});$$

applying this n times, we get

$$\mathcal{E}_{\alpha,\varepsilon}(\mu_n * \ldots * \mu_1 * \nu) < \max(\lambda^n \mathcal{E}_{\alpha,\varepsilon}(\nu), \widetilde{C}).$$

Recall that Lemma 3.2.6 gives a uniform upper bound

$$\mathcal{E}_{\alpha,\varepsilon}(\nu) \leq (\omega_k + c_{\alpha,k})(-\log(\varepsilon))^\alpha,$$

choose

$$\varepsilon := e^{-\left(\frac{1}{\lambda}\right)^{\frac{n}{\alpha}}},$$

then we have

$$\mathcal{E}_{\alpha,\varepsilon}(\mu_n * \ldots * \mu_1 * \nu) < \max\left((\omega_k + c_{\alpha,k}), \widetilde{C}\right).$$

Now, applying Lemma 3.2.7 for any $r > \varepsilon = e^{-\kappa^n}$, where $\kappa := (1/\lambda)^{1/\alpha} > 1$, we deduce that

$$(\mu_n * \ldots * \mu_1 * \nu)(B_r(x)) < C_{\alpha,k}\sqrt{\max\left((\omega_k + c_{\alpha,k}), \widetilde{C}\right)} \cdot (-\log(r))^{-\frac{\alpha}{2}}.$$

Renaming $\alpha/2$ as α completes the proof of Theorem 3.1.4.

□

3.3 Lipschitz homeomorphisms

In this section we will outline the proof of Theorem 3.1.9. Since it utilizes the same method as the proof of Theorem 3.1.4 we will give modified versions of key propositions avoiding technical details. In this section we will work with a fixed $\alpha > 0$.

First, we will need an analog of Proposition 3.2.8 to have an estimate for $\mathcal{E}_{\alpha,\varepsilon}(f_*\nu)$ for a Lipschitz continuous f. It can be formulated as follows:

Proposition 3.3.1. *For every $\delta > 0$ there exists $\varepsilon_0 > 0$ and $B < \infty$, such that every $0 < \varepsilon < \varepsilon_0$, every $f \in \mathrm{Lip}(M)$ and every $\nu \in \mathcal{M}$ the following holds:*

$$\mathcal{E}_{\alpha,\varepsilon}(f_*\nu) \leq (1 + \delta \log(\mathfrak{L}(f)))^{\alpha}(\mathcal{E}_{\alpha,\varepsilon}(\nu) + B). \tag{3.3.1}$$

In this case useful corollaries can be formulated as follows:

Corollary 3.3.2. *For every $\delta > 0$ there exists $C < \infty$ such that for every $\varepsilon > 0$, every $f \in \mathrm{Lip}(M)$ and every $\nu \in \mathcal{M}$, such that $\mathcal{E}_{\alpha,\varepsilon}(\nu) > C$ the following holds:*

$$\frac{1}{1+\delta}(1+\delta\log(\mathfrak{L}(f)))^{-\alpha} < \frac{\mathcal{E}_{\alpha,\varepsilon}(f_*\nu)}{\mathcal{E}_{\alpha,\varepsilon}(\nu)} < (1+\delta)(1+\delta\log(\mathfrak{L}(f)))^{\alpha}. \tag{3.3.2}$$

Corollary 3.3.3. *There exists $C < \infty$ such that for every $\varepsilon > 0$, every $f \in \mathrm{Lip}(M)$ and every $\nu \in \mathcal{M}$ the following holds:*

$$\mathcal{E}_{\alpha,\varepsilon}(f_*\nu) < \max(2(1+\log(\mathfrak{L}(f)))^{\alpha}\mathcal{E}_{\alpha,\varepsilon}(\nu), C).$$

Notice, that Proposition 3.2.14 does not depend on the regularity of f, hence it remains unchanged. That, once again, allows us to establish an estimate for $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(f_*\nu)$ (analog of Corollary 3.2.17):

Corollary 3.3.4. *For every $\delta > 0$ and every $R < \infty$ there exists $C > 0$ such that for every $f \in \mathrm{Lip}(M)$ with $\mathfrak{L}(f) < R$, every $\varepsilon > 0$ and every ν such that $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) > C$ or $\mathcal{E}_{\alpha,\varepsilon}(\nu) > C$ one has*

$$\frac{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(f_*\nu)}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)} \in (1-\delta, 1+\delta).$$

Next on the list is the analog of Proposition 3.2.22, which is formulated below.

Proposition 3.3.5. *For any $\delta > 0$, any $R < \infty$ and any $\alpha \in (0,1]$ there exists $C > 0$ such that for any $\varepsilon > 0$, any $f \in \mathrm{Lip}(M)$, such that $\mathfrak{L}(f) < R$ and any $\nu \in \mathcal{M}$ such that $\mathcal{E}_{\alpha,\varepsilon}(\nu) > C$ or $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) > C$ one has*

$$W(f_*\theta_{\alpha,\varepsilon}[\nu], \theta_{\alpha,\varepsilon}[f_*\nu]) < \delta.$$

Once again, the proof repeats the one of Proposition 3.2.22 up to minor computational details. Finally, the uniform estimate (3.1.3) allows us to establish the following analog of Proposition 3.2.25:

Proposition 3.3.6. *If a probability measure μ on $\mathrm{Lip}(M)$ satisfies assumption (3.1.3) then for every $\delta > 0$ there exists $C > 0$ such that for every $\varepsilon > 0$ and every measure $\nu \in \mathcal{M}$ such that $\mathcal{E}_{\alpha,\varepsilon}(\nu) > C$ the following formula holds:*

$$\frac{\mathbb{E}_\mu\left[\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(f_*\nu)\right]}{\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu)} \in (1-\delta, 1+\delta).$$

Utilizing these modified statements we are able to repeat the proof of the contracting property for $\widetilde{\mathcal{E}}_{\alpha,\varepsilon}$ (analog of Proposition 3.2.27, but with a fixed α taken from assumption (3.1.4)):

Proposition 3.3.7. *Under the assumptions of Theorem 3.1.9 there exists $\lambda < 1$ and $C < \infty$, such that for every $\mu \in K$, every $\varepsilon > 0$ and every $\nu \in \mathcal{M}$ if $\mathcal{E}_{\alpha,\varepsilon}(\nu) > C$ then*

$$\widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\mu * \nu) < \lambda \widetilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu).$$

Finally, we can formulate a similar estimate for all measures and not just those with high energy. This statement will also be used in Section 3.5 to prove a technical generalization of Theorem 3.1.9, which was mentioned at the end of Section 1.1.2.

Corollary 3.3.8. *In the assumptions of Proposition 3.3.7 one can conclude that there exist $\widetilde{C} < \infty, \lambda < 1$, such that for every $\mu \in K$, every $\varepsilon > 0$ and every measure ν on M*

$$\mathcal{E}_{\alpha,\varepsilon}(\mu * \nu) < \max(\lambda \mathcal{E}_{\alpha,\varepsilon}(\nu), \widetilde{C}).$$

From here Theorem 3.1.9 (and hence Theorem 3.1.8) follows.

3.4 Technical lemmata

In this section we establish some properties of functions φ_α and $\varphi_{\alpha,\varepsilon}$ and then prove Proposition 3.2.4.

A straightforward computation allows us to prove the following

Lemma 3.4.1. *For any* $0 < r < \frac{1}{e}$

$$\int_{\mathbf{x}\in\mathbb{R}^k, r<|\mathbf{x}|<\frac{1}{e}} \varphi_\alpha(|\mathbf{x}|)^2 \, d\mathbf{x} = c_{\alpha,k}((-\log(r))^\alpha - 1),$$

where $c_{\alpha,k} = \frac{k\omega_k}{\alpha}$*, and* ω_k *is the volume of a unit ball in* $\mathbb{R}^k$*.*

Proof. Since the function $\varphi_\alpha(|\mathbf{x}|)$ depends only on the radius $|\mathbf{x}|$ we conclude that its integral over a sphere of radius y is equal to

$$\int_{\mathbf{x}\in\mathbb{R}^k, |\mathbf{x}|=y} \varphi_\alpha(|\mathbf{x}|)^2 \, \mathrm{d}S(\mathbf{x}) = k\omega_k y^{k-1} \varphi_\alpha(y)^2$$

(ω_k is the volume of a unit ball in $\mathbb{R}^k$, hence $k\omega_k y^{k-1}$ is the $(k-1)$-dimensional volume of a sphere of radius y). Integrating over the radius y we get:

$$\begin{aligned}\int_{\mathbf{x}\in\mathbb{R}^k, r<|\mathbf{x}|<\frac{1}{e}} \varphi_\alpha(|\mathbf{x}|)^2 \, \mathrm{d}\mathbf{x} &= \int_r^{1/e} k\omega_k y^{k-1} \frac{|\log(y)|^{\alpha-1}}{y^k} \, \mathrm{d}y = \\ &= k\omega_k \int_r^{1/e} \frac{(-\log(y))^{\alpha-1}}{y} \, \mathrm{d}y = \frac{k\omega_k}{\alpha}(|\log(r)|^\alpha - 1).\end{aligned}$$

□

Let us define a function

$$V_{\alpha,\varepsilon}(r) = \begin{cases} c_{\alpha,k}|\log(r)|^{\alpha}, & \text{for } \varepsilon < r < 1/e; \\ c_{\alpha,k}|\log(\varepsilon)|^{\alpha}, & \text{for } r \le \varepsilon. \end{cases}$$

Our next goal is to prove

Proposition 3.4.2. *For every $\alpha_0 > 0$ and every $\delta > 0$ there exists $\varepsilon_0 > 0$ and $r_0 > 0$, such that for every $0 < \alpha < \alpha_0$, for every $0 < \varepsilon < \varepsilon_0$, and every $0 \le r < r_0$ we have*

$$\frac{1}{1+\delta} < \frac{U_{\alpha,\varepsilon}(r)}{V_{\alpha,\varepsilon}(r)} < 1 + \delta,$$

where the function $U_{\alpha,\varepsilon}(r)$ is defined by equation (3.2.7).

Proof. Given $\alpha_0 > 0$ and $\delta > 0$, take some constant $c > 0$; its exact value will be chosen later (and will depend on δ and α_0, but not r). Now, write the integral (3.2.7):

$$U_{\alpha,\varepsilon}(r) = \int_{\mathbb{R}^k} \varphi_{\alpha,\varepsilon}(|\mathbf{x}|)\varphi_{\alpha,\varepsilon}(|\mathbf{x} - \bar{r}|)\,\mathrm{d}\mathbf{x},$$

splitting the domain of integration into several parts:

(i) Balls $B_{r/2}(0)$ and $B_{r/2}(\bar{r})$;

(ii) Difference $B_{cr}(0) \setminus (B_{r/2}(0) \cup B_{r/2}(\bar{r}))$;

(iii) Outer region $\mathbb{R}^k \setminus B_{\frac{1}{2e}}(0)$;

(iv) Spherical layer between two spheres $B_{\frac{1}{2e}}(0) \setminus B_{cr}(0)$.

We will show that, for an appropriate choice of the constant c, ε_0, and r_0 the integrals over all these regions except for the last one do not exceed $\frac{\delta}{10}V_{\alpha,\varepsilon}(r)$, and the integral over the

spherical layer will differ from $V_{\alpha,\varepsilon}(r)$ by at most $\frac{\delta}{10}V_{\alpha,\varepsilon}(r)$:

$$1-\frac{\delta}{10}<\frac{1}{V_{\alpha,\varepsilon}(r)}\int_{cr<|\mathbf{x}|<\frac{1}{2e}}\varphi_{\alpha,\varepsilon}(|\mathbf{x}|)\varphi_{\alpha,\varepsilon}(|\mathbf{x}-\bar{r}|)\,\mathrm{d}\mathbf{x}<1+\frac{\delta}{10}. \tag{3.4.1}$$

Adding such upper bounds once they are established would give

$$1-\frac{\delta}{2}<\frac{U_{\alpha,\varepsilon}(r)}{V_{\alpha,\varepsilon}(r)}<1+\frac{\delta}{2},$$

so that would conclude the proof.

Let us start by establishing the required inequality for the outer region. If $|\mathbf{x}|>cr$ then

$$1-\frac{1}{c}<\frac{|\mathbf{x}-\bar{r}|}{|\mathbf{x}|}<1+\frac{1}{c}.$$

Hence, for c big enough we have

$$-\frac{2}{c}<\log|\mathbf{x}-\bar{r}|-\log|\mathbf{x}|<\frac{1}{c},$$

and for any $cr<|\mathbf{x}|<1/e$ we get

$$1-\frac{2}{c}<\frac{\log|\mathbf{x}-\bar{r}|}{\log|\mathbf{x}|}<1+\frac{1}{c}.$$

Recalling the definition of $\varphi_{\alpha,\varepsilon}$ (Definition 3.2.2) we conclude that given $\delta>0$ and α_0 we can pick $r_0<\frac{1}{2e}$ and c big enough so that

$$1-\frac{\delta}{100}<\frac{\varphi_{\alpha,\varepsilon}(|\mathbf{x}|)\varphi_{\alpha,\varepsilon}(|\mathbf{x}-\bar{r}|)}{\varphi_{\alpha,\varepsilon}(|\mathbf{x}|)^2}<1+\frac{\delta}{100}$$

for every $\mathbf{x}$ in spherical layer $B_{\frac{1}{2e}}(0) \setminus B_{cr}(0)$. Integrating this inequality we get

$$1 - \frac{\delta}{100} < \frac{\int_{cr<|\mathbf{x}|<\frac{1}{2e}} \varphi_{\alpha,\varepsilon}(|\mathbf{x}|)\varphi_{\alpha,\varepsilon}(|\mathbf{x}-\bar{r}|)\,\mathrm{d}\mathbf{x}}{\int_{cr<|\mathbf{x}|<\frac{1}{2e}} \varphi_{\alpha,\varepsilon}(|\mathbf{x}|)^2\,\mathrm{d}\mathbf{x}} < 1 + \frac{\delta}{100}. \tag{3.4.2}$$

Using Lemma 3.4.1 it's not too hard to establish that given $\alpha_0 > 0$, $\delta > 0$, and c we can choose $r_0 > 0$ and $\varepsilon_0 > 0$ so that for every $0 < \alpha < \alpha_0$, every $0 < \varepsilon < \varepsilon_0$, and $0 < r < r_0$ we have

$$1 - \frac{\delta}{100} < \frac{\int_{cr<|\mathbf{x}|<\frac{1}{2e}} \varphi_{\alpha,\varepsilon}(|\mathbf{x}|)^2\,\mathrm{d}\mathbf{x}}{V_{\alpha,\varepsilon}(r)} < 1 + \frac{\delta}{100}. \tag{3.4.3}$$

Namely, to prove the inequality above one needs to treat cases $r \geq \varepsilon$ and $r < \varepsilon$ separately. If $r \geq \varepsilon$ then by Lemma 3.4.1 we have

$$\begin{aligned} V_{\alpha,\varepsilon}(r) &= \int_{r<|\mathbf{x}|<1/e} \varphi_{\alpha,\varepsilon}(|\mathbf{x}|)^2\,\mathrm{d}\mathbf{x} + c_{\alpha,k} = \\ &= \int_{r<|\mathbf{x}|<cr} \varphi_{\alpha,\varepsilon}(|\mathbf{x}|)^2\,\mathrm{d}\mathbf{x} + \int_{cr<|\mathbf{x}|<\frac{1}{2e}} \varphi_{\alpha,\varepsilon}(|\mathbf{x}|)^2\,\mathrm{d}\mathbf{x} + \int_{\frac{1}{2e}<|\mathbf{x}|<\frac{1}{e}} \varphi_{\alpha,\varepsilon}(|\mathbf{x}|)^2\,\mathrm{d}\mathbf{x} + c_{\alpha,k} = \\ &= c_{\alpha,k}((-\log(r))^\alpha - (-\log(cr))^\alpha) + c_{\alpha,k}(1+\log(2))^\alpha + \int_{cr<|\mathbf{x}|<\frac{1}{2e}} \varphi_{\alpha,\varepsilon}(|\mathbf{x}|)^2\,\mathrm{d}\mathbf{x}. \end{aligned}$$

Now we notice that all summands except for the last one are small compared to $V_{\alpha,\varepsilon}(r) = c_{\alpha,k}(-\log(r))^\alpha$ for r close to zero, i. e.

$$\frac{c_{\alpha,k}((-\log(r))^\alpha - (-\log(cr))^\alpha) + c_{\alpha,k}(1+\log(2))^\alpha}{c_{\alpha,k}(-\log(r))^\alpha} \to_{r\to 0} 0,$$

and the convergence is uniform in $\alpha \in (0, \alpha_0)$. Hence, by taking small enough r_0 we guarantee (3.4.3). If $r < \varepsilon$ we can represent $V_{\alpha,\varepsilon}(r)$ in the following way:

$$V_{\alpha,\varepsilon}(r) = V_{\alpha,\varepsilon}(\varepsilon) = \int_{\varepsilon<|\mathbf{x}|<\frac{1}{e}} \varphi_{\alpha,\varepsilon}(|\mathbf{x}|)^2\,\mathrm{d}\mathbf{x} + c_{\alpha,k}.$$

The domain of that integral intersects with $(cr, \frac{1}{2e})$ by at least $(c\varepsilon, \frac{1}{2e})$ (since $c > 1$ and

$r < \varepsilon$). Hence the following estimate holds:

$$\left| \int_{cr<|\mathbf{x}|<\frac{1}{2e}} \varphi_{\alpha,\varepsilon}(|\mathbf{x}|)^2 \, \mathrm{d}\mathbf{x} - V_{\alpha,\varepsilon}(r) \right| \leq$$
$$\leq \int_{0<|\mathbf{x}|<\varepsilon} \varphi_\alpha(\varepsilon)^2 \, \mathrm{d}\mathbf{x} + \int_{\varepsilon<|\mathbf{x}|<c\varepsilon} \varphi_\alpha(|\mathbf{x}|)^2 \, \mathrm{d}\mathbf{x} + \int_{\frac{1}{2e}<|\mathbf{x}|<\frac{1}{e}} \varphi_\alpha(|\mathbf{x}|)^2 \, \mathrm{d}\mathbf{x} + c_{\alpha,k} =$$
$$= \omega_k \varepsilon^k \frac{(-\log(\varepsilon))^{\alpha-1}}{\varepsilon^k} + c_{\alpha,k}((-\log(\varepsilon))^\alpha - (-\log(c\varepsilon))^\alpha) + c_{\alpha,k}(1 + \log(2)).$$

Dividing both sides by $V_{\alpha,\varepsilon}(r) = c_{\alpha,k}(-\log(\varepsilon))^\alpha$ and picking ε_0 small enough we establish (3.4.3).

Now equations (3.4.2) and (3.4.3) together imply (3.4.1).

We organize the rest of the proof of Proposition 3.4.2 in a sequence of Lemmata estimating integrals over domains mentioned in (i), (ii), and (iii).

Lemma 3.4.3. *For every $\alpha_0 > 0$, and $\delta > 0$ there exists $r_0 > 0$ and $\varepsilon_0 > 0$ such that for every $0 < \alpha < \alpha_0$, $0 < \varepsilon < \varepsilon_0$, and every $0 < r < r_0$ we have*

$$\int_{B_{r/2}(0) \cup B_{r/2}(\bar{r})} \varphi_{\alpha,\varepsilon}(|\mathbf{x}|)\varphi_{\alpha,\varepsilon}(|\mathbf{x} - \bar{r}|) \, d\mathbf{x} < \frac{\delta}{10} V_{\alpha,\varepsilon}(r). \tag{3.4.4}$$

Proof. Due to the central symmetry of $\varphi_{\alpha,\varepsilon}(|\mathbf{x}|)$ it will be enough to estimate just the integral over $B_{r/2}(0)$. If $r < \varepsilon$ as we have seen before

$$\int_{B_{r/2}(0)} \varphi_{\alpha,\varepsilon}(|\mathbf{x}|)\varphi_{\alpha,\varepsilon}(|\mathbf{x} - \bar{r}|) \, \mathrm{d}\mathbf{x} < \int_{|\mathbf{x}|<\varepsilon} \varphi_{\alpha,\varepsilon}(\varepsilon)^2 \, \mathrm{d}\mathbf{x} = \omega_k(-\log(\varepsilon))^{\alpha-1}$$

and for small enough ε_0 the estimate (3.4.4) follows.

If $r \geq \varepsilon$ the desired inequality will follow from the uniform in $\alpha \in (0, \alpha_0)$ convergence

$$\frac{\int_{|\mathbf{x}|<r/2} \varphi_\alpha(|\mathbf{x}|)\varphi_\alpha(|\bar{r} - \mathbf{x}|) \, \mathrm{d}\mathbf{x}}{(-\log(r))^\alpha} \to_{r\to 0} 0.$$

To prove said convergence we start with the following computation:

$$\int_{|\mathbf{x}|<r/2} \varphi_\alpha(|\mathbf{x}|)\varphi_\alpha(|\bar{r}-\mathbf{x}|)\,\mathrm{d}\mathbf{x} < \varphi_\alpha(r/2)\int_{|\mathbf{x}|<r/2}\varphi_\alpha(|\mathbf{x}|)\,\mathrm{d}\mathbf{x} = \\ = k\omega_k \frac{(-\log(r/2))^{\frac{\alpha-1}{2}}}{(r/2)^{k/2}}\int_0^{r/2}\frac{(-\log(t))^{\frac{\alpha-1}{2}}}{t^{k/2}}t^{k-1}\,\mathrm{d}t. \quad (3.4.5)$$

We will focus on the last integral. If $\alpha \leq 1$ then

$$\int_0^{r/2}(-\log(t))^{\frac{\alpha-1}{2}}t^{k/2-1}\,\mathrm{d}t \leq \int_0^{r/2}t^{k/2-1}\,\mathrm{d}t = \frac{2}{k}\left(\frac{r}{2}\right)^{k/2}.$$

Combining that with formula (3.4.5) we get

$$\frac{\int_{|\mathbf{x}|<r/2}\varphi_\alpha(|\mathbf{x}|)\varphi_\alpha(|\bar{r}-\mathbf{x}|)\,\mathrm{d}\mathbf{x}}{(-\log(r))^\alpha} < 2\omega_k\frac{(-\log(r/2))^{\frac{\alpha-1}{2}}}{(-\log(r))^\alpha}$$

and uniform convergence follows.

If $\alpha > 1$ after taking the last integral in formula (3.4.5) by parts we obtain:

$$\int_0^{r/2}(-\log(t))^{\frac{\alpha-1}{2}}t^{k/2-1}\,\mathrm{d}t = \\ = \left.\frac{2(-\log(t))^{\frac{\alpha-1}{2}}t^{k/2}}{k}\right|_0^{r/2} + \frac{\alpha-1}{k}\int_0^{r/2}(-\log(t))^{\frac{\alpha-3}{2}}t^{k/2-1}\,\mathrm{d}t. \quad (3.4.6)$$

Since r goes to zero we can assume that $-\log(r/2) > 2(\alpha_0 - 1)/k$. In that case we can write

$$\begin{aligned}
&\frac{\alpha-1}{k}\int_0^{r/2}(-\log(t))^{\frac{\alpha-3}{2}}t^{k/2-1}\,\mathrm{d}t = \\
&= \frac{\alpha-1}{k}\int_0^{r/2}(-\log(t))^{\frac{\alpha-3}{2}}(-\log(t))(-\log(t))^{-1}t^{k/2-1}\,\mathrm{d}t < \\
&< \frac{\alpha-1}{k}(-\log(r/2))^{-1}\int_0^{r/2}(-\log(t))^{\frac{\alpha-1}{2}}t^{k/2-1}\,\mathrm{d}t < \\
&\qquad\qquad < \frac{1}{2}\int_0^{r/2}(-\log(t))^{\frac{\alpha-1}{2}}t^{k/2-1}\,\mathrm{d}t.
\end{aligned}$$

Together with formula (3.4.6) the last inequality implies

$$\int_0^{r/2}(-\log(t))^{\frac{\alpha-1}{2}}t^{k/2-1}\,\mathrm{d}t < \frac{4}{k}\left(-\log\left(\frac{r}{2}\right)\right)^{\frac{\alpha-1}{2}}\left(\frac{r}{2}\right)^{k/2}$$

and the uniform convergence follows. □

Lemma 3.4.4. *For every $\alpha_0 > 0$, $\delta > 0$, and $c > 0$ there exists $r_0 > 0$ and $\varepsilon_0 > 0$ such that for every $0 < \alpha < \alpha_0$, $0 < \varepsilon < \varepsilon_0$, and every $0 < r < r_0$ we have*

$$\int_{B_{cr}(0)\setminus(B_{r/2}(0)\cup B_{r/2}(\bar{r}))}\varphi_{\alpha,\varepsilon}(|\mathbf{x}|)\varphi_{\alpha,\varepsilon}(|\mathbf{x}-\bar{r}|)\,d\mathbf{x} < \frac{\delta}{10}V_{\alpha,\varepsilon}(r). \tag{3.4.7}$$

Proof. Let us denote the domain of integration by D:

$$D = B_{cr}(0) \setminus (B_{r/2}(0) \cup B_{r/2}(\bar{r})).$$

Using Cauchy-Schwarz inequality we conclude that

$$\int_D \varphi_{\alpha,\varepsilon}(|\mathbf{x}|)\varphi_{\alpha,\varepsilon}(|\mathbf{x}-\bar{r}|)\,\mathrm{d}\mathbf{x} \leq \sqrt{\int_D \varphi_{\alpha,\varepsilon}(|\mathbf{x}|)^2\,\mathrm{d}\mathbf{x}\int_D \varphi_{\alpha,\varepsilon}(|\mathbf{x}-\bar{r}|)^2\,\mathrm{d}\mathbf{x}}.$$

By increasing domains of both integrals we arrive to

$$\int_D \varphi_{\alpha,\varepsilon}(|\mathbf{x}|)\varphi_{\alpha,\varepsilon}(|\mathbf{x}-\bar{r}|)\,\mathrm{d}\mathbf{x} \le \int_{r/2<|\mathbf{x}|<2cr} \varphi_{\alpha,\varepsilon}(|\mathbf{x}|)^2\,\mathrm{d}\mathbf{x}.$$

Now we have to consider two cases, depending on the relation of r and ε. We treat these cases similarly to our proof of estimate (3.4.3). Assume that $r \ge \varepsilon$. Then by Lemma 3.4.1 we have

$$\int_{r/2<|\mathbf{x}|<2cr} \varphi_{\alpha,\varepsilon}(|\mathbf{x}|)^2\,\mathrm{d}\mathbf{x} \le c_{\alpha,k}((-\log(r/2))^\alpha - (-\log(cr))^\alpha).$$

Choosing r_0 small enough we guarantee inequality (3.4.7).

If $r < \varepsilon$ we have

$$\int_{r/2<|\mathbf{x}|<2cr} \varphi_{\alpha,\varepsilon}(|\mathbf{x}|)^2\,\mathrm{d}\mathbf{x} < \omega_k \varepsilon^k \frac{|\log(\varepsilon)|^{\alpha-1}}{\varepsilon^k} + \int_{\varepsilon<|\mathbf{x}|<2c\varepsilon} \varphi_{\alpha,\varepsilon}(|\mathbf{x}|)^2\,\mathrm{d}\mathbf{x},$$

and choosing ε_0 small enough we establish (3.4.7).

□

To conclude the proof of Proposition 3.4.2 it remains to notice that the integral over domain (iii) is uniformly bounded.

□

Finally, using Proposition 3.4.2 we are able to prove Proposition 3.2.4

Proof of Proposition 3.2.4. **(I)** First, notice that the function $\varphi_{\alpha,\varepsilon}(r)$ is nonincreasing. Using that fact Part **(I)** follows from the definition of $U_{\alpha,\varepsilon}$, for details see Lemma 2.5.6.

(II) Choosing $\delta = 1$ in Proposition 3.4.2 we obtain inequality (3.2.9).

(III) Choosing $\delta = 1$ in Proposition 3.4.2 we obtain inequality (3.2.10).

(IV) Thanks to Proposition 3.4.2 it is enough to prove inequality (3.2.11) for $V_{\alpha,\varepsilon}(r_1)$ and $V_{\alpha,\varepsilon}(r_2)$, which easily follows from the definition.

□

3.5 Joint regularity of two random measures

In this section we present a tailored version of Theorem 3.1.9 that we intend to use in order to prove Central Limit Theorem for non-stationary products of random $\mathrm{SL}(2,\mathbb{R})$ matrices with optimal assumption on the distributions' tails (see Section 1.2 and Chapter 4). Namely, the following result allows us to control the probability that two points independently generated by iterating two random dynamical systems will end up close to one another.

Theorem 3.5.1. *Let K be a compact set (with respect to weak-* topology) in the space of Borel probability measures on Homeo(M), satisfying the following assumptions:*

- *For every $\mu \in K$ we have* $\mathrm{supp}(\mu) \subset \mathrm{Lip}(M)$.
- *There exist $\alpha > 0$ and C_0 such that for every $\mu \in K$*

$$\int_{\mathrm{Lip}(M)} (\log(\mathfrak{L}(f)))^{\alpha} \, d\mu(f) < C_0. \tag{3.5.1}$$

- *For every $\mu \in K$ there are no measures $m_1, m_2 \in \mathcal{M}$, such that $f_* m_1 = m_2$ for μ-a.e. f.*

Then there exist C and $\kappa > 1$ such that for any initial measures $\nu_0^{(1)}, \nu_0^{(2)}$, every number of iterations $n_1, n_2 \in \mathbb{N}$, and every choice of two sequences of measures $\mu_1^{(1)}, \mu_2^{(1)}, \ldots, \mu_{n_1}^{(1)} \in K$

and $\mu_1^{(2)}, \mu_2^{(2)}, \ldots, \mu_{n_2}^{(2)} \in K$ *one has:*

$$\iint_{M\times M} |\log\left(\max\{d(x,y), r\}\right)|^{\alpha}\, d\nu_1(x)\, d\nu_2(y) < C, \tag{3.5.2}$$

where $r > e^{-\kappa^{\min(n_1,n_2)}}$, $\nu_1 = \mu_{n_1}^{(1)} * \mu_{n_1-1}^{(1)} * \ldots * \mu_1^{(1)} * \nu_0^{(1)}$, *and* $\nu_2 = \mu_{n_2}^{(2)} * \mu_{n_2-1}^{(2)} * \ldots * \mu_1^{(2)} * \nu_0^{(2)}$.

Proof. For this proof let us fix $k = \dim(M)$ and $\alpha > 0$ from inequality (3.5.1). According to Part **(III)** of Proposition 3.2.4 there exists $\varepsilon_0 > 0$ and C_U such that for any $\varepsilon < r < \varepsilon_0$ we have

$$U_{\alpha,\varepsilon}(r) > C_U |\log(r)|^{\alpha}.$$

Fix $\kappa = \lambda^{-\frac{1}{\alpha}}$, some $r > e^{-\kappa^{\min(n_1,n_2)}}$, and measures ν_1, ν_2 from the statement. In order to prove (3.5.2) it is enough to show that for some constant $\tilde{C}$ we have

$$\iint_{M\times M} U_{\alpha,r}(d(x,y))\, \mathrm{d}\nu_1(x)\, \mathrm{d}\nu_2(y) < \tilde{C}.$$

Notice that according to the Definition 3.2.5:

$$\mathcal{E}_{\alpha,r}\left(\frac{\nu_1+\nu_2}{2}\right) = \iint_{M\times M} U_{\alpha,r}(d(x,y))\, \mathrm{d}\frac{\nu_1+\nu_2}{2}(x)\, \mathrm{d}\frac{\nu_1+\nu_2}{2}(y) \geq$$
$$\geq \frac{1}{2}\iint_{M\times M} U_{\alpha,r}(d(x,y))\, \mathrm{d}\nu_1(x)\, \mathrm{d}\nu_2(y),$$

so all we need is to come up with a constant upper bound for $\mathcal{E}_{\alpha,r}\left(\frac{\nu_1+\nu_2}{2}\right)$. Thanks to Proposition 3.2.14 it is sufficient to estimate $\widetilde{\mathcal{E}}_{\alpha,r}\left(\frac{\nu_1+\nu_2}{2}\right)$. Now we are able to employ a geometric inequality in $L^2(M, \mathrm{Leb})$ using vectors $\rho_{\alpha,r}[\nu_1]$ and $\rho_{\alpha,r}[\nu_2]$. Namely, recall that

due to the Definition 3.2.12 we have

$$\widetilde{\mathcal{E}}_{\alpha,r}\left(\frac{\nu_1+\nu_2}{2}\right)=\left\|\frac{\rho_{\alpha,r}[\nu_1]+\rho_{\alpha,r}[\nu_2]}{2}\right\|^2_{L^2(M)}\leq$$
$$\leq\frac{\|\rho_{\alpha,r}[\nu_1]\|^2_{L^2(M)}+\|\rho_{\alpha,r}[\nu_2]\|^2_{L^2(M)}}{2}=\frac{1}{2}(\widetilde{\mathcal{E}}_{\alpha,r}(\nu_1)+\widetilde{\mathcal{E}}_{\alpha,r}(\nu_2)).$$

Using Proposition 3.2.14 one more time we conclude that for big enough energies we have

$$\frac{1}{2}(\widetilde{\mathcal{E}}_{\alpha,r}(\nu_1)+\widetilde{\mathcal{E}}_{\alpha,r}(\nu_2))\leq\mathcal{E}_{\alpha,r}(\nu_1)+\mathcal{E}_{\alpha,r}(\nu_2).$$

Starting from here we are mimicking the end of the proof of Theorem 3.1.4. Applying Corollary 3.3.8 we arrive to

$$\mathcal{E}_{\alpha,r}(\nu_1)\leq\max\left(\lambda^{n_1}\mathcal{E}_{\alpha,r}\left(\nu_0^{(1)}\right),\widetilde{C}\right)\quad\text{and}\quad\mathcal{E}_{\alpha,r}(\nu_2)\leq\max\left(\lambda^{n_2}\mathcal{E}_{\alpha,r}\left(\nu_0^{(2)}\right),\widetilde{C}\right).$$

From Lemma 3.2.6 we know that

$$\max\left(\mathcal{E}_{\alpha,r}\left(\nu_0^{(1)}\right),\mathcal{E}_{\alpha,r}\left(\nu_0^{(2)}\right)\right)<(\omega_k+c_{\alpha,k})(-\log(r))^\alpha,$$

so since

$$r>e^{-\kappa^{\min(n_1,n_2)}}=e^{-\lambda^{-\frac{\min(n_1,n_2)}{\alpha}}}$$

we have

$$\mathcal{E}_{\alpha,r}(\nu_1)+\mathcal{E}_{\alpha,r}(\nu_2)<2\max\left(\omega_k+c_{\alpha,k},\widetilde{C}\right).$$

From here Theorem 3.5.1 follows.

□

Chapter 4

Central Limit Theorem for non-stationary random products of $\mathrm{SL}(2,\mathbb{R})$ matrices

4.1 Notations and plan of the proof of the main result

Recall that $\mathcal{K}$ is a compact subset in the set of probability measures on the group $\mathrm{SL}(2,\mathbb{R})$. We will say that **the measures condition** is satisfied if for every measure $\mu \in \mathcal{K}$ there are no Borel probability measures ν_1, ν_2 on $\mathbb{RP}^1$ such that $(f_A)_*\nu_1 = \nu_2$ for μ-almost every $A \in \mathrm{SL}(2,\mathbb{R})$, where f_A is the projectivisation of the matrix A (see Eq. (4.5.1) below).

Let us fix some sequence $\{\mu_i\}_{i\in\mathbb{N}}$, $\mu_i \in \mathcal{K}$, and let $A_i \in \mathrm{SL}(2,\mathbb{R})$ be independent matrix-valued random variables, with A_i being distributed w.r.t. μ_i. Set

$$T_n = A_n A_{n-1} \dots A_1,$$

and denote (following Section 1.2.2)

$$L_n = \mathbb{E} \log \|T_n\|.$$

Moreover, let

$$T_{(n_1,n_2]} := A_{n_2} A_{n_2-1} \dots A_{n_1+1}$$

be the part of the product of our random matrices A_i, where the index varies from $n_1 + 1$ to n_2. Also, denote

$$\xi_n = \log \|T_n\|, \quad \xi_{(n_1,n_2]} = \log \left\|T_{(n_1,n_2]}\right\|.$$

Note that if two intervals of indices $(n_1, n_2]$ and $(n_1', n_2']$ are disjoint, then the corresponding products $T_{(n_1,n_2]}$ and $T_{(n_1',n_2']}$ are independent, and thus so are their log-norms $\xi_{(n_1,n_2]}$ and $\xi_{(n_1',n_2']}$.

Now, a long product of matrices can be split into two parts (that we will later choose to be of comparable lengths): for any n, n' one has

$$T_{n+n'} = T_{(n,n+n']} T_n;$$

in particular, this implies

$$\xi_{n+n'} = \log \|T_{(n,n+n']} T_n\| \le \log \|T_n\| + \log \|T_{(n,n+n']}\| = \xi_n + \xi_{(n,n+n']}. \tag{4.1.1}$$

The right hand side of the inequality in (4.1.1) is a sum of two independent random variables; let us introduce the random variable $R_{n,n'}$ that measures the difference between the right and left hand sides of (4.1.1):

$$R_{n,n'} = \log \|T_n\| + \log \left\|T_{(n,n+n']}\right\| - \log \|T_{n+n'}\| = (\xi_n + \xi_{(n,n+n']}) - \xi_{n+n'}. \tag{4.1.2}$$

Remark 4.1.1. The idea of using estimates for $R_{n,n'}$ in order to establish CLT is similar in spirit to the results and methods used in [7], [8], and [33].

If due to some reason all the discrepancies $R_{n,n'}$ would vanish, then the problem would reduce to analysis of a sum of independent random variables. Therefore, if we could control the random variables $R_{n,n'}$ and knew that the variances of $\{\xi_n\}$ grow, we could try to show that the presence of these controlled discrepancies does not change the behaviour of $\{\xi_n\}$ too much.

We thus split the proof of Theorem 1.2.6 into a "probabilistic" and a "geometric" parts. In the "probabilistic" part (Sections 4.2–4.4), we pass to a general framework of a family of abstract random variables $\{\widetilde{\xi}_{a;n}\}$, independent if the corresponding intervals $(a, a+n]$ do not intersect; they will be later related to the random variables $\{\xi_{(n_1,n_2]}\}$ by (4.2.1). Our main result in this part, Theorem 4.2.1, is a tool that establishes sufficient conditions for the CLT to hold for such a family. These conditions, properties ((a))–((e)) listed in Section 4.2.1, are stated in terms of the moments of these random variables and the behaviour of the

corresponding additivity discrepancies

$$\widetilde{R}_{a;n,n'} := \widetilde{\xi}_{a;n} + \widetilde{\xi}_{a+n;n'} - \widetilde{\xi}_{a;n+n'}.$$

To prove Theorem 4.2.1, we start by obtaining under these assumptions an upper bound for the moments of $\widetilde{\xi}_{a;n}$; it is Proposition 4.2.2 in Section 4.2.2. In Section 4.2.3 (see Lemma 4.2.5) we show that the variances of $\{\widetilde{\xi}_{a;n}\}$ under our assumptions must grow linearly.

Section 4.3 contains the core part of the proof of Theorem 4.2.1, the bootstrapping argument. Namely, the sum of two independent random variables is closer to the Gaussian behavior than the summands separately. We introduce the quantitative way (4.3.1) of measuring how close the distribution is to the Gaussian, and establish the corresponding inequality in Section 4.3.3 (see Lemma 4.3.4). Then, we control how an additional perturbation, coming from the $\widetilde{R}_{a;n,n'}$ terms, can worsen the bounds. This is done in Section 4.3.4; see Proposition 4.3.7 and Corollary 4.3.8 therein. Finally, in Section 4.4, these bootstrapping arguments are applied to prove the Abstract CLT, Theorem 4.2.1.

After that we switch to our initial setting and the random variables $\{\xi_{(n_1,n_2]}\}$, corresponding to the random matrix products, and start the "geometric" part of the proof of Theorem 1.2.6. Our goal is now to check that after centering, the variables $\widetilde{\xi}_{a;n} = \xi_{(a,a+n]} - \mathbb{E}\xi_{(a,a+n]}$ satisfy the assumptions of Theorem 4.2.1.

Specifically, in Section 4.5 we establish uniform bounds on moments of the discrepancies $R_{n,n'}$, see Proposition 4.5.6. To do so, we have to show that it is (sufficiently) improbable that the most expanded vector for the product T_n is sent to the direction close to the one that is contracted by $T_{(n,n+n']}$. This can be reformulated in terms of the action on the projective line $\mathbb{RP}^1$: in these terms, it is the probability of two sequences of iterations sending two given points close to each other. We use results from Chapter 3, where such estimates (log-Hölder bounds after a finite number of non-stationary iterations) were established.

In Section 4.6 we verify that the assumptions ((a))–((d)) of the Abstract CLT are satisfied for the random variables $\widetilde{\xi}_{a;n} = \xi_{(a,a+n]} - \mathbb{E}\xi_{(a,a+n]}$. The assumption ((e)) is verified in Section 4.7. Together with Theorem 4.2.1, this proves the main result, Theorem 1.2.6, see the end of Section 4.6.

Section 4.8 provides the proof of the CLT for vectors and matrix elements of the non-stationary random matrix products, Theorem 1.2.8. Finally, in Section 4.9 we give an example showing that the assumption of finiteness of $(2+\varepsilon)$-th moment in Theorem 1.2.6 is optimal, and cannot be replaced by an assumption of finiteness of 2nd moment, unlike in the case of random products of iid matrices.

4.2 Abstract probabilistic setup

4.2.1 Abstract Central Limit Theorem

Assume that we are given an abstract family $\{\widetilde{\xi}_{a;n}\}_{a\geq 0,\, n\geq 1}$ of random variables; we will list the assumptions, under which we claim the CLT for this family. As we have already mentioned, it is linked to the random variables $\xi_{(a,a+n]} = \log \left\| T_{(a,a+n]} \right\|$ originating from the products of random matrices by setting

$$\widetilde{\xi}_{a;n} := \xi_{(a,a+n]} - \mathbb{E}\xi_{(a,a+n]}. \tag{4.2.1}$$

Note that the second index in this new family corresponds to the "*length*" of the product, and not to the last index; this choice is made due to the importance of this length (the estimates will be uniform in a as n tends to infinity).

Define the difference (additivity discrepancy) random variables

$$\widetilde{R}_{a;n,n'} := \widetilde{\xi}_{a;n} + \widetilde{\xi}_{a+n;n'} - \widetilde{\xi}_{a;n+n'}, \tag{4.2.2}$$

and assume that the following conditions hold:

(a) **Centering:** $\mathbb{E}\widetilde{\xi}_{a;n} = 0$ for all a, n.

(b) **Independence**: for any a, a', n, n' such that $a' \geq a + n$, the random variables $\widetilde{\xi}_{a;n}$ and $\widetilde{\xi}_{a',n'}$ are independent.

(c) **Initial moments:** The variables $\widetilde{\xi}_{a;n}$ have finite $2+\varepsilon$-th moments, and

$$\exists C_M : \quad \forall a = 0, 1, 2, \ldots, \quad \mathbb{E}|\widetilde{\xi}_{a;1}|^{2+\varepsilon} < C_M^{2+\varepsilon}. \tag{4.2.3}$$

(d) **Moments of differences:** There exists C_R such that for any a and any n, n' with $\frac{n'}{2} \leq n \leq 2n'$, one has

$$\mathbb{E}|\widetilde{R}_{a;n,n'}|^{2+\varepsilon} \leq C_R^{2+\varepsilon}. \tag{4.2.4}$$

(e) **Growth of variance**: For every c there exists $n_1 \in \mathbb{N}$ such that

$$\forall n \geq n_1 \quad \forall a \quad \operatorname{Var} \widetilde{\xi}_{a;n} \geq c. \tag{4.2.5}$$

As we will see in Section 4.6, for the family $\{\widetilde{\xi}_{a;n}\}$ associated to the products of random matrices by (4.2.1) these assumptions indeed hold; see Lemma 4.6.1 and Proposition 4.6.2.

The following theorem is our main abstract probability result.

Theorem 4.2.1 (Abstract CLT)**.** *Let the family $\widetilde{\xi}_{a;n}$ satisfy the assumptions ((a))–((e)).*

Then there exist constants C_1, C_2 and a number n_0 such that for every $n > n_0$

$$\forall a \quad C_1^2\, n \le \operatorname{Var} \widetilde{\xi}_{a;n} \le C_2^2\, n, \tag{4.2.6}$$

and random variables

$$\frac{\widetilde{\xi}_{0;n}}{\sqrt{\operatorname{Var} \widetilde{\xi}_{0;n}}} \tag{4.2.7}$$

converge to $\mathcal{N}(0,1)$ in distribution as $n \to \infty$.

As we have already declared, Theorem 1.2.6 will be deduced from Theorem 4.2.1 by checking that the family (4.2.1) satisfies the assumptions ((a))–((e)) above. The rest of this section and Sections 4.3 and 4.4 are devoted to the proof of Theorem 4.2.1.

4.2.2 Upper bounds for moments of $\widetilde{\xi}_{a;n}$

We start by establishing some upper moment estimates for $\widetilde{\xi}_{a;n}$. Note that in the case of $\widetilde{\xi}_{a;n}$ being sums of independent $\mathcal{N}(0,1)$-summands (and in particular, all $R_{a;n,n'} = 0$), one would have $\widetilde{\xi}_{a;n} \sim \mathcal{N}(0,n)$, and thus the p-th moment of $\widetilde{\xi}_{a;n}$ would scale exactly as $n^{p/2}$. The conclusion of Proposition 4.2.2 below, which is the main result of this section, extends upper bounds with such a scaling to the more general setting of the assumptions above.

Proposition 4.2.2. *Assume that the family $\widetilde{\xi}_{a;n}$ satisfies the assumptions ((a))–((d)) above. Then, there exists a constant C_2 such that for all a and n,*

$$\mathbb{E}|\widetilde{\xi}_{a;n}|^{2+\varepsilon} < C_2^{2+\varepsilon} n^{1+\frac{\varepsilon}{2}}. \tag{4.2.8}$$

As a corollary, for any $p \in [0, 2+\varepsilon]$, any a and n, one has

$$\mathbb{E}|\widetilde{\xi}_{a;n}|^p < C_2^p n^{p/2}. \tag{4.2.9}$$

Remark 4.2.3. In the additive case, when $\widetilde{R}_{a;n,n'} = 0$, Proposition 4.2.2 turns into Rosenthal type estimates, e.g. see [37, 36].

Proof of Proposition 4.2.2. We will start by establishing the conclusion (4.2.9) for $p = 2$ (and thus for all smaller values of p), in other words, a linear upper bound for the growth of the variance of $\widetilde{\xi}_{a;n}$:

$$\exists C_\sigma : \quad \forall a, \quad \forall n \quad \operatorname{Var} \widetilde{\xi}_{a;n} \le C_\sigma^2 \cdot n. \tag{4.2.10}$$

To do so, we will recurrently define the sequence c_m, such that for each $m = 1, 2, \dots$ one has

$$\forall a \quad \operatorname{Var} \widetilde{\xi}_{a;m} \le c_m^2 \cdot m. \tag{4.2.11}$$

Namely, we start by setting $c_1 := C_M$; the estimate (4.2.11) is then satisfied due to initial moments assumption ((c)). Now, to construct c_m with $m > 1$, take $n = \left\lfloor \frac{m}{2} \right\rfloor$ and $n' = m - n$. Then for any a,

$$\operatorname{Var}(\widetilde{\xi}_{a;n} + \widetilde{\xi}_{a+n;n'}) = \operatorname{Var} \widetilde{\xi}_{a;n} + \operatorname{Var} \widetilde{\xi}_{a+n;n'} \le c_n^2 \cdot n + c_{n'}^2 \cdot n' \le \left(\max(c_n, c_{n'})\right)^2 \cdot m.$$

Now, the L_2-norm triangle inequality (or Cauchy-Schwarz inequality)

$$\sqrt{\operatorname{Var}(X - Y)} \le \sqrt{\operatorname{Var} X} + \sqrt{\operatorname{Var} Y}$$

holds for any two random variables X, Y with finite second moment; applying it for

$$\widetilde{\xi}_{a;m} = \underbrace{(\widetilde{\xi}_{a;n} + \widetilde{\xi}_{a+n;n'})}_{X} - \underbrace{\widetilde{R}_{a;n,n'}}_{Y},$$

and using assumption ((d)) that implies $\sqrt{\operatorname{Var} \widetilde{R}_{a;n,n'}} \leq 2C_R$, one gets

$$\frac{1}{\sqrt{m}} \sqrt{\operatorname{Var} \widetilde{\xi}_{a;m}} \leq \max(c_n, c_{n'}) + \frac{2C_R}{\sqrt{m}}. \tag{4.2.12}$$

Hence, it suffices to take c_m to be the right hand side of (4.2.12):

$$c_m := \max(c_n, c_{n'}) + \frac{2C_R}{\sqrt{m}}, \tag{4.2.13}$$

to ensure that (4.2.11) holds for this m. For the sequence defined by the recurrence relation (4.2.13), it is easy to check by induction that for all $m = 2^k + 1, \ldots, 2^{k+1}$ we have

$$c_m \leq c_1 + \sum_{j=0}^{k} \frac{2C_R}{\sqrt{2^j}},$$

which in turn implies a uniform bound

$$c_m \leq C_M + \frac{2C_R}{1 - \frac{1}{\sqrt{2}}} =: C_\sigma,$$

thus concluding the proof of (4.2.10). Note also that due to the Jensen inequality, the established estimate also implies

$$\forall a, \quad \forall n \quad \mathbb{E}|\widetilde{\xi}_{a;n}|^p \leq (C_\sigma \cdot \sqrt{n})^p \tag{4.2.14}$$

for every $p \in [0, 2]$.

Let us now pass to the estimate of the $(2 + \varepsilon)$-th moments. We are going again to define

recurrently a sequence c'_m, such that for all $m = 1, 2, \dots$ one has

$$\forall a \quad \mathbb{E}|\widetilde{\xi}_{a;m}|^{2+\varepsilon} \leq (c'_m \cdot \sqrt{m})^{2+\varepsilon}. \tag{4.2.15}$$

As before, we start by setting $c'_1 := C_M$; the estimate (4.2.15) is then satisfied due to initial moments assumption ((c)). Now, to construct c_m with $m > 1$, we take $n = \left\lfloor \frac{m}{2} \right\rfloor$ and $n' = m - n$. We will use the following lemma (postponing its proof until the end of the proof of Proposition 4.2.2):

Lemma 4.2.4. *Let X, Y be two independent random variables. Then*

$$\mathbb{E}|X+Y|^{2+\varepsilon} \leq \mathbb{E}|X|^{2+\varepsilon} + \mathbb{E}|Y|^{2+\varepsilon} + 2^{2+\varepsilon}\left(\mathbb{E}|X| \cdot \mathbb{E}|Y|^{1+\varepsilon} + \mathbb{E}|X|^{1+\varepsilon} \cdot \mathbb{E}|Y|\right). \tag{4.2.16}$$

Now, let us apply Lemma 4.2.4 to $X = \widetilde{\xi}_{a;n}$ and $Y = \widetilde{\xi}_{a+n;n'}$. Note that due to the upper bound (4.2.14) for first and $1+\varepsilon$-th moments, we get

$$\mathbb{E}|X|^{1+\varepsilon} \cdot \mathbb{E}|Y| \leq C_\sigma^{1+\varepsilon} \sqrt{n}^{1+\varepsilon} \cdot C_\sigma \sqrt{n'} \leq C_\sigma^{2+\varepsilon} \sqrt{m}^{2+\varepsilon},$$

and the same estimate applies to $\mathbb{E}|X| \cdot \mathbb{E}|Y|^{1+\varepsilon}$. From the estimate (4.2.16) we thus get

$$\begin{aligned} \mathbb{E}\left|\widetilde{\xi}_{a;n} + \widetilde{\xi}_{a+n;n'}\right|^{2+\varepsilon} &\leq \mathbb{E}\left|\widetilde{\xi}_{a;n}\right|^{2+\varepsilon} + \mathbb{E}\left|\widetilde{\xi}_{a+n;n'}\right|^{2+\varepsilon} + 2(2C_\sigma)^{2+\varepsilon} \cdot \sqrt{m}^{2+\varepsilon} \\ &\leq (\max(c'_n, c'_{n'}))^{2+\varepsilon} \cdot \left(n^{1+\frac{\varepsilon}{2}} + (n')^{1+\frac{\varepsilon}{2}}\right) + (4C_\sigma)^{2+\varepsilon} \cdot m^{1+\frac{\varepsilon}{2}} \end{aligned} \tag{4.2.17}$$

Now, let

$$\lambda_m := \left(\frac{n}{m}\right)^{1+\frac{\varepsilon}{2}} + \left(\frac{n'}{m}\right)^{1+\frac{\varepsilon}{2}}.$$

Taking $2+\varepsilon$-th power root, and using the concavity inequality $\sqrt[2+\varepsilon]{a+b} \leq \sqrt[2+\varepsilon]{a} + \sqrt[2+\varepsilon]{b}$, we

obtain

$$\left(\mathbb{E}\left|\widetilde{\xi}_{a;n}+\widetilde{\xi}_{a+n;n'}\right|^{2+\varepsilon}\right)^{1/(2+\varepsilon)} \leq (\lambda_m^{1/(2+\varepsilon)}\cdot\max(c'_n,c'_{n'})+4C_\sigma)\cdot\sqrt{m}. \tag{4.2.18}$$

Now, using the $L_{2+\varepsilon}$–triangle inequality for $(\widetilde{\xi}_{a;n}+\widetilde{\xi}_{a+n;n'})$ and $\widetilde{R}_{a;n,n'}$, we get

$$\left(\mathbb{E}\left|\widetilde{\xi}_{a;m}\right|^{2+\varepsilon}\right)^{1/(2+\varepsilon)} \leq (\lambda_m^{1/(2+\varepsilon)}\cdot\max(c'_n,c'_{n'})+4C_\sigma)\cdot\sqrt{m}+2C_R.$$

Hence, it suffices to take

$$c'_m := \lambda_m^{1/(2+\varepsilon)}\cdot\max(c'_n,c'_{n'})+4C_\sigma+\frac{2C_R}{\sqrt{m}} \tag{4.2.19}$$

for the upper bound $\mathbb{E}\left|\widetilde{\xi}_{a;m}\right|^{2+\varepsilon} \leq (c'_m\cdot\sqrt{m})^{2+\varepsilon}$ to hold.

Now, for every m one has $\lambda_m<1$, and the values λ_m converge to $2^{-\frac{\varepsilon}{2}}<1$ as $m\to\infty$. Hence, there exists $\lambda<1$ such that for all m, one has $\lambda_m^{1/(2+\varepsilon)}<\lambda$. Taking

$$C_2 := \frac{4C_\sigma+2C_R}{1-\lambda}$$

to be the fixed point of the map $c\mapsto\lambda c+(4C_\sigma+2C_R)$, we see by recurrence that the sequence (c'_m) satisfies $c'_m\leq C_2$ for all m. We have thus obtained the desired upper bound

$$\mathbb{E}|\widetilde{\xi}_{a;m}|^{2+\varepsilon} \leq C_2^{2+\varepsilon}\cdot\sqrt{m}^{2+\varepsilon}.$$

Let us now prove Lemma 4.2.4:

Proof of Lemma 4.2.4. Let us first show that for any $a,b\in\mathbb{R}$,

$$|a+b|^{2+\varepsilon} \leq |a|^{2+\varepsilon}+|b|^{2+\varepsilon}+2^{2+\varepsilon}\cdot(|a|^{1+\varepsilon}|b|+|a|\cdot|b|^{1+\varepsilon}). \tag{4.2.20}$$

Indeed, Jensen's inequality for the function $(1+x)^{2+\varepsilon}$ on $[0,1]$ implies that

$$\forall x \in [0,1] \quad (1+x)^{2+\varepsilon} \leq (1-x)\cdot 1 + x \cdot 2^{2+\varepsilon} \leq 1 + 2^{2+\varepsilon} x. \tag{4.2.21}$$

Now, if $|a| \geq |b|$, then

$$|a+b|^{2+\varepsilon} \leq |a|^{2+\varepsilon} \cdot \left(1 + \frac{|b|}{|a|}\right)^{2+\varepsilon} \leq |a|^{2+\varepsilon} + 2^{2+\varepsilon} \cdot |a|^{1+\varepsilon}|b|, \tag{4.2.22}$$

where we have applied (4.2.21) with $x = \frac{|b|}{|a|}$. In the same way, if $|b| \geq |a|$, then

$$|a+b|^{2+\varepsilon} \leq |b|^{2+\varepsilon} + 2^{2+\varepsilon} \cdot |b|^{1+\varepsilon}|a|. \tag{4.2.23}$$

Taking the sum of the right hand sides of (4.2.22) and (4.2.23), we obtain the desired upper bound (4.2.20). It now suffices to apply this inequality for X and Y, and to take the expectation. □

With Lemma 4.2.4 proven, the proof of Proposition 4.2.2 is now complete. □

4.2.3 Linear growth of variances

In this section we prove a lower growth bound for the variances of $\widetilde{\xi}_{a;n}$. Together with already established Proposition 4.2.2, it implies the linear growth conclusion (4.2.6) of Theorem 4.2.1; the latter corresponds to inequality (1.2.4) in Theorem 1.2.6.

Lemma 4.2.5. *Under the assumptions ((a))–((e)) above, there exists $C_1 > 0$ and n_0 such that*

$$\forall a \quad \forall n \geq n_0 \quad \operatorname{Var} \widetilde{\xi}_{a;n} \geq C_1^2 n. \tag{4.2.24}$$

The proof of Lemma 4.2.5 uses the same ideas of controlling the influence of the discrepancy $\widetilde{R}_{a;n,n'}$ as those that appear in the proof of Proposition 4.2.2.

Proof of Lemma 4.2.5. Recall that

$$\widetilde{\xi}_{a;n+n'} = (\widetilde{\xi}_{a;n} + \widetilde{\xi}_{a+n;n'}) - \widetilde{R}_{a;n,n'};$$

Cauchy-Schwarz inequality thus implies that

$$\sqrt{\operatorname{Var} \widetilde{\xi}_{a;n+n'}} \geq \sqrt{\operatorname{Var}(\widetilde{\xi}_{a;n} + \widetilde{\xi}_{a+n;n'})} - \sqrt{\operatorname{Var} \widetilde{R}_{a;n,n'}}. \tag{4.2.25}$$

We will use (4.2.25) to establish by induction that for some n_0 and $q > 0$ for all $m \geq n_0$ and all a one has

$$\sqrt{\operatorname{Var} \widetilde{\xi}_{a;m}} \geq \sqrt{q(m+1)} + 3C_R. \tag{4.2.26}$$

To do that, set n_0 to be equal to n_1 for which (4.2.5) holds with $c = 16C_R^2$, and let $q := \frac{C_R^2}{2n_0}$, where C_R is given by the assumption ((d)). This choice guarantees that (4.2.26) holds for any a and every $m = n_0, \dots, 2n_0 - 1$: we have

$$\sqrt{\operatorname{Var} \widetilde{\xi}_{a;m}} \geq \sqrt{16C_R^2} = 4C_R \geq \sqrt{\frac{C_R^2}{2n_0} \cdot (m+1)} + 3C_R.$$

Now, we proceed by induction to show that (4.2.26) actually holds for every $m \geq n_0$. Indeed, let $m \geq 2n_0$ be the first number for which (4.2.26) has not yet been established; decompose it as $m = n + n'$, where $n = \lfloor \frac{m}{2} \rfloor$, $n' = \lceil \frac{m}{2} \rceil = m - n$. Then, each of the variances $\operatorname{Var} \widetilde{\xi}_{a;n}$, $\operatorname{Var} \widetilde{\xi}_{a+n;n'}$ in (4.2.25) is bounded from below by $\left(\sqrt{q(n+1)} + 3C_R\right)^2$, and hence, due to

independence of $\widetilde{\xi}_{a;n}$ and $\widetilde{\xi}_{a+n;n'}$ we have

$$\sqrt{\operatorname{Var}\widetilde{\xi}_{a;m}} \geq \sqrt{\operatorname{Var}\widetilde{\xi}_{a;n} + \operatorname{Var}\widetilde{\xi}_{a+n;n'}} - \sqrt{\operatorname{Var}\widetilde{R}_{a;n,n'}} \geq$$
$$\geq \sqrt{2}\cdot(\sqrt{q(n+1)} + 3C_R) - C_R \geq$$
$$\geq \sqrt{q(m+1)} + (3\sqrt{2}-1)C_R,$$

where we have used $2(n+1) \geq m+1$, as $n = \lfloor \frac{m}{2} \rfloor$. As $3\sqrt{2}-1 > 3$, this proves the induction step. In particular, for every $m \geq n_0$ and any a we have $\operatorname{Var}\widetilde{\xi}_{a;m} \geq qm$, that proves the lemma with $C_1 = \sqrt{q} = \frac{C_R}{\sqrt{2n_0}}$.

□

Now, define the normalised family

$$\eta_{a;n} := \frac{1}{\sqrt{\operatorname{Var}\widetilde{\xi}_{a;n}}}\widetilde{\xi}_{a;n}. \tag{4.2.27}$$

Combining the upper bound for the $2+\varepsilon$-th moment of $\widetilde{\xi}_{a;n}$ by $(C_2\sqrt{n})^{2+\varepsilon}$ from Proposition 4.2.2 with a lower bound for its variance by $(C_1\sqrt{n})^2$ from Lemma 4.2.5, we get the following uniform upper bound.

Corollary 4.2.6. *Under the assumptions ((a))–((e)) above, there exists $\overline{C}$, such that for all $n \geq n_0$, with n_0 given by Lemma 4.2.5, and any a, the normalised variable $\eta_{a;n}$, defined by* (4.2.27)*, satisfies*

$$\mathbb{E}|\eta_{a;n}|^{2+\varepsilon} < \overline{C}^{2+\varepsilon}.$$

Proof. Indeed, one has

$$\mathbb{E}|\eta_{a;n}|^{2+\varepsilon} < \frac{(C_2\sqrt{n})^{2+\varepsilon}}{(C_1\sqrt{n})^{2+\varepsilon}} = \frac{C_2^{2+\varepsilon}}{C_1^{2+\varepsilon}},$$

hence it suffices to take $\overline{C} := C_2/C_1$. $\square$

4.3 Bootstrapping: distance to the Gaussian distribution

Let us now pass to the core part of the proof of Theorem 4.2.1: in this section we present the bootstrapping arguments that allow to show the convergence to Gaussian law.

4.3.1 Preliminaries

Let ξ, ξ' be two independent random variables with finite $(2+\varepsilon)$-th moments and comparable variances: for a given constant $C > 1$, we have

$$C^{-1} < \frac{\operatorname{Var}\xi}{\operatorname{Var}\xi'} < C.$$

We will provide a value $N'_\rho(\xi)$, measuring quantitatively "non-Gaussianity" of the law of ξ, such that (under appropriate assumptions) it will be smaller for the sum $\xi + \xi'$ than for the summands separately.

Definition 4.3.1. For a random variable ξ we denote by $\varphi_\xi(t)$ its *characteristic function*:

$$\varphi_\xi(t) = \mathbb{E}e^{it\xi}.$$

Now, let

$$N_\rho(\eta) = \sup_{0<|t|<\rho} \frac{\left|\log\left(\varphi_\eta(t)e^{t^2/2}\right)\right|}{|t|^{2+\varepsilon}}, \quad N'_\rho(\xi) = N_\rho\left(\frac{\xi - \mathbb{E}\xi}{\sqrt{\operatorname{Var}\xi}}\right). \tag{4.3.1}$$

Then $\eta \sim \mathcal{N}(0,1)$ if and only if $N_\rho(\eta) = 0$ for all $\rho > 0$ (as the distribution of a random

variable is uniquely determined by its characteristic function). Note also that $N_\rho(\eta)$ might be infinite if the corresponding characteristic function φ_η vanishes somewhere on $[-\rho, \rho]$. Finally, the logarithm here is a function of a complex variable (as the characteristic function might be, and most often is, non-real). As soon as $\varphi_\eta(t)$ does not vanish on $[-\rho, \rho]$, we define the composition $\log\left(\varphi_\eta(t)e^{t^2/2}\right)$ by a continuous extension, starting with the value $\log 1 = 0$ at $t = 0$.

4.3.2 Initial estimates

To start a bootstrapping argument, one needs some initial bounds, in this case, for the norms $N'_\rho(\xi)$ for some $\rho > 0$. The first step to obtain these is the following bound for the characteristic functions:

Lemma 4.3.2. *Let X be a random variable with*

$$\mathbb{E}X = 0, \quad \operatorname{Var} X = 1, \quad \mathbb{E}|X|^{2+\varepsilon} < C_X^{2+\varepsilon},$$

for some $\varepsilon \in (0, 1)$. Then its characteristic function satisfies

$$\left|\varphi_X(t) - \left(1 - \frac{t^2}{2}\right)\right| \le C_X^{2+\varepsilon} \cdot |t|^{2+\varepsilon} \quad \textit{for all } t \in \mathbb{R}. \tag{4.3.2}$$

Proof. Note that it suffices to establish the following estimate for the second derivative $\varphi''(t)$: for all $t \in \mathbb{R}$,

$$|\varphi''_X(t) + 1| \le 2C_X^{2+\varepsilon}\,|t|^\varepsilon. \tag{4.3.3}$$

Indeed, integrating (4.3.3) two times then suffices to obtain (4.3.2):

$$\varphi_X(t) - \left(1 - \frac{t^2}{2}\right) = \int_0^t dt_1 \int_0^{t_1} (\varphi''_X(t_2) + 1)\, dt_2.$$

Now, let us rewrite the estimated expression in (4.3.3):

$$|\varphi_X''(t) + 1| = |\mathbb{E}(X^2(e^{itX} - 1))| \leq \mathbb{E}(X^2|e^{itX} - 1|) \leq \mathbb{E}(X^2 \cdot 2|tX|^{\varepsilon}); \tag{4.3.4}$$

here we have used $\operatorname{Var} X = 1$ for the first equality, and the upper bound

$$|e^{ix} - 1| \leq \min(2, |x|) \leq 2|x|^{\varepsilon} \tag{4.3.5}$$

for the last inequality. Finally, the right hand side of (4.3.4) can be rewritten as

$$2|t|^{\varepsilon} \cdot \mathbb{E}|X|^{2+\varepsilon} < 2C_X^{2+\varepsilon}|t|^{\varepsilon},$$

thus completing the proof. □

Joining Lemma 4.3.2 with the estimate from Corollary 4.2.6, we get an initial bound lemma for the family $\{\widetilde{\xi}_{a;n}\}$.

Lemma 4.3.3. *Under the assumptions ((a))–((e)) above, for the normalised variables $\eta_{a;n}$, defined by* (4.2.27)*, the following estimate holds. There exists $\rho_0 > 0$, such that for any $n \geq n_0$, where n_0 is given by Lemma 4.2.5, and any a the value $N_{\rho_0}(\eta_{a;n})$ is well-defined and satisfies*

$$N_{\rho_0}(\eta_{a;n}) < 3\overline{C}^{2+\varepsilon} \quad \textit{and} \quad \rho_0^{2+\varepsilon} N_{\rho_0}(\eta_{a;n}) < \frac{1}{100},$$

where $\overline{C}$ is the constant defined in Corollary 4.2.6.

Moreover, the choice of ρ_0 depends only on the constants under the assumptions ((a))–((e)).

Proof. Applying Lemma 4.3.2, and multiplying (4.3.2) by $e^{\frac{t^2}{2}}$, we get

$$\left| e^{\frac{t^2}{2}} \varphi_{\eta_{a;n}}(t) - e^{\frac{t^2}{2}} \left(1 - \frac{t^2}{2}\right) \right| \leq \overline{C}^{2+\varepsilon} |t|^{2+\varepsilon} \cdot e^{\frac{t^2}{2}} \tag{4.3.6}$$

Now,

$$e^{\frac{t^2}{2}} \left(1 - \frac{t^2}{2}\right) = 1 + o(|t|^{2+\varepsilon}),$$

thus for sufficiently small ρ one has

$$\forall |t| \leq \rho \qquad \left| e^{\frac{t^2}{2}} \left(1 - \frac{t^2}{2}\right) - 1 \right| \leq \frac{1}{2} \overline{C}^{2+\varepsilon} |t|^{2+\varepsilon}, \tag{4.3.7}$$

as well as

$$\forall |t| \leq \rho \qquad e^{\frac{t^2}{2}} \leq e^{\frac{\rho^2}{2}} < \frac{3}{2},$$

hence the right hand side of (4.3.6) can be replaced by $\frac{3}{2} \overline{C}^{2+\varepsilon} |t|^{2+\varepsilon}$. Joining this with (4.3.7), we get

$$\forall |t| \leq \rho \qquad \left| e^{\frac{t^2}{2}} \varphi_{\eta_{a;n}}(t) - 1 \right| \leq \left(\frac{1}{2} + \frac{3}{2} \right) \cdot \overline{C}^{2+\varepsilon} |t|^{2+\varepsilon} = 2(\overline{C} \cdot |t|)^{2+\varepsilon}.$$

Taking

$$\rho_0 := \min \left(\rho, \frac{1}{20\overline{C}} \right),$$

we ensure that $2(\overline{C}\rho_0)^{2+\varepsilon} < \frac{1}{100}$. Hence the function $\log z$ is $\frac{3}{2}$-Lipschitz in the (complex) disc $U_{2(\overline{C}\rho_0)^{2+\varepsilon}}(1)$, and therefore,

$$\forall |t| \leq \rho_0 \quad \left| \log \left(e^{\frac{t^2}{2}} \varphi_{\eta_{a;n}}(t) \right) \right| \leq \frac{3}{2} \cdot 2\overline{C}^{2+\varepsilon} |t|^{2+\varepsilon},$$

which implies the desired conclusions

$$N_{\rho_0}(\eta_{a;n}) < 3\overline{C}^{2+\varepsilon} \quad \text{and} \quad \rho_0^{2+\varepsilon} N_{\rho_0}(\eta_{a;n}) < \frac{1}{(20\overline{C})^{2+\varepsilon}} \cdot 3\,\overline{C}^{2+\varepsilon} < \frac{1}{100}.$$

□

4.3.3 Sum of two independent variables

The following is the first step of the bootstrapping argument, estimating the decrease of N'-values for the sum of two independent random variables. Notice that in addition to the decrease by a linear factor, the parameter ρ (describing the size of the domain) gets increased.

Lemma 4.3.4. *For any $C > 0$ there exist $\lambda < 1$ and $L > 1$ such that if for some $\rho > 0$ for some independent random variables ξ, ξ' one has*

$$C^{-1} < \frac{\operatorname{Var} \xi}{\operatorname{Var} \xi'} < C$$

and values $N'_\rho(\xi), N'_\rho(\xi')$ are finite, then

$$N'_{L\rho}(\xi + \xi') \le \lambda \cdot \max(N'_\rho(\xi), N'_\rho(\xi')). \tag{4.3.8}$$

Proof. Let

$$\eta = \frac{\xi - \mathbb{E}\xi}{\sqrt{\operatorname{Var} \xi}}, \quad \eta' = \frac{\xi' - \mathbb{E}\xi'}{\sqrt{\operatorname{Var} \xi'}}, \quad \eta'' = \frac{(\xi + \xi') - (\mathbb{E}(\xi + \xi'))}{\sqrt{\operatorname{Var}(\xi + \xi')}},$$

Also, denote

$$c = \sqrt{\frac{\operatorname{Var} \xi}{\operatorname{Var} \xi + \operatorname{Var} \xi'}}, \quad c' = \sqrt{\frac{\operatorname{Var} \xi'}{\operatorname{Var} \xi + \operatorname{Var} \xi'}};$$

then, one has

$$\eta'' = c\eta + c'\eta',$$

with the coefficients that satisfy

$$c^2 + (c')^2 = 1, \quad c, c' \leq \sqrt{\frac{C}{C+1}} < 1.$$

By definition, we have for any L

$$N'_{L\rho}(\xi + \xi') = N_{L\rho}(\eta'') = N_{L\rho}(c\eta + c'\eta').$$

Now, for the characteristic functions we have

$$\varphi_{c\eta + c'\eta'}(t) = \varphi_\eta(ct) \cdot \varphi_{\eta'}(c't),$$

and as $c^2 + (c')^2 = 1$, we have

$$e^{\frac{1}{2}t^2}\varphi_{c\eta + c'\eta'}(t) = e^{\frac{1}{2}(ct)^2}\varphi_\eta(ct) \cdot e^{\frac{1}{2}(c't)^2}\varphi_{\eta'}(c't). \tag{4.3.9}$$

If $cL, c'L \leq 1$, taking the logarithm of (4.3.9) and dividing by $|t|^{2+\varepsilon}$, we get

$$N_{L\rho}(\eta'') = \sup_{0<|t|<L\rho} \frac{\left|\log\left(e^{\frac{1}{2}t^2}\varphi_{\eta''}(t)\right)\right|}{|t|^{2+\varepsilon}} \leq$$
$$\leq \sup_{0<|t|<L\rho} \frac{\left|\log\left(e^{\frac{1}{2}(ct)^2}\varphi_\eta(ct)\right)\right|}{|t|^{2+\varepsilon}} + \sup_{0<|t|<L\rho} \frac{\left|\log\left(e^{\frac{1}{2}(c't)^2}\varphi_{\eta'}(c't)\right)\right|}{|t|^{2+\varepsilon}}$$

Making the ct and $c't$ variable change in the first and second expressions respectively in the

right hand side, we obtain

$$N_{L\rho}(\eta'') \le c^{2+\varepsilon} N_{cL\rho}(\eta) + (c')^{2+\varepsilon} N_{c'L\rho}(\eta') \le$$
$$\le (c^{2+\varepsilon} + (c')^{2+\varepsilon}) \cdot \max(N_\rho(\eta), N_\rho(\eta')) \le$$
$$\le \max(c^\varepsilon, (c')^\varepsilon) \cdot \max(N_\rho(\eta), N_\rho(\eta')).$$

Taking $L = \sqrt{\frac{C+1}{C}}$ and $\lambda = L^{-\varepsilon}$ concludes the proof. □

4.3.4 Correction by an additional term

For the family $\{\widetilde{\xi}_{a;n}\}$, one has

$$\widetilde{\xi}_{a;n+n'} = (\widetilde{\xi}_{a;n} + \widetilde{\xi}_{a+n;n'}) - \widetilde{R}_{a;n,n'},$$

with an additional term $\widetilde{R}_{a;n,n'}$ present in addition to the sum of independent random variables

$$\widetilde{S}_{a;n,n'} := \widetilde{\xi}_{a;n} + \widetilde{\xi}_{a+n;n'}, \tag{4.3.10}$$

Due to this, an additional (and possibly non-independent) term is added to the normalized random variable: for

$$X = \frac{\widetilde{S}_{a;n,n'}}{\sqrt{\operatorname{Var} \widetilde{S}_{a;n,n'}}}, \quad Y = \frac{\widetilde{\xi}_{a;n+n'}}{\sqrt{\operatorname{Var} \widetilde{\xi}_{a;n+n'}}} \tag{4.3.11}$$

this term is the difference

$$r = r_{a;n,n'} = Y - X.$$

In this section we control influence of this difference between normalized variables on the corresponding value $N'_\rho(\cdot)$. First, note that $\mathbb{E}r = 0$, and its $2+\varepsilon$-th moment $\mathbb{E}|r|^{2+\varepsilon}$ satisfies the following upper bound.

Lemma 4.3.5. *Under the assumptions ((a))–((e)) above, and with n_0 given by Lemma 4.2.5, there exists $Q_r < \infty$, such that for every $n, n' \geq n_0$, satisfying*

$$\frac{n}{2} \leq n' \leq 2n,$$

we have

$$\mathbb{E}|r_{a;n,n'}|^{2+\varepsilon} < \left(\frac{Q_r}{\sqrt{n+n'}}\right)^{2+\varepsilon} \tag{4.3.12}$$

Proof. Note that $r_{a;n,n'}$ can be expressed as

$$r = r_{a;n,n'} = \frac{\sqrt{\operatorname{Var}\widetilde{S}_{a;n,n'}} - \sqrt{\operatorname{Var}\widetilde{\xi}_{a;n+n'}}}{\sqrt{\operatorname{Var}\widetilde{\xi}_{a;n+n'}}} \cdot X - \frac{1}{\sqrt{\operatorname{Var}\widetilde{\xi}_{a;n+n'}}}\widetilde{R}_{a;n,n'}. \tag{4.3.13}$$

Now, the $L_{2+\varepsilon}$-norms of both X and $\widetilde{R}_{a;n,n'}$ are uniformly bounded (due to Corollary 4.2.6 and assumption ((d)) respectively). On the other hand, from the Cauchy-Schwartz inequality one has

$$\left|\sqrt{\operatorname{Var}\widetilde{S}_{a;n,n'}} - \sqrt{\operatorname{Var}\widetilde{\xi}_{a;n+n'}}\right| \leq \sqrt{\operatorname{Var}\widetilde{R}_{a;n,n'}},$$

and hence, due to Lemma 4.2.5, both coefficients admit upper bounds as $\frac{\text{const}}{\sqrt{n+n'}}$. □

Our next step is to control the influence of such a "small" change by $r = Y - X$ on the non-Gaussianity value $N_\rho(\cdot)$. We first estimate the change of the characteristic function. Namely, assuming that we are given random variables X and Y with a bound on their $2+\varepsilon$-th moment, we provide an estimate that tends to zero as a power of $2+\varepsilon$-th moment of their

difference.

Lemma 4.3.6. *Let X, Y be two random variables with*

$$\mathbb{E}X = \mathbb{E}Y = 0, \quad \operatorname{Var} X = \operatorname{Var} Y = 1, \quad \mathbb{E}|X|^{2+\varepsilon}, \mathbb{E}|Y|^{2+\varepsilon} < C_X^{2+\varepsilon},$$

for some $\varepsilon \in (0,1)$. Assume that for $r = Y - X$ one has $\mathbb{E}|r|^{2+\varepsilon} < C_r^{2+\varepsilon}$. Then for any $t \in \mathbb{R}$ one has

$$|\varphi_X(t) - \varphi_Y(t)| \le 5 C_r^{\varepsilon} C_X^2 \cdot |t|^{2+\varepsilon} \tag{4.3.14}$$

Proof. Note that it suffices to obtain an estimate

$$|(\varphi_X - \varphi_Y)''(t)| \le 10\, C_r^{\varepsilon} C_X^2 \cdot |t|^{\varepsilon}; \tag{4.3.15}$$

indeed, using

$$\varphi_X(t) - \varphi_Y(t) = \int_0^t dt_1 \int_0^{t_1} (\varphi_X - \varphi_Y)''(t_2)\, dt_2,$$

and integrating (4.3.15) two times, we get the desired (4.3.14).

Now, the second derivative of the difference of the characteristic functions can be rewritten in the following way.

$$(\varphi_X - \varphi_Y)''(t) = -\mathbb{E}(X^2 e^{itX} - Y^2 e^{itY}) = -\mathbb{E}(X^2(e^{itX} - 1) - Y^2(e^{itY} - 1)), \tag{4.3.16}$$

where the second equality follows from $\operatorname{Var} X = \operatorname{Var} Y$.

In order to obtain the estimate (4.3.15), let us decompose the right hand side of (4.3.16):

$$|\varphi_X''(t) - \varphi_Y''(t)| \le \mathbb{E}|(X^2 - Y^2)(e^{itX} - 1)| + \mathbb{E}(Y^2 |e^{itY} - e^{itX}|).$$

Using (4.3.5), we can estimate the second summand:

$$\mathbb{E}(Y^2|e^{itY} - e^{itX}|) \le \mathbb{E}(Y^2 \cdot 2|t(X-Y)|^\varepsilon) = |t|^\varepsilon \cdot \mathbb{E}(2|r|^\varepsilon \cdot Y^2);$$

the right hand side now does not exceed $2C_r^\varepsilon C_X^2 |t|^\varepsilon$ due to the Hölder inequality with the exponents $p = \frac{2+\varepsilon}{\varepsilon}$ and $q = \frac{2+\varepsilon}{2}$.

To estimate the first summand, note that $Y^2 - X^2 = r(X+Y)$, hence it does not exceed

$$\mathbb{E}|(X^2 - Y^2)(e^{itX} - 1)| \le \mathbb{E}\left(|r| \cdot (|X| + |Y|) \cdot 2|tX|^\varepsilon\right).$$

Moreover, as $|r| \le |X| + |Y|$ and hence $|r| \le |r|^\varepsilon (|X| + |Y|)^{1-\varepsilon}$, this summand does not exceed

$$\mathbb{E}\left(|r|^\varepsilon \cdot 2(|X| + |Y|)^2\right) \cdot |t|^\varepsilon,$$

and due to the Hölder inequality with the same exponents p, q as above it does not exceed $8\, C_r^\varepsilon C_X^2 \cdot |t|^\varepsilon$. Adding these estimates together, we obtain the desired (4.3.15). □

Now, apply these estimates to estimate the change of $N_\rho(\cdot)$:

Proposition 4.3.7. *Under the assumptions of Lemma 4.3.6, let $K_r := 5\, C_r^\varepsilon C_X^2$ be the factor that appears in its conclusion* (4.3.14)*. Assume additionally that for some $\rho > 0$ one has*

$$\rho^{2+\varepsilon} N_\rho(X) \le \frac{1}{100} \tag{4.3.17}$$

and

$$K_r \rho^{2+\varepsilon} e^{\frac{\rho^2}{2}} \le \frac{1}{100}. \tag{4.3.18}$$

Then

$$N_\rho(Y) \le N_\rho(X) + 2K_r e^{\frac{\rho^2}{2}}.$$

Proof. Note first that due to (4.3.17), for any $|t| \le \rho$ we have

$$\left|\log\left(\varphi_X(t)e^{\frac{t^2}{2}}\right)\right| \le \frac{1}{100}$$

and hence

$$\left|\varphi_X(t)e^{\frac{t^2}{2}} - 1\right| \le \frac{1}{50}$$

(as the exponent function is 2-Lipschitz in $U_{1/100}(0)$).

At the same time, the conclusion of Lemma 4.3.6 and the assumption (4.3.18) imply that for any $|t| \le \rho$ one has

$$\left|\varphi_X(t)e^{\frac{t^2}{2}} - \varphi_Y(t)e^{\frac{t^2}{2}}\right| = |\varphi_X(t) - \varphi_Y(t)| \cdot e^{\frac{t^2}{2}} \le K_r\rho^{2+\varepsilon} \cdot e^{\frac{\rho^2}{2}} \le \frac{1}{100},$$

hence altogether

$$\left|\varphi_Y(t)e^{\frac{t^2}{2}} - 1\right| \le \left|\varphi_X(t)e^{\frac{t^2}{2}} - \varphi_Y(t)e^{\frac{t^2}{2}}\right| + \left|\varphi_X(t)e^{\frac{t^2}{2}} - 1\right| \le \frac{1}{100} + \frac{1}{50} \le \frac{1}{25}.$$

Finally, the logarithm function is 2-Lipschitz in $U_{\frac{1}{25}}(1)$, and thus for such t

$$\begin{aligned}\left|\log\left(\varphi_Y(t)e^{\frac{t^2}{2}}\right)\right| &\le \left|\log\left(\varphi_X(t)e^{\frac{t^2}{2}}\right)\right| + 2\cdot\left|\varphi_Y(t)e^{\frac{t^2}{2}} - \varphi_X(t)e^{\frac{t^2}{2}}\right| \\ &\le N_\rho(X)|t|^{2+\varepsilon} + 2K_r e^{\frac{t^2}{2}}|t|^{2+\varepsilon} \le (N_\rho(X) + 2K_r e^{\frac{t^2}{2}})\cdot|t|^{2+\varepsilon}. \quad (4.3.19)\end{aligned}$$

This implies the desired

$$N_\rho(Y) \le N_\rho(X) + 2K_r e^{\frac{\rho^2}{2}}.$$

□

Let us now apply Proposition 4.3.7 to the family $\{\widetilde{\xi}_{a;n}\}$ that occurs in the setting of Theorem 4.2.1.

Corollary 4.3.8. *Under the assumptions of Theorem 4.2.1, there exists a constant K such that if for some $\rho > 0$, some a, some $n, n' \ge n_0$, with n_0 given by Lemma 4.2.5, as soon as $\frac{n}{2} \le n' \le 2n$, one has*

$$\rho^{2+\varepsilon} N'_\rho(\widetilde{\xi}_{a;n} + \widetilde{\xi}_{a+n;n'}) \le \frac{1}{100} \tag{4.3.20}$$

and

$$\rho^{2+\varepsilon} \frac{K}{(n+n')^{\varepsilon/2}} e^{\frac{\rho^2}{2}} \le \frac{1}{100}, \tag{4.3.21}$$

then

$$N'_\rho(\widetilde{\xi}_{a;n+n'}) \le N'_\rho(\widetilde{\xi}_{a;n} + \widetilde{\xi}_{a+n;n'}) + 2\frac{K}{(n+n')^{\varepsilon/2}} e^{\frac{\rho^2}{2}}.$$

Proof. Take X and Y, defined by (4.3.11), so that

$$N_\rho(X) = N'_\rho(\widetilde{\xi}_{a;n} + \widetilde{\xi}_{a+n;n'}), \quad N_\rho(Y) = N'_\rho(\widetilde{\xi}_{a;n+n'}).$$

Applying Lemma 4.3.5, from its conclusion (4.3.12) we get an estimate for the $L_{2+\varepsilon}$-norm of

their difference $r = r_{a;n,n'} = Y - X$:

$$C_r := \left(\mathbb{E}|r|^{2+\varepsilon}\right)^{1/(2+\varepsilon)} \leq \frac{Q_r}{\sqrt{n+n'}}.$$

Hence, the value $K_r = 5C_r^{\varepsilon}C_X^2$ in Proposition 4.3.7 is bounded from above by

$$K_r = 5\, C_r^{\varepsilon}C_X^2 \leq 5\, \left(\frac{Q_r}{\sqrt{n+n'}}\right)^{\varepsilon} \cdot C_X^2 = \frac{K}{(n+n')^{\varepsilon/2}},$$

where

$$K := 5\, Q_r^{\varepsilon}C_X^2.$$

The conclusion then immediately follows from Proposition 4.3.7. □

4.4 Proof of the abstract CLT

Joining the results of the previous sections, we obtain the following proposition:

Proposition 4.4.1. *Under the assumptions of Theorem 4.2.1, there exist sequences $\rho_n \to \infty$, $\delta_n \to 0$, such that for any $n \geq n_0$, where n_0 is given by Lemma 4.2.5, and any a, one has*

$$N'_{\rho_n}(\widetilde{\xi}_{a;n}) \leq \delta_n. \tag{4.4.1}$$

Prior to proving it, note that Theorem 4.2.1 follows from Proposition 4.4.1 almost immediately.

Proof of Theorem 4.2.1. Assume that the assumptions of Theorem 4.2.1 are satisfied. Due

to Proposition 4.4.1, for any $\rho > 0$ for all sufficiently large n we have $\rho_n \geq \rho$, and hence

$$\lim_{n\to\infty} N'_\rho(\widetilde{\xi}_{a;n}) = \lim_{n\to\infty} N_\rho(\eta_{a;n}) = 0,$$

where

$$\eta_{a;n} = \frac{\widetilde{\xi}_{a;n}}{\sqrt{\operatorname{Var}\widetilde{\xi}_{a;n}}}.$$

In particular, the characteristic functions $\varphi_{\eta_{a;n}}(t)$ of the normalized variables converge uniformly on compact sets to $e^{-\frac{t^2}{2}}$. As the weak convergence of random variables is equivalent (Lévy's continuity theorem) to the pointwise convergence of their characteristic functions, we have the desired weak convergence

$$\eta_{a;n} = \frac{\widetilde{\xi}_{a;n}}{\sqrt{\operatorname{Var}\widetilde{\xi}_{a;n}}} \to \mathcal{N}(0,1), \quad n \to \infty.$$

Moreover, this convergence is actually uniform in a, as the convergence of the characteristic functions is uniform in a on any compact interval $[-\rho, \rho]$ due to the uniform estimate (4.4.1).

□

Proof of Proposition 4.4.1. We are going to construct the sequence $(\rho_n, \delta_n)_{n\geq n_0}$ so that the desired property (4.4.1) can be established by induction on n. Namely, we have the following

Lemma 4.4.2 (Bootstrapping)**.** *Let the sequence $(\rho_n, \delta_n)_{n\geq n_0}$, where n_0 is given by Lemma 4.2.5, be chosen in such a way that the following conditions hold:*

- *As $n \to \infty$, one has $\rho_n \to \infty$ and $\delta_n \to 0$.*

- *For some $n_1 \geq 2n_0$, we have*

$$\rho_n = \rho_0', \quad \delta_n = 3\overline{C}^{2+\varepsilon} \quad \textit{for all} \quad n = n_0, \ldots, n_1,$$

where $\overline{C}$ is given by Corollary 4.2.6,

$$\rho_0' = \min\left(\rho_0, \frac{1}{20\overline{C}}\right)$$

and ρ_0 is given by Lemma 4.3.3.

- *For every $m > 2n_0$, taking $n = \lfloor \frac{m}{2} \rfloor$ and $n' = m - n$, one has*

$$\rho_m \leq L \min(\rho_n, \rho_{n'}), \tag{4.4.2}$$

$$\rho_m^{2+\varepsilon}\delta_m \leq \frac{1}{100}, \tag{4.4.3}$$

$$\left(\lambda \max(\delta_n, \delta_{n'}) + 2\,\frac{K e^{\frac{\rho_m^2}{2}}}{m^{\varepsilon/2}}\right) \leq \delta_m, \tag{4.4.4}$$

where constants L and λ are defined by the conclusion of Lemma 4.3.4 for $C = 2\left(\frac{C_\sigma}{C_1}\right)^2$, where C_1 is given by Lemma 4.2.5, and C_σ is given by (4.2.10).

Then the conclusion of Proposition 4.4.1 holds for this sequence.

Proof. The proof of (4.4.1) is by induction. Namely, the base of the induction is formed by $m = n_0, \ldots, n_1$, where the inequality

$$N'_{\rho_m}(\widetilde{\xi}_{a;m}) \leq \delta_m$$

follows from the choice of ρ_0' and Lemma 4.3.3.

Let us make the induction step. Namely, for $m > n_1$ let $n = \lceil \frac{m}{2} \rceil$ and $n' = m - n$; then,

$n_0 \le n, n' < m$, so the conclusion is already established for these indices. Hence, due to the induction assumption and inequality (4.4.2),

$$N'_{\rho_m/L}(\widetilde{\xi}_{a;n}) \le N'_{\rho_n}(\widetilde{\xi}_{a;n}) \le \delta_n \quad \text{and} \quad N'_{\rho_m/L}(\widetilde{\xi}_{a+n;n'}) \le N'_{\rho_{n'}}(\widetilde{\xi}_{a+n;n'}) \le \delta_{n'}.$$

and thus due to the conclusion (4.3.8) of Lemma 4.3.4,

$$N'_{\rho_m}(\widetilde{\xi}_{a;n} + \widetilde{\xi}_{a+n;n'}) \le \lambda \max(\delta_n, \delta_{n'}). \tag{4.4.5}$$

Now, we are going to apply Corollary 4.3.8 for ρ_m, n, n'. First, let us check that its assumptions are satisfied. Indeed, multiplying (4.4.5) by $\rho_m^{2+\varepsilon}$ and applying (4.4.3) and (4.4.4), we get

$$\rho_m^{2+\varepsilon} N'_{\rho_m}(\widetilde{\xi}_{a;n} + \widetilde{\xi}_{a+n;n'}) \le \rho_m^{2+\varepsilon}\delta_m \le \frac{1}{100},$$

so the assumption (4.3.20) is satisfied. Next, assumption (4.3.21) follows again from (4.4.4) and (4.4.3):

$$\frac{K}{m^{\varepsilon/2}}\rho_m^{2+\varepsilon} e^{\frac{\rho_m^2}{2}} \le \rho_m^{2+\varepsilon}\delta_m \le \frac{1}{100}.$$

Corollary 4.3.8 is thus applicable, and hence (again using (4.4.4)) we get

$$N'_{\rho_m}(\widetilde{\xi}_{a;m}) \le \left(\lambda \max(\delta_n, \delta_{n'}) + 2\frac{K e^{\frac{\rho_m^2}{2}}}{m^{\varepsilon/2}}\right) \le \delta_m.$$

The induction step is complete. □

To complete the proof of Proposition 4.4.1, it remains to construct the sequences

$$\rho_m \to \infty, \quad \delta_m \to 0$$

that satisfy the assumptions of Lemma 4.4.2. Roughly speaking, if we were keeping the radii ρ_m constant, the contraction with the factor λ would effectively allow to bring δ_m to zero as the additional term $\frac{2Ke^{\frac{\rho_m^2}{2}}}{m^{\varepsilon/2}}$ then also tends to zero. It suffices now to make the radii ρ_m increase extremely slowly, so that the exponent $e^{\frac{\rho_m^2}{2}}$ would not break this asymptotic vanishing.

Following this idea, we will choose ρ_m so that

$$\rho_m \leq \frac{1}{2}\sqrt{\varepsilon \log m};$$

such a restriction allows to use

$$\frac{Ke^{\frac{\rho_m^2}{2}}}{m^{\varepsilon/2}} \leq \frac{Km^{\varepsilon/8}}{m^{\varepsilon/2}} < \frac{K}{m^{\varepsilon/4}} \tag{4.4.6}$$

when checking that inequality (4.4.4) holds. Next, we let

$$\delta_m = Am^{-\beta}, \quad m > n_1, \tag{4.4.7}$$

where the constant A is chosen so that at $m = n_1$ this value coincides with $3\overline{C}^{2+\varepsilon}$,

$$A = 3\overline{C}^{2+\varepsilon} \cdot n_1^{\beta}, \tag{4.4.8}$$

and the (sufficiently small) power $\beta > 0$ and the (sufficiently large) initial index n_1 are yet to be fixed.

Now, choose the exponent $\beta > 0$ sufficiently small so that

$$\lambda \cdot 2^{\beta} < 1, \quad \beta < \frac{\varepsilon}{4},$$

and fix $\lambda' \in (2^{\beta}\lambda, 1)$.

Then, for all sufficiently large n_1 the condition (4.4.4) holds and can be proved by induction. Indeed, in the left hand side the first summand is

$$\lambda \max(\delta_n, \delta_{n'}) \leq \lambda \cdot A \left(\frac{m-1}{2}\right)^{-\beta} = 2^{\beta}\lambda \cdot Am^{-\beta} \cdot \left(\frac{m-1}{m}\right)^{-\beta} < \lambda'\delta_m.$$

The second summand due to (4.4.6) is at most $K \cdot m^{-\varepsilon/4}$, thus it suffices to check for $m > n_1$ the inequality

$$\lambda' Am^{-\beta} + 2Km^{-\varepsilon/4} < Am^{-\beta},$$

which can be rewritten as

$$(1-\lambda')Am^{-\beta} > 2Km^{-\varepsilon/4}. \tag{4.4.9}$$

As $\beta < \frac{\varepsilon}{4}$, due to the monotonicity it suffices to check (4.4.9) for $m = n_1$. Now, recall that (4.4.8) is used to determine A for given n_1 and β; substituting the value of A from (4.4.8), we see that (4.4.9) holds for $m = n_1$ once n_1 is sufficiently large to ensure

$$(1-\lambda') \cdot 3\overline{C}^{2+\varepsilon} > 2Kn_1^{-\varepsilon/4}$$

We fix a sufficiently large n_1 so that (4.4.9) holds, fix the corresponding A (defined by (4.4.8)) and the sequence (δ_m), defined for $m > n_1$ by (4.4.7). Then, we use (4.4.3) and (4.4.2) to

choose the sequence (ρ_m). Namely, for $m > n_1$ we let

$$\rho_m = \min\left(L\min(\rho_n, \rho_{n'}), \frac{1}{2}\sqrt{\varepsilon \log m}, (100\delta_m)^{-1/(2+\varepsilon)} \right). \tag{4.4.10}$$

Then the inequality $\rho_m \le (100\delta_m)^{-1/(2+\varepsilon)}$ implies (4.4.3), and (4.4.2) is satisfied automatically. Finally, as

$$\min\left(\frac{1}{2}\sqrt{\varepsilon \log m}, (100\delta_m)^{-1/(2+\varepsilon)} \right) \to \infty, \quad m \to \infty,$$

the sequence (ρ_m) defined by (4.4.10) also tends to infinity; actually, it will coincide with $\frac{1}{2}\sqrt{\varepsilon \log m}$ for all sufficiently large m. This completes the proof of Proposition 4.4.1. □

4.5 Moment estimates for $R_{n,n'}$

Let us now return to the setting of Theorem 1.2.6, the one of random matrix products. We start by establishing the bounds for the moments of discrepancies $R_{n,n'}$ that are defined by (4.1.2). To do so, we first discuss some properties of $\mathrm{SL}(2,\mathbb{R})$ matrices in Section 4.5.1, and then state and prove the main estimate, Proposition 4.5.6, in Section 4.5.2.

4.5.1 Preliminaries: action of $\mathrm{SL}(2,\mathbb{R})$ matrices

Let $[\cdot]$ be the canonical projection $[\cdot] : \mathbb{R}^2\backslash\{0\} \to \mathbb{RP}^1$, and denote by f_B the projectivization of the matrix B, namely,

$$f_B : \mathbb{RP}^1 \to \mathbb{RP}^1, \quad \text{such that } f_B \circ [\cdot] = [\cdot] \circ B. \tag{4.5.1}$$

We consider $\mathbb{RP}^1$ to be equipped with the metric $\operatorname{dist}(\cdot,\cdot)$ that is the angle between the corresponding lines.

Recall that every matrix $B \in \mathrm{SL}(2,\mathbb{R})$ can be written as a product

$$B = \mathrm{Rot}_{\beta_1} \begin{pmatrix} \|B\| & 0 \\ 0 & \|B\|^{-1} \end{pmatrix} \mathrm{Rot}_{\beta_2}, \tag{4.5.2}$$

where Rot_β is a rotation by the angle β. Let e_1, e_2 be the standard basis vectors of $\mathbb{R}^2$; for a matrix B in the form (4.5.2), denote by $\mathbf{r}(B) = [\mathrm{Rot}_{\beta_2}^{-1} e_2] \in \mathbb{RP}^1$ the direction that is *contracted* the most by B. Finally, for a vector $v \in \mathbb{R}^2$ and a matrix $B \in \mathrm{SL}(2,\mathbb{R})$, let

$$\Theta(B, v) = \log(\|B\| \cdot |v|) - \log(|Bv|) = \log \frac{\|B\| \cdot |v|}{|Bv|};$$

in other words, $\Theta(B, v)$ is a function that compares the expansion by B of the vector v with the maximal possible expansion by B over all the nonzero vectors.

We have the following estimate:

Lemma 4.5.1. *For any $B \in \mathrm{SL}(2,\mathbb{R})$ and any nonzero vector $v \in \mathbb{R}^2$, one has*

$$\Theta(B, v) \le -\log \sin \operatorname{dist}([v], \mathbf{r}(B)). \tag{4.5.3}$$

Proof. We can assume the vector v to be of unit length. Distance $\operatorname{dist}([v], \mathbf{r}(B))$ is equal to the angle between $\mathrm{Rot}_{\beta_2}(v)$ and e_2, thus the component of $\mathrm{Rot}_{\beta_2}(v)$ that is parallel to e_1 is equal to $\sin \operatorname{dist}([v], \mathbf{r}(B))$; see Fig. 4.1. After the application of the diagonal matrix in the representation (4.5.2), this component gets multiplied by $\|B\|$, and provides a lower bound for the length of the image $|Bv|$. This immediately implies (4.5.3). □

Lemma 4.5.1 immediately implies the following estimate:

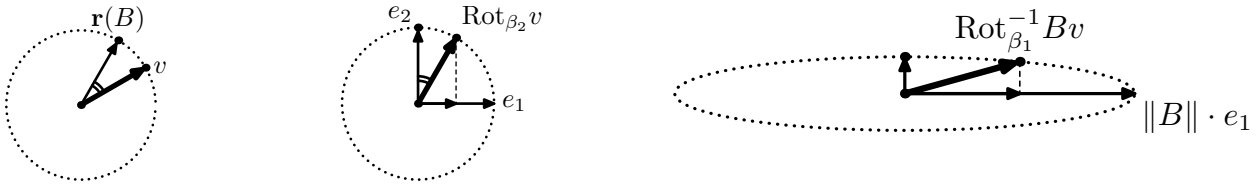

Figure 4.1: Left: mostly contracted direction $\mathbf{r}(B)$ and a vector v. Center: their images after rotation by Rot_{β_2}. Right: the images after the application of the diagonal matrix in the decomposition (4.5.2).

Corollary 4.5.2. *For any $B \in \mathrm{SL}(2,\mathbb{R})$ for at least one of two coordinate vectors e_i one has $\Theta(B, e_i) \le \log\sqrt{2}$.*

Proof. Indeed, at least one of the points $[e_1], [e_2]$ is at the distance at least $\frac{\pi}{4}$ from the direction $\mathbf{r}(B)$, and the estimate follows from (4.5.3). □

Lemma 4.5.1 involves the location of the direction $\mathbf{r}(B)$; the following statement gives a way to approximate its location:

Lemma 4.5.3. *For any $B \in \mathrm{SL}(2,\mathbb{R})$ for at least one of two coordinate vectors e_j one has*

$$\mathrm{dist}(f_B^{-1}[e_j], \mathbf{r}(B)) \le \frac{1}{\|B\|^2}. \tag{4.5.4}$$

Proof. One of the vectors $\mathrm{Rot}_{\beta_1}^{-1} e_j$ is at the angle $\theta \ge \frac{\pi}{4}$ with the direction e_1. Now,

$$\begin{pmatrix} \|B\| & 0 \\ 0 & \|B\|^{-1} \end{pmatrix}^{-1} \mathrm{Rot}_{\beta_1}^{-1} e_j = \begin{pmatrix} \|B\| & 0 \\ 0 & \|B\|^{-1} \end{pmatrix}^{-1} \begin{pmatrix} \cos\theta \\ \pm\sin\theta \end{pmatrix} = \begin{pmatrix} \frac{1}{\|B\|}\cos\theta \\ \pm\|B\|\cdot\sin\theta \end{pmatrix}. \tag{4.5.5}$$

The tangent of the angle θ' between the vector (4.5.5) and the direction of e_2 is thus equal to $\tan\theta' = \frac{\mathrm{cotan}\,\theta}{\|B\|^2} \le \frac{1}{\|B\|^2}$. An application of $\mathrm{Rot}_{\beta_2}^{-1}$ then concludes the proof:

$$\mathrm{dist}(f_B^{-1}[e_j], \mathbf{r}(B)) = \theta' \le \tan\theta' \le \frac{1}{\|B\|^2}.$$

□

Next, note that one can estimate the decrease in the log-norm in the product of two matrices B_1 and B_2 using their action on any (in particular, well-chosen) vector v:

Lemma 4.5.4. *For any $B_1, B_2 \in \mathrm{SL}(2,\mathbb{R})$ and any nonzero vector v,*

$$\log \|B_1\| + \log \|B_2\| - \log \|B_2 B_1\| \le \Theta(B_1, v) + \Theta(B_2, B_1 v). \tag{4.5.6}$$

Proof. We can assume the vector v to be a unit one. Then,

$$\begin{aligned}\log \|B_2 B_1\| \ge \log |B_2 B_1 v| = \log \|B_2\| + \log |B_1 v| - \Theta(B_2, B_1 v) \\ = \log \|B_2\| + \log \|B_1\| - \Theta(B_1, v) - \Theta(B_2, B_1 v).\end{aligned}$$

□

Finally, the previous lemmas can be joined together in order to obtain a good estimate for the right hand side of (4.5.6). Namely, for any $B_1, B_2 \in \mathrm{SL}(2,\mathbb{R})$ consider the images of the coordinate directions $f_{B_1}([e_i])$, $i = 1, 2$, and preimages $f_{B_2}^{-1}([e_j])$, $j = 1, 2$. Let Δ_{B_1,B_2} be the minimal distance between these pairs:

$$\Delta_{B_1,B_2} := \min_{i=1,2} \min_{j=1,2} \operatorname{dist}(f_{B_1}([e_i]), f_{B_2}^{-1}([e_j])). \tag{4.5.7}$$

We then have the following estimate:

Proposition 4.5.5. *For any $B_1, B_2 \in \mathrm{SL}(2,\mathbb{R})$, one has*

$$\log \|B_1\| + \log \|B_2\| - \log \|B_2 B_1\| \le \log 4\sqrt{2} + \log \min(\Delta_{B_1,B_2}^{-1}, \|B_2\|^2). \tag{4.5.8}$$

Proof. Due to Corollary 4.5.2, we can choose a coordinate vector e_i so that $\Theta(B_1, e_i) \le$

$\log\sqrt{2}$. Due to Lemma 4.5.4, the left hand side of (4.5.8) does not exceed

$$\Theta(B_1, e_i) + \Theta(B_2, B_1 e_i) \le \log\sqrt{2} + \Theta(B_2, B_1 e_i). \tag{4.5.9}$$

Note that for any nonzero vector v (in particular, for $v = B_1 e_i$), one has $|B_2 v| \ge \frac{1}{\|B_2\|}|v|$ and hence

$$\Theta(B_2, v) \le \log \|B_2\|^2.$$

Thus, if $\Delta^{-1}_{B_1,B_2} \ge \frac{1}{2}\|B_2\|^2$, the estimate (4.5.9) implies the desired (4.5.8) immediately: its right hand side then does not exceed

$$\log\sqrt{2} + \Theta(B_2, B_1 e_i) \le \log\sqrt{2} + \log(2 \cdot \frac{1}{2}\|B_2\|^2) \le \log 2\sqrt{2} + \log\min(\Delta^{-1}_{B_1,B_2}, \|B_2\|^2).$$

Given that, from now on we can assume that $\Delta^{-1}_{B_1,B_2} < \frac{1}{2}\|B_2\|^2$, and the proof will be complete once we establish that

$$\Theta(B_2, B_1 e_i) \le \log 4 + \log \Delta^{-1}_{B_1,B_2}. \tag{4.5.10}$$

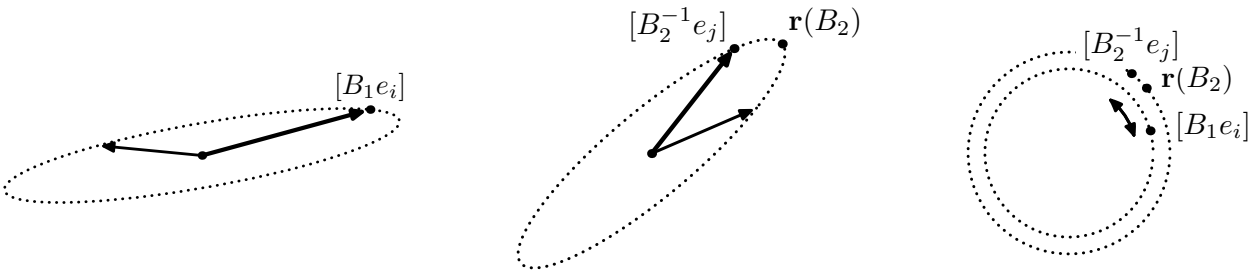

Figure 4.2: Top left: image $[B_1 e_i]$ provided by Corollary 4.5.2. Top right: the direction $\mathbf{r}(B_2)$ and the preimage $[B_2^{-1} e_j]$, sufficiently close to it, that is provided by Lemma 4.5.3. Bottom: these three directions and an arc of length at least Δ_{B_1,B_2}.

Now, let e_j be the coordinate vector for which the estimate (4.5.4) from Lemma 4.5.3 holds

(see Fig. 4.2). Then for $v = f_{B_1}(e_i)$, by triangle inequality and definition of Δ_{B_1,B_2} one has

$$\begin{aligned}\operatorname{dist}([v], \mathbf{r}(B_2)) &\geq \operatorname{dist}(f_{B_1}([e_i]), f_{B_2}^{-1}([e_j])) - \operatorname{dist}(f_{B_2}^{-1}([e_j]), \mathbf{r}(B_2)) \\ &\geq \Delta_{B_1,B_2} - \frac{1}{\|B_2\|^2} \geq \frac{1}{2}\Delta_{B_1,B_2},\end{aligned}$$

where we have used $\Delta_{B_1,B_2} > \frac{2}{\|B_2\|^2}$ for the last inequality. Applying Lemma 4.5.1, we finally get

$$\Theta(B_2, B_1 e_i) \leq -\log\sin\operatorname{dist}([v], \mathbf{r}(B_2)) \leq -\log\sin\frac{1}{2}\Delta_{B_1,B_2} \leq \log 4 + \log\Delta_{B_1,B_2}^{-1},$$

where we have used the inequality $\sin x \geq \frac{2}{\pi}x \geq \frac{x}{2}$ for $x \in [0, \frac{\pi}{2}]$. This completes the proof of (4.5.10) and thus of the proposition. □

4.5.2 Estimates

The main statement of this section is the following estimate for the moments of the random variables $R_{n,n'}$. As $\gamma > 2$, we can choose $\varepsilon \in (0, 1]$ such that

$$2 + \varepsilon < \gamma.$$

We fix a choice of such ε from now on until the end of the paper.

Proposition 4.5.6. *Under the assumptions of Theorem 1.2.6, there exists C_R, such that for every $n, n' \in \mathbb{N}$ with $\frac{n'}{2} \leq n \leq 2n'$, and any measures $\mu_1, \ldots, \mu_{n+n'} \in \mathcal{K}$, one has*

$$\mathbb{E}R_{n,n'} < C_R, \quad \mathbb{E}R_{n,n'}^2 < C_R^2, \quad \textit{and} \quad \mathbb{E}R_{n,n'}^{2+\varepsilon} < C_R^{2+\varepsilon}.$$

Proof of Proposition 4.5.6. First of all, notice that it suffices to prove the estimate for $\mathbb{E}R_{n,n'}^{2+\varepsilon}$,

as it implies the other two by using Jensen inequality: for every $p \in [1, 2+\varepsilon]$ one has

$$\mathbb{E}R_{n,n'}^p = \mathbb{E}(R_{n,n'}^{2+\varepsilon})^{\frac{p}{2+\varepsilon}} \leq (\mathbb{E}R_{n,n'}^{2+\varepsilon})^{\frac{p}{2+\varepsilon}} < (C_R^{2+\varepsilon})^{\frac{p}{2+\varepsilon}} = C_R^p.$$

Now, the random variable $R_{n,n'}$ has the form

$$R_{n,n'} = \log \|B_1\| + \log \|B_2\| - \log \|B_2 B_1\|,$$

where $B_1 = T_n$ and $B_2 = T_{(n,n+n']}$. Proposition 4.5.5 then implies

$$R_{n,n'} \leq \log 4\sqrt{2} + \log \min(\Delta^{-1}_{T_n, T_{(n,n+n']}}, \|T_{(n,n+n']}\|^2) \tag{4.5.11}$$

In what follows we will use the following regularity result from Chapter 3. The setting of Chapter 3 is a general setting of non-stationary random dynamics: one assumes that a compact set $\mathcal{K}_{\mathcal{M}}$ of probability measures on Homeo($\mathcal{M}$) for some compact manifold $\mathcal{M}$ is given, and that these measures are concentrated on the bi-Lipschitz maps. It is also assumed that all these measures satisfy no deterministic image condition, and that the Lipschitz constant admits a uniformly bounded γ-th log-moment:

$$\forall \mu \in \mathcal{K}_{\mathcal{M}} \quad \int_{\mathrm{Lip}(\mathcal{M})} (\log \max(\mathrm{Lip}(f), \mathrm{Lip}(f^{-1})))^\gamma \, \mathrm{d}\mu(f) < C_{\mathcal{K}_{\mathcal{M}}}.$$

Then, the following estimates hold:

Theorem 4.5.7 (partial case of Theorem 3.5.1)**.** *Under the assumptions above, there exist constants $\kappa > 1$ and $C_{\mathcal{M}}$, such that for any two initial probability measures $\nu_0^{(1)}$ and $\nu_0^{(2)}$ on $\mathcal{M}$ and any two sequences of iterations,*

$$\mu_1^{(1)}, \ldots, \mu_{n_1}^{(1)}, \mu_1^{(2)}, \ldots, \mu_{n_2}^{(2)} \in \mathcal{K}_{\mathcal{M}},$$

for the random images

$$\nu_1 = \mu_{n_1}^{(1)} * \cdots * \mu_1^{(1)} * \nu_0^{(1)}, \quad \nu_2 = \mu_{n_2}^{(2)} * \cdots * \mu_1^{(2)} * \nu_0^{(2)} \tag{4.5.12}$$

one has a uniform bound for γ-th log moment for the distance between the random points with a cut-off at radius $\theta_{n_1,n_2} := \exp(-\kappa^{\min(n_1,n_2)})$:

$$\iint_{\mathcal{M}\times\mathcal{M}} |\log \max(d(x,y), \theta_{n_1,n_2})|^\gamma \; d\nu_1 \, d\nu_2 < C_{\mathcal{M}}. \tag{4.5.13}$$

In particular, the Markov inequality for (4.5.13) immediately implies a log-Hölder-type bound for the distances between these images

$$(\nu_1 \times \nu_2)\{(x,y) \mid d(x,y) \le r\} < C_{\mathcal{M}} \cdot |\log r|^{-\gamma} \tag{4.5.14}$$

for every $r \ge \theta_{n_1,n_2}$.

In our case, $\mathcal{M}$ will be the projective line $\mathbb{RP}^1$ and the maps will be the projectivisations of linear maps of $\mathbb{R}^2$. Namely, to a probability measure $\mu \in \mathcal{K}$ on $\mathrm{SL}(2,\mathbb{R})$ we associate *two* probability measures that are its pushforwards by the maps from $\mathrm{SL}(2,\mathbb{R})$ to $\mathrm{Homeo}(\mathbb{RP}^1)$,

$$F_1 : A \mapsto f_A \quad \text{and} \quad F_2 : A \mapsto f_A^{-1},$$

and we let the set $\mathcal{K}_{\mathcal{M}}$ be formed by these images,

$$\mathcal{K}_{\mathcal{M}} := \{(F_1)_*\mu \mid \mu \in \mathcal{K}\} \cup \{(F_2)_*\mu \mid \mu \in \mathcal{K}\}.$$

Then it satisfies the assumptions of Theorem 4.5.7: the set $\mathcal{K}_{\mathcal{M}}$ is compact as a union of two continuous images of a compact $\mathcal{K}$, the moments assumption follows from the moments assumption for $\mathcal{K}$, and absence of a deterministic image follows from the similar condition

for the measures from $\mathcal{K}$.

We will take initial measures

$$\nu_0^{(1)} = \nu_0^{(2)} = \frac{1}{2}\delta_{[e_1]} + \frac{1}{2}\delta_{[e_2]}.$$

Then, taking $n_1 := n$ and the sequence $\mu_j^{(1)} := (F_1)_*\mu_j$, we see that the image measure ν_1 is the half-sum of the laws of $[T_n e_1]$ and of $[T_n e_2]$. In the same way, taking $n_2 := n'$ and the sequence

$$\mu_j^{(2)} := (F_2)_*\mu_{n+n'-(j-1)}, \quad j = 1, \dots, n',$$

we see that the image measure ν_2 is the half-sum of the laws of $[T^{-1}_{(n,n+n']}e_1]$ and of $[T^{-1}_{(n,n+n']}e_2]$. An application of Theorem 4.5.7 thus yields a uniform upper bound

$$\frac{1}{4}\sum_{i=1,2}\sum_{j=1,2}\mathbb{E}\left|\log\max\left\{\operatorname{dist}\left([T_n e_i], [T^{-1}_{(n,n+n']}e_j]\right), \theta_{n,n'}\right\}\right|^\gamma \le C_{\mathcal{M}},$$

and hence (replacing the sum over i and j with the maximal summand)

$$\mathbb{E}\left|\log\max\left\{\Delta_{T_n, T_{(n,n+n']}}, \theta_{n,n'}\right\}\right|^\gamma \le 4C_{\mathcal{M}},$$

Now, in the right hand side of (4.5.11) we have a different maximum: instead of a cut-off at a fixed threshold $\theta_{n,n'}$, the norm of the product $T_{(n,n+n']}$ appears. To handle it, note that if $\Delta^{-1}_{T_n, T_{(n,n+n']}} \le \theta^{-1}_{n,n'}$, then

$$\min(\Delta^{-1}_{T_n, T_{(n,n+n']}}, \|T_{(n,n+n']}\|^2) \le \Delta^{-1}_{T_n, T_{(n,n+n']}} = \min(\Delta^{-1}_{T_n, T_{(n,n+n']}}, \theta^{-1}_{n,n'}),$$

while if $\Delta^{-1}_{T_n,T_{(n,n+n']}} > \theta^{-1}_{n,n'}$, then

$$\min(\Delta^{-1}_{T_n,T_{(n,n+n']}}, \|T_{(n,n+n']}\|^2) \le \|T_{(n,n+n']}\|^2;$$

applying the logarithm and joining these two inequalities, we get

$$\log\min(\Delta^{-1}_{T_n,T_{(n,n+n']}}, \|T_{(n,n+n']}\|^2) \le \left|\log\max(\Delta_{T_n,T_{(n,n+n']}}, \theta_{n,n'})\right| + \\ + (\log\|T_{(n,n+n']}\|^2)\cdot \mathbb{1}_{\{\Delta_{T_n,T_{(n,n+n']}}<\theta_{n,n'}\}}. \quad (4.5.15)$$

The $(2+\varepsilon)$-th moment of second summand can be then estimated by the Hölder inequality with the exponents $p = \frac{\gamma}{2+\varepsilon} > 1$ and $q = \frac{1}{1-\frac{1}{p}}$:

$$\mathbb{E}\left[(2\log\|T_{(n,n+n']}\|)^{2+\varepsilon}\cdot \mathbb{1}_{\{\Delta_{T_n,T_{(n,n+n']}}<\theta_{n,n'}\}}\right] \\ \le \left(\mathbb{E}(2\log\|T_{(n,n+n']}\|)^{\gamma}\right)^{1/p}\cdot\left(\mathbb{P}\left\{\Delta_{T_n,T_{(n,n+n']}} < \theta_{n,n'}\right\}\right)^{1/q} \\ \le (2n')^{2+\varepsilon}M^{1/p}\cdot(4C_{\mathcal{M}}\cdot|\log\theta_{n,n'}|^{-\gamma})^{1/q} \quad (4.5.16)$$

Here we have noticed that $\log\|T_{(n,n+n']}\|$ does not exceed a sum of n' summands of the form $\log\|A_j\|$, the γ-th moment of each of which does not exceed M, and used the Markov inequality (4.5.14) to estimate the probability in the second factor. Now, as $|\log\theta_{n,n'}| = \kappa^{\min(n,n')}$ grows *exponentially* in n' (recall that $n \le 2n'$ due to the assumption of the proposition), the right hand side of (4.5.16) is uniformly bounded (and actually tends to 0 as n' grows).

We have shown that both random variables in the right hand side of (4.5.15) have uniformly bounded $(2+\varepsilon)$-th moments, and hence the same applies to $R_{n,n'}$: the upper bound (4.5.11) for it differs by an addition of a constant $\log 4\sqrt{2}$. □

4.6 Proof of Theorem 1.2.6

In this section we verify the assumptions ((a))–((e)) in the setting of Theorem 1.2.6, except for the unboundedness of variance, which is stated in Proposition 4.6.2 below and is proven in the next section. We then deduce Theorem 1.2.6 from Theorem 4.2.1.

Lemma 4.6.1. *Under the assumptions of Theorem 1.2.6, the family of random variables* $\{\widetilde{\xi}_{a;n}\}$*, defined by* (4.2.1)*, satisfies the assumptions ((a))–((e)) above.*

Proof. The only non-trivial conclusions are the estimate ((d)) and the lower bound for the variance ((e)). To ensure ((d)), recall that Proposition 4.5.6 guarantees for $\frac{n'}{2} \le n \le 2n'$ the bound $\mathbb{E}R_{a;n,n'}^{2+\varepsilon} \le C_R^{2+\varepsilon}$, where

$$R_{a;n,n'} := \xi_{(a,a+n]} + \xi_{(a+n,a+n+n']} - \xi_{(a,a+n+n']} \ge 0.$$

Now, this (due to the Jensen's inequality for $x^{1/(2+\varepsilon)}$) implies $\mathbb{E}R_{a;n,n'} \le C_R$, and hence the random variable $\widetilde{R}_{a;n,n'} = R_{a;n,n'} - \mathbb{E}R_{a;n,n'}$ is the difference between two random variables, whose $L_{2+\varepsilon}(\mathbb{P})$-norm does not exceed C_R. The desired (4.2.4) then follows from the triangle inequality for $L_{2+\varepsilon}$-norm after we redefine $C_R := 2C_R$ for consistency.

Verifying assumption ((e)) turns out to be a surprisingly technical endeavour. We give the corresponding formal statement about the matrices here and postpone its proof until Section 4.7.

Proposition 4.6.2. *Under the assumptions of Theorem 1.2.6, for any* $c > 0$ *there exists* $n_1 \in \mathbb{N}$ *such that for any* $n \ge n_1$ *and any collection of distributions* $\mu_1, \dots, \mu_n \in \mathcal{K}$ *one has*

$$\mathrm{Var}_{\mu_1,\dots,\mu_n} \xi_n \ge c.$$

From Proposition 4.6.2 and formula (4.2.1) the assumption ((e)) follows immediately.

□

We conclude this section with the proof of our main result, Theorem 1.2.6.

Proof of Theorem 1.2.6. Assume that the assumptions of Theorem 1.2.6 are satisfied. Then, due to Lemma 4.6.1, the family $\widetilde{\xi}_{a;n}$, defined by (4.2.1), satisfies the assumptions ((a))–((d)), while Proposition 4.6.2 guarantees that this family satisfies the assumption ((e)). Hence, for this family the conclusions of Theorem 4.2.1 hold, implying that random variables

$$\frac{\xi_n - \mathbb{E}\xi_n}{\sqrt{\operatorname{Var}\xi_n}} = \frac{\widetilde{\xi}_{0;n}}{\sqrt{\operatorname{Var}\widetilde{\xi}_{0;n}}}$$

weakly converge to $\mathcal{N}(0,1)$. Moreover, the speed of convergence in Theorem 4.2.1 is regulated by the sequences (ρ_m, δ_m), constructed in Lemma 4.4.2. These sequences require for their construction only constants C_R, C_2, n_0, etc., that can be chosen independently of the sequence of measures $\mu_1, \mu_2, \dots \in \mathcal{K}$, as such uniformity holds for all the statements in Section 4.5. Hence, the convergence of normalised measures in Theorem 1.2.6 is also uniform in the choice of the sequence $\mu_1, \mu_2, \dots \in \mathcal{K}$. □

4.7 Unboundedness of Variance: proof of Proposition 4.6.2

In this section we prove Proposition 4.6.2, i.e. show that under the assumptions of Theorem 1.2.6 variances of ξ_n become arbitrarily large. This completes the proof of Theorem 1.2.6.

In order to do so, we will assume that n is quite large, and will decompose the full product $A_n \dots A_1$ into a several "long" groups $D_{m+1}, \dots, D_1$, between which some "short" composi-

tions are applied:

$$A_n \dots A_1 = D_{m+1}(B_{\mu_{n_0,m}} \dots B_{\mu_{1,m}})D_m \dots D_2(B_{\mu_{n_0,1}} \dots B_{\mu_{1,1}})D_1.$$

We will show that (for an appropriate choice of lengths) even conditionally to all $D_1, \dots, D_{m+1}$, the distribution of the log-norm of the product (with high probability) has sufficiently high variance. At the same time, dividing by the product of norms of D_j, we get the composition

$$\frac{D_{m+1}}{\|D_{m+1}\|}(B_{\mu_{n_0,m}} \dots B_{\mu_{1,m}})\frac{D_m}{\|D_m\|} \dots \frac{D_2}{\|D_2\|}(B_{\mu_{n_0,1}} \dots B_{\mu_{1,1}})\frac{D_1}{\|D_1\|},$$

where all the quotients $\frac{D_j}{\|D_j\|}$ are almost rank-one matrices.

Therefore, we first consider the variance of a distribution of images of a given vector under random linear maps of rank one. In this case it is easier to show that the variance grows, see Lemma 4.7.1 and Lemma 4.7.6 below. By continuity, if one replaces random rank one linear maps by random linear maps of large norm, and uses the fact that for a matrix $D \in \mathrm{SL}(2, \mathbb{R})$ with large norm, $\frac{D}{\|D\|}$ is close to a linear map of rank one and norm one, then a lower bound on variances still holds, see Lemma 4.7.7 and Corollary 4.7.8. Finally, we can complete the proof of Proposition 4.6.2 by applying the fact that with large probability a composition of a long enough sequence of random $\mathrm{SL}(2, \mathbb{R})$ matrices has a large norm.

Let us now realize this strategy.

Let $Y \subseteq \mathrm{Mat}_2(\mathbb{R})$ be the space of all linear maps $\mathbb{R}^2 \to \mathbb{R}^2$ of norm 1 and of rank 1. Notice that Y is homeomorphic to the torus $\mathbb{T}^2$; indeed, it follows from the fact that any such map can be represented as a composition of an orthogonal projection to a one-dimensional subspace and a rotation.

Lemma 4.7.1. *There exist $\varepsilon_0 > 0$ and $n_0 \in \mathbb{N}$ such that for any non-zero vector $v \in \mathbb{R}^2$,*

any $p \in Y$, and any $\mu_1, \mu_2, \ldots, \mu_{n_0} \in \mathcal{K}$ we have

$$\operatorname{Var} \log \left| p \circ (B_{\mu_{n_0}} \ldots B_{\mu_1}) v \right| \geq \varepsilon_0.$$

To prove Lemma 4.7.1 we will use a statement from [22] that was called *Atom Dissolving Theorem* there. We will start with a couple of definitions.

Definition 4.7.2. Denote by $\mathfrak{Max}(\nu)$ the weight of a maximal atom of a probability measure ν. In particular, if ν has no atoms, then $\mathfrak{Max}(\nu) = 0$.

Definition 4.7.3. Let X be a metric compact. For a measure μ on the space of homeomorphisms $\mathrm{Homeo}(X)$, we say that there is

- *no finite set with a deterministic image*, if there are no two finite sets $F, F' \subset X$ such that $f(F) = F'$ for μ-a.e. $f \in \mathrm{Homeo}(X)$;
- *no measure with a deterministic image*, if there are no two probability measures ν, ν' on X such that $f_*\nu = \nu'$ for μ-a.e. $f \in \mathrm{Homeo}(X)$.

The following statement is a general statement for non-stationary dynamics, ensuring the "dissolving of atoms": decrease of the probability of a given point being sent to any particular point.

Theorem 4.7.4 (Atoms Dissolving Theorem 1.13 from [22])**.** *Let $\mathbf{K}_X$ be a compact set of probability measures on Homeo(X).*

- *Assume that for any $\mu \in \mathbf{K}_X$ there is no finite set with a deterministic image. Then for any $\varepsilon > 0$ there exists n such that for any probability measure ν on X and any sequence $\mu_1, \ldots, \mu_n \in \mathbf{K}_X$ we have*

$$\mathfrak{Max}\left(\mu_n * \cdots * \mu_1 * \nu\right) < \varepsilon.$$

In particular, for any probability measure ν on X and any sequence $\mu_1, \mu_2, \ldots \in \mathbf{K}_X$ we have

$$\lim_{n\to\infty} \mathfrak{Max}\,(\mu_n * \cdots * \mu_1 * \nu) = 0.$$

- *If, moreover, for any $\mu \in \mathbf{K}_X$ there is no measure with a deterministic image, then the convergence is exponential and uniform over all sequences $\mu_1, \mu_2, \ldots$ from $\mathbf{K}_X^{\mathbb{N}}$ and all probability measures ν. That is, there exists $\lambda < 1$ such that for any n, any ν and any $\mu_1, \mu_2, \cdots \in \mathbf{K}_X$*

$$\mathfrak{Max}\,(\mu_n * \cdots * \mu_1 * \nu) < \lambda^n.$$

In the proof below we will only be using the first part of Theorem 4.7.4.

Proof of Lemma 4.7.1. Due to Theorem 4.7.4 and our assumptions regarding the measures from $\mathcal{K}$, there exists $n' \in \mathbb{N}$ such that for any $\mu_1, \mu_2, \ldots, \mu_{n'} \in \mathcal{K}$ we have

$$\mathfrak{Max}\,(\mu_{n'} * \cdots * \mu_1 * \nu) < \frac{1}{2}$$

for any probability measure ν on $\mathbb{RP}^1$. To prove Lemma 4.7.1 it is enough to choose $n_0 = n' + 1$.

Since

$$\operatorname{Var}\log\left|p \circ (B_{\mu_{n_0}} \ldots B_{\mu_1})v\right| = \operatorname{Var}\log\left|p \circ (B_{\mu_{n_0}} \ldots B_{\mu_1})\frac{v}{|v|}\right|,$$

without loss of generality we can assume that $|v| = 1$ and, slightly abusing the notation, consider it an element of $\mathbb{RP}^1$. For given $p \in Y$, $v \in \mathbb{R}^2, |v| = 1$, $\{\mu_1, \mu_2, \ldots, \mu_{n_0}\} \in \mathcal{K}^{n_0}$ consider the probability distribution χ on $[0, +\infty)$ of the random images $\left|p \circ (B_{\mu_{n_0}} \ldots B_{\mu_1})v\right|$.

Lemma 4.7.5. *The function* $\Phi : \mathbb{RP}^1 \times Y \times \mathcal{K}^{n_0} \to \mathbb{R} \cup \{\infty\}$ *defined by*

$$\Phi(v, p, \mu_1, \mu_2, \dots, \mu_{n_0}) = \begin{cases} \infty, & \text{if } \chi(\{0\}) > 0; \\ \operatorname{Var} \log \left| p \circ (B_{\mu_{n_0}} \dots B_{\mu_1}) v \right|, & \text{if } \chi(\{0\}) = 0, \end{cases}$$

is lower semicontinuous.

Proof. Notice that χ depends continuously on $(v, p, \mu_1, \mu_2, \dots, \mu_{n_0})$ in weak-$*$ topology.

Let us consider the cases when $\chi(\{0\}) > 0$ and when $\chi(\{0\}) = 0$ separately.

Assume first that $\chi(\{0\}) > 0$. We want to show that given $M > 0$, for any sufficiently small perturbation χ' of χ we have $\operatorname{Var} \log \chi' > M$ (here we slightly abuse notation by thinking of χ and χ' as both distributions and random variables with these distributions). Notice that the measures condition implies that χ cannot be concentrated exclusively at $0 \in \mathbb{R}$. Hence for some $\tau > 0$ we have $\chi[\tau, +\infty) > 0$. If χ' is a probability distribution that is sufficiently close to χ, then $\chi'[\tau/2, +\infty)$ is not less than $\frac{1}{2}\chi[\tau, +\infty)$, and the χ'-weight of a small neighborhood of the origin is at least $\frac{1}{2}\chi(\{0\})$. Choosing that neighborhood small enough guarantees that $\operatorname{Var} \log \chi' > M$.

Assume now that $\chi(\{0\}) = 0$ and $\operatorname{Var} \log \chi < \infty$. Then

$$\operatorname{Var} \log \chi = \lim_{T \to \infty} \operatorname{Var} \left[(\log \chi)|_{[-T,T]} \right],$$

so for any $\varepsilon > 0$, for some large enough $T > 0$ we have

$$\operatorname{Var} \left[(\log \chi)|_{[-T,T]} \right] > \operatorname{Var} \log \chi - \frac{\varepsilon}{2}.$$

Therefore, for any χ' that is sufficiently close to χ we have

$$\operatorname{Var}\log\chi' \geq \operatorname{Var}\left[(\log\chi')|_{[-2T,2T]}\right] \geq \operatorname{Var}\left[(\log\chi)|_{[-T,T]}\right] - \frac{\varepsilon}{2} > \operatorname{Var}\log\chi - \varepsilon.$$

The case when $\chi(\{0\}) = 0$ and $\operatorname{Var}\log\chi = \infty$ can be treated similarly. □

The space $\mathbb{RP}^1 \times Y \times \mathcal{K}^{n_0}$ is compact. Hence, Lemma 4.7.5 implies that it is enough to show that $\Phi > 0$ to ensure that for some $\varepsilon_0 > 0$ we have $\Phi \geq \varepsilon_0 > 0$.

Suppose this is not the case, and for some unit vector v, a linear map $p \in Y$, and $\mu_1, \ldots, \mu_{n_0} \in \mathcal{K}$ we have

$$\operatorname{Var}\log\left|p \circ (B_{\mu_{n_0}} \ldots B_{\mu_1})v\right| = 0.$$

Then for some $d \geq 0$ with probability 1 we have $\left|p \circ (B_{\mu_{n_0}} \ldots B_{\mu_1})v\right| = d$. That means that $B_{\mu_{n_0}} \ldots B_{\mu_1} v$ has to belong to $L = \{u \mid |p(u)| = d\}$, which is a line (if $d = 0$) or a union of two lines (if $d > 0$). This implies that $\mu_1 \times \mu_2 \times \ldots \times \mu_{n_0-1}$-almost surely the image $B_{\mu_{n_0-1}} \ldots B_{\mu_1} v$ must belong to the set $\cap_{A \in \operatorname{supp}(\mu_{n_0})} A^{-1}(L)$. Since the measure μ_{n_0}

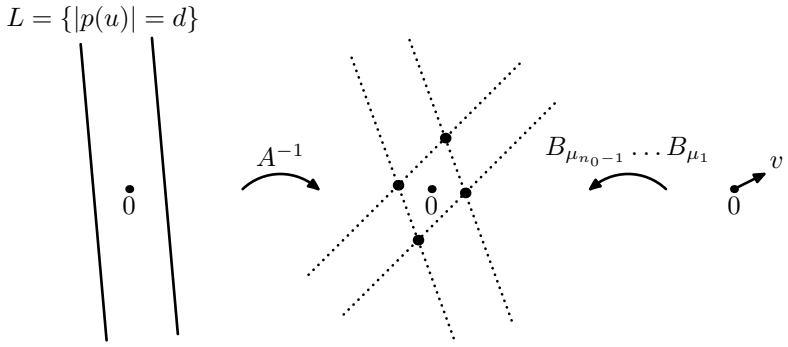

Figure 4.3: The set L and its preimages

must satisfy the measure condition, this intersection must consist of at most four points (see Fig. 4.3), whose projectivization gives at most two points on $\mathbb{RP}^1$. But this would imply that if ν is an atomic measure on $\mathbb{RP}^1$ at the point corresponding to the initial vector v, then $\mu_{n_0-1} * \cdots * \mu_1 * \nu$ is a measure supported on at most two points, which contradicts the

choice of n_0 above. This completes the proof of Lemma 4.7.1. □

Lemma 4.7.6. *For any $m \in \mathbb{N}$, any $\{\mu_{1,i}, \dots, \mu_{n_0,i}\}_{i=1,\dots,m} \in \mathcal{K}^{mn_0}$, and any $\{p_1, \dots, p_{m+1}\} \in Y^{m+1}$ we have*

$$\operatorname{Var} \log \|p_{m+1}(B_{\mu_{n_0,m}} \dots B_{\mu_{1,m}})p_m \dots p_2(B_{\mu_{n_0,1}} \dots B_{\mu_{1,1}})p_1\| \geq \varepsilon_0 m.$$

Proof. As each p_j is a unit norm rank 1 matrix, it can be written as

$$p_j = v_j \otimes \ell_j, \quad \text{where} \quad v_j \in \mathbb{R}^2, \quad \ell_j \in (\mathbb{R}^2)^*, \quad |v_j| = |\ell_j| = 1.$$

Now, let

$$\tilde{B}_j := B_{\mu_{n_0,j}} \dots B_{\mu_{1,j}}$$

be the j-th intermediate product. Then for the product

$$T = p_{m+1}\tilde{B}_m p_m \dots p_2 \tilde{B}_1 p_1$$

one has for any $v \in \mathbb{R}^2$

$$T(v) = v_{m+1} \cdot \ell_{m+1}(\tilde{B}_m v_m) \cdot \dots \cdot \ell_2(\tilde{B}_1 v_1) \cdot \ell_1(v),$$

and hence

$$\log \|T\| = \sum_{j=1}^{m} \log |\ell_{j+1}(\tilde{B}_j v_j)|. \tag{4.7.1}$$

Right hand side of (4.7.1) is a sum of m independent random variables, and the variance of each of them is at least ε_0 due to Lemma 4.7.1. Thus, the variance of $\log \|T\|$ is at least $m\varepsilon_0$. □

Lemma 4.7.7. *There exists a neighborhood U of the compact $\mathcal{K}^{n_0 m} \times Y^{m+1}$ in $\mathcal{K}^{n_0 m} \times Mat_2(\mathbb{R})^{m+1}$ such that for any*

$$\bar{\mu} \times \{D_j\}_{j=1,\dots,m+1} \in U,$$

where $\bar{\mu} = \{\mu_{1,i}, \dots, \mu_{n_0,i}\}_{i=1,\dots,m}$ and $D_j \in Mat_2(\mathbb{R})$, we have

$$\operatorname{Var} \log \|D_{m+1}(B_{\mu_{n_0,m}} \dots B_{\mu_{1,m}})D_m \dots D_2(B_{\mu_{n_0,1}} \dots B_{\mu_{1,1}})D_1\| \geq \frac{\varepsilon_0 m}{2}$$

Proof. On $\mathcal{K}^{n_0 m} \times Y^{m+1}$ this variance is bounded from below by $\varepsilon_0 m$ due to Lemma 4.7.6. As the set $\mathcal{K}^{n_0 m} \times Y^{m+1}$ is compact, and the variance is a lower-semicontinuous function of a distribution, there exists a neighbourhood U of this compact on which the variance is at least $\frac{m\varepsilon_0}{2}$. □

Corollary 4.7.8. *There exists Q such that for any $D_1, \dots, D_{m+1} \in \mathrm{SL}(2, \mathbb{R})$ with $\|D_j\| \geq Q$, $j = 1, \dots, m+1$ one has*

$$\operatorname{Var} \log \|D_{m+1}(B_{\mu_{n_0,m}} \dots B_{\mu_{1,m}})D_m \dots D_2(B_{\mu_{n_0,1}} \dots B_{\mu_{1,1}})D_1\| \geq \frac{\varepsilon_0 m}{2}$$

Proof.

$$\log \|D_{m+1}(B_{\mu_{n_0,m}} \dots B_{\mu_{1,m}})D_m \dots D_2(B_{\mu_{n_0,1}} \dots B_{\mu_{1,1}})D_1\| =$$
$$= \log \|\tilde{D}_{m+1}(B_{\mu_{n_0,m}} \dots B_{\mu_{1,m}})\tilde{D}_m \dots \tilde{D}_2(B_{\mu_{n_0,1}} \dots B_{\mu_{1,1}})\tilde{D}_1\| + \sum_{j=1}^{m+1} \log \|D_j\|, \quad (4.7.2)$$

where $\tilde{D}_j := \frac{D_j}{\|D_j\|}$. On the other hand, as $\|D\| \to \infty$ for $D \in \mathrm{SL}(2, \mathbb{R})$, one has $\frac{D}{\|D\|} \to Y$, so it suffices to choose Q sufficiently large to ensure that

$$(\{\mu_{i,k}\}_{1 \leq i \leq n_0, 1 \leq k \leq m}, (\tilde{D}_1, \dots, \tilde{D}_{m+1})) \in U$$

once $\|D_j\| \geq Q$ for all $j = 1, \dots, m+1$, where U is provided by Lemma 4.7.7. $\square$

Proof of Proposition 4.6.2. First, fix n_0 and ε_0 given by Lemma 4.7.1. Then, choose and fix m such that $\frac{m\varepsilon_0}{4} > c$.

Now, take a sufficiently large Q provided by Corollary 4.7.8. It follows from [18, Theorem 2.2] that for a sufficiently large n_2 one has

$$\forall n' \geq n_2 \quad \forall \mu_1, \dots, \mu_{n'} \in \mathcal{K} \quad \mathbb{P}_{\mu_1,\dots,\mu_{n'}}(\|A_{n'} \dots A_1\| \geq Q) \geq 1 - \frac{1}{2(m+1)}.$$

Now, take $n_3 := n_2(m+1) + n_0 m$. Then, for any $n \geq n_3$ and any $\mu_1, \dots, \mu_n \in \mathcal{K}$ we can decompose the product $A_n \dots A_1$ as

$$A_n \dots A_1 = D_{m+1} \tilde{B}_m D_m \dots D_2 \tilde{B}_1 D_1,$$

where each D_j is a product of at least n_2 matrices A_i, and each $\tilde{B}_j$ is a product of n_0 of A_i's.

This implies that with the probability at least $\frac{1}{2}$ one has $\|D_j\| \geq Q$ for all j, and hence the variance of the distribution conditional to such D_j is at least $\frac{m\varepsilon_0}{2}$. Thus, we finally have

$$\operatorname{Var} \xi_n \geq \mathbb{E}_{D_1,\dots D_{m+1}} \operatorname{Var}(\xi_n \mid D_1, \dots D_{m+1}) \geq \frac{1}{2} \cdot \frac{m\varepsilon_0}{2} = \frac{m\varepsilon_0}{4} > c.$$

$\square$

4.8 CLT for images of vectors and matrix elements

Let us now deduce Theorem 1.2.8 from Theorem 1.2.6.

Proof of Theorem 1.2.8. We start with establishing the first part of the theorem, convergence

for normalised log-lengths $\log|T_n v|$. Note that it suffices to show that for every $s > 0$ there exists a constant C_s such that for all sufficiently large n,

$$\mathbb{P}\left(\log\|T_n\| - \log|T_n v| > C_s\right) < s. \tag{4.8.1}$$

Indeed, as $\mathrm{Var}(\log\|T_n\|)$ tends to infinity, once (4.8.1) is established, it would imply that the difference

$$\frac{\log\|T_n\| - L_n}{\sqrt{\mathrm{Var}(\log\|T_n\|)}} - \frac{\log|T_n v| - L_n}{\sqrt{\mathrm{Var}(\log\|T_n\|)}} = \frac{\log\|T_n\| - \log|T_n v|}{\sqrt{\mathrm{Var}(\log\|T_n\|)}}$$

converges to zero in probability. And as adding a random variable that converges in probability to zero does not affect the weak limit, this would imply the convergence

$$\frac{\log|T_n v| - L_n}{\sqrt{\mathrm{Var}(\log\|T_n\|)}} \to \mathcal{N}(0,1).$$

Now, without loss of generality we can assume that the vector v is a vector of unit length. The difference $\log\|T_n\| - \log|T_n v| = \Theta(T_n, v)$ can then be estimated using Lemma 4.5.1: its conclusion (4.5.3) implies that it suffices to show that for a sufficiently small angle $\alpha = \alpha(s)$, one has

$$\mathbb{P}\left(\mathbf{r}(T_n) \in U_\alpha[v]\right) < s. \tag{4.8.2}$$

Indeed, once such angle α is found, we can take $C_s := -\log\sin\alpha$.

We will now proceed along the same lines as in the proof of Lemma 4.5.4. Namely, Lemma 4.5.3 allows us to approximate the location of $\mathbf{r}(T_n)$: it is $\frac{1}{\|T_n\|^2}$-close to at least one of the directions $[T_n^{-1}e_1]$, $[T_n^{-1}e_2]$.

Now, for every $\alpha > 0$ the probability that such a preimage direction belongs to a given 2α-

neighbourhood of a given direction can be estimated using the log-Hölder estimates (4.5.14) that are implied by Theorem 4.5.7. Namely, we apply it with

$$\nu_0^{(1)} = \nu_0^{(2)} = \delta_{[e_i]}, \quad \mu_j^{(1)} = \mu_j^{(2)} = (F_2)_* \mu_{n+1-j},$$

where $F_2 : A \mapsto f_A^{-1}$, so that the measures $\nu_1 = \nu_2$ (defined by (4.5.12)) are the distributions of preimages $[T_n^{-1} e_i]$. Note that

$$U_{2\alpha}([v]) \times U_{2\alpha}([v]) \subset \{(x, y) \mid \operatorname{dist}(x, y) < 4\alpha\},$$

and thus

$$\nu_1(U_{2\alpha}([v]))^2 \leq (\nu_1 \times \nu_2)(\{(x, y) \mid \operatorname{dist}(x, y) < 4\alpha\}).$$

Hence, from (4.5.14) we get that for every $i = 1, 2$

$$\mathbb{P}\left([T_n^{-1} e_i] \in U_{2\alpha}([v])\right) \leq \sqrt{C_{\mathcal{M}}} \cdot |\log(4\alpha)|^{-\gamma/2} \tag{4.8.3}$$

Fixing $\alpha > 0$ sufficiently small so that the right hand side of (4.8.3) does not exceed $s/3$, we see that with the probability at least $1 - \frac{2}{3}s$ neither of the preimages $[T_n^{-1} e_1]$, $[T_n^{-1} e_2]$ belongs to $U_{2\alpha}([v])$. Finally, take the number n of matrices sufficiently large so that $\|T_n\|^2 > \frac{1}{\alpha}$ with the probability at least $1 - \frac{s}{3}$ and the allowed radius $\theta_{n,n} < 4\alpha$; by Lemma 4.5.3 it implies that at least one of the distances $\operatorname{dist}(\mathbf{r}(T_n), [T_n^{-1} e_i])$ does not exceed α. Then altogether with the probability at least $1 - s$ we have

$$\operatorname{dist}(\mathbf{r}(T_n), [v]) \geq \max_{i=1,2} (\operatorname{dist}([v], [T_n^{-1} e_i]) - \operatorname{dist}(\mathbf{r}(T_n), [T_n^{-1} e_i])) \geq 2\alpha - \alpha = \alpha,$$

implying the desired (4.8.2).

The same argument applies to the random variables $\log|(T_n)_{i,j}|$: it suffices to show that for every $s > 0$ there exists a constant C'_s such that for all sufficiently large n,

$$\mathbb{P}\left(\log|T_n e_j| - \log|(T_n)_{i,j}| > C'_s\right) < s. \tag{4.8.4}$$

Now, for any pair of indices $i, j = 1, 2$,

$$|(T_n)_{i,j}| = |T_n e_j| \cdot \sin \operatorname{dist}([T_n e_j], [e_{3-i}]).$$

Using the same log-Hölder estimate (4.5.14) for the forward dynamics, we choose α' such that $[T_n e_j]$ belongs to α'-neighborhoods of either of $[e_1], [e_2]$ with the probability at most s and $\alpha' > \theta_{n,n}$ for sufficiently large n, providing (4.8.4) and thus completing the proof. □

4.9 Examples

In this section, we discuss examples, showing that in the non-stationary setting, assuming only second moments to be uniformly bounded does not suffice to obtain the convergence to the Gaussian distribution. In particular, it shows that the moment assumption in Theorem 1.2.6 is optimal.

We start with an example addressing the classical CLT setting on sums of independent random variables. For every $\varepsilon > 0$, consider a random variable ξ_ε, taking values

$$\xi_\varepsilon = \begin{cases} 1 & \text{with probability } \frac{1-\varepsilon}{2} \\ -1 & \text{with probability } \frac{1-\varepsilon}{2} \\ \frac{10}{\sqrt{\varepsilon}} & \text{with probability } \varepsilon. \end{cases} \tag{4.9.1}$$

Now, the second moments of these random variables are uniformly bounded:

$$\mathbb{E}\xi_\varepsilon^2 < 1 + \varepsilon \cdot \frac{10^2}{\varepsilon} = 101.$$

Example 4.9.1. Take a sequence n_j, satisfying $n_{j+1} > n_j^6$ for all j, and let $\varepsilon_j := \frac{1}{n_j}$. Consider the sequence of independent random variables

$$\xi_{1,1}, \dots, \xi_{1,n_1}; \xi_{2,1}, \dots, \xi_{2,n_2}; \dots; \xi_{j,1} \dots, \xi_{j,n_j}; \dots$$

where all the random variables $\xi_{j,i}$, $i = 1, \dots, n_j$ have the distribution (4.9.1) with $\varepsilon = \varepsilon_j$. Then, though this is a sequence of independent random variables with uniformly bounded second moments, their sums are not asymptotically normal.

Proof. Consider the sums within the groups,

$$S_j := \xi_{j,1} + \dots + \xi_{j,n_j}.$$

Note that $\frac{1}{\sqrt{n_j}} S_j$ converge in law to a law that is the sum of the Gaussian law $\mathcal{N}(0,1)$ and of the 10-scaled Poisson distribution with the parameter 1. Indeed, the Poisson component comes from the variables taking values $\frac{10}{\sqrt{\varepsilon_j}}$, and conditionally to the places where these "large" values occur, the part that is left satisfies the classical CLT.

Thus, the limit law is multi-modal (see Fig. 4.4), and hence the sums S_j are not asymptotically normal.

Finally, the growth condition $n_j > n_{j-1}^6$ shows that $\frac{S_1 + \dots + S_{j-1}}{\sqrt{n_j}}$ converges to zero uniformly, hence the same non-normal distribution is the limit of the sequence

$$\frac{\xi_{1,1} + \dots + \xi_{j,n_j}}{\sqrt{n_j}} = \frac{S_1 + \dots + S_j}{\sqrt{n_j}}.$$

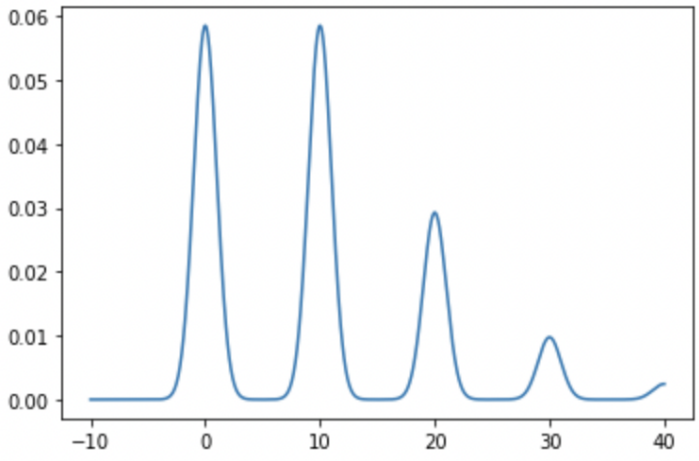

Figure 4.4: Density for the limit law in Example 4.9.1

□

Now, for every $\varepsilon \in [0, \frac{1}{2}]$ let μ_ε be a measure on $\mathrm{SL}(2, \mathbb{R})$ that is the law of the random matrix

$$B_\varepsilon = R_{\beta_1} \begin{pmatrix} r_\varepsilon & 0 \\ 0 & r_\varepsilon^{-1} \end{pmatrix} R_{\beta_2},$$

where β_1, β_2 and r_ε are independent, β_1 and β_2 are uniformly distributed on $[0, 2\pi]$, and

$$r_\varepsilon = \begin{cases} 2 & \text{with probability } 1-\varepsilon \\ e^{\frac{10}{\sqrt{\varepsilon}}} & \text{with probability } \varepsilon. \end{cases} \tag{4.9.2}$$

These measures then form a compact set $\mathcal{K}$ of probability measures on $\mathrm{SL}(2, \mathbb{R})$, on which the second logarithmic moment is uniformly bounded:

$$\mathbb{E}(\log \|B_\varepsilon\|)^2 < (\log 2)^2 + \varepsilon \cdot \frac{10^2}{\varepsilon} < 101.$$

Moreover, due to the presence of random rotations by α and β the measures condition (stated in Sec. 1.2.2) is also satisfied.

Now, denote by λ the Lyapunov exponent for the product of the matrices corresponding to μ_0, and by σ^2 the corresponding variance in the CLT. We then have the following example.

Example 4.9.2. Take lengths n_j and the associated values ε_j as in Example 4.9.1, and take the sequence in which these measures are repeated by groups,

$$\underbrace{\mu_{\varepsilon_1}, \dots, \mu_{\varepsilon_1}}_{n_1 \text{ times}}; \underbrace{\mu_{\varepsilon_2}, \dots, \mu_{\varepsilon_2}}_{n_2 \text{ times}}; \dots; \underbrace{\mu_{\varepsilon_j}, \dots, \mu_{\varepsilon_j}}_{n_j \text{ times}}; \dots.$$

Let T_n be the associated products of independent random matrices, and denote $N_j := n_1 + \dots + n_j$ the index at which the j-th group ends. Then the random variables

$$\frac{1}{\sqrt{n_j}}(\log \|T_{N_j}\| - n_j\lambda)$$

converge in law to the sum of a Gaussian distribution $\mathcal{N}(0, \sigma^2)$ and of the 10-scaled Poisson distribution with parameter 1. In particular, this limit is non-normal, and thus Theorem 1.2.6 cannot hold for such a product.

Proof. Let us first instead of the log-norms of matrices $\log \|T_n\|$ fix an initial vector v_0 and consider log-lengths $\log |T_n v_0|$ of its images $v_n := T_n v_0$. Note that increments $\eta_n := \log |v_n| - \log |v_{n-1}|$ of this sequence are then *independent* random variables. Indeed, due to the pre-composition with the rotation R_{β_2}, the distribution of the increment of the log-norm,

$$\log |B_\varepsilon v| - \log |v| = \log \frac{|B_\varepsilon v|}{|v|}, \tag{4.9.3}$$

does not depend on the choice of the initial nonzero vector v. Hence, each new (n-th) increment $\log \frac{|v_{n+1}|}{|v_n|}$, conditioned to all the previous matrices (that determine the corresponding image v_n and all the preceding increments) has the same distribution, and thus these increments are independent.

Now, the distribution of the increment (4.9.3) is very close to the one in Example 4.9.1: it is a mix of a given non-degenerate bounded law (corresponding to the distribution μ_0) with the probability $1-\varepsilon$ (the case if r_ε takes value 2), and of the law that is associated to the application of the diagonal matrix associated to $e^{\frac{10}{\sqrt{\varepsilon}}}$. The latter law is concentrated around $\frac{10}{\sqrt{\varepsilon}}$: as $\varepsilon \to 0$, this law, rescaled by $\sqrt{\varepsilon}$, converges to the constant 10. Meanwhile, to the former law the usual Central Limit Theorem is applicable, and the sum of n_j such independent random variables, from which $n_j\lambda$ is subtracted, after division by $\sqrt{n_j}$ converges to $\mathcal{N}(0,\sigma^2)$.

Also, note that in the same way as in Example 4.9.1, it suffices to study the product only of matrices corresponding to the j-th group. Indeed, the log-norm of the product of the previous ones does not exceed $2n_{j-1}\cdot 10\sqrt{n_{j-1}} = 20n_{j-1}^{3/2}$, while $n_{j-1} < n_j^{1/6}$, and thus this upper bound is $o(\sqrt{n_j})$.

Hence, we have the convergence in law of the rescaled log-laws

$$\frac{1}{\sqrt{n_j}}(\log|T_{N_j}v_0| - n_j\lambda) \tag{4.9.4}$$

to the sum of $\mathcal{N}(0,\sigma^2)$ and of the 10-scaled Poisson distribution with the parameter 1; we denote this distribution $\mathcal{D}$.

Finally, take the initial vector v_0 to be chosen randomly uniformly on the unit circle (and independently from the product T_{N_j}). Conditionally to each choice of v_0, the law of (4.9.4) is the same and converges to $\mathcal{D}$ as $j\to\infty$. At the same time, the log-norm $\log\|T_n\|$ is close to $\log|T_nv_0|$ for most (in the sense of the Lebesgue measure) vectors v_0 on the circle. Namely, due to Lemma 4.5.1,

$$|\log|T_nv_0| - \log\|T_n\|\,| \le |\log\sin\operatorname{dist}([v_0],\mathbf{r}(T_n))|, \tag{4.9.5}$$

and as v_0 is independent from T_n and the distribution of its direction $[v_0]$ is uniform, the distribution of the random variable in the right hand side of (4.9.5) does not depend on n. In particular, for every R the probability that the right hand side exceeds R is at most e^{-R}.

Hence, the convergence in law of the log-lengths (4.9.4) implies also the convergence of

$$\frac{1}{\sqrt{n_j}}(\log \|T_{N_j}\| - n_j\lambda) \tag{4.9.6}$$

to the same distribution $\mathcal{D}$. Indeed, for v_0 chosen independently from T_{N_j}, taking $R_j = \sqrt[4]{n_j}$, we see that the random variables (4.9.4) and (4.9.6) differ at most by $\frac{R_j}{\sqrt{n_j}} = n_j^{-1/4} = o(1)$ on the set of the probability $1 - e^{-R_j} = 1 - o(1)$. In particular, this convergence implies that the sequence of random variables $\log \|T_n\|$ is not asymptotically normal.

□

Bibliography

[1] R. Aoun and Y. Guivarc'h. Random matrix products when the top Lyapunov exponent is simple. *J. Eur. Math. Soc. (JEMS)*, 22(7):2135–2182, 2020.

[2] L. Arnold. Random dynamical systems. In *Dynamical systems (Montecatini Terme, 1994)*, volume 1609 of *Lecture Notes in Math.*, pages 1–43. Springer, Berlin, 1995.

[3] J. Atnip, G. Froyland, C. Gonzalez-Tokman, and A. Quas. Universal gap growth for Lyapunov exponents of perturbed matrix products. 2023. arXiv:2312.03181.

[4] A. Avila and J. Bochi. Lyapunov exponents: Part I, notes of mini-course given at the school of dynamical systems, ICTP. 2008. Lecture Notes.

[5] R. Bellman. Limit theorems for non-commutative operations. I. *Duke Math. J.*, 21:491–500, 1954.

[6] Y. Benoist and J. Quint. *Random walks on reductive groups*, volume 62 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics*. Springer, Cham, 2016. xi+323 pp.

[7] Y. Benoist and J. F. Quint. Central limit theorem for linear groups. *The Annals of Probability*, 44:1308–1340, 2016.

[8] Y. Benoist and J. F. Quint. Central limit theorem on hyperbolic groups. *Izv. Math.*, 80(1):3–23, 2016.

[9] Y. Benoist and J. F. Quint. *Random walks on reductive groups*. Springer International Publishing, 2016.

[10] P. Bougerol and J. Lacroix. *Products of Random Matrices with Applications to Schrödinger Operators*. Birkhauser, Boston, 1985.

[11] J. Bourgain, A. Furman, E. Lindenstrauss, and S. Mozes. Stationary measures and equidistribution for orbits of nonabelian semigroups on the torus. *J. Amer. Math. Soc.*, 24:231–280, 2011.

[12] A. Brown and F. Hertz. Measure rigidity for random dynamics on surfaces and related skew products. *J. Amer. Math. Soc.*, 30:1055–1132, 2017.

[13] D. Feng. Dimension of invariant measures for affine iterated function systems. *Duke Math. J.*, 172:701–774, 2023.

[14] A. Furman. Random walks on groups and random transformations. In *Handbook of dynamical systems*, volume 1A, pages 931–1014. North-Holland, Amsterdam, 2002.

[15] A. Furman. What is ... a stationary measure? *Notices Amer. Math. Soc.*, 58(9):1276–1277, 2011.

[16] H. Furstenberg. Noncommuting random products. *Trans. Amer. Math. Soc.*, 108:377–428, 1963.

[17] H. Furstenberg and K. Kesten. Products of random matrices. *Ann. Math. Statist.*, 31:457–469, 1960.

[18] I. Goldsheid. Exponential growth of products of non-stationary Markov-dependent matrices. *Int. Math. Res. Not. (IMRN)*, 2022:6310–6346, 2022.

[19] I. Goldsheid and Y. Guivarc'h. Zariski closure and the dimension of the Gaussian law of the product of random matrices. I. *Probab. Th. Rel. Fields*, 105:109–142, 1996.

[20] I. Gol'dsheĭd and G. Margulis. Lyapunov exponents of a product of random matrices. *Uspekhi Mat. Nauk*, 44(5):13–60, 1989.

[21] A. Gorodetski and V. Kleptsyn. Non-stationary Anderson localization. *Comm. Amer. Math. Soc.*, 5:81–143, 2025.

[22] A. Gorodetski and V. Kleptsyn. Non-stationary version of Furstenberg's theorem on random matrix products. *Journal of the Association for Mathematical Research*, 4(1):1–41, Mar. 2026.

[23] A. Gorodetski, V. Kleptsyn, and G. Monakov. Central limit theorem for non-stationary random products of $SL(2,\mathbb{R})$ matrices. 2026. preprint, arXiv:2411.12003.

[24] A. Gorodetski, V. Kleptsyn, and G. Monakov. Hölder regularity of stationary measures. *Inventiones mathematicae*, 243:1037–1077, 2026.

[25] Y. Guivarc'h. Produits de matrices aleatoires et applications aux proprietes geometriques des sous-groupes du groupe lineaire. *Ergodic Theory Dynam. Systems*, 10:483–512, 1990.

[26] Y. Guivarc'h and A. Raugi. Products of random matrices: convergence theorems. In *Random matrices and their applications (Brunswick, Maine, 1984)*, volume 50 of *Contemp. Math.*, pages 31–54. Amer. Math. Soc., Providence, RI, 1986.

[27] C. Jan. Vitesse de convergence dans le TCL pour des chaines de Markov et certains processus associes a des systemes dynamiques. *C. R. Acad. Sci. Paris Ser. I Math.*, 331:395–398, 2000.

[28] Y. Kifer. *Ergodic theory of random transformations*, volume 10 of *Progress in Probability and Statistics*. Birkhäuser Boston, Inc., Boston, MA, 1986. x+210 pp.

[29] Y. Kifer. Random dynamics and its applications. In *Proceedings of the International Congress of Mathematicians*, volume II (Berlin, 1998), pages 809–818, 1998. Doc. Math. 1998, Extra Vol. II.

[30] A. Kolmogorov. Sur la loi forte des grands nombres. *Comptes Rendus des Seances de l'Academie des Sciences*, 191:910–912, 1930.

[31] P. Liu and M. Qian. *Smooth ergodic theory of random dynamical systems*, volume 1606 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1995. xii+221 pp.

[32] D. Malicet. Random walks on homeo(S^1). *Commun. Math. Phys.*, 356:1083–1116, 2017.

[33] P. March and A. Sznitman. Some connections between excursion theory and the discrete Schrödinger equation with random potentials. *Probab. Theory Related Fields*, 75(1):11–53, 1987.

[34] G. Monakov. Log-Hölder regularity of stationary measures. *Commun. Math. Phys.*, 406(244), 2025.

[35] E. L. Page. Théorèmes limites pour les produits de matrices aléatoires. In H. Heyer, editor, *Probability Measures on Groups*. Springer-Verlag, New York, 1982.

[36] I. Pinelis. Optimum bounds for the distributions of martingales in Banach spaces. *Ann. Probab.*, 22:1679–1706, 1994.

[37] A. Rapaport. Exact dimensionality and Ledrappier-Young formula for the Furstenberg measure. *Trans. Amer. Math. Soc.*, 374:5225–5268, 2021.

[38] V. N. Tutubalin. On limit theorems for products of random matrices. *Teor. Veroyatnost. i Primenen.*, 10:19–32, 1965.

[39] V. N. Tutubalin. A central limit theorem for products of random matrices and some of its applications. In *Symposia Mathematica, Vol. XXI*, pages 101–116. Academic Press, London-New York, 1977.

[40] C. Villani. *Optimal Transport. Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. xxii + 973 pp.

[41] K. Zieber. Localization of the one-dimensional non-stationary anderson model. *Journal of Mathematical Physics*, 67(4):042702, 04 2026.