HÖLDER REGULARITY OF STATIONARY MEASURES

ANTON GORODETSKI, VICTOR KLEPTSYN, AND GRIGORII MONAKOV

Abstract. 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. The result is a vast generalization of the classical statement on Hölder continuity of stationary measures of random walks on linear groups.

1. Introduction

Let $M$ be a smooth closed Riemannian manifold, and $\mu$ be a Borel probability measure on $\text{Diff}^1(M)$, 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 $f_i \in \text{Diff}^1(M)$ are chosen randomly and independently, with respect to the distribution $\mu$.

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

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

The measure $\nu$ is called $\mu$-stationary if $\mu \ast \nu = \nu$. Stationary measures of random dynamical systems are analogues of invariant measures of deterministic maps, and their properties are of crucial importance for many results in random dynamics, see [A, BH, BQ, BL, Fu1, Fu2, Kif1, Kif2, LQ, M] and references therein.

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

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

Here is the main result of this paper:

A. G. and G. M. were supported in part by NSF grant DMS–1855541.
V.K. was supported in part by ANR Gromeov (ANR-19-CE40-0007) and by Centre Henri Lebesgue (ANR-11-LABX-0020-01).
Theorem 1.1. Suppose that \( \mu \) is a probability distribution on \( \text{Diff}^1(M) \) such that \( \int L(f)^\gamma d\mu(f) < \infty \) for some \( \gamma > 0 \). Suppose also that there is no probability measure \( m \) on the manifold \( M \) invariant under every map \( f \in \text{supp} \mu \). Then every stationary measure of a random dynamical system defined by the distribution \( \mu \) is Hölder continuous.

In Section 2 below we provide formal definitions and more general and stronger versions of this result. Specifically, in Theorem 2.6 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 2.8 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.

Let us now discuss two classical statements (on Hölder continuity of stationary measures for random matrix products and for iterated function systems) that can be considered partial cases of Theorem 1.1.

1.1. Stationary measures for random matrix products. Consider random products of iid matrices \( A_n A_{n-1} \ldots A_1 \), where each \( A_i \) is a random matrix chosen with respect to a probability distribution \( \tilde{\mu} \) on \( SL(d, \mathbb{R}) \), \( d \geq 2 \). Let \( S_{\tilde{\mu}} \) be the closed semigroup in \( SL(d, \mathbb{R}) \) generated by matrices from \( \text{supp} \tilde{\mu} \). Then \( S_{\tilde{\mu}} \) 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 \( A(\bigcup_{i=1}^{N} V_i) = \bigcup_{i=1}^{N} V_i \) for all \( A \in S_{\tilde{\mu}} \).

and proximal (or contracting) if there exists a sequence \( A_1, A_2, \ldots, \in S_{\tilde{\mu}} \) and a sequence of real numbers \( a_1, a_2, \ldots \) such that the sequence of operators \( \{a_i A_i\} \) converges in norm to a linear endomorphism of \( \mathbb{R}^d \) of rank one, see [BL] for a detailed discussion of these notions.

Every \( A \in SL(d, \mathbb{R}) \) induces a projective map \( f : \mathbb{R}^{d-1} \rightarrow \mathbb{R}^{d-1} \). Let us denote by \( \mu \) the distribution on projective maps induced by the distribution \( \tilde{\mu} \) on \( SL(d, \mathbb{R}) \).

Theorem 1.2 (Guivarc’h, [G90]). Suppose that, in the setting above, \( S_{\tilde{\mu}} \) is strongly irreducible and proximal, and

\[
\int \|A\|^\gamma d\tilde{\mu} < \infty
\]

for some \( \gamma > 0 \). Then the random dynamical system given by the distribution \( \mu \) on the projective maps of \( \mathbb{R}^{d-1} \) has unique stationary measure \( \nu \) on \( \mathbb{R}^{d-1} \), and \( \nu \) is Hölder continuous.
An alternative proof of Theorem 1.2 (inspired by [BFLM]) can be found in [BQ]. A proof of Hölder continuity of stationary measures under somewhat weaker assumptions was obtained in [AG].

Notice that Theorem 1.2 can be considered as a partial case of Theorem 1.1. 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 [BL, Chapter III]), and finiteness of momentum condition implies the corresponding condition in Theorem 1.1.

Notice that positivity of Lyapunov exponents for random matrix products due to Furstenberg Theorem holds under weaker conditions. In particular, instead of finiteness of the integral \( \int \|A\|^\gamma d\tilde{\mu} \) it is enough to require that \( \int \log \|A\| d\tilde{\mu} < \infty \). It is interesting to notice that it is not sufficient to ensure Hölder continuity of the stationary measure, see Appendix A below. At the same time, weaker statements on moduli of continuity can be established under these weaker assumptions, see Proposition 4.5 in [BQ16] and Theorem 1.4 in [DKW].

Other properties of the stationary measure in the setting of Theorem 1.2 were studied, sometimes under additional assumptions on the distribution \( \tilde{\mu} \). For example, in [KL] it was conjectured that if the distribution \( \tilde{\mu} \) in Theorem 1.2 is finitely supported, then the stationary measure must be singular. A counterexample to that conjecture with absolutely continues stationary measure was constructed in [BPS], and with stationary measure having smooth density – in [B]. A related question is whether in the case of a distribution \( \tilde{\mu} \) that generates a Fuchsian group the corresponding stationary measure must be singular. It was settled for non-cocompact actions by Guivarc’h and Le Jan in [GL] (see also [DKN] and [GMT]), but is still open in full generality.

1.2. Stationary measures for iterated function systems. An iterated function system is defined by a finite collection of contractions of a complete metric space, usually of \( \mathbb{R}^d \) or a compact subset of \( \mathbb{R}^d \). In [FL, Proposition 2.2] it is shown that in the case of an iterated function system generated by a finite collection of similarities in \( \mathbb{R}^d \) (without a common fixed point) applied with given prescribed probabilities the unique stationary measure must be Hölder continuous. Let us generalize this statement by allowing infinite families of contractions.

**Theorem 1.3.** Let \( B \subset \mathbb{R}^d \) be a closed ball, and \( \mu \) be a probability distribution on the space of all \( C^1 \) maps (contractions) \( f : B \to B \) such that \( f \) is a diffeomorphims of \( B \) onto \( f(B) \), and \( \max_{x \in B} \|Df(x)\| < 1 \).
Suppose that
\[ \int \left[ \text{Lip}(f^{-1}|_{f(B)}) \right]^\gamma d\mu < \infty \text{ for some } \gamma > 0, \]
and there is no common fixed point for all maps \( f \in \text{supp} \mu \). Then the unique stationary measure of the random dynamical system generated by the distribution \( \mu \) is Hölder continuous.

Uniqueness of a stationary measure in this context is well known, e.g. see [Hu] or [Sz]. Hölder continuity of the stationary measure in Theorem 1.3 is an immediate consequence of Theorem 1.1. Indeed, due to the contraction mapping principle, any contraction must have exactly one invariant measure, namely an atomic measure at a fixed point. Therefore, existence of a common invariant measure implies existence of a common fixed point. Since Theorem 1.1 can be applied to the maps of a manifold with boundary (see Remark 2.9 below), Theorem 1.3 follows.

We believe that Theorem 1.3 is well known to the experts, but to the best of our knowledge it has not appeared in print in this generality.

A famous specific example of one-dimensional stationary measures are those given by the Bernoulli convolutions. These are stationary measures for linearly contracting self-maps of the interval \([0, 1]\),
\[ x \mapsto \lambda x \text{ and } x \mapsto 1 - \lambda(1 - x), \]
where \( \lambda \in \left[ \frac{1}{2}, 1 \right] \). These measures were extensively studied for at least 70 years; on the one hand, it is known [E] that for some special values of \( \lambda \) (for instance, for inverse golden ratio) the stationary measure \( \nu_\lambda \) is singular and, moreover, has its Hausdorff dimension strictly less than 1. On the other hand, a famous result by Solomyak [Sol] establishes absolute continuity of \( \nu_\lambda \) for almost every \( \lambda \in \left[ \frac{1}{2}, 1 \right] \). Moreover, due to an improvement by Shmerkin [Shm] the set of exceptional values of the parameter \( \lambda \) is in fact of zero Hausdorff dimension. For general surveys we refer the reader to [PSS, Va1, Va2]. For recent results on a lower bound on the Hausdorff dimension of \( \nu_\lambda \) see [FF, KPV, Va3].

Questions about exact dimensionality of stationary measures of iterated function systems and explicit formulas for their dimension (usually of Ledrappier-Young type, i.e. in terms of entropy and Lyapunov exponents) were heavily studied, e.g. see [BK, F, FJ, Ho, HR], and references there. Similar questions on stationary measures of random matrix products were studied in [HS, L, LL, R]. Notice that while any Hölder continuous measure must have positive Hausdorff dimension,
without any additional assumptions it does not have to be exact dimensional, and, other way around, exact dimensional measure does not necessarily have to be Hölder continuous.

2. Definitions and main results

Definition 2.1. We say that the measure $\nu$ on $M$ is $(\alpha,C)$-Hölder, if
\begin{equation}
\forall x \in M \quad \forall r > 0 \quad \nu(B_r(x)) < Cr^\alpha.
\end{equation}

Definition 2.2. For a diffeomorphism $f \in \text{Diff}^1(M)$ we define
\begin{equation}
\text{Lip}(f) = \sup_{x,y \in M} \left( \frac{d(f(x), f(y))}{d(x, y)} \right)
\end{equation}
and
\begin{equation}
\mathcal{L}(f) = \max(\text{Lip}(f), \text{Lip}(f^{-1})).
\end{equation}

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.

Theorem 2.3. Let $\mu$ be a probability measure on $\text{Diff}^1(M)$, satisfying the following assumptions:

- **(finite moment condition)** There exists $\gamma > 0$ and $C_0 > 0$ such that
\begin{equation}
\int \mathcal{L}(f)^\gamma \, d\mu(f) < C_0;
\end{equation}

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

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

Remark 2.4. Denote by $\mathcal{M}$ the space of all probability measures on $\text{Diff}^1(M)$ equipped with weak-* topology. If $\mathcal{K} \subset \mathcal{M}$ is a compact set, such that the assumptions of Theorem 2.3 (with uniform $\gamma > 0$ and $C_0 > 0$) hold for all $\mu \in \mathcal{K}$, then $\alpha$ and $C$ in the conclusion can be chosen uniformly in $\mu \in \mathcal{K}$.

Remark 2.5. 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 supp $\mu$ preserve the same non-Hölder measure, e.g. some atomic measure.
Theorem 2.3 states that 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 \( \nu_0 \) become “more and more Hölder” as the number \( n \) of iterations increases.

However, if both initial measure \( \nu_0 \) and the measure \( \mu \) are atomic (for instance, if \( \nu_0 \) is a Dirac measure, and the measure \( \mu \) 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 2.6.** Assume that \( \mu \) satisfies the assumptions of Theorem 2.3. Then there exist \( \alpha > 0, C \) and \( \kappa < 1 \) such that for any initial measure \( \nu_0 \), any number of iterations \( n \in \mathbb{N} \), and any \( x \in M \) one has:

\[
(2.3) \quad \text{if } r > \kappa^n \text{ then } (\mu^n * \nu_0)(B_r(x)) < Cr^\alpha.
\]

**Remark 2.7.** Equivalently, the conclusion (2.3) in Theorem 2.6 can be replaced by the following one (perhaps, with different values of the constants):

for any \( r > 0 \) we have \( (\mu^n * \nu_0)(B_r(x)) < C(r^\alpha + \kappa^n) \).

Notice that Theorem 2.3 is an immediate consequence of Theorem 2.6. Indeed, if \( \nu_0 \) is a stationary measure, then \( \mu^n * \nu_0 = \nu_0 \), and since \( \kappa^n \to 0 \) as \( n \to \infty \), applying Theorem 2.6 to the measure \( \nu_0 \) shows that \( \nu_0 \) is Hölder continuous, and hence Theorem 2.3 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 \( \mu_i \). One example of such situation is non-stationary version of the Furstenberg theorem \([\text{GK],[Go]}\), 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 assumptions the random (averaged) image of any given probability measure on \( M \) after \( n \) iterations satisfies the Hölder property on not-too-small scales.

**Theorem 2.8.** Let \( K \subset M \) 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

\[
\int \mathcal{L}(f)^\gamma \, d\mu(f) < C_0;
\]

• **(no deterministic images)** For any \( \mu \in \mathcal{K} \) there are no probability measures \( \nu, \nu' \) on \( M \) such that \( f_*\nu = \nu' \) for \( \mu \)-almost all \( f \in \text{Diff}_1(M) \).

Then there exist \( \alpha > 0 \), \( C, \kappa < 1 \) such that for any initial measure \( \nu_0 \), any \( n \), and any distributions \( \mu_1, \ldots, \mu_n \in \mathcal{K} \) the \( n \)-th image of \( \nu_0 \) satisfies \((\alpha, C)\)-Hölder property on the scales up to \( \kappa^n \):

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

**Remark 2.9.** Theorems 2.3, 2.6 and 2.8 also hold for the case of \( M \) being a compact manifold with a boundary. In this case we consider \( \mu \) that is a probability measure on the space of all diffeomorphisms of \( M \) on the image.

**Remark 2.10.** While we formulate the results for \( C^1 \)-diffeomorphisms, smoothness is not really used in the proof. We only need the maps to be bi-Lipschitz, so we notice here that Theorems 2.3, 2.6 and 2.8 also hold for random dynamical systems defined by bi-Lipschitz maps, for probability measure \( \mu \) defined on the space of bi-Lipschitz homeomorphisms:

\[
\{ f \in \text{Homeo}(M) \mid \mathcal{L}(f) < \infty \}.
\]

2.1. **“No invariant measure” vs “no deterministic images” conditions.** The “no deterministic images” assumption in Theorem 2.8 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 \( f_i \mapsto \tilde{f}_i := g_i \circ f_i \circ g_i^{-1} \) 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 2.11.** Let \( \mu \in \mathcal{M} \) be a probability measure on \( \text{Diff}^1(M) \) such that there is no common invariant measure for all \( f \in \text{supp} \mu \). Then there exists \( k \in \mathbb{N} \) such that \( \mu^k \) satisfies “no deterministic images” condition (i.e. there are no probability measures \( \nu, \nu' \) 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 \( (f_1, f_2, \ldots, f_k) \in \left( \text{Diff}^1(M) \right)^k \)).

**Proof of Proposition 2.11.** Assume that for arbitrarily large \( k \) there exist measures \( \nu_k, \nu'_k \) such that \( f_*, \nu_k = \nu'_k \) for \( \mu^k \)-a.e. \( f \in \text{Diff}^1(M) \). For any such \( k \) consider the intermediate images (note that intermediate images must be deterministic, too):

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

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 \( \mu\)-a.e. \( f \in \text{Diff}^1(M) \). Namely, for \( \mu\)-a.e. \( f \in \text{Diff}^1(M) \) one has

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

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

Notice that Proposition 2.11 implies that it suffices to prove Theorems 2.3 and 2.6 under a more restrictive “no deterministic images” assumption instead of the “no common invariant measures” condition. In particular, Theorem 2.6 can be considered a corollary of Theorem 2.8.
3. 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 5.

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 $\nu$ on $M$ define its energy as

$$(3.1) \quad E_\alpha(\nu) := \int \int_M d(x,y)^{-\alpha} d\nu(x) d\nu(y)$$

An immediate application of Markov’s inequality shows that when $E_\alpha(\nu)$ 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

$$(3.2) \quad E_\alpha(\nu) \geq (2r)^{-\alpha} \nu(B_r(x))^2,$$

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.

The key step here is to show that for a sufficiently small $\alpha$, convolution with $\mu$ 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

$$(3.3) \quad \mathcal{E}_\alpha(\mu * \nu) \leq \lambda \mathcal{E}_\alpha(\nu) + C_0.$$

To do so, we re-interpret the energy (3.1) in terms similar to $L_2$-norm of a convenient function. Namely, for a given measure $\nu$, consider the function

$$(3.4) \quad \rho_\alpha[\nu](y) := \int_M \varphi_\alpha(d(x,y)) d\nu(x),$$

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,

$$(3.5) \quad \tilde{\mathcal{E}}_\alpha(\nu) := \int_M \rho_\alpha[\nu](y)^2 d\text{Leb}(y).$$

It turns out that there exists constant $c_\alpha$ such that for any $\nu$

$$(3.6) \quad \tilde{\mathcal{E}}_\alpha(\nu) = (c_\alpha + o(1)) \cdot \mathcal{E}_\alpha(\nu),$$
where \( o(1) \) tends to zero as either of energies \( \mathcal{E}_\alpha \), \( \tilde{\mathcal{E}}_\alpha \) tend to infinity. Indeed, the right hand side of (3.5) can be rewritten as

\[
(3.7) \quad \tilde{\mathcal{E}}_\alpha(\nu) = \int_M \rho_\alpha^2[\nu](y) \, d\text{Leb}(y)
\]

\[
= \iint_{M \times M \times M} \varphi_\alpha(d(x,y)) \varphi_\alpha(d(z,y)) \, d\nu(x) d\nu(z) \, d\text{Leb}(y)
\]

\[
= \iint_{M \times M} K_\alpha(x,z) \, d\nu(x) d\nu(z),
\]

where

\[
(3.8) \quad K_\alpha(x,z) := \int_M \varphi_\alpha(d(x,y)) \varphi_\alpha(d(z,y)) \, d\text{Leb}(y).
\]

If \( M \) was replaced by the Euclidean space \( \mathbb{R}^k \), the kernel (3.8) would have exactly the form \( c_\alpha |x - z|^{-\alpha} \), where \( c_\alpha \) is a constant, due to the symmetry and scaling arguments (\( K_\alpha(x,z) \)) can depend only on the distance between \( x \) and \( z \) and should scale as its \( \alpha \)'s inverse power). Thus, on \( \mathbb{R}^k \) one would have exact proportionality \( c_\alpha \mathcal{E}_\alpha = \tilde{\mathcal{E}}_\alpha \). 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 \( \alpha \) the energy \( \mathcal{E}_\alpha \) is almost unchanged by applying a given diffeomorphism \( f \), as the distances are changed at most \( \mathcal{L}(f) \) times, and their \( \alpha \)'s powers are changed by a factor close to 1. Due to (3.6), this implies that for a large value of energy the same applies to \( \tilde{\mathcal{E}}_\alpha \).

However, the latter energy admits an \( L_2 \)-norm interpretation: the definition (3.5) is the squared length \( \langle \rho, \rho \rangle_{L_2(M, \text{Leb})} \) of a vector (function) \( \rho \in L_2(M) \), linearly associated to the measure \( \nu \). In particular, if the \( L_2 \)-norm of the average of images \( f_* \nu \) (where \( f \) is distributed w.r.t. \( \mu \)) does not decrease, it means that (most of) these vectors are essentially (almost) aligned. Considering the non-probability measure

\[
(3.9) \quad \Theta_\alpha[\nu] := \rho_\alpha[\nu]^2(y) \, d\text{Leb}(y),
\]

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

The relation (3.3) that we have (non-rigorously) established implies that when the energy \( \mathcal{E}_\alpha(\nu) \) is large, the passage to the averaged image
μ * ν decreases it. This already suffices to establish the existence of a Hölder stationary measure. Indeed, take any measure ν₀ of finite α-energy (for instance, the Lebesgue measure on M). Consider the sequence of its averaged images,

\[ νₙ := μ * ν_{n-1}, \quad n = 1, 2, \ldots; \]

their $E_α$ energies then stay uniformly bounded. Hence, the same bound for the energy applies to their Cesaro averages $\bar{ν}_n := \frac{1}{n} \sum_{j=0}^{n-1} ν_j$. 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 (3.3) reads as

\[ E_α(ν) \leq \lambda E_α(ν) + C₀, \]

thus implying that either the energy $E_α(ν) \leq \frac{C₀}{1-λ}$, 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 2.3–2.8 definitions of energy in (3.1) and (3.5) are to be modified to exclude the infinity from the list of possibilities.

Such a modification is done by replacing the interaction kernel $d(x, y)^{-α}$ in the definition of the energy $E_α$ 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 4.5 below)

\[ E_{α,ε}(ν) := \int\int_M U_{α,ε}(d(x, y))dν(x)dν(y). \]

for some bounded interaction potential $U_{α,ε}(r)$. Unfortunately, bluntly taking $[\max(r, ε)]^{-α}$ in the role of $U_{α,ε}(r)$ makes it harder to work with the $L₂$-interpretation. Thus we first modify the $L₂$-definition, applying a cut-off in the construction of the function $ρ$, and then use it to define the energy.

3.1. Structure of the paper. We start by providing rigorous definitions for the cut-off energies $\tilde{E}_{α,ε}$ and $E_{α,ε}$ in Sections 4.1 and 4.2 (see Definitions 4.2 and 4.5 below). We then state some of their properties in Section 4.3. In particular, we show that when the energies are large, they are comparable to each other (see Proposition 4.10 which is analogous to (3.6) above).
Next, in Section 4.4 we introduce measures \( \theta_{\alpha,\epsilon} [\nu] \) associated with the measure \( \nu \), and describe the effect of an action of “not too distorting” diffeomorphisms (see Proposition 4.21).

We provide the formal proof of Theorems 2.3-2.8 in Section 5. The main intermediate result is Proposition 5.1 on exponential decrease of large energies; it is an analog of (3.3) and is established in Section 5.1. The final step in the proof of Theorems 2.3-2.8 is made in Section 5.2.

To make an exposition more reader friendly, the detailed proofs of the most technical statements are postponed till Section 6. Namely, in Section 6.1 we discuss the properties of the functions \( \varphi_{\alpha,\epsilon}(r) \) and \( U_{\alpha,\epsilon}(r) \), and prove Proposition 4.6. In Section 6.2 we provide the proof of Proposition 4.10 that claims that the energies \( E_{\alpha,\epsilon}(\nu) \) and \( \tilde{E}_{\alpha,\epsilon}(\nu) \) cannot differ too much. The proof of Proposition 4.21 is given in Section 6.4. Proposition 4.16 is proven in Section 6.3, and in Section 6.5 the proof of Lemma 5.3 is given.

Finally, in Appendix A we give an example that shows that the finite moment condition (2.2) cannot be replaced by the condition \( \int \log L(f) \, d\mu(f) < \infty \) that would be an analog of the condition on distribution in the classical Furstenberg Theorem on random matrix products.

### 4. Tools for the proof

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

#### 4.1. \( L_2 \)-type energy \( \tilde{E}_{\alpha,\epsilon} \)

To implement the strategy described in Section 3 in a way applicable to any initial probability measure on the manifold, let us start by modifying the definitions of the energy given by (3.1) and (3.5). As \( L_2 \)-point of view is the essential component of the proof, we start by modifying the function used to realise it.

**Definition 4.1.** We define (see Fig. 1)

\[
\varphi_{\alpha,\epsilon}(r) := \begin{cases} 
 r^{-\frac{k+\alpha}{2}}, & r \geq \epsilon \\
 \frac{1}{\epsilon^n} \cdot r^{-\frac{k+\alpha}{2}}, & r < \epsilon.
\end{cases}
\]  

The following definition should be considered an analog of (3.5) and (3.9) from Section 3.
Figure 1. Function $\varphi_{\alpha,\varepsilon}(r)$: it coincides with $r^{-\frac{k+\alpha}{2}}$ (graph drawn by dotted line) for $r > \varepsilon$, and scales as a slower-growing power of $r$ for $r < \varepsilon$

**Definition 4.2.** For a measure $\nu \in \mathcal{M}$, we define a function

\[
\rho_{\alpha,\varepsilon}[\nu](y) := \int_{\mathcal{M}} \varphi_{\alpha,\varepsilon}(d(x, y))d\nu(x).
\]

and a non-probability measure

\[
\Theta_{\alpha,\varepsilon}[\nu] := \rho_{\alpha,\varepsilon}^2[\nu](y) \, d\text{Leb}(y).
\]

We also define

\[
\tilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) := \Theta_{\alpha,\varepsilon}[\nu](M) = \int_{\mathcal{M}} \rho_{\alpha,\varepsilon}^2[\nu](y) \, d\text{Leb}(y).
\]

In the same way as in Section 3, we rewrite the definition (4.3) as a pairwise interaction with some kernel $K_{\alpha,\varepsilon}$:

**Lemma 4.3.**

\[
\tilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) = \int_{\mathcal{M} \times \mathcal{M}} K_{\alpha,\varepsilon}(x, z)d\nu(x)d\nu(z),
\]

where

\[
K_{\alpha,\varepsilon}(x, z) := \int_{\mathcal{M}} \varphi_{\alpha,\varepsilon}(d(x, y))\varphi_{\alpha,\varepsilon}(d(z, y)) \, d\text{Leb}(y).
\]

**Proof.** In the same way as in (3.7), it suffices to substitute (4.2) that defines $\rho_{\alpha,\varepsilon}^2[\nu](y)$, into (4.3), obtaining a triple integral

\[
\tilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) = \int_{\mathcal{M} \times \mathcal{M} \times \mathcal{M}} \varphi_{\alpha,\varepsilon}(d(x, y))\varphi_{\alpha,\varepsilon}(d(z, y)) \, d\nu(x)d\nu(z) \, d\text{Leb}(y),
\]

and then change the order of integration. \(\square\)
4.2. 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 $\tilde{\mathcal{E}}_{\alpha,\varepsilon}$ 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 3, we first replace the manifold $M$ with the Euclidean space $\mathbb{R}^k$ in (4.4):

**Definition 4.4.** For any two points $x, z \in \mathbb{R}^k$ consider the integral

(4.6) \[ I_{\alpha,\varepsilon}(x, z) := \int_{\mathbb{R}^k} \varphi_{\alpha,\varepsilon}(|y - x|)\varphi_{\alpha,\varepsilon}(|y - z|) \, d\text{Leb}(y). \]

Due to the symmetry of the problem, this integral depends only on the distance $r = |x - z|$ between these two points:

\[ I_{\alpha,\varepsilon}(x, z) = U_{\alpha,\varepsilon}(|x - z|) \]

for some function $U_{\alpha,\varepsilon}$ of $r = |x - z|$. We take this as a definition of $U_{\alpha,\varepsilon}(r)$.

Now we are ready to give a definition of energy $\mathcal{E}_{\alpha,\varepsilon}(\nu)$ that is a modified version of (3.1):

**Definition 4.5.** We define

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

![Figure 2. The function $U_{\alpha,\varepsilon}(r)$](image)

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).

**Proposition 4.6.** 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

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

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

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

also, for any $r > 0$

$$\frac{U_{\alpha,\varepsilon}(r)}{c_\alpha r^{-\alpha}} \rightarrow 1, \quad \text{as } \varepsilon \rightarrow 0.$$

3. **($\alpha$-slow changing)** For any $\varepsilon$ and any $r' > r > 0$, one has

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

We postpone the proof of these properties until Section 6.1. Specifically, conclusion 1 will be proven in Lemma 6.4 and Lemma 6.6, conclusion 2 in Corollary 6.2 and Lemma 6.5, and conclusion 3 in Proposition 6.3.

Meanwhile, we will be using Proposition 4.6 to prove our main results. We will assume that for any given $\alpha \in (0, \frac{1}{2})$ the constants $c_\alpha, c'_\alpha$ are as in the conclusions of this proposition.

Note that the value of $U_{\alpha,\varepsilon}(0)$, appearing in conclusion 1 of Proposition 4.6, can be calculated explicitly:

**Lemma 4.7.** $U_{\alpha,\varepsilon}(0) = \frac{L}{\alpha} \varepsilon^{-\alpha}$, where $L = 2 \text{Leb} B_1(0)$ is a constant.

**Proof.**

$$U_{\alpha,\varepsilon}(0) = \int_{\mathbb{R}^k} \varphi_{\alpha,\varepsilon}(|x|)^2 \text{dLeb}(x) =$$

$$= \int_{|x|<\varepsilon} \left(\frac{|x|^\frac{\alpha}{2} - \frac{\alpha}{2}}{\varepsilon^\alpha}\right)^2 \text{dLeb}(x) + \int_{|x|>\varepsilon} \left(|x|^\frac{\alpha}{2} - \frac{\alpha}{2}\right)^2 \text{dLeb}(x) \leq$$

$$= \text{Leb}(B_1(0)) \left(\frac{r^\alpha}{\alpha \varepsilon^{2\alpha}} \bigg|_{\epsilon}^{+\infty} \right) = \frac{L}{\alpha} \varepsilon^{-\alpha}.$$

Due to the monotonicity conclusion 1 of Proposition 4.6 one gets the following uniform upper bound:
Corollary 4.8. For any $\alpha \in (0, \frac{1}{2})$ there exists $C''_\alpha$ such that for any measure $\nu$ 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

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

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

Proposition 4.6 also allows to have Hölder type bounds in terms of the energy, the statement that is an analogue to (3.2) above:

Lemma 4.9. For any $\alpha, \varepsilon > 0$, any probability measure $\nu$ on $M$ one has

$$(4.10) \quad \forall y \in M \quad \forall r > \varepsilon \quad \nu(B_r(y)) \leq \sqrt{\frac{\mathcal{E}_{\alpha, \varepsilon}(\nu)^{2\alpha}}{c'_\alpha (2r)^{-\alpha}}} \cdot r^{\frac{2\alpha}{\alpha}},$$

where $c'_\alpha$ is given by (4.8) from Proposition 4.6.

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 (4.8). Dividing by $c'_\alpha (2r)^{-\alpha}$ and taking square root, we obtain the desired (4.10). □

4.3. Properties of the energies $\mathcal{E}_{\alpha, \varepsilon}$ and $\tilde{\mathcal{E}}_{\alpha, \varepsilon}$. In the same way as for (3.6), the two energies, $\mathcal{E}_{\alpha, \varepsilon}$ and $\tilde{\mathcal{E}}_{\alpha, \varepsilon}$, are comparable. Namely, we have the following statement.

Proposition 4.10. For any $\alpha \in (0, 1/2)$ and any $\delta > 0$ there exists $C > 0$ such that for any $\varepsilon > 0$ and any measure $\nu$ on $M$ such that $\tilde{\mathcal{E}}_{\alpha, \varepsilon}(\nu) > C$ or $\mathcal{E}_{\alpha, \varepsilon}(\nu) > C$ one has

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

The idea of the proof is the same as for (3.6): 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 6.2.

It is convenient to re-formulate Proposition 4.10 using the following notation:
Definition 4.11. We say that two positive numbers, \( A \) and \( A' \), are \((\delta,C)\)-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 4.12. Proposition 4.10 can be equivalently reformulated in the following way, that includes all possible probability measures \( \nu \), not necessarily the high energy ones. For any \( \alpha \in (0, 1/2) \) and any \( \delta > 0 \) there exists \( C > 0 \) such that
\[
\forall \varepsilon > 0 \quad \forall \nu \quad E_{\alpha,\varepsilon}(\nu) \approx_{(\delta,C)} \tilde{E}_{\alpha,\varepsilon}(\nu).
\]

The next proposition and its corollary give a bound on how the energy \( E_{\alpha,\varepsilon} \) can change under an action of a diffeomorphism. Namely, it is changed by a factor that is at most \( \mathcal{L}(f)^{\alpha} \), thus once \( \alpha \) is close to 0, the factor is close to 1.

Proposition 4.13. For any \( \alpha \), any \( \varepsilon > 0 \) and any \( f \in \text{Diff}^1(M) \) one has
\[
\mathcal{L}(f)^{-\alpha} \leq \frac{E_{\alpha,\varepsilon}(f_*\nu)}{E_{\alpha,\varepsilon}(\nu)} \leq \mathcal{L}(f)^{\alpha}.
\]

Proof. Indeed, by definition of \( \mathcal{L}(f) \) we know that
\[
d(f(x), f(z)) \leq \mathcal{L}(f)d(x,z) \quad \text{for any } x, z \in M.
\]
Proposition 4.6 thus implies (see (4.9)) that
\[
\mathcal{L}(f)^{-\alpha} U_{\alpha,\varepsilon}(d(x,z)) \leq U_{\alpha,\varepsilon}(d(f(x), f(z)))
\]
Integrating over \( d\nu(x) d\nu(z) \), one gets the desired
\[
\mathcal{L}(f)^{-\alpha} E_{\alpha,\varepsilon}(\nu) \leq E(f_*\nu).
\]
The second inequality follows in the same way from
\[
\mathcal{L}(f)^{-1}d(x,z) \leq d(f(x), f(z))
\]
□

Corollary 4.14. 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 \( f \in \text{Diff}^1(M) \) with \( \mathcal{L}(f) < R \) one has
\[
\frac{E_{\alpha,\varepsilon}(f_*\nu)}{E_{\alpha,\varepsilon}(\nu)} \in (1 - \delta, 1 + \delta).
\]
Joining this statement with Proposition 4.10, we get the same statement for the energy $\tilde{\mathcal{E}}_{\alpha,\varepsilon}$:

**Corollary 4.15.** 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 $f \in \text{Diff}^1(M)$ with $\mathcal{L}(f) < R$, any $\varepsilon > 0$ and any $\nu$ such that $\tilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) > C$ or $\mathcal{E}_{\alpha,\varepsilon}(\nu) > C$ one has

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

A final concluding remark in this direction is that the statement of Corollary 4.15 survives if one takes an expectation of energy of a random image, provided that the moment condition (2.2) is satisfied.

**Proposition 4.16.** If a probability measure $\mu \in \mathcal{M}$ satisfies assumption (2.2) 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 $\nu$ on the manifold $M$ with $\tilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) > C$ one has:

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

For any given individual diffeomorphism $f$ and any measure $\nu$, the quotient $\mathcal{E}_{\alpha,\varepsilon}(f_*\nu)/\mathcal{E}_{\alpha,\varepsilon}(\nu)$ does not exceed $\mathcal{L}(f)^\alpha$ and hence converges to 1 as $\alpha$ tends to 0 uniformly in $\varepsilon$. Now, conclusion (4.13) is in a sense bringing the expectation under the limit, that can be justified via the Lebesgue dominated convergence theorem. We provide a formal proof of this proposition in Section 6.3.

4.4. **Normalizations $\theta_{\alpha,\varepsilon}$, Wesserstein distance.** In the same way as in the sketch of the proof, we consider the normalizations $\theta_{\alpha,\varepsilon}$ of measures $\Theta_{\alpha,\varepsilon}$. 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 4.17.**

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

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 4.18. Let \( \nu_1, \nu_2 \) be two probability measures on a measure space \( (M, \mathcal{B}) \). Then the Wasserstein distance between them is defined as

\[
W(\nu_1, \nu_2) = \inf_{\gamma} \int_{M \times M} d(x, y) \, d\gamma(x, y),
\]

where the infimum is taken over all probability measures \( \gamma \) 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 4.19. Let \( \nu_1, \nu_2 \) be two probability measures on a measure space \( (M, \mathcal{B}) \). Then the total variation distance between them is

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

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

Lemma 4.20. For any two probability measures \( \nu_1, \nu_2 \) on a manifold \( M \) we have

\[
W(\nu_1, \nu_2) \leq \text{diam}(M) \cdot TV(\nu_1, \nu_2).
\]

We are going to use the following statement. Assuming that \( \alpha \) is sufficiently small, the energy of a measure \( \nu \) is sufficiently high, and a diffeomorphism \( f \) is “not too distorting”, not only are the energies of \( \nu \) 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 4.21. 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 \( f \in \text{Diff}^1(M) \) with \( \mathcal{L}(f) < R \), any \( \varepsilon > 0 \), and any \( \nu \) such that \( \tilde{\mathcal{E}}_{\alpha,\varepsilon}(\nu) > C \) or \( \mathcal{E}_{\alpha,\varepsilon}(\nu) > C \) one has

\[
W(f_*\theta_{\alpha,\varepsilon}(\nu), \theta_{\alpha,\varepsilon}(f_*\nu)) < \delta.
\]

The idea of the proof of this proposition is close to the one for Proposition 4.10: large energy means that most of it comes from points close to each other. Moreover, considering the energy as a triple integral (4.5), 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 \( U_{\alpha,\varepsilon}(x, z) \). Finally, the latter almost does not change if \( \mathcal{L}(f)^\alpha \) is close to 1, thus implying the desired closeness of two measures.

Once again, we postpone the formal proof of this proposition until Section 6.4.
5. PROOFS OF MAIN RESULTS

In this section we prove Theorems 2.3, 2.6, and 2.8, except for the technically complicated parts, that we moved to Section 6 (specifically, the proofs of Propositions 4.6, 4.10, 4.16, 4.21, together with the proof of Lemma 5.3).

5.1. Exponential decrease of large energies. The key step of the proof of our main results is the following proposition:

Proposition 5.1. If the measure $\mu$ satisfies finite moment condition and no deterministic images then there exist $\alpha > 0, C < \infty, \lambda < 1$, such that for any $\varepsilon > 0$ if $\tilde{E}_{\alpha, \varepsilon}(\nu) > C$ then

$$\tilde{E}_{\alpha, \varepsilon}(\mu \ast \nu) < \lambda \tilde{E}_{\alpha, \varepsilon}(\nu).$$

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 $\nu$ with energy $\tilde{E}_{\alpha, \varepsilon}(\nu) > C$, such that

$$\tilde{E}_{\alpha, \varepsilon}(\mu \ast \nu) \geq (1 - \delta)\tilde{E}_{\alpha, \varepsilon}(\nu).$$

Recall the following standard statement:

Lemma 5.2. 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

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

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 (5.3): 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 $L^2(M, \text{Leb})$. Namely, for a given measure $\nu$ its averaged image $\mu \ast \nu$ is the expectation of $f_*\nu$, where the diffeomorphism $f$ of $M$ is taken randomly w.r.t. the measure $\mu$. Hence, the same applies for the density $\rho_{\alpha, \varepsilon}[\nu]$, given by (4.2) (see Definition 4.2):

$$\rho_{\alpha, \varepsilon}[\mu \ast \nu] = \mathbb{E}_\mu \rho_{\alpha, \varepsilon}[f_*\nu].$$
Substituting this into (5.3) (with \( v = \rho_{\alpha, \varepsilon}[f_* \nu] \)), and taking into account the definition (4.3), we get

\[
\begin{align*}
\mathbb{E}_\mu \left[ \int_M (\rho_{\alpha, \varepsilon}[f_* \nu](y) - \rho_{\alpha, \varepsilon}[\mu \ast \nu](y))^2 \, d\text{Leb}(y) \right] &=
\mathbb{E}_\mu \left[ \tilde{\mathcal{E}}_{\alpha, \varepsilon}(f_* \nu) \right] - \tilde{\mathcal{E}}_{\alpha, \varepsilon}(\mu \ast \nu);
\end{align*}
\]

in particular,

\[
\tilde{\mathcal{E}}_{\alpha, \varepsilon}(\mu \ast \nu) \leq \mathbb{E}_\mu \left[ \tilde{\mathcal{E}}_{\alpha, \varepsilon}(f_* \nu) \right].
\]

By Proposition 4.16 we can find sufficiently small \( \alpha > 0 \) and sufficiently large \( C < \infty \), such that for any \( \varepsilon > 0 \) and any measure \( \nu \) on \( M \) with \( \tilde{\mathcal{E}}_{\alpha, \varepsilon}(\nu) > C \) the inequality (4.13) holds, and thus (taking only the upper bound)

\[
\mathbb{E}_\mu \left[ \tilde{\mathcal{E}}_{\alpha, \varepsilon}(f_* \nu) \right] \leq (1 + \delta) \tilde{\mathcal{E}}_{\alpha, \varepsilon}(\nu).
\]

Denote by \( \nu \) a probability measure with \( \tilde{\mathcal{E}}_{\alpha, \varepsilon}(\nu) > C \), for which the inequality (5.2) holds. Substituting (5.7) and (5.2) in the right hand side of (5.5), we get

\[
\mathbb{E}_\mu \left[ \int_M (\rho_{\alpha, \varepsilon}(f_* \nu)(y) - \rho_{\alpha, \varepsilon}(\mu \ast \nu)(y))^2 \, d\text{Leb}(y) \right] < 2 \delta \tilde{\mathcal{E}}_{\alpha, \varepsilon}(\nu).
\]

and thus

\[
\mathbb{E}_\mu \left[ \int_M (\rho_{\alpha, \varepsilon}(f_* \nu)(y) - \rho_{\alpha, \varepsilon}(\mu \ast \nu)(y))^2 \frac{d\text{Leb}(y)}{\tilde{\mathcal{E}}_{\alpha, \varepsilon}(\nu)} \right] < 2 \delta.
\]

The final step is to show that measures

\[
\theta_{\alpha, \varepsilon}(f_* \nu) = \frac{\rho_{\alpha, \varepsilon}(f_* \nu)(y)^2}{\tilde{\mathcal{E}}_{\alpha, \varepsilon}(f_* \nu)} \, d\text{Leb}(y)
\]

and

\[
\theta_{\alpha, \varepsilon}(\mu \ast \nu) = \frac{\rho_{\alpha, \varepsilon}(\mu \ast \nu)(y)^2}{\tilde{\mathcal{E}}_{\alpha, \varepsilon}(\mu \ast \nu)} \, d\text{Leb}(y)
\]

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

**Lemma 5.3.** Under the assumptions above

\[
\mathbb{P}_\mu \left[ \text{TV}(\theta_{\alpha, \varepsilon}(f_* \nu), \theta_{\alpha, \varepsilon}(\mu \ast \nu)) > 10 \sqrt{\delta} \right] < 4 \sqrt{\delta},
\]

and hence

\[
\mathbb{P}_\mu \left[ W(\theta_{\alpha, \varepsilon}(f_* \nu), \theta_{\alpha, \varepsilon}(\mu \ast \nu)) > 10 \text{diam}(M) \sqrt{\delta} \right] < 4 \sqrt{\delta}.
\]
Proof of Lemma 5.3 is the most technical part of the proof of Proposition 5.1, and we postpone it till Section 6.5.

We are now ready to conclude the proof of Proposition 5.1. Namely, the conclusion (5.11) 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 $f_*$-image of a given measure, $\theta_{\alpha,\varepsilon}(\nu)$.

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

$$P_{\mu}[L(f) < R] > 1 - \sqrt{\delta}$$

By Proposition 4.21 we can choose $C$ big enough, so that for any $f \in \text{Diff}^1(M)$, such that $L(f) < R$ inequality (4.15) holds, and using triangle inequality we arrive to

$$(5.12) \quad P_{\mu} \left[ W(f_*\theta_{\alpha,\varepsilon}(\nu), \theta_{\alpha,\varepsilon}(\mu_*\nu)) > 10\text{diam}(M)\sqrt{\delta} + \delta \right] < 5\sqrt{\delta}.$$ 

Now consider the sequence $\delta_n = \frac{1}{n}$ and denote by $\nu_n$ the corresponding measure and corresponding parameters by $\alpha_n, \varepsilon_n$, for which (5.2) 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 $\theta_{\alpha_n,\varepsilon_n}(\mu_*\nu_n)$ also converge. It remains to notice that due to inequality (5.12) the limit measure

$$m = \lim_{k \to \infty} \theta_{\alpha_n,\varepsilon_n}(\nu_n)$$

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

$$\tilde{m} = \lim_{k \to \infty} \theta_{\alpha_n,\varepsilon_n}(\mu_*\nu_n),$$

which contradicts our assumption. \qed

Finally, Proposition 5.1 can be generalized by the following statement.

**Proposition 5.4.** 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

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

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


Then the constants $\alpha, \lambda$ and $C$ in Proposition 5.1 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 4 the constants can be chosen uniformly for all $\mu \in \mathcal{K}$. Now, repeating the proof of Proposition 5.1 we see that the only part to be modified is passage to the limit. Again taking $\delta_n = \frac{1}{n}$ one finds $\alpha_n, \varepsilon_n$ and a measure $\mu_n \in \mathcal{K}$ such that (5.2) holds. Passing to a subsequence $(n_k)$, one can ensure the existence of three limits:

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

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

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

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 5.1 can be modified to include all possible measures $\nu$ (without assuming the energy higher than $C$). To do so, we need to adjust the upper bound (5.1) in order to include low-energy measures $\nu$:

Lemma 5.5. Assume that the measure $\mu$ satisfies the finite moment condition (2.2) with some $\gamma$, and that $\alpha < \gamma$ and $C$ are given. Then there exists $C'$ such that for any measure $\nu$ on $M$ with $\mathcal{E}_{\alpha, \varepsilon}(\nu) \leq C$ one has

$$\mathcal{E}_{\alpha, \varepsilon}(\mu * \nu) \leq C'. \quad (5.13)$$

Corollary 5.6. In the assumptions of Proposition 5.1 one can conclude that there exist $\alpha > 0$, $\tilde{C} < \infty$, $\lambda < 1$, such that for any $\varepsilon > 0$ and any measure $\nu$ on $M$

$$\mathcal{E}_{\alpha, \varepsilon}(\mu * \nu) < \max(\lambda \mathcal{E}_{\alpha, \varepsilon}(\nu), \tilde{C}) \quad (5.14)$$

Proof of Lemma 5.5. Note first that due to Proposition 4.10 for any $\alpha$ and $C_1$ there exists $C_2$ such that for any $\varepsilon$ and any measure $\nu$ one has

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

and vice versa,

$$\mathcal{E}_{\alpha, \varepsilon}(\nu) < C_1 \Rightarrow \tilde{\mathcal{E}}_{\alpha, \varepsilon}(\nu) < C_2. \quad (5.15)$$

In other words, having an upper bound on one of the energies $\mathcal{E}_{\alpha, \varepsilon}(\nu)$, $\tilde{\mathcal{E}}_{\alpha, \varepsilon}(\nu)$ implies a bound for the other one.
Next, the finite moment condition implies the finiteness of the expectation
\[
C_I := \mathbb{E}_\mu \mathcal{L}(f)^\alpha < \infty.
\]
Now, (5.5) implies that for any measure \( \nu \) we have
\[
\tilde{\mathcal{E}}_{\alpha,\epsilon}(\mu * \nu) \leq \mathbb{E}_\mu \tilde{\mathcal{E}}_{\alpha,\epsilon}(f_*\nu);
\]
At the same time Proposition 4.10 for \( \delta = \frac{1}{2} \) implies that for some constant \( C_1 \) one has for any measure \( \nu' \)
\[
\tilde{\mathcal{E}}_{\alpha,\epsilon}(\nu') \leq \max(2\mathcal{E}_{\alpha,\epsilon}(\nu'), C_1).
\]
Applying this for \( \nu' = f_*\nu \) and joining it with Proposition 4.13, we get
\[
\tilde{\mathcal{E}}_{\alpha,\epsilon}(f_*\nu) \leq \max(2\mathcal{E}_{\alpha,\epsilon}(f_*\nu), C_1) \leq 2\mathcal{L}(f)^\alpha \mathcal{E}_{\alpha,\epsilon}(\nu) + C_1.
\]
Taking the expectation w.r.t. \( \mu \) and using (5.16), we finally get a uniform bound
\[
\tilde{\mathcal{E}}_{\alpha,\epsilon}(\mu * \nu) \leq 2C_I C + C_1 =: C'.
\]

**Proof of Corollary 5.6.** In the proof of Proposition 5.1 we can take \( \alpha \) arbitrarily small, so we can assume that \( \alpha < \gamma \). Applying Proposition 4.10 to both sides of (5.1), where we take \( \delta \) sufficiently small so that \( \frac{1}{1+\delta} > \lambda \), allows to conclude that there exists some constant \( C_3 \) such that for all \( \epsilon \) and \( \nu \)
\[
\mathcal{E}_{\alpha,\epsilon}(\mu * \nu) < \lambda' \mathcal{E}_{\alpha,\epsilon}(\nu) \quad \text{if} \quad \mathcal{E}_{\alpha,\epsilon}(\nu) > C_3,
\]
where \( \lambda' = \frac{1+\delta}{1+\delta}\lambda < 1 \). Meanwhile, Lemma 5.5 implies that there exists some \( C_4 \) such that
\[
\mathcal{E}_{\alpha,\epsilon}(\mu * \nu) < C_4 \quad \text{if} \quad \mathcal{E}_{\alpha,\epsilon}(\nu) \leq C_3.
\]
Taking \( \tilde{C} := \max(C_3, C_4) \), we get the desired (5.14). \( \square \)

### 5.2. Hölder bounds: proofs of Theorems 2.3–2.8

Now we are ready to complete the proof of Theorem 2.6 (and thus of Theorem 2.3).

**Proof of Theorem 2.6.** Let \( \alpha, \check{C} \) be as in Corollary 5.6; for any \( \epsilon \) and for any measure \( \nu \) one has
\[
\mathcal{E}_{\alpha,\epsilon}(\mu * \nu) < \max(\lambda \mathcal{E}_{\alpha,\epsilon}(\nu), \check{C});
\]
applying this \( n \) times, we get
\[
\mathcal{E}_{\alpha,\epsilon}(\mu^n * \nu) < \max(\lambda^n \mathcal{E}_{\alpha,\epsilon}(\nu), \check{C}).
\]
Recall that Lemma 4.7 gives a uniform upper bound $E_{\alpha,\varepsilon}(\nu) \leq C''_{\alpha} \varepsilon^{-\alpha}$, where $C''_{\alpha} = \frac{L}{\alpha}$; choose $\varepsilon := \lambda^n \alpha$, then we have

\[ E_{\alpha,\varepsilon}(\mu^n \ast \nu) < \max (2C, C''_{\alpha}) \cdot r^{\frac{\alpha}{n}}, \]

which completes the proof of Theorem 2.6.

Now, applying Lemma 4.9 for any $r > \varepsilon = \kappa^n$, where $\kappa := \lambda^{\frac{1}{2}}$, we deduce that

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

Arguing in the same way as in the proof of Theorem 2.6, Theorem 2.8 can be deduced from Proposition 5.4.

**Remark 5.7.** The same proof works for the setting mentioned in Remark 2.9. One has to consider an embedding $M \hookrightarrow \tilde{M}$ into a closed manifold $\tilde{M}$, to define energies $E_{\alpha,\varepsilon}$ and $\tilde{E}_{\alpha,\varepsilon}$ with respect to the Leb measure on $\tilde{M}$ and to repeat the original proof verbatim.

### 6. Proofs of used properties

Here we provide the proofs of the technical statements that were formulated and used above, as described in Section 3.1.

#### 6.1. Properties of $\varphi_{\alpha,\varepsilon}(r)$ and $U_{\alpha,\varepsilon}(r)$

Let us start with some scaling properties of $\varphi_{\alpha,\varepsilon}$ and of $U_{\alpha,\varepsilon}$:

**Lemma 6.1.** For any $\lambda > 0$ and any $\alpha, \varepsilon > 0$ we have

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

and

\[ \forall r > 0 \quad U_{\alpha,\lambda \varepsilon}(\lambda r) = \lambda^{-\alpha} U_{\alpha,\varepsilon}(r). \]

In particular, one has

\[ U_{\alpha,\varepsilon}(r) = r^{-\alpha} U_{\alpha,\varepsilon}(1) = \varepsilon^{-\alpha} U_{\alpha,1}(r/\varepsilon). \]

**Proof.** Relation (6.1) follows directly from (4.1) in the definition of $\varphi_{\alpha,\varepsilon}$. For (6.2), it suffices to make a change of variable $y = \lambda y'$ in the integral (4.6). Finally, (6.3) immediately follows from (6.2). \qed

**Corollary 6.2.** 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 \(\varepsilon\), the function \(\varphi_{\alpha,\varepsilon}(r)\) and hence \(U_{\alpha,\varepsilon}(r)\) is decreasing. The conclusion thus directly follows from (6.3). \(\square\)

**Proposition 6.3.**

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

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

Proof. Note first that, on each of the intervals \((0, \varepsilon]\) and \([\varepsilon, \infty)\) the function \(\varphi_{\alpha,\varepsilon}\) takes the form \(\text{const} \cdot r^{-\frac{k}{2} + \frac{\alpha}{2}}\), 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 \((r, \varepsilon]\) and for \((\varepsilon, r']\).

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 (4.6) for \(I_{\alpha,\varepsilon}(x, z)\):

\[
U_{\alpha,\varepsilon}(r) = I_{\alpha,\varepsilon}(x, z) = \int_{\mathbb{R}^k} \varphi_{\alpha,\varepsilon}(|y - x|) \varphi_{\alpha,\varepsilon}(|y - z|) \, d\text{Leb}(y) = \int_{\mathbb{R}^k} \varphi_{\alpha,\varepsilon}(|\lambda y' - \lambda x'|) \varphi_{\alpha,\varepsilon}(|\lambda y' - \lambda z'|) \, d\text{Leb}(\lambda y') = \lambda^k \int_{\mathbb{R}^k} \varphi_{\alpha,\varepsilon}(\lambda |y' - x'|) \varphi_{\alpha,\varepsilon}(\lambda |y' - z'|) \, d\text{Leb}(y')
\]

Now, due to (6.4),

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

Hence, \(U_{\alpha,\varepsilon}(r)\) differs from

\[
\lambda^k \int_{\mathbb{R}^k} \varphi_{\alpha,\varepsilon}(|y' - x'|) \varphi_{\alpha,\varepsilon}(|y' - z'|) \, d\text{Leb}(y') = \lambda^k U(r')
\]

by a factor that is between \(\lambda^{-(k - \alpha)}\) and \(\lambda^{-(k + \alpha)}\). As \(\lambda^k\) cancels out, we get the desired

\[
\lambda^\alpha U(r') \leq U(r) \leq \lambda^{-\alpha} U(r').
\]

\(\square\)

A first immediate remark is the finiteness of \(U\):
Lemma 6.4. For any $\alpha \in (0, \frac{1}{2})$, $\varepsilon > 0$, and any $r \geq 0$ the value $U_{\alpha, \varepsilon}(r)$ is finite, and
\begin{equation}
\forall r \geq 0 \quad U_{\alpha, \varepsilon}(r) \leq \langle \varphi_{\alpha, \varepsilon}, \varphi_{\alpha, \varepsilon} \rangle_{L^2(\mathbb{R}^k)} = U_{\alpha, \varepsilon}(0).
\end{equation}

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, $U(r)$ is a scalar product between two $r$-shifted copies of $\varphi_{\alpha, \varepsilon}$. Hence, it is finite for all $r \geq 0$, and we get the desired (4.7). \qed

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 6.5. For any $\alpha \in (0, \frac{1}{2})$ the function $U_{\alpha, \varepsilon}(r)$ is decreasing in $\varepsilon$, and one has
\begin{equation}
\forall r > 0 \quad \lim_{\varepsilon \to 0} U_{\alpha, \varepsilon}(r) = c_\alpha r^{-\alpha}
\end{equation}
where
\[ c_\alpha := \int_{\mathbb{R}^k} |y|^{-\frac{k+\alpha}{2}} |y - (1, 0, \ldots, 0)|^{-\frac{k+\alpha}{2}} \, d\text{Leb}(y) \]
is a constant depending only on $\alpha$.

Proof. The monotonicity part immediately comes from the monotonicity of $\varphi_{\alpha, \varepsilon}$ in $\varepsilon$. As the functions under the integral in (4.6) are positive and monotonous in $\varepsilon$, 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 $c_\alpha$ comes from substituting $r = 1$. \qed

Lemma 6.6. 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 $1_{I_{B_R(0)}}(x) = 1_{|x|<R}$ 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. 3). But that is a partial case of the following simple statement:

Lemma 6.7. Suppose $A \subset \mathbb{R}^k$ is convex, $B \subset A$ is measurable, and $\text{Leb}(B) < \infty$. Then for any non-zero vector $\bar{v} \in \mathbb{R}^k$ the measure $\text{Leb}(A \cap (B + t\bar{v}))$ as a function of $t \geq 0$ is non-increasing.
Indeed, monotonicity in Lemma 6.7 follows from the fact that due to convexity of $A$, for any point $b \in B$, as soon as $b + t\tilde{v} \notin A$, we must have $b + t\tilde{v} \notin A$ for all $t > t_1$. \hfill \square

6.2. **Comparing $E_{\alpha,\varepsilon}(\nu)$ and $\tilde{E}_{\alpha,\varepsilon}(\nu)$: Proof of Proposition 4.10.** We will need the following statement:

**Lemma 6.8.** 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:

\begin{equation}
\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)), \text{ i.e.}
\end{equation}

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

Postponing its proof for the moment, note that it implies Proposition 4.10. \hfill \square

**Proof of Proposition 4.10.** It suffices to integrate (6.8) w.r.t. $\nu \times \nu(x, z)$. As constant $C$ does not depend on these points nor on the measure $\nu$, one gets

\[
\frac{1}{1 + \delta}(E(\nu) - C) < \tilde{E}(\nu) < (1 + \delta)E(\nu) + C
\]

with the same constant $C$. As it was noticed in Remark 4.12, this is an equivalent form of Proposition 4.10. \hfill \square

**Proof of Lemma 6.8.** Let $\alpha > 0$ be given. Take any $r_0 > 0$ and divide the integral (4.4) defining $K_{\alpha,\varepsilon}(x, z)$ into two parts, depending on
whether the distance $d(x, y)$ exceeds $r_0$:

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

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

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:

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

The second summand in (6.9) does not exceed $\text{vol}(M) \cdot \varphi_{\alpha}(r_0)$, 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 uniformly bounded, and the integral of $r^{-\frac{k+\alpha}{2}}$ on $B_{r_0}(0) \subset \mathbb{R}^k$ converges.

We thus have

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

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

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 $\psi$ are $\delta_1$-close to 1:

$$\mathcal{L}(\psi) < 1 + \delta_1, \quad \text{Jac}(\psi) \in \left(\frac{1}{1 + \delta_1}, 1 + \delta_1\right).$$
Making a change of variables $y = \psi(y')$ in the integral for $K_{\alpha,\varepsilon}^{(\nu)}(x, z)$, we get

$$K_{\alpha,\varepsilon}^{(\nu)}(x, z) = \int_{B_{r_0}(x)} \varphi_{\alpha,\varepsilon}(d(x, y))\varphi_{\alpha,\varepsilon}(d(z, y))\,d\text{Leb}_M(y)$$

$$= \int_{B_{r_0}(x')} \varphi_{\alpha,\varepsilon}(d(\psi(x'), \psi(y'))\varphi_{\alpha,\varepsilon}(d(\psi(z'), \psi(y')))\,d\text{Leb}_M(\psi(y'));$$

all the three quotients

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

are close to 1: the first two by a factor $(1 + \delta_1)^{\frac{k + \alpha}{\alpha}}$ due to Proposition 6.3 and the last one by the factor $(1 + \delta_1)$. Thus, $K_{\alpha,\varepsilon}^{(\nu)}(x, z)$ differs from

$$(6.11) \quad I_{\alpha,\varepsilon}^{(\nu)}(x', z') := \int_{B_{r_0}(x')} \varphi_{\alpha,\varepsilon}(|x' - y'|)\varphi_{\alpha,\varepsilon}(|z' - y'|)\,d\text{Leb}_{\mathbb{R}^k}(y')$$

by the factor at most $(1 + \delta_1)^{k+\alpha+1}$. Now, $I_{\alpha,\varepsilon}^{(\nu)}(x', z')$ is also a part of $I_{\alpha,\varepsilon}(x', z')$, that differs from it by

$$\tilde{T}_{\alpha,\varepsilon}^{(\nu)}(x', z') := \int_{\mathbb{R}^k \setminus B_{r_0}(x')} \varphi_{\alpha,\varepsilon}(|x' - y'|)\varphi_{\alpha,\varepsilon}(|z' - y'|)\,d\text{Leb}_{\mathbb{R}^k}(y'),$$

that is bounded uniformly in $\varepsilon, x', z'$ (for $|x' - z'| < \frac{\sigma_2}{2}$ due to the convergence of the integral of $r^{-k-\alpha}$ at infinity, for $|x' - z'| \geq \frac{\sigma_2}{2}$ due to Lemma 6.5). Thus, first choosing $\delta_1$ so that $(1 + \delta_1)^{k+\alpha+2} < 1 + \delta$ and then accordingly choosing $r_0$, we obtain the desired estimate (6.8). \hfill \Box

6.3. $E_\mu[\tilde{E}_{\alpha,\varepsilon}(f_*\nu)]$ vs $\tilde{E}_{\alpha,\varepsilon}(\nu)$: Proof of Proposition 4.16.

Proof of Proposition 4.16. Let us start by establishing that an estimate (4.13) holds with $\tilde{E}_{\alpha,\varepsilon}$ replaced by $E_{\alpha,\varepsilon}$. Indeed, from Proposition 4.13 we know that for any $f \in \text{Diff}^1(M)$ one has

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

Taking the expectation with respect to measure $\mu$ we get

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

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

$$\mathbb{E} \left[ \mathcal{L}(f)^\alpha \right] \leq (\mathbb{E} \left[ \mathcal{L}(f)^{\gamma} \right])^{\alpha/\gamma}$$
and by Jensen’s inequality we also have
\[
(\mathbb{E} [\mathcal{L}(f)^\alpha])^{-1} \leq \mathbb{E} [\mathcal{L}(f)^{-\alpha}] .
\]
Hence, taking \( C_0 \) such that \( \mathbb{E}_\mu \mathcal{L}(f)^\gamma < C_0 \), one gets that for any \( \alpha < \gamma \),
\[
\frac{\mathbb{E}_\mu [E_{\alpha,\varepsilon}(f_*\nu)]}{E_{\alpha,\varepsilon}(\nu)} \in (C_0^{-\frac{\alpha}{\gamma}}, C_0^{\frac{\alpha}{\gamma}}).
\]
Since \( C_0^{\frac{\alpha}{\gamma}} \) tends to 1 as \( \alpha \) tends to 0, for any \( \delta_E > 0 \) there exists \( \alpha_0 \) such that for any \( \alpha \in (0, \alpha_0) \) one has
\[
(6.12) \quad \frac{\mathbb{E}_\mu [E_{\alpha,\varepsilon}(f_*\nu)]}{E_{\alpha,\varepsilon}(\nu)} \in \left( \frac{1}{1 + \delta_E}, 1 + \delta_E \right);
\]
Let us now return back to the original (4.13). Namely, let \( \delta > 0 \) be given; choose and fix \( \delta_E > 0 \) such that \( (1 + \delta_E)^3 < 1 + \delta \), and let \( \alpha_0 \) be chosen w.r.t. \( \delta_E \) so that (6.12) holds.

By Proposition 4.10 there exists \( C_E > 0 \) such that for any measure \( \nu' \) on \( M \) and any \( \varepsilon > 0 \) one has
\[
(6.13) \quad \frac{1}{1 + \delta_E} E_{\alpha,\varepsilon}(\nu') - \frac{C_E}{1 + \delta_E} < \bar{E}_{\alpha,\varepsilon}(\nu') < (1 + \delta_E)E_{\alpha,\varepsilon}(\nu') + C_E
\]
Applying this for \( \nu' = f_*\nu \) and taking the expectation w.r.t. \( \mu \) provides
\[
\frac{1}{1 + \delta_E} \mathbb{E}_\mu [E_{\alpha,\varepsilon}(f_*\nu)] - \frac{C_E}{1 + \delta_E} < \mathbb{E}_\mu \left[ \bar{E}_{\alpha,\varepsilon}(f_*\nu) \right] < (1 + \delta_E)\mathbb{E}_\mu \left[ E_{\alpha,\varepsilon}(f_*\nu) \right] + C_E;
\]
using (6.12), we thus get
\[
\frac{1}{(1 + \delta_E)^2} E_{\alpha,\varepsilon}(\nu) - \frac{C_E}{1 + \delta_E} < \mathbb{E}_\mu \left[ \bar{E}_{\alpha,\varepsilon}(f_*\nu) \right] < (1 + \delta_E)^2 E_{\alpha,\varepsilon}(\nu) + C_E.
\]
Finally, using (6.13) with \( \nu' = \nu \) to estimate \( E_{\alpha,\varepsilon} \) via \( \bar{E}_{\alpha,\varepsilon} \), we get
\[
\frac{1}{(1 + \delta_E)^2} \cdot \frac{1}{(1 + \delta_E)} (E_{\alpha,\varepsilon}(\nu) - C_E) - \frac{C_E}{1 + \delta_E} < \mathbb{E}_\mu \left[ \bar{E}_{\alpha,\varepsilon}(f_*\nu) \right] < (1 + \delta_E)^2 \cdot ((1 + \delta_E)E_{\alpha,\varepsilon}(\nu) + C_E) + C_E;
\]
as \( (1 + \delta_E)^3 < 1 + \delta \), we get the desired
\[
(1 + \delta)(E_{\alpha,\varepsilon}(\nu) - C') < \mathbb{E}_\mu \left[ \bar{E}_{\alpha,\varepsilon}(f_*\nu) \right] < (1 + \delta)E_{\alpha,\varepsilon}(\nu) + C'
\]
for some constant \( C' \). As \( \delta > 0 \) was arbitrary, and taking (6.13) into account again, (4.13) follows. \( \square \)
6.4. Estimating $W(f, \theta_{\alpha,\varepsilon}(\nu), \theta_{\alpha,\varepsilon}(f_\ast \nu))$: Proof of Proposition 4.21

Proof of Proposition 4.21. For given $\alpha, \varepsilon > 0$ and measure $\nu$, consider a (non-probability) measure $m_{\alpha,\varepsilon}(\nu)$ on $M \times M \times M$, given by

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

Denote by $\pi_1, \pi_2$ 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

$$(6.14) \quad \hat{\Theta}_{\alpha,\varepsilon}(\nu) := (\pi_{1,3})_* m_{\alpha,\varepsilon}(\nu) = K_{\alpha,\varepsilon}(x,z) \nu(dx) \nu(dz);$$

the second equality is due to (4.4). Finally, consider the measure

$$(6.15) \quad \Theta'_{\alpha,\varepsilon}(\nu) := (\pi_1)_* m_{\alpha,\varepsilon}(\nu)$$

as well as the normalizations of these measures,

$$(6.16) \quad \hat{\theta}_{\alpha,\varepsilon}(\nu) := \frac{1}{\tilde{E}_{\alpha,\varepsilon}(\nu)} \hat{\Theta}_{\alpha,\varepsilon}(\nu), \quad \theta'_{\alpha,\varepsilon}(\nu) := \frac{1}{\tilde{E}_{\alpha,\varepsilon}(\nu)} \Theta'_{\alpha,\varepsilon}(\nu).$$

For a high-energy measure $\nu$ 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 6.9. For any $\delta_1, \alpha$ there exists $C'$ such that for any $\varepsilon > 0$ and $\nu$ with $\tilde{E}_{\alpha,\varepsilon}(\nu) > C'$ 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 6.8 we have

$$m_{\alpha,\varepsilon}(\nu)(\overline{A}_{r_0}) = \int K_{\alpha,\varepsilon}^{(r_0)}(x,z) d\nu(x) d\nu(z) \leq C_{\alpha}^K(r_0),$$

where the second inequality is due to (6.10). Thus, the non-normalized measure $m_{\alpha,\varepsilon}(\nu)(\overline{A}_{r_0})$ does not exceed a constant $C_{\alpha}^K(r_0)$.

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

\begin{equation}
W(\theta_{\alpha,\varepsilon}(\nu),\theta'_{\alpha,\varepsilon}(\nu)) \leq r_0 \cdot \frac{m_{\alpha,\varepsilon}(\nu)(A_{r_0})}{\tilde{E}_{\alpha,\varepsilon}(\nu)} \quad \text{+ diam}(M) \cdot \frac{m_{\alpha,\varepsilon}(\nu)(\overline{A}_{r_0})}{\tilde{E}_{\alpha,\varepsilon}(\nu)} \leq r_0 + \frac{C^K(r_0)}{\tilde{E}_{\alpha,\varepsilon}(\nu)},
\end{equation}

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

\[\tilde{E}_{\alpha,\varepsilon}(\nu) > \frac{2C^K(r_0)}{\delta_1} =: C'\]

to ensure that the total Wasserstein distance does not exceed \(\delta_1\).

On the other hand, for the measures \(\theta'_{\alpha,\varepsilon}(\nu)\) the analogue of Proposition 4.21 can be established directly, and one can even estimate the total variations distance:

**Lemma 6.10.** 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 \(f \in \text{Diff}^1(M)\) with \(\mathcal{L}(f) < R\), any \(\varepsilon > 0\), and any \(\nu\) such that \(\tilde{E}_{\alpha,\varepsilon}(\nu) > C\) or \(E_{\alpha,\varepsilon}(\nu) > C\) one has

\begin{equation}
\text{TV}(f_*\theta'_{\alpha,\varepsilon}(\nu),\theta'_{\alpha,\varepsilon}(f_*\nu)) < \delta_2,
\end{equation}

and, hence,

\begin{equation}
W(f_*\theta'_{\alpha,\varepsilon}(\nu),\theta'_{\alpha,\varepsilon}(f_*\nu)) < \delta_2 \cdot \text{diam}(M).
\end{equation}

**Proof.** We will first use Lemma 6.8 together with Proposition 6.3 to compare \(K_{\alpha,\varepsilon}(x,z)\) with \(K_{\alpha,\varepsilon}(f(x),f(z))\). Namely, choose \(\alpha_0\) and \(\delta_3\) so small that

\[(1 + \delta_3)^2 \cdot R^{\alpha_0} < 1 + \frac{\delta_2}{10}.
\]

Then from Lemma 6.8 we know that there exists \(C > 0\) such that for any \(x, z, \varepsilon\) we have

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

while from (6.5) in Proposition 6.3 we get

\[\frac{U_{\alpha,\varepsilon}(f(x),f(z))}{U_{\alpha,\varepsilon}(x,z)} \in (\mathcal{L}(f)^{-\alpha}, \mathcal{L}(f)^{\alpha}) \subset (R^{-\alpha_0}, R^{\alpha_0}).
\]

Joining these three estimates together, we obtain

\begin{equation}
K_{\alpha,\varepsilon}(x,z) \approx_{(\delta_3,C')} K_{\alpha,\varepsilon}(f(x),f(z))
\end{equation}

for some explicit constant \(C''\).
Now, applying $f^{-1}_*$ does not change the total variations distance, so instead of (6.18) we can show the equivalent statement

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

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

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

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

where $\pi_1$ is the projection on the first coordinate. To obtain the desired (6.21), we will actually show that for sufficiently high energy $\tilde{E}_{\alpha,\varepsilon}(\nu)$ even before projection one has

$$TV(\hat{\theta}_{\alpha,\varepsilon}(\nu), f^{-1}_*\hat{\theta}_{\alpha,\varepsilon}(f_*\nu)) < \delta_2.$$  

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

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

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

Now, (6.20) implies that once $K_{\alpha,\varepsilon}(x, z) > C_2 =: \frac{20C'\nu}{\delta_2}$, 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 4.15, once the energy $\tilde{E}_{\alpha,\varepsilon}(\nu)$ is sufficiently large, the normalization constants are also close to each other,

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

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 $\left(\frac{1}{(1 + \frac{\delta_2}{6})^2}, (1 + \frac{\delta_2}{6})^2\right)$.  

Finally, the $\hat{\theta}_{\alpha,\varepsilon}(\nu)$-measure of its complement does not exceed $\frac{C_2}{\tilde{E}_{\alpha,\varepsilon}(\nu)}$. Hence, if the energy $\tilde{E}_{\alpha,\varepsilon}(\nu)$ is sufficiently high to make sure that
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 (6.22) indeed does not exceed $\delta_2$.

An application of Lemma 4.20 concludes the proof of the upper bound (6.19) for the Wasserstein distance. □

Lemma 6.9 and Lemma 6.10 together imply Proposition 4.21. Indeed, we have:

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

The first and the third summands can be estimated directly using Lemma 6.9. Indeed, for the first summand, it suffices to note that the application of $f$ increases the Wasserstein distance at most $\mathcal{L}(f) < R$ times; for the last summand, to apply this lemma, we note that an upper bound on $\mathcal{L}(f)$ implies that $f_* \nu$ is of sufficiently high energy provided that $\nu$ is of high energy.

Finally, the second summand is estimated directly by Lemma 6.10. □

6.5. Estimating $W(\theta_{\alpha,\varepsilon}(f_* \nu), \theta_{\alpha,\varepsilon}(\mu \ast \nu))$: Proof of Lemma 5.3.

Proof of Lemma 5.3. We start by working with the non-normalized measures $\Theta_{\alpha,\varepsilon}(f_* \nu), \Theta_{\alpha,\varepsilon}(\mu \ast \nu)$. To simplify the notation, denote

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

Therefore, we have

\[
\Theta_{\alpha,\varepsilon}[f_* \nu] = (g_f(y))^2 d\text{Leb}(y) \quad \text{and} \quad \Theta_{\alpha,\varepsilon}[\mu \ast \nu] = (\overline{g}(y))^2 d\text{Leb}(y).
\]

From (5.2) we have

\[
\tilde{E}_{\alpha,\varepsilon}(\nu) \leq \frac{1}{1 - \delta} \tilde{E}_{\alpha,\varepsilon}(\mu \ast \nu).
\]

Joining with (5.8), we have

\[
(6.23) \quad \mathbb{E}_f \left[ \int_M |g_f(y) - \overline{g}(y)|^2 d\text{Leb}(y) \right] \leq \frac{2\delta}{1 - \delta} \tilde{E}_{\alpha,\varepsilon}(\mu \ast \nu).
\]
By Markov inequality, with the probability at least $1 - \frac{2\sqrt{\delta}}{\delta}$, one has

$$
(6.24) \quad \int_M |g_f(y) - \bar{g}(y)|^2 d\text{Leb}(y) \leq \sqrt{\delta} \tilde{E}_{\alpha,\varepsilon}(\mu * \nu)
$$

$$
= \sqrt{\delta} \Theta_{\alpha,\varepsilon}[\mu * \nu](M) = \sqrt{\delta} \int_M \bar{g}^2(y) d\text{Leb}(y).
$$

Now, for any $f$ for which (6.24) holds, consider the set

$$
X_f := \{y \in M : |g_f(y) - \bar{g}(y)| \leq \sqrt{\delta} \cdot \bar{g}(y)\}.
$$

Again from Markov inequality type argument, one has

$$
(6.25) \quad \Theta_{\alpha,\varepsilon}[\mu * \nu](X_f) \geq \left(1 - \frac{\sqrt{\delta}}{\delta}\right) \Theta_{\alpha,\varepsilon}[\mu * \nu](M).
$$

Indeed, on the complement $M \setminus X_f$ one has $|g_f(y) - \bar{g}(y)|^2 > \sqrt{\delta} \cdot \bar{g}^2(y)$; integrating, one gets

$$
\int_{M \setminus X_f} |g_f(y) - \bar{g}(y)|^2 d\text{Leb}(y) > \sqrt{\delta} \cdot \Theta_{\alpha,\varepsilon}[\mu * \nu](M \setminus X_f).
$$

Thus, if (6.25) did not hold, it would imply

$$
\Theta_{\alpha,\varepsilon}[\mu * \nu](M \setminus X_f) > \sqrt{\delta} \cdot \Theta_{\alpha,\varepsilon}[\mu * \nu](M),
$$

hence providing a contradiction with (6.24).

Now, (6.25) implies that the non-normalized measures $\Theta_{\alpha,\varepsilon}[f * \nu]$ and $\Theta_{\alpha,\varepsilon}[\mu * \nu]$ share a common part

$$
\Xi_f := (1 - \sqrt{\delta})^2 \bar{g}^2(y) \cdot 1_{X_f}(y) d\text{Leb}(y)
$$

of measure at least

$$
(6.26) \quad \Xi_f(M) \geq (1 - \sqrt{\delta})^2 (1 - \sqrt{\delta}) \cdot \Theta_{\alpha,\varepsilon}[\mu * \nu](M).
$$

Finally, note that the normalization constants are also close with high probability, namely, for $f$ such that the inequality (6.24) holds. Indeed, the inequality (6.24) can be rewritten as

$$
\|g_f - \bar{g}\|_{L^2(M)}^2 \leq \sqrt{\delta} \cdot \|\bar{g}\|_{L^2(M)}^2;
$$

hence, for any $f$ for which it holds, the triangle inequality implies

$$
(6.27) \quad \Theta_{\alpha,\varepsilon}[f * \nu](M) = \|g_f\|_{L^2(M)} \leq (1 + \sqrt{\delta})^2 \Theta_{\alpha,\varepsilon}[\mu * \nu](M).
$$

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 (6.26) and (6.27), we see that this part has measure at least
\[
\frac{1}{C_f} \Xi_f(M) \geq \frac{(1 - \sqrt{\delta})^2(1 - \sqrt{\delta})}{(1 + \sqrt{\delta})^2} \geq 1 - 10 \sqrt{\delta}.
\]
Hence, for any \(f\) satisfying (6.24) the total variation distance between the normalized measures does not exceed
\[
TV(\theta_{\alpha,\varepsilon}(f_*\nu), \theta_{\alpha,\varepsilon}(\mu * \nu)) \leq 10 \sqrt{\delta}.
\]
As (6.24) holds with the probability at least \(1 - \frac{2\sqrt{\delta}}{1 - \delta} > 1 - 4\sqrt{\delta}\), this establishes conclusion (5.10). Finally, this immediately implies conclusion (5.11) due to Lemma 4.20.

\[\Box\]

**Appendix A. Example**

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 [G90], as we discussed in Section 1.1 above. The condition on finiteness of moments (used both in our main result and in [G90]) 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 [Fur1]. 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 A.1.** There exists a probability distribution \(\mu\) on \(\text{SL}(2, \mathbb{R})\) such that

1) The smallest subgroup generated by \(\text{supp} \mu\) 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
\[
A_n = R_{\frac{1}{2^n}} \cdot \begin{pmatrix} 2n^2 & 0 \\ 0 & 2^{-n^2} \end{pmatrix} \cdot R_{-\frac{1}{2^n}},
\]
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 \( \mu \) on \( \text{SL}(2, \mathbb{R}) \) 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 A.1 are satisfied. Let us show that 4) also holds. Let us denote by \( f_n : \mathbb{P}^1 \to \mathbb{P}^1 \) a projectivization of \( A_n : \mathbb{R}^2 \to \mathbb{R}^2 \). Slightly abusing the notation, let us use the same symbol \( \mu \) 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 \( \nu \) be a stationary probability measure on \( \mathbb{P}^1 \). Due to the explicit form of the matrices \( A_n \) we have \( \text{supp} \nu \subseteq [0, 2/5] \) (here we identify \( \mathbb{P}^1 \) with \( [0, 1]/\{0\sim1\} \)). If \( x \in [0, 2/5] \subset \mathbb{P}^1 \) and \( v \in \mathbb{R}^2 \) is the corresponding unit vector, then

\[
|A_n v| \geq 2^{n^2} \cos \frac{2\pi}{5} > \frac{2^{n^2}}{10},
\]

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

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

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

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

hence the measure \( \nu \) cannot be Hölder continuous. \( \square \)

**Acknowledgments**

The authors are grateful to B. Barany, S. Cantat, B. Deroin, Y. Guivarc’h, F. Ledrappier, and B. Solomyak for fruitful discussions and for providing helpful references.

**References**

[A] L. Arnold, Random dynamical systems. *Dynamical systems (Montecatini Terme, 1994)*, pp. 1–43, Lecture Notes in Math., **1609**, Springer, Berlin, 1995.

[AG] R. Aoun, Y. Guivarc’h, Random matrix products when the top Lyapunov exponent is simple, *J. Eur. Math. Soc. (JEMS)* **22** (2020), pp. 2135–2182.

[BK] B. Barany, A. Käenmäki, Ledrappier-Young formula and exact dimensionality of self-affine measures, *Adv. Math.* **318** (2017), pp. 88–129.
HÖLDER REGULARITY OF STATIONARY MEASURES

[BPS] B. Barany, M. Pollicott, K. Simon, Stationary measures for projective transformations: The Blackwell and Furstenberg measures, *J. Stat. Phys.* **148** (2012), pp. 393–421.

[BQ18] Y. Benoist, J. Quint, On the regularity of stationary measures, *Israel J. Math.* **226** (2018), pp. 1–14.

[BQ] Y. Benoist, J. Quint, Random walks on reductive groups. *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics* **62**, Springer, Cham, 2016, xi+323 pp.

[BQ16] Y. Benoist, J. Quint, Central Limit Theorem for Linear Groups, *The Annals of Probability*, **44** (2016), pp. 1308–1340.

[BL] P. Bougerol and J. Lacroix, *Products of Random Matrices with Applications to Schrödinger Operators*, Birkhäuser, Boston, 1985.

[B] J. Bourgain, Finitely supported measures on $SL_2(ℝ)$ which are absolutely continuous at infinity, *Geometric aspects of functional analysis*, pp. 133–141, Lecture Notes in Math., **2050**, Springer, Heidelberg, 2012.

[BFLM] J. Bourgain, A. Furman, E. Lindenstrauss, S. Mozes, Stationary measures and equidistribution for orbits of nonabelian semigroups on the torus, *J. Amer. Math. Soc.* **24** (2011), pp. 231–280.

[BH] A. Brown, F. Hertz, Measure rigidity for random dynamics on surfaces and related skew products, *J. Amer. Math. Soc.* **30** (2017), pp. 1055–1132.

[CD] S. Cantat, R. Dujardin, Random dynamics on real and complex projective surfaces, preprint, arXiv:2006.04394.

[DKN] B. Deroin, V. Kleptsyn, A. Navas, On the question of ergodicity for minimal group actions on the circle, *Mosc. Math. J.* **9** (2009), pp. 263–303.

[DKW] T. Dinh, L. Kaufmann, H. Wu, Products of random matrices: a dynamical point of view, *Pure and Applied Mathematics Quarterly* **17** (2021), pp. 933–969.

[E] P. Erdős, On a family of symmetric Bernoulli convolutions, *Amer. J. Math.* **61** (1939), pp. 974–976.

[FJ] K. Falconer, X. Jin, Exact dimensionality and projections of random self-similar measures and sets, *J. Lond. Math. Soc.* **90** (2014), pp. 388–412.

[F] D. Feng, Dimension of invariant measures for affine iterated function systems, preprint, arXiv:1901.01691.

[FF] D. Feng, Zh. Feng, Estimates on the dimension of self-similar measures with overlaps, *J. Lond. Math. Soc.* **105** (2022), pp. 2104–2135.

[FL] D. Feng, K. Lau, Multifractal formalism for self-similar measures with weak separation condition, *J. Math. Pures Appl.* **9** (2009), pp. 407–428.

[Fu1] A. Furman, Random walks on groups and random transformations, *Handbook of dynamical systems*, vol. 1A, 931–1014, North-Holland, Amsterdam, 2002.

[Fu2] A. Furman, What is . . . a stationary measure? *Notices Amer. Math. Soc.* **58** (2011), no. 9, pp. 1276–1277.

[Fur1] H. Furstenberg, Noncommuting random products, *Trans. Amer. Math. Soc.*, **108** (1963), pp. 377–428.

[Fur2] H. Furstenberg, Random walks and discrete subgroups of Lie groups, *1971 Advances in Probability and Related Topics*, Vol. 1, pp. 1–63, Dekker, New York.

[Fur3] H. Furstenberg, Boundary theory and stochastic processes on homogeneous spaces, *Harmonic analysis on homogeneous spaces* (**Proc. Sympos. Pure Math.**, **29**, AMS, Providence, 1977, pp. 1–49).
40 ANTON GORODETSKI, VICTOR KLEPTSYN, AND GRIGORII MONAKOV

Vol. XXVI, Williams Coll., Williamstown, Mass., 1972), 193–229. Amer. Math. Soc., Providence, R.I., 1973.

[FurK] H. Furstenberg, H. Kesten, Products of random matrices, Ann. Math. Statist. 31 (1960), pp. 457–469.

[FurKif] H. Furstenberg, Y. Kifer, Random matrix products and measures on projective spaces, Israel Journal of Mathematics 46 (1983), pp. 12–32.

[GMT] V. Gadre, J. Maher, G. Tiozzo, Word length statistics for Teichmüller geodesics and singularity you harmonic measures, Comment. Math. Helv. 92 (2017), pp. 1–30.

[Go] I. Goldsheid, Exponential growth of products of non-stationary Markov-dependent matrices, Int. Math. Res. Not. (IMRN), (2022), pp. 6310–6346.

[GK] A. Gorodetski, V. Kleptsyn, Non-stationary version of Furstenberg Theorem on random matrix products, work in progress.

[G90] Y. Guivarc’h, Produits de matrices aléatoires et applications aux propriétés géométriques des sous-groupes du groupe lineaire, Ergodic Theory Dynam. Systems 10 (1990), pp. 483–512.

[GR] Y. Guivarc’h, A. Raugi, Products of random matrices: convergence theorems, Random matrices and their applications (Brunswick, Maine, 1984), 31–54, Contemp. Math., 50, Amer. Math. Soc., Providence, RI, 1986.

[GL] Y. Guivarc’h, Y. Le Jan, Y., Asymptotic winding of the geodesic flow on modular surfaces and continued fractions, Ann. Sci. Ecole Norm. Sup. 26 (1993), pp. 23–50.

[Ho] M. Hochman, On self-similar sets with overlaps and inverse theorems for entropy, Ann. of Math. (2) 180 (2014), pp. 773–822.

[HR] M. Hochman, A. Rapaport, Hausdorff dimension of planar self-affine sets and measures with overlaps, J. Eur. Math. Soc. (JEMS), 24 (2022), pp. 2361–2441.

[HS] M. Hochman, B. Solomyak, On the dimension of Furstenberg measure for $SL_2(\mathbb{R})$ random matrix products, Invent. Math. 210 (2017), pp. 815–875.

[Hu] J. Hutchinson, Fractals and self-similarity, Indiana Univ. Math. J. 30 (1981), pp. 713–747.

[KL] V. Kaimanovich, V. Le Prince, Matrix random products with singular harmonic measure, Geom. Ded. 150 (2011), pp. 257–279.

[Kif1] Yu. Kifer, Random dynamics and its applications, Proceedings of the International Congress of Mathematicians, Vol. II (Berlin, 1998). Doc. Math. 1998, Extra Vol. II, pp. 809–818.

[Kif2] Yu. Kifer, Ergodic theory of random transformations, Progress in Probability and Statistics, 10, Birkhäuser Boston, Inc., Boston, MA, 1986. x+210 pp.

[KPV] V. Kleptsyn, M. Pollicott, P. Vytnova, Uniform lower bounds on the dimension of Bernoulli convolutions, Adv. Math. 395 (2022), Paper No. 108090.

[L] F. Ledrappier, Quelques proprietes des exposants caracteristiques. Ecole d’ete de probabilites de Saint-Flour, XII—1982 (Lecture Notes in Mathematics, 1097), Springer, Berlin, 1984, pp. 305–396.

[LL] F. Ledrappier, P. Lessa, Exact dimension of Furstenberg measures, preprint, ArXiv 2105.11712

[LQ] P. Liu, M. Qian, Smooth ergodic theory of random dynamical systems, Lecture Notes in Mathematics, 1606, Springer-Verlag, Berlin, 1995, xii+221 pp.

[M] D. Malicet, Random Walks on Homeo($S^1$), Commun. Math. Phys. 356 (2017), pp. 1083–1116.
[PSS] Y. Peres, W. Schlag, B. Solomyak, Sixty years of Bernoulli convolutions, Fractal geometry and stochastics, II (Greifswald/Koserow, 1998), pp. 39–65, Progr. Probab., 46, Birkhäuser, Basel, 2000.

[R] A. Rapaport, Exact dimensionality and Ledrappier-Young formula for the Furstenberg measure, Trans. Amer. Math. Soc. 374 (2021), pp. 5225–5268.

[Shm] P. Shmerkin, On the exceptional set for absolute continuity of Bernoulli convolutions, Geom. Funct. Anal. 24 (2014), pp. 946–958.

[Sol] B. Solomyak, On the random series \( \sum \pm \lambda^n \) (an Erdős problem), Ann. of Math. (2) 142 (1995), pp. 611–625.

[Sz] T. Szarek, Invariant measures for iterated function systems, Ann. Polon. Math. 75 (2000), pp. 87–98.

[Va1] P. Varju, Recent progress on Bernoulli convolutions, European Congress of Mathematics, pp. 847–867, Eur. Math. Soc., Zürich, 2018

[Va2] P. Varju, Self-similar sets and measures on the line, preprint, arXiv:2109.10629

[Va3] P. Varju, On the dimension of Bernoulli convolutions for all transcendental parameters, Ann. of Math. 189 (2019), pp. 1001–1011.

[V] C. Villani, Optimal Transport. Old and New, Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 338, Springer-Verlag, Berlin, 2009, xxii + 973 pp.

Anton Gorodetski, Department of Mathematics, University of California, Irvine, CA 92697, USA
Email address: asgor@uci.edu

Victor Kleptsyn, Univ Rennes, CNRS, IRMAR - UMR 6625, F-35000 Rennes, France
Email address: victor.kleptsyn@univ-rennes1.fr

Grigorii Monakov, Department of Mathematics, University of California, Irvine, CA 92697, USA
Email address: gmonakov@uci.edu