Received: 5 May 2021 | Revised: 7 April 2022 | Accepted: 19 April 2022

DOI: 10.1112/jlms.12649

RESEARCH ARTICLE

Journal of the London Mathematical Society

Random-like properties of chaotic forcing

Paolo Giulietti[1] | **Stefano Marmi**[2] | **Matteo Tanzi**[3]

[1]Department of Mathematics, Università di Pisa, Pisa, Italy

[2]Scuola Normale Superiore, Pisa, Italy

[3]Courant Institute of Mathematical Sciences, New York University, New York, USA

Correspondence
Paolo Giulietti, Department of Mathematics, Università di Pisa, Largo Bruno Pontecorvo, 5, Pisa I-56127, Italy.
Email: paolo.giulietti@unipi.it

Funding information
INDAM - GNFM; UniCredit Bank R&D; PRIN, Grant/Award Number: 2017S35EHN; Dynamics and Information Theory Institute; Marie Skłodowska-Curie Actions, Grant/Award Number: 843880

Abstract
We prove that skew systems with a sufficiently expanding base have approximate exponential decay of correlations, meaning that the exponential rate is observed modulo an error. The fiber maps are only assumed to be Lipschitz regular and to depend on the base in a way that guarantees diffusive behaviour on the vertical component. The assumptions do not imply an hyperbolic picture and one cannot rely on the spectral properties of the transfer operators involved. The approximate nature of the result is the inevitable price one pays for having so mild assumptions on the dynamics on the vertical component. However, the error in the approximation goes to zero when the expansion of the base tends to infinity. The result can be applied beyond the original setup when combined with acceleration or conjugation arguments, as our examples show.

MSC 2020
37A25 (primary), 37A30, 37H05 (secondary)

1 | INTRODUCTION

One of the main questions of modern dynamical systems theory is: *to which extent a deterministic chaotic system resembles a random process?* This question has been addressed in various contexts from different point of views (see [53] for a review). Here we study it in relation to forcing, and in particular we investigate the similarities between random and (sufficiently chaotic) deterministic forcing focusing on the statistical properties of the forced system.

A forced system is a system whose intrinsic dynamics is affected by an external influence typically coming from the interaction with another system or the surrounding environment. The forcing can be modelled to be random, for example, obtained by adding to the dynamics a noise term independent in time, or deterministic, that is, dependent on a variable that evolves in time following a deterministic law[†].

In the random case, classical results from the theory of Markov chains show that if there is enough diffusion, for example, if the forcing adds smooth unbounded noise to the dynamics, then the forced system has a stationary measure that describes its asymptotic statistical behaviour, and exhibits memory loss and annealed exponential decay of correlations (among others [4, 21]). In contrast, if the forcing is deterministic, it is well known that even just to prove existence of a physically relevant invariant measure one needs to impose strong assumptions both on the intrinsic dynamic and on the forcing, often leading to some degree of hyperbolicity of the system and/or a good spectral picture of the operators involved (see literature below).

In this paper we prove that, if the forcing has a "diffusive effect" and is generated by a uniformly expanding map with high expansion, then the deterministic system has an *approximate* stationary measure and exhibits *approximate* decay of correlations. We postpone rigorous definitions to later sections. Loosely speaking, an approximate stationary measure describes the asymptotic statistical properties of the system modulo a controlled error, and by an approximate exponential decay of correlations we mean that measurements of observables along orbits exhibit exponential decay of correlations also modulo an error. Most importantly, these errors go to zero when the expansion of the map generating the forcing goes to infinity. In other words we could say that, when the expansion of the map generating the forcing goes to infinity, the deterministic forcing becomes indistinguishable from random forcing with respect to the statistical properties we analyze.

It is important to remark that our requirements do not ensure global hyperbolic properties or a good spectral picture, and even the existence of a physically relevant invariant measure cannot be deduced from the assumptions. The price that we pay is the approximate nature of the result. Its relevance, however, is clear when having an eye to applications; here decorrelation estimates come from observations of real-world systems and are intrinsically affected by a measurement error: if this error is larger than the approximation error in the decorrelation estimate, exact and approximate decay of correlations are indistinguishable.

Our approach is quite flexible and we expect it to be adaptable to a variety of situations beyond the current working assumptions, for example in situations with lower regularity, or in combination with various conjugations arguments (see Section 5 for some generalizations).

1.1 | Literature review

In mathematical terms, a forced system in discrete time can be described by a *skew-product* transformation which is a map $F : \Omega \times X \to \Omega \times X$ such that

$$F(\omega, x) = (g(\omega), f(\omega, x)) \tag{1.1}$$

where $g : \Omega \to \Omega$ and $f : \Omega \times X \to X$. The set Ω is called the *base* of the skew-product, while X is referred to as the *vertical fiber*. The main characteristic of a skew-product is that the evolution on the base Ω does not depend on the vertical fiber X.

The literature on skew-products is vast to the extent that there are entire research trends studying particular aspects of these systems (e.g., iterated function systems, random dynamical systems,

[†] For precise definitions and a comparison between deterministic and random forcing see Section A.3 in Appendix.

smoothness of invariant graphs over skew-products, etc.). Here we focus on those works dealing with statistical properties of skew products that have a “deterministic” base, such as [6–8, 11, 22–26, 28, 29, 34, 38, 41, 44, 47, 51] and references therein. These works usually only require g to be a measure preserving ergodic transformation or, at most, to exhibit some uniform hyperbolicity. However, they restrict the fiber map f to one of some particular classes to ensure contraction or hyperbolic properties (exact or averaged) of the vertical fiber. Skew-products with nonuniform hyperbolicity can still be studied but in a more qualitative sense [5, 12]. In contrast, our results make only mild regularity assumptions on f, but require that g is uniformly expanding with large minimal expansion.

As a consequence of our requirements, the map F is likely to have a dominated splitting of the tangent space and be *partially hyperbolic* (see, e.g., [32, 40]) with an expanding direction roughly aligned with the base dominating the other invariant directions. To put our work under this perspective, let us remind that available results on existence of physical measures and decay of correlations for partially hyperbolic systems often assume low dimensional geometry either of the phase space or of some invariant directions, and/or nonvanishing Lyapunov exponents [1, 2, 15, 18, 19, 46, 48] which, in general, are not granted in our setup. More recent results give sufficient conditions for partially hyperbolic systems to have exponential decay of correlations by turning qualitative topological conditions such as accessibility [10], into quantitative properties of the operators involved [13, 27]. The systems we consider do not fit in these results due to lack of smoothness, but it is even unclear if these results can be applied to those systems in our setup that have the required regularity.

As the base map is much more chaotic than the vertical fibers, our setup is reminiscent of fast-slow systems (see [13, 14, 17, 33, 35] among many others). However, the dynamic of our skew-products does not present separation of time-scales since at each time step it can produce displacements of the same order both in the base system and in the vertical fibers.

1.2 | Organization of the paper

In Section 2 we present the setting, the results, some examples, and a sketch of the proof. In Section 3 we prove our result in the simpler situation where the map in the base has no distortion and the phase space is 2D. In Section 4 we prove our main theorem in full generality. In Section 5 we discuss some generalizations and limits of our approach. In the appendices we gather some background material and results on Markov chains (in Appendix A), disintegration of measures (in Appendix B), and some computations involving the Kantorovich–Wasserstein distance that are used throughout the proofs (in Appendix C).

2 | SETTING AND RESULTS

2.1 | Setting

Let’s consider a map F as in (1.1) where we set $\Omega = \mathbb{T}^{m_1}$ and $X = \mathbb{T}^{m_2}$, here $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ is the 1D torus and m_1, m_2 two positive integers. In the following we will denote by $|p_1 - p_2|$ the distance between $p_1, p_2 \in \mathbb{T}^N$ regardless of the specific $N \in \mathbb{N}$. For $I \subseteq \mathbb{T}^m$, we denote by $\mathrm{Op}(I)$ its open part.

2.1.1 | The base map g

Consider $g : \mathbb{T}^{m_1} \to \mathbb{T}^{m_1}$ a C^2 local diffeomorphism. In particular, there is $d \in \mathbb{N}$ and $\mathcal{I} = \{I_i\}_{i=1}^d$ a partition of $\mathbb{T}^{m_1}$ such that: $\mathrm{Op}(I_i) = I_i \mod 0$, $\{g_i := g|_{I_i}\}_{i=1}^d$ with $g_i : I_i \to \mathbb{T}^{m_1}$ are invertible branches of g, and $g_i|_{\mathrm{Op}(I_i)}$ is C^2. Call $\{h_i := g_i^{-1}\}_{i=1}^d$ the corresponding inverses.

We assume that g satisfies the following assumptions:

$$\exists \sigma > 1 \text{ s.t. } \|\mathrm{D}g_\omega v\| \geqslant \sigma \|v\| \qquad \forall \omega \in \mathbb{T}^{m_1},\ v \in \mathbb{R}^{m_1}, \tag{H0.1}$$

where $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^{m_1}$, and

$$\exists D \geqslant 0 \text{ s.t. } \frac{|\mathrm{D}g_{h_i(\omega_1)}|}{|\mathrm{D}g_{h_i(\omega_2)}|} \leqslant e^{D|\omega_1 - \omega_2|} \qquad \forall \omega_1, \omega_2 \in \mathbb{T}^{m_1} \text{ and } \forall i \tag{H0.2}$$

where $|\mathrm{D}g_{h_i(\omega_1)}|$ denotes the determinant of $\mathrm{D}g_{h_i(\omega_1)}$. Condition (H0.1) states that the differential of g expands vectors in tangent space of a factor at least $\sigma > 1$, while (H0.2) imposes a uniform bound on the distortion. It is well known that g has a unique absolutely continuous invariant probability (a.c.i.p.) measure (see [9],[49] and references therein). We call ν_g this measure and $\rho_g := \frac{d\nu_g}{d\,\mathrm{Leb}_{\mathbb{T}^{m_1}}}$ its density, where $\mathrm{Leb}_{\mathbb{T}^{m_1}}$ is the Lebesgue measure on $\mathbb{T}^{m_1}$.

2.1.2 | The vertical fiber maps f

We assume $f : \mathbb{T}^{m_1} \times \mathbb{T}^{m_2} \to \mathbb{T}^{m_2}$ to be at least Lipschitz, and denote by $L \geqslant 0$ the Lipschitz constant, namely

$$L := \inf_{(\omega_1,x_1)\neq(\omega_2,x_2)} \frac{|f(\omega_1, x_1) - f(\omega_2, x_2)|}{|(\omega_1, x_1) - (\omega_2, x_2)|}. \tag{H0.3}$$

Let $\{f_\omega\}_{\omega \in \mathbb{T}^{m_1}}$ be the collection of maps $f_\omega : \mathbb{T}^{m_2} \to \mathbb{T}^{m_2}$, that is, $f_\omega(\cdot) := f(\omega, \cdot)$. We write $f(\cdot, x)$ for the maps $f(\cdot, x) : \mathbb{T}^{m_1} \to \mathbb{T}^{m_2}$ obtained by fixing $x \in \mathbb{T}^{m_2}$ and letting $\omega \in \mathbb{T}^{m_1}$ vary. We let $\pi_1 : \mathbb{T}^{m_1} \times \mathbb{T}^{m_2} \to \mathbb{T}^{m_1}$ be the projection onto the *horizontal* $\mathbb{T}^{m_1}$-coordinate and, given a measure μ on $\mathbb{T}^{m_1} \times \mathbb{T}^{m_2}$, we refer to $\pi_{1*}\mu$ as the *horizontal marginal* of μ. We also denote by $\pi_2 : \mathbb{T}^{m_1} \times \mathbb{T}^{m_2} \to \mathbb{T}^{m_2}$ the projection onto the *vertical* $\mathbb{T}^{m_2}$-coordinate and refer to $\Pi\mu := \pi_{2*}\mu$ as the *vertical marginal* of the measure μ.

2.1.3 | $\mathcal{P}$, the random counterpart of F

In the following, $\mathcal{M}_1(Y)$ denotes the space of Borel probability measures on the compact metric space Y.

For $\mu \in \mathcal{M}_1(\mathbb{T}^{m_2})$, define the push-forward $f_{\omega*}\mu(A) = \mu(f_\omega^{-1}(A))$ for any measurable $A \subseteq \mathbb{T}^{m_2}$, and define the operator $\mathcal{P} : \mathcal{M}_1(\mathbb{T}^{m_2}) \to \mathcal{M}_1(\mathbb{T}^{m_2})$

$$\mathcal{P}\mu := \int_{\mathbb{T}^{m_1}} d\nu_g(\omega) f_{\omega*}\mu = \int_{\mathbb{T}^{m_1}} d\omega \rho_g(\omega) f_{\omega*}\mu. \tag{2.1}$$

Notice that $\mathcal{P}$ is the generator for a *discrete time stationary Markov process* with transition kernel

$$P(x, A) := \int_{\mathbb{T}^{m_1}} d\omega \delta_{f_\omega(x)}(A)\rho_g(\omega).$$

where $\delta_{f_\omega(x)}$ denotes the Dirac mass at $f_\omega(x)$. These operators are well studied in the literature and sufficient conditions under which $\mathcal{P}$ has a spectral gap in various functional spaces are known (see, e.g., [30, 42, 45] and Appendix A).

It is important to notice that if at each time step one was to apply a map $\{f_\omega\}_{\omega\in\mathbb{T}^{m_1}}$ sampled *independently* with respect to ν_g, then the operator $\mathcal{P}$ would describe the evolution of the vertical marginal. In other terms, one can think of the Markov chain generated by $\mathcal{P}$ as the "random counterpart" of the deterministic evolution given by F which instead selects the map f_ω at each time-step according to the deterministic process ω, $g(\omega)$, $g^2(\omega)$, ... generated by g.

2.2 | Main assumption

Assumption H below requires that the Markov chain generated by $\mathcal{P}$ is geometrically ergodic with respect to the Total Variation (TV) distance d_{TV} (see Appendix A for definitions).

Assumption H. There are $C > 0$ and $\lambda \in (0, 1)$ such that

$$d_{TV}(\mathcal{P}^n\mu, \mathcal{P}^n\nu) \leqslant C\lambda^n d_{TV}(\mu, \nu),$$

for all μ, $\nu \in \mathcal{M}_1(\mathbb{T}^{m_2})$.

By a standard Krylov–Bogolyubov argument, it follows that there is a unique $\eta_0 \in \mathcal{M}_1(\mathbb{T}^{m_2})$ invariant under $\mathcal{P}$, that is, such that $\mathcal{P}\eta_0 = \eta_0$, which is called a *stationary measure* for the Markov chain generated by $\mathcal{P}$. Also notice that Assumption H is a condition on $\mathcal{P}$, and therefore it depends on $f : \mathbb{T}^{m_1} \times \mathbb{T}^{m_2} \to \mathbb{T}^{m_2}$ and ν_g only.

2.3 | Main results

For a randomly forced system whose evolution is given by geometrically ergodic Markov chain, any initial measure evolves exponentially fast towards the stationary measure of the Markov chain. Our first result shows that something similar happens also for the deterministic skew-product F, modulo a controlled approximation error. Loosely speaking, under Assumption H, we show that there is a probability measure $\overline{\eta} \in \mathcal{M}_1(\mathbb{T}^{m_2})$ such that, under iteration of the push-forward F_*, the vertical marginal of any sufficiently regular initial measure enters a small neighbourhood of $\overline{\eta}$ (w.r.t. d_W) at exponential speed. The size of the neighbourhood can be made arbitrarily small increasing the minimal expansion of g. To rigorously states the result we need two definitions given below.

Recall that the Kantorovich–Wasserstein distance between two probability measures μ_1, μ_2 on $\mathbb{T}^{m_2}$ is defined as

$$d_W(\mu_1, \mu_2) := \inf_{\gamma\in C(\mu_1,\mu_2)} \int_{\mathbb{T}^{m_2}\times\mathbb{T}^{m_2}} |x - y| d\gamma(x, y) \tag{2.2}$$

where $C(\mu_1, \mu_2)$ is the set of couplings of μ_1 and μ_2, that is, the set of all probability measures on $\mathbb{T}^{m_2} \times \mathbb{T}^{m_2}$ with marginals μ_1 and μ_2 respectively on the first and second factor.

Definition 2.1. Given $\mu \in \mathcal{M}_1(\mathbb{T}^{m_1} \times \mathbb{T}^{m_2})$, we say that μ has *Lipschitz disintegration along vertical fibres*, or simply *Lipschitz disintegration*, if there is a disintegration $\{\mu_\omega\}_{\omega \in \mathbb{T}^{m_1}}$ of μ, with respect to the measurable partition $\{\{\omega\} \times \mathbb{T}^{m_2}\}_{\omega \in \mathbb{T}^{m_1}}$, such that the map $\omega \mapsto \mu_\omega$ from $(\mathbb{T}^{m_1}, |\cdot|)$ to $(\mathcal{M}_1(\mathbb{T}^{m_2}), d_W)$ is Lipschitz. Let

$$\mathrm{Lip}(\mu) := \inf_{\omega_1 \neq \omega_2} \frac{d_W(\mu_{\omega_1}, \mu_{\omega_2})}{|\omega_1 - \omega_2|}$$

the Lispchitz constant of $\omega \mapsto \mu_\omega$.[†]

The class of probability measures with Lipschitz disintegration plays a central role in this paper. In Appendix B we gather statements about disintegration of measures that will be used throughout. We are now ready to state our first result.

Theorem 2.1. *Let F satisfy assumptions (H0.1)-(H0.3) and Assumption (H) with datum m_1, $m_2 \in \mathbb{N}$, D, L, $C > 0$, $\sigma > 1$, $\lambda \in (0, 1)$. Then there is a probability measure $\overline{\eta} \in \mathcal{M}_1(\mathbb{T}^{m_2})$ and constants $\widetilde{\lambda} \in (0, 1)$, $\widetilde{C} > 0$, $\ell_0 > 0$ satisfying: for every $\varepsilon > 0$, there is $\sigma_0 > 0$ (depending on ε and all the datum but σ) such that if $\sigma > \sigma_0$,*

$$d_W\left(\Pi F_*^n \mu, \overline{\eta}\right) \leqslant \widetilde{C}\widetilde{\lambda}^n + \varepsilon$$

for all $n \in \mathbb{N}$ and any $\mu \in \mathcal{M}_1(\mathbb{T}^{m_1} \times \mathbb{T}^{m_2})$ with horizontal marginal equal to $\mathrm{Leb}_{\mathbb{T}^{m_1}}$, Lipschitz disintegration, and $\mathrm{Lip}(\mu) \leqslant \ell_0$.

The measure $\overline{\eta}$ plays a role analogous to that of the stationary measure of an ergodic Markov chain, however is neither the vertical marginal of some invariant measure for F_* nor is an exact limit for $\Pi F_*^n \mu$. For these reasons, we call it an *approximate stationary measure*. In the case with no distortion, for example, $g(\omega) = \sigma\omega \bmod 1$ with $\sigma \geqslant 2$, $\overline{\eta}$ can be taken equal to η_0, the stationary measure of the Markov chain $\mathcal{P}$, and $\widetilde{C}$ and $\widetilde{\lambda}$ equal C and λ from Assumption (H) (see Section 3). As shown in Section 4.3, when g has nonzero distortion, $\overline{\eta}$ can be different from η_0 and is related to the fixed point of another operator, called $\boldsymbol{\mathcal{L}}$, introduced in Section 4.2.

Remark 2.1. In Theorem 2.1, the assumption that μ has horizontal marginal equal to the Lebesgue measure can be relaxed and one can consider μ with horizontal marginal absolutely continuous with respect to Lebesgue and with sufficiently regular density (see Proposition 4.5). For what concerns ℓ_0, an explicit expression is given in (4.3).

From Theorem 2.1 one can deduce information on the statistical properties of the dynamics defined by F. When describing the statistical properties of a skew-product such as F, we adopt the *annealed* point of view, that is, we assume to have access to observations of measurable functions $\varphi : \mathbb{T}^{m_2} \to \mathbb{R}$ along the orbits of the system. Picking as reference measure on $\mathbb{T}^{m_1} \times \mathbb{T}^{m_2}$ the

[†] It is easy to check that if μ admits a Lipschitz disintegration, this is the only Lipschitz disintegration admitted by μ, therefore $\mathrm{Lip}(\mu)$ is well defined

Lebesgue measure, $\mathrm{Leb}_{\mathbb{T}^{m_1}\times\mathbb{T}^{m_2}}$, gives rise to the sequence of dependent random variables

$$\{\varphi\circ\pi_2\circ F^n\}_{n=1}^{+\infty}$$

on $(\mathbb{T}^{m_1}\times\mathbb{T}^{m_2}, \mathrm{Leb}_{\mathbb{T}^{m_1}\times\mathbb{T}^{m_2}})$.

For $\varphi,\psi:\mathbb{T}^{m_2}\to\mathbb{R}$ in suitable functional spaces, we ask if there are constants $A\in\mathbb{R}$ and $\widetilde{\lambda}\in(0,1)$ such that

$$\left|\int_{\mathbb{T}^{m_1}\times\mathbb{T}^{m_2}}\varphi(\pi_2F^n(\omega,x))\psi(x)\ d\omega dx - A\right| = \mathcal{O}(\widetilde{\lambda}^n) \tag{2.3}$$

When (2.3) is satisfied, the system is said to have exponential annealed decay of correlations. The term *annealed* refers to the fact that the observables φ,ψ depend on the vertical $\mathbb{T}^{m_2}$-coordinate only, and therefore the correlations are averaged with respect to the horizontal $\mathbb{T}^{m_1}$-coordinate.

As already argued in the introduction, our systems have little hope to satisfy (2.3), but the following theorem shows that F exhibits exponential annealed decay of correlations, up to a given precision that depends on the expansion of the base system.

Theorem 2.2. *Under the assumptions of Theorem 2.1, for $\widetilde{\lambda},\widetilde{C},\overline{\eta}$ given there, it holds*

$$\left|\int_{\mathbb{T}^{m_1}\times\mathbb{T}^{m_2}}\varphi(\pi_2F^n(\omega,x))\psi(x)\ d\omega dx - \int_{\mathbb{T}^{m_2}}\varphi(x)d\overline{\eta}(x)\int_{\mathbb{T}^{m_2}}\psi(x)dx\right| \leqslant C_{\varphi,\psi}(\widetilde{C}\widetilde{\lambda}^n+\varepsilon)$$

for all $\psi\in L^1(\mathbb{T}^{m_2};\mathbb{R})$ and $\varphi\in\mathrm{Lip}(\mathbb{T}^{m_2};\mathbb{R})$ where $C_{\varphi,\psi}>0$ depends on φ,ψ but not from n,ε.

Remark 2.2.

- Given D and L, one might need a large minimal expansion σ_0 to ensure that $\varepsilon>0$ in the theorems above is small. Examples of base maps g with given distortion, and arbitrarily large minimal expansions σ_0 can be constructed easily by fixing any map $g_0:\mathbb{T}^{m_1}\to\mathbb{T}^{m_1}$ satisfying (H0.1)-(H0.2), and considering $g:=g_0^n$ with high $n\in\mathbb{N}$. With this definition, g has minimal expansion equal to the minimal expansion of g raised to the power $n\in\mathbb{N}$, and distortion uniformly bounded with respect to n.
- In the simpler case $g(\omega)=\sigma\omega \bmod 1$, one gets that ε can be chosen of the order of $\sigma^{-\gamma}$ with γ depending on C, L, and λ (see also Remark 3.1).
- Existence of an invariant measure which is physical or with some smoothness such as an SRB measure (see [52] for definitions) has little hope in general. One reason is the low regularity of F which is only Lipschitz. However, imposing higher regularity, for example, F globally $C^{1+\alpha}$, would not be enough: The domination that (possibly) results from the high expansion in the base, even if it can lead to existence of positive Lyapunov exponents, cannot ensure existence of an SRB or physical measure by itself – all the more reasons not to expect exact exponential decay of correlations for the dynamical system F.
- We can give an explicit bound for the constant $C_{\varphi,\psi}$. Letting $\psi-\int_{\mathbb{T}^{m_2}}\psi=\psi_1-\psi_2$ with $\psi_1,\psi_2\geqslant 0$ being the positive and negative components of $\psi-\int_{\mathbb{T}^{m_2}}\psi$,

$$C_{\varphi,\psi}\leqslant 2\|\psi\|_{L^1}(\mathrm{Lip}(\varphi)+1).$$

- As mentioned in the introduction, whenever one has additional information on the fiber maps $\{f_\omega\}_{\omega\in\mathbb{T}}$, other approaches could lead to more precise statements.

2.4 | Examples

One way to ensure that Assumption H holds is by imposing two main regularity requirements on f with respect to the horizontal variable ω, that is, with respect to the forcing: 1) *Regularity condition:* f is C^k in the variable ω for a sufficiently large k. 2) *Non-degeneracy condition:* the differential of f with respect to ω is invertible which, for every $x \in \mathbb{T}^{m_2}$, makes the function $f(\cdot, x) : \mathbb{T}^{m_1} \to \mathbb{T}^{m_2}$ a local diffeomorphism on its range (notice that for this requirement to hold m_1 has to be equal to m_2).

Example 2.1. Let's consider $m_1 = m_2 = m$, and assume that for any $x \in \mathbb{T}^m$ $f(\cdot, x) : \mathbb{T}^m \to \mathbb{T}^m$ is a C^2 local diffeomorphism or, equivalently, $\{f_\omega\}_{\omega\in\mathbb{T}^m}$ is a family of maps with C^2 dependence on the parameter ω such that the differential $(\mathrm{D}f(\cdot, x))_\omega$ is bijective for every $x, \omega \in \mathbb{T}^m$.

From Equation (2.1) one can deduce that

$$\mathcal{P}\delta_x = f(\cdot, x)_* \nu_g$$

and since $f(\cdot, x)$ is a non-singular transformation, the expression of its Perron-Frobenius operator gives

$$\frac{d\mathcal{P}\delta_x}{d\,\mathrm{Leb}_{\mathbb{T}^m}}(y) = \sum_k \frac{\rho_g(y_k)}{|\mathrm{D}f(\cdot, x)_{y_k}|}$$

where the sum is over all the preimages y_k of y under the map $f(\cdot, x)$. $\frac{d\mathcal{P}\delta_x}{d\,\mathrm{Leb}_{\mathbb{T}^m}}$ is in C^1 since $|\mathrm{D}f(\cdot, x)|$ and ρ_g are C^1 functions. It is also uniformly bounded away from zero, as there is $c_1 > 0$ such that $\rho_g > c_1$ (see, e.g., [49]), and there is $K_1 > 0$ such that $|\mathrm{D}f(\cdot, x)_\omega| \leqslant K_1$ for every $\omega, x \in \mathbb{T}^m$. This implies that for every $x \in \mathbb{T}^m$, $\frac{d\mathcal{P}\delta_x}{d\,\mathrm{Leb}_{\mathbb{T}^m}}(y) > cK_1$, that is, the densities of the transition probabilities are all uniformly bounded away from zero. It is well known that the Markov chain generated by $\mathcal{P}$ is geometrically ergodic, that is, satisfies Assumption (H) (see Theorem A.1 in the Appendix).

The following example is a subcase of the example above and shows one of the simplest nontrivial setups.

Example 2.2 (System with additive deterministic noise). $f(\omega, x) = T(x) + h(\omega)$, where $T : \mathbb{T}^m \to \mathbb{T}^m$ is any Lipschitz map, and $h : \mathbb{T}^m \to \mathbb{T}^m$ is a local C^2 diffeomorphism.

Let us stress that these sufficient conditions for Assumption H: 1) are by no means necessary; 2) give no control on a single fiber map f_ω beyond the requirement that it is Lipschitz regular; 3) do not imply good spectral properties for F_*.

2.5 | Sketch of the proof

To prove Theorem 2.1 and Theorem 2.2, we are going to study the evolution of probability measures on $\mathbb{T}^{m_1} \times \mathbb{T}^{m_2}$ that have a Lipschitz disintegration with a focus on the evolution of their vertical marginals. To do so we follow the steps below.

1) First of all we show that under the assumptions of the main theorem, if $\mu \in \mathcal{M}_1(\mathbb{T}^{m_1} \times \mathbb{T}^{m_2})$ has Lipschitz disintegration, so does $F_*^n\mu$ for any $n \in \mathbb{N}$, and $\mathrm{Lip}(F_*^n\mu)$ is bounded uniformly in $n \in \mathbb{N}$ (see Proposition 4.2). This is a consequence of the uniform (high) expansion of the map g. This result shows the existence of an invariant class of measures whose disintegration is smooth along the $\mathbb{T}^{m_1}$-coordinate, and is proved by using an explicit expression for a disintegration of $F_*\mu$ in terms of a disintegration of μ (see Proposition 4.1).
2) Next, we use the above fact to show that the vertical marginal of $F_*^n\mu$ can be approximated by looking at the action of an auxiliary operator, $\boldsymbol{\mathcal{L}}$, that acts on a suitable decomposition of μ and that, unlike F_*, has good contraction properties (see (4.6) for the definition of $\boldsymbol{\mathcal{L}}$, Proposition 4.3, and Proposition 4.4). The contraction properties of $\boldsymbol{\mathcal{L}}$ are inherited from those of $\mathcal{P}$ and imply a nice spectral picture (see Remark 4.5). Loosely speaking, when $g(\omega) = \sigma\omega \bmod 1$, $\boldsymbol{\mathcal{L}}$ can be taken to be equal to $\mathcal{P}$, see Section 3.
3) The above approximation allows to show that under application of F_*, the system exhibits approximate exponential memory loss[†] on its vertical marginal. By this we mean that given any two probability measures $\mu_1, \mu_2 \in \mathcal{M}_1(\mathbb{T}^{m_1} \times \mathbb{T}^{m_2})$ with Lipschitz disintegration, the Kantorovich–Wasserstein distance between vertical marginals $\Pi F_*^n\mu_1$ and $\Pi F_*^n\mu_2$ shrinks exponentially fast modulo an approximation error (see Proposition 4.5).
4) Finally, we use the above approximate memory loss to prove the existence of an approximate stationary measure and of approximate decay of correlations.

Remark 2.3. Picking a metric on $\mathcal{M}_1(\mathbb{T}^{m_2})$ as weak as the Wasserstein metric d_W is crucial to our arguments. Measures having Lipschitz disintegration with respect to other stronger metrics – for example the TV distance – may not be invariant under F_* without further assumptions on $\{f_\omega\}_{\omega\in\Omega}$.

In Section 3, we give a proof of Theorem 2.1 and Theorem 2.2 in the simpler case where: $m_1 = m_2 = 1$; the dynamics in the base is smooth and has no distortion, that is, $D = 0$. Under these assumptions $g : \mathbb{T} \to \mathbb{T}$ can be written as

$$g(\omega) = \sigma\omega \mod 1, \qquad \sigma \in \mathbb{N}\backslash\{1\}. \tag{H0+}$$

This is for the sake of presentation since in this case the treatment of points 2) and 3) does not require the introduction of the auxiliary operator $\boldsymbol{\mathcal{L}}$, and ΠF_*^n can be approximated directly, in some sense made precise below, with $\mathcal{P}$. This makes the proof much easier than in the general case with $D > 0$ and it allows for explicit bounds for ε w.r.t σ.

3 | CASE WITHOUT DISTORTION

In this section we work under Assumption H0+ and Assumption H. Namely we consider $g : \mathbb{T} \to \mathbb{T}$ defined as $g(\omega) = \sigma\omega \mod 1$, where $\sigma \in \mathbb{N}\backslash\{1\}$. Recall that under these assumptions $\nu_g = \mathrm{Leb}_\mathbb{T}$ and ρ_g is constant equal to one. Take $\mu \in \mathcal{M}_1(\mathbb{T}\times\mathbb{T})$ having horizontal marginal $\mathrm{Leb}_\mathbb{T}$. We study the evolution of μ under applications of F_*.

First of all notice that as a consequence of the skew-product structure of F and invariance of $\mathrm{Leb}_\mathbb{T}$ under g, also $F_*\mu$ has horizontal marginal equal to $\mathrm{Leb}_\mathbb{T}$. Secondly, the evolution of the disintegration along vertical fibres, is given by the following proposition.

[†] For a definition and an example of *exact* memory loss see, for example, [39].

Proposition 3.1. *Let μ be a probability measure on $\mathbb{T}\times\mathbb{T}$ with horizontal marginal equal to* $\mathrm{Leb}_{\mathbb{T}}$. *Let $\{\mu_\omega\}_{\omega\in\mathbb{T}}$ be a disintegration of μ along vertical fibres, then a disintegration of $F_*\mu$ along vertical fibres is given by $\{(F_*\mu)_\omega\}_{\omega\in\mathbb{T}}$ with*

$$(F_*\mu)_\omega = \frac{1}{\sigma}\sum_{i=0}^{\sigma-1} f_{\frac{\omega+i}{\sigma}\,*}\mu_{\frac{\omega+i}{\sigma}}. \tag{3.1}$$

This statement is a particular case of Proposition 4.1 below, therefore we omit the proof. It is sufficient to say that, in this setting one has

$$(F_*\mu)_\omega(I) = \lim_{\delta\to 0}\frac{(F_*\mu)([\omega-\delta,\omega+\delta]\times I)}{2\delta}$$

for any interval I, where the numerator can be easily controlled.

Next, recall Definition 2.1. In the proposition below we use (3.1) to deduce that if μ has Lipschitz disintegration, then so do all its iterates $F_*^n\mu$, and if σ is sufficiently large, then their Lipschitz constants are all uniformly bounded and small when $\sigma\to\infty$.

Before moving to the next proposition we recall for the reader's convenience a property of the Kantorovich–Wasserstein distance (see, e.g., [50] for details). Given a Borel signed measure ξ on $\mathbb{T}$ with $\xi(\mathbb{T})=0$, consider the Wasserstein norm

$$\|\xi\|_W := \sup_{\varphi\in\mathrm{Lip}^1(\mathbb{T})}\int_{\mathbb{T}}\varphi\, d\xi.$$

Recall that we denoted by $\mathrm{Lip}^1(\mathbb{T}) := \{\varphi : \mathbb{T}\to\mathbb{R} : \mathrm{Lip}(\varphi)\leqslant 1\}$ the Lipschitz functions on $(\mathbb{T}, |\cdot|)$ with Lipschitz constant less or equal to one (we write Lip^1 when there is no risk of confusion). The Kantorovich–Wasserstein distance defined in (2.2) can be rewritten as

$$d_W(\mu,\nu) = \|\mu-\nu\|_W = \sup_{\varphi\in\mathrm{Lip}^1}\int_{\mathbb{T}}\varphi(x)\, d(\mu-\nu)(x).$$

This characterization will simplify the notation later in the proofs.[†]

Proposition 3.2. *Let μ be a probability measure on $\mathbb{T}\times\mathbb{T}$ with horizontal marginal* $\mathrm{Leb}_{\mathbb{T}}$ *and Lipschitz disintegration $\{\mu_\omega\}_{\omega\in\mathbb{T}}$. Then the disintegration of $F_*\mu$ defined in Equation (3.1) is also Lipschitz and*

$$\mathrm{Lip}\,(F_*\mu)\leqslant L\sigma^{-1}\,\mathrm{Lip}(\mu)+L\sigma^{-1}$$

Proof. For $\omega,\omega'\in\mathbb{T}$

$$\begin{aligned} d_W((F_*\mu)_\omega,(F_*\mu)_{\omega'}) &= \sup_{\varphi\in\mathrm{Lip}^1}\int_{\mathbb{T}}\varphi\, d\left(\frac{1}{\sigma}\sum_{i=0}^{\sigma-1} f_{\frac{\omega+i}{\sigma}\,*}\mu_{\frac{\omega+i}{\sigma}} - \frac{1}{\sigma}\sum_{i=0}^{\sigma-1} f_{\frac{\omega'+i}{\sigma}\,*}\mu_{\frac{\omega'+i}{\sigma}}\right) \\ &\leqslant \frac{1}{\sigma}\sum_{i=0}^{\sigma-1}\sup_{\varphi\in\mathrm{Lip}^1}\int_{\mathbb{T}}\varphi d\left(f_{\frac{\omega+i}{\sigma}\,*}\mu_{\frac{\omega+i}{\sigma}} - f_{\frac{\omega'+i}{\sigma}\,*}\mu_{\frac{\omega'+i}{\sigma}}\right) \end{aligned}$$

[†] In this section, the absence of distortion keeps the proofs of the next propositions rather classical and expert readers could skip to the general case. However, we included the details here to highlight and estimate the dependence of ε on σ.

Calling $\omega_i := \frac{\omega+i}{\sigma}$, and $\omega_i' := \frac{\omega'+i}{\sigma}$ for brevity, we have

$$\sup_{\varphi\in\mathrm{Lip}^1}\int_{\mathbb{T}}\varphi d\Big(f_{\omega_i *}\mu_{\omega_i} - f_{\omega_i' *}\mu_{\omega_i'}\Big) = d_W(f_{\omega_i *}\mu_{\omega_i}, f_{\omega_i' *}\mu_{\omega_i'})$$

$$\leqslant d_W(f_{\omega_i *}\mu_{\omega_i}, f_{\omega_i' *}\mu_{\omega_i}) + d_W(f_{\omega_i' *}\mu_{\omega_i}, f_{\omega_i' *}\mu_{\omega_i'}).$$

For the first term above, notice that for any $\xi \in \mathcal{M}_1(\mathbb{T})$ and $\varphi \in \mathrm{Lip}^1$

$$\int_{\mathbb{T}}\varphi d(f_{\omega_i *}\xi - f_{\omega_i' *}\xi) = \int_{\mathbb{T}}(\varphi\circ f_{\omega_i}(x) - \varphi\circ f_{\omega_i'}(x))d\xi(x) \leqslant L\sigma^{-1}|\omega-\omega'|,$$

where L is the Lipschitz constant of f, implying

$$d_W(f_{\omega_i *}\mu_{\omega_i}, f_{\omega_i' *}\mu_{\omega_i}) \leqslant L\sigma^{-1}|\omega-\omega'|.$$

The second term can be bounded using an analogous computation

$$d_W(f_{\omega_i' *}\mu_{\omega_i}, f_{\omega_i' *}\mu_{\omega_i'}) \leqslant L d_W(\mu_{\omega_i}, \mu_{\omega_i'}) \leqslant L\,\mathrm{Lip}(\mu)\sigma^{-1}|\omega-\omega'|.$$

where we used that the Lipschitz constant of $f_{\omega *}$ is equal to the Lipschitz constant of f_ω (see Lemma C.1 in the Appendix) which is upper bounded by L as in (H0.3).

Putting all the estimates back together we obtain

$$d_W((F_*\mu)_\omega, (F_*\mu)_{\omega'}) \leqslant L\sigma^{-1}[1+\mathrm{Lip}(\mu)]\ |\omega-\omega'|. \qquad \square$$

As a corollary to the previous proposition, for σ sufficiently large, we obtain the existence of an invariant class of measures whose disintegration has Lipschitz dependence on the variable $\omega \in \mathbb{T}$, and whose Lipschitz constant goes to zero as $\sigma \to \infty$. More precisely, let us define the set $\mathcal{M}_{1,\mathrm{Leb}_{\mathbb{T}}}(\mathbb{T}\times\mathbb{T})$ of probability measures on $\mathbb{T}\times\mathbb{T}$ with horizontal marginal $\mathrm{Leb}_{\mathbb{T}}$. Let's call $\Gamma_\ell \subset \mathcal{M}_{1,\mathrm{Leb}_{\mathbb{T}}}$ the set of those probability measures that have Lipschitz disintegration with Lipschitz constant at most ℓ:

$$\Gamma_\ell := \Big\{\mu \in \mathcal{M}_{1,\mathrm{Leb}_{\mathbb{T}}} \ :\ \mathrm{Lip}(\mu) \leqslant \ell\Big\}.$$

Corollary 3.1. *If $\sigma > L$, then the set Γ_ℓ is invariant under the push-forward F_* for every $\ell \geqslant \ell_0$ with*

$$\ell_0 := \frac{L\sigma^{-1}}{1-L\sigma^{-1}}.$$

Actually one can show more: For $\ell > \ell_0$, Γ_ℓ is mapped by F_* into $\Gamma_{\ell'}$ with $\ell' < \ell$ and so F_* "regularizes" the Lipschitz constant of the disintegration.

The following proposition controls the evolution of vertical marginals for two probability measures in Γ_{ℓ_0} under application of F_*. In the statements below, the constants C and λ are the same as those in Assumption (H).

Proposition 3.3 (Approximate Memory Loss). *For every $\varepsilon > 0$ there is $\sigma_0(\varepsilon) > L$ such that if $\sigma > \sigma_0(\varepsilon)$ then*

i) $$d_W(\Pi F_*^n \mu_1, \Pi F_*^n \mu_2) \leqslant C\lambda^n + \varepsilon, \qquad \forall \mu_1, \mu_2 \in \Gamma_{\ell_0};$$

ii) $$d_W(\Pi F_*^n \mu, \eta_0) \leqslant C\lambda^n + \varepsilon, \qquad \forall \mu \in \Gamma_{\ell_0};$$

where η_0 is the stationary measure for $\mathcal{P}$.

Proof. Let $\mu := \mu_1 - \mu_2$ and recall that $\Pi\mu = \int_{\mathbb{T}} \mu_\omega \, d\omega$ is the vertical marginal of μ. Since

$$d_W(\mu_\omega, \mu_{\omega'}) \leqslant \ell_0 |\omega - \omega'| \leqslant \ell_0$$

then $d_W(\mu_\omega, \Pi\mu) \leqslant \ell_0$ (see Lemma C.2 in the Appendix). Therefore,

$$\Pi F_* \mu = \int_{\mathbb{T}} d\omega f_{\omega *} \mu_\omega = \mathcal{P}(\Pi\mu) + \int_{\mathbb{T}} d\omega f_{\omega *}(\mu_\omega - \Pi\mu),$$

where $\mathcal{P}$ is defined in Equation (2.1) and, by Lemma C.1,

$$\left\| \int_{\mathbb{T}} d\omega f_{\omega *}(\mu_\omega - \Pi\mu) \right\|_W \leqslant L\ell_0.$$

For higher iterates, one gets the telescopic sum

$$\begin{aligned} \Pi F_*^n \mu &= \int_{\mathbb{T}} d\omega_{n-1} f_{\omega_{n-1} *}((F_*^{n-1}\mu)_{\omega_{n-1}} - \Pi F_*^{n-1}\mu) + \int_{\mathbb{T}} d\omega_{n-1} f_{\omega_{n-1} *}(\Pi F_*^{n-1}\mu) \\ &= \mathcal{P}^n(\Pi\mu) + \sum_{i=0}^{n-1} \int_{\mathbb{T}} d\omega_{n-1} f_{\omega_{n-1} *} \dots \int_{\mathbb{T}} d\omega_i f_{\omega_i *}((F_*^i \mu)_{\omega_i} - \Pi F_*^i \mu) \end{aligned} \tag{3.2}$$

and by triangle inequality

$$\|\Pi F_*^n \mu\|_W \leqslant \|\mathcal{P}^n \Pi\mu\|_W + \left\| \sum_{i=0}^{n-1} \int_{\mathbb{T}} d\omega_{n-1} f_{\omega_{n-1} *} \dots \int_{\mathbb{T}} d\omega_i f_{\omega_i *}((F_*^i \mu)_{\omega_i} - \Pi F_*^i \mu) \right\|_W. \tag{3.3}$$

For the first term in the above inequality,

$$\|\mathcal{P}^n(\Pi\mu)\|_W \leqslant C\lambda^n$$

since $d_W \leqslant d_{TV}$ (see Lemma C.3) together with Assumption H. For the second term, each summand can be treated as follows

$$\begin{aligned} &\left\| \int_{\mathbb{T}} d\omega_{n-1} f_{\omega_{n-1} *} \dots \int_{\mathbb{T}} d\omega_i f_{\omega_i *} \left[(F_*^i \mu_1)_{\omega_i} - (F_*^i \mu_2)_{\omega_i} - \Pi F_*^i \mu_1 + \Pi F_*^i \mu_2 \right] \right\|_W \leqslant \\ &\leqslant \sup_\omega \mathrm{Lip}(f_{\omega *})^{n-1-i} \sup_\omega \left\| (F_*^i \mu_1)_{\omega_i} - (F_*^i \mu_2)_{\omega_i} - \Pi F_*^i \mu_1 + \Pi F_*^i \mu_2 \right\|_W \\ &\leqslant 2L^{n-1-i} \ell_0. \end{aligned} \tag{3.4}$$

Now, one can pick $n_0 \in \mathbb{N}$ such that $C\lambda^{n_0} \leqslant \varepsilon/2$, and $\sigma_0 > 0$ so that

$$\frac{L\sigma_0^{-1}}{1 - L\sigma_0^{-1}} \sum_{i=0}^{n_0-1} L^{n_0-1-i} = \ell_0 \sum_{i=0}^{n_0-1} L^{n_0-1-i} \leqslant \varepsilon/2.$$

This way, if $n \leqslant n_0$

$$\|\Pi F_*^n \mu\|_W \leqslant C\lambda^n + 2\ell_0 \sum_{i=0}^{n-1} L^{n-1-i} \leqslant C\lambda^n + \varepsilon/2$$

and if $n \geqslant n_0$,

$$d_W(\Pi F_*^n \mu_1, \Pi F_*^n \mu_2) = d_W(\Pi F_*^{n_0} F_*^{n-n_0} \mu_1, \Pi F_*^{n_0} F_*^{n-n_0} \mu_2) \leqslant C\lambda^{n_0} + \varepsilon/2 \leqslant C\lambda^n + \varepsilon,$$

since $F_*^{n-n_0}\mu_1$ and $F_*^{n-n_0}\mu_2$ both belong to Γ_{ℓ_0} which proves point i).

For point ii), going back to (3.2) and picking n_0 and σ_0 as above, for any $\mu \in \Gamma_{\ell_0}$ and $n \leqslant n_0$

$$d_W(\Pi F_*^n \mu, \eta_0) \leqslant d_W(\Pi F_*^n \mu, \mathcal{P}^n \Pi \mu) + d_W(\mathcal{P}^n \Pi \mu, \eta_0) \leqslant C\lambda^n + \varepsilon/2$$

while for $n \geqslant n_0$ we use an analogous computation and get

$$d_W(\Pi F_*^n \mu, \eta_0) \leqslant C\lambda^n + \varepsilon$$

□

We can now proceed with the proof of the main theorems in the case without distortion.

Proof of Theorem 2.1. *under condition* (H0+). The statement of the theorem is an immediate corollary of point ii) of the proposition above. □

Proof of Theorem 2.2. *under condition* (H0+). Assume that $\int \psi(x)dx = 0$. Then $\psi = \psi_1 - \psi_2$ where $\psi_1, \psi_2 \geqslant 0$ are the positive and negative parts of ψ and $\int \psi_1 = \int \psi_2 =: M$. Take μ the measure on $\mathbb{T} \times \mathbb{T}$ defined as

$$d\mu(\omega, x) = M^{-1}(\psi_1(x) - \psi_2(x))d\omega dx. \tag{3.5}$$

It follows that $\mu = \mu_1 - \mu_2$ where μ_1, μ_2 are probability measures having constant disintegrations $\mu_{1,\omega} = M^{-1}\psi_1(x)dx$ and $\mu_{2,\omega} = M^{-1}\psi_2(x)dx$. In particular, $\mu_1,\ \mu_2 \in \Gamma_{\ell_0}$.

Now, picking σ_0 as in Proposition 3.3, if $\sigma > \sigma_0$,

$$\begin{aligned}\left|\int_{\mathbb{T}\times\mathbb{T}} \varphi \circ F^n(\omega, x)\psi(x)dxd\omega\right| &= M\left|\int_{\mathbb{T}} \varphi(x)d(\Pi F_*^n \mu)(x)\right| \\ &\leqslant M \operatorname{Lip}(\varphi)(C\lambda^n + \varepsilon)\end{aligned} \tag{3.6}$$

where we used that φ does not depend on $\omega \in \mathbb{T}$, and Proposition 3.3.

If $\int \psi \neq 0$ consider $\widetilde{\psi} := \psi - \int \psi$.

$$\begin{aligned}\int_{\mathbb{T}\times\mathbb{T}} \varphi \circ F^n(\omega,x)\psi(x)dxd\omega &= \int_{\mathbb{T}\times\mathbb{T}} \varphi \circ F^n(\omega,x)\widetilde{\psi}(x)dxd\omega \\ &\quad + \left(\int_{\mathbb{T}\times\mathbb{T}} \varphi \circ F^n(\omega,x)d\omega dx - \int_{\mathbb{T}} \varphi(x)d\eta_0(x)\right)\left(\int_{\mathbb{T}} \psi(x)dx\right) \\ &\quad + \left(\int \varphi(x)d\eta_0(x)\right)\left(\int_{\mathbb{T}} \psi(x)dx\right).\end{aligned}$$

For the first term, we use (3.6); for the second term

$$\begin{aligned}\left|\int_{\mathbb{T}\times\mathbb{T}} \varphi \circ F^n(\omega,x)d\omega dx - \int_{\mathbb{T}} \varphi(x)d\eta_0(x)\right| &= \left|\int_{\mathbb{T}} \varphi(x)d(\Pi F^n_* \operatorname{Leb}_{\mathbb{T}\times\mathbb{T}} - \eta_0)(x)\right| \\ &\leqslant \operatorname{Lip}(\varphi)d_W(\Pi F^n_* \operatorname{Leb}_{\mathbb{T}\times\mathbb{T}}, \eta_0)\end{aligned}$$

and from point ii) of Proposition 3.3 the above is less than $\operatorname{Lip}(\varphi)[C\lambda^n + \varepsilon]$. By triangle inequality

$$\left|\int_{\mathbb{T}\times\mathbb{T}} \varphi \circ F^n(\omega,x)\psi(x)dxd\omega - \left(\int_{\mathbb{T}} \varphi(x)d\eta_0(x)\right)\left(\int_{\mathbb{T}} \psi(x)dx\right)\right| \leqslant C_{\varphi,\psi}(C\lambda^n + \varepsilon)$$

where $C_{\varphi,\psi} \leqslant \frac{3}{2}\|\psi\|_{L^1}(\operatorname{Lip}(\varphi) + 1)$. □

Remark 3.1. As a remark, note that if one tries to estimate the quantifier ε in Theorem 2.2 for a given datum, by inspecting the proof of this simpler case, one realizes that a bound for ε is proportional to the smallest number one gets from the sequence $\{\max\{C\lambda^n, 2\sigma^{-1}L^n\}\}_{n\in\mathbb{N}}$. Since the first sequence is decreasing, while the second is increasing, the optimal trade off is achieved when they are of about the same size. Imposing $C\lambda^n = 2\sigma^{-1}L^n$ gives the estimate $\varepsilon \lesssim \sigma^{-\gamma}$ for some $\gamma > 0$ which depends on C, λ, and L.

4 | GENERAL CASE: PROOF OF THEOREM 2.2

4.1 | Control on the disintegration along vertical fibres

Below we assume that F satisfies assumptions (H0.1)-(H0.3) and Assumption (H).

Take a measure μ_0 on $\mathbb{T}^{m_1} \times \mathbb{T}^{m_2}$ with horizontal marginal equal to $\nu_0 \in \mathcal{M}_1(\mathbb{T}^{m_1})$ which is absolutely continuous with respect to Lebesgue, and let $\mu_1 := F_*\mu_0$. It follows from the skew-product structure of F that the horizontal marginal of μ_1 equals $\nu_1 := g_*\nu_0$. We will denote by ρ_1 the density of ν_1 †. Recall from Section 2 that g is a local diffeomorphism, g_i are its invertible branches, and h_i their inverses. Then an explicit expression of ρ_1 in terms of ρ_0 is given by

$$\rho_1(\omega) = \sum_{i=1}^{d} \frac{\rho_0(\omega_i)}{|\mathrm{D}g_{\omega_i}|} \quad \forall \omega \in \mathbb{T}^{m_1}$$

where we denote by $\omega_i = h_i\omega$ the preimages of ω and $|\mathrm{D}g_{\omega_i}|$.

† Since g is a local diffeomorphism is in particular *nonsingular*, its push-forward sends absolutely continuous measures to absolutely continuous measures.

For $k \in \{0,1\}$, let $\{\mu_{k,\omega}\}_{\omega\in\mathbb{T}^{m_1}}$ be a disintegration of μ_k w.r.t. the measurable partition $\{\{\omega\}\times\mathbb{T}^{m_2}\}_{\omega\in\mathbb{T}^{m_1}}$. For a definition and some results on disintegrations see Appendix B.

Proposition 4.1. *A disintegration $\{\mu_{1,\omega}\}_{\omega\in\mathbb{T}^{m_1}}$ of μ_1 is given by*

$$\mu_{1,\omega} = \frac{1}{\rho_1(\omega)}\sum_{i=1}^{d}\frac{\rho_0(\omega_i)}{|\mathrm{D}g_{\omega_i}|}f_{\omega_i *}\mu_{0,\omega_i}. \tag{4.1}$$

Proof. Let $B_\delta(\omega)\subset\mathbb{T}^{m_1}$ be the Euclidean ball centered at ω of radius δ. By Theorem B.1 in Appendix B, for $\mathrm{Leb}_{\mathbb{T}^{m_1}}$-a.e. ω

$$\mu_{1,\omega} = \lim_{\delta\to 0}\frac{\int_{B_\delta(\omega)} ds\rho_1(s)\mu_{1,s}}{\int_{B_\delta(\omega)} ds\rho_1(s)}, \tag{4.2}$$

where the limit is with respect to the weak$*$ topology. Using the definition of disintegration and that $\mu_1(B_\delta(\omega)\times I) = \mu_0(F^{-1}(B_\delta(\omega)\times I))$, for every measurable set I on $\mathbb{T}^{m_2}$ and $\delta > 0$ sufficiently small, one gets

$$\int_{B_\delta(\omega)} ds\rho_1(s)\mu_{1,s} = \sum_{i=1}^{d}\int_{h_i(B_\delta(\omega))} ds\rho_0(s)f_{s*}\mu_{0,s}.$$

By changing variables, $s = h_i(s')$, and multiplying and dividing by $\rho_1(s)$, the above equals

$$\int_{B_\delta(\omega)} ds'\rho_1(s')\left[\frac{1}{\rho_1(s')}\sum_{i=1}^{d}\frac{\rho_0(s'_i)}{|\mathrm{D}g_{s'_i}|}f_{s'_i *}\mu_{0,s'_i}\right],$$

where we denoted $s'_i = h_i(s')$. Applying Lebesgue's differentiation theorem, Equation (4.2) becomes

$$\mu_{1,\omega} = \frac{1}{\rho_1(\omega)}\sum_{i=1}^{d}\frac{\rho_0(\omega_i)}{|\mathrm{D}g_{\omega_i}|}f_{\omega_i *}\mu_{0,\omega_i}.$$

□

The formula for the evolution of disintegrations in (4.1) depends on ν_0 and ν_1, the horizontal marginals of the measures μ_0 and μ_1. Thanks to assumptions on g, the evolution of the horizontal can be controlled (see Lemma 4.1 below).

Consider for $a \geqslant 0$, the cone of log-Lipschitz functions,

$$\mathcal{V}_a := \left\{\varphi : \mathbb{T}^{m_1}\to\mathbb{R}^+ : \frac{\varphi(\omega)}{\varphi(\omega')} \leqslant e^{a|\omega-\omega'|}\right\}.$$

The following lemma gathers some standard facts about uniformly expanding maps with bounded distortion, such as g.

Lemma 4.1. *Let $g : \mathbb{T}^{m_1}\to\mathbb{T}^{m_1}$ be a C^2 local diffeomorphism satisfying (H0.1)-(H0.2), and let ρ_0 and ρ_1 be as above.*

i) *If* $\rho_0 \in \mathcal{V}_a$, *then* $\rho_1 \in \mathcal{V}_{\sigma^{-1}a+D}$. *In particular, if* $a \geqslant a_0 := \frac{D}{1-\sigma^{-1}}$, *then* $\rho_1 \in \mathcal{V}_a$*;*
ii) *If* $\rho_0 \in \mathcal{V}_{a_0}$, *calling* ρ_n *the density of* $\nu_n := g_*^n \nu_0$, *there are* $C_g > 0$ *and* $\lambda_g \in (0,1)$ *such that*

$$\|\rho_n - \rho_g\|_\infty := \sup_{\omega \in \mathbb{T}^{m_1}} |\rho_n(\omega) - \rho_g(\omega)| \leqslant C_g \lambda_g^n.$$

Proof. See, for example, [36], [49]. □

Remark 4.1. In point ii) of Lemma 4.1, one can choose $C_g := C_g(D,\sigma)$ and $\lambda_g := \lambda_g(D,\sigma)$. Moreover, fixed D, C_g and λ_g can be chosen to be decreasing with respect to σ. This implies that fixed $D > 0$ and $\sigma_0 > 1$, there are constants $\overline{C}$ and $\overline{\lambda} \in (0,1)$ such that $C_g < \overline{C}$ and $\lambda_g < \overline{\lambda}$ for any g satisfying (H0.1)-(H0.2) with $\sigma \geqslant \sigma_0$.

From now on we will restrict our analysis to probability measures on $\mathbb{T}^{m_1} \times \mathbb{T}^{m_2}$ whose horizontal marginals belong to $\mathcal{V}_a$ for some $a > 0$.

Proposition 4.2. *Assume* $\rho_0 \in \mathcal{V}_a$ *for some* $a \geqslant a_0$ *and that* μ_0 *has Lipschitz disintegration. Then the disintegration of* μ_1 *given in (4.1) is Lipschitz and*

$$\mathrm{Lip}\,(\mu_1) \leqslant \sigma^{-1} L \,\mathrm{Lip}\,(\mu_0) + \left[C_a + \sigma^{-1} L\right]$$

where $C_a := e^{(a+\sigma^{-1}a+D)C_1}$ *and* C_1 *is the diameter of* $\mathbb{T}^{m_1}$.

Proof. The proof is analogous to that of Proposition 3.2, although one has to work with (4.1), rather then the simpler formula (3.1). For $\omega, \omega' \in \mathbb{T}^{m_1}$,

$$\begin{aligned}
d_W(\mu_{1,\omega}, \mu_{1,\omega'}) = \sup_{\varphi \in \mathrm{Lip}^1} \int \varphi d\Bigg(\sum_{i=1}^d \frac{1}{\rho_1(\omega)} \frac{\rho_0(\omega_i)}{|\mathrm{D}g_{\omega_i}|} f_{\omega_i *}\mu_{0,\omega_i} - \frac{1}{\rho_1(\omega')} \frac{\rho_0(\omega_i')}{|\mathrm{D}g_{\omega_i'}|} f_{\omega_i' *}\mu_{0,\omega_i'}\Bigg) \\
\leqslant \sup_{\varphi \in \mathrm{Lip}^1} \int \varphi d\Bigg(\sum_{i=1}^d \frac{1}{\rho_1(\omega)} \frac{\rho_0(\omega_i)}{|\mathrm{D}g_{\omega_i}|} f_{\omega_i *}\mu_{0,\omega_i} - \frac{1}{\rho_1(\omega')} \frac{\rho_0(\omega_i')}{|\mathrm{D}g_{\omega_i'}|} f_{\omega_i *}\mu_{0,\omega_i}\Bigg) + \\
+ \sup_{\varphi \in \mathrm{Lip}^1} \sum_{i=1}^d \int \varphi d\Bigg(\frac{1}{\rho_1(\omega')} \frac{\rho_0(\omega_i')}{|\mathrm{D}g_{\omega_i'}|} f_{\omega_i *}\mu_{0,\omega_i} - \frac{1}{\rho_1(\omega')} \frac{\rho_0(\omega_i')}{|\mathrm{D}g_{\omega_i'}|} f_{\omega_i' *}\mu_{0,\omega_i'}\Bigg) \\
=: A + B,
\end{aligned}$$

where to get the inequality we added and subtracted the same quantity and distributed the sup.

Upper bound for A. To bound the first term

$$\begin{aligned}
A = \sup_{\varphi \in \mathrm{Lip}^1} \sum_{i=1}^d \frac{1}{\rho_1(\omega)} \frac{\rho_0(\omega_i)}{|\mathrm{D}g_{\omega_i}|} \left(1 - \frac{\frac{1}{\rho_1(\omega')} \frac{\rho_0(\omega_i')}{|\mathrm{D}g_{\omega_i'}|}}{\frac{1}{\rho_1(\omega)} \frac{\rho_0(\omega_i)}{|\mathrm{D}g_{\omega_i}|}} \right) \int \varphi d(f_{\omega_i *}\mu_{0,\omega_i}) \\
\leqslant \frac{1}{\rho_1(\omega)} \sum_{i=1}^d \frac{\rho_0(\omega_i)}{|\mathrm{D}g_{\omega_i}|} \left| 1 - e^{[a+\sigma^{-1}a+D]|\omega-\omega'|} \right| \\
\leqslant e^{[a+\sigma^{-1}a+D]C_1} |\omega - \omega'|,
\end{aligned}$$

where $C_1 > 0$ is the diameter of $\mathbb{T}^{m_1}$. To estimate the ratio in parenthesis, we used that: $\rho_0 \in \mathcal{V}_a$ with $a \geqslant a_0$ implies $\rho_1 \in \mathcal{V}_a$, $|\varphi| \leqslant 1$, and (H0.2).

Upper bound for B. The second term can be bounded by

$$d_W\left(\sum_{i=1}^{d} \frac{1}{\rho_1(\omega')} \frac{\rho_0(\omega_i')}{|\mathrm{D}g_{\omega_i'}|} f_{\omega_i *}\mu_{0,\omega_i}, \sum_{i=1}^{d} \frac{1}{\rho_1(\omega')} \frac{\rho_0(\omega_i')}{|\mathrm{D}g_{\omega_i'}|} f_{\omega_i' *}\mu_{0,\omega_i'}\right) \leqslant \max_i d_W(f_{\omega_i *}\mu_{0,\omega_i}, f_{\omega_i' *}\mu_{0,\omega_i'}),$$

where we used that $\sum_{i=1}^{d} \frac{1}{\rho_1(\omega')} \frac{\rho_0(\omega_i')}{|\mathrm{D}g_{\omega_i'}|} = 1$, and Lemma C.4 on the Wasserstein distance between convex combinations of measures. The distance $d_W(f_{\omega_i *}\mu_{0,\omega_i}, f_{\omega_i' *}\mu_{0,\omega_i'})$ can be estimated as in the proof of Proposition 3.2 and gives

$$B \leqslant \sigma^{-1} L[1 + \mathrm{Lip}(\mu_0)]\, |\omega - \omega'|.$$

Putting together the estimates for A and B

$$\mathrm{Lip}\,(\mu_1) \leqslant \sigma^{-1} L\, \mathrm{Lip}\,(\mu_0) + \left[C_a + \sigma^{-1} L\right].$$

□

As a corollary to the previous proposition we obtain the existence of an invariant class of measures, $\Gamma_{\ell,a}$, that have a Lipschitz disintegration with constant at most ℓ and horizontal marginal with density in $\mathcal{V}_a$:

$$\Gamma_{\ell,a} := \left\{\mu \in \mathcal{M}_1(\mathbb{T}^{m_1} \times \mathbb{T}^{m_2}) : \ \mathrm{Lip}\,(\mu) \leqslant \ell \ \text{ and } \ \frac{d\pi_{1*}\mu}{d\,\mathrm{Leb}_{\mathbb{T}^{m_1}}} \in \mathcal{V}_a\right\}.$$

Corollary 4.1.

i) *If $\rho \in \mathcal{V}_a$ with $a \leqslant a_0 := \frac{D}{1-\sigma^{-1}}$ and $\sigma > L$, then $F_*(\Gamma_{\ell,a}) \subset \Gamma_{\ell,a_0}$ for every $\ell \geqslant \ell_0$ with*

$$\ell_0 := \frac{\sigma^{-1} L + C_{a_0}(1 + D)}{\left(1 - \sigma^{-1} L\right)}. \tag{4.3}$$

ii) *If there are $a > 0$ and $\ell > 0$ such that $\mu \in \Gamma_{\ell,a}$, then for every $\delta > 0$ there is $N \in \mathbb{N}$ such that*

$$\mathrm{Lip}(F_*^n \mu) \leqslant \ell_0 + \delta$$

for all $n > N$.

Proof. To prove i), recall that the horizontal marginal of $F_*\mu$ is the push-forward under g of the horizontal marginal of μ. Since the horizontal marginal of μ has density in $\mathcal{V}_{a_0}$, by the inclusion $\mathcal{V}_a \subset \mathcal{V}_{a_0}$ and Lemma 4.1, the horizontal marginal of $F_*\mu$ belongs to $\mathcal{V}_{a_0}$. By Proposition 4.2 and the choice of ℓ_0, it follows that if $\mathrm{Lip}(\mu) \leqslant \ell_0$ then also $\mathrm{Lip}(F_*\mu) \leqslant \ell_0$.

For point ii) notice that $F_*^n\mu$ belongs to Γ_{ℓ_n,a_n} for some a_n and ℓ_n such that $a_n \to a_0$ as $n \to \infty$ by Lemma 4.1, and $\ell_n \to \ell_0$ as $n \to \infty$ by Proposition 4.2. The claim follows easily. □

4.2 | Tracking the evolution of the vertical marginal

Let's consider $\mathcal{M}_{1,\nu_g}(\mathbb{T}^{m_1}\times\mathbb{T}^{m_2})$ the set of Borel probability measures on $\mathbb{T}^{m_1}\times\mathbb{T}^{m_2}$ having horizontal marginal equal to ν_g, the invariant measure for g, and recall that for $\mu\in\mathcal{M}_{1,\nu_g}(\mathbb{T}^{m_1}\times\mathbb{T}^{m_2})$, the vertical marginal is given by

$$\Pi\mu=\int_{\mathbb{T}^{m_2}} d\omega\rho_g(\omega)\mu_\omega.$$

For every $i=1,\dots,d$, call

$$\overline{\rho}_i := \nu_g(I_i)=\int_{I_i} d\omega\rho_g(\omega)$$

the measure of I_i with respect to the invariant measure of g. Define the map $\boldsymbol{\Delta}:\mathcal{M}_{1,\nu_g}(\mathbb{T}^{m_1}\times\mathbb{T}^{m_2})\to(\mathcal{M}_1(\mathbb{T}^{m_2}))^d$ in the following way

$$(\boldsymbol{\Delta}\mu)_i := \overline{\rho}_i^{-1}\int_{I_i} d\omega\rho_g(\omega)\mu_\omega,$$

that is, $(\boldsymbol{\Delta}\mu)_i$ is the average of the disintegration $\{\mu_\omega\}_{\omega\in\mathbb{T}^{m_1}}$ on I_i with respect to the invariant measure of g. The map $\boldsymbol{\Delta}$ gives a decomposition of μ which can be viewed as a *coarse-graining* of the disintegration of $\{\mu_\omega\}_{\omega\in\mathbb{T}^{m_1}}$. Moreover, for any $\mu\in\mathcal{M}_{1,\nu_g}(\mathbb{T}^{m_1}\times\mathbb{T}^{m_2})$

$$\Pi\mu=\sum_{i=1}^d\overline{\rho}_i(\boldsymbol{\Delta}\mu)_i,$$

therefore, by keeping track of $\boldsymbol{\Delta}(F_*^n\mu)$, we can keep track of $\Pi F_*^n\mu$.

Consider also

$$\mathcal{F}_i := \overline{\rho}_i^{\,-1}\int_{I_i} d\omega\rho_g(\omega)f_{\omega *}, \tag{4.4}$$

which is the average of the operators $\{f_{\omega *}\}_{\omega\in\mathbb{T}^{m_1}}$ on I_i w.r.t. ν_g restricted to I_i and normalized. A lemma below shows that $\mathcal{F}_i$ is an approximation of $f_{\omega *}$ for $\omega\in I_i$. The smaller is the size of I_i, i.e. the larger is $\sigma>0$, the better is the approximation.

For every $1\leqslant i,j\leqslant d$, define the operators

$$\mathcal{L}_{ij} := \overline{\rho}_i^{\,-1}\left(\int_{I_i} d\omega\frac{\rho_g(\omega_j)}{|\mathrm{D}g_{\omega_j}|}\right)\mathcal{F}_j; \tag{4.5}$$

and consider the operator $\boldsymbol{\mathcal{L}}:(\mathcal{M}_1(\mathbb{T}^{m_2}))^d\to(\mathcal{M}_1(\mathbb{T}^{m_2}))^d$

$$(\boldsymbol{\mathcal{L}}\boldsymbol{\mu})_i=\sum_{j=1}^d\mathcal{L}_{ij}(\boldsymbol{\mu})_j. \tag{4.6}$$

Remark 4.2. Before moving on, let us stress why the above mappings $\mathbf{\Delta}$ and $\boldsymbol{\mathcal{L}}$ are important: $\mathbf{\Delta}(F_*\mu)$ and $\boldsymbol{\mathcal{L}}(\mathbf{\Delta}\mu)$ are very close when the expansion of g is very large†. This will let us prove that for fixed n, we can approximate $\mathbf{\Delta}(F_*^n\mu)$ with $\boldsymbol{\mathcal{L}}^n(\mathbf{\Delta}\mu)$ when the expansion of g is sufficiently large, with the advantage that $\boldsymbol{\mathcal{L}}$ has good contraction properties (thanks to its relation with $\mathcal{P}$, see Remark 4.5).

Remark 4.3. Notice that both operators $\mathbf{\Delta}$ and $\boldsymbol{\mathcal{L}}$ depend on d, the number of branches of g, and therefore on σ. Since we ultimately want to let σ be large in order for the RHS of (4.7) to be small, we will have to keep σ and d explicit in all our estimates.

The remarks above is formalized in the following proposition. For $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in (\mathcal{M}_1(\mathbb{T}^{m_2}))^d$ we define

$$d_W(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) = \max_{i=1,\dots,d} d_W((\boldsymbol{\mu}_1)_i, (\boldsymbol{\mu}_2)_i).$$

Proposition 4.3. *If $\mu \in \Gamma_{\ell_0, a_0}$, with ℓ_0 and a_0 as in Corollary 4.1, then there is a constant $K_\# > 0$ uniform in σ, and there is $C_3 : (1, +\infty) \to \mathbb{R}^+$ decreasing such that*

$$d_W(\mathbf{\Delta}(F_*^n\mu), \boldsymbol{\mathcal{L}}^n(\mathbf{\Delta}\mu)) < K_\# \frac{L^{n+1} - L}{L - 1}\left(\ell_0\sigma^{-1} + C_3(\sigma)\|\rho_0 - \rho_g\|_\infty\right), \tag{4.7}$$

where ρ_0 is the density of the horizontal marginal of μ.

Proof. Let's call ν_0 the horizontal marginal of μ, and $\rho_0 \in \mathcal{V}_{a_0}$ its density. Let's denote by $\nu_n := F_*^n\nu_0$ and by $\rho_n := \frac{d\nu_n}{d\,\mathrm{Leb}}$. By Lemma 4.1, $\rho_n \in \mathcal{V}_{a_0}$ for every $n \in \mathbb{N}_0$.

First, let's prove (4.7) for $n = 1$. Recalling the disintegration (4.1),

$$\begin{aligned}
(\mathbf{\Delta}(F_*\mu))_i &= \bar{\rho}_i^{-1} \int_{I_i} d\omega \rho_g(\omega)(F_*\mu)_\omega \\
&= \bar{\rho}_i^{-1} \int_{I_i} d\omega \frac{\rho_g(\omega)}{\rho_1(\omega)} \sum_{j=1}^d \frac{\rho_0(\omega_j)}{|\mathrm{D}g_{\omega_j}|} f_{\omega_j *}\mu_{\omega_j} \\
&= \bar{\rho}_i^{-1} \int_{I_i} d\omega \left(\frac{\rho_g(\omega)}{\rho_1(\omega)} - 1\right) \sum_{j=1}^d \frac{\rho_0(\omega_j)}{|\mathrm{D}g_{\omega_j}|} f_{\omega_j *}\mu_{\omega_j} && (4.8)\\
&\quad + \bar{\rho}_i^{-1} \int_{I_i} d\omega \sum_{j=1}^d \frac{\rho_0(\omega_j) - \rho_g(\omega_j)}{|\mathrm{D}g_{\omega_j}|} f_{\omega_j *}\mu_{\omega_j} && (4.9)\\
&\quad + \bar{\rho}_i^{-1} \int_{I_i} d\omega \sum_{j=1}^d \frac{\rho_g(\omega_j)}{|\mathrm{D}g_{\omega_j}|} f_{\omega_j *}\mu_{\omega_j}, && (4.10)
\end{aligned}$$

where in the last equality we added and subtracted the same terms. We denote by A the term in (4.8) and by B the term in (4.9). Call $\mathrm{Lip}_0^1(\mathbb{T}^{m_2}; \mathbb{R})$, Lip_0^1 for brevity, the set of Lipschitz functions

† One can think of $\boldsymbol{\mathcal{L}}$ as an approximation for the action of F_* on measures μ having constant disintegration on the intervals I_i.

from $\mathbb{T}^{m_2}$ to $\mathbb{R}$ with zero integral. When computing $\|\cdot\|_W$, taking the supremum over Lip^1 or Lip^1_0 the same, as the integrals of φ and that of $\varphi - \int \varphi$ are the same. This has the advantage that for $\varphi \in \mathrm{Lip}^1_0$, $|\varphi| \leqslant C_2$, where C_2 is the diameter of $\mathbb{T}^{m_2}$.

$$d_W((\mathbf{\Delta}(F_*\mu))_i, (\boldsymbol{\mathcal{L}}\mathbf{\Delta}\mu)_i) = \sup_{\varphi\in\mathrm{Lip}^1_0} \int \varphi d\Bigg(A + B + \sum_{j=1}^{d} \overline{\rho}_i^{\,-1} \int_{I_i} d\omega \frac{\rho_g(\omega_j)}{|\mathrm{D}g_{\omega_j}|}(f_{\omega_j *} - \mathcal{F}_j)\mu_{\omega_j} \\ + \sum_{j=1}^{d} \overline{\rho}_i^{\,-1} \int_{I_i} d\omega \frac{\rho_g(\omega_j)}{|\mathrm{D}g_{\omega_j}|}\mathcal{F}_j\Big(\mu_{\omega_j} - (\mathbf{\Delta}\mu)_j\Big)\Bigg). \tag{4.11}$$

Let us call

$$\delta_n := \|\rho_n - \rho_g\|_\infty = \sup_{\omega\in\mathbb{T}^{m_1}} |\rho_n(\omega) - \rho_g(\omega)|.$$

Since $\rho_1 \in \mathcal{V}_a$, $|\rho_1| \geqslant e^{-C_1 D}$ where $C_1 > 0$ is the diameter of $\mathbb{T}^{m_1}$. Therefore

$$\left|1 - \frac{\rho_g(\omega)}{\rho_1(\omega)}\right| \leqslant \frac{1}{\rho_1(\omega)}|\rho_g(\omega) - \rho_1(\omega)| \leqslant e^{C_1 D}\delta_1.$$

Now we distribute the sup among the four terms on the RHS of (4.11), and estimate each of them separately.

$$\begin{aligned}
\sup_{\varphi\in\mathrm{Lip}^1_0} \int_{\mathbb{T}^{m_2}} \varphi dA &\leqslant \overline{\rho}_i^{\,-1} \int_{I_i} d\omega \sum_{j=1}^{d} \frac{\rho_0(\omega_j)}{|\mathrm{D}g_{\omega_j}|} \sup_{\varphi\in\mathrm{Lip}^1_0} \left| \int_{\mathbb{T}^{m_2}} \varphi(x) d\left[\left(\frac{\rho_g(\omega)}{\rho_1(\omega)} - 1\right) f_{\omega_j *}\mu_{\omega_j}\right](x)\right| \\
&\leqslant \overline{\rho}_i^{\,-1} \int_{I_i} d\omega \sum_{j=1}^{d} \frac{\rho_0(\omega_j)}{|\mathrm{D}g_{\omega_j}|} C_2 e^{C_1 D}\delta_1 \\
&= C_2 e^{C_1 D}\delta_1 \overline{\rho}_i^{\,-1} \int_{I_i} d\omega \rho_1(\omega) \\
&= C_2 e^{C_1 D}\delta_1 \frac{\nu_1(I_i)}{\nu_g(I_i)} \\
&\leqslant C_2 e^{3C_1 D}\delta_1.
\end{aligned}$$

$$\begin{aligned}
\sup_{\varphi\in\mathrm{Lip}^1_0} \int \varphi dB &= \sup_{\varphi\in\mathrm{Lip}^1_0} \int_{\mathbb{T}^{m_2}} \varphi(x) d\left[\overline{\rho}_i^{\,-1} \int_{I_i} d\omega \sum_{j=1}^{d} \frac{\rho_0(\omega_j) - \rho_g(\omega_j)}{|\mathrm{D}g_{\omega_j}|} f_{\omega_j *}\mu_{\omega_j}\right](x) \\
&\leqslant \overline{\rho}_i^{\,-1} \int_{I_i} d\omega \sum_{j=1}^{d} \frac{1}{|\mathrm{D}g_{\omega_j}|} \sup_{\varphi\in\mathrm{Lip}^1_0} \left| \int \varphi(x)(\rho_0(\omega_j) - \rho_g(\omega_j)) d f_{\omega_j *}\mu_{\omega_j}(x)\right| \\
&\leqslant \overline{\rho}_i^{\,-1} \int_{I_i} d\omega \sum_{j=1}^{d} \frac{1}{|\mathrm{D}g_{\omega_j}|} C_2\delta_0
\end{aligned}$$

$$= \frac{g_* \operatorname{Leb}_{\mathbb{T}^{m_1}}(I_i)}{\nu_g(I_i)} C_2\delta_0$$
$$\leqslant e^{2C_1 D} C_2 \delta_0$$

For the third term in the big parenthesis of Equation (4.11), using the definition of $\mathcal{F}_j$

$$\begin{aligned}
&\left| \int \varphi(x) \sum_{j=1}^{d} \overline{\rho}_i^{\,-1} \int_{I_i} d\omega \frac{\rho_g(\omega_j)}{|\mathrm{D}g_{\omega_j}|} d(f_{\omega_j *} - \mathcal{F}_j)\mu_{\omega_j}(x) \right| \\
&= \left| \sum_{j=1}^{d} \int_{I_j} d\omega' \rho_g(\omega') \overline{\rho}_j^{\,-1} \overline{\rho}_i^{\,-1} \int_{I_i} d\omega \frac{\rho_g(\omega_j)}{|\mathrm{D}g_{\omega_j}|} \int \varphi(x) d(f_{\omega_j *} - f_{\omega' *})\mu_{\omega_j}(x) \right| \\
&= \left| \sum_{j=1}^{d} \int_{I_j} d\omega' \rho_g(\omega') \overline{\rho}_j^{\,-1} \overline{\rho}_i^{\,-1} \int_{I_i} d\omega \frac{\rho_g(\omega_j)}{|\mathrm{D}g_{\omega_j}|} \int (\varphi \circ f_{\omega_j}(x) - \varphi \circ f_{\omega'}(x)) d\mu_{\omega_j}(x) \right| \\
&\leqslant \sum_{j=1}^{d} \int_{I_j} d\omega' \rho_g(\omega') \overline{\rho}_j^{\,-1} \overline{\rho}_i^{\,-1} \int_{I_i} d\omega \frac{\rho_g(\omega_j)}{|\mathrm{D}g_{\omega_j}|} \int |\varphi \circ f_{\omega_j}(x) - \varphi \circ f_{\omega'}(x)| d\mu_{\omega_j}(x) \\
&\leqslant \sum_{j=1}^{d} \int_{I_j} d\omega' \rho_g(\omega') \overline{\rho}_j^{\,-1} \overline{\rho}_i^{\,-1} \int_{I_i} d\omega \frac{\rho_g(\omega_j)}{|\mathrm{D}g_{\omega_j}|} L \operatorname{diam}(I_j) \\
&\leqslant L\sigma^{-1} C_1 \sum_{j=1}^{d} \overline{\rho}_j^{\,-1} \nu_g(I_j) \overline{\rho}_i^{\,-1} \int_{I_i} d\omega \frac{\rho_g(\omega_j)}{|\mathrm{D}g_{\omega_j}|} \\
&\leqslant L\sigma^{-1} C_1 \sum_{j=1}^{d} \overline{\rho}_i^{\,-1} \int_{I_i} d\omega \frac{\rho_g(\omega_j)}{|\mathrm{D}g_{\omega_j}|} \\
&\leqslant L\sigma^{-1} C_1,
\end{aligned}$$

where we used that $\overline{\rho}_j = \nu_g(I_j)$; that

$$|\varphi \circ f_{\omega_j}(x) - \varphi \circ f_{\omega'}(x)| \leqslant |f_{\omega_j}(x) - f_{\omega'}(x)| \leqslant L|\omega_j - \omega'| \leqslant L \operatorname{diam}(I_j) \leqslant L\sigma^{-1} C_1,$$

recall that L is the Lipschitz constant of f and $C_1 = \operatorname{diam}(\mathbb{T}^{m_1})$; and that

$$\overline{\rho}_i^{\,-1} \int_{I_i} d\omega \sum_{j=1}^{d} \frac{\rho_g(\omega_j)}{|\mathrm{D}g_{\omega_j}|} = \overline{\rho}_i^{\,-1} \int_{I_i} d\omega \rho_g(\omega) = 1. \tag{4.12}$$

For the last term on the RHS of Equation (4.11),

$$\int \varphi(x) d\left(\sum_{j=1}^{d} \overline{\rho}_i^{\,-1} \int_{I_i} d\omega \frac{\rho_g(\omega_j)}{|\mathrm{D}g_{\omega_j}|} \mathcal{F}_j(\mu_{\omega_j} - (\boldsymbol{\Delta}\mu)_j) \right)(x)$$

$$= \sum_{j=1}^{d} \overline{\rho}_i^{-1} \int_{I_i} d\omega \frac{\rho_g(\omega_j)}{|\mathrm{D}g_{\omega_j}|} \int \varphi(x) d\mathcal{F}_j(\mu_{\omega_j} - (\boldsymbol{\Delta}\mu)_j)(x)$$

$$\leqslant \sum_{j=1}^{d} \overline{\rho}_i^{-1} \int_{I_i} d\omega \frac{\rho_g(\omega_j)}{|\mathrm{D}g_{\omega_j}|} \mathrm{Lip}(\mathcal{F}_j) d_W(\mu_{\omega_j}, (\boldsymbol{\Delta}\mu)_j) \tag{4.13}$$

$$\leqslant L\ell_0\sigma^{-1}C_1, \tag{4.14}$$

where in the last step we used that $\mathrm{Lip}(\mathcal{F}_j) \leqslant \sup_\omega \mathrm{Lip}(f_{\omega *}) \leqslant L$ and that

$$d_W(\mu_{\omega_j}, (\boldsymbol{\Delta}\mu)_j) \leqslant \ell_0 \operatorname{diam}(I_j) \leqslant \ell_0\sigma^{-1}C_1.$$

Putting all of the above together we conclude that there is $K_\# > 0$ (independent of $\sigma > 0$) such that

$$d_W((\boldsymbol{\Delta}(F_*\mu))_i, (\boldsymbol{\mathcal{L}}\boldsymbol{\Delta}\mu)_i) \leqslant K_\#(\sigma^{-1} + \delta_0 + \delta_1).$$

Now, since for every $k \in \mathbb{N}$, $F_*^k\mu \in \mathcal{M}_{1,\nu_k}(\mathbb{T}^{m_1} \times \mathbb{T}^{m_1}) \cap \Gamma_{\ell_0}$, by repeated applications of the triangle inequality

$$\begin{aligned}
d_W(\boldsymbol{\Delta}F_*^n\mu, \boldsymbol{\mathcal{L}}^n\boldsymbol{\Delta}\mu) &\leqslant \sum_{k=0}^{n-1} d_W(\boldsymbol{\mathcal{L}}^{n-k-1}\boldsymbol{\Delta}(F_*F_*^k\mu), \boldsymbol{\mathcal{L}}^{n-k}\boldsymbol{\Delta}(F_*^k\mu)) \\
&\leqslant \sum_{k=0}^{n-1} \left(\sup_\omega \mathrm{Lip}(f_{\omega *})\right)^{n-k-1} d_W(\boldsymbol{\Delta}(F_*F_*^k\mu), \boldsymbol{\mathcal{L}}\boldsymbol{\Delta}(F_*^k\mu)) \\
&\leqslant K_\# \sum_{k=0}^{n-1} L^{n-k}\left[\ell_0\sigma^{-1} + \delta_k + \delta_{k+1}\right] \\
&\leqslant K_\# \left[\ell_0\sigma^{-1} + \sum_{k=0}^{n-1} \delta_k + \delta_{k+1}\right] \sum_{k=0}^{n-1} L^{n-k} \\
&\leqslant K_\#(\ell_0\sigma^{-1} + C_3\delta_0) \sum_{k=0}^{n-1} L^{n-k},
\end{aligned}$$

where we used that by Lemma 4.1, $\delta_k \leqslant C_g\lambda_g^k\delta_0$, and $C_3 := \frac{2C_g}{1-\lambda_g}$. □

The operator $\boldsymbol{\mathcal{L}}$ has good spectral properties. To prove it, we are going to need the lemma below.

Lemma 4.2. *For every $i, j = 1, \dots, d$,*

$$\overline{\rho}_i^{-1}\overline{\rho}_j^{-1} \int_{I_i} d\omega \frac{\rho_g(\omega_j)}{|\mathrm{D}g_{\omega_j}|} > p,$$

with

$$p := e^{-C_1 D\left[\frac{3}{1-\sigma^{-1}}+1\right]} \in (0,1),$$

where D is the bound on the distortion of the map g, and C_1 is the diameter of $\mathbb{T}^{m_1}$.

Proof. Recall that ω_j is shorthand notation for $h_j(\omega)$. Since

$$\int_{\mathbb{T}^{m_1}} d\omega \frac{1}{|\mathrm{D}g_{\omega_j}|} = \mathrm{Leb}_{\mathbb{T}^{m_1}}(I_j) =: |I_j|,$$

and $|\mathrm{D}g \circ h_j|$ is continuous, there is ω_0 such that

$$\frac{1}{|\mathrm{D}g_{h_j(\omega_0)}|} = |I_j|.$$

Recalling the notation $\omega_j = h_j(\omega)$, the bound on the distortion (H0.2) gives

$$|I_j|^{-1}|\mathrm{D}g_{\omega_j}|^{-1} = \frac{|\mathrm{D}g_{h_j(\omega_0)}|}{|\mathrm{D}g_{h_j(\omega)}|} \geqslant e^{-D|\omega-\omega_0|} \geqslant e^{-DC_1},$$

where C_1 equals the diameter of $\mathbb{T}^{m_1}$ w.r.t. the Euclidean distance.

Also, recall that $\rho_g \in \mathcal{V}_{a_0}$ with $a_0 = \frac{D}{1-\sigma^{-1}}$, therefore $e^{-C_1 a_0} \leqslant \rho_g \leqslant e^{C_1 a_0}$ and $\overline{\rho}_i \leqslant e^{C_1 a_0}|I_i|$. Putting the above considerations together

$$\begin{aligned}
\overline{\rho}_i^{\,-1}\overline{\rho}_j^{\,-1} \int_{I_i} d\omega \frac{\rho_g(\omega_j)}{|\mathrm{D}g_{\omega_j}|} &\geqslant e^{-C_1 a_0}|I_j|^{-1} e^{-C_1 a_0}|I_i|^{-1} \int_{I_i} d\omega \frac{e^{-C_1 a_0}}{|\mathrm{D}g_{\omega_j}|} \\
&\geqslant e^{-3C_1 a_0}|I_i|^{-1} \int_{I_i} d\omega e^{-DC_1} \\
&\geqslant e^{-3C_1 a_0 - DC_1}.
\end{aligned}$$

□

Remark 4.4. Notice that p depends on σ, but for D fixed, p increases with $\sigma > 1$. In particular, assuming that $\sigma \geqslant \sigma_0 > 1$, we get the bound

$$p \geqslant e^{-C_1 D\left[1+\frac{3}{1-\sigma_0^{-1}}\right]},$$

which is independent of σ.

Lemma 4.2 implies that $\boldsymbol{\mathcal{L}}$ can be decomposed in the following way: there are $p_{ij} \geqslant 0$ such that

$$(\boldsymbol{\mathcal{L}})_{ij} = \mathcal{L}_{ij} = p\int_{I_j} d\omega \rho_g(\omega) f_{\omega *} + p_{ij}\mathcal{F}_j. \tag{4.15}$$

Define $\mathcal{L}_1 : (\mathcal{M}_1(\mathbb{T}^{m_2}))^d \to (\mathcal{M}_1(\mathbb{T}^{m_2}))^d$ as

$$(\mathcal{L}_1)_{ij} = \int_{I_j} d\omega \rho_g(\omega) f_{\omega *},$$

and $\mathcal{L}_2 : (\mathcal{M}_1(\mathbb{T}^{m_2}))^d \to (\mathcal{M}_1(\mathbb{T}^{m_2}))^d$ as

$$(\mathcal{L}_2)_{ij} := p_{ij} \mathcal{F}_j,$$

so that $\mathcal{L} = p\mathcal{L}_1 + \mathcal{L}_2$.

Remark 4.5. Notice that $(\mathcal{L}_{\mathbf{1}})_{ij}$ does not depend on i, and summing over j

$$\sum_j (\mathcal{L}_{\mathbf{1}})_{ij} = \mathcal{P}.$$

When $\mathcal{L}$ acts on a vector of measures $\boldsymbol{\nu} = \nu(1, \dots, 1)$ having all identical entries, one has

$$(\mathcal{L}\boldsymbol{\nu})_i = p\mathcal{P}\nu + \sum_j (\mathcal{L}_2)_{ij} \nu.$$

The spectral properties of $\mathcal{P}$ and the lower bound on p found in Lemma 4.2, will be used in a coupling argument, Proposition 4.4 below, that ultimately yields the spectral properties of $\mathcal{L}$.

In a proposition below we show that the operator $\mathcal{L}$ has good contracting properties with respect to d_{TV}. First we state a couple of lemmas.

Lemma 4.3. *For every* $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in (\mathcal{M}_1(\mathbb{T}))^d$

$$d_{TV}(\mathcal{L}\boldsymbol{\mu}_1, \mathcal{L}\boldsymbol{\mu}_2) \leqslant d_{TV}(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2). \tag{4.16}$$

Proof. By definition of Total Variation distance, transfer operators are weak contractions with respect to d_{TV}; in particular, for any $\eta_1, \eta_2 \in \mathcal{M}_1(\mathbb{T})$ and any $\omega \in \mathbb{T}^{m_1}$

$$d_{TV}(f_{\omega *}\eta_1, f_{\omega *}\eta_2) \leqslant d_{TV}(\eta_1, \eta_2),$$

and therefore

$$d_{TV}(\mathcal{F}_j \eta_1, \mathcal{F}_j \eta_2) \leqslant d_{TV}(\eta_1, \eta_2) \tag{4.17}$$

for any j.

By formula 4.12 one gets that for every i

$$\begin{aligned} & d_{TV}((\mathcal{L}\boldsymbol{\mu}_1)_i, (\mathcal{L}\boldsymbol{\mu}_2)_i) \\ & = d_{TV}\left(\sum_j \overline{\rho}_i^{\,-1} \left(\int_{I_i} d\omega \frac{\rho_g(\omega_j)}{|\mathrm{D}g_{\omega_j}|} \right) \mathcal{F}_j(\boldsymbol{\mu}_1)_j, \sum_j \overline{\rho}_i^{\,-1} \left(\int_{I_i} d\omega \frac{\rho_g(\omega_j)}{|\mathrm{D}g_{\omega_j}|} \right) \mathcal{F}_j(\boldsymbol{\mu}_2)_j \right) \end{aligned}$$

$$\leqslant \sum_j \overline{\rho}_i^{-1}\left(\int_{I_i} d\omega \frac{\rho_g(\omega_j)}{|\mathrm{D}g_{\omega_j}|}\right) d_{TV}(\mathcal{F}_j(\boldsymbol{\mu}_1)_j, \mathcal{F}_j(\boldsymbol{\mu}_2)_j)$$

$$\leqslant \sum_j \overline{\rho}_i^{-1}\left(\int_{I_i} d\omega \frac{\rho_g(\omega_j)}{|\mathrm{D}g_{\omega_j}|}\right) d_{TV}((\boldsymbol{\mu}_1)_j, (\boldsymbol{\mu}_2)_j)$$

$$\leqslant d_{TV}(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2). \qquad \square$$

Lemma 4.4. *For every $n \in \mathbb{N}$, the following decomposition holds*

$$\boldsymbol{\mathcal{L}}^n \boldsymbol{\mu} := p^n \boldsymbol{\mathcal{L}}_1^n \boldsymbol{\mu} + (1 - p^n)\boldsymbol{R}_n \boldsymbol{\mu}.$$

where $\boldsymbol{R}_n : (\mathcal{M}_1(\mathbb{T}^{m_2}))^d \to (\mathcal{M}_1(\mathbb{T}^{m_2}))^d$ is such that

$$d_{TV}(\boldsymbol{R}_n\boldsymbol{\mu}_1, \boldsymbol{R}_n\boldsymbol{\mu}_2) \leqslant d_{TV}(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2),$$

for all $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in (\mathcal{M}_1(\mathbb{T}^{m_2}))^d$.

Proof. Let's start noticing that, by definition, $\sum_j p_{ij} = (1 - p)$ for every i, in fact comparing equations (4.15) and (4.5) follows that

$$p\overline{\rho}_i + p_{ij} = \overline{\rho}_j^{-1}\left(\int_{I_i} d\omega \frac{\rho_g(\omega_j)}{|\mathrm{D}g_{\omega_j}|}\right),$$

and

$$\sum_j (p\overline{\rho}_j + p_{ij}) = p + \sum_j p_{ij} = \sum_j \overline{\rho}_i^{-1}\left(\int_{I_i} d\omega \frac{\rho_g(\omega_j)}{|\mathrm{D}g_{\omega_j}|}\right) = 1.$$

Now we prove the statement of the lemma by induction on $n \in \mathbb{N}$. For $n = 1$, $\boldsymbol{R}_1 = (1 - p)^{-1}\boldsymbol{\mathcal{L}}_2$ and recalling (4.17)

$$d_{TV}((1 - p)^{-1}\boldsymbol{\mathcal{L}}_2\boldsymbol{\mu}_1, (1 - p)^{-1}\boldsymbol{\mathcal{L}}_2\boldsymbol{\mu}_2)$$

$$= \max_i d_{TV}\left(\sum_j (1 - p)^{-1} p_{ij}\mathcal{F}_j(\boldsymbol{\mu}_1)_j, \sum_j (1 - p)^{-1} p_{ij}\mathcal{F}_j(\boldsymbol{\mu}_2)_j\right)$$

$$\leqslant \max_i \sum_j (1 - p)^{-1} p_{ij} d_{TV}((\boldsymbol{\mu}_1)_j, (\boldsymbol{\mu}_2)_j)$$

$$\leqslant d_{TV}(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2).$$

Now assume that the statement is true for $n - 1$.

$$\boldsymbol{\mathcal{L}}^n = \boldsymbol{\mathcal{L}}\boldsymbol{\mathcal{L}}^{n-1} = p^n \boldsymbol{\mathcal{L}}_1{}^n + p^{n-1}(1 - p)\boldsymbol{R}_1\boldsymbol{\mathcal{L}}_1{}^{n-1} + (1 - p^{n-1})\boldsymbol{\mathcal{L}}\boldsymbol{R}_{n-1}.$$

Define

$$\boldsymbol{R}_n := \frac{(1-p)p^{n-1}\boldsymbol{R}_1\boldsymbol{\mathcal{L}}_1{}^{n-1} + (1-p^{n-1})\boldsymbol{\mathcal{L}}\boldsymbol{R}_{n-1}}{1-p^n}$$

and by Lemma 4.3 applied to $\boldsymbol{\mathcal{L}}$ and $\boldsymbol{\mathcal{L}_1}$, by the inductive step – $(1-p)p^{n-1} + (1-p^{n-1}) = 1 - p^n$– and by Lemma C.4

$$d_{TV}(\boldsymbol{R}_n\boldsymbol{\mu}_1, \boldsymbol{R}_n\boldsymbol{\mu}_2) \leqslant d_{TV}(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2). \qquad \square$$

We are now ready to show that $\boldsymbol{\mathcal{L}}$ has good contraction properties with respect to the Total Variation distance. The proof uses a coupling argument.

Proposition 4.4. *There are* $C_{\mathcal{L}} > 0$ *and* $\lambda_{\mathcal{L}} \in (0,1)$ *such that for any* $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in (\mathcal{M}_1(\mathbb{T}^{m_2}))^d$

$$d_{TV}(\boldsymbol{\mathcal{L}}^n\boldsymbol{\mu}_1, \boldsymbol{\mathcal{L}}^n\boldsymbol{\mu}_2) \leqslant C_{\mathcal{L}}\lambda_{\mathcal{L}}^n d_{TV}(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2). \tag{4.18}$$

Proof. Notice that all the rows of the operator $\boldsymbol{\mathcal{L}}_1$ are equal, therefore, for any $\boldsymbol{\mu} \in (\mathcal{M}_1(\mathbb{T}^{m_2}))^d$, also all the components of $\boldsymbol{\mathcal{L}}_1\boldsymbol{\mu}$ are equal, that is, there is $\mu' \in \mathcal{M}_1(\mathbb{T}^{m_2})$ such that $(\boldsymbol{\mathcal{L}}_1\boldsymbol{\mu})_i = \mu'$. By definition of $\boldsymbol{\mathcal{L}}_1$ follows that

$$(\boldsymbol{\mathcal{L}}_1^2\boldsymbol{\mu})_i = \sum_{j=1}^d (\boldsymbol{\mathcal{L}}_1)_{ij}\mu' = \sum_{j=1}^d \int_{I_j} d\omega\rho_j(\omega) f_{\omega *}\mu' = \mathcal{P}\mu',$$

and by induction

$$(\boldsymbol{\mathcal{L}}_1^n\boldsymbol{\mu})_i = \mathcal{P}^{n-1}\mu', \tag{4.19}$$

for every i and $n > 1$.

Pick $n_0 > 1$ such that $C\lambda^{n_0-1} \leqslant \frac{1}{2}$. Then it follows from (4.19) and Assumption (H) that for all $n > 1$

$$d_{TV}(\boldsymbol{\mathcal{L}}_1^n\boldsymbol{\mu}_1, \boldsymbol{\mathcal{L}}_1^n\boldsymbol{\mu}_2) \leqslant \frac{1}{2} d_{TV}(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2).$$

which implies

$$\begin{aligned} d_{TV}(\boldsymbol{\mathcal{L}}^{n_0}\boldsymbol{\mu}_1, \boldsymbol{\mathcal{L}}^{n_0}\boldsymbol{\mu}_2) &\leqslant p^{n_0} d_{TV}(\boldsymbol{\mathcal{L}}_1^{n_0}\boldsymbol{\mu}_1, \boldsymbol{\mathcal{L}}_1^{n_0}\boldsymbol{\mu}_2) + (1-p^{n_0}) d_{TV}(R_{n_0}\boldsymbol{\mu}_1, R_{n_0}\boldsymbol{\mu}_2) \\ &\leqslant \left(1 - \frac{1}{2}p^{n_0}\right) d_{TV}(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2). \end{aligned}$$

Define $\lambda_{\mathcal{L}} := (1 - \frac{1}{2}p^{n_0})^{\frac{1}{n_0}}$ and $C_{\mathcal{L}} := \lambda_{\mathcal{L}}^{-n_0}$. For every $n \in \mathbb{N}$ there are $k \in \mathbb{N}$ and $0 \leqslant r < n_0$ such that $n = kn_0 + r$, and by Lemma 4.3

$$\begin{aligned} d_{TV}(\boldsymbol{\mathcal{L}}^n\boldsymbol{\mu}_1, \boldsymbol{\mathcal{L}}^n\boldsymbol{\mu}_2) &= d_{TV}(\boldsymbol{\mathcal{L}}^r\mathcal{L}^{kn_0}\boldsymbol{\mu}_1, \boldsymbol{\mathcal{L}}^r\mathcal{L}^{kn_0}\boldsymbol{\mu}_2) \\ &\leqslant d_{TV}(\boldsymbol{\mathcal{L}}^{kn_0}\boldsymbol{\mu}_1, \boldsymbol{\mathcal{L}}^{kn_0}\boldsymbol{\mu}_2) \end{aligned}$$

$$
\begin{aligned}
&\leqslant \lambda_{\mathcal{L}}^{n_0 k} d_{TV}(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) \\
&\leqslant C_{\mathcal{L}} \lambda_{\mathcal{L}}^{n} d_{TV}(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2). \qquad \square
\end{aligned}
$$

The contraction properties of $\mathcal{L}$, (4.18), and the weak*-compactness of $(\mathcal{M}_1(\mathbb{T}^{m_2}))^d$ imply the existence of $\boldsymbol{\eta}_0 \in (\mathcal{M}_1(\mathbb{T}^{m_2}))^d$ such that $\mathcal{L}\boldsymbol{\eta}_0 = \boldsymbol{\eta}_0$.

The following proposition is the analogous of Proposition 3.3 in the case without distortion and proves approximated memory loss for the vertical marginals under application of F_*. In this case, there is an extra difficulty as, in order to prove Theorem 2.2, the class of probability measures we start from should include those having horizontal marginal equal to $\mathrm{Leb}_{\mathbb{T}^{m_1}}$ which in general can be different from the invariant measure ν_g.

Proposition 4.5 (Approximate Memory Loss). *Fix $D, L > 0$. There is $C''_{\mathcal{L}} > 0$ such that given any $\varepsilon > 0$, there is $\sigma_0 > L$ such that for any $\sigma > \sigma_0$, and any F satisfying (H0.1)-(H0.3) and Assumption (H)*

i)

$$
d_W(\Pi F_*^t \mu_1, \Pi F_*^t \mu_2) \leqslant C''_{\mathcal{L}} \lambda_{\mathcal{L}}^t + \varepsilon, \qquad \forall t \in \mathbb{N}
$$

for any $\mu_1, \mu_2 \in \Gamma_{\ell_0, a_0}$, ℓ_0 defined in (4.3);

ii)

$$
d_W\left(\Pi F_*^t \mu, \sum_{i=1}^{d} \overline{\rho}_i (\boldsymbol{\eta}_0)_i \right) \leqslant C''_{\mathcal{L}} \lambda_{\mathcal{L}}^t + \varepsilon, \qquad \forall t \in \mathbb{N}
$$

for any $\mu \in \Gamma_{a_0, \ell_0}$, ℓ_0 defined in (4.3).

Proof. Pick any two probability measures $\mu_1, \mu_2 \in \Gamma_{\ell_0}$. Then for $n, m \in \mathbb{N}$

$$
\begin{aligned}
d_W(\boldsymbol{\Delta} F_*^n(F_*^m \mu_1), \boldsymbol{\Delta} F_*^n(F_*^m \mu_2)) &\leqslant d_W(\mathcal{L}^n \boldsymbol{\Delta} F_*^m \mu_1, \mathcal{L}^n \boldsymbol{\Delta} F_*^m \mu_2) + \\
&\quad + d_W(\mathcal{L}^n \boldsymbol{\Delta} F_*^m \mu_1, \boldsymbol{\Delta} F_*^n F_*^m \mu_1) + \\
&\quad + d_W(\mathcal{L}^n \boldsymbol{\Delta} F_*^m \mu_2, \boldsymbol{\Delta} F_*^n F_*^m \mu_2) \\
&\leqslant C_2 d_{TV}(\mathcal{L}^n \boldsymbol{\Delta} F_*^m \mu_1, \mathcal{L}^n \boldsymbol{\Delta} F_*^m \mu_2) + \\
&\qquad + 2K_{\#} L^{n+1} (\sigma^{-1} + C_3(\sigma) \|\rho_m - \rho_g\|_\infty) \\
&\leqslant C_2 C_{\mathcal{L}} \lambda_{\mathcal{L}}^n + 2K_{\#} L^{n+1} (\sigma^{-1} + C_3(\sigma) \|\rho_m - \rho_g\|_\infty).
\end{aligned}
$$

where we used triangle inequality, Lemma C.3 (recall that C_2 is the diameter of $\mathbb{T}^{m_2}$), and Proposition 4.3.

For every $\varepsilon > 0$, pick $n_0 \in \mathbb{N}$, $m_0 \in \mathbb{N}$, and σ_0 large enough so that $C_2 C_{\mathcal{L}} \lambda_{\mathcal{L}}^{n_0} \leqslant \varepsilon/2$, and

$$
2K_{\#} L^{n+1} (\sigma^{-1} + C_3(\sigma) \|\rho_{m_0} - \rho_g\|_\infty) \leqslant \frac{\varepsilon}{2}
$$

Notice that m_0 is a transient one waits for the horizontal marginal to get sufficiently close to ν_g while n_0 is the time one waits for $\boldsymbol{\mathcal{L}}$ to contract by the desired amount.

Calling $C'_{\boldsymbol{\mathcal{L}}} := C_2 C_{\boldsymbol{\mathcal{L}}} \lambda_{\boldsymbol{\mathcal{L}}}^{-m_0}$

$$d_W(\boldsymbol{\Delta} F_*^{n+m_0}\mu_1, \boldsymbol{\Delta} F_*^{n+m_0}\mu_2) \leqslant C_2 C_{\boldsymbol{\mathcal{L}}} \lambda_{\boldsymbol{\mathcal{L}}}^{n_0} + \varepsilon/2 \leqslant C'_{\boldsymbol{\mathcal{L}}} \lambda_{\boldsymbol{\mathcal{L}}}^{n+m_0} + \varepsilon.$$

for every n, in fact if $n \leqslant n_0$

$$d_W(\boldsymbol{\Delta} F_*^{n+m_0}\mu_1, \boldsymbol{\Delta} F_*^{n+m_0}\mu_2) \leqslant C'_{\boldsymbol{\mathcal{L}}} \lambda_{\boldsymbol{\mathcal{L}}}^{n} + \varepsilon/2$$

and if $n \geqslant n_0$

$$d_W(\boldsymbol{\Delta} F_*^{n+m_0}\mu_1, \boldsymbol{\Delta} F_*^{n+m_0}\mu_2) \leqslant C'_{\boldsymbol{\mathcal{L}}} \lambda_{\boldsymbol{\mathcal{L}}}^{n_0} + \varepsilon/2 \leqslant C'_{\boldsymbol{\mathcal{L}}} \lambda_{\boldsymbol{\mathcal{L}}}^{n} + \varepsilon.$$

Recall that

$$\Pi F_*^{n+m_0}\mu_j = \sum_{i=1}^{d} \overline{\rho}_i (\boldsymbol{\Delta} F_*^{n+m_0}\mu_j)_i$$

and since the above is a convex combination, using Lemma C.4

$$d_W(\Pi F_*^{n+m_0}\mu_1, \Pi F_*^{n+m_0}\mu_2) \leqslant d_W(\boldsymbol{\Delta} F_*^{n+m_0}\mu_1, \boldsymbol{\Delta} F_*^{n+m_0}\mu_2) \leqslant C'_{\boldsymbol{\mathcal{L}}} \lambda_{\boldsymbol{\mathcal{L}}}^{n+m_0} + \varepsilon$$

which proves point i) with $t \geqslant m_0$. If $t \leqslant m_0$, by the definition of d_W

$$d_W(\Pi F_*^{m}\mu_1, \Pi F_*^{m}\mu_2) \leqslant C_2$$

the diameter of $\mathbb{T}^{m_2}$. Therefore, picking

$$C''_{\boldsymbol{\mathcal{L}}} := \max \left\{ C'_{\boldsymbol{\mathcal{L}}},\ C_2 \lambda_{\boldsymbol{\mathcal{L}}}^{-m_0} \right\}$$

we get

$$d_W(\Pi F_*^{t}\mu_1, \Pi F_*^{t}\mu_2) \leqslant C''_{\boldsymbol{\mathcal{L}}} \lambda_{\boldsymbol{\mathcal{L}}}^{t} + \varepsilon$$

for all $t \in \mathbb{N}$ which concludes the proof of point i).

To prove point ii), recall that $\boldsymbol{\eta}_0 \in (\mathcal{M}_1(\mathbb{T}^{m_2}))^d$ is fixed by $\boldsymbol{\mathcal{L}}$. Now

$$\begin{aligned} d_W(\boldsymbol{\Delta} F_*^{n+m_0}\mu, \boldsymbol{\eta}_0) &\leqslant d_W(\boldsymbol{\Delta} F_*^{n+m_0}\mu, \boldsymbol{\mathcal{L}}^n \boldsymbol{\Delta} F_*^{m_0}\mu) + d_W(\boldsymbol{\mathcal{L}}^n \boldsymbol{\Delta} F_*^{m_0}\mu, \boldsymbol{\mathcal{L}}^n \boldsymbol{\eta}_0) \\ &\leqslant d_W(\boldsymbol{\Delta} F_*^{n+m_0}\mu, \boldsymbol{\mathcal{L}}^n \boldsymbol{\Delta} F_*^{m_0}\mu) + C_2 d_{TV}(\boldsymbol{\mathcal{L}}^n \boldsymbol{\Delta} F_*^{m_0}\mu, \boldsymbol{\mathcal{L}}^n \boldsymbol{\eta}_0) \\ &\leqslant K_\# L^{n+1}(\sigma^{-1} + C_3(\sigma)\|\rho_m - \rho_g\|_\infty) + C_2 C_{\boldsymbol{\mathcal{L}}} \lambda_{\boldsymbol{\mathcal{L}}}^{n}. \end{aligned}$$

In a way completely analogous to the proof of point i) one can show that for every $\varepsilon > 0$ there are σ_0 sufficiently large and $C''_{\boldsymbol{\mathcal{L}}} > 0$ such that for $\sigma > \sigma_0$

$$d_W(\boldsymbol{\Delta} F_*^{t}\mu, \boldsymbol{\eta}_0) \leqslant C''_{\boldsymbol{\mathcal{L}}} \lambda_{\boldsymbol{\mathcal{L}}}^{t} + \varepsilon.$$

By definition of d_W, the above means that $d_W((\boldsymbol{\Delta} F^t_* \mu)_i, (\boldsymbol{\eta}_0)_i) \leqslant C''_{\mathcal{L}} \lambda^t_{\mathcal{L}} + \varepsilon$ for all i, which implies that

$$d_W\left(\sum_i \overline{\rho}_i \, (\boldsymbol{\Delta} F^t_* \mu)_i, \sum_i \overline{\rho}_i \, (\boldsymbol{\eta}_0)_i \right) \leqslant C''_{\mathcal{L}} \lambda^t_{\mathcal{L}} + \varepsilon.$$

Since $\Pi F^n_* \mu = \sum_i \overline{\rho}_i \, (\boldsymbol{\Delta} F^n_* \mu)_i$, the statement follows. □

Proof of Theorem 2.1. The statement of the theorem is the content of point ii) of the proposition above. □

Proof of Theorem 2.2. With all the work above done, the proof of the theorem is almost identical to the case without distortion. The only difference is that instead of η_0 in the proof of the case without distortion, one has to substitute $\sum_i \overline{\rho}_i \, (\boldsymbol{\eta}_0)_i$, and apply Proposition 4.5 in place of Proposition 3.3. □

4.3 | Fixed point for $\mathcal{L}$ and fixed point for $\mathcal{P}$

Point ii) of Proposition 4.5 shows that if σ is sufficiently large, then the vertical marginal of $F^n_* \mu$ becomes close to $\overline{\eta} = \sum_{i=1}^d \overline{\rho}_i (\boldsymbol{\eta}_0)_i$. The purpose of this section is to remark that, in general, $\overline{\eta}$ is different (and possibly quite far) from η_0, the stationary measure of $\mathcal{P}$. We prove this fact in an indirect way by showing that the unique fixed point of $\mathcal{P} := \int_{\mathbb{T}} d\omega \rho_g(\omega) f_{\omega *}$, η_0, and the unique fixed point of $\mathcal{P}' := \int_{\mathbb{T}} d\omega \rho_{g^{k-1}}(\omega)(f_{g^{k-1}(\omega)} \circ \ldots \circ f_\omega)_*$, that we will call η'_0, can be in general very different for some $k > 1$. If this is the case, $\Pi F^{nk}_* \mathrm{Leb}_{\mathbb{T}^{m_1} \times \mathbb{T}^{m_2}}$ cannot become close to both η_0 and η'_0, and since $\mathcal{P}'$ is the random counterpart of F^k, it implies that $\sum_{i=1}^d \overline{\rho}_i (\boldsymbol{\eta}_0)_i$ can be far from the fixed points of $\mathcal{P}$ or/and $\mathcal{P}'$. At the end of the section we also give numerical evidence that $\overline{\eta}$ can be different from η_0 when the map g has nonzero distortion.

For simplicity of exposition, we are going to present an example that does not satisfy the smoothness requirements of Theorem 2.2. However, with a small modification on a set of arbitrarily small measure, the system can be made as smooth as one likes and all the considerations below carry over to the smoothed version.

First of all, we define the map $g := g_{M,\kappa} : \mathbb{T} \to \mathbb{T}$ where $M \in \mathbb{N}$ and $\kappa \in (0, 1)$ are parameters. We identify $\mathbb{T}$ with [0,1] in the usual way and divide [0,1] into $2M$ intervals of equal length

$$I_j := \left[\frac{j-1}{2M}, \frac{j}{2M} \right].$$

Let $\kappa' = 1 - \kappa$. Define for $0 \leqslant j \leqslant M - 1$

$$g_{M,\kappa}(\omega) := \begin{cases} \frac{M}{\kappa} \omega - \frac{j}{2\kappa} & \omega \in [j/2M, (j+\kappa)/2M] \\ \frac{M}{\kappa'} \omega - \frac{j+\kappa-\kappa'}{2\kappa'} & \omega \in [(j+\kappa)/2M, (j+1)/2M] \end{cases} \tag{4.20}$$

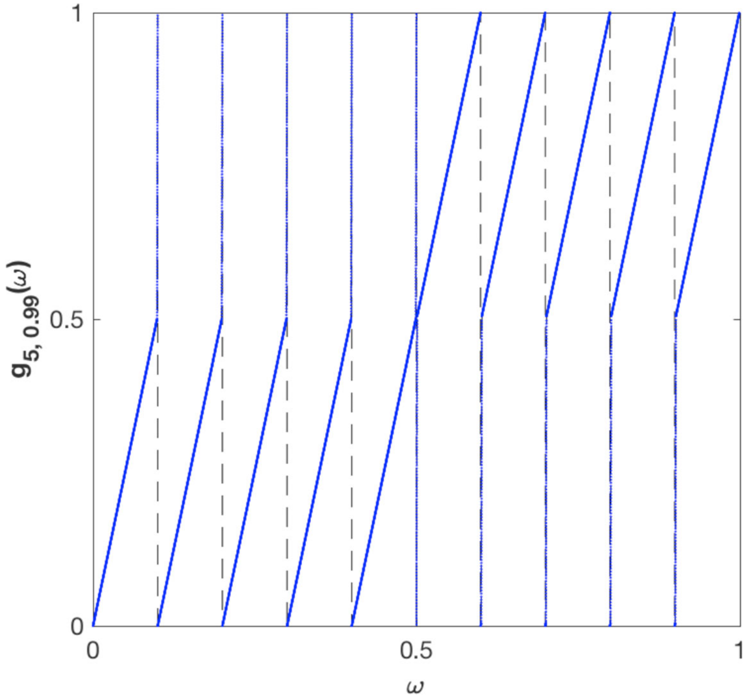

FIGURE 1 Graph of $g_{5,0.99}$

and for $M \leqslant j \leqslant 2M-1$

$$g_{M,\kappa}(\omega) := \begin{cases} \frac{M}{\kappa'}\omega - \frac{j}{2\kappa'} & \omega \in [j/2M, (j+\kappa')/2M] \\ \frac{M}{\kappa}\omega - \frac{j+\kappa'-\kappa}{2\kappa} & \omega \in [(j+\kappa')/2M, (j+1)/2M] \end{cases} \tag{4.21}$$

The graph of $g_{5,0.99}$ is presented in Figure 1.

It is easy to verify that $g_{M,\kappa}$ is piecewise affine, uniformly expanding, and keeps the Lebesgue measure invariant. Also, the minimal expansion of $g_{M,\kappa}$ can be made arbitrarily large by letting $M \to \infty$.

Notice that for $1 \leqslant j \leqslant M$,

$$g_{M,\kappa}([j/2M, (j+\kappa)/2M]) = [0, 1/2] \text{ and } g_{M,\kappa}([(j+\kappa)/2M, (j+1)/2M]) = [1/2, 1]$$

while for $M+1 \leqslant j \leqslant 2M$

$$g_{M,\kappa}([j/2M, (j+\kappa')/2M]) = [0, 1/2] \text{ and } g_{M,\kappa}([(j+\kappa')/2M, (j+1)/2M]) = [1/2, 1].$$

Picking $\kappa \approx 1$, most of the points in the interval $[0, 1/2]$ are mapped back to $[0, 1/2]$, and also most of the points of $[1/2, 1]$ are mapped back to $[1/2, 1]$. More precisely, defining $V_1 := [0, 1/2]$, $V_2 := [1/2, 1]$ and

$$V_{i,n} := \{\omega \in V_i : g^k_{M,\kappa}(\omega) \in V_i \text{ for } 0 \leqslant k \leqslant n-1\};$$

$V_{i,n} \subset V_i$ is such that, for any $n \in \mathbb{N}$,

$$|V_{i,n}| \to 1/2 \quad \text{as} \quad \varkappa \to 1. \tag{4.22}$$

Fix $\varepsilon > 0$ a small number. Pick $\varphi : \mathbb{T} \to \mathbb{T}$ a $N - S$ diffeomorphism such that $|\varphi(x) - x| \leqslant \varepsilon$[†], and define

$$f_\omega(x) = \begin{cases} 2\omega & \omega \in I_1 \\ \varphi + a\omega & \omega \in I_2. \end{cases} \tag{4.23}$$

One can check that $\mathcal{P} := \int_{\mathbb{T}} d\omega f_{\omega *}$ maps a small closed ball around $\mathrm{Leb}_{\mathbb{T}}$ into itself, with the diameter of the ball going to zero (in Total Variation distance) when $a \to 0$. This implies that the unique fixed point of $\mathcal{P}$ is close to $\mathrm{Leb}_{\mathbb{T}}$.

To ease the notation, from now on we write g in place of $g_{M,\varkappa}$. Let's look at $f_\omega^{n-1} := f_{g^{n-1}(\omega)} \circ \ldots \circ f_\omega$ and study $\mathcal{P}' := \int_{\mathbb{T}} d\omega (f_\omega^{n-1})_*$. Fix $\Delta > 0$ small. For any $\overline{x}_0 \in \mathbb{T}$ and $(\omega_k)_{k=0}^{n-1}$ with $\omega_k \in V_2$, consider $(\overline{x}_k)_{k=0}^{n-1}$ with $\overline{x}_{k+1} = \varphi(\overline{x}_k) + a\omega_k$. Pick $n \in \mathbb{N}$ large and $a > 0$ small so that for any $\overline{x}_0 \in [N - \Delta, N + \Delta]^c$ and $(\omega_k)_{k=0}^{n-1}$ as above, $\overline{x}_{n-1} \in [S - \Delta, S + \Delta]$. One can find $\varkappa$ close enough to one so that $|V_{1,n}| = |V_{2,n}| = 0.49$, which implies

$$\begin{aligned}
\mathcal{P}'\eta &= \int_{V_{1,n}} d\omega (f_\omega^{n-1})_*\eta + \int_{V_{2,n}} d\omega (f_\omega^{n-1})_*\eta + \int_{(V_{1,n}\cup V_{2,n})^c} d\omega (f_\omega^{n-1})_*\eta \\
&= 0.49\,\mathrm{Leb}_{\mathbb{T}} + \int_{V_{2,n}} d\omega (f_\omega^{n-1})_*\eta + \int_{(V_{1,n}\cup V_{2,n})^c} d\omega (f_\omega^{n-1})_*\eta.
\end{aligned}$$

Given the expression of $\mathcal{P}'$, if η_0' is such that $\mathcal{P}'\eta_0' = \eta_0'$ then, $\eta_0' = 0.49\,\mathrm{Leb} + 0.51\eta_1$, where η_1 is some probability measure. This implies that

$$\begin{aligned}
\mathcal{P}'\eta_0([S - \Delta, S + \Delta]) &= (0.49\mathcal{P}'\,\mathrm{Leb} + 0.51\mathcal{P}'\eta_1)([S - \Delta, S + \Delta]) \\
&= 0.49 \int_{V_{2,n}} d\omega (f_\omega^{n-1})_*\,\mathrm{Leb}([S - \Delta, S + \Delta]) \\
&\quad + (1 - 0.49^2)\eta_2([S - \Delta, S + \Delta]) \\
&> 0.49^2(1 - 2\Delta)
\end{aligned}$$

where η_2 above is some probability measure. Since $\Delta > 0$ is arbitrary, $\eta_0'([S - \Delta, S + \Delta]) \approx 1/4$ while $\eta_0([S - \Delta, S + \Delta]) \approx 2\Delta$ which makes η_0 and η_0' two very far apart measures with respect to most metrics (e.g., d_{TV}, d_W,...).

In Figure 2 below we compare numerical simulations of the distribution of mass on the vertical marginal after several iterations of skew-products F with different base maps g. For each such map, we consider several initial conditions sampled randomly and uniformly on $[0, 1] \times [0, 1]$, let F act for a while on these points, then take their vertical coordinates, and plot them on a histogram.

[†] A North-South (NS) diffeomorphism is a diffeomorphism with exactly two fixed points: one attracting, the South Pole (S), and one repelling, the North Pole (N), such that for any $x \neq N$, $\varphi^n(x) \to S$. Furthermore, $\varphi'(N) > 1$ and $\varphi'(S) < 1$ so that the two fixed points are hyperbolic.

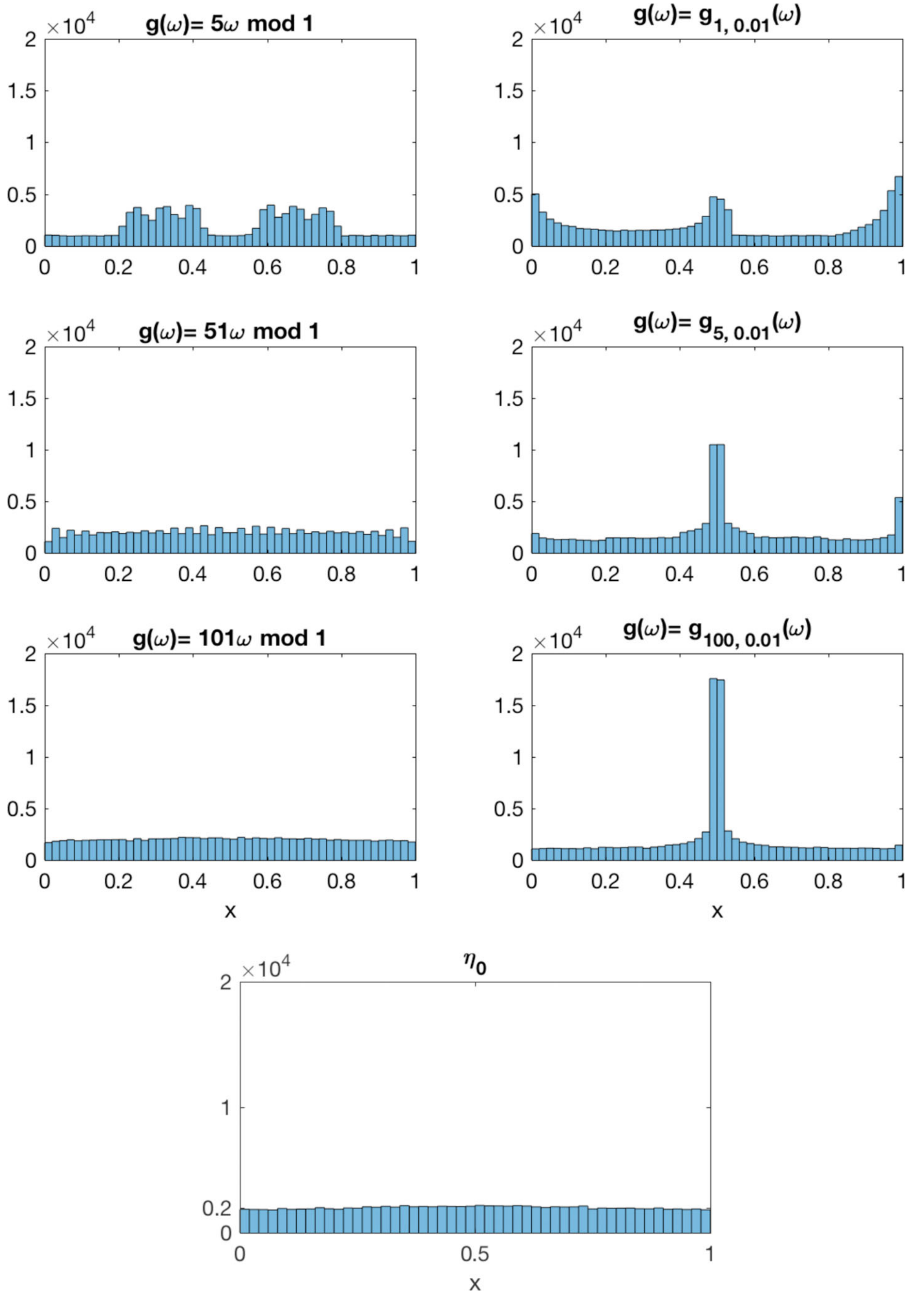

FIGURE 2 For different base maps g, we consider 10^4 initial conditions $\{(\omega_k, x_k)\}_{k=1}^{10^4}$ sampled randomly and uniformly on $[0,1] \times [0,1]$, let $F(\omega, x) = (g(\omega), f(\omega, x))$ act for 100 time steps to obtain $\{F^{100}(\omega_k, x_k)\}_{k=1}^{10^4}$, take the vertical x-coordinates of these points, and plot them on a histogram. The different g maps used are indicated above the histograms. The fiber maps are the same throughout and as in (4.23) with $\varphi(x) = x - 0.01\sin(2\pi x)$ and $a = 0.001$. The last panel shows a numerical approximation for η_0 obtained as in the deterministic case by applying F to $\{(\omega_k, x_k)\}_{k=1}^{10^4}$ but where, instead of having g in the base, we sampled the ω-coordinate at random independently (both w.r.t. time and initial conditions) and uniformly on [0,1] using the random number generator built in the programming language

When the expansion in the base is large, we expect the distribution given by the histogram to be close to $\overline{\eta}$. We compare the case of base maps with no distortion, $g(\omega) = \sigma\omega \bmod 1$, against base maps $g_{M,\kappa}$ defined above. We also simulate numerically η_0, the stationary measure for $\mathcal{P}$ (as given by the random number generator of the programme). The fiber maps f_ω are of the kind described in (4.23).

In the case without distortion, when the minimal expansion in the base increases, we can see that the simulated $\overline{\eta}$ becomes very close to η_0 (as per Propostion 3.3 point ii)), while in the case with distortion, $\overline{\eta}$ and η_0 are different.

5 | GENERALIZATIONS AND LIMITATIONS

In this section we discuss a few generalizations of the results and the techniques presented above, and also some of the limitations. Before proceeding with the generalizations, we would like to stress that the goal of this paper was not to give a result in its greatest generality possible, but rather to present some techniques that we believe can be applied (with different levels of additional effort) to various setups.

5.1 | Assumptions on the base

The regularity assumptions on the map g can be revised to fit other situations. For example, Ω could be a compact manifold with boundary such as $\Omega = [0,1]^{m_1}$ with g piecewise C^2 with onto branches. By this we mean that there are open sets $\{I_i\}_{i=1}^d$ partitioning Ω modulo sets of measure zero, and such that $g|_{I_i} : I_i \to (0,1)^{m_1}$ is a C^2 uniformly expanding diffeomorphism with bounded distortion.

For the system in Section 3, that is, when $m_1 = 1$ and no distortion, this corresponds to considering maps $g : [0,1] \to [0,1]$ for which there are $n \in \mathbb{N}$ and $0 =: a_0 < a_1 < \ldots < a_n < a_{n+1} := 1$ such that $g|_{(a_i,a_{i+1})}$ is C^2 and onto (0,1). It is easy to check that all the proof of statements in Section 3 hold, *mutatis mutandis*, for maps g satisfying these assumptions.

Also the assumption that g must be C^2 (or piecewise C^2) is not necessary, and can be substituted by g being $C^{1+\alpha}$ (or piecewise $C^{1+\alpha}$), meaning that g is once differentiable and with $\alpha-$Hölder differential (or same property, but piecewise).

In fact, we expect that our results hold for more general bases. What is needed is $g : \Omega \to \Omega$ where Ω is a compact Riemannian manifold (with or without boundary), and g is a piecewise uniformly expanding map with bounded distortion admitting a Markov partition and a unique absolutely continuous invariant manifold.

5.2 | Robustness under conjugacy

Consider a map $\hat{g} : \Omega \to \Omega$ and assume that there is an invertible map $h : \Omega \to \mathbb{T}^{m_1}$ which is measurable and with measurable inverse, such that $\hat{g} := h^{-1} \circ g \circ h$ for a map $g : \mathbb{T}^{m_1} \to \mathbb{T}^{m_1}$. Consider $\hat{f} : \Omega \times \mathbb{T}^{m_2} \to \mathbb{T}^{m_2}$ and the skew-product system $\hat{F} : \Omega \times \mathbb{T}^{m_2} \to \Omega \times \mathbb{T}^{m_2}$

$$\hat{F}(\omega, x) = (\hat{g}(\omega), \hat{f}(\omega, x)).$$

Then, if one can show that the skew-system $F : \mathbb{T}^{m_1} \times \mathbb{T}^{m_2} \to \mathbb{T}^{m_1} \times \mathbb{T}^{m_2}$

$$F(\omega, x) = (g(\omega), f(\omega, x))$$

with $f(\omega, x) := \hat{f}(h^{-1}\omega, x)$, satisfies an approximate decay of correlations (as in Theorem 2.2), then so does $\hat{F}$. This is made precise in the following proposition.

Proposition 5.1. *Suppose $\hat{F} : \Omega \times \mathbb{T}^{m_2} \to \Omega \times \mathbb{T}^{m_2}$ and $F : \mathbb{T}^{m_1} \times \mathbb{T}^{m_2} \to \mathbb{T}^{m_1} \times \mathbb{T}^{m_2}$ are as above, and assume that for some $\varepsilon > 0$, η a probability measure, $\widetilde{C} > 0$ and $\widetilde{\lambda} \in (0, 1)$ the conclusion of Theorem 2.2 holds for F. Then, defining $\nu := (h^{-1})_* \operatorname{Leb}_{\mathbb{T}^{m_1}}$*

$$\left| \int_{\Omega \times \mathbb{T}^{m_2}} \varphi(\pi_2 \hat{F}^n(\omega, x))\psi(x) dx d\nu(\omega) - \int_{\mathbb{T}^{m_2}} \varphi(x) d\eta(x) \int_{\mathbb{T}^{m_2}} \psi(x) dx \right| \leqslant C_{\varphi,\psi}(\widetilde{C}\widetilde{\lambda}^n + \varepsilon)$$

for all $\psi \in L^1(\mathbb{T}^{m_2}; \mathbb{R})$ and $\varphi \in \operatorname{Lip}(\mathbb{T}^{m_2}; \mathbb{R})$.

Proof. Take $\psi \in L^1(\mathbb{T}^{m_2}; \mathbb{R})$ and $\varphi \in \operatorname{Lip}(\mathbb{T}^{m_2}; \mathbb{R})$. Define $\nu = (h^{-1})_* \operatorname{Leb}_{\mathbb{T}^{m_1}}$ a probability measure on Ω. Let's call $H := h \times \mathrm{id}$ which is invertible with inverse $H^{-1} = h^{-1} \times \mathrm{id}$.

$$\begin{aligned} \int_{\Omega \times \mathbb{T}^{m_2}} \psi\ \varphi \circ \pi_2 \circ \hat{F}^n d\nu \otimes \operatorname{Leb} &= \int_{\Omega \times \mathbb{T}^{m_2}} \psi \circ \pi_2 \circ H^{-1}\ \varphi \circ \pi_2 \circ \hat{F}^n \circ H^{-1} dH_*(\nu \otimes \operatorname{Leb}) \\ &= \int_{\Omega \times \mathbb{T}^{m_2}} \psi\ \varphi \circ \pi_2 H \circ \hat{F}^n \circ H^{-1} d \operatorname{Leb} \\ &= \int_{\Omega \times \mathbb{T}^{m_2}} \psi\ \varphi \circ \pi_2 \circ F^n d \operatorname{Leb}. \end{aligned}$$

Therefore, from the assumptions, there is $C_{\varphi,\psi} > 0$ such that

$$\left| \int_{\Omega \times \mathbb{T}^{m_2}} \psi\ \varphi \circ \pi_2 \circ \hat{F}^n d\nu \otimes \operatorname{Leb} - \int_{\mathbb{T}^{m_2}} \varphi(x) d\eta(x) \int_{\mathbb{T}^{m_2}} \psi(x) dx \right| \leqslant C_{\varphi,\psi}(\widetilde{C}\widetilde{\lambda}^n + \varepsilon). \qquad \square$$

As an example, one can use Theorem 2.2 to prove approximate decay of correlation in case the forcing is driven by a power of the logistic map $\hat{g}_0(x) = 4x(1 - x)$. In fact, it is well known $\hat{g}_0$ is conjugate to the tent map

$$g_0 = \begin{cases} 2x & x \in [0, 1/2) \\ 1 - 2x & x \in [1/2, 1] \end{cases},$$

via a C^1 map $h : [0, 1] \to [0, 1]$. Analogously, for any $n \in \mathbb{N}$, also $\hat{g} := \hat{g}_0^n$ is conjugate to $g := g_0^n$ via h, and g is in the class of maps admitted by the generalization in Section 5.1 for which one can apply Theorem 2.2.

5.3 | Fiber generalization

We expect that the choice $X = \mathbb{T}^{m_2}$ for the vertical fiber can be relaxed. As long as an analogue of Proposition 4.1 holds, one can study $F : \mathbb{T}^{m_1} \times X \to \mathbb{T}^{m_1} \times X$ where X is a compact metric space.

What is needed is that topological conditional measures on vertical fibers $\{\omega\} \times X$ give a disintegration w.r.t the vertical foliation. From the results in [43], a sufficient condition is for X to be a compact Riemanninan manifold or a compact separable ultrametric space.

5.4 | More or less regular disintegrations

In Definition 2.1 we have given the definition of Lipschitz disintegration $\{\mu_\omega\}_{\omega\in\Omega}$ and later we have shown how, under the hypotheses of Theorem 2.2, certain classes of measures with Lipschitz disintegration were kept invariant by the dynamics. Measures having Hölder disintegration can be defined in a completely analogous way, and they can be used to define classes of invariant measures for example in the case where $f : \Omega \times X \to X$ is only Hölder and not Lipschitz.

Analogously, one could think of defining measures having disintegrations of higher regularity, for example, differentiable for a suitable notion of differentiability for curves in $\mathcal{M}_1(X)$, and exploit these classes.

5.5 | Limitations of the approach

A more substantial and also natural step forward from Theorem 2.2, would be considering g an invertible uniformly hyperbolic map, like an Anosov diffeomorphism or a map with an Axiom A attractor. Unfortunately it seems hard to extend the techniques in this paper to this case. The main reason is that one needs the contraction properties of the inverse branches of g: for invertible uniformly hyperbolic systems, instead, some directions are contracted when taking preimages, but others are expanded and this spoils the arguments.

For the same reason our approach is evidently ill-suited to treat skew-products with quasi-periodic base (see, e.g., [20]).

APPENDIX A: MARKOV CHAINS AND RANDOM DYNAMICS

In this section we report some classical results about geometric ergodicity of Markov chains [16, 21 31, 37], and we relate this to random dynamical systems in discrete time.

A.1 | Markov chains and geometric ergodicity

Definition A.1. Given a Polish space S, the *state space*, endowed with a countably generated $\sigma-$field $\mathcal{B}(S)$, a discrete time Markov process is a sequence of random variables $\{X_t\}_{t\in\mathbb{N}_0}$ defined on a probability space $(\Omega, \Sigma, \mathbb{P})$ such that for all $n \in \mathbb{N}$

$$\mathbb{E}_{\mathbb{P}}[X_n|X_{n-1}, \dots, X_0] = \mathbb{E}_{\mathbb{P}}[X_n|\, X_{n-1}].$$

The Markov process is called *stationary* if $\mathbb{E}_{\mathbb{P}}[X_n|\, X_{n-1}]$ does not depend on n, and $P : S \times \mathcal{B}(S) \to \mathbb{R}^+$ is the associated transition kernel satisfying

$$\mathbb{P}\big(X_{n+1} \in A|\, X_n = x\big) = P(x, A).$$

For every $x \in S$, $P(x, \cdot)$ defines a probability measure with the following meaning: $P(x, A)$ is the probability that $X_{n+1} \in A$ given that at time n one has observed $X_n = x$.

Given a stationary Markov process and $n \in \mathbb{N}$, one can extend the notion of kernel to higher iterates: Define $P^m : S \times \mathcal{B}(S) \to \mathbb{R}^+$

$$P^m(x, A) = \mathbb{P}\big(X_{n+m} \in A | X_n = x\big).$$

For any $n \in \mathbb{N}$, P^n generates an action on the set of measures on $(S, \mathcal{B}(S))$ in the following way. Given μ a measure on S, define

$$\mathcal{P}\mu(A) := \int_S P(x, A) d\mu(x),$$

and

$$\mathcal{P}^n\mu(A) := \int_S P^n(x, A) d\mu(x).$$

Using the properties of transition kernels one can prove that, $\{\mathcal{P}^n\}$ satisfies the semi-group property

$$\mathcal{P}^n \circ \mathcal{P}^m = \mathcal{P}^{n+m}$$

making $\mathcal{P}$ the generator of a semi-group action on probability measures on $(S, \mathcal{B}(S))$.

Definition A.2. A stationary Markov chain is said to be *geometrically ergodic* if there are $C > 0$ and $\lambda \in (0, 1)$ such that

$$d_{TV}(P^n(x_1, \cdot), P^n(x_2, \cdot)) \leqslant C\lambda^n, \qquad \forall x_1, x_2 \in S.$$

For a definition of the Total Variation distance d_{TV} see the beginning of Sec. C. From Definition A.2 follows that if a Markov chain is *geometrically ergodic*, then there is a probability measure η_0 such that, for every probability measure μ on $(S, \mathcal{B}(S))$,

$$d_{TV}(\mathcal{P}^n\mu, \eta_0) \leqslant C\lambda^n.$$

The measure η_0 satisfies $\mathcal{P}(\eta_0) = \eta_0$ and is also called a *stationary distribution* or *stationary measure*.

A.2 | Sufficient conditions for geometric ergodicity

In this subsection we give a sufficient condition that ensures geometric ergodicity of a stationary Markov chain. Weaker conditions working in more general setups are available and involve *petite sets* [21] or Lyapunov functions [31].

Theorem A.1 [37]. *Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a stationary Markov chain on $(S, \mathcal{B}(S))$ with transition kernel $P : S \times \mathcal{B}(S) \to \mathbb{R}^+$. Assume there is ν a probability measure, $\varepsilon > 0$ and $n_0 \in \mathbb{N}$ such that*

$$P^{n_0}(x, \cdot) \geqslant \varepsilon\nu(\cdot), \qquad \forall x \in S.$$

Then the Markov chain is geometrically ergodic.

A.3 | Randomly forced systems and Markov chains

In this section we discuss the difference, in terms of mathematical definitions, between random and deterministic forcing.

By *random forcing*, we mean that given a probability space (Ω, ν) and $f : \Omega \times X \to X$, at the $n-$th iteration we apply the map $f_{\zeta_n} := f(\zeta_n, \cdot) : X \to X$, where $\{\zeta_n\}_{n\in\mathbb{N}}$ is an i.i.d sequence of random variables defined on some probability space $(\Xi, \mathbb{P})$ with values in Ω and distributed according to ν. Fixed $w \in \Xi$, the forward orbits of the system are given by

$$O(x) := \left\{ f_{\zeta_n(w)} \circ \ldots \circ f_{\zeta_1(w)} \circ f_{\zeta_0(w)}(x) : \ n \in \mathbb{N}_0 \right\}, \qquad \forall x \in X.$$

An important example of random forcing is given by additive i.i.d. noise: Consider $X = \mathbb{T}^m$, or any other set with an additive structure, a map $T : X \to X$ and $\{\zeta_n\}_{n\in\mathbb{N}_0}$ an i.i.d. sequence of random variables with values in X and distributed according to ν, then taking $\Omega = X$ define $f : \Omega \times X \to X$ as

$$f(\omega, x) := T(x) + \omega.$$

Composing at time $n \in \mathbb{N}$ by f_{ζ_n} corresponds to considering the recursive equation

$$\mathcal{X}_{n+1} = T(\mathcal{X}_n) + \zeta_n, \qquad \forall n \in \mathbb{N}_0,$$

where $\mathcal{X}_n$ is the state of the system at time n. What the above means is that, calling $(\Xi, \mathbb{P})$ the underlying probability space where $\{\zeta_n\}_{n\in\mathbb{N}_0}$ are defined, $\{\mathcal{X}_n\}_{n\in\mathbb{N}_0}$ are random variables satisfying

$$\mathbb{P}\big(\mathcal{X}_{n+1} \in A | \mathcal{X}_n = x_n\big) = \mathbb{P}(\xi_n \in (A - T(x_n))),$$

and thus $\{\mathcal{X}_n\}_{n\in\mathbb{N}_0}$ is a Markov chain. In the above, T denotes the intrinsic dynamics, that is, the dynamics the system would have if it did not receive any forcing, while ξ_n is the random forcing noise term.

Deterministic forcing is also represented as application at time $n \in \mathbb{N}$ of the map $f(\zeta_n, \cdot) : X \to X$. However, in this case the sequence $\{\zeta_n\}_{n\in\mathbb{N}}$ is not required to be independent, but it should satisfy $\zeta_{n+1} = g(\zeta_n)$ for some transformation $g : \Omega \to \Omega$ that preserves the measure ν. This corresponds also to the general definition of *random dynamical system* usually given in the literature (see [3]).

The difference between random and deterministic forcing is not a stark one. In fact one can show that random forcing is a particular case of deterministic forcing where g is an appropriate shift map: Given $f : \Omega \times X \to X$ and $\{\zeta_n\}_{n\in\mathbb{N}_0}$ an i.i.d sequence with values in Ω distributed as ν, we can construct the probability space (Ω', ν') with $\Omega' := \Omega^{\mathbb{N}_0}$ and $\nu' := \nu^{\otimes\mathbb{N}_0}$. Now define the sequence of identically distributed random variables $\{\boldsymbol{\zeta}_k\}_{k\in\mathbb{N}_0}$ in Ω' with $\boldsymbol{\zeta}_k := \{\zeta_{n+k}\}_{n\in\mathbb{N}_0}$, and $f' : \Omega' \times X \to X$

$$f'(\boldsymbol{\zeta}, x) := f((\boldsymbol{\zeta})_0, x),$$

where $(\boldsymbol{\zeta})_0$ denotes the first term of the sequence $\boldsymbol{\zeta} \in \Omega'$. With this definition we also have

$$\boldsymbol{\zeta}_{k+1} = \sigma^{k+1}(\{\zeta_n\}_{n\in\mathbb{N}_0}) = \sigma(\boldsymbol{\zeta}_{k-1})$$

where $\sigma : \Omega' \to \Omega'$ is the left shift which is easy to check that keeps the measure ν' invariant.

APPENDIX B: DISINTEGRATION OF MEASURE AND ROHLIN'S THEOREM

The following definitions and results are taken from [43], adapted to the level of generality needed in this paper.

Definition B.1. Let (X, μ) be a topological probability space, Y a metric space and $\pi : X \to Y$ a measurable function. Call $\hat{\mu} := \pi_* \mu$. A system of conditional measures of μ with respect to (X, π, Y) is a collection of measures $\{\mu_y\}_{y \in Y}$ such that

1) For $\hat{\mu}$−almost every $y \in Y$, μ_y is a probability measure on $\pi^{-1}(y)$.
2) For every measurable subset $B \subset X$, $y \mapsto \mu_y(B)$ is measurable and

$$\mu(B) = \int \mu_{\pi^{-1}(y)}(B) d\hat{\mu}(y).$$

When Y in the above definition is a measurable partition of X and $\pi(x)$ is the unique element of the partition to which x belongs, then we also call $\{\mu_y\}_{y \in Y}$ a disintegration of μ.

Definition B.2. In the same setup of Definition B.1, the topological conditional measure of μ with respect to (X, π, y, Y) is the weak$*$ limit (if it exists)

$$\mu_y := \lim_{\varepsilon \to 0^+} \mu_{\pi^{-1}(B(y,\varepsilon))}$$

where $B(y, \varepsilon)$ is the ball centered at y with radius ε with respect to the metric on Y and

$$\mu_{\pi^{-1}(B(y,\varepsilon))}(I) = \frac{\mu(\pi^{-1}(B(y, \varepsilon)) \cap I)}{\mu(\pi^{-1}(B(y, \varepsilon))}.$$

Theorem B.1 (Theorem 2.2 [43]). *Let (X, μ) be a compact metric probability space, let Y be a separable Riemannian manifold. Let $\pi : X \to Y$ be measurable.Then for $\hat{\mu}$−almost every $y \in Y$, the topological conditional measure of μ with respect to (X, π, y, Y) exists as in Definition B.2. Furthermore the collection of measures $\{\mu_y\}_{y \in Y}$ is a system of conditional measures as in Definition B.1. (If μ_y does not exist, set $\mu_y = 0$).*

APPENDIX C: WASSERSTEIN DISTANCE: SOME COMPUTATIONS

Consider a compact metric space (Y, d). Then the Kantorovich–Wasserstein between $\mu_1, \mu_2 \in \mathcal{M}_1(Y)$ is defined as

$$d_W(\mu_1, \mu_2) := \sup_{\gamma \in C(\mu_1, \mu_2)} \int_{Y \times Y} d(s, s') d\gamma(s, s')$$

where $C(\mu_1, \mu_2)$ is the set of all couplings between μ_1 and μ_2. If we consider instead of the metric d the discrete metric d_{dis} defined as

$$d_{dis}(s, s') = \begin{cases} 1 & s = s' \\ 0 & s \neq s' \end{cases}$$

We have that

$$d_{TV}(\mu_1, \mu_2) := \sup_{\gamma \in C(\mu_1,\mu_2)} \int_{Y\times Y} d_{dis}(s,s')d\gamma(s,s').$$

Lemma C.1. *Let (Y,d) be a metric space, $T : Y \to Y$ a Lipschitz transformation with Lipschitz constant* Lip(T)*. Then for any $\xi_1, \xi_2 \in \mathcal{M}_1(\mathbb{T})$*

$$d_W(T_*\xi_1, T_*\xi_2) \leqslant \mathrm{Lip}(T) d_W(\xi_1, \xi_2).$$

Proof.

$$\begin{aligned} d_W(T_*\xi_1, T_*\xi_2) &= \sup_{\varphi\in \mathrm{Lip}^1(Y)} \int_Y \varphi(y) d(T_*\xi_1 - T^*\xi_2)(y) \\ &= \sup_{\varphi\in \mathrm{Lip}^1(Y)} \int_Y \varphi \circ T d(\xi_1 - \xi_2)(y) \\ &\leqslant \sup_{\varphi\in \mathrm{Lip}^1(Y)} \mathrm{Lip}(\varphi\circ T) d_W(\xi_1,\xi_2) \end{aligned}$$

and $\mathrm{Lip}(\varphi \circ T) \leqslant \mathrm{Lip}(\varphi)\,\mathrm{Lip}(T)$. □

Remark C.1. The above lemma can be read in the following way: If $T : (Y,d) \to (Y,d)$ is Lipschitz, then $T_* : (\mathcal{M}_1(Y), d_W) \to (\mathcal{M}_1(Y), d_W)$ is Lipschitz with $\mathrm{Lip}(T_*) = \mathrm{Lip}(T)$.

Lemma C.2. *Consider (S,ν) a measurable space with ν a probability measure, and Y a compact metric space. Assume that $\{\mu_s\}_{s\in S}$ is a family of measures belonging to $\mathcal{M}_1(Y)$ and that $\exists \ell > 0$ s.t. $d_W(\mu_s, \mu_{s'}) \leqslant \ell$ for for every $s, s' \in S$. Then the measure $\overline{\mu} \in \mathcal{M}_1(Y)$ defined as*

$$\overline{\mu}(A) := \int_S d\nu(s)\mu_s(A)$$

is such that $d_W(\overline{\mu}, \mu_s) \leqslant \ell$ for all $s \in S$.

Proof. Pick $s \in S$

$$\begin{aligned} d_W(\overline{\mu}, \mu_s) &= \sup_{\varphi\in \mathrm{Lip}^1(Y)} \int_Y \varphi(y) d(\overline{\mu} - \mu_s)(y) \\ &= \sup_{\varphi\in \mathrm{Lip}^1(Y)} \int_Y \int_S d\nu(s')\varphi(y) d(\mu_{s'} - \mu_s)(y) \\ &\leqslant \int_S d\nu(s') \sup_{\varphi\in \mathrm{Lip}^1(Y)} \int_Y \varphi(y) d(\mu_{s'} - \mu_s)(y) \\ &\leqslant \int_S d\nu(s') d_W(\mu_s, \mu_{s'}) \\ &\leqslant \ell. \end{aligned}$$

□

Lemma C.3. *Let* (Y, d) *be a bounded metric space and call* $\operatorname{diam}(Y)$ *its diameter. Then*

$$d_W(\mu_1, \mu_2) \leqslant \operatorname{diam}(Y) d_{TV}(\mu_1, \mu_2).$$

Proof.

$$\begin{aligned} d_W(\mu_1, \mu_2) &= \sup_{\gamma \in C(\mu_1, \mu_2)} \int_{Y \times Y} d(s, s') d\gamma(s, s') \\ &\leqslant \sup_{\gamma \in C(\mu_1, \mu_2)} \int_{Y \times Y} \operatorname{diam}(Y) d_{dis}(s, s') d\gamma(s, s') \\ &= \operatorname{diam}(Y) d_{TV}(\mu_1, \mu_2). \end{aligned}$$

□

Lemma C.4. *Assume* $\{\mu_i\}_{i=1}^n$ *and* $\{\mu_i'\}_{i=1}^n$ *are probability measures in* $\mathcal{M}_1(Y)$ *and* $\{b_i\}_{i=1}^n$, $b_i > 0$, *are weights with* $\sum_{i=1}^n b_i = 1$. *Then*

$$d_W\left(\sum_{i=1}^n b_i \mu_i, \sum_{i=1}^n b_i \mu_i'\right) \leqslant \max_i d_W(\mu_i, \mu_i').$$

Proof.

$$\begin{aligned} d_W\Big(\sum_{i=1}^n b_i \mu_i, \sum_{i=1}^n b_i \mu_i'\Big) &\leqslant \sup_{\varphi \in \mathrm{Lip}^1} \int_Y \varphi d\left(\sum_{i=1}^n b_i \mu_i - \sum_{i=1}^n b_i \mu_i'\right) \\ &\leqslant \sum_{i=1}^n b_i \sup_{\varphi \in \mathrm{Lip}^1} \int_Y \varphi d(\mu_i - \mu_i') \\ &\leqslant \sum_{i=1}^n b_i d_W(\mu_i, \mu_i') \\ &\leqslant \max_i d_W(\mu_i, \mu_i'). \end{aligned}$$

□

ACKNOWLEDGEMENTS

The first author is partially supported by the PRIN Grant 2017S35EHN “Regular and stochastic behaviour in dynamical systems” of MIUR, Italy and by INDAM - GNFM 2020 project “Deterministic and stochastic dynamical systems for climate studies”. The second author acknowledges the support of UniCredit Bank R&D group through the “Dynamics and Information Theory Institute” at the Scuola Normale Superiore. The third author acknowledges funding from the H2020 Marie Skłodowska-Curie Actions “Ergodic Theory of Complex Systems” project no. 843880. He is also grateful for the hospitality of the Centro di Ricerca Matematica Ennio de Giorgi and Scuola Normale Superiore where part of this work was carried out. The authors thank the anonymous referee, Lai-Sang Young and Stefano Galatolo for comments and discussions on preliminary versions.

Open Access Funding provided by Universita degli Studi di Pisa within the CRUI-CARE Agreement.

JOURNAL INFORMATION

The *Journal of the London Mathematical Society* is wholly owned and managed by the London Mathematical Society, a not-for-profit Charity registered with the UK Charity Commission.

All surplus income from its publishing programme is used to support mathematicians and mathematics research in the form of research grants, conference grants, prizes, initiatives for early career researchers and the promotion of mathematics.

REFERENCES

1. J. F. Alves, C. Bonatti, and M. Viana, *SRB measures for partially hyperbolic systems whose central direction is mostly expanding*, Invent. Math. **140** (2000), no. 2, 351–398 (English).
2. J. F. Alves, C. L. Dias, S. Luzzatto, and V. Pinheiro, *SRB measures for partially hyperbolic systems whose central direction is weakly expanding*, J. Eur. Math. Soc. **19** (2017), no. 10, 2911–2946.
3. L. Arnold, *Random dynamical systems*, Springer, Berlin, 1998 (English).
4. V. Baladi and L.-S. Young, *On the spectra of randomly perturbed expanding maps*, Comm. Math. Phys. **156** (1993), no. 2, 355–385.
5. P. Berger and P. D. Carrasco, *Non-uniformly hyperbolic diffeomorphisms derived from the standard map*, Commun. Math. Phys. **329** (2014), no. 1, 239–262 (English).
6. K. Bjerklöv, *A note on circle maps driven by strongly expanding endomorphisms on t*, Dynam. Sys. **33** (2018), 361–368.
7. A. Blumenthal, J. Xue, and L.-S. Young, *Lyapunov exponents and correlation decay for random perturbations of some prototypical 2d maps*, Commun. Math. Phys. **359** (2018), no. 1, 347–373 (English).
8. C. Bose, A. Quas, and M. Tanzi, *Random composition of L-S-V maps sampled over large parameter ranges*, Nonlinearity **34** (2021), no. 6, 3641–3675 (English).
9. A. Boyarsky and P. Góra, *Laws of chaos. Invariant measures and dynamical systems in one dimension*, Birkhäuser, Boston, MA, 1997 (English).
10. K. Burns and A. Wilkinson, *On the ergodicity of partially hyperbolic systems*, Ann. Math. (2) **171** (2010), no. 1, 451–489 (English).
11. O. Butterley and P. Eslami, *Exponential mixing for skew products with discontinuities.*, Trans. Am. Math. Soc. **369** (2017), no. 2, 783–803 (English).
12. P. D. Carrasco, *Random products of standard maps*, Commun. Math. Phys. **377** (2020), no. 2, 773–810 (English).
13. R. Castorrini and C. Liverani, *Quantitative statistical properties of two-dimensional partially hyperbolic systems*, https://arxiv.org/abs/2007.05602 (2020).
14. I. Chevyrev, P. K. Friz, A. Korepanov, and I. Melbourne, *Superdiffusive limits for deterministic fast-slow dynamical systems*, Probab. Theory Relat. Fields **178** (2020), no. 3-4, 735–770 (English).
15. B. Christian and V. Marcelo, *SRB measures for partially hyperbolic systems whose central direction is mostly contracting*, Isr. J. Math. **115** (2000), 157–193.
16. M. Cloez, B. Hairer, *Exponential ergodicity for Markov processes with random switching*, Bernoulli **21** (2015), 505–536.
17. J. De Simoi and C. Liverani, *Limit theorems for fast-slow partially hyperbolic systems.*, Invent. Math. **213** (2018), no. 3, 811–1016 (English).
18. D. Dolgopyat, *Limit theorems for partially hyperbolic systems*, Trans. Am. Math. Soc. **356** (2004), no. 4, 1637–1689 (English).
19. D. Dolgopyat, *On differentiability of SRB states for partially hyperbolic systems*, Invent. Math. **155** (2004), no. 2, 389–449 (English).
20. D. Dolgopyat and B. Fayad, *Limit theorems for toral translations*, Hyperbolic dynamics, fluctuations and large deviations **89** (2015), 227–277.
21. D. Down, S. P. Meyn, and R. L. Tweedie, *Exponential and uniform ergodicity of Markov processes*, Ann. Probab. **23** (1995), no. 4, 1671–1691.
22. D. Dragičević, G. Froyland, C. Gonzalez-Tokman, and S. Vaienti, *A spectral approach for quenched limit theorems for random expanding dynamical systems*, Comm. Math. Phys. **360** (2018), no. 3, 1121–1187.
23. D. Dragičević, G. Froyland, C. González-Tokman, and S. Vaienti, *A spectral approach for quenched limit theorems for random hyperbolic dynamical systems*, Trans. Amer. Math. Soc. **373** (2020), no. 1, 629–664.
24. S. Galatolo, *Quantitative statistical stability, speed of convergence to equilibrium and partially hyperbolic skew products*, J. Éc. Polytech., Math. **5** (2018), 377–405 (English).

25. S. Galatolo and R. Lucena, *Spectral gap and quantitative statistical stability for systems with contracting fibers and Lorenz-like maps*, Discrete Contin. Dyn. Syst. **40** (2020), no. 3, 1309–1360 (English).
26. S. Galatolo, J. Rousseau, and B. Saussol, *Skew products, quantitative recurrence, shrinking targets and decay of correlations.*, Ergodic Theory Dyn. Syst. **35** (2015), no. 6, 1814–1845 (English).
27. P. Giulietti, A. Hammerlindl, and D. Ravotti, *Quantitative global-local mixing for accessible skew products*, Ann. Henri Poincaré **23** (2022), no. 3, 923–971 (English).
28. S. Gouëzel, *Statistical properties of a skew product with a curve of neutral points*, Ergodic Theory Dynam. Systems **27** (2007), no. 1, 123–151.
29. Y. Hafouta, *Limit theorems for some skew products with mixing base maps*, Ergodic Theory Dynam. Systems **41** (2019), 241–271.
30. M. Hairer and J. C. Mattingly, *Spectral gaps in Wasserstein distances and the 2D stochastic Navier-Stokes equations.*, Ann. Probab. **36** (2008), no. 6, 2050–2091 (English).
31. M. Hairer and J. C. Mattingly, *Yet another look at Harris' ergodic theorem for Markov chains*, Seminar on Stochastic Analysis, Random Fields and Applications VI (Basel) (R. Dalang, M. Dozzi, F. Russo, eds.), Springer, Basel, 2011, pp. 109–117.
32. B. Hasselblatt and Y. Pesin, *Partially hyperbolic dynamical systems*, Handbook of Dynamical Systems, vol. 1B, Elsevier, Amsterdam, 2006, pp. 1–55 (English).
33. D. Kelly and I. Melbourne, *Smooth approximation of stochastic differential equations*, Ann. Probab. **44** (2016), no. 1, 479–520 (English).
34. B. Kloeckner, *Extensions with shrinking fibers*, Ergodic Theory Dynam. Systems **41** (2021), 1795–1834.
35. A. Korepanov, Z. Kosloff, and I. Melbourne, *Explicit coupling argument for non-uniformly hyperbolic transformations*, Proc. R. Soc. Edinb., Sect. A, Math. **149** (2020), 101–130.
36. C. Liverani, *Decay of correlations*, Ann. Math. (2) **142** (1995), no. 2, 239–301 (English).
37. S. P. Meyn and R. L. Tweedie, *Markov chains and stochastic stability*, Springer-Verlag, London, 1993.
38. M. Nicol, A. Török, and S. Vaienti, *Central limit theorems for sequential and random intermittent dynamical systems*, Ergodic Theory Dynam. Systems **38** (2018), no. 3, 1127–1153.
39. W. Ott, M. Stenlund, and L.-S. Young, *Memory loss for time-dependent dynamical systems*, Math. Res. Lett. **16** (2009), no. 3, 463–475 (English).
40. M. Sambarino, *A (short) survey on dominated splittings*, *Mathematical Congress of the Americas: First Mathematical Congress of the Americas*, American Mathematical Society (AMS), Providence, RI, 2016, pp. 149–183 (English).
41. S. I. Santos and C. Walkden, *Distributional and local limit laws for a class of iterated maps that contract on average*, Stoch. Dyn. **13** (2013), no. 02, 1250019.
42. A. N. Shiryaev and R. P. Boas, *Probability*, 2nd ed., Springer-Verlag, Berlin, Heidelberg, 1995.
43. D. Simmons, *Conditional measures and conditional expectation; Rohlin's disintegration theorem*, Discrete Contin. Dyn. Syst. - A **32** (2012), 2565.
44. M. Stenlund, *Non-stationary compositions of Anosov diffeomorphisms*, Nonlinearity **24** (2011), no. 10, 2991.
45. D. W. Stroock, *An introduction to Markov processes*, 2nd ed., Graduate Texts in Mathematics, vol. 230, Springer, Berlin Heidelberg, 2014.
46. M. Tanzi and L.-S. Young, *Nonuniformly hyperbolic systems arising from coupling of chaotic and gradient-like systems*, Discrete Contin. Dyn. Syst.-A **40** (2020), no. 10, 6015.
47. M. Tanzi and L.-S. Young, *Existence of physical measures in some excitation-inhibition networks*, Nonlinearity **35** (2022), no. 2, 889–915 (English).
48. M. Tsujii, *Physical measures for partially hyperbolic surface endomorphisms*, Acta Math. **194** (2005), no. 1, 37–132 (English).
49. M. Viana, *Stochastic dynamics of deterministic systems*, IMPA, Rio de Janeiro, 1997.
50. C. Villani, *Optimal transport: Old and new*, vol. 338, Springer, Berlin, 2009 (English).
51. C. Walkden and T. Withers, *Invariant graphs of a family of non-uniformly expanding skew products over Markov maps*, Nonlinearity **31** (2018), no. 6, 2726.
52. L.-S. Young, *What are SRB measures, and which dynamical systems have them?* J. Stat. Phys. **108** (2002), no. 5-6, 733–754 (English).
53. L.-S. Young, *Comparing chaotic and random dynamical systems*, J. Math. Phys. **60** (2019), no. 5, 052701.