STATISTICAL PROPERTIES OF MOSTLY CONTRACTING FAST-SLOW PARTIALLY HYPERBOLIC SYSTEMS.

JACOPO DE SIMOI AND CARLANGELO LIVERANI

ABSTRACT. We consider a class of $\mathcal{C}^4$ partially hyperbolic systems on $\mathbb{T}^2$ described by maps $F_\varepsilon(x,\theta) = (f(x,\theta), \theta + \varepsilon\omega(x,\theta))$ where $f(\cdot,\theta)$ are expanding maps of the circle. For sufficiently small ε and ω generic in an open set, we precisely classify the SRB measures for F_ε and their statistical properties, including exponential decay of correlation for Hölder observables with explicit and nearly optimal bounds on the decay rate.

1. INTRODUCTION

There has been a lot of attention lately to the properties of *partially hyperbolic* systems and their perturbations. The main emphasis has been on geometric properties and on stable ergodicity. In the latter field many deep results have been obtained starting with [24, 39]. Nevertheless, it is well known, at least since the work of Krylov [31], that for many applications ergodicity is not sufficient and some type of mixing (usually in the form of effective quantitative estimates) is of paramount importance. Unfortunately, very few results are known regarding stronger statistical properties for partially hyperbolic systems. More precisely, we have some results in the case of mostly expanding central direction [2], and mostly contracting central direction [12, 7]. For central direction with zero Lyapunov exponents (or close to zero) there exist quantitative results on exponential decay of correlations only for group extensions of Anosov maps and Anosov flows [13, 5, 11, 32, 43], but none of them apply to an open class (although some form of rapid mixing is known to be typical for large classes of flows [20, 35]). It would then be of great interest in the field of Dynamical Systems, but also, e.g., for *non-equilibrium Statistical Mechanics*, to extend the class of systems for which statistical properties are well understood. See [38, 33] for a discussion of some aspects of these issues and [40] for an interesting application to non-equilibrium Statistical Mechanics.

Another argument of renewed interest is *averaging theory*, due to new powerful results [16, 18] and, among others, new applications of clear relevance for non-equilibrium Statistical Mechanics [19]. Yet, averaging theory only provides information on a given time scale; a natural and very relevant question is *what happens at longer, possibly infinitely long, time scales.* Such information would be

2000 *Mathematics Subject Classification.* 37A25, 37C30, 37D30, 37A50, 60F17.

Key words and phrases. Averaging, partially hyperbolic, decay of correlations, metastability.

This work would not exist without the many and fruitful discussions which both authors had with Dmitry Dolgopyat, who would very well deserve to be listed among the authors. We thank Ian Morris for suggesting the simple argument we use in the proof of Lemma 6.2. We also thank the anonymous referees for many very helpful suggestions among which to include the examples and discussion in Section 3, for pointing out several inaccuracies in previous versions of this paper and for suggesting a simpler and shorter argument for the proof of Theorem 6.4. The authors are also glad to thank Piermarco Cannarsa, Bastien Fernandez, Christophe Poquet and Ke Zhang for their very useful comments. This work was supported by the European Advanced Grant Macroscopic Laws and Dynamical Systems (MALADY) (ERC AdG 246953) and by NSERC. Both authors are pleased to thank the Fields Institute in Toronto, Canada, (where this work started) for the excellent hospitality and working conditions provided during the spring semester 2011.

encoded in the SRB measure and its statistical properties. Hence we have a natural connection with the above mentioned open problem in partially hyperbolic systems (as indeed the slow variables can often be considered as central directions).

The above program can be carried out for deterministic systems subject to small random perturbations (e.g. stochastic differential equations with vanishing diffusion coefficient), where the fast variable (modeled by Brownian Motion), is (in some sense) infinitely fast [21]. In our setting, on the one hand the motion of the fast variable is deterministic, although chaotic; on the other hand, the motion of the slow variable is not hyperbolic (hence one cannot implement strategies based on the strong chaoticity of the unperturbed motion and the essential irrelevance of the perturbation, where many powerful technical tool are available, starting with [27]). It is then not so surprising that a preliminary step needed to carry out the above program is to establish, in a very precise technical sense, to which extent the motion of the fast variable can be confused with the motion of a random variable. In particular, this requires to go well beyond the known results on averaging and deviations from the average that can be found in [28, 16]. One needs the analog of a Local Central Limit Theorem for the process of the fluctuations around the average. This is in itself a non trivial endeavor which has been first accomplished, for a simple but relevant class of systems, in [9].

Finally, in analogy with the stochastic case, see [21], one can expect *metastable* behavior.[1] Indeed, metastability is a phenomenon that has been widely investigated in the stochastic setting, see [36] for a detailed account. Yet, to our knowledge, no results whatsoever exist in the deterministic setting. The strongest results in such a direction can be found in [30] where it is proven, for a fairly large class of systems, that the transition between basins of attractions takes place only at exponentially long time scales, thus one has a clear indication of the existence of, at least, two time scales in such systems. Yet, the results in [30] are not sufficient to investigate the longer times needed to establish a full metastability scenario (in the sense of Footnote 1). It is then natural to ask if metastability results hold in the present deterministic setting. Of course, to answer to such questions, one needs to combine good *Large Deviation Estimates*[2] with a precise quantitative understanding of the mixing properties of the local dynamics. This is the topic of this paper and it clarifies the connection of metastability with the above mentioned general problems. Accordingly, metastability (together with partial hyperbolicity, non-equilibrium statistical mechanics and averaging) constitutes a fourth natural and important line of research among the ones that motivate and converge in this paper.

To carry out the above research program it turns out to be necessary a preliminary understanding of the long time properties of the averaged motion. In general, this is an impossible task, since the averaged system can be essentially any ordinary differential equation (ODE). To simplify matters, as a first step, we consider the simplest possible averaged dynamics: a one dimensional ODE on the circle with finitely many, non degenerate, equilibrium points.

[1] By a *metastable* system here we mean a situation in which two time scales are present: one, the short one, in which the system seems to have several invariant measures, and hence to lack ergodicity, and a longer time scale in which it turns out that the system has indeed only one relevant, mixing, invariant measure. This can be seen experimentally by the presence of two time scales in the decay of correlations.

[2] These were first derived in [30]. But a much more refined and quantitative version can be found in [9].

2. The model and the results

2.1. **The model.** Let us now introduce the class of systems we are going to investigate. For $\varepsilon > 0$ we consider the family of maps $F_\varepsilon \in \mathcal{C}^4(\mathbb{T}^2, \mathbb{T}^2)$ defined by

$$F_\varepsilon(x, \theta) = (f(x, \theta), \theta + \varepsilon\omega(x, \theta) \mod 1) \tag{2.1}$$

where $f \in C^4(\mathbb{T}^2, \mathbb{T})$ and $\omega \in C^4(\mathbb{T}^2, \mathbb{R})$. We assume that $f(\cdot, \theta) = f_\theta : \mathbb{T} \to \mathbb{T}$ is an orientation-preserving expanding map for each $\theta \in \mathbb{T}$; moreover, by possibly replacing F_ε by a suitable iterate, we will always assume that $\partial_x f \geq \lambda \geq 2$.

Remark 2.1. *In the sequel, we will take ε to be fixed and sufficiently small depending on f and ω. We could indeed regard (2.1) as an arbitrary perturbation of the map $\widetilde{F}(x, \theta) = (\tilde{f}(x, \theta), \theta)$ by $\varepsilon(g(x, \theta), \omega(x, \theta))$; in fact, since a sufficiently small perturbation of a family of expanding maps is still a family of expanding maps, one could always set $f = \tilde{f} + \varepsilon g$. However, in order to treat this slightly more general case, we would need to show that our conditions on the smallness of ε depend on ω uniformly in a neighborhood of $\tilde{f}$. We do not pursue this for sake of simplicity.*

Since f_θ are expanding maps of the circle, there exists a unique family of absolutely continuous (SRB) f_θ-invariant probability measures whose densities we denote by ρ_θ. By our regularity assumptions on F_ε it follows (see e.g. [23, Section 8]) that ρ_θ is a $\mathcal{C}^3$-smooth family of $\mathcal{C}^3$-densities. Let us now define $\bar{\omega}(\theta) = \int_{\mathbb{T}} \omega(x, \theta)\rho_\theta(x)dx$. Observe that our earlier considerations concerning the smoothness of the family ρ_θ imply that $\bar{\omega} \in \mathcal{C}^3(\mathbb{T})$.

Our first standing assumption reads

(A1) $\bar{\omega}$ has a non-empty discrete set of non-degenerate zeros.

In particular, we assume the set of zeros to be given by $\{\theta_{i,\pm}\}_{i \in \mathbb{Z}/n_Z\mathbb{Z}}$ with $n_Z \in \mathbb{N}$, $\bar{\omega}'(\theta_{i,+}) > 0$ and $\bar{\omega}'(\theta_{i,-}) < 0$; we assume, having fixed an orientation of $\mathbb{T}$, that the indexing is so that for any k, $\theta_{k,+} < \theta_{k,-} < \theta_{k+1,+}$, where all indices k are taken $\bmod n_Z$.

In Section 4 we will see that the map F_0 has an invariant center distribution generated by vectors $(s_*(x, \theta), 1)$. Let us now denote with ψ_* the directional derivative of ω in the center direction, or, more precisely, with respect to the vector $(s_*(x, \theta), 1)$, that is:

$$\psi_*(x, \theta) = \partial_x\omega(x, \theta)s_*(x, \theta) + \partial_\theta\omega(x, \theta). \tag{2.2}$$

Let us also define its average $\bar{\psi}_*(\theta) = \int_{\mathbb{T}} \psi_*(x, \theta)\rho_\theta(x)dx$. As it will be made clear by (4.7) and subsequent discussion, $1 + \varepsilon\psi_*$ is the one step-contraction (or expansion) in the center direction. Remark that the system is non-uniformly hyperbolic and it is far from obvious how to compute the central Lyapunov exponent for Lebesgue-a.e. point. Our second assumption will, eventually, allow us to prove that the center Lyapunov exponent is Lebesgue-a.s. negative:

(A2) $\max_{k \in \{1, \cdots, n_Z\}} \bar{\psi}_*(\theta_{k,-}) = -1$.

Remark 2.2. *Observe that if $\max_{k \in \{1, \cdots, n_Z\}} \bar{\psi}_*(\theta_{k,-}) < 0$, it is always possible to rescale*[3] *ω and ε so that (A2) holds. In other words, the -1 on the right hand side is just a normalization which can be achieved without loss of generality.*

Remark 2.3. *Observe moreover that one can explicitly compute an arbitrarily precise approximation of ψ_* (see (4.10) and Remark 4.1). Thus, in view of the above remark, Assumption (A2) is in principle explicitly checkable for a given map.*

[3] That is, for $\varrho_r \neq 0$, we let $\omega \mapsto \varrho_r\omega$ and $\varepsilon \mapsto \varrho_r^{-1}\varepsilon$ so that the product $\varepsilon\omega$ is left unchanged, together with all other dynamically defined quantities. Observe that under this rescaling, (2.2) gives $\psi_* \mapsto \varrho_r\psi_*$.

The above condition is not optimal, it implies that the center direction is mostly contracting on average (see Lemma 7.2) in a neighborhood of *every* sink. Of course, a negative Lyapunov exponent in the center direction could also emerge from the interaction between different sinks but this would be much harder to investigate.

Remark 2.4. *It is quite possible that (A2) is not necessary and our Main Theorem holds also in the case of zero or positive central Lyapunov exponent. Yet, its proof clearly would require a different approach and it remains the subject of further studies, see also Section 11 on this issue.*

Remark 2.5. *Observe that in case $\partial_\theta f$ is identically 0, then (A2) follows by (A1) since $s_* = 0$, $\partial_\theta \rho_\theta = 0$ and hence $\bar\omega'(\theta) = \bar\psi_*(\theta)$. Thus, in such a case, the center Lyapunov exponent turns out to be determined by the averaged system and it is always negative. This is not true in general if $\partial_\theta f \neq 0$ and motivates our need for Assumption (A2). In particular, in Section 3.4 we provide an explicit example of a system for which (A2) does not hold and $\bar\omega'(\theta_{k,-}) < 0 < \bar\psi_*(\theta_{k,-})$, i.e., despite the fact that the $\theta_{k,-}$ is a sink for the averaged dynamics, center vectors close to $\{\theta = \theta_{k,-}\}$ are, on average, expanded by the dynamics. This phenomenon is related to the potential lack of absolute continuity of the central foliation of partially hyperbolic systems (see e.g. [41]).*

In order to state our Main Theorem, it is necessary to introduce a few more definitions, which force us to take a (very minor) detour through non-smooth analysis (see e.g. [6]). A Lipschitz function $h \in \mathcal{C}^{\mathrm{Lip}}([0,T],\mathbb{T})$ is said to be a (θ^0,θ^1)-*path* if it satisfies the boundary conditions $h(0) = \theta^0$, $h(T) = \theta^1$; T will be referred to as the *length* of h. Recall that Rademacher's Theorem implies that a Lipschitz function h is differentiable everywhere except on a set of zero Lebesgue measure which we denote with E_h. For each $s \in [0,T]$ let us define the *Clarke generalized derivative* of h as the set-valued function:

$$\boldsymbol{\partial} h(s) = \mathrm{hull}\{\lim_{k\to\infty} h'(s_k) \, : \, s_k \to s \text{ and } \{s_k\} \subset [0,T] \setminus E_h\}.$$

The set $\boldsymbol{\partial} h(s)$ is compact and non-empty for any $s \in [0,T]$ (see e.g. [6, Proposition 2.1.5]) and so is its graph, i.e. the set $\bigcup_{s\in[0,T]}\{s\} \times \boldsymbol{\partial} h(s) \subset [0,T]\times\mathbb{R}$ (this follows from the definition and from a standard Cantor diagonal argument). Moreover if $s \notin E_h$, then $h'(s) \in \boldsymbol{\partial} h(s)$ and if h' is continuous at s we have $\boldsymbol{\partial} h(s) = \{h'(s)\}$ (see [6, Proposition 2.3.1]).

We say that a path h of length T is *admissible* if for any $s \in [0,T]$, $\boldsymbol{\partial} h(s) \subset \mathrm{int}\,\Omega(h(s))$, where for any $\theta \in \mathbb{T}$, we define the (non-empty, convex and compact) set

$$\Omega(\theta) = \{\mu(\omega(\cdot,\theta)) \,|\, \mu \text{ is a } f_\theta\text{-invariant probability}\}. \tag{2.3}$$

Observe that the set of admissible paths might be empty if $\mathrm{int}\,\Omega(\theta) = \emptyset$ for some θ; this potential degeneracy is excluded by our next assumption. Recall that a function $\phi \in \mathcal{C}^0(\mathbb{T})$ is said to be a *(continuous) coboundary (with respect to a map $f : \mathbb{T} \to \mathbb{T}$)* if there exists $\beta \in \mathcal{C}^0(\mathbb{T})$ so that

$$\phi = \beta - \beta \circ f.$$

Two functions $\phi_1, \phi_2 \in \mathcal{C}^0(\mathbb{T})$ are said to be *cohomologous (with respect to f)* if their difference $\phi_2 - \phi_1$ is a coboundary (with respect to f). Our standing non-degeneracy[4] assumption is:

(A3) for each $\theta \in \mathbb{T}$, the function $\omega(\cdot,\theta)$ is not cohomologous to a constant function with respect to f_θ.

[4] In [9] it is shown that assumption (A3) is in fact generic in $\mathcal{C}^2$. Observe moreover that the condition can be easily checked on periodic orbits.

We finally state two more conditions:

(A4) there exists $i \in \{1, \cdots, n_Z\}$ so that for any $\theta \in \mathbb{T}$, there exists an admissible $(\theta, \theta_{i,-})$-path. We can always assume, without loss of generality, that $i = 1$.

Observe that, under conditions (A1) and (A3), condition (A4) is trivially satisfied if $n_Z = 1$ (see Section 6.4). In cases where (A4) does not hold, we can still obtain interesting results under the following additional condition:

(A5) the set of zeros $\{\theta_{i,\pm}\}_{i=1,\cdots,n_Z}$ of $\bar{\omega}$ cuts $\mathbb{T}$ in $2n_Z$ open intervals: any such interval J satisfies one of the following two properties
  i. for any $\theta \in J$, $0 \in \operatorname{int}\Omega(\theta)$
  ii. there exists $\theta \in J$ so that $0 \notin \Omega(\theta)$.

The above condition is simply a non-degeneracy condition; more precisely (A5) is an open and dense property among the maps enjoying (A1), (A2) and (A3) (see Lemma 6.12). Section 2.3 contains further comments on this last assumption.

2.2. The result.

We are now finally ready to state our main result.

Main Theorem. *Under assumptions (A1), (A2), (A3) and (A5), if $\varepsilon > 0$ is sufficiently small, then F_ε admits at most n_Z SRB measures.*

Under assumptions (A1), (A2), (A3) and (A4), if $\varepsilon > 0$ is sufficiently small, then F_ε admits a unique SRB measure μ_ε. This measure enjoys exponential decay of correlations for Hölder observables. More precisely: there exist $C_1, C_2, C_3, C_4 > 0$ (independent of ε) such that, for any $\alpha \in (0, 3]$ and $\beta \in (0, 1]$, any two functions $A \in \mathcal{C}^\alpha(\mathbb{T}^2)$ and $B \in \mathcal{C}^\beta(\mathbb{T}^2)$:

$$|\mathrm{Leb}(A \cdot B \circ F_\varepsilon^n) - \mathrm{Leb}(A)\mu_\varepsilon(B)| \le C_1 \sup_\theta \|A(\cdot, \theta)\|_{\mathcal{C}^\alpha} \sup_x \|B(x, \cdot)\|_{\mathcal{C}^\beta} e^{-\alpha\beta c_\varepsilon n},$$

where

$$c_\varepsilon = \begin{cases} C_2 \varepsilon / \log \varepsilon^{-1} & \text{if } n_Z = 1, \\ C_3 \exp(-C_4 \varepsilon^{-1}) & \text{otherwise.} \end{cases} \tag{2.4}$$

The proof of our Main Theorem will be given in Section 9.4. In fact, in Section 9.4 we prove a slightly stronger version (detailing the exact number and properties of the SRB measures in the case (A4) does not hold, but (A5) does) which, to be properly stated, needs the introduction of several extra notations (see Theorem 9.9 and Corollary 9.10 for more details). Next, we provide a number of remarks to clarify and put into context the above result.

Remark 2.6. *There is a long lasting controversy regarding the definition of SRB measures (see e.g. [47]). Since endomorphisms do not have an unstable foliation, we will follow common practice (see e.g. [14, Corollary 2]) and say that μ_ε is an SRB measure if it is F_ε-invariant and its* ergodic basin

$$B(\mu_\varepsilon) = \left\{ p \in \mathbb{T}^2 \,:\, \frac{1}{n}\sum_{k=0}^{n-1} \delta_{F_\varepsilon^k(p)} \to \mu_\varepsilon \text{ weakly as } n \to \infty \right\}$$

has positive Lebesgue measure. These measures are also called physical measures.

Remark 2.7. *As a particular (and non generic) case of (A5) (i.e. every interval J satisfies property i) let us introduce the following condition*[5]

(A4*) *for any $\theta \in \mathbb{T}$, $0 \in \operatorname{int}\Omega(\theta)$;*

[5] In [30], ω is said to be *complete* at θ if this condition holds at θ.

Condition (A4) immediately implies (A4) (it is strictly stronger and assuming it in our Main Theorem would imply the existence of a unique SRB measure with ε-dense support). Most importantly, it can in principle be checked in concrete examples as it suffices*[6] *to find, for every $\theta \in \mathbb{T}$, two periodic orbits of f_θ so that the average of $\omega(\cdot, \theta)$ is positive on one of them and negative on the other one. Moreover, it is obvious to observe that, for any given F_0, the set $\{\omega : (A4^*) \text{ holds}\}$ contains an open set in the $\mathcal{C}^4$-topology. Finally, it is immediate to check that Condition (A4*) also implies Condition (A3).*

Remark 2.8. *A natural question is whether the values of c_ε in (2.4) are optimal or not. The answer is "essentially yes". To clarify this, in Section 3 we give some explicit examples to which our Theorem applies and we provide a lower bound for the decay of correlations in such examples. Also we take the opportunity to compare our situation with the case of small stochastic perturbations discussed by Wentzell-Freidlin [21] uncovering both strong similarities and fundamental differences. We summarize our findings in Remarks 3.2, 3.3, 3.4 and 3.5.*

Remark 2.9. *For simplicity our Main Theorem is stated for the Lebesgue measure. In fact it holds for a much wider class of measures, i.e. measures that can be obtained as weak limit of standard families (see Section 5 for details). Such measures include, in particular, SRB measures as a special example. Also, note that for the SRB measure it is certainly possible for the decay of correlations to be much faster also in the case $n_Z > 1$ since, the mass being already distributed in equilibrium, one may not have metastable states.*

Remark 2.10. *Several related results are available in the literature: first of all Tsujii in [42] proves that a generic*[7] *family of type (2.1) has a finite number of SRB measures absolutely continuous with respect to Lebesgue. We believe that such a result applies to the present context, as Tsujii's genericity condition should reduce to our hypotheses (A3) and (A5), but this is not obvious to check. Next, exponential decay of correlations has been proven in the case of* mostly expanding *and* mostly contracting *center foliations. The mostly contracting case is studied in [3, 7, 8, 12], but see [34] for a recent overview on the subject; the mostly expanding case in [1, 2, 22]. Unfortunately, to apply such results it is necessary to either show that the central Lyapunov exponent is negative or to estimate the Lebesgue measure of the points that have not expanded up to time n. On the one hand this is rather tricky to do,*[8] *on the other hand the estimates on the rate of correlation decay provided by these papers are not quantitative. In particular, such results do not provide any information on how the rate of decay depends on ε, hence they completely miss the issue of metastability. On the contrary, Kifer's papers [29, 30] address very clearly the metastability issue, but, as already remarked, the results there do not allow to investigate the longer time scales, that is the SRB measures and their statistical properties.*

Finally let us remark that, contrary to most of the current literature (which discusses "generic" systems), our conditions are explicit and, often, checkable by just studying few periodic orbits of the system.

2.3. Overview and structure of the paper.

Let us now sketch the strategy of our proof and outline the structure of this paper. In Section 3 we present some explicit class of examples to which our Main Theorem applies and an interesting

---

[6] The equivalence holds since the measures supported on periodic orbits are weakly dense in the set of the invariant measures [37].

[7] The exact meaning of *generic* is a bit technical and we refer to [42] for the details.

[8] We essentially prove that the Lyapunov exponents are negative, but this takes a good part of this paper.

case to which it does not. For some simple situations we compute how the SRB measure looks like, we investigate metastability and compare it with the Wentzell-Freidlin case.

Our system is an example of fast-slow system (see Section 4): averaging theory (see Section 6) implies that the slow variable θ undergoes a diffusion around the dynamics of the averaged system, which is described by the ODE $\dot{\theta} = \bar{\omega}(\theta)$. Assumption (A3) implies, by the results of [9], that the diffusion is non-degenerate and indeed satisfies precise Large Deviation Estimates and a Local Central Limit Theorem. In turn, Assumption (A1) implies that the averaged system has n_Z pairs of sinks and sources: the set $\{\theta_{k,+}\}$ partitions (mod 0) the torus $\mathbb{T}$ in n_Z intervals $I_{k,-} = [\theta_{k,+}, \theta_{k+1,+}]$, whose interiors are the basins of attraction of $\theta_{k,-}$, i.e. the averaged dynamics pushes every point in int $I_{k,-}$ to $\theta_{k,-}$ exponentially fast.

In particular, if $n_Z = 1$, then the averaged dynamics pushes almost every initial condition to the unique sink θ_-. Introducing a suitable notion of standard pairs (see Section 5), we can prove that the true dynamics closely follows the averaged one with high probability (Section 7). Thanks to this fact we can establish a coupling argument (see Section 8 for the basic facts on coupling, Section 9 for the setup of the argument and Section 10 for proofs and the details) among sufficiently close standard pairs: this implies exponential decay of correlations with a rate that is essentially given by the time-scale of the averaged motion.

On the other hand, if $n_Z > 1$, then the averaged dynamics will push initial conditions belonging to different basins to the corresponding sink; we thus need to rely on large deviations to prove that standard pairs (i.e. mass) are allowed to move from one basin to another, although with very small probability. Such events are called *adiabatic transitions* and their typical time-scale is exponentially small in ε^{-1}. If the diffusion were purely stochastic and unbounded (i.e., in a Wentzell-Freidlin system [21]), then all transitions between different basins would be allowed. On the contrary, in our deterministic realization, some of the transitions might not be actually possible since the "noise" is bounded (see Section 3.2 for an explicit example of this phenomenon and Section 6.4 for an accurate description). Hence some sinks could act as traps for the real dynamics: this constitutes an obstruction to ergodicity. We need assumption (A4) to guarantee that no such obstructions occur. In case (A4) does not hold, we need (A5) to exclude borderline situations in which the number of different SRB measures could change in an arbitrarily small neighborhood of F_0. Finally, in Section 11 we discuss the strengths and shortcomings of our approach and we illustrate several open problems that must be addressed to push forward the research program started by this paper.

Notational remark 2.11. *We will henceforth fix f and ω to satisfy all properties enumerated before; all values that we declare to be* constant *below will depend on this choice. We will often use $C_\#, c_\#$ to designate some constants (again possibly depending on f and ω), whose actual value is irrelevant and can thus change form one instance to the next.*

3. Examples

To provide a better understanding of the results obtained in the present paper we first (Sections 3.1, 3.2 and 3.3) discuss in some detail a few examples to which our theory applies. Then (in Section 3.4) we briefly mention an example that does not satisfy our conditions since the central direction seems to be, unexpectedly, mostly expanding. Along the way, we take the occasion to carry out a precise comparison with the case of small random perturbations of a dynamical system (the so-called Wentzell–Freidlin systems). The conclusions of such a comparison are summarized in Remarks 3.2, 3.3, 3.4 and 3.5.

Carrying out explicit computations in a specified example can be rather laborious. We thus prefer to give examples belonging to a particularly simple class of systems in which such explicit computations can be done fairly easily: skew-products[9] over the doubling map:

$$F_\varepsilon(x,\theta) = (2x, \theta + \varepsilon\bar{\omega}(\theta) + \varepsilon\hat{\omega}(x)) \mod 1, \tag{3.1}$$

where, consistently with our notation, $\int_{\mathbb{T}} \hat{\omega}(x)dx = 0$. Also, to further simplify matters, we will always assume that $\int_{\mathbb{T}} \bar{\omega}(\theta)d\theta = 0$, so that the associated Wentzell–Freidlin system is reversible.

Note that in this case the fast dynamics does not depend on θ, hence making the example very simple, although still far from trivial. Thus the SRB measure for the fast dynamics is the Lebesgue measure m for every $\theta \in \mathbb{T}$. Recall that the limit theorems proved in [9] describe statistical properties of the process θ_n where $(x_n, \theta_n) = F_\varepsilon^n(x_0, \theta_0)$ and the initial conditions (x_0, θ_0) are distributed according to an (invariant) class of measures called *standard pairs*. Such measures will only be properly defined in Section 5: for the time being, the reader can pretend such initial conditions to be given by $\theta_0 =$ constant and x_0 distributed with a smooth density. As already remarked, averaging theory implies that the process $\theta_\varepsilon(t) = \theta_{\varepsilon^{-1}t}$ is close to the solution of the ODE $\dot{\theta} = \bar{\omega}(\theta)$. If we consider small random perturbations of such an ODE we obtain the corresponding Wentzell–Freidlin scenario, which should describe more accurately our rescaled process $\theta_\varepsilon(t)$,[10]

$$dw = \bar{\omega}(w)dt + \sqrt{\varepsilon}\hat{\boldsymbol{\sigma}}(w)dB \tag{3.2}$$

where B is the standard Brownian motion. Note that, due to the fact that we have a skew product and the simple form of ω, in this particular case $\hat{\boldsymbol{\sigma}}(\theta) = \hat{\boldsymbol{\sigma}}$ is independent of θ.

3.1. **Skew-products over the doubling map–one sink.** To further simplify matters, we assume that $\hat{\omega}(0) = 3$ and $\|\bar{\omega}\|_\infty \leq 1$. Since the Dirac measure δ_0 is an invariant measure for the doubling map, we have that $\hat{\omega}$ cannot be a coboundary, hence (A3) is satisfied. We assume (A1). Since $\partial_\theta f = 0$, we have (see Remark 2.5) $\psi_*(x,\theta) = \bar{\omega}'(\theta) = \bar{\psi}_*(\theta)$ and we can then assume, without loss of generality, that (A2) is satisfied.

Let us start with the case in which $\bar{\omega}$ has only one sink θ_-, hence $\bar{\omega}'(\theta_-) = -1$ by (A2). Observe moreover that assumption (A4) is automatically verified since we have only one sink. First of all let us understand the SRB measure μ_ε. Let $\hat{H}$ be a suitable neighborhood of θ_- (see Section 6.3 for more details) and let $1-p = \mu_\varepsilon(\hat{H})$; setting $B_\varepsilon = [\theta_- - C_\#\sqrt{\varepsilon}\log\varepsilon^{-1}, \theta_- + C_\#\sqrt{\varepsilon}\log\varepsilon^{-1}]$, $q = \mu_\varepsilon(\hat{H}\setminus B_\varepsilon)$. Then Lemma 7.5 implies that $p \leq \varepsilon^\beta$. Lemma 7.4 implies that at least $\frac{2}{3}$ of the mass of a standard pair in $\hat{H}$ moves to B_ε is a time of order $\varepsilon^{-1}\log\varepsilon^{-1}$. While [9, Theorem 2.8] implies that at time $T\varepsilon^{-1}$ the mass on any standard pair ℓ in B_ε will be distributed according to a Gaussian centered in $\theta_- + e^{-2T}(\theta_\ell - \theta_-)$ and with variance $\frac{1-e^{-4T}}{2}\varepsilon\hat{\boldsymbol{\sigma}}^2$ apart from a mass $\varepsilon^{2\beta}$, eventually making β smaller. Iterating this $\log\varepsilon^{-1}$ times we have that the mass is distributed according to a Gaussian centered in θ_- and with variance $\varepsilon\hat{\boldsymbol{\sigma}}^2$ apart from a mass ε^β. This implies that,

$$1 - p - q \geq \frac{2}{3}q - (1-p-q)(1-\varepsilon^\beta) - C_\#\varepsilon^\beta$$

[9] The reader can easily work along the lines we suggest and construct more elaborate examples which are not skew-products, and yet feature all properties described in our examples.

[10] It is possible to make this correspondence quantitatively precise for times of order $\varepsilon^{-\alpha}$ for some $\alpha > 0$. We refrain from doing it to keep the length of the paper under control and we postpone it to further work.

hence $q \le C_\# \varepsilon^\beta$. It follows that μ_ε consists of a Gaussian of variance $\varepsilon\hat{\boldsymbol\sigma}^2$ centered at θ_-, a part from a mass of order ε^β.[11]

Now that we have a good understanding of the SRB measure μ_ε, we can address the issue of the decay of correlations, in particular it is natural to wonder if the results of our Main Theorem is optimal or not. Let us choose A supported on a δ neighborhood of $\{\theta_+\}\times\mathbb{T}$, where $\bar\omega(\theta_+)=0$ and $\bar\omega'(\theta_+)>0$, and B supported in a δ neighborhood of $\{\theta_-\}\times\mathbb{T}$, then we ask what is the maximal c_ε such that

$$\begin{aligned}|\mathrm{Leb}(A\cdot B\circ F_\varepsilon^n)-\mathrm{Leb}(A)\mu_\varepsilon(B)| &\le C_1\sup_\theta\|A(\cdot,\theta)\|_{\mathcal{C}^2}\sup_x\|B(x,\cdot)\|_{\mathcal{C}^2}\exp(-c_\varepsilon n)\\ &\le C_\#\exp(-c_\varepsilon n).\end{aligned}$$

If we choose δ small enough, there will be a distance bigger than $\frac12|\theta_- - \theta_+|$ between the support of A and B. Moreover, $\mathrm{Leb}(A)=\delta$, while $\mu_\varepsilon(B)\ge 1-C_\#\varepsilon^\beta$. In addition, since θ can move only of steps of order ε, we will have that $A\cdot B\circ F_\varepsilon^n=0$ for all $n\le C_\#\varepsilon^{-1}$. Hence, $\frac12\le C_\#\exp(-c_\varepsilon\varepsilon^{-1})$ which implies that it must be $c_\varepsilon\le C_\#\varepsilon$.

We have thus seen that in the present case our Main Theorem is, at least, close to optimal. Whether or not the $\log\varepsilon^{-1}$ is really there, or it is an artifact of our method of proof, it remains to be seen.

To gain some more insight, let us compare the above situation with the Wentzell–Freidlin system (3.2). First of all, note that, taking advantage of the fact that we are in dimension one, we can write the generator associated to the process as[12]

$$\begin{aligned}L_\varepsilon\varphi&=\bar\omega\varphi'+\frac{\varepsilon}{2}\hat{\boldsymbol\sigma}^2\varphi''=\frac{\varepsilon}{2\rho_\varepsilon}(\hat{\boldsymbol\sigma}^2\rho_\varepsilon\varphi')',\\ \rho_\varepsilon(\theta)&=Z\hat{\boldsymbol\sigma}^{-2}e^{2\varepsilon^{-1}\int_0^\theta\frac{\bar\omega}{\hat{\boldsymbol\sigma}^2}}\ ;\quad\int_{\mathbb{T}}\rho_\varepsilon=1.\end{aligned}\tag{3.3}$$

One can then easily check that L_ε is reversible with respect to the probability measure $d\nu_\varepsilon=\rho_\varepsilon dx$. Thus ν_ε is the invariant probability measure of (3.2) and it is a direct computation to see that its Wasserstein distance from μ_ε is less that $C_\#\varepsilon^\beta$. Next, we need an estimate of the spectral gap of L_ε in $L^2(\mathbb{T},\nu_\varepsilon)$. As we were unable to locate it in the literature, we provide it here (and since we are at it, we do it for the general case in which $\hat{\boldsymbol\sigma}$ is not constant, but still strictly positive).

Lemma 3.1. *If $\bar\omega,\hat{\boldsymbol\sigma}\in\mathcal{C}^2(\mathbb{T},\mathbb{R})$, $\bar\omega$ has only two non degenerate zeroes θ_-,θ_+ and $\inf\hat{\boldsymbol\sigma}>0$, then there exists $c_0>0$ such that, for each ε small enough, L_ε has a spectral gap larger than c_0.*

Proof. We follow the logic used in [44, Theorem 1.2, Appendix A.19]. Setting $V_\varepsilon(\theta)=-2\int_0^\theta\frac{\bar\omega}{\varepsilon\hat{\boldsymbol\sigma}^2}+2\log\hat{\boldsymbol\sigma}-\log Z$, we have $\rho_\varepsilon=e^{-V_\varepsilon}$. Next, let

$$W_\varepsilon=\hat{\boldsymbol\sigma}^2(V_\varepsilon')^2/2-(\hat{\boldsymbol\sigma}^2V_\varepsilon')',$$

then for each $\varphi\in\mathcal{C}^2$ we have (see [10, Proof of Theorem 6.2.21])[13]

$$\int_{\mathbb{T}}W_\varepsilon\varphi^2\rho_\varepsilon\le 2\int_{\mathbb{T}}\hat{\boldsymbol\sigma}^2(\varphi')^2\rho_\varepsilon.\tag{3.4}$$

---

[11] Of course, we mean this in the sense of [9, Theorem 2.8], on a scale smaller than ε the SRB could have some complicated fine structure. This issue is here left open.

[12] Note that, by hypotheses, $\rho_\varepsilon(1)=Z\hat{\boldsymbol\sigma}^{-2}e^{2\varepsilon^{-1}\int_0^1\frac{\bar\omega}{\hat{\boldsymbol\sigma}^2}}=Z\hat{\boldsymbol\sigma}^{-2}=\rho_\varepsilon(0)$, hence ρ_ε is a smooth function on $\mathbb{T}$.

[13] For the reader convenience, here is how to argue: compute using

$$0\le\int_{\mathbb{T}}\left[\hat{\boldsymbol\sigma}\left(e^{-V/2}\varphi\right)'\right]^2\ \text{and}\ \int_{\mathbb{T}}\hat{\boldsymbol\sigma}^2V'\varphi'\varphi e^{-V}=\frac12\int_{\mathbb{T}}\left[(\hat{\boldsymbol\sigma}V')^2-(\hat{\boldsymbol\sigma}^2V')'\right]\varphi^2e^{-V}.$$

Note that the right hand side of (3.4) is nothing else than the Dirichlet form associated to L_ε. In addition, $W_\varepsilon = \frac{2}{\varepsilon^2 \hat{\boldsymbol{\sigma}}^2}\bar\omega^2 - 4\frac{\bar\omega\hat{\boldsymbol{\sigma}}'}{\varepsilon\hat{\boldsymbol{\sigma}}} + \frac{2}{\varepsilon}\bar\omega' + 2(\hat{\boldsymbol{\sigma}}')^2 - 2(\hat{\boldsymbol{\sigma}}'\hat{\boldsymbol{\sigma}})'$ is always positive apart from a neighborhood of θ_-. Indeed, let $A_\varepsilon = [\theta_- - a\sqrt{\varepsilon}, \theta_- - a\sqrt{\varepsilon}]$, then, if a is chosen large enough, $\inf_{A_\varepsilon^c} W_\varepsilon \geq \varepsilon^{-1}$. Moreover, if $\theta, \theta' \in A_\varepsilon$, then

$$\frac{\rho_\varepsilon(\theta)}{\rho_\varepsilon(\theta')} \leq e^{c_\# \varepsilon^{-1}|\theta-\theta'|^2} \leq e^{c_\# a^2}. \tag{3.5}$$

Let K to be chosen later (large enough) and choose a so that $\int_{A_\varepsilon^c} \rho_\varepsilon \leq K^{-1}$. Note that there exist a constant $C_a > 0$ such that $C_a^{-1}\varepsilon^{-\frac{1}{2}} \leq \rho_\varepsilon(\theta) \leq C_a\varepsilon^{-\frac{1}{2}}$, for all $\theta \in A_\varepsilon$.

The last needed ingredient is the standard Poincaré inequality in A_ε: there exists $b > 0$ such that, for all ε small enough,[14]

$$\int_{A_\varepsilon} \varphi^2 \rho_\varepsilon \leq b\varepsilon \int_{\mathbb{T}} (\varphi')^2 \rho_\varepsilon + 2\left(\int_{A_\varepsilon} \varphi\rho_\varepsilon\right)^2.$$

Thus, for each $\varphi \in \mathcal{C}^2$, we have, for K large enough,

$$\begin{aligned}\int_{\mathbb{T}} \varphi^2 \rho_\varepsilon &\leq C_\# \varepsilon \int_{\mathbb{T}} W_\varepsilon \varphi^2 e^{-V_\varepsilon} + \frac{K}{4}\int_{A_\varepsilon} \varphi^2\rho_\varepsilon \\ &\leq C_\#\varepsilon \int_{\mathbb{T}} \hat{\boldsymbol{\sigma}}^2 (\varphi')^2\rho_\varepsilon + \frac{K}{2}\left(\int_{A_\varepsilon}\varphi\rho_\varepsilon\right)^2.\end{aligned}$$

To conclude, assume that $\int_{\mathbb{T}} \varphi\rho_\varepsilon = 0$, hence $\int_{A_\varepsilon}\varphi\rho_\varepsilon = -\int_{A_\varepsilon^c}\varphi\rho_\varepsilon$. Thus

$$\left(\int_{A_\varepsilon}\varphi\rho_\varepsilon\right)^2 \leq \nu_\varepsilon(A_\varepsilon^c)\int_{A_\varepsilon^c}\varphi^2\rho_\varepsilon \leq K^{-1}\int_{\mathbb{T}}\varphi^2\rho_\varepsilon.$$

We are thus ready to compute the spectral gap: let $\varphi \in \mathcal{C}^2$ such that $\nu_\varepsilon(\varphi) = 0$, then

$$\int_{\mathbb{T}} \varphi(-L_\varepsilon\varphi)\rho_\varepsilon = \frac{\varepsilon}{2}\int_{\mathbb{T}} \hat{\boldsymbol{\sigma}}^2(\varphi')\rho_\varepsilon \geq C_\# \int_{\mathbb{T}} \varphi^2\rho_\varepsilon.$$ □

The above Lemma implies that,

$$|\nu_\varepsilon(A(w(0))B(w(t)) - \nu_\varepsilon(A)\nu_\varepsilon(B)| \leq C_\#\|A\|_{L^2(\nu_\varepsilon)}\|B\|_{L^2(\nu_\varepsilon)}e^{-c_0 t}.$$

If the Wentzell–Freidlin process (3.2) is a good predictor of what happens for the process $\theta_{\varepsilon^{-1}t}$ (as we, in this case, conjecture), then the factor $\log\varepsilon^{-1}$, in our estimate for the rate decay of correlations (2.4), should be absent if one starts from the SRB measure rather then the Lebesgue measure. Note however that if we start from a measure non absolutely continuous with respect to ν_ε (e.g. a δ at some θ_0) or with an exponentially large Radon-Nikodyn derivative (e.g. Lebesgue), then it will take a time at least $C_\#\log\varepsilon^{-1}$ before the measure becomes uniformly absolutely continuous with respect to ν_ε. Hence, the question if such a factor is present or not when starting from a more general measure remains unclear, see also Remark 2.9.

Remark 3.2. *We have thus seen that, in this simple case, our deterministic process and the Wentzell–Freidlin process are remarkably similar. In fact, we conjecture that they have the same exact statistical properties.*

[14] Again, for the reader convenience, here is how to argue: first of all note that (3.5) implies that, on A_ε the ratio between the sup and inf to ρ_ε is bounded by $e^{c_\# a^2}$, and remember that $\nu_\varepsilon(A_\varepsilon) \geq 1/2$. Then

$$\int_{A_\varepsilon}\varphi^2\rho_\varepsilon = \frac{1}{2\nu_\varepsilon(A_\varepsilon)}\int_{A_\varepsilon^2}[\varphi(x)-\varphi(y)]^2\rho_\varepsilon(x)\rho_\varepsilon(y)dxdy + \frac{1}{\nu_\varepsilon(A_\varepsilon)}\left(\int_{A_\varepsilon}\varphi\rho_\varepsilon\right)^2.$$

While $[\varphi(x)-\varphi(y)]^2 \leq (\int_{A_\varepsilon}|\varphi'|)^2 \leq |A_\varepsilon|\int_{A_\varepsilon}(\varphi')^2 \leq C_\#\varepsilon\int_{\mathbb{T}}(\varphi')^2\rho_\varepsilon$.

3.2. **Skew-products over the doubling map–two sinks (non ergodic case).** Next, let us consider the case in which $n_Z = 2$. To start with, we assume that $|\bar{\omega}|$ reaches the value one in each of the intervals $(\theta_{1,-}, \theta_{1,+})$, $(\theta_{1,+}, \theta_{2,-})$, $(\theta_{2,-}, \theta_{2,+})$, $(\theta_{2,+}, \theta_{1,-})$. Also we assume that $|\hat{\omega}| \leq \frac{1}{2}$. Note that now assumption (A4) is **not** satisfied while it is satisfied hypotheses (A5) where alternative ii holds for any interval J. In the language of Subsection 6.4 this implies that there are two trapping sets $(\theta_{2,+}, \theta_{1,+}) \supset \mathcal{T}_{\epsilon,1} \ni \theta_{1,-}$ and $(\theta_{1,+}, \theta_{2,+}) \supset \mathcal{T}_{\epsilon,2} \ni \theta_{2,-}$.[15] Hence, the dynamics has two attractors with basins that contain the respective trapping sets and there are two SRB measures $\mu_{i,\varepsilon}$ supported in $\{\mathcal{T}_{\epsilon,i}\}$, respectively. In fact, the SRB measure $\mu_{1,\varepsilon}$ charges any ε-ball in a fixed (i.e. independent of ε) neighborhood[16] of $\theta_{1,-}$. Yet, by arguments similar to the ones used above, such measures are ε^{β}-close to two Gaussians, with variance of order ε and centered at $\{\theta_{1,-}, \theta_{2,-}\}$, respectively. That is, the system looks superficially like it has two attractors contained in a $\sqrt{\varepsilon}$ neighborhood of $\{\theta_{1,-}, \theta_{2,-}\}$.

Remark 3.3. *Note that in this case we have a drastic difference with the Wentzell–Freidlin process which, on the contrary, is ergodic. For the Wentzell–Freidlin process the measures $\mu_{i,\varepsilon}$ are essentially the* metastable states. *On the contrary, in the deterministic case they are stable (i.e. invariant). As already remarked, this is due to the substantial difference in the large deviations rate function of the two processes.*

3.3. **Skew-products over the doubling map–two sinks (ergodic case).** In this case we choose $\bar{\omega}$ as in our previous example , but with $-\frac{1}{3} < \hat{\omega} \leq 3$ with $\hat{\omega}(0) = 3$. Next, remember that the set of invariant probability measures is a closed convex set and so $\Omega(\theta) = \{\mu(\omega(\cdot, \theta)) \mid \mu \text{ invariant for } 2x \mod 1\}$ is a closed interval for each $\theta \in \mathbb{T}$. Thus $\Omega(\theta) \supset [\bar{\omega}(\theta), \bar{\omega}(\theta) + 3] \supset [1, 2]$. Accordingly, the path $h(t) = \frac{3}{2}t$ is admissible and visits all the circle, hence also assumption (A4) is satisfied. We are thus again in a setting to which our results apply, hence the map has a unique SRB measure that can be represented as a standard family.

As in the previous case the invariant measure will essentially consist of two Gaussians of variance of order ε centered at $\{\theta_{1,-}, \theta_{2,-}\}$, respectively. Yet, to really understand how the SRB looks like we must know the mass of the two Gaussians, let us call them $\{p_i\}$, respectively. Of course, $1 - p_1 - p_2 \leq C_{\#}\varepsilon^{\beta}$.

In general, to figure out the latter we can use [9, Theorem 2.4] to compute the probability for a trajectory to go from a neighborhood of $\theta_{1,-}$ to a neighborhood of $\theta_{2,-}$ and viceversa. Since the ratio of p_1 and p_2 depends on the probability of going from one sink to the other. This entails some work and more information on $\bar{\omega}$.

In order to keep things as simple as possible we choose $\bar{\omega}(\theta) = \sin(4\pi\theta)$. Note that if we set $R(x, \theta) = (x, \theta + \frac{1}{2} \mod 1)$, then $F_\varepsilon \circ R = R \circ F_\varepsilon$. But such a symmetry implies that, for each continuous function φ,

$$R_*\mu_\varepsilon(\varphi) = \mu_\varepsilon(\varphi \circ R \circ F_\varepsilon) = \mu_\varepsilon(\varphi \circ F \circ R) = R_*\mu_\varepsilon(\varphi \circ F_\varepsilon).$$

Thus $R_*\mu_\varepsilon$ is invariant and, as μ_ε, can be written in terms of a standard family. Since such a measure is unique, it must be $\mu_\varepsilon = R_*\mu_\varepsilon$, that is $p_1 = p_2 = \frac{1}{2} + \mathcal{O}(\varepsilon^{\beta})$.

Now that we have identified the SRB measure, we can discuss the issue of the decay of correlations. We choose A to be supported in a δ neighborhood of $\theta_{1,-}$, for δ small enough, such that $\mathrm{Leb}(A) = 1$ and B to be supported on a δ neighborhood

[15] The choice of ϵ is rather arbitrary, it suffices that it is small enough so that $\mathcal{T}_{\epsilon,i} \neq \emptyset$. In the present case $\epsilon = 1/4$ will do.

[16] Again, in the language of Section 6.4, one can take the neighborhood to be $\bigcap_{\epsilon>0} A^{+}_{\epsilon,\theta_{1,-}}$

of $\theta_{2,-}$ such that $\mu_\varepsilon(B) = 1$. We then ask what is the maximal c_ε such that

$$\begin{aligned}|\mathrm{Leb}(A \cdot B \circ F_\varepsilon^n) - \mathrm{Leb}(A)\mu_\varepsilon(B)| &\leq C_1 \sup_\theta \|A(\cdot,\theta)\|_{\mathcal{C}^2} \sup_x \|B(x,\cdot)\|_{\mathcal{C}^2} \exp(-c_\varepsilon n) \\ &\leq C_\# \exp(-c_\varepsilon n).\end{aligned}$$

We know that $\mathrm{Leb}(A)\mu_\varepsilon(B) = 1$. It remains to compute $\mathrm{Leb}(A \cdot B \circ F_\varepsilon^n)$. Note that at $1/8$ and $7/8$ we have $\bar\omega = 1$, on the other hand, by hypotheses $\hat\omega \geq -1/2$, thus $\bar\omega + \hat\omega \geq \frac{1}{2}$. That is, around $1/8$ and $7/8$ the motion can take place only from left to right.

Let $I_{i,-} = \{\theta \in \mathbb{T} \; : \; |\theta - \theta_{i,-}| \leq \delta\}$. Consider a process starting from a standard pair ℓ with $\theta_\ell \in I_{1,-}$. Our aim is to compute the probability of the event $Q_T = \{\theta_\varepsilon(T) \in I_{2,-}\}$. Note that, by choosing T_0 large enough, we have $|\bar\theta(T_0) - \theta_{1,-}| \leq \frac{1}{4}\delta$. Thus if $\gamma(T_0) \notin I_-$, then $\sup_{t\in[0,T_0]} |\gamma(t) - \bar\theta(t)| > \delta/2$. Then, using Theorem 6.1, we have

$$\mathbb{P}_\varepsilon(\gamma(T_0) \in I_{2,-}) \leq \mathbb{P}_\varepsilon(\{\gamma(T_0) \notin I_{1,-}\}) \leq \mu_\ell(Q(\delta/2, 1)) \leq e^{-c_\# \varepsilon^{-1}}.$$

Since the distribution at time T_0 is still made of standard pairs we can apply the same argument and obtain that

$$\mathbb{P}_\varepsilon(\{\gamma(T_0 n) \notin I_{1,-}\}) \leq n e^{-c_\# \varepsilon^{-1}}$$

It follows that, for each $n \leq e^{c_\# \varepsilon^{-1}}$ we have

$$\mathrm{Leb}(A \cdot B \circ F_\varepsilon^n) \leq \|A\|_\infty \|B\|_\infty \mathbb{P}_\varepsilon(\{\gamma(\varepsilon n) \notin I_{1,-}\}) \leq e^{-c_\# \varepsilon^{-1}}.$$

It follows that

$$c_\varepsilon \leq e^{-c_\# \varepsilon^{-1}}.$$

Thus the estimate in our Main Theorem has the right dependence on ε, even though, of course, the value of the constants are very hard to determine.

This means that, if we start with an initial distribution with mass, say, $1/4$ in a neighborhood of $\theta_{1,-}$ and $3/4$ in a neighborhood of $\theta_{2,-}$, then for an exponentially long time we will see a situation very similar to what we have seen in Section 3.2: it looks like the system is distributed according to an invariant measure. Yet, if we look at a longer exponential time, we will see the ratio of the masses of the two Gaussian change till it reaches the values approximately $1/2, 1/2$ which characterize the true invariant measure. Hence the metastability phenomena we have announced.

Remark 3.4. *We have seen that, in this case, we have metastable states as in the Wentzell–Freidlin case. Only, the attentive reader has certainly noticed that, in absence of a symmetry, there is no reason for the masses in the two sinks to be the same. Also we have seen by our large deviation computations that the probabilities to transit from one sink to the other are always exponentially small in ε^{-1}. It is thus to be expected that in a non symmetric (generic) case one of the two masses will be exponentially smaller than the other. Thus the SRB measure will look very much like a single Gaussian centered at the "winning" sink. Of course, the same occurs for Wentzell–Freidlin, yet who is the winning sink is decided by the large deviation functionals which are very different. So, again, we should expect cases in which the invariant measures of the two processes looks completely different, being centered at different sinks a part for an exponentially small mass.*

3.4. **An interesting non-example.** Let $\ell \in \mathbb{N}$, $\ell > 1$, and consider the family

$$F_\varepsilon(x, \theta) = (\ell x + \sin(2\pi\theta)\left[\alpha \sin(2\pi x) + \beta \sin(2\ell\pi x)\right], \theta + \varepsilon \cos(2\pi x)) \mod \mathbb{Z}^2.$$

In the above example, assuming $\ell - 2\pi(\alpha + \ell\beta) > 2$, Assumptions (A1) and (A3) are satisfied; moreover if ℓ is odd, we have that $x = 0$ and $x = 1/2$ are fixed points of f_θ for any $\theta \in \mathbb{T}$; since $\omega(0) = 1$ and $\omega(1/2) = -1$, we have $\Omega(\theta) \supset [-1, 1]$ for any $\theta \in \mathbb{T}$, hence Assumption (A4*) is satisfied (see Remark 2.7) and in particular (A4) holds. However, Assumption (A2) is not obvious to verify. This whole subsection is devoted to the discussion of this issue. In doing so we will uncover the possibility of a most surprising feature: an "attractor" with all Lyapunov exponents almost surely positive (with respect to the SRB measure).[17] To actually prove this would take some non trivial work; here we content ourselves by showing that for $\alpha, \beta > 0$ the average dynamics has a sink and yet the true dynamics near such a sink has center vectors that are mostly expanding.

Observe that if $\theta = 0$ or $\theta = 1/2$ (so that $\sin(2\pi\theta) = 0$), then $f_\theta(x) = \ell x$, thus $\rho_\theta = 1$, and $\bar\omega(\theta) = 0$. Let us now compute $\bar\omega'(\theta)$ at $\theta = 0$:

$$\begin{aligned}
\bar\omega'(\theta) &= \frac{d}{d\theta}\int_{\mathbb{T}} \omega(x)\rho_\theta(x)dx = \sum_{k=1}^{\infty}\int_{\mathbb{T}} (\omega\circ f_\theta^k(x))' \frac{\partial_\theta f(x,\theta)}{f'_\theta(x)}\rho_\theta(x)dx \\
&= -(2\pi)^2\sum_{k=1}^{\infty} \ell^{k-1}\int_{\mathbb{T}} \sin(2\ell^k\pi x)[\alpha\sin(2\pi x) + \beta\sin(2\ell\pi x)] \\
&= -2\pi^2\beta,
\end{aligned}$$

where we have used the perturbative results detailed in [9, Appendix A.3]. Thus if $\beta > 0$, then $\theta = 0$ is a sink.

On the other hand, as will be shown in (4.7), the expansion of center vectors at time n is

$$\log \mu_n(p) = \varepsilon \sum_{k=0}^{n-1}[\partial_\theta\omega(p_k) + \partial_x\omega(p_k)s_{n-k}(p_k)] + \mathcal{O}(\varepsilon^2 n), \tag{3.6}$$

where s_n is defined in (4.6); we thus have the formula:

$$\begin{aligned}
s_n(p_0) = s_n(x_0,\theta_0) &= -\sum_{k=0}^{n-1}\frac{\partial_\theta f(x_k,\theta_k)}{\prod_{j=0}^{k}\partial_x f(x_j,\theta_j)} + \mathcal{O}(\varepsilon) \\
&= -\sum_{k=0}^{n-1}\frac{\partial_\theta f(f_{\theta_0}^k(x_0),\theta_0)}{(f_{\theta_0}^{k+1})'(x_0)} + \mathcal{O}(\varepsilon\log\varepsilon^{-1})
\end{aligned}$$

where we have used [9, Lemma 4.2]. Substituting this is (3.6) yields

$$\begin{aligned}
\log\mu_n(p) &= \varepsilon\sum_{k=0}^{n-1}[\partial_\theta\omega(p_k) + \partial_x\omega(p_k)s_*(p_k)] + \mathcal{O}(n\varepsilon^2\log\varepsilon^{-1} + \varepsilon) \\
&= \varepsilon\sum_{k=0}^{n-1}\psi_*(p_k) + \mathcal{O}(n\varepsilon^2\log\varepsilon^{-1} + \varepsilon).
\end{aligned}$$

The above formula illustrates the announced relation between the function ψ_* and the central Lyapunov exponent. Also we have seen that

$$\psi_*(p_0) = -\sum_{k=0}^{\infty}\frac{\partial_\theta f(f_{\theta_0}^k(x_0),\theta_0)}{(f_{\theta_0}^{k+1})'(x_0)} + \mathcal{O}(\varepsilon\log\varepsilon^{-1}).$$

[17] Of course, technically speaking, there is no attractor as condition (A4*) guarantees that the dynamics will visit an ε-dense set in configuration space. Yet, for small ε and each $\beta \in (0, 1/2)$ a portion $1 - e^{-c_\# \varepsilon^{-1+2\beta}}$ of the mass is concentrated in a $\mathcal{O}(\varepsilon^\beta)$-neighborhood of $\theta = 0$. So the situation differs indeed very little from an attractor. In passing, this example shows that a purely topological description of the dynamics can fail miserably in capturing the relevant properties of the motion.

The average of the logarithm of the expansion of center vectors, at $\theta_0 = 0$, is

$$\begin{aligned}\bar\psi_*(\theta_0) &= -\sum_{k=0}^{\infty}\int_{\mathbb{T}} \partial_x\omega(x,\theta_0)\frac{\partial_\theta f(f^k_{\theta_0}(x),\theta_0)}{(f^{k+1}_{\theta_0})'(x)}dx + \mathcal{O}(\varepsilon\log\varepsilon^{-1}) = \\ &= 2\pi\sum_{k=0}^{\infty}\int_{\mathbb{T}}\sin(2\pi x)\frac{\alpha\sin(2\ell^k\pi x)+\beta\sin(2\ell^{k+1}\pi x)}{\ell^{k+1}}dx + \mathcal{O}(\varepsilon\log\varepsilon^{-1}) = \\ &= 2\pi^2\frac{\alpha}{\ell} + \mathcal{O}(\varepsilon\log\varepsilon^{-1}).\end{aligned}$$

Consequently, if $\alpha < 0$, then assumption (A2) is satisfied and our Main Theorem applies. On the contrary, if $\alpha > 0$, then assumption (A2) is violated and, as announced, we have a map that we expect to have positive central Lyapunov exponent.

Remark 3.5. *We have just seen another drastic difference between the Wentzell–Freidlin process and the deterministic process: in the Wentzell–Freidlin process the Lyapunov exponent associated to the slow variable is always negative*[18] *while we have seen that for the deterministic process it can be positive. This depends on the fact that the stochastic process does not reflect completely the interplay between the slow and the fast variable which can be much more subtle in the deterministic case.*

4. Geometry

Throughout this article, $\pi : \mathbb{T}^2 \to \mathbb{T}$ denotes the projection on the x-coordinate. We denote a point in $\mathbb{T}^2$ by $p = (x, \theta)$; we use the notation $p_n = (x_n, \theta_n) = F^n_\varepsilon p$. Our first task is to find invariant cones for the dynamics: for $\gamma^{\mathrm{u}}, \gamma^{\mathrm{c}} > 0$ to be specified later, let us define the *unstable cone* and the *center cone* as, respectively:

$$\mathfrak{C}^{\mathrm{u}} = \{(\xi,\eta)\in\mathbb{R}^2 \;:\; |\eta| \le \varepsilon\gamma^{\mathrm{u}}|\xi|\} \qquad \mathfrak{C}^{\mathrm{c}} = \{(\xi,\eta)\in\mathbb{R}^2 \;:\; |\xi| \le \gamma^{\mathrm{c}}|\eta|\}. \tag{4.1}$$

We claim that there exist $\gamma^{\mathrm{u}}, \gamma^{\mathrm{c}}$ such that, if ε is small enough, $dF_\varepsilon\mathfrak{C}^{\mathrm{u}} \subset \mathfrak{C}^{\mathrm{u}}$ and $dF_\varepsilon^{-1}\mathfrak{C}^{\mathrm{c}} \subset \mathfrak{C}^{\mathrm{c}}$. In fact, let us compute the differential of F_ε:

$$dF_\varepsilon = \begin{pmatrix}\partial_x f & \partial_\theta f\\ \varepsilon\partial_x\omega & 1+\varepsilon\partial_\theta\omega\end{pmatrix}; \tag{4.2}$$

consequently, if we consider the vector $(1, \varepsilon u)$

$$\begin{aligned} d_pF_\varepsilon(1,\varepsilon u) &= (\partial_x f(p) + \varepsilon u\partial_\theta f(p), \varepsilon\partial_x\omega(p) + \varepsilon u + \varepsilon^2 u\partial_\theta\omega(p))\\ &= \partial_x f(p)\left(1+\varepsilon\frac{\partial_\theta f(p)}{\partial_x f(p)}u\right)\cdot(1,\varepsilon\Xi_p(u)) \end{aligned}\tag{4.3}$$

where

$$\Xi_p(u) = \frac{\partial_x\omega(p) + (1+\varepsilon\partial_\theta\omega(p))u}{\partial_x f(p) + \varepsilon\partial_\theta f(p)u}, \tag{4.4}$$

from which we obtain our claim, choosing for instance

$$\gamma^{\mathrm{u}} = 2\|\partial_x\omega\|_\infty \qquad\text{and}\qquad \gamma^{\mathrm{c}} = 2\|\partial_\theta f\|_\infty. \tag{4.5}$$

From the above computations it is easy to see that F_ε is a partially hyperbolic map with expanding direction in $\mathfrak{C}^{\mathrm{u}}$ and central direction in $\mathfrak{C}^{\mathrm{c}}$.

It follows that, for any $p \in \mathbb{T}^2$ and $n \in \mathbb{N}$, we can define the real quantities μ_n, ν_n, u_n and s_n as follows:

$$d_pF^n_\varepsilon(1,0) = \nu_n(1,\varepsilon u_n) \qquad d_pF^n_\varepsilon(s_n,1) = \mu_n(0,1) \tag{4.6}$$

with $|u_n| \le \gamma^{\mathrm{u}}$ and $|s_n| \le \gamma^{\mathrm{c}}$. For each n the slope field s_n is smooth, therefore integrable; given any (small) $h > 0$ and $p_* = (x_*, \theta_*) \in \mathbb{T}^2$, define $\mathcal{W}^{\mathrm{c}}_n(p_*, h)$ the

[18] This follows from a direct computation.

local n-step center manifold of size h as the connected component containing p_* of the intersection with the strip $\{\theta \in B(\theta_*, h)\}$ of the integral curve of $(s_n, 1)$ passing through p_*. Observe that, by definition, any vector tangent to a local n-step center manifold belongs to the center cone.

Moreover, notice that, by definition,

$$d_pF_\varepsilon(s_n(p), 1) = \mu_n(p)/\mu_{n-1}(F_\varepsilon p)(s_{n-1}(F_\varepsilon p), 1);$$

a direct application of (4.2) yields

$$\frac{\mu_n(p)}{\mu_{n-1}(F_\varepsilon p)} = 1 + \varepsilon\left[\partial_\theta\omega(p) + \partial_x\omega(p)s_n(p)\right] \tag{4.7}$$

Observe that the above expression implies

$$\log\mu_m(p) - \log\mu_{m-n}(F_\varepsilon^n p) \leq \Psi n\varepsilon, \text{ where } \Psi = \|\partial_\theta\omega\| + \gamma^c\|\partial_x\omega\|. \tag{4.8}$$

Moreover,

$$s_n(p) = \frac{(1 + \varepsilon\partial_\theta\omega(p))s_{n-1}(F_\varepsilon(p)) - \partial_\theta f(p)}{\partial_x f(p) - \varepsilon\partial_x\omega(p)s_{n-1}(F_\varepsilon(p))} =: \Xi_p^-(s_{n-1}(F_\varepsilon(p)). \tag{4.9}$$

Note that

$$\frac{d}{ds}\Xi_p^-(s) = \frac{(1 + \varepsilon\partial_\theta\omega(p))\partial_x f(p) - \varepsilon\partial_x\omega(p)\partial_\theta f(p)}{\left[\partial_x f(p) - \varepsilon\partial_x\omega(p)s\right]^2}.$$

Accordingly, for each $|s| \leq \gamma^c$ and ε small enough, we have that there exists $\sigma_c \in (0, 1)$ such that

$$\left|\frac{d}{ds}\Xi_p^-(s)\right| \leq \sigma_c.$$

This implies that s_n is a converging sequence: let s_∞ be its limit. Then, for all $p \in \mathbb{T}^2$, $|s_n(p) - s_\infty(p)| \leq C_\#\sigma_c^n$; we have thus a formula for the center slope of F_ε.

For any $p = (x, \theta) \in \mathbb{T}^2$ let us now define

$$s_{*,n}(p) = -\sum_{k=0}^{n}\frac{\partial_\theta f(f_\theta^k(x))}{\left(f_\theta^{k+1}\right)'(x)}$$

It is clear from (4.9) and from the above comments that $s_* = \lim_{n\to\infty} s_{*,n}$ is the center slope of F_0. Moreover it is not difficult to observe that $\|s_{*,n} - s_n\| \leq C_\#\varepsilon n$, which then (using the fact that both $s_{*,n}$ and s_n converge exponentially fast) implies that $\|s_* - s_\infty\| \leq C_\#\varepsilon\log\varepsilon^{-1}$. Yet, it is well known that, in general, s_* is not a very regular function of the point. This could create trouble while using our assumption (A2) since typically we would need to use s_* in formulae that require some regularity. To overcome this problem we define a regularized function ψ that approximates ψ_* (see (2.2)). First, let $\psi_\infty = \partial_\theta\omega(p) + \partial_x\omega(p)s_\infty(p)$ and define

$$\psi(p) = \partial_\theta\omega(p) + \partial_x\omega(p)s_{\bar{n}}(p) \qquad \bar{\psi}(\theta) = \int_{\mathbb{T}}\psi(x, \theta)\rho_\theta(x)dx \tag{4.10}$$

where $\bar{n}$ is so that for any $p \in \mathbb{T}^2$

$$\|\psi(p) - \psi_\infty(p)\| < \varrho < 1/16$$

for some ϱ small to be specified in due course.

Remark 4.1. *Due to the uniformity in p of all estimates involved, we have uniform bounds on the norms of ψ, i.e.: $\|\psi\|_{\mathcal{C}^0} < \Psi$ and $\|\psi\|_{\mathcal{C}^1} < \exp(C_\#\bar{n})$. Moreover, we can always assume ε to be so small that $\|\psi_* - \psi_\infty\| < C_\#\varepsilon\log\varepsilon^{-1} < 1/16$, so that, under assumption (A2),*

$$\max_{k\in\{1,\cdots,n_Z\}}\bar{\psi}(\theta_{k,-}) \in [-9/8, -7/8].$$

Define the function ζ_n as:

$$\zeta_n = \varepsilon \sum_{k=0}^{n-1} \psi \circ F_\varepsilon^k. \tag{4.11}$$

Lemma 4.2 (Distortion)**.** *For any $T > 0$ there exists C_T so that, for any $p \in \mathbb{T}^2$ and $h > 0$ sufficiently small, let $N = \lfloor T\varepsilon^{-1} \rfloor$*

$$\sup_{q \in \mathcal{W}_N^c(p,h)} \mu_N(q) \le \exp(\zeta_N(p) + C_T h + 2T\varrho + 2\bar{n}\Psi\varepsilon).$$

Proof. Let us introduce the convenient function $\psi_n(p) = \partial_\theta\omega(p) + \partial_x\omega(p)s_n(p)$ (and likewise let $\psi_\infty(p) = \partial_\theta\omega(p) + \partial_x\omega(p)s_\infty(p)$); then by (4.7) we can write $\mu_N(p) \le \exp\left(\varepsilon \sum_{n=0}^{N-1} \psi_n\right)$. On the other hand, by construction and the triangle inequality we have $\|\psi - \psi_n\| < 2\varrho$ if $n \ge \bar{n}$ (otherwise the trivial bound $\|\psi - \psi_n\| < 2\Psi$ holds); we conclude that $\mu_N(p) \le \exp(\zeta_N(p) + 2T\varrho + 2\bar{n}\Psi\varepsilon)$; next, we need to compute the derivative of ζ_N along the N-step central direction. By (4.11) it follows

$$d\zeta_N(s_N, 1) = \varepsilon \sum_{k=0}^{N-1} \langle \nabla\psi \circ F_\varepsilon^k, dF_\varepsilon^k(s_N, 1)\rangle;$$

hence, (4.6) and (4.8) imply that $dF_\varepsilon^k(s_N, 1) = e^{\mathcal{O}(k\Psi\varepsilon)}(s_{N-k}, 1)$. Thus there exists $b_T \sim \exp(c_\# T) > 0$ such that

$$|d\zeta_N(s_N, 1)| \le b_T\varepsilon \sum_{k=0}^{N-1} e^{k\Psi\varepsilon} \le C_\# b_T \Psi^{-1}.$$

Accordingly,

$$\sup_{q \in \mathcal{W}_N^c(p,h)} \mu_N(q) \le \exp(\zeta_N(p) + C_\# b_T\Psi^{-1}h + 2T\varrho + 2\bar{n}\Psi\varepsilon). \qquad \square$$

5. Standard pairs, families and couplings

5.1. Definitions and basic facts.

In this section we recap the standard families formalism, first introduced by Dolgopyat (see e.g. [14, 15, 16]) to study statistical properties of partially hyperbolic dynamical systems.[19]

Remark 5.1. *The educated reader will certainly notice that our regularity assumptions are stronger than the ones which are usually required to apply the coupling argument (see e.g. [4]). The stronger regularity conditions are in fact needed in order to obtain the refined statistical properties (i.e., the Local Central Limit Theorem) that we use to set up the coupling argument in an efficient manner. Consequently, they are crucial to obtain the near-optimal bounds on the rate of decay of correlations that we seek.*

5.1.1. *Standard pairs.* Let us fix a small $\delta > 0$, and $D_1, D_1' > 0$ large to be specified later; for any $c_1 > 0$ let us define the set of functions

$$\Sigma_{c_1} = \{G \in \mathcal{C}^3([a,b], \mathbb{T}) : a, b \in \mathbb{T}, b - a \in [\delta/2, \delta],$$
$$\|G'\| \le \varepsilon c_1, \ \|G''\| \le \varepsilon D_1 c_1, \ \|G'''\| \le \varepsilon D_1' c_1\}.$$

Let us associate to any $G \in \Sigma_{c_1}$ the map $\mathbb{G}(x) = (x, G(x))$; the graph of any such G (i.e. the image of $\mathbb{G}$) will be called a *proper c_1-standard curve.* With a little abuse of terminology, we refer to the quantity $b - a$ as the *length of the curve.* If we do not require the lower bound for the length of the curve, we obtain the definition of a *short c_1-standard curve*; for ease of exposition we adopt the convention that

[19] See also [9, Section 3.2] for a similar, but more general, account of the framework in this context.

all standard curves are assumed to be proper unless otherwise specified. Also, with another convenient abuse of terminology, we use the term c_1-*standard curve* to indicate also the function G or the map $\mathbb{G}$. Two c_1-standard curves G^0 and G^1 are said to be *stacked* if their projection on the x axis coincide; we say that G^0 and G^1 are Δ-*stacked* if they are stacked and $\|G^0 - G^1\|_{\mathcal{C}^1} < \Delta$.

Let us fix $D_2 > 0$ once again to be specified in due course. For any $c_2 > 0$ define the set of c_2-*standard* probability densities on the standard curve G as

$$D_{c_2}(G) = \left\{ \rho \in \mathcal{C}^2([a,b], \mathbb{R}_+) \, : \, \int_a^b \rho(x)dx = 1, \, \left\| \frac{\rho'}{\rho} \right\| \le c_2, \, \left\| \frac{\rho''}{\rho} \right\| \le D_2 c_2 \right\}.$$

A (c_1, c_2)-*standard pair* ℓ is given by $\ell = (\mathbb{G}, \rho)$, where $G \in \Sigma_{c_1}$ and $\rho \in D_{c_2}(G)$. We similarly define *short* (c_1, c_2)-*standard pairs*, by allowing G to be a short c_1-standard curve. We define $|\ell| = b - a$ to be the *length of* ℓ. A (c_1, c_2)-standard pair $\ell = (\mathbb{G}, \rho)$ uniquely identifies a probability measure μ_ℓ on $\mathbb{T}^2$ defined as follows: for any Borel-measurable function g on $\mathbb{T}^2$ let

$$\mu_\ell(g) := \int_a^b g(\mathbb{G}(x))\rho(x)dx.$$

Let L_{c_1,c_2} denote the set of all (c_1, c_2)-standard pairs.

5.1.2. *Standard families.* A standard family can be conveniently regarded as a *random standard pair*. More precisely: a (c_1, c_2)-*standard family* $\mathfrak{L}$ is given by a Lebesgue probability space[20] $\mathscr{A} = (\mathcal{A}, \mathcal{F}, \nu)$ and a $\mathcal{F}$-measurable[21] map $\ell : \mathcal{A} \to L_{c_1,c_2}$.

For simplicity's sake, in this paper we will mostly restrict to standard families such that $\mathscr{A} = (\mathcal{A}, \mathcal{F}, \nu)$ is a discrete probability space (i.e., $\mathcal{A}$ is at most countable and $\mathcal{F}$ is the power set of $\mathcal{A}$). We will thus imply that $\mathcal{A}$ is at most countable, and simply write $\mathscr{A} = (\mathcal{A}, \nu)$, otherwise explicitly stated. We will denote the set of all (c_1, c_2)-standard families by $\mathbb{L}_{(c_1,c_2)}$.

A (c_1, c_2)-standard family $\mathfrak{L}$ identifies a unique probability measure $\hat{\mu}_{\mathfrak{L}}$ on the product space $\mathcal{A} \times \mathbb{T}^2$ (with the product σ-algebra): for any measurable function $\hat{g}$ on $\mathcal{A} \times \mathbb{T}^2$ let

$$\hat{\mu}_{\mathfrak{L}}(\hat{g}) := \int_{\mathcal{A}} \mu_{\ell(\alpha)}(\hat{g}(\alpha, \cdot))d\nu.$$

Define the *support* of $\mathfrak{L}$ as $\operatorname{supp} \mathfrak{L} = \operatorname{supp} \hat{\mu}_{\mathfrak{L}} \subset \mathcal{A} \times \mathbb{T}^2$. The natural projection $\boldsymbol{\pi} : \mathcal{A} \times \mathbb{T}^2 \to \mathbb{T}^2$ induces a probability measure on $\mathbb{T}^2$ which we denote by $\mu_{\mathfrak{L}} = \boldsymbol{\pi}_* \hat{\mu}_{\mathfrak{L}}$; in other words, for any Borel-measurable function g of $\mathbb{T}^2$, let

$$\mu_{\mathfrak{L}}(g) := \int_{\mathcal{A}} \mu_{\ell(\alpha)}(g)d\nu.$$

Clearly, we have $\operatorname{supp} \mu_{\mathfrak{L}} = \boldsymbol{\pi} \operatorname{supp} \mathfrak{L}$.[22] We therefore obtain a correspondence between (c_1, c_2)-standard families and probabilities on $\mathbb{T}^2$; we denote by $\sim$ the equivalence relation induced by the above correspondence i.e. we let $\mathfrak{L} \sim \mathfrak{L}'$ if

---

[20] Recall that a probability space is a Lebesgue space if it is isomorphic to the disjoint union of an interval $[0, a]$ with Lebesgue measure and (at most) countably many atoms.

[21] The set L_{c_1,c_2} of (c_1, c_2)-standard pairs is in fact a space of smooth functions; it is thus a measurable space with the Borel σ-algebra. More in detail, if $\mathbb{G} : [a, b] \to \mathbb{T}^2$ and $\rho : [a, b] \to \mathbb{R}^+$ are defined as above, let $\hat{\mathbb{G}}$ and $\hat{\rho}$ be defined by precomposing $\mathbb{G}$ and ρ respectively with the affine orientation-preserving map $[0, 1] \to [a, b]$. A standard pair-valued function is thus $\mathcal{F}$-measurable if both maps $(\alpha, s) \mapsto \hat{\mathbb{G}}_\alpha(s)$ and $(\alpha, s) \mapsto \hat{\rho}_\alpha(s)$ are jointly measurable. In particular, for any Borel set $E \subset \mathbb{T}^2$, the function $\alpha \mapsto \mu_{\ell(\alpha)}(E)$ is $\mathcal{F}$-measurable.

[22] This concept can be obviously applied to a single standard pair, considering it a family with just one element. In such case, the support of the standard pair and the support of the associated measure can be trivially identified.

and only if $\mu_{\mathfrak{L}} = \mu_{\mathfrak{L}'}$. We denote with $[\mathfrak{L}]$ the corresponding equivalence class, which therefore uniquely identifies a probability measure. We say that a probability measure μ *admits a* (c_1, c_2)*-standard disintegration* if there exists a (c_1, c_2)-standard family $\mathfrak{L}$ so that $\mu_{\mathfrak{L}} = \mu$; we write $\mathfrak{L} \in \mathbb{L}_{c_1,c_2}(\mu)$.

5.1.3. *Conditioning.* Let $\mathfrak{L} = ((\mathcal{A}, \nu), \ell) \in \mathbb{L}_{(c_1,c_2)}$; a family $\mathfrak{L}' = ((\mathcal{A}', \nu'), \ell')$ is said to be a *subfamily of* $\mathfrak{L}$ (denoted with $\mathfrak{L}' \subset \mathfrak{L}$) if

- $\operatorname{supp} \mathfrak{L}' \subset \operatorname{supp} \mathfrak{L}$, that is: $\mathcal{A}' \subset \mathcal{A}$ and $\forall\, \alpha \in \mathcal{A}'$ we have $\operatorname{supp} \ell'(\alpha) \subset \operatorname{supp} \ell(\alpha)$;
- for any measurable set $E \subset \mathcal{A} \times \mathbb{T}^2$, $\hat\mu_{\mathfrak{L}'}(E) = \hat\mu_{\mathfrak{L}}(E \cap \operatorname{supp} \mathfrak{L}')/\hat\mu_{\mathfrak{L}}(\operatorname{supp} \mathfrak{L}')$.

Given $\mathcal{A}' \subset \mathcal{A}$, we define the *subfamily conditioned on* $\mathcal{A}'$ to be $\mathfrak{L}\big|\mathcal{A}' = ((\mathcal{A}', \nu'), \ell\big|\mathcal{A}')$, where $\nu'(E) = \nu(E|\mathcal{A}')$ and $\ell\big|\mathcal{A}'$ is the restriction of ℓ on $\mathcal{A}'$.

5.1.4. *Convex combinations of pairs and families.* We call a real number κ a *weight* if $\kappa \in [0, 1]$. Given a (at most countable) collection of (c_1, c_2)-standard families $\{\mathfrak{L}_j = (\mathscr{A}_j, \ell_j)\}$ together with a collection of weights $\{\kappa_j\}$ such that $\sum_j \kappa_j = 1$, we can define the *convex combination* $\sum_j \kappa_j \mathfrak{L}_j$ as the (c_1, c_2)-standard family $\mathfrak{L} = (\mathscr{A}, \ell)$ obtained by "choosing a standard family $\mathfrak{L}_j$ at random with probability κ_j". More precisely, let $\mathscr{A} = (\mathcal{A}, \nu)$ be the discrete probability space given by $\mathcal{A} = \{(j, \alpha) : \alpha \in \mathcal{A}_j\}$ and measure $\nu = \sum_j \kappa_j \cdot \boldsymbol{\iota}_{j*}\nu_j$, where $\boldsymbol{\iota}_j$ is the natural injection $\boldsymbol{\iota}_j : \mathcal{A}_j \to \mathcal{A}$. Last, let us define the random element ℓ as $\ell(j, \alpha) = \ell_j(\alpha)$; clearly $\mu_{\mathfrak{L}} = \sum_j \kappa_j \mu_{\mathfrak{L}_j}$. With this in mind, observe that we can recover the components of a convex combination by conditioning with respect to the events $\bar{\mathcal{A}}_k = \{(j, \alpha) : j = k, \alpha \in \mathcal{A}_k\}$. Observe, moreover, that standard families can naturally be regarded as convex combinations of standard pairs.

5.2. **Standard pairs and dynamics.** Having made precise the concept of standard pair and families, our next step is to illustrate their relation with the dynamics generated by the map F_ε.

5.2.1. *Invariance.* As a first step we study the evolution of a (c_1, c_2)-standard pair.

Proposition 5.2 (Invariance)**.** *There exist* c_1*,* c_2 *such that, if* ε *is sufficiently small and* ℓ *is a* (c_1, c_2)*-standard pair,* $F_{\varepsilon*}\mu_\ell$ *admits a* (c_1, c_2)*-standard disintegration.*

Remark 5.3. *The above proposition is a simplified version of the corresponding* [9, Proposition 3.3] *where it is proved in a more general setting. Since there are a few differences in the notation and terminology between this version and the one of* [9]*, we prefer to give an adapted proof below for the reader's convenience. Despite its technical nature, the proof is instrumental for a few definitions which will be given later. We thus prefer to give it now rather than relegating it to some appendix.*

Proof. Let $\ell = (\mathbb{G}, \rho)$ be a (c_1, c_2)-standard pair. For any sufficiently smooth function A on $\mathbb{T}^2$, by the definition of standard curve, it is trivial to check that:

$$\|(A \circ \mathbb{G})'\| \le \|dA\|(1 + \varepsilon c_1) \tag{5.1a}$$

$$\|(A \circ \mathbb{G})''\| \le \varepsilon\|dA\|D_1 c_1 + \|dA\|_{\mathcal{C}^1}(1 + \varepsilon c_1)^2 \tag{5.1b}$$

$$\|(A \circ \mathbb{G})'''\| \le \varepsilon\|dA\|D_1' c_1 + \|dA\|_{\mathcal{C}^2}(1 + \varepsilon(1 + D_1)c_1)^3. \tag{5.1c}$$

Let us then introduce the maps $f_{\mathbb{G}} = f \circ \mathbb{G}$ and $\omega_{\mathbb{G}} = \omega \circ \mathbb{G}$. Recall that $\lambda > 2$, defined in Section 2 denotes the minimal expansion of f_θ; we will assume ε to be small enough (depending on our choice of c_1) so that $f_{\mathbb{G}}' \ge \lambda - \varepsilon c_1\|\partial_\theta f\| > 3/2$; in particular, $f_{\mathbb{G}}$ is an expanding map. Provided δ has been chosen small enough, $f_{\mathbb{G}}$ is invertible. Let $\varphi(x) = f_{\mathbb{G}}^{-1}(x)$. Differentiating we obtain

$$\varphi' = \frac{1}{f'_{\mathbb{G}}} \circ \varphi \qquad \varphi'' = -\frac{f''_{\mathbb{G}}}{f'^3_{\mathbb{G}}} \circ \varphi \qquad \varphi''' = \frac{3f''^2_{\mathbb{G}} - f'''_{\mathbb{G}} f'_{\mathbb{G}}}{f'^5_{\mathbb{G}}} \circ \varphi. \tag{5.2}$$

We can thus write:

$$\begin{aligned} F_{\varepsilon *}\mu_\ell(g) = \mu_\ell(g \circ F_\varepsilon) &= \int_a^b g(f_{\mathbb{G}}(x), \bar{G}(x))\rho(x)dx \\ &= \int_{f_{\mathbb{G}(a)}}^{f_{\mathbb{G}(b)}} g(x, \bar{G}(\varphi(x))) \cdot \rho(\varphi(x))\varphi'(x)dx, \end{aligned}$$

where $\bar{G}(x) := G(x) + \varepsilon\omega_{\mathbb{G}}(x)$. Fix a partition (mod 0) of $[f_{\mathbb{G}}(a), f_{\mathbb{G}}(b)] = \bigcup_{j\in\mathcal{J}}[a_j, b_j]$, with $b_j - a_j \in [\delta/2, \delta]$ and $b_j = a_{j+1}$. We can thus write

$$F_{\varepsilon *}\mu_\ell(g) = \sum_j Z_j \int_{a_j}^{b_j} g(x, G_j(x)) \cdot \rho_j(x)dx = \sum_j Z_j \mu_{(\mathbb{G}_j, \rho_j)}(g).$$

provided that $G_j = \bar{G} \circ \varphi_j$ and $\rho_j = Z_j^{-1} \cdot \rho \circ \varphi_j \cdot \varphi'_j$ where $\varphi_j = \varphi|_{[a_j, b_j]}$ and $Z_j = \int_{a_j}^{b_j} \rho(\varphi_j(x))\varphi'_j(x)dx$. Observe that, by construction, we have $\sum_j Z_j = 1$. Differentiating the above definitions and using (5.2) we obtain

$$G'_j = \frac{\bar{G}'}{f'_{\mathbb{G}}} \circ \varphi_j \tag{5.3a}$$

$$G''_j = \frac{\bar{G}''}{f'^2_{\mathbb{G}}} \circ \varphi_j - G'_j \cdot \frac{f''_{\mathbb{G}}}{f'^2_{\mathbb{G}}} \circ \varphi_j \tag{5.3b}$$

$$G'''_j = \frac{\bar{G}'''}{f'^3_{\mathbb{G}}} \circ \varphi_j - 3G''_j \cdot \frac{f''_{\mathbb{G}}}{f'^2_{\mathbb{G}}} \circ \varphi_j - G'_j \cdot \frac{f'''_{\mathbb{G}}}{f'^3_{\mathbb{G}}} \circ \varphi_j \tag{5.3c}$$

and similarly

$$\frac{\rho'_j}{\rho_j} = \frac{\rho'}{\rho \cdot f'_{\mathbb{G}}} \circ \varphi_j - \frac{f''_{\mathbb{G}}}{f'^2_{\mathbb{G}}} \circ \varphi_j \tag{5.4a}$$

$$\frac{\rho''_j}{\rho_j} = \frac{\rho''}{\rho \cdot f'^2_{\mathbb{G}}} \circ \varphi_j - 3\frac{\rho'_j}{\rho_j} \cdot \frac{f''_{\mathbb{G}}}{f'^2_{\mathbb{G}}} \circ \varphi_j - \frac{f'''_{\mathbb{G}}}{f'^3_{\mathbb{G}}} \circ \varphi_j. \tag{5.4b}$$

Using (5.3a), the definition of $\bar{G}$ and (5.1a) we obtain, for small enough ε:

$$\begin{aligned} \|G'_j\| &\le \left\| \frac{G' + \varepsilon\omega'_{\mathbb{G}}}{f'_{\mathbb{G}}} \right\| \le \frac{2}{3}(1 + \varepsilon\|d\omega\|)\varepsilon c_1 + \frac{2}{3}\varepsilon\|d\omega\| \\ &\le \frac{3}{4}\varepsilon c_1 + \varepsilon D_1 \end{aligned}$$

where $D_1 = \frac{2}{3}\|d\omega\|$. We can then fix c_1 large enough so that the right hand side of the above inequality is less than c_1. Next we will use C_* for a generic constant depending on c_1, D_1, D'_1, c_2, D_2 and $C_\#$ for a generic constant depending only on F_ε. Then, we find[23]

$$\|G''_j\| \le \frac{3}{4}\varepsilon[c_1 D_1 + C_\#] + \varepsilon^2 C_*\,; \quad \|G'''_j\| \le \frac{3}{4}\varepsilon\left[c_1(D'_1 + D_1 C_\# + C_\#) + C_\#\right] + \varepsilon^2 C_*$$

$$\left\|\frac{\rho'_j}{\rho_j}\right\| \le \frac{3}{4}c_2 + C_\# + \varepsilon C_*\,; \qquad \left\|\frac{\rho''_j}{\rho_j}\right\| \le \frac{3}{4}c_2[D_2 + C_\#] + C_\# + \varepsilon C_*.$$

We can then fix c_1, D'_1, c_2, D_2 sufficiently large and then ε sufficiently small to ensure that the $(\mathbb{G}_j, \rho_j)$ are standard pairs. We have thus obtained a decomposition of $F_{\varepsilon *}\mu_\ell$ given by the discrete standard family $\mathfrak{L}' = ((\mathcal{J}, Z_j), \ell_j)$. □

[23] The reader can easily fill in the details of the computations.

Remark 5.4. *The construction described in the above proposition yields more than just a standard disintegration of* $F_{\varepsilon*}\mu_\ell$. *In fact, it gives an* invertible *map* $\hat F_\varepsilon : \operatorname{supp}\ell \to \operatorname{supp}\mathfrak{L}$ *such that* $F_\varepsilon = \boldsymbol{\pi}\circ\hat F_\varepsilon$ *and* $\hat\mu_{\mathfrak{L}} = \hat F_{\varepsilon*}\mu_\ell$ *(such map does not exist in general for a standard disintegration of* $F_{\varepsilon*}\mu_\ell$*).*

It is immediate to extend the above proposition to standard families: let $\mathfrak{L} = ((\mathcal{A},\nu),\ell)$ be a standard family; then by definition we have, for any measurable function g:

$$F_{\varepsilon*}\mu_{\mathfrak{L}}(g) = F_{\varepsilon*}\sum_{\alpha\in\mathcal{A}}\nu_\alpha\mu_{\ell_\alpha}(g) = \sum_{\alpha\in\mathcal{A}}\nu_\alpha F_{\varepsilon*}\mu_{\ell_\alpha}(g) = \sum_{\alpha\in\mathcal{A}}\nu_\alpha\mu_{\mathfrak{L}'_\alpha}(g)$$

where $\mathfrak{L}'_\alpha$ is the standard family obtained by applying Proposition 5.2 to ℓ_α. We conclude that the convex combination

$$\mathfrak{L}' = \sum_{\alpha\in\mathcal{A}}\nu_\alpha\mathfrak{L}'_\alpha$$

is a standard disintegration of $F_{\varepsilon*}\mu_{\mathfrak{L}}$; moreover there exists an invertible map (which we still denote) $\hat F_\varepsilon : \operatorname{supp}\mathfrak{L}\to\operatorname{supp}\mathfrak{L}'$ so that $\boldsymbol{\pi}\circ\hat F_\varepsilon = F_\varepsilon\circ\boldsymbol{\pi}$ and $\hat F_{\varepsilon*}\hat\mu_{\mathfrak{L}} = \hat\mu_{\mathfrak{L}'}$.

5.2.2. *Pushforwards and filtrations.* A standard disintegration of $F_{\varepsilon*}\mu_{\mathfrak{L}}$ equipped with a map $\hat F_\varepsilon$ as above is called a (c_1,c_2)-*standard pushforward* of $\mathfrak{L}$. A (finite or countable) sequence $\{\mathfrak{L}_n\}$ is said to be a *sequence of* (c_1,c_2)-*standard pushforwards* of $\mathfrak{L}_0$ if for each $n\geq 0$, $\mathfrak{L}_{n+1}$ is a (c_1,c_2)-standard pushforward of $\mathfrak{L}_n$. At times, when some confusion might arise, we will write $\mathfrak{L}_n(\mathfrak{L})$ to make clear that $\mathfrak{L}_n$ is a pushforward of the family $\mathfrak{L}$.

Let us comment on the above important definition

Remark 5.5. *Consider a sequence of* (c_1,c_2)*-standard pushforwards of a standard pair* ℓ*; it is instructive to consider the sequence* $\mathfrak{L}_n$ *as a random process. For each* $p\in\operatorname{supp}\ell$*, let* $\boldsymbol{\alpha}_n : \operatorname{supp}\ell\to\mathcal{A}_n$ *be the map* $\boldsymbol{\alpha}_n = \boldsymbol{\pi}_{\mathcal{A}}\circ\hat F^n_\varepsilon$*.*[24] *Next, let us introduce the shorthand (abusing but suggestive) notation* $\ell_n(p) = \ell_n(\boldsymbol{\alpha}_n(p))$*. Accordingly, the sequence of functions* $\{\ell_n\}$ *can be regarded as a random process on the standard pair* ℓ *with values in the space of standard pairs.*

Observe moreover that our construction of $\hat F_\varepsilon$ implies the following important property: given $\alpha\in\mathcal{A}_n$ let $U_n(\alpha)$ be the connected subcurve $\boldsymbol{\alpha}_n^{-1}(\alpha)\subset\operatorname{supp}\ell$ whose n-image is $\ell_n(\alpha)$; then let $\mathcal{F}_n$ be the σ-algebra generated by the collection $\{U_n(\alpha)\}_{\alpha\in\mathcal{A}_n}$ (i.e., the σ-algebra generated by $\boldsymbol{\alpha}_n$). The sequence $\{\mathcal{F}_n\}$ is a filtration and the process $\{\boldsymbol{\alpha}_n\}$ (or, loosely speaking, $\{\ell_n\}$) is (naturally) adapted to such a filtration.

For each $p\in\operatorname{supp}\ell$ let us also introduce the shorthand notation $U_n(p) = U_n(\boldsymbol{\alpha}_n(p))$: observe that standard distortion arguments yield:

$$C_\#{}^{-1}\Lambda_n(p)^{-1}\leq|U_n(p)|\leq C_\#\Lambda_n(p)^{-1}, \tag{5.5}$$

where $\Lambda_n(p) = \frac{dx_n}{dx_0}$ and the derivative is taken along the curve; in particular $|U_n(p)|\leq C_\# 2^{-n}$.

Henceforth we assume c_1, c_2 to be fixed in order for Proposition 5.2 to hold and we fix δ to be so small that $\delta c_2 < 1/50$. Moreover, since c_1 and c_2 are now fixed, we will refer to a (c_1,c_2)-standard pair (resp. family, pushforward) simply as a *standard pair* (resp. family, pushforward); we let –with a further slight abuse of notation– $[F_\varepsilon\mathfrak{L}] = \mathbb{L}_{c_1,c_2}(F_{\varepsilon*}\mu_{\mathfrak{L}})$.

The proof of Proposition 5.2 in fact shows the existence of a standard pushforward of any standard family $\mathfrak{L}$. A pair ℓ is said to be N-*prestandard* if $F^N_{\varepsilon*}\mu_\ell$ admits

[24] Obviously, $\boldsymbol{\pi}_{\mathcal{A}}(\alpha,p) = \alpha$, for each $\alpha\in\mathcal{A}, p\in\mathbb{T}^2$.

a standard decomposition; we say that ℓ is *prestandard* if it is N-prestandard for some N. We say that a family $\mathfrak{L}$ is N*-prestandard* (resp. *prestandard*) if every $\ell \in \mathfrak{L}$ is N-prestandard (resp. prestandard).

Remark 5.6. *Consider a short standard pair ℓ of length at least δ_*: the proof of Proposition 5.2 implies that standard curves are expanded at an exponential rate. We can conclude that ℓ is N_{R}-prestandard with $N_{\mathrm{R}} \sim C_\# |\log \delta_*|$. We call N_{R} the* recovery time *of ℓ.*

Remark 5.7. *Let ℓ be a $(c_1, \gamma c_2)$-standard pair with $\gamma > 1$: the proof of Proposition 5.2 implies that densities on standard curves are regularized by the dynamics at an exponential rate; hence ℓ is N_{R}-prestandard with $N_{\mathrm{R}} \sim C_\# \log \gamma$. Again, we call N_{R} the* recovery time *of ℓ.*

Remark 5.8. *Consider a standard pair $\ell = (\mathbb{G}, \rho)$; by definition of standard density, we have, for any $x \in [a, b]$:*

$$\frac{\exp(-2c_2\delta)}{|\ell|} \le \rho(x) \le \frac{\exp(2c_2\delta)}{|\ell|}. \tag{5.6}$$

Consequently, for any constant $m_* \le 1/2$, we can define $\hat\rho(x)$ so that $\rho(x) = m_*/|\ell| + \hat\rho(x)$, and by the above estimate and our choice for δ we have $\hat\rho(x) \ge \rho(x)/3$. Consequently, since $\hat\rho' = \rho'$ (and thus $\hat\rho'' = \rho''$), we have:

$$\left\|\frac{\hat\rho'}{\hat\rho}\right\| \le 3\left\|\frac{\rho'}{\rho}\right\| \le 3c_2' \qquad\qquad \left\|\frac{\hat\rho''}{\hat\rho}\right\| \le 3\left\|\frac{\rho''}{\rho}\right\| \le 3c_2''$$

i.e. $\hat\rho(x) \in D_{3c_2}(G)$. The standard pair ℓ can thus be split *as:*

$$\ell \sim m_*\ell_* + (1 - m_*)\hat\ell,$$

where $\ell_* = (\mathbb{G}, 1/|\ell|)$ is a standard pair and $\hat\ell = (\mathbb{G}, \hat\rho/(1 - m_*))$ is a $\mathcal{O}(1)$-prestandard pair.

A fundamental property of standard families is that any SRB measure is a weak limit of a sequence of measures that can be disintegrated into standard families; we do not give the proof of this fact here, since we will prove a slightly stronger statement in Lemma 9.8.

6. Averaged dynamics

Standard pairs are a very convenient way to describe initial conditions which are, in a sense, well distributed with respect to the dynamics (see the discussion at the beginning of Section 8 for further comments). Let us start making this vague statement more concrete by stating some results which follow from the ones that are proved in [9]. First of all, let us introduce some useful notation; recall that for $p \in \mathbb{T}^2$ we denote $(x_n(p), \theta_n(p)) = F_\varepsilon^n(p)$; recall moreover the definition of ζ_n given in (4.11); let $z_n(p) = (\theta_n(p), \zeta_n(p))$ and define the polygonal interpolation

$$z_\varepsilon(t; p) = z_{\lfloor t\varepsilon^{-1}\rfloor}(p) + (t\varepsilon^{-1} - \lfloor t\varepsilon^{-1}\rfloor)(z_{\lfloor t\varepsilon^{-1}\rfloor+1}(p) - z_{\lfloor t\varepsilon^{-1}\rfloor}(p)).$$

Let us also introduce the functions θ_ε, ζ_ε so that $z_\varepsilon(t; p) = (\theta_\varepsilon(t; p), \zeta_\varepsilon(t; p))$. Note that if $p = (x_0, \theta_0)$ is distributed according to some measure μ (e.g. $\mu = \mu_\ell$, where ℓ is a standard pair), then for any $T > 0$ z_ε is naturally a random variable with values in $\mathcal{C}^0([0, T], \mathbb{T} \times \mathbb{R})$ (and likewise θ_ε and ζ_ε) and thus the pushforward $z_{\varepsilon *}\mu_\ell$ is a probability on $\mathcal{C}^0([0, T], \mathbb{T} \times \mathbb{R})$.

For any $t \ge 0$ and $\theta_* \in \mathbb{T}$, we define the function $\bar z(t; \theta_*) = (\bar\theta(t; \theta_*), \bar\zeta(t; \theta_*))$ to be the solution of the ODE problem

$$\frac{d}{dt}\bar z(t; \theta_*) = \left(\bar\omega(\bar\theta(t; \theta_*)), \bar\psi(\bar\theta(t; \theta_*))\right), \text{ with } \bar z(0; \theta_*) = (\theta_*, 0), \tag{6.1}$$

where $\bar{\omega}(\theta) = \int_{\mathbb{T}} \omega(x,\theta)\rho_\theta(x)dx$, $\bar{\psi}(\theta) = \int_{\mathbb{T}} \psi(x,\theta)\rho_\theta(x)dx$, ψ is defined in (4.10), and ρ_θ is the density of the unique absolutely continuous invariant measure for the expanding map f_θ. Observe that (6.1) admits a unique solution since we use the regularized function $\bar{\psi}$, which is smooth by Remark 4.1.

Then, the *Averaging Principle* (see [9, Theorem 2.1]) states that for any $T > 0$, if the initial conditions (x_0, θ_0) are distributed on a standard pair[25] that crosses $\{\theta = \theta_*\}$, the random variable z_ε converges in probability to $\bar{z}(\cdot;\theta_*)$ on $[0, T]$ as $\varepsilon \to 0$.

6.1. **Large and moderate deviations.** Given the above facts, it is then natural to attempt a description of the behavior of deviations from the averaged dynamics. For $p = (x_0, \theta_0)$, let us define:[26]

$$\Delta z(t;p) = (\Delta\theta(t;p), \Delta\zeta(t;p)) := z_\varepsilon(t;p) - \bar{z}(t;\theta_0). \tag{6.2}$$

A first rough (but useful) result that follows from [9] is the following

Theorem 6.1. *There exists $\varepsilon_0 > 0$ such that if we fix $T > 0$, then there exist $\bar{C} > 0$ such that for any $\varepsilon \leq \varepsilon_0$, $R \geq \bar{C}\sqrt{\varepsilon}$ and standard pair ℓ, we have*

$$\mu_\ell\left(\sup_{t\in[0,T]} \|\Delta z(t;\cdot)\| \geq R\right) \leq \exp(-c_T R^2 \varepsilon^{-1})$$

where c_T is a constant which depends on T only and $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^2$.

Proof. The proof immediately follows from [9, Proposition 2.3], which however is stated using different notations. In our case we let $A = (\omega, \psi)$; $\gamma_{A,\varepsilon}$ is the random element of $\mathcal{C}^0([0,T],\mathbb{R}^2)$ obtained by lifting $z_\varepsilon(\cdot) - z_\varepsilon(0)$ to $\mathbb{R}^2$ and $\bar{\gamma}_A(t,\theta_0)$ is the lift[27] of $\bar{z}(\cdot\,;\theta_0) - \bar{z}(0;\theta_0)$ to $\mathbb{R}^2$. Finally, the statement of [9, Proposition 2.3] involves a probability $\mathbb{P}_{A,\varepsilon}$ on the space $\mathcal{C}^0([0,T],\mathbb{R}^2)$ defined by $\mathbb{P}_{A,\varepsilon} = (\gamma_{A,\varepsilon})_*\mu_\ell$. Hence, by definition,

$$\mu_\ell\left(\sup_{t\in[0,T]} \|\Delta z(t;\cdot)\| \geq R\right) = \mathbb{P}_{A,\varepsilon}\left(\{\gamma \in \mathcal{C}^0([0,T],\mathbb{T}) \,:\, \|\gamma(\cdot) - \bar{\gamma}_A(\cdot,\theta^*_\ell)\|_\infty \geq R\}\right).$$

Now we can apply [9, Proposition 2.3], which gives the desired result. □

We now proceed to prove two other results (Theorems 6.3 and 6.4) which also follow from the Large Deviations estimates obtained in [9]. Loosely speaking these result state that if there exists an admissible (θ^0, θ^1)-path, then there exists an orbit of the real system connecting a neighborhood of $\{\theta = \theta^0\}$ to a neighborhood of $\{\theta = \theta^1\}$; conversely if all (θ^0, θ^1)-paths are not admissible, we would like to say that no orbit of the real system connects the two said sets. Indeed such statements could only have a chance to hold true in the limit $\varepsilon \to 0$, and even in this case there would be borderline admissible paths for which none of our statements would hold. In order to properly state our results in the case $\varepsilon > 0$ we need to refine the notion of admissibility that has been introduced in Section 2; recall the definition of $\Omega(\theta)$ given in (2.3); in particular $\Omega(\theta)$ is a closed interval for any $\theta \in \mathbb{T}$. For $\epsilon > 0$ and $\theta \in \mathbb{T}$, introduce the notations

$$\Omega^+_\epsilon(\theta) = \Omega(\theta) \cup \partial_\epsilon\Omega(\theta) \qquad\qquad \Omega^-_\epsilon(\theta) = \Omega(\theta) \setminus \partial_\epsilon\Omega(\theta),$$

[25] Notice that the definition of standard pair in fact depends on ε, therefore this convergence holds for any sequence of standard pairs which in turn weakly converges to the flat standard pair $\{\theta = \theta_*\}$.

[26] The function Δz (and thus $\Delta\theta$ and $\Delta\zeta$) indeed depend on ε (since so does z_ε); however, we do not explicitly add a subscript ε to ease notation.

[27] Such lifts are uniquely determined by the condition $\gamma_{A,\varepsilon}(0) = 0$ and $\bar{\gamma}_A(0) = 0$.

where $\partial_\epsilon\Omega(\theta) = \{b : \mathrm{dist}(b, \partial\Omega(\theta)) < \epsilon\}$. It is immediate to observe that if $\epsilon' < \epsilon$, $\Omega^+_{\epsilon'} \subset \Omega^+_\epsilon$ and $\Omega^-_{\epsilon'} \supset \Omega^-_\epsilon$; moreover $\mathrm{int}\,\Omega(\theta) = \bigcup_{\epsilon>0} \Omega^-_\epsilon$ and $\Omega(\theta) = \bigcap_{\epsilon>0} \Omega^+_\epsilon$. We say that a (θ^0, θ^1)-path h of length T is *ϵ-admissible* if for any $s \in [0, T]$ we have $\boldsymbol{\partial} h(s) \subset \mathrm{int}\,\Omega^-_\epsilon(h(s))$; likewise we say that h is *ϵ-forbidden* if for some $s \in [0, T]$ we have $\boldsymbol{\partial} h(s) \not\subset \Omega^+_\epsilon(h(s))$. Observe that, by definition, and by compactness of the graph of $\boldsymbol{\partial} h(s)$, h is admissible if and only if it is ϵ-admissible for some $\epsilon > 0$; likewise if h is ϵ-admissible (resp. not ϵ-forbidden), then there exists some $\epsilon' > \epsilon$ (resp. $\epsilon' < \epsilon$) so that h is ϵ'-admissible (resp. not ϵ'-forbidden).

First of all let us prove an auxiliary

Lemma 6.2. *Let $\bar\omega^+(\theta) = \max \Omega(\theta)$ and $\bar\omega^-(\theta) = \min \Omega(\theta)$; then $\bar\omega^\pm$ are (uniformly) continuous functions.*

Proof. Let us prove the statement for $\bar\omega^+(\theta)$ (the statement for $\bar\omega^-$ follows by noting that $\min \Omega(\theta) = -\max[-\Omega(\theta)]$). It is possible to characterize $\bar\omega^+$ as follows (see [25, Proposition 2.1]):

$$\bar\omega^+(\theta) = \sup_{x\in\mathbb{T}} \limsup_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \omega(f^n_\theta(x), \theta).$$

Observe that for any $\varrho > 0$, if $|\theta^0 - \theta^1|$ is sufficiently small, there exists a homeomorphism $\Phi : \mathbb{T} \to \mathbb{T}$ with $\|\Phi - \mathbb{1}\|_{\mathcal{C}^0} < \varrho$ so that $f_{\theta^1} = \Phi \circ f_{\theta^0} \circ \Phi^{-1}$ (see e.g. [26, Lemma 2]). We gather that

$$\bar\omega^+(\theta^1) = \sup_{x\in\mathbb{T}} \limsup_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N} \omega(f^n_{\theta^1}(x), \theta^1) = \sup_{x\in\mathbb{T}} \limsup_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N} \omega(\Phi \circ f^n_{\theta^0}(x), \theta^1).$$

Since $\|\omega(\cdot, \theta^1) - \omega(\Phi(\cdot), \theta^1)\|_{\mathcal{C}^0}$ can be made arbitrarily small by choosing ϱ arbitrarily small, and since ω is a smooth function, our lemma follows. $\square$

Let us now prove the following lower bound.

Theorem 6.3. *Assume that there exists an ϵ-admissible (θ^0, θ^1)-path of length $T > 0$ for some $T > 0$ and $\epsilon > 0$; if $\varepsilon > 0$ is sufficiently small (depending on T and ϵ), for any standard pair ℓ whose support intersects $\{\theta = \theta^0\}$ we have, if C is sufficiently large,*

$$\mu_\ell(\theta_{\lfloor T\varepsilon^{-1}\rfloor} \in B(\theta^1, C\varepsilon^{5/12})) \geq \exp(-C_T\varepsilon^{-1}). \tag{6.3}$$

Proof. Once again we plan to apply Large Deviations estimates from [9], which are stated using different notations. Define $A = \omega$, and let $\gamma_{A,\varepsilon}$ be the lift of the random element $\theta_\varepsilon(\cdot) - \theta_\varepsilon(0)$ to $\mathbb{R}$ starting at 0; let $\mathbb{P}_{A,\varepsilon} = \gamma_{A,\varepsilon*}\mu_\ell$. Let $\bar h$ be an ϵ-admissible (θ^0, θ^1)-path of length T (which exists by hypothesis); let $\bar\gamma$ be the lift of $\bar h - \theta^0$ to $\mathbb{R}$ starting at 0. Let us assume ε is sufficiently small (with respect to ϵ and T) to be specified later and define the set

$$Q_\varepsilon = \{\gamma \in \mathcal{C}^0([0, T], \mathbb{T}) : \gamma(T) \in B(\bar\gamma(T), C\varepsilon^{5/12})\}.$$

Then, by construction and by [9, Theorem 2.4], for all $\Delta_* > 0$ if ε is sufficiently small (depending on Δ_*), we conclude that

$$\mu_\ell(\theta_{\lfloor T\varepsilon^{-1}\rfloor} \in B(\theta^1, C\varepsilon^{5/12})) \geq \mathbb{P}_{A,\varepsilon}(Q_\varepsilon) \geq \exp(-\varepsilon^{-1}C_{T,\Delta_*} \inf_{\gamma\in Q^-_\varepsilon} \mathscr{I}^+_{\theta^0,\Delta_*}(\gamma))$$

where $Q^-_\varepsilon = \{\gamma : B_{\mathcal{C}^0([0,T])}(\gamma, C_\#\varepsilon^{5/12}) \subset Q_\varepsilon\}$ for some universal $C_\#$ and the *rate function* $\mathscr{I}^+_{\theta^0,\Delta_*}(\gamma)$ is finite, provided that γ is Lipschitz, $\gamma(0) = 0$ and $h = \gamma + \theta^0$ is ϵ-admissible for some $\epsilon > \Delta_*$ (see [9, Section 6] for additional details). Let us thus choose (e.g.) $\Delta_* = \epsilon/2$, so that $\mathscr{I}^+_{\theta^0,\Delta_*}(\bar\gamma) < \infty$; hence, provided that $\bar\gamma \in Q^-_\varepsilon$ (which holds true if $C > C_\#$), we can conclude

$$\mu_\ell(\theta_{\lfloor T\varepsilon^{-1}\rfloor} \in B(\theta^1, C\varepsilon^{5/12})) \geq \exp(-C_{T,\bar\gamma}\varepsilon^{-1}).$$

Observe that, in the above inequality, the right hand side depends on the path $\bar h$ (via $\bar\gamma$), but (6.3) requires a uniform bound. However, Proposition 5.2 implies that if some point in the n-th image of a standard pair ℓ belongs to $\{\theta \in B(\theta^1, C\varepsilon^{5/12})\}$, then there is a whole standard pair in any n-pushforward of ℓ which belongs to (a $\mathcal{O}(\varepsilon)$-neighborhood of) the given set. Hence the above inequality indeed implies (6.3), where $\exp(C_\#)$ is proportional to the maximal expansion of F_ε along a standard curve. □

We now prove what can be regarded as a converse of the above theorem.

Theorem 6.4. *For any $\epsilon > 0$, there exist $\varrho > 0$ and $T_F > 0$ so that if ε is sufficiently small (depending on ϵ) the following holds: if every (θ^0, θ^1)-path is ϵ-forbidden, then for any $N \geq \lfloor T_F\varepsilon^{-1}\rfloor$*

$$F^N_\varepsilon(\{\theta \in B(\theta^0, \varrho)\}) \cap \{\theta \in B(\theta^1, \varrho)\} = \emptyset. \tag{6.4}$$

We first need to prove two auxiliary lemmata

Lemma 6.5. *Every (θ^0, θ^1)-path is ϵ-forbidden if and only if*

$$\min_{\theta\in[\theta^0,\theta^1]} \bar\omega^+(\theta) \leq -\epsilon \qquad \textit{and} \qquad \max_{\theta\in[\theta^1,\theta^0]} \bar\omega^-(\theta) \geq \epsilon. \tag{6.5}$$

Proof. First, let $\min_{\theta\in[\theta^0,\theta^1]} \bar\omega^+(\theta) > -\epsilon$; in particular there exists $0 < \epsilon_* < \epsilon$ and $\varrho_* > 0$ so that $\min_{\theta\in[\theta^0,\theta^1]} \bar\omega^+(\theta) + \epsilon_* > \varrho_*$. Let h_* be a[28] path solving the differential equation $h_*'(s) = \bar\omega^+(h_*(s)) + \epsilon_*$ with initial condition $h_*(0) = \theta^0$; since $h_*'(s) \geq \varrho_*$ there exists $T \leq 1/\varrho_*$ so that $h_*(T) = \theta^1$, so h_* is a (θ^0, θ^1)-path. Our construction moreover guarantees that $\bar\omega^+(h(s)) \leq h_*'(s) < \bar\omega^+(h(s)) + \epsilon$, which implies that h_* is not ϵ-forbidden and contradicts our assumptions. If we assume $\max_{\theta\in[\theta^1,\theta^0]} \bar\omega^-(\theta) < \epsilon$, a similar argument also allows to construct a non ϵ-forbidden (θ^0, θ^1)-path. This concludes the proof of the direct implication.

Let us prove the reverse implication. Let h be a (θ^0, θ^1)-path of length T; without loss of generality we can assume that $h(s) \notin \{\theta^0, \theta^1\}$ if $s \in (0, T)$. Then either $h([0, T]) = [\theta^0, \theta^1]$ or $h([0, T]) = [\theta^1, \theta^0]$. Let us assume the first possibility; the second case can be completed by the reader following an analogous argument. Assume by contradiction that h is not ϵ-forbidden; then there exists $\epsilon' < \epsilon$ so that h is not ϵ'-forbidden. By (6.5), there exist an interval $[\theta_-, \theta_+] \subset (\theta^0, \theta^1)$ so that

$$\max \Omega_{\epsilon'}(\theta) < 0 \qquad\qquad \text{if } \theta \in [\theta_-, \theta_+]. \tag{6.6}$$

Then by our assumptions there exist $0 < s_- < s_+ < T$ so that $h(s_-) = \theta_-$, $h(s_+) = \theta_+$ and $h([s_-, s_+]) = [\theta_-, \theta_+]$. By Borg's Mean Value Theorem (see [6, Theorem 2.2.4]) there exists $s \in [s_-, s_+]$ so that $\boldsymbol{\partial} h(s) \ni 2\varrho/(s_+ - s_-) > 0$. This gives a contradiction with (6.6), since we assume h to be not ϵ'-forbidden.

□

Remark 6.6. *The same argument used in the above proof indeed shows that no (θ^0, θ^1)-path is ϵ-admissible if and only if*

$$\min_{\theta\in[\theta^0,\theta^1]} \bar\omega^+(\theta) \leq \epsilon \qquad \textit{and} \qquad \max_{\theta\in[\theta^1,\theta^0]} \bar\omega^-(\theta) \geq -\epsilon. \tag{6.7}$$

[28] The function $\bar\omega^+(\theta)$ is continuous, therefore a solution of the given differential equation exists (by Cauchy–Peano Theorem), but in general is not unique.

Lemma 6.7. *For any $\bar{\epsilon}$, if ε is sufficiently small, for any $\varepsilon^{-1/3} < n < C_\# \bar{\epsilon}^{1/2}\varepsilon^{-1/2}$, we have, for any (x_0, θ_0):*

$$\bar{\omega}^-(\theta_0) - \bar{\epsilon} < \frac{\theta_n - \theta_0}{\varepsilon n} < \bar{\omega}^+(\theta_0) + \bar{\epsilon}$$

Proof. We will only prove the right inequality (the left one follows by replacing ω with $-\omega$ in our argument below). Observe that, by [25, Proposition 2.1] we have:

$$\bar{\omega}^+(\theta) = \limsup_{n\to\infty} \frac{1}{n} \sup_{x\in\mathbb{T}} \sum_{k=0}^{n-1} \omega(f_\theta^k(x), \theta).$$

Hence, there exists $\bar{n}$ so that if $n > \bar{n}$

$$\frac{1}{n} \sup_{x\in\mathbb{T}} \sum_{k=0}^{n-1} \omega(f_\theta^k(x), \theta) \le \bar{\omega}^+(\theta) + \bar{\epsilon}/2. \tag{6.8}$$

By compactness (remember Lemma 6.2), $\bar{n}$ can be chosen to be uniform in θ. Next, assume $\varepsilon < \bar{n}^{-3}$. By definition

$$\frac{\theta_n - \theta_0}{\varepsilon n} = \frac{1}{n}\sum_{k=0}^{n-1} \omega(x_k, \theta_k).$$

By [9, Lemma 4.2], for any x_0 there exists x_* so that $|f_{\theta_0}^k(x_*) - x_k| < \varepsilon k^2$; moreover by definition $|\theta_k - \theta_0| < \varepsilon k$. Accordingly, provided $C_\#\varepsilon n^2 \le \bar{\epsilon}/2$,

$$\left|\frac{1}{n}\sum_{k=0}^{n-1} \omega(x_k, \theta_k) - \frac{1}{n}\sum_{k=0}^{n-1} \omega(f_\theta^k(x_*), \theta_0)\right| \le C_\#\varepsilon n^2 \le \bar{\epsilon}/2.$$

The above estimate, together with equation (6.8), concludes the proof of the Lemma since $n > \varepsilon^{-1/3} > \bar{n}$. $\square$

We can now give the

Proof of Theorem 6.4. Uniform continuity of $\bar{\omega}^\pm$ (proved in Lemma 6.2) guarantees that there exists $\varrho > 0$ so that if $|\theta - \theta'| < \varrho$, $|\bar{\omega}^\pm(\theta) - \bar{\omega}^\pm(\theta')| < \epsilon/4$; by Lemma 6.5 there exists $\theta_+ \in [\theta^0, \theta^1]$ (resp. $\theta_- \in [\theta^1, \theta^0]$) so that $\bar{\omega}^+(\theta_+) \le -\epsilon$ (resp. $\bar{\omega}^-(\theta_-) \ge \epsilon$); moreover $|\theta_+ - \theta_-| > 8\varrho$. Let us define: $U^+ = [\theta_- - \varrho, \theta_+ + \varrho]$, $U = [\theta_- + \varrho, \theta_+ - \varrho]$ and $U^- = [\theta_- + 2\varrho, \theta_+ - 2\varrho]$. Let $p = (x_0, \theta_0)$, and as usual let $(x_l, \theta_l) = F_\varepsilon^l(x_0, \theta_0)$; then we claim:

(a) if $\theta_0 \in U^+$, then $\theta_l \in U^-$ for some $0 \le l \le T_{\mathrm{F}}\varepsilon^{-1}$
(b) if $\theta_0 \in U^-$, then $\theta_l \in U$ for any $l \ge 0$.

Let us prove (a): if $\theta_0 \in U^-$ there is nothing to prove, otherwise we have either $\theta_0 \in B(\theta_+, 2\varrho)$ or $\theta_0 \in B(\theta_-, 2\varrho)$. Assume the first possibility (the second one can be dealt with similarly). Observe that by definition, for any $\theta \in B(\theta_+, 2\varrho)$ we have $\bar{\omega}^+(\theta) < \epsilon/2$. Let n and ε satisfy the hypotheses of Lemma 6.7 with $\bar{\epsilon} = \epsilon/4$; and let $\bar{n} = \inf\{n : \theta_n \notin B(\theta_+, 2\varrho)\}$ then if $kn < \bar{n}$ Lemma 6.7 implies that $\theta_{(k+1)n} - \theta_{kn} < -\varepsilon n\epsilon/4$; we conclude that $\theta_{kn} - \theta_0 < -\varepsilon kn\epsilon/4$, so (a) holds by choosing $T_{\mathrm{F}} > 16\varrho/\epsilon$.

Let us prove (b): assume by contradiction that there exists (x_0, θ_0) so that $\theta_0 \in U^-$ and $\theta_n \notin U$ for some n; let $n' = \min\{k : \theta_l \notin U^- \text{ for any } k \le l \le n\}$. observe that by construction $n' - n > \varrho\varepsilon^{-1}/\|\omega\|$; assume $\theta_{n'} \in B(\theta_+, 2\varrho)$ (the possibility $\theta_{n'} \in B(\theta_-, 2\varrho)$ can be dealt with similarly). Then by Lemma 6.7 we know that if ε is sufficiently small $\theta_{n'+\varepsilon^{-1/3}} - \theta_{n'} < -\varepsilon^{2/3}\epsilon/4$, which implies that $\theta_{n'+\varepsilon^{-1/3}} \in U^-$, contradicting the definition of n', since $n' < n' + \varepsilon^{-1/3} \le n$: hence we proved (b).

We can now conclude: in fact the existence of (x_0, θ_0) so that $\theta_0 \in B(\theta^0, \varrho)$ and $\theta_n \in B(\theta^1, \varrho)$ for some $n > T_{\mathrm{F}}\varepsilon^{-1}$ would contradict (a) and (b), since by construction $\theta_0 \in U^+$ and $\theta_n \notin U$. $\square$

6.2. **Local Central Limit Theorem.** In [9] we also obtained a Local Central Limit Theorem (see [9, Theorem 2.8]):

Theorem 6.8. *For any $T > 0$, there exists $\varepsilon_0 > 0$ and $0 < \alpha_0 < 1$ so that the following holds. For any compact interval $I \subset \mathbb{R}$, real numbers $\kappa > 0$, $\varepsilon \in (0, \varepsilon_0)$ and $t \in [\varepsilon^{1/2000}, T]$, any standard pair ℓ which intersects $\{\theta = \theta_0\}$, we have:*

$$\varepsilon^{-1/2}\mu_\ell(\Delta\theta(t;\cdot) \in \varepsilon I + \kappa\varepsilon^{1/2}) = \mathrm{Leb}\, I \cdot \frac{e^{-\kappa^2/(2\boldsymbol{\sigma}_t^2)}}{\boldsymbol{\sigma}_t\sqrt{2\pi}} + \mathcal{O}(\varepsilon^{\alpha_0}). \tag{6.9}$$

where the variance $\boldsymbol{\sigma}_t^2 = \boldsymbol{\sigma}_t^2(\theta_0)$ is given by

$$\boldsymbol{\sigma}_t^2 = \int_0^t e^{2\int_s^t \bar\omega'(\bar\theta(r,\theta_0))dr}\hat{\boldsymbol{\sigma}}^2(\bar\theta(s,\theta_0))ds, \tag{6.10}$$

and $\hat{\boldsymbol{\sigma}}^2(\theta)$ is given by the usual Green–Kubo formula

$$\hat{\boldsymbol{\sigma}}^2(\theta) = \int_{\mathbb{T}}\left[\hat\omega^2(x,\theta) + 2\sum_{m=1}^{\infty}\hat\omega(f_\theta^m(x),\theta)\hat\omega(x,\theta)\right]\rho_\theta(x)dx,$$

where $\hat\omega(x,\theta) = \omega(x,\theta) - \bar\omega(\theta)$.

Observe that, $\hat{\boldsymbol{\sigma}}$ defined above is uniformly bounded away from 0 by Assumption (A3) and compactness of $\mathbb{T}$; hence we conclude that

$$C_\# t \le \boldsymbol{\sigma}_t^2 \le C_\# \exp(c_\# t)t \tag{6.11}$$

6.3. **Averaged dynamics: description.** In this section we will describe the averaged dynamics of the variable θ_n and of the auxiliary variable ζ_n. As we noted earlier, assumption (A1) enables us to give a simple description of the averaged dynamics $\bar\theta$. Let us start by fixing some terminology and notation. As already briefly mentioned in Section 2, we define the intervals:

$$I_{k,-} := [\theta_{k,+}, \theta_{k+1,+}] \ni \theta_{k,-} \qquad I_{k,+} := [\theta_{k-1,-}, \theta_{k,-}] \ni \theta_{k,+}.$$

By (6.1), any point in $\operatorname{int} I_{k,-}$ (resp. $\operatorname{int} I_{k,+}$) converges in forward time (resp. backward time) to $\theta_{k,-}$ (resp. $\theta_{k,+}$): we thus call $\operatorname{int} I_{k,-}$ (resp. $\operatorname{int} I_{k,+}$) the *(forward) basin of attraction* of $\theta_{k,-}$ (resp. *backward basin of attraction* of $\theta_{k,+}$). In particular, any sufficiently small ball B_k containing $\theta_{k,-}$ is forward-invariant, that is:

$$\forall\ k \in \{1, \cdots, n_{\mathrm{Z}}\},\, t > 0,\, \theta_0 \in B_k \qquad |\bar\theta(t,\theta_0) - \theta_{k,-}| \le |\theta_0 - \theta_{k,-}|. \tag{6.12}$$

Let us now define the sets

$$W_{k,-} := \{\theta \in I_{k,-} \,:\, \bar\omega'(\theta) < \bar\omega'(\theta_{k,-})/2;\ \bar\psi(\theta) < -3/4\} \tag{6.13a}$$

$$W_{k,+} := \{\theta \in I_{k,+} \,:\, \bar\omega'(\theta) > \bar\omega'(\theta_{k,+})/2\}; \tag{6.13b}$$

observe that $W_{k,-} \neq \emptyset$ by Assumption (A2) and Remark 4.1. For fixed $r_-, r_+ > 0$ small, define $H_k = B(\theta_{k,-}, r_-)$ and $S_k = B(\theta_{k,+}, r_+)$. We prescribe r_- (resp. r_+) to be small enough so that $H_k \subset W_{k,-}$ (resp. $S_k \subset W_{k,+}$) for any k. Define moreover

$$\mathbb{H} = \bigcup_k H_k \qquad \mathbb{S} = \bigcup_k S_k. \tag{6.14}$$

Finally, let us define $\hat H_k = B(\theta_{k,-}, 3r_-/4)$ and $\hat{\mathbb{H}} = \bigcup_k \hat H_k$.

By (6.12), each of the sets H_k is invariant for the averaged dynamics; using (6.1) we thus conclude that ζ has an average negative drift on H_k whose rate is strictly less than $-1/2$. This will imply that center vectors are, *on average*, contracted at an exponential rate, as long as the trajectory stays in one of the H_k's. We will then use Large Deviation estimates (i.e. Theorem 6.1), to obtain similar result for the real dynamics.

6.4. **Averaged dynamics: further properties.** Let us now introduce a few additional notions which we will need, in particular, when (A4) does not hold.

Remark 6.9. *The goal of this section is to define trapping sets for the dynamics in an abstract manner. The reason for this is twofold: first it makes very clear where assumption (A5) is actually used (see Lemma 6.14); second, it allows to prove all our results with little or no reference to the actual geometry of the trapping sets, which we think would be useful for further generalization to higher dimensional settings.*

For each $\theta \in \mathbb{T}$ and $T > 0$ we define the sets

$$A^+_{\epsilon,\theta,T} = \{\theta' \in \mathbb{T} : \exists\, (\theta,\theta')\text{-path of length} \leq T \text{ that is not } \epsilon\text{-forbidden})\};$$
$$A^-_{\epsilon,\theta,T} = \{\theta' \in \mathbb{T} : \exists\, \epsilon\text{-admissible } (\theta',\theta)\text{-path of length} \leq T)\}.$$

Observe that $A^+_{\epsilon,\theta,T}$ is given by *end points* of paths *starting from* θ, while $A^-_{\epsilon,\theta,T}$ is given by *starting points* of paths *ending at* θ. We denote with $A^\pm_{\epsilon,\theta} = \bigcup_{T>0} A^\pm_{\epsilon,\theta,T}$.

Lemma 6.10. *The following properties hold for any $\theta \in \mathbb{T}$ and $\epsilon > 0$:*

(a) *$A^\pm_{\epsilon,\theta}$ is connected (i.e. an interval) for any $\theta \in \mathbb{T}$;*
(b) *$\theta \in A^-_{\epsilon,\theta'} \Rightarrow A^+_{\epsilon,\theta} \cap A^-_{\epsilon,\theta'} \neq \emptyset \Rightarrow \theta' \in A^+_{\epsilon,\theta}$;*
(c) *if $\theta' \in A^\pm_{\epsilon,\theta}$, then $A^\pm_{\epsilon,\theta'} \subset A^\pm_{\epsilon,\theta}$;*
(d) *if $0 \in \operatorname{int}\Omega(\theta)$ then $\theta \in A^\pm_{\epsilon,\theta}$, provided that ϵ has been chosen small enough.*

The proof of the above properties readily follows from the definition and it is left to the reader

Lemma 6.11. *$A^\pm_{\epsilon,\theta,T}$ is an open set for any $T > 0$, $\epsilon > 0$ and $\theta \in \mathbb{T}$.*

Proof. Let us prove the statement for $A^+_{\epsilon,\theta}$; the proof for $A^-_{\epsilon,\theta}$ is similar. Assume that $\theta' \in A^+_{\epsilon,\theta}$; then there exists a (θ,θ')-path h which is not ϵ-forbidden. Recall that the graph $\{s, \boldsymbol{\partial} h(s)\}_{s\in[0,T]}$ is compact and that by definition of a ϵ-forbidden path, for any s $\boldsymbol{\partial} h(s) \subset \Omega^+_\epsilon(s)$, which is an open set. We conclude that if $|\varrho|$ is sufficiently small, the path $h_\varrho(s) = h(s) + \varrho\epsilon s$ is also not ϵ-forbidden. Then our statement holds since $h_\varrho(T) = \theta' + \varrho\epsilon T$. □

Notice that if $\epsilon' < \epsilon$ we have $A^+_{\epsilon',\theta} \subset A^+_{\epsilon,\theta}$ and $A^-_{\epsilon',\theta} \supset A^-_{\epsilon,\theta}$; in particular we can define $A^+_\theta = \bigcap_{\epsilon>0} A^+_{\epsilon,\theta}$ and $A^-_\theta = \bigcup_{\epsilon>0} A^-_{\epsilon,\theta}$.

Further, by (A3), we know that for any $\theta \in \mathbb{T}$ we have $\bar\omega(\theta) \in \operatorname{int}\Omega(\theta)$ (see e.g. the discussion before [9, Proposition 2.2]). In particular, there exists $\varrho > 0$ so that $\Omega(\theta_{i,\pm}) \supset (-\varrho,\varrho)$ for each $i = 1,\cdots,n_{\mathrm{Z}}$. Since f_θ is a smooth family of expanding maps, we conclude that, possibly by choosing a smaller ϱ, we can find open neighborhoods $\Theta_{i,\pm} \ni \theta_{i,\pm}$ so that for each $i = 1,\cdots,n_{\mathrm{Z}}$ and $\theta \in \Theta_{i,\pm}$, we have $\Omega(\theta) \supset (-\varrho,\varrho)$. We conclude (unsurprisingly) that $I_{i,-} \subset A^-_{\epsilon,\theta_{i,-}}$, provided that ϵ is small enough. Moreover, observe that by definition $\forall\,\theta,\theta' \in \Theta_{i,\pm}$ we have $A^\pm_{\epsilon,\theta} = A^\pm_{\epsilon,\theta'}$. Finally, observe that by possibly decreasing ϱ we can assume that if θ does not belong to any of the $\Theta_{i,\pm}$'s, $|\bar\omega(\theta)| > \varrho$. We conclude that for sufficiently small ϵ, $I_{i,-} \subset A^-_{\epsilon,\theta_{i,-},T_I}$ where we define $T_I := 2\varrho^{-1}$.

We are now in the position to prove genericity of Condition (A5):

Lemma 6.12. *Condition (A5) holds for a set that is $\mathcal{C}^4$-open and dense in the set of ω which satisfy (A1), (A2) and (A3).*

Proof. Let F_ε be so that (A1), (A2) and (A3) are satisfied. Assume that there exists an interval J so that neither property i nor ii holds. To fix ideas let us assume that $\bar\omega(\theta) \geq 0$ for any $\theta \in J$ (the other case can be treated analogously) i.e. $J = [\theta_{k,+}, \theta_{k,-}]$ for some $k \in \{1, \cdots, n_Z\}$. By assumption there exists $\theta_* \in J$ so that $0 \in \partial\Omega(\theta_*)$. By our previous discussion we know that $\theta_* \notin \Theta_{k,+} \cup \Theta_{k,-}$; thus let $\bar\omega_J(\theta)$ be a C^∞ bump function which is 0 on $\mathbb{T}\setminus J$ and 1 on $J\setminus\Theta_{k,+}\cup\Theta_{k,-}$. Then for any $\epsilon > 0$, if we let $\omega(x,\theta) \mapsto \omega(x,\theta) + \epsilon\bar\omega_J(\theta)$, we obtain a dynamical system which satisfies property ii in J, since $0 \notin \operatorname{cl}\Omega(\theta_*)$. Since ω_J is supported away from $\mathbb{T}\setminus J$, the same construction can be applied independently to all other J's for which (A5) is not satisfied, which concludes the proof. Observe in fact that our construction does not interfere with assumptions (A3) (since our perturbation is a function that is constant in x), (A1) and (A2) (since our perturbation is supported away from the set $\{\theta_{i,\pm}\}_{i=1,\cdots,n_Z}$). □

For $i = 1, \cdots, n_Z$ define the *ϵ-trapping set* of $\theta_{i,-}$

$$\mathcal{T}_{\epsilon,i} = \{\theta \in \mathbb{T} : A^+_{\epsilon,\theta} \subset A^-_{\epsilon,\theta_{i,-}}\}.$$

Observe that if $\epsilon' < \epsilon$, we have $\mathcal{T}_{\epsilon',i} \supset \mathcal{T}_{\epsilon,i}$. We say that a sink $\theta_{i,-}$ is *recurrent* if $\mathcal{T}_{\epsilon,i} \neq \emptyset$ for sufficiently small ϵ and *transient* otherwise.

Lemma 6.13 (Properties of trapping sets)**.** *If ϵ is sufficiently small, the following properties hold:*

(a) *There exists $T_\mathcal{T} > 0$ so that $\mathcal{T}_{\epsilon,i} \subset A^-_{\epsilon,\theta_{i,-},T_\mathcal{T}}$;*
(b) *if $\theta \in \mathcal{T}_{\epsilon,i}$, then $A^+_{\epsilon,\theta} \subset \mathcal{T}_{\epsilon,i}$;*
(c) *either $\mathcal{T}_{\epsilon,i} \cap \mathcal{T}_{\epsilon,j} = \emptyset$ or $\mathcal{T}_{\epsilon,i} = \mathcal{T}_{\epsilon,j}$;*
(d) *$\theta_{i,-}$ is recurrent if and only if $\mathcal{T}_{\epsilon,i} \supset \Theta_{i,-}$;*
(e) *$\theta_{i,-}$ is transient if and only if $\exists\, j \in \{1, \cdots, n_Z\}$ s.t. $\theta_{j,-} \in A^+_{\epsilon,\theta_{i,-}} \setminus A^-_{\epsilon,\theta_{i,-}}$;*

Proof. Choose an arbitrary $\theta \in \mathcal{T}_{\epsilon,i}$ and let $j \in \{1, \cdots, n_Z\}$ so that $I_{j,-} \ni \theta$; if ϵ is sufficiently small, $\theta \in A^-_{\epsilon,\theta_{j,-},T_I}$; then, by definition of $\mathcal{T}_{\epsilon,i}$, we have $\theta_{j,-} \in A^-_{\epsilon,\theta_{i,-}}$. Observe that since there are only finitely many pairs of sinks, there exists $T' > 0$ so that for any $i, j \in \{1, \cdots, n_Z\}$, either $\theta_{i,-} \in A^-_{\epsilon,\theta_{j,-},T'}$ or $\theta_{i,-} \notin A^-_{\epsilon,\theta_{j,-}}$. Hence, by Lemma 6.10(c) we have $\theta \in A^-_{\epsilon,\theta_{i,-},T_\mathcal{T}}$, where we set $T_\mathcal{T} = T_I + T'$; this proves (a). On the other hand, (b) follows from the fact that if $\theta' \in A^+_{\epsilon,\theta}$, we have $A^+_{\epsilon,\theta'} \subset A^+_{\epsilon,\theta} \subset A^-_{\epsilon,\theta_{i,-}}$. Assume now $\theta \in \mathcal{T}_{\epsilon,i} \cap \mathcal{T}_{\epsilon,j} \neq \emptyset$: then by (b) we have $\theta_{i,-} \in \mathcal{T}_{\epsilon,j}$ and $\theta_{j,-} \in \mathcal{T}_{\epsilon,i}$; by (a), we conclude that $\theta_{i,-} \in A^-_{\epsilon,\theta_{j,-}}$ and $\theta_{j,-} \in A^-_{\epsilon,\theta_{i,-}}$, which by definition imply respectively that $\mathcal{T}_{\epsilon,i} \subset \mathcal{T}_{\epsilon,j}$ and $\mathcal{T}_{\epsilon,j} \subset \mathcal{T}_{\epsilon,i}$, proving (c).

Now assume $\mathcal{T}_{\epsilon,i} \neq \emptyset$ and let $\theta \in \mathcal{T}_{\epsilon,i}$: by (a) $\theta_{i,-} \in A^+_{\epsilon,\theta}$ and thus, by (b), $\theta_{i,-} \in \mathcal{T}_{\epsilon,i}$. Then, by construction, $\forall\,\theta' \in \Theta_{i,-}$, $A^+_{\epsilon,\theta'} = A^+_{\epsilon,\theta_{i,-}}$, which in particular proves (d). In turn (d) implies that $\mathcal{T}_{\epsilon,i} = \emptyset$ if and only if $A^+_{\epsilon,\theta_{i,-}} \setminus A^-_{\epsilon,\theta_{i,-}} \neq \emptyset$; let $\theta \in A^+_{\epsilon,\theta_{i,-}} \setminus A^-_{\epsilon,\theta_{i,-}}$. Then $\theta \in I_{j,-} \subset A^-_{\epsilon,\theta_{j,-}}$ for some $j \neq i$, which in turn implies (e). □

In general it is possible for a system to have no recurrent sinks. However, as the following lemma shows, Condition (A5) guarantees that this cannot happen.

Lemma 6.14. *Assume that Condition (A5) holds and ϵ is sufficiently small: then*

(a) *for any $i, j \in \{1, \cdots, n_Z\}$ we have $\theta_{j,-} \in A^+_{\epsilon,\theta_{i,-}}$ if and only if $\theta_{i,-} \in A^-_{\epsilon,\theta_{j,-}}$.*
(b) *for any $\theta \in \mathbb{T}$ there exists a recurrent sink $\theta_{i,-}$ so that $\theta \in A^-_{\epsilon,\theta_{i,-},T_\mathcal{T}}$.*

Proof. By Lemma 6.10(b) we have that if $\theta_{i,-} \in A^-_{\epsilon,\theta_{j,-}}$ then $\theta_{j,-} \in A^+_{\epsilon,\theta_{i,-}}$ thus we only need to prove the direct implication. Assume by contradiction that one can find arbitrarily small ϵ so that there exists a non ϵ-forbidden $(\theta_{i,-}, \theta_{j,-})$-path but yet all $(\theta_{i,-},\theta_{j,-})$-paths are not ϵ-admissible. By Lemma 6.5 we gather that $\min_{\theta\in[\theta_{i,-},\theta_{j,-}]}\bar\omega^+(\theta) > -\epsilon$ or $\max_{\theta\in[\theta_{j,-},\theta_{i,-}]}\bar\omega^-(\theta) < \epsilon$. Indeed by the same argument used in the proof of Lemma 6.5 we can conclude that if every $(\theta_{i,-},\theta_{j,-})$-path is not ϵ-admissible, then $\min_{\theta\in[\theta_{i,-},\theta_{j,-}]}\bar\omega^+(\theta) \le \epsilon$ and $\max_{\theta\in[\theta_{j,-},\theta_{i,-}]}\bar\omega^-(\theta) \ge -\epsilon$. Since ϵ is arbitrarily small, we conclude that $\min_{\theta\in[\theta_{i,-},\theta_{j,-}]}\bar\omega^+(\theta) = 0$ or $\max_{\theta\in[\theta_{j,-},\theta_{i,-}]}\bar\omega^-(\theta) = 0$. In either case, assumption (A5) is violated, which is a contradiction. This proves (a).

Let now $\theta\in\mathbb{T}$ be arbitrary and assume by contradiction that every sink $\theta_{i,-}$ so that $A^-_{\epsilon,\theta_{i,-}} \ni \theta$ is transient. Let i_0 so that $\theta\in I_{i_0,-}$: in particular $\theta\in A^-_{\epsilon,\theta_{i_0,-},T_I}$. Since $\theta_{i_0,-}$ is transient, by 6.13(e) there exists another sink $\theta_{i_1,-}\in A^+_{\epsilon,\theta_{i_0,-}}\setminus A^-_{\epsilon,\theta_{i_0,-}}$; by part (a) and Lemma 6.10(c) we conclude that $\theta_{i_0,-}\in A^-_{\epsilon,\theta_{i_1,-}}$, and therefore $\theta\in A^-_{\epsilon,\theta_{i_1,-}}$ by Lemma 6.10(c). Hence $\theta_{i_1,-}$ is also transient and we can again apply 6.13(e). By repeating this construction, we obtain a sequence of sinks $\{\theta_{i_k,-}\}$; since there are only finitely many sinks, eventually we have $\theta_{i_k,-}=\theta_{i_l,-}$ for some $l>k$, which in particular implies $\theta_{i_{k+1},-}\in A^-_{\epsilon,\theta_{i_k,-}}$, which contradicts 6.13(e).

Hence, we conclude that $\theta_{i_k,-}$ is recurrent; by definition we have $\theta_{i_0,-}\in A^-_{\epsilon,\theta_{i_k,-},T'}$ (where T' was defined in the proof of Lemma 6.13), which gives $\theta\in A^-_{\epsilon,\theta_{i_k},T_{\mathcal{T}},-}$, as we needed to show. □

Remark 6.15. *In this language, (A4) states that $A^-_{\theta_{1,-}}=\mathbb{T}$; by Lemma 6.11 and compactness of $\mathbb{T}$ we conclude that if ϵ is sufficiently small, $A^-_{\epsilon,\theta_{1,-}}=\mathbb{T}$, which gives $\mathcal{T}_{\epsilon,1}=\mathbb{T}$. In particular, by Lemma 6.13(c), there can be only one trapping set: for any $i=1,\cdots,n_Z$, either $\mathcal{T}_{\epsilon,i}=\emptyset$ or $\mathcal{T}_{\epsilon,i}=\mathcal{T}_{\epsilon,1}$.*

On the other hand, (A4) implies that $A^\pm_{\epsilon,\theta}=\mathbb{T}$ for any $\theta\in\mathbb{T}$.*

We will henceforth fix $\epsilon>0$ so small that all above results hold true. Observe that Lemma 6.14(b), together with Theorem 6.3 immediately implies that for any standard pair ℓ:

$$\mu_\ell\left(\theta_{\lfloor T\varepsilon^{-1}\rfloor}\notin\bigcup_{i=1,\cdots,n_Z}\mathcal{T}_{\epsilon,i}\right)\le(1-\exp(-c_\#\varepsilon^{-1}))^{\lfloor T/T_{\mathcal{T}}\rfloor};\tag{6.15}$$

in other words: any point on a standard pair will eventually be trapped by some $\mathcal{T}_{\epsilon,i}$. Observe moreover that Theorem 6.4 implies that if $\theta_{i,-}$ is recurrent:

$$F^n_\varepsilon(\mathbb{T}\times\mathcal{T}^+_{\epsilon,i})\subset\mathbb{T}\times\mathcal{T}^-_{\epsilon,i}\text{ for any }n>\lfloor T_{\mathrm{F}}\varepsilon^{-1}\rfloor.\tag{6.16}$$

where $\mathcal{T}^+_{\epsilon,i}=B(\mathcal{T}_{\epsilon,i},\varrho)$, $\mathcal{T}^-_{\epsilon,i}=\{\theta:B(\theta,\varrho)\subset\mathcal{T}_{\epsilon,i}\}$ and ϱ and T_{F} are the constants appearing in the statement of Theorem 6.4.

Corollary 6.16. *If $\theta_{i,-}$ is recurrent, there exists an F_ε-invariant $X_i\subset\mathbb{T}\times\mathcal{T}_{\epsilon,i}$ which attracts every point in $\mathbb{T}\times\mathcal{T}_{\epsilon,i}$.*

Proof. Let us define $X_i^{(0)}=\mathbb{T}\times\operatorname{cl}\mathcal{T}_{\epsilon,i}$; by (6.16) we have

$$F^n_\varepsilon X_i^{(0)}\subset\operatorname{int}X_i^{(0)}\text{ for any }n>\lfloor T\varepsilon^{-1}\rfloor.$$

Let us define $X_i^{(s)}=F_\varepsilon^{s\lfloor T\varepsilon^{-1}\rfloor}X_i^{(0)}$; then $X_i^{(s)}\supset X_i^{(s+1)}$; define $X_i=\bigcap_{s\ge0}X_i^{(s)}\ne\emptyset$. By definition X_i is invariant for $F_\varepsilon^{\lfloor T\varepsilon^{-1}\rfloor}$. In fact, it is invariant by F_ε: let $p\in X_i$, then in particular $F_\varepsilon p\in F_\varepsilon X_i^{(s)}$ for any $s>0$; then by (6.16) we conclude that $F_\varepsilon p\in X_i^{(s)}$ for any $s\ge0$, that is $F_\varepsilon p\in X_i$. □

We will from now on assume r_- so small that $H_k\subset\Theta_{k,-}$ for all $k=1,\cdots,n_Z$.

7. From averaged to true dynamics

In this section we show that the true dynamics behaves similarly to the averaged one with very high probability. To this end we will follow the dynamics in rather long time steps. This strategy will be employed also in the following sections, using possibly even longer time steps. Unfortunately, this requires a somewhat cumbersome notation. To guide the reader through the various future constructions we establish the following conventions:

Notational remark 7.1. *In the following we will introduce constants $T_\sharp$, where $\sharp$ stands for some generic subscript, to designate a* macroscopic *time step i.e., a time step of order* 1 *for the averaged motion. To such times we will associate the corresponding* microscopic *time steps for the map F_ε which we will consistently denote with $N_\sharp = \lfloor T_\sharp \varepsilon^{-1} \rfloor$.*

We will also need to consider $\mathcal{O}(\log \varepsilon^{-1})$ multiples of such macroscopic times: to this end we will introduce various constants denoted with $\mathcal{R}_\sharp$ and we will let $\mathcal{K}_\sharp = \lfloor \mathcal{R}_\sharp \log \varepsilon^{-1} \rfloor$.

In this way the reader will be able to immediately distinguish shorter time steps (e.g. $N_\sharp$) from the (logarithmically) longer ones (e.g. $\mathcal{K}_\sharp N_{\sharp'}$).

7.1. Escape and contraction.

Lemma 7.2 below essentially states that if ℓ is supported on some set $\{\theta \in H_k\}$, the $O(\varepsilon^{-1})$-image of ℓ will escape from $\{\theta \in H_k\}$ with exponentially small probability. Additionally, we have some bounds on the random variable ζ_n, which controls the contraction in the center direction. Recall, from the previous section, that $(x_n(p), \theta_n(p)) = F_\varepsilon^n(p)$ and $\zeta_n(p)$, defined in (4.11), are considered to be random variables when $p \in \mathbb{T}^2$ is distributed on a standard pair ℓ. Given a standard pair ℓ, define $\theta_\ell^* = \mu_\ell(\theta_0)$; given a set $P \subset \mathbb{T}$, we say that *ℓ is located at P* if $\theta_\ell^* \in P$.

Let us fix at this point $T_{\mathrm{S}} > 0$ sufficiently large[29] and recall that, following the convention introduced in the above Notational Remark 7.1, we let $N_{\mathrm{S}} = \lfloor T_{\mathrm{S}} \varepsilon^{-1} \rfloor$.

Lemma 7.2. *If T_{S} is sufficiently large and ε sufficiently small, then for any standard pair ℓ located at H_k for some k,*[30] *we have*

$$\mu_\ell(\theta_{N_{\mathrm{S}}} \in \hat{H}_k, \zeta_{N_{\mathrm{S}}} \le -9T_{\mathrm{S}}/16) \ge 1 - \exp(-c_\# \varepsilon^{-1}).$$

Proof. Fix $TCn > 0$ to be specified later and define the set

$$R = \{p \in \operatorname{supp} \ell : \sup_{t \in [0, T_{\mathrm{S}}]} |\Delta\theta(t, p)| < r_-/8, \sup_{t \in [0, T_{\mathrm{S}}]} |\Delta\zeta(t, p)| < T_{\mathrm{S}}/16\}.$$

We claim that we can choose ε sufficiently small so that for any $p \in R$, we have $\theta_{N_{\mathrm{S}}}(p) \in \hat{H}_k$ and $\zeta_{N_{\mathrm{S}}}(p) \le -9T_{\mathrm{S}}/16$. This would then prove our lemma, since Theorem 6.1 implies that $\mu_\ell(R) \ge 1 - \exp(-c_\# \varepsilon^{-1})$.

To prove our claim, it is convenient to make our set H_k fuzzy; for $\varkappa \in (1/2, 2)$, define $H_{k,\varkappa} = B(\theta_{k,-}, \varkappa r_-)$. First, we assume T_{S} to be so large that, for any k, $\bar{\theta}(T_{\mathrm{S}}; H_k) \subset H_{k,1/2}$. Then, since $\theta_\ell^* \in H_k$, we can assume ε to be small enough to ensure that $\theta_{N_{\mathrm{S}}}(R) \subset H_{k,3/4} = \hat{H}_k$, which proves the first part of our claim.

Additionally, observe that $\theta_n(R) \subset H_{k,5/4}$ for any $0 \le n \le N_{\mathrm{S}}$; by choosing a smaller r_- if necessary we can assume that $\bar{\psi}(\theta) < -5/8$ for any $\theta \in H_{k,5/4}$. Thus, for any $p \in R$, $\bar{\psi}(\theta_n(p)) < -5/8$: hence (6.1) and the definition of R then imply that:

$$\zeta_n(p) \le -\frac{5}{8} n\varepsilon + \frac{T_{\mathrm{S}}}{16},$$

which concludes the proof of our claim. □

[29] The choice for T_{S} depends on a number of assumptions in Lemmata 7.2, 7.4 and 7.5; it is however important to observe that such requirements depend on f and ω only.

[30] Recall that H_k and $\hat{H}_k$ are defined in Subsection 6.3.

Lemma 7.2 implies that as long as a standard pair is located at $\mathbb{H}$, it will stay there with large probability for an exponentially long time.

Corollary 7.3. *Let ℓ be a standard pair located at H_k for some k. For any $l > 0$:*

$$\mu_\ell(\theta_{lN_{\mathrm{S}}} \in \hat{H}_k) \geq (1 - \exp(-c_\# \varepsilon^{-1}))^l.$$

Proof. The proof follows by induction on l: Lemma 7.2 proves the base step $l = 1$. Assume now that the statement holds for $l - 1$. Let $\mathcal{A}'_{N_{\mathrm{S}}} = \boldsymbol{\alpha}_{N_{\mathrm{S}}}(\theta_{N_{\mathrm{S}}} \in \hat{H}_k)$, where $\boldsymbol{\alpha}_{N_{\mathrm{S}}}$ was defined in Remark 5.5; by definition of standard curve, for any $\alpha \in \mathcal{A}'_{N_{\mathrm{S}}}$ and $p, q \in U_{N_{\mathrm{S}}}(\alpha)$ (recall that $U_{N_{\mathrm{S}}}(\alpha) = \boldsymbol{\alpha}^{-1}_{N_{\mathrm{S}}}(\alpha)$), we have $|\theta_{N_{\mathrm{S}}}(q) - \theta_{N_{\mathrm{S}}}(p)| < C_\# \varepsilon$; this in turn implies that $\theta^*_{\ell(\alpha)} \in H_k$. Then, by the inductive assumption and Lemma 7.2,

$$\begin{aligned}\mu_\ell(\theta_{lN_{\mathrm{S}}} \in \hat{H}_k) &\geq \mu_{\mathfrak{L}_{N_{\mathrm{S}}}}(\theta_{(l-1)N_{\mathrm{S}}} \in \hat{H}_k | \mathcal{A}'_{N_{\mathrm{S}}})\mu_\ell(\mathcal{A}'_{N_{\mathrm{S}}}) \\ &\geq (1 - \exp(c_\# \varepsilon^{-1}))^{l-1} \mu_\ell(\theta_{N_{\mathrm{S}}} \in \hat{H}_k) \geq (1 - \exp(c_\# \varepsilon^{-1}))^l. \qquad \square\end{aligned}$$

The above corollary allows to obtain sharper information on the θ variable by means of the following lemma.

Lemma 7.4. *If T_{S} is sufficiently large, there exists $C, \mathcal{R}_D > 0$ so that, provided ε is sufficiently small, for any standard pair ℓ located at H_k and for any $\mathcal{R} \geq \mathcal{R}_D$, letting $\mathcal{K} = \lfloor \mathcal{R} \log \varepsilon^{-1} \rfloor$:*

$$\mu_\ell(\theta^*_{\ell_{\mathcal{K}N_{\mathrm{S}}}(\cdot)} \notin B(\theta_{k,-}, C\sqrt{\varepsilon})) < \frac{1}{3}, \tag{7.1}$$

where, recall, we consider $\ell_{\mathcal{K}N_{\mathrm{S}}}(\cdot)$ to be a random standard pair according to Remark 5.5.

Proof. Define the function $\mathcal{V} : \mathbb{T} \to \mathbb{R}_+$:

$$\mathcal{V}(\theta) = \min\{|\theta - \theta_{k,-}|, r_-\}.$$

We will use $\mathcal{V}$ as a sort of Lyapunov function, namely, we claim that if ℓ is located at H_k, $\mathcal{V}$ satisfies the following geometric drift condition:

$$\mu_\ell(\mathcal{V} \circ \theta_{N_{\mathrm{S}}}) \leq \frac{1}{2}\mu_\ell(\mathcal{V}) + C_{T_{\mathrm{S}}}\sqrt{\varepsilon}, \tag{7.2}$$

where, in the above expression, we regard θ_n as a random variable on ℓ and to simplify the exposition we will abuse notation and write $\mu_\ell(\mathcal{V})$ instead of $\mu_\ell(\mathcal{V} \circ \theta_0)$. Moreover $C_{T_{\mathrm{S}}}$ is a constant which depends on T_{S} only. In fact, first observe that, by Theorem 6.1, since $\|\mathcal{V}\|_\infty \leq 1$:

$$\mu_\ell(\mathcal{V} \circ \theta_{N_{\mathrm{S}}}) = \mu_\ell((\mathcal{V} \cdot \mathbb{1}_{B(\bar{\theta}(T_{\mathrm{S}}; \theta^*_\ell), \varepsilon^{1/2 - \alpha_0/2})}) \circ \theta_{N_{\mathrm{S}}}) + \mathcal{O}(\exp(-C_{T_{\mathrm{S}}} \varepsilon^{-\alpha_0})).$$

Now let us subdivide the interval $B(\bar{\theta}(T_{\mathrm{S}}; \theta^*_\ell), \varepsilon^{1/2 - \alpha_0/2})$ in $\mathcal{O}(\varepsilon^{-1/2 - \alpha_0/2})$ intervals I_j of size $\mathcal{O}(\varepsilon)$, so that we can write

$$\mu_\ell((\mathcal{V} \cdot \mathbb{1}_{B(\bar{\theta}(T_{\mathrm{S}}; \theta^*_\ell), \varepsilon^{1/2 - \alpha_0/2})}) \circ \theta_{N_{\mathrm{S}}}) = \sum_j \mu_\ell((\mathcal{V} \cdot \mathbb{1}_{I_j}) \circ \theta_{N_{\mathrm{S}}}).$$

Using Theorem 6.8 and the fact that $\mathcal{V}$ is Lipschitz yields, on each interval I_j,

$$\begin{aligned}\mu_\ell((\mathcal{V} \cdot \mathbb{1}_{I_j}) \circ \theta_{N_{\mathrm{S}}}) = &\int_{I_j} \frac{e^{-(y - \bar{\theta}(T_{\mathrm{S}}; \theta^*_\ell))^2/(2\varepsilon \boldsymbol{\sigma}^2_{T_{\mathrm{S}}})}}{\boldsymbol{\sigma}_{T_{\mathrm{S}}}\sqrt{2\pi\varepsilon}} \mathcal{V}(y) dy \\ &+ \mathcal{O}\left(\varepsilon^{\alpha_0 - \frac{1}{2}} \int_{I_j} \mathcal{V}(y) dy + \varepsilon^{\frac{3}{2}}\right).\end{aligned}$$

Hence, summing over all intervals and using standard Large Deviations bounds for the Normal Distribution, we obtain

$$\mu_\ell(\mathcal{V}\circ\theta_{N_{\mathrm{S}}}) = \int \frac{e^{-(y-\bar\theta(T_{\mathrm{S}};\theta^*_\ell))^2/(2\varepsilon\boldsymbol{\sigma}^2_{T_{\mathrm{S}}})}}{\boldsymbol{\sigma}_{T_{\mathrm{S}}}\sqrt{2\pi\varepsilon}}\mathcal{V}(y)dy + \mathcal{O}(\varepsilon^{1-\alpha_0/2}) \\ + \mathcal{O}\left(\varepsilon^{\alpha_0-1/2}\int_{\bar\theta(T_{\mathrm{S}};\theta^*_\ell)-\varepsilon^{1/2-\alpha_0/2}}^{\bar\theta(T_{\mathrm{S}};\theta^*_\ell)+\varepsilon^{1/2-\alpha_0/2}}\mathcal{V}(y)dy\right)$$

and, again, since $\mathcal{V}$ is Lipschitz:

$$\mu_\ell(\mathcal{V}\circ\theta_{N_{\mathrm{S}}}) = (1+\mathcal{O}(\varepsilon^{\alpha_0/2}))\mathcal{V}(\bar\theta(T_{\mathrm{S}};\theta^*_\ell)) + (\boldsymbol{\sigma}_{T_{\mathrm{S}}}+1)\mathcal{O}(\sqrt{\varepsilon}). \tag{7.3}$$

Then, let us assume that T_{S} has been chosen sufficiently large that for any $\theta\in H_k$:

$$|\bar\theta(T_{\mathrm{S}};\theta)-\theta_{k,-}| \le \frac{1}{3}|\theta-\theta_{k,-}|,$$

so that in particular we have $\mathcal{V}(\bar\theta(T_{\mathrm{S}};\theta)) < \mathcal{V}(\theta)/3$. Since $\mu_\ell(\mathcal{V}) = \mathcal{V}(\theta^*_\ell) + \mathcal{O}(\varepsilon)$, (7.3) reads:

$$\mu_\ell(\mathcal{V}\circ\theta_{N_{\mathrm{S}}}) < \frac{1+\varepsilon^{\alpha_0/2}}{3}\mu_\ell(\mathcal{V}) + C_{T_{\mathrm{S}}}\sqrt{\varepsilon}$$

which, gives (7.2), provided ε is chosen small enough.

Observe now that by Corollary 7.3, $\mu_\ell(\theta^*_{\ell_{lN_{\mathrm{S}}}(\cdot)}\notin H_k) = \nu_{lN_{\mathrm{S}}}(\theta^*_\ell\notin H_k) < l\exp(-c_\#\varepsilon^{-1})$. Using (7.2) we can then conclude that there exists $\mathcal{R}_{\mathrm{D}}>0$ sufficiently large so that for any $\mathcal{R}\ge\mathcal{R}_{\mathrm{D}}$ and $\mathcal{K}=\lfloor\mathcal{R}\log\varepsilon^{-1}\rfloor$:

$$\begin{aligned}\mu_{\mathfrak{L}_{\mathcal{K}N_{\mathrm{S}}}}(\mathcal{V}) &= \mu_{\mathfrak{L}_{(\mathcal{K}-1)N_{\mathrm{S}}}}(\mathcal{V}\circ\theta_{N_{\mathrm{S}}})\\ &= \mu_{\mathfrak{L}_{(\mathcal{K}-1)N_{\mathrm{S}}}}(\mathcal{V}\circ\theta_{N_{\mathrm{S}}}|\theta^*_\ell\in H_k)\nu_{(\mathcal{K}-1)N_{\mathrm{S}}}(\theta^*_\ell\in H_k) + C_\#\mathcal{K}\exp(-c_\#\varepsilon^{-1})\\ &\le \frac{1}{2}\mu_{\mathfrak{L}_{(\mathcal{K}-1)N_{\mathrm{S}}}}(\mathcal{V}|\theta^*_\ell\in H_k) + C_{T_{\mathrm{S}}}\sqrt{\varepsilon} + C_\#\mathcal{K}\exp(-c_\#\varepsilon^{-1})\\ &\le \frac{1}{2}\mu_{\mathfrak{L}_{(\mathcal{K}-1)N_{\mathrm{S}}}}(\mathcal{V}) + C_{T_{\mathrm{S}}}\sqrt{\varepsilon} + C_\#\mathcal{K}\exp(-c_\#\varepsilon^{-1}).\end{aligned}$$

Iterating the above estimates $\mathcal{K}$ times yields

$$\mu_{\mathfrak{L}_{\mathcal{K}N_{\mathrm{S}}}}(\mathcal{V}) \le C_{T_{\mathrm{S}}}\sqrt{\varepsilon}(1+\mu_\ell(\mathcal{V})) + C_\#\mathcal{K}^2\exp(-c_\#\varepsilon^{-1}) \le C_{T_{\mathrm{S}}}\sqrt{\varepsilon}.$$

Markov Inequality thus implies (7.1) e.g. choosing $C=3C_{T_{\mathrm{S}}}$. □

7.2. **Attractors: return to $\mathbb{H}$.** We now proceed to describe the dynamics outside $\mathbb{H}$ (for its definition, see (6.14)); indeed we will not need very refined results in this region; essentially we will only prove that the dynamics comes to $\mathbb{H}$ with very large probability in time $\mathcal{O}(\log\varepsilon^{-1})$.

Lemma 7.5. *If T_{S} is sufficiently large, there exists $\beta>0$ and $\mathcal{R}_{\mathrm{A}}>1$ such that, provided ε is sufficiently small, for any standard pair ℓ we have:*

$$\mu_\ell(\theta_{\mathcal{K}_{\mathrm{A}}N_{\mathrm{S}}}\notin\hat{\mathbb{H}}) < \varepsilon^\beta. \tag{7.4}$$

where, according to Notational Remark 7.1, $\mathcal{K}_{\mathrm{A}} = \lfloor\mathcal{R}_{\mathrm{A}}\log\varepsilon^{-1}\rfloor$.

Proof. Fix $\mathcal{R}_{\mathrm{A}}>1$ sufficiently large to be specified later. We will prove the lemma in two steps; first let us show the following auxiliary result:

Sub-lemma 7.6. *There exists $T_0>0$, and $c=c(\mathcal{R}_{\mathrm{A}})$ so that if ε is sufficiently small, for any $\lfloor T_0\varepsilon^{-1}\rfloor = N_0\le N\le\mathcal{K}_{\mathrm{A}}N_{\mathrm{S}}$ and standard pair ℓ not located at $\mathbb{S}$:*

$$\mu_\ell(\theta_N\notin\hat{\mathbb{H}}) < \exp(-c\varepsilon^{-1}).$$

Proof. Our stipulations on $\bar\omega$ guarantee that, if $\theta \notin \mathbb{S}$, there exists $T_0 > 0$ such that if $T > T_0$, then $\bar\theta(T;\theta) \in \hat{\mathbb{H}}$. According to Notational Remark 7.1, let $N_0 = \lfloor T_0\varepsilon^{-1}\rfloor$ and write $N = lN_{\mathrm{S}}+M$ where $N_0 \le M < N_0+N_{\mathrm{S}}$. By Large Deviations arguments analogous to the ones used in the proof of Lemma 7.2 we conclude that

$$\mu_\ell(\theta_M \notin \hat{\mathbb{H}}) < \exp(-\bar c\varepsilon^{-1}).$$

Let $\mathfrak{L}_M$ be a standard M-pushforward of ℓ; observe that if $\theta_M(p) \in \hat{\mathbb{H}}$, necessarily $\ell_M(p)$ is located at $\mathbb{H}$ (recall that $\ell_M(p)$ was defined in Remark 5.5); we thus conclude that $\mu_\ell(\theta^*_{\ell_M} \notin \mathbb{H}) < \exp(-\bar c\varepsilon^{-1})$. Hence, since

$$\begin{aligned}\mu_\ell(\theta_N \notin \hat{\mathbb{H}}) &\le \mu_\ell(\theta_N \notin \hat{\mathbb{H}}|\theta^*_{\ell_M} \in \mathbb{H}) + \exp(-\bar c\varepsilon^{-1})\\ &\le \mu_{\mathfrak{L}_M}(\theta_{N-M} \notin \hat{\mathbb{H}}|\theta^*_{\ell} \in \mathbb{H}) + \exp(-\bar c\varepsilon^{-1});\end{aligned}$$

our result then follows by applying Corollary 7.3. □

Observe that Sub-Lemma 7.6 proves Lemma 7.5 in the particular case $\theta^*_\ell \notin \mathbb{S}$. If this is not the case, we have ℓ is located at $\mathbb{S} = \bigcup_k S_k$: for ease of exposition, let k be fixed so that $\theta^*_\ell \in S_k$ and let us drop k from our notations; that is, let $\theta_+ = \theta_{k,+}$, $S = S_k$. In order to prove our lemma we claim that it suffices to show that for some $\beta' > 0$:

$$\mu_\ell(\theta_N \in \hat S \text{ for all } 0 \le N \le \mathcal{K}_{\mathrm{A}}N_{\mathrm{S}} - N_0) < \varepsilon^{\beta'} \tag{7.5}$$

where $\hat S$ is an $\mathcal{O}(\varepsilon^{1/4})$-neighborhood of S. In fact, by (7.5), with probability $1-\varepsilon^{\beta'}$, there exists some $0 \le N \le \mathcal{K}_{\mathrm{A}}N_{\mathrm{S}} - N_0$ so that $\theta^*_{\ell_N(p)} \notin \mathbb{S}$; applying Sub-Lemma 7.6 to such standard pair then guarantees that the $(\mathcal{K}_{\mathrm{A}}N_{\mathrm{S}} - N)$-iterate of ℓ_N will be supported in $\hat{\mathbb{H}}$ with probability $1 - \exp(-\bar c\varepsilon^{-1})$, which in turn implies (7.4) and concludes our proof.

We are thus left to prove (7.5): fix $c_S > 0$ to be specified later and define the function $\mathcal{V} : \mathbb{T} \to \mathbb{R}$:

$$\mathcal{V}(\theta) = \begin{cases} 0 & \text{if } \theta \notin \hat S \\ \frac{1}{c_S\sqrt{\varepsilon}} & \text{if } |\theta - \theta_+| \le c_S\sqrt{\varepsilon} \\ \frac{1}{|\theta-\theta_+|} & \text{otherwise.}\end{cases}$$

We claim there exists $\vartheta \in (0, 1)$ so that for any $0 < k < \mathcal{K}_{\mathrm{A}}$:

$$\mu_\ell(\mathcal{V} \circ \theta_{kN_{\mathrm{S}}}) \le C_\#\vartheta^k\mu_\ell(\mathcal{V}) + C_\# \exp(-c_\#\varepsilon^{-1}), \tag{7.6}$$

Then, if $\bar{\mathcal{R}} \sim \log\varepsilon^{-1}$ is so that $\vartheta^{\bar{\mathcal{R}}} = \mathcal{O}(\varepsilon^{\beta'})$, by (7.6) we gather $\mu_\ell(\mathcal{V} \circ \theta_{\bar{\mathcal{R}}N_{\mathrm{S}}}) \le C_\#\varepsilon^{\beta'}$. Markov Inequality then implies that

$$\mu_\ell(\theta_{\bar{\mathcal{R}}N_{\mathrm{S}}} \in \hat S) = \mu_\ell\left(\mathcal{V} \circ \theta_{\bar{\mathcal{R}}N_{\mathrm{S}}} > 1/(2r_+)\right) < C_\#\varepsilon^{\beta'},$$

provided that ε is sufficiently small, which in turn implies (7.5), choosing $\mathcal{R}_{\mathrm{A}}$ sufficiently large (e.g., $\mathcal{R}_{\mathrm{A}} \ge 2\bar{\mathcal{R}}$ is certainly enough).

Thus, to conclude our proof, it suffices to prove (7.6). We have three cases:

(a) $\theta^*_\ell \notin S$
(b) $\theta^*_\ell \in S$, $|\theta^*_\ell - \theta_+| \ge c_S\sqrt{\varepsilon}$
(c) $|\theta^*_\ell - \theta_+| < c_S\sqrt{\varepsilon}$.

Let us first consider case (a): in this case, for any $N_0 < N < \mathcal{K}_{\mathrm{A}}N_{\mathrm{S}}$, we claim that:

$$\mu_\ell(\mathcal{V} \circ \theta_N) \le \exp(-c_\#\varepsilon^{-1}). \tag{7.7a}$$

The above estimates immediately follows by Sub-Lemma 7.6 if $\theta^*_\ell \notin \mathbb{S}$, and by a similar large deviations argument otherwise.[31]

[31] In fact $\hat S$ is a repelling set for the averaged dynamics, hence if $\theta^*_\ell \in \mathbb{S} \setminus S$, the averaged dynamics will certainly keep θ away from $\hat S$.

Let us now consider case (b): by definition of W_+ (see (6.13)), we know that if $\theta_0 \in W_+$, $|\bar\omega(\theta_0)| \geq \bar\omega'(\theta_+)|\theta_0 - \theta_+|/2$. We assume T_{S} so large that for any $\theta_0 \in S$, we have either $\bar\theta(T_{\mathrm{S}};\theta_0) \notin \hat S$ or $|\bar\theta(T_{\mathrm{S}};\theta_0) - \theta_+| \geq 2|\theta_0 - \theta_+|$. Hence, we have

$$\mu_\ell\left(|\theta_{N_{\mathrm{S}}} - \theta_+| \leq \frac{3}{2}|\theta^*_\ell - \theta_+|\right) \leq \mu_\ell\left(|\theta_{N_{\mathrm{S}}} - \bar\theta(T_{\mathrm{S}};\theta^*_\ell)| \geq \frac{1}{2}|\theta^*_\ell - \theta_+|\right).$$

Assuming c_S sufficiently large, we can apply Theorem 6.1 and gather that

$$\mu_\ell\left(|\theta_{N_{\mathrm{S}}} - \bar\theta(T_{\mathrm{S}};\theta^*_\ell)| \geq \frac{1}{2}|\theta^*_\ell - \theta_+|\right) \leq \exp(-C_{T_{\mathrm{S}}}|\theta^*_\ell - \theta_+|^2\varepsilon^{-1}).$$

Consequently:

$$\mu_\ell(\mathcal{V}\circ\theta_{N_{\mathrm{S}}}) \leq \frac{2}{3}\mu_\ell(\mathcal{V}) + \frac{1}{c_S\sqrt{\varepsilon}}\exp(-C_{T_{\mathrm{S}}}|\theta^*_\ell - \theta_+|^2\varepsilon^{-1}).$$

Choosing c_S to be so large[32] that

$$\frac{1}{c_S\sqrt{\varepsilon}}\exp(-C_{T_{\mathrm{S}}}|\theta^*_\ell - \theta_+|^2\varepsilon^{-1}) \leq \frac{1}{12|\theta^*_\ell - \theta_+|},$$

we obtain

$$\mu_\ell(\mathcal{V}\circ\theta_{N_{\mathrm{S}}}) \leq \frac{5}{6}\mu_\ell(\mathcal{V}). \tag{7.7b}$$

Finally, we need to consider case (c): first of all by definition of $\mathcal{V}$ we can immediately conclude that for any $n \geq 0$:

$$\mu_\ell(\mathcal{V}\circ\theta_n) \leq \frac{7}{6}\mu_\ell(\mathcal{V}). \tag{7.7c}$$

Moreover, using once again Theorem 6.8 and the lower bound in (6.11) we can choose T_1 to be so large that, for any $\kappa \in \mathbb{R}$, $\mu_\ell(\Delta\theta(T_1;\cdot)\varepsilon^{-1/2} \in B(\kappa, 2c_S)) < 1/3$. We thus conclude that for any standard pair ℓ so that $|\theta^*_\ell - \theta_+| < c_S\sqrt{\varepsilon}$:

$$\mu_\ell(\mathcal{V}\circ\theta_{N_1}) \leq \frac{1}{3}\cdot\frac{1}{c_S\sqrt{\varepsilon}} + \frac{1}{2c_S\sqrt{\varepsilon}} \leq \frac{5}{6}\mu_\ell(\mathcal{V}). \tag{7.7d}$$

Observe that by possibly increasing T_1, we can guarantee $N_1 = pN_{\mathrm{S}}$ for some $p \in \mathbb{N}$. Collecting bounds (7.7) we can therefore conclude that, for any $k \geq 0$ and for any sequence $\mathfrak{L}_n$ of pushforwards of ℓ:

$$\begin{aligned}\mu_\ell(\mathcal{V}\circ\theta_{kpN_{\mathrm{S}}}) &= \mu_{\mathfrak{L}_{(k-1)pN_{\mathrm{S}}}}(\mathcal{V}\circ\theta_{pN_{\mathrm{S}}}) \leq \vartheta_*\mu_{\mathfrak{L}_{(k-1)pN_{\mathrm{S}}}}(\mathcal{V}) + \exp(-c_\#\varepsilon^{-1})\\ &\leq \vartheta_*^k\mu_\ell(\mathcal{V}) + C_\#\exp(-c_\#\varepsilon^{-1}),\end{aligned}$$

for some $\vartheta_* \in (0,1)$ (e.g., $\vartheta_* = (5/6)(7/6) = 35/36$ works). The above inequality immediately implies (7.6), choosing $\vartheta = \vartheta_*^{1/p}$. □

The above results show quantitatively that the dynamics tends to concentrate around the sinks of the averaged dynamics, where most of the center vectors are contracted at an exponential rate. This fact will be the crucial ingredient in our arguments.

[32] Observe that the choice of c_S depends on $C_{T_{\mathrm{S}}}$ and thus on T_{S}.

8. Coupling: basic facts and definitions

We are now ready to start the discussion of statistical properties of the map F_ε. As anticipated, we will classify its SRB measures (in the sense of Remark 2.6) and study their statistical properties using the framework of standard pairs.

The main advantage of using standard pairs is that we are in a sense able to separate the deterministic behavior (at the level of standard pairs) from the stochastic behavior (regarding standard pairs as "atoms").

Let us start by recalling the main ideas: the crucial observation (due to Dolgopyat, see e.g. [14, 15, 16]) is that SRB measures can be written as weak limits of measures that admit a standard decomposition (see also Lemma 9.8).

Then, given any two such measures let $\mathfrak{L}^0$ and $\mathfrak{L}^1$ be the associated standard families. If we can prove that the measures induced by $F_\varepsilon^n\mathfrak{L}^0$ and $F_\varepsilon^n\mathfrak{L}^1$ are asymptotically equal; then this would imply uniqueness of the SRB measure for the system. If, additionally, we can control the rate at which $F_\varepsilon^n\mathfrak{L}^0$ and $F_\varepsilon^n\mathfrak{L}^1$ are approaching, then we are able to retrieve precise information about the rate of mixing of sufficiently smooth observables.

This project can be carried out provided that there exists only one trapping set for the dynamics (that is guaranteed if (A4) holds). Otherwise, if (A5) holds, it is still possible to prove uniqueness of SRB measure supported on each set $\{\theta \in \mathcal{T}_{\epsilon,i}\}$, and obtain information about the rate of mixing of sufficiently smooth observables supported in each of these sets.

The strategy which is most commonly employed in order to study the above mentioned asymptotic equality, and estimate the speed of mixing, is the *coupling technique*.[33]

8.1. **Basic coupling definitions.** Let us start by recalling some useful definitions: a *coupling* of two probability measures μ_0 and μ_1 (on the measurable space $\mathbb{T}^2$) is given by a probability measure $\boldsymbol{\mu}$ on the product space $\mathbb{T}^2 \times \mathbb{T}^2$ whose marginals on the first and second factor coincide with μ_0 and μ_1 respectively. We denote with $\Gamma(\mu_0, \mu_1)$ the set of couplings of the two probability measures. The *Wasserstein distance* of two probability measures μ_0, μ_1 is defined as

$$\mathrm{d_W}(\mu_0,\mu_1) = \inf_{\boldsymbol{\mu}\in\Gamma(\mu_0,\mu_1)} \mathbb{E}_{\boldsymbol{\mu}}(\mathrm{dist_V}). \tag{8.1}$$

Note that the definition of $\mathrm{d_W}$ depends on the choice of the distance. In this paper, we find convenient to employ the *vertical distance* $\mathrm{dist_V}$ given by:

$$\mathrm{dist_V}((x,\theta),(x',\theta')) = \begin{cases} 1 & \text{if } x \neq x' \\ |\theta-\theta'| & \text{otherwise.} \end{cases} \tag{8.2}$$

Observe that $\mathrm{dist_V}$ controls the standard Euclidean distance (which we denote by dist), i.e. $\mathrm{dist}(p,p') \leq \sqrt{2}\,\mathrm{dist_V}(p,p')$ for any p, $p' \in \mathbb{T}^2$.

When two measures admit a standard decomposition, it is convenient to describe their couplings in terms of standard families. Let us start by considering standard pairs: by a (c_1', c_2')-*standard couple* $\underline{\ell} = (\underline{\ell}^{\,0}, \underline{\ell}^{\,1})$ we mean a couple of (c_1', c_2')-standard pairs $\underline{\ell}^{\,0}$ and $\underline{\ell}^{\,1}$ and a measure $\boldsymbol{\mu}$ on $\mathbb{T}^2$ such that the marginals are the measures are $\mu_{\underline{\ell}^0}$ and $\mu_{\underline{\ell}^1}$ respectively.[34] Let now $\mathfrak{L}^0$ and $\mathfrak{L}^1$ be two

[33] Coupling has been long used in abstract Ergodic Theory under the name of *joining*, but it has been re-introduced in the the study of the statistical properties of smooth systems (smooth Ergodic Theory) by Lai-Sang Young [46], borrowing it from the theory of Markov chains. The version we are going to present here has been developed by Dmitry Dolgopyat in the standard pair framework.

[34] We do not include explicitly $\boldsymbol{\mu}$ in the notation to make it more readable and as it does not create confusion.

(pre)standard families; a (c_1', c_2')-*standard coupling* of $\mathfrak{L}^0$ and $\mathfrak{L}^1$ is a random element $\underline{\mathfrak{L}} = ((\mathcal{A}, \nu), \underline{\ell})$ where $\underline{\ell} : \mathcal{A} \to L_{c_1',c_2'} \times L_{c_1',c_2'}$ is a random couple $\underline{\ell} = (\underline{\ell}^0, \underline{\ell}^1)$ of (c_1', c_2')-standard pairs so that $\underline{\mathfrak{L}}^i = ((\mathcal{A}, \nu), \underline{\ell}^i) \sim \mathfrak{L}^i$.

Given two (pre)standard families $\mathfrak{L}^0$ and $\mathfrak{L}^1$, there is no canonical choice of a standard coupling. A simple, but important example is given by the *independent* coupling: we let $(\mathcal{A}, \nu) = \mathsf{X}_{i=0,1}(\mathcal{A}^i, \nu^i)$ and we define $\underline{\ell}((\alpha_0, \alpha_1)) = (\ell^0(\alpha_0), \ell^1(\alpha_1)) = (\underline{\ell}^0(\alpha_0, \alpha_1), \underline{\ell}^1(\alpha_0, \alpha_1))$ equipped with the product measure. Observe that, in this case, the families $\underline{\mathfrak{L}}^0$ and $\underline{\mathfrak{L}}^1$ are independent random variables. As for standard families, we will declare two coupling equivalent if their marginal measures $\mu_{\underline{\mathfrak{L}}^i}$ are the same, and we will designate the equivalence classes by $[\underline{\mathfrak{L}}]$.

Let $\underline{\mathfrak{L}} = ((\mathcal{A}, \nu), \underline{\ell})$; given $\mathcal{A}' \subset \mathcal{A}$ we define the *subcoupling conditioned on* $\mathcal{A}'$ to be $\underline{\mathfrak{L}}|\mathcal{A}' = ((\mathcal{A}', \nu'), \underline{\ell}|\mathcal{A}')$, where, once again, $\nu'(E) = \nu(E|\mathcal{A}')$.

We say that $\underline{\ell} = (\underline{\ell}^0, \underline{\ell}^1)$ is a *matched couple* (resp. Δ-*matched couple*) if ℓ^0 and ℓ^1 are stacked (resp. Δ-stacked), have equal densities (see Section 5.1.1 for the definition of "stacked") and are coupled along the vertical direction, that is: $\boldsymbol{\mu}(g) = \int g(\mathbb{G}^0(x), \mathbb{G}^1(x))\rho(x)dx$ where ρ is their common density. We will call this the *canonical coupling* for matched pairs. Note that

$$\mathrm{d_W}(\mu_{\underline{\ell}^0}, \mu_{\underline{\ell}^1}) \leq \Delta. \tag{8.3}$$

Recall that, given the (pre)standard families $\{\mathfrak{L}^i\}$ it is defined for all n the pushforward $\left[F_\varepsilon^n \mathfrak{L}^i\right] = \left[\mathfrak{L}_n^i\right]$. Then, given any standard coupling $\underline{\mathfrak{L}}$ of $\mathfrak{L}^0, \mathfrak{L}^1$, we define $[F_\varepsilon^n \underline{\mathfrak{L}}]$ as the equivalence class of the product coupling of $\left[\mathfrak{L}_n^i\right]$. A sequence of (pre)standard couplings $(\underline{\mathfrak{L}}_n)_n$ is said to be a *pushforward* the (pre)standard coupling $\underline{\mathfrak{L}}$ if $\underline{\mathfrak{L}}_n \in [F_\varepsilon^n \underline{\mathfrak{L}}]$. Note that such a definition is much less stringent than the notion of pushforward for a standard family, it is then not surprising that we will need a more stringent definition for standard couplings as well.

If $\underline{\mathfrak{L}}_n^0$ is a pushforward of a standard family $\underline{\mathfrak{L}}^0$ and, for each $\alpha \in \mathcal{A}_n$, $\underline{\ell}_n(\alpha)$ is a Δ-matched pair, for some Δ, then we say that we have a *matched pushforward*.[35]

Remark 8.1. *Note that, if we have a matched pushforward $\underline{\mathfrak{L}}_n$ of $\underline{\ell}$, then Remark 5.5 applies to the families $\underline{\ell}_n^0$ and, by the matching property, to $\underline{\ell}_n^1$ as well. It makes thus sense to write $\underline{\ell}_n(p)$.*

Notational remark 8.2. *In the sequel, to simplify our notation, we adopt the convention that the symbols $\mathfrak{L}, \mathcal{A}, \nu, \ell$ will carry subscripts and superscripts in the natural consistent way.*

8.2. **The holonomy map.** Loosely speaking, in the hyperbolic setting, the coupling technique is based on the dynamical idea of "linking mass of standard pairs to nearby ones along stable manifold". In our setting, since we lack a stable direction, we will "link mass" along curves that approximate the center direction for at least $N_\mathrm{S} = \mathcal{O}(\varepsilon^{-1})$ iterates (as it turns out, we will use local N_S-step center manifolds $\mathcal{W}_{N_\mathrm{S}}^\mathrm{c}$, that have been defined in Section 4). In the next section we will show that (A2) guarantees average contraction along such curves and, as a consequence, modulo large deviations, they can indeed serve the purpose of stable manifolds. The crucial issue, however, is that the regularity of the holonomy map along the curves $\mathcal{W}_{N_\mathrm{S}}^\mathrm{c}$ deteriorates very quickly compared to the average contraction rate on $\mathcal{W}_{N_\mathrm{S}}^\mathrm{c}$. This fact is the main obstacle to set up an efficient coupling strategy and, in fact, will force us to use very short center manifolds (see Remark 8.4).

To make precise the above issue let us start by properly defining the *holonomy map* along center manifolds: for some small $\Delta > 0$, let $\mathbb{G}^0 = (x, G^0(x))$ and $\mathbb{G}^1 = (x, G^1(x))$ be two Δ-stacked standard curves above $[a, b]$ (recall the appropriate

[35] Note that, with the above definition, $\underline{\mathfrak{L}}^1$ will not be, in general, a standard family. Yet, it will be n-prestandard and that is all is needed in the following.

definitions given in Section 5.1.1). Then, for $s \in [0,1]$ define the interpolating curves $\mathbb{G}^s$ by convex combination, i.e.

$$G^s(x) = (1-s)G^0(x) + sG^1(x) \qquad\qquad \mathbb{G}^s = (x, G^s(x)).$$

Let $h_n(s;x)$ be the unique solution of the following non-autonomous ODE:

$$\frac{d}{ds}h_n(s;x) = \frac{G^1(h_n(s;x)) - G^0(h_n(s;x))}{1 - s_n(\mathbb{G}^s(h_n(s;x)))G^{s\prime}(h_n(s;x))} s_n(\mathbb{G}^s(h_n(s;x)))$$

with initial condition $h_n(0;x) = x$. By invariance of the center cone and by definition of standard curve it is immediate to show that the above problem is well-defined[36] provided that $x \in [a', b']$ where $a' = a + 2\Delta\gamma^{\rm c}$ and $b' = b - 2\Delta\gamma^{\rm c}$ (recall the definition of $\gamma^{\rm c}$ given in (4.5)). Likewise, it is not difficult to check that $\pi F_\varepsilon^n(\mathbb{G}^0(x)) = \pi F_\varepsilon^n(\mathbb{G}^s(h_n(s;x)))$ for all $x \in [a', b']$ and all $s \in [0,1]$, where π is the projection on the x-coordinate. Let us define the n*-step holonomy map* $\mathcal{H}_n : [a', b'] \to [a, b]$ as $\mathcal{H}_n(x) = h_n(1;x)$; observe that:

$$\pi F_\varepsilon^n(\mathbb{G}^0(x)) = \pi F_\varepsilon^n(\mathbb{G}^1(\mathcal{H}_n(x))). \tag{8.4}$$

The map $\mathcal{H}_n$ is an orientation preserving diffeomorphism; geometrically, if $p^0 \in \operatorname{supp} \ell^0$ and $p^1 = \mathbb{G}^1(\mathcal{H}_n(\pi p^0)) \in \operatorname{supp} \ell^1$, then p^0 and p^1 are joined by a local n-step center manifold $\mathcal{W}_n^{\rm c}$

We are now ready to state the relevant properties of the holonomy map $\mathcal{H}_n$.

Proposition 8.3. *Let $\mathbb{G}^0$ and $\mathbb{G}^1$ be two $\Delta\varepsilon$-stacked standard curves above $[a,b]$. For any $T > 0$ let $N = \lfloor T\varepsilon^{-1} \rfloor$ (according to Notational Remark 7.1), and $\mathcal{H}_N$ be the N-step holonomy map between the two curves;*

(a) *for any $p^0 \in \mathbb{G}^0([a + 2\Delta\gamma^{\rm c}\varepsilon, b - 2\Delta\gamma^{\rm c}\varepsilon])$, let $p^1 = \mathbb{G}^1(\mathcal{H}_N(\pi p^0))$. Then, recalling the notation $p_n^i = F_\varepsilon^n p^i$:*

$$\operatorname{dist}_{\rm V}(p_N^0, p_N^1) \le (1 + C_T\varepsilon + 2T\varrho)\exp(\zeta_N(p^0))\Delta\varepsilon. \tag{8.5}$$

(b) *let $u^0(x) = G^{0\prime}(x)\varepsilon^{-1}$ and $u^1(x) = G^{1\prime}(\mathcal{H}_N(x))\varepsilon^{-1}$ and define $u_n^i(x) = \Xi_{p^i(x)}^{(n)} u^i(x)$, where $\Xi_p^{(n)} = \Xi_{p_{n-1}} \circ \cdots \circ \Xi_{p_0}$ and Ξ_p was defined in Section 4; then $u_n^i(x)$ are the rescaled slopes of the image curve at the point $p_n^i(x) = F_\varepsilon^n \circ \mathbb{G}^i(x)$ and they satisfy, for each $0 \le n \le N$:*

$$\|u_n^1 - u_n^0\|_\infty \le C_\# \exp(\Psi T)\Delta[\varepsilon + (2\lambda^{-1})^n]; \tag{8.6}$$

where recall that $\lambda > 2$ is the minimal expansion of f_θ (see Section 2) and $\Psi = \|\partial_\theta\omega\| + \gamma^{\rm u}\|\partial_x\omega\|$ (see (4.8));

(c) *there exists $\mathcal{D}_T \sim C_\# T\exp(\Psi T)$ such that:*

$$\|\log \mathcal{H}_N'\| \le \mathcal{D}_T\Delta. \tag{8.7}$$

Proof. Since the standard curves $\mathbb{G}^i$'s are $\Delta\varepsilon$-stacked and p^0 and p^1 are joined by a center manifold (which we will denote by $\mathcal{W}_N^{\rm c}(p^0)$) whose tangent belongs to the center cone, we gather that $\operatorname{dist}(p^0, p^1) \le C_\#\Delta\varepsilon$; moreover $\pi p_N^0 = \pi p_N^1$ as already observed earlier. Moreover, let $h \le (1 + C_\#\varepsilon)\Delta\varepsilon$ be the distance of the projection of p^0 and p^1 on the θ coordinate; by (4.6) and definition of the center cone (4.1), we obtain that:

$$\operatorname{dist}(p_n^0, p_n^1) \le (1 + \gamma^{\rm c}) \sup_{q \in \mathcal{W}_N(p^0, p^1)} \frac{\mu_N(q)}{\mu_{N-n}(F_\varepsilon^n q)} \cdot h$$

from the above and (4.8) we conclude that for any $0 \le n \le N$

$$\operatorname{dist}(p_n^0, p_n^1) \le C_\# \Psi n\varepsilon\Delta\varepsilon. \tag{8.8}$$

[36] The reader can easily check that the differential equation admits a unique solution; recall (4.6) for the definition of s_n.

If $n = N$, since $\pi p_N^0 = \pi p_N^1$, we have the better estimate

$$\operatorname{dist}(p_N^0, p_N^1) = \operatorname{dist}_{\mathrm{V}}(p_N^0, p_N^1) \le \sup_{q \in \mathcal{W}_N(p^0,p^1)} \mu_N(q) \cdot h$$

which using Lemma 4.2 immediately implies (8.5) and proves item (a). On the other hand, since $0 \le n \le N$, (4.4) and (8.8) yield, for small enough ε,

$$\begin{aligned}|u_n^1(x) - u_n^0(x)| &= |\Xi_{p_{n-1}^1(x)}(u_{n-1}^1(x)) - \Xi_{p_{n-1}^0(x)}(u_{n-1}^0(x))| \\ &\le C_\# \exp(\Psi T)\Delta\varepsilon + 2\lambda^{-1}|u_{n-1}^1(x) - u_{n-1}^0(x)| \le \\ &\le \frac{C_\# \exp(\Psi T)\Delta\varepsilon}{1 - 2\lambda^{-1}} + (2\lambda^{-1})^n |u^1(x) - u^0(x)|;\end{aligned}$$

by the definition of stacked curves we obtain $|u^1(x) - u^0(x)| \le C_\#\Delta$, which implies (8.6) and proves item (b).

Let us now prove item (c): differentiating (8.4) yields:

$$\underbrace{d\pi\, dF_\varepsilon^N(\mathbb{G}^0(x))(1, G^{0\prime}(x))}_{x\text{-expansion along } \mathbb{G}^0} = \underbrace{d\pi\, dF_\varepsilon^N(\mathbb{G}^1(\mathcal{H}_N(x)))(1, G^{1\prime}(\mathcal{H}_N(x)))}_{x\text{-expansion along } \mathbb{G}^1} \mathcal{H}_N'(x),$$

which we can rewrite, using (4.3) and letting $\Gamma_N = \prod_{k=0}^{N-1} \partial_x f \circ F_\varepsilon^k$, as

$$\mathcal{H}_N'(x) = \frac{\Gamma_N(p^0(x))}{\Gamma_N(p^1(x))} \prod_{n=0}^{N-1} \frac{1 + \varepsilon \frac{\partial_\theta f}{\partial_x f}(p_n^0(x)) u_n^0(x)}{1 + \varepsilon \frac{\partial_\theta f}{\partial_x f}(p_n^1(x)) u_n^1(x)}. \tag{8.9}$$

Then, using (8.8) we gather that:

$$\left| \log \frac{\Gamma_N(p^0(x))}{\Gamma_N(p^1(x))} \right| \le C_\# T \exp(\Psi T)\Delta, \tag{8.10}$$

and using item (b) we can conclude:

$$\left| \log \prod_{n=0}^{N-1} \frac{1 + \varepsilon \frac{\partial_\theta f}{\partial_x f}(p_n^0(x)) u_n^0(x)}{1 + \varepsilon \frac{\partial_\theta f}{\partial_x f}(p_n^1(x)) u_n^1(x)} \right| \le C_\#(T+1)\exp(\Psi T)\Delta\varepsilon.$$

The above estimates imply (8.7) and consequently conclude the proof of our lemma. □

Remark 8.4. *Item (c) in the above lemma implies in particular that the N-step holonomy map $\mathcal{H}_N$ has good regularity properties for $T = \mathcal{O}(1)$ provided that $\Delta = \mathcal{O}(1)$, i.e. if the stacked pairs are at a distance $\mathcal{O}(\varepsilon)$. A Local Central Limit Theorem is thus crucial for the effectiveness of the coupling procedure, since it provides information about the distribution of standard pairs at the $\mathcal{O}(\varepsilon)$-scale.*

9. The coupling argument

9.1. **An informal exposition of the global strategy.** For simplicity let us first assume that $n_{\mathrm{Z}} = 1$ so that $\{\theta = \theta_{1,-}\}$ is the only attractor for the averaged dynamics. Since we expect the real dynamics to be well approximated by a $\sqrt{\varepsilon}$-diffusion around the averaged dynamics, we will be able to conclude (see the Bootstrap Lemma 9.6) that, if we let any two standard families evolve for a sufficiently long time (which turns out to be $\mathcal{O}(\varepsilon^{-1}\log\varepsilon^{-1})$), then a substantial portion of their mass will be carried by standard pairs which are supported in a $\mathcal{O}(\sqrt{\varepsilon})$ neighborhood of $\theta_{1,-}$. Using a Local Central Limit Theorem (see Theorem 6.8) we can control effectively the distribution of such standard pairs with a $\mathcal{O}(\varepsilon)$-resolution. Once two standard pairs are stacked at a distance ε, since we are close to a sink and by (A2), the averaged system will make them (slowly) approach to each other (see Lemma 9.1). Once they are sufficiently close (e.g. $\mathcal{O}(\varepsilon^{1+\tau})$ for some $\tau > 0$)

we can show that the real dynamics follows the average one almost all the time (a part from rare large deviations) and that the distance between the standard pairs keeps contracting forever with positive probability. Thus we can couple (see the Coupling procedure, Lemma 9.3) almost all their mass forever. We then conclude by iteratively applying the same argument to the mass in the leftover pieces (see Lemma 9.7).

If $n_{\mathrm{Z}} > 1$ there are two possibilities: if Assumption (A4) holds, then our Large Deviations results (see Theorem 6.3) allow us to prove that any standard pair will have some positive (although exponentially small in ε^{-1}) probability of being close to $\theta_{1,-}$ after time $\mathcal{O}(\varepsilon^{-1}\log\varepsilon^{-1})$; we can then conclude the proof by applying the argument above to this tiny amount of mass at each step.

Otherwise, if (A4) does not hold but (A5) holds, Lemma 6.14(b) guarantees the existence of at least one and at most n_{Z} recurrent sinks. Then, once again using Large Deviations arguments, any standard pair that is supported on $\{\theta \in \mathcal{T}_{\epsilon,i}\}$ will have positive (although exponentially small in ε^{-1}) probability of being close to $\theta_{i,-}$ after time $\mathcal{O}(\varepsilon^{-1}\log\varepsilon^{-1})$; moreover (6.16) guarantees that any pushforward will also enjoy the same property. As before, we can conclude by iteratively applying the same argument to this tiny amount of mass at each step. Of course, this procedure does not necessarily yield a unique invariant measure, but rather as many distinct invariant measures as the number of distinct trapping sets.

9.2. **The basic Coupling Step.** We now describe the core of our coupling argument, i.e., we describe how to actually couple two standard pairs (or more precisely, the processes they generate) for $\mathcal{O}(\varepsilon^{-1})$ iterations.

Recall that at the beginning of the previous section we fixed the constant $T_{\mathrm{S}} > 0$ (and correspondingly N_{S}) sufficiently large; fix ϱ so that $T_{\mathrm{S}}\varrho < 1/64$ (recall that ϱ was introduced in the definition (see (4.10)) of ψ). Recall also the definitions of a matched couple and of a matched pushforward given in Section 8.1

Lemma 9.1 (Coupling Step). *For any $\bar\Delta > 0$, there exist $\bar\varepsilon > 0$ so that the following holds. For any $0 \le N \le N_{\mathrm{S}}$, $\varepsilon \in (0,\bar\varepsilon)$, $\Delta \in (0,\bar\Delta)$ and $\Delta\varepsilon$-matched standard couple $\underline{\ell}$, there exist sequences of pushforwards $\left(\underline{\mathfrak{L}}^{\mathrm{C}}_n\right)_{n=0}^{N}$ and $\left(\underline{\mathfrak{L}}^{\mathrm{U}}_n\right)_{n=0}^{\infty}$ so that $\underline{\mathfrak{L}}^{\mathrm{C}}_N$ is a matched pushforward and*

$$[F^n_\varepsilon\underline{\ell}\,] \ni m_{\mathrm{C}}\underline{\mathfrak{L}}^{\mathrm{C}}_n + (1-m_{\mathrm{C}})\underline{\mathfrak{L}}^{\mathrm{U}}_n.$$

In addition,

(a) *$\mathfrak{L}^{\mathrm{C},0}_0 = \{\{0\},\ell_0\}$, $\ell_0 \subset \underline{\ell}^{\,0}$ and*

$$\hat\mu_{\mathfrak{L}^0}(\operatorname{supp}\mathfrak{L}^{\mathrm{C},0}_0) = m_{\mathrm{C}} = m_{\mathrm{C}}(\Delta) = (1-c_*\Delta\varepsilon)\cdot\exp(-4\mathcal{D}_{T_{\mathrm{S}}}\Delta),$$

where c_* is a constant which does not depend on $\underline{\ell}$ and $\mathcal{D}_{T_{\mathrm{S}}}$ is defined in Proposition 8.3(c);

(b) *any $\underline{\ell}_{\,N}(p) \in \underline{\mathfrak{L}}^{\mathrm{C}}_N$ is a $\Delta\varepsilon\exp(\zeta_N(p)+C_\#\varepsilon+1/32)$-matched standard couple;*

(c) *$\underline{\mathfrak{L}}^{\mathrm{U}}_n$ is a standard coupling provided that $n \ge N + C_\#|\log(c_\#\Delta)|$.*

Proof. To fix ideas, for $i = 0,1$, let $\underline{\ell}^{\,i} = (\mathbb{G}^i,\rho)$ and $[a,b] = \pi(\operatorname{supp}\underline{\ell}^{\,0}) = \pi(\operatorname{supp}\underline{\ell}^{\,1})$; let $\mathcal{H}_N$ be the N-step holonomy map between $\underline{\ell}^{\,0}$ and $\underline{\ell}^{\,1}$, defined in Section 8.2.

Fix c_* to be specified later and define a^0_* and b^0_* so that

$$\int_a^{a^0_*}\rho(x)dx = \int_{b^0_*}^{b}\rho(x)dx = \frac{1}{2}c_*\Delta\varepsilon.$$

By (5.6), the definition of $\mathcal{H}_N$ and our estimates for the center cone, we can choose c_* to be so large that the interval $[a^0_*, b^0_*]$ is in the domain of definition of $\mathcal{H}_N$; we let $[a^1_*, b^1_*] = \mathcal{H}_N[a^0_*, b^0_*]$. Moreover, eventually by further increasing c_*, we also guarantee that for $i \in \{0, 1\}$:

$$a^i_* - a, b - b^i_* \geq \Delta\varepsilon. \tag{9.1}$$

Finally, we assume c_* to be sufficiently large so that

$$\int_a^{a^1_*} \rho(x)dx = \int_{b^1_*}^b \rho(x)dx \leq c_*\Delta\varepsilon.$$

For $i \in \{0, 1\}$ let us cut the standard pair $\underline{\ell}^{\,i}$ at the points a^i_* and b^i_*; in doing so we obtain two (very) short standard pairs (which we denote by ℓ^i_{L} and ℓ^i_{R}) whose lengths are bounded below by (9.1) and a (possibly short) standard pair, which we denote by ℓ^i_*, with $|\ell^i_*| \geq \delta/2 - C_\# \Delta\varepsilon$ (see Figure 1 for a sketch of our setup).

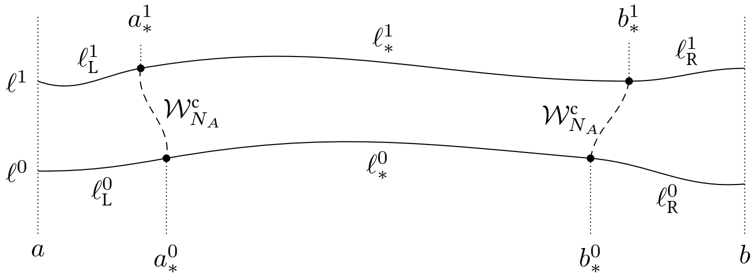

Figure 1. Setup for our decomposition.

Let us introduce some notation: we define $\ell^i_* = (\mathbb{G}^i_*, \rho^i_*)$ (resp. $\ell^i_{\mathrm{L}} = (\mathbb{G}^i_{\mathrm{L}}, \rho^i_{\mathrm{L}})$, $\ell^i_{\mathrm{R}} = (\mathbb{G}^i_{\mathrm{R}}, \rho^i_{\mathrm{R}})$), where $\mathbb{G}^i_*$ (resp. $\mathbb{G}^i_{\mathrm{L}}$, $\mathbb{G}^i_{\mathrm{R}}$) is the restriction of $\mathbb{G}^i$ to the interval $[a^i_*, b^i_*]$ (resp. $[a, a^i_*]$, $[b^i_*, b]$) and $\rho^i_* = \rho/m^i_*$ (resp. $\rho^i_{\mathrm{L}} = \rho/m^i_{\mathrm{L}}$, $\rho^i_{\mathrm{R}} = \rho/m^i_{\mathrm{R}}$) where $m^i_* = \int_{a^i_*}^{b^i_*} \rho(x)dx$ (resp. $m^i_{\mathrm{L}} = \int_a^{a^i_*} \rho(x)dx$, $m^i_{\mathrm{R}} = \int_{b^i_*}^b \rho(x)dx$). In particular, our construction yields $m^0_{\mathrm{L}} = m^0_{\mathrm{R}} = c_*\Delta\varepsilon/2$ and $m^0_* = 1 - c_*\Delta\varepsilon$.

Let $\rho^0_{\mathrm{C}} = \rho^0_*$ and define ρ^1_{C} on $[a^1_*, b^1_*]$ as the push-forward $\rho^1_{\mathrm{C}} = \mathcal{H}_{N*}\rho^0_{\mathrm{C}}$. More explicitly, for any $x^1 \in [a^1_*, b^1_*]$ let $x^0 = \mathcal{H}^{-1}_N x^1$, then:

$$\rho^1_{\mathrm{C}}(x^1) = \frac{\rho^0_{\mathrm{C}}(x^0)}{\mathcal{H}'_N(x^0)}; \tag{9.2}$$

observe that ρ^1_{C} is not necessarily a standard density. We now claim that

$$\rho^1_{\mathrm{C}}(x^1) \leq \exp(2\mathcal{D}_{T_{\mathrm{S}}}\Delta) m^1_* \rho^1_*(x^1). \tag{9.3}$$

In fact, by (9.2) and since by definition $m^i_*\rho^i_* = \rho$:

$$\begin{aligned}\left|\log \frac{\rho^1_{\mathrm{C}}(x^1)}{m^1_*\rho^1_*(x^1)}\right| &= \left|\log \frac{1}{m^0_*} \frac{m^0_*\rho^0_*(x^0)}{m^1_*\rho^1_*(x^1)} \frac{1}{\mathcal{H}'_N(x^0)}\right| \\ &\leq \left|\log m^0_*\right| + \left|\log \frac{\rho(x^0)}{\rho(x^1)}\right| + \left|\log \mathcal{H}'_N(x^0)\right| \leq C_\#\Delta\varepsilon + \mathcal{D}_{T_{\mathrm{S}}}\Delta,\end{aligned}$$

where the first two terms can be bounded using invariance of the center cone and the definition of standard density, and the third one using Proposition 8.3 (where the reader can also find the definition of $\mathcal{D}_{T_{\mathrm{S}}}$). Thus, provided that $\bar{\varepsilon}$ is sufficiently small, (9.3) holds, which in turn implies that there exist positive densities $\rho^i_{\mathrm{U}*}$ so that, letting $m_{\mathrm{C}} = m^0_* \exp(-4\mathcal{D}_{T_{\mathrm{S}}}\Delta)$:

$$m^0_*\rho^0_*(x^0) = m_{\mathrm{C}}\rho^0_{\mathrm{C}}(x^0) + (m^0_* - m_{\mathrm{C}})\rho^0_{\mathrm{U}*}(x^0) \tag{9.4a}$$

$$m^1_*\rho^1_*(x^1) = \underbrace{m_{\mathrm{C}}\rho^1_{\mathrm{C}}(x^1)}_{\text{coupled}} + \underbrace{(m^1_* - m_{\mathrm{C}})\rho^1_{\mathrm{U}*}(x^1)}_{\text{uncoupled}}; \tag{9.4b}$$

where, in particular $\rho^0_{\mathrm{U}*} = \rho^0_{\mathrm{C}} = \rho^0_*$.

Let us now define $\ell^{\mathrm{C},i} = (\mathbb{G}^i_*, \rho^i_{\mathrm{C}})$ and $\ell^i_{\mathrm{U}*} = (\mathbb{G}^i_*, \rho^i_{\mathrm{U}*})$; let furthermore

$$\mathfrak{L}^{\mathrm{U},i} = \frac{m^i_{\mathrm{L}}}{1-m_{\mathrm{C}}}\ell^i_{\mathrm{L}} + \frac{m^i_{\mathrm{R}}}{1-m_{\mathrm{C}}}\ell^i_{\mathrm{R}} + \frac{m^i_* - m_{\mathrm{C}}}{1-m_{\mathrm{C}}}\ell^i_{\mathrm{U}*}.$$

We let $\underline{\ell}^{\,\mathrm{C}}$ be the coupling of $\ell^{\mathrm{C},0}$ and $\ell^{\mathrm{C},1}$ given by

$$\boldsymbol{\mu}(g) = \int g(\mathbb{G}^0_*(x), \mathbb{G}^1_*(H_N(x))\rho^0_{\mathrm{C}}(x)dx,$$

and $\underline{\mathfrak{L}}^{\mathrm{U}}$ be the independent coupling of $\mathfrak{L}^{\mathrm{U},0}$ and $\mathfrak{L}^{\mathrm{U},1}$. Also, let $\underline{\mathfrak{L}}^{\mathrm{C}}_n$ and $\underline{\mathfrak{L}}^{\mathrm{U}}_n$ be pushforwards of $\underline{\ell}^{\,\mathrm{C}}$ and $\underline{\mathfrak{L}}^{\mathrm{U}}$, respectively. We claim that these couplings satisfy properties (a)–(c).

In fact, (a) follows by our construction. Then, observe that, by Proposition 8.3(a), the couples $\underline{\ell}_{\,N}(p)$ are in fact matched. Since $\ell^{\mathrm{C},0}$ is standard, $\underline{\ell}^{\,0}_{\,N}(p)$ will also be standard, and consequently so will be $\underline{\ell}^{\,1}_{\,N}(p)$, since the two pairs have equal densities; item (b) then follows by estimates (8.5).

We now proceed to prove item (c): let $\underline{\ell}^{\,i}_{\,N}(\alpha)$ be a pair in $\mathfrak{L}^{\mathrm{U},i}_N$; there are two possibilities:

i. $\underline{\ell}^{\,i}_{\,N}(\alpha)$ belongs to the N-th image of either ℓ^i_{R} or ℓ^i_{L}

ii. $\underline{\ell}^{\,i}_{\,N}(\alpha)$ belongs to the N-th image of $\ell^i_{\mathrm{U}*}$.

In the first case, we know by (9.1) that the length of the short curves ℓ^i_{L} or ℓ^i_{R} is bounded below by $c_\#\Delta\varepsilon$; Remark 5.6 then implies that the pairs ℓ^i_{L} and ℓ^i_{R} are $C_\#|\log(c_\#\Delta\varepsilon)|$-prestandard, which in particular proves item (c) in case i. We are left with case ii: by definition $\ell^0_{\mathrm{U}*}$ is a standard pair, hence so will be $\ell^0_N(\alpha)$. We therefore only need to prove our statement for pairs in the image of $\ell^1_{\mathrm{U}*} = (G^1_*, \rho^1_{\mathrm{U}*})$. By (9.4b) we have:

$$(m^1_* - m_{\mathrm{C}})\rho^1_{\mathrm{U}*} = m^1_*\rho^1_* - m_{\mathrm{C}}\rho^1_{\mathrm{C}}$$

Let us denote by $\rho^1_{\mathrm{U}*,N}$ the pushforward of $\rho^1_{\mathrm{U}*}$ by $F^N_{\varepsilon*}$; we obtain:

$$\begin{aligned}(m^1_* - m_{\mathrm{C}})\rho^1_{\mathrm{U}*,N}(x_N) &= (m^1_* - m_{\mathrm{C}})\rho^1_{\mathrm{U}*}(x^1_0(x_N))\frac{dx^1_0}{dx_N}\\ &= m^1_*\rho^1_*(x^1_0(x_N))\frac{dx^1_0}{dx_N} - m_{\mathrm{C}}\rho^1_{\mathrm{C}}(x^1_0(x_N))\frac{dx^1_0}{dx_N},\end{aligned} \tag{9.5}$$

where we denote with $x^i_0(x_N)$ the x-coordinate of the point $p^i_0(x_N) \in U^i_N(\alpha) \subset \operatorname{supp}\ell^i_0$ so that $\pi F^N_\varepsilon(p^i_0(x_N)) = x_N$ (recall the definition of U^i_N given in Section 5.2.2). The first term is the push-forward of a standard density (and thus a standard density); the second term is also a standard density, since by our construction $\rho^1_{\mathrm{C}}(x^1_0(x_N))\frac{dx^1_0}{dx_N} = \rho^0_{\mathrm{C}}(x^0_0(x_N))\frac{dx^0_0}{dx_N}$, which is the push-forward of a standard density. We now take derivatives of (9.5) and obtain:

$$\left\|\frac{{\rho^1_{\mathrm{U}*,N}}'}{\rho^1_{\mathrm{U}*,N}}\right\| \leq \left\|\frac{m^1_*\rho^1_* + m_{\mathrm{C}}\rho^1_{\mathrm{C}}}{m^1_*\rho^1_* - m_{\mathrm{C}}\rho^1_{\mathrm{C}}}\right\| c_2 \qquad \left\|\frac{{\rho^1_{\mathrm{U}*,N}}''}{\rho^1_{\mathrm{U}*,N}}\right\| \leq \left\|\frac{m^1_*\rho^1_* + m_{\mathrm{C}}\rho^1_{\mathrm{C}}}{m^1_*\rho^1_* - m_{\mathrm{C}}\rho^1_{\mathrm{C}}}\right\| D_2c_2,$$

and using (9.3):

$$\left\|\frac{m^1_*\rho^1_* + m_{\mathrm{C}}\rho^1_{\mathrm{C}}}{m^1_*\rho^1_* - m_{\mathrm{C}}\rho^1_{\mathrm{C}}}\right\| \leq \frac{2}{1-\exp(-2\mathcal{D}_{T_{\mathrm{S}}}\Delta)} \leq 2\left(1 + \frac{1}{2\mathcal{D}_{T_{\mathrm{S}}}\Delta}\right).$$

Hence, $\rho^1_{\mathrm{U}*,N} \in D_{2\left(1+1/2\mathcal{D}_{T_{\mathrm{S}}}\Delta\right)c_2}(G^1_{\mathrm{U}*,N})$ and by Remark 5.7 we can thus conclude that any pair in case ii is a $C_\#|\log(c_\#\Delta)|$-prestandard pair. $\square$

Corollary 9.2. *For any $\bar\Delta > 0$, there exist $\bar\varepsilon > 0$ so that the following holds. For any $N \le N_{\rm S}$, $\varepsilon \in (0, \bar\varepsilon)$, $\Delta \in (0, \bar\Delta)$ and $\Delta\varepsilon$-matched standard couple $\underline{\ell}$ we have*

$$\mathrm{d}_W(F^N_{\varepsilon *}\mu_{\ell^0}, F^N_{\varepsilon *}\mu_{\ell^1}) \le C_\# \Delta.$$

Proof. Applying Lemma 9.1 to $\underline{\ell}$ we obtain

$$\left[F^N_\varepsilon \underline{\ell}\right] \ni m_{\rm C}\mathfrak{L}^{\rm C}_N + (1 - m_{\rm C})\mathfrak{L}^{\rm U}_N.$$

By item (b) we gather that $\mathfrak{L}^{\rm C}_N$ is a $C_\#\Delta\varepsilon$-matched coupling; thus:

$$\mathrm{d}_{\rm W}(F^N_{\varepsilon *}\mu_{\ell^0}, F^N_{\varepsilon *}\mu_{\ell^1}) \le C_\# m_{\rm C}\Delta\varepsilon + (1 - m_{\rm C}),$$

from which we conclude using the estimate for $m_{\rm C}$ given in item (a). □

9.3. **The global Coupling procedure.** The idea is now to iterate Lemma 9.1 with $N = N_{\rm S}$ and discard those couples which at step k are not exponentially close in k. The crucial fact to prove is that if we start coupling pairs which are sufficiently close, this strategy can be carried out with probability arbitrarily close to 1; this, together with other useful estimates, is the content of the following lemma, whose proof will be given in Section 10.1. Recall the definition of conditioned subcouplings given in Section 8.1.

Lemma 9.3. *For any $\gamma > 0$, provided that ε is small enough, there exists $\tau > 0$ so that for any $\varepsilon^{1+\tau}$-matched standard couple $\underline{\ell}$, there exists a sequence $\mathfrak{L}_{[k]} \in \left[F^{kN_{\rm S}}_\varepsilon \underline{\ell}\right]$, $k \in \mathbb{N}$, and random variables*[37] *$U_{[k]} : \mathcal{A}_{[k]} \to \mathbb{Z} \cup \{\infty\}$ satisfying the following properties:*

(a) *for any $k \ge 0$ and $\alpha \in \mathcal{A}_{[k]}$ such that $U_{[k]}(\alpha) = \infty$, the standard couple $\underline{\ell}_{[k]}(\alpha)$ is $C_\# \exp(-c_\# k)\varepsilon^{1+\tau/2}$-matched.*
(b) *for any $l < k \le k'$ we have $\nu_{[k]}(\{U_{[k]} = l\}) = \nu_{[k']}(\{U_{[k']} = l\})$; moreover $\mathfrak{L}_{[k']}|\{U_{[k']} = l\} \in \left[F^{(k'-k)N_{\rm S}}_\varepsilon \mathfrak{L}_{[k]}|\{U_{[k]} = l\}\right]$; finally, the family $\mathfrak{L}_{[l]}|\{U_{[l]} = l - 1\}$ is $lN_{\rm S}$-prestandard.*
(c) *$M_{{\rm C}_k} = \nu_{[k]}(U_{[k]} = \infty)$ is a non-increasing sequence in $[0, 1]$ so that, for all $k \in \mathbb{N}$,*

$$M_{{\rm C}_k} \ge \exp(-\gamma). \tag{9.6}$$

Moreover, if $k' \ge k$, we have

$$M_{{\rm C}_k} - M_{{\rm C}_{k'}} \le \gamma \exp(-c_\# k/\log \varepsilon^{-1}). \tag{9.7}$$

Remark 9.4. *As we already explained, the lack of uniform hyperbolicity implies that the dynamics might fail to bring together at a uniform rate two standard pairs which started close together. When such a failure happens, we declare the couple to* break up *and we give up tracking them in the future. The above lemma tells us that if two standard pairs are sufficiently close, such break ups are relatively unlikely. The random variable $U_{[k]}$ in the above statement keeps track of the coupling step at which the corresponding couple broke up (up to the k-th step); if it is ∞, it means that the couple did not break up (yet). Thus (9.6) guarantees that a break up will happen with probability which is arbitrarily small with γ; similarly (9.7) gives an exponential tail bound (with rate $\mathcal{O}(\varepsilon/\log\varepsilon^{-1})$) on the probability of a break up occurring after k steps.*

Observe that Lemma 9.3 requires the standard couple $\underline{\ell}$ to be $C_\#\varepsilon^{1+\tau}$-matched. The following Lemma specifies under which conditions in $\mathcal{O}(\varepsilon^{-1}\log\varepsilon^{-1})$ iterations the dynamics will bring a portion of the image of two standard pairs in such a convenient position. Recall that a standard pair ℓ *is located at the trapping set* $\mathcal{T}_{\epsilon,i}$

[37] Recall that, according to Notational Remark 8.2, $\mathcal{A}_{[k]}$ is the index set of $\mathfrak{L}_{[k]}$ and $\nu_{[k]}$ the corresponding measure.

if $\theta^*_\ell \in \mathcal{T}_{\epsilon,i}$; likewise a standard family $\mathfrak{L}$ is said to *be located at* $\mathcal{T}_{\epsilon,i}$ if for any $\alpha \in \mathcal{A}$, $\theta^*_{\ell(\alpha)}$ is located at to $\mathcal{T}_{\epsilon,i}$. A standard couple $\underline{\ell} = (\ell^0, \ell^1)$ is said to *be located at the trapping set* $\mathcal{T}_{\epsilon,i}$ if both ℓ^0 and ℓ^1 are located at $\mathcal{T}_{\epsilon,i}$. Finally, we denote with $n_{\mathrm{Z},i}$ the number of sinks $\theta_{j,-}$ that are contained in $\mathcal{T}_{\epsilon,i}$.[38]

Remark 9.5. *Observe that by (6.16), if ℓ is located at $\mathcal{T}_{\epsilon,i}$, then any n-pushforward of ℓ is located at $\mathcal{T}_{\epsilon,i}$ provided that $n \geq \lfloor T_F \varepsilon^{-1} \rfloor$.*

Lemma 9.6 (Bootstrap)**.** *Let $\theta_{i,-}$ be a recurrent sink; for any $\tau > 0$, there exist $\mathcal{R}_{\mathrm{B}}$ and $\bar{\varepsilon} > 0$ so that for any $\mathcal{R} \geq \mathcal{R}_{\mathrm{B}}$, $\mathcal{K} = \lfloor \mathcal{R} \log \varepsilon^{-1} \rfloor$, $\varepsilon \in (0, \bar{\varepsilon})$ and any standard couple $\underline{\ell}$ located at $\mathcal{T}_{\epsilon,i}$, we have*

$$[F_{\varepsilon *}^{\mathcal{K} N_{\mathrm{S}}} \mu_{\underline{\ell}}] \ni m_{\mathrm{B}} \underline{\mathfrak{L}}^{\mathrm{B}} + (1 - m_{\mathrm{B}}) \underline{\mathfrak{L}}^{\mathrm{R}},$$

where:

(a) *$\underline{\mathfrak{L}}^{\mathrm{B}} = (\underline{\mathfrak{L}}^{\mathrm{B},0}, \underline{\mathfrak{L}}^{\mathrm{B},1})$ is an $\varepsilon^{1+\tau}$-matched standard coupling;*
(b) *$\mathfrak{L}^{\mathrm{R},0}$ and $\mathfrak{L}^{\mathrm{R},1}$ are $\mathcal{O}(\tau \log \varepsilon^{-1})$-prestandard families;*
(c) *$m_{\mathrm{B}} = m_{\mathrm{B}}(\mathcal{R})$ is a non-increasing function of $\mathcal{R}$; moreover if $n_{\mathrm{Z},i} = 1$, m_{B} can be chosen to be uniform in ε; otherwise $m_{\mathrm{B}} \sim \exp(-c_\# \varepsilon^{-1})$.*

The proof of Lemma 9.6 will be given in Section 10.2. Recall the definition of Wasserstein distance given in (8.1). We now see how the previous results allow us to prove the following

Lemma 9.7 (Coupling Lemma)**.** *There exist $\bar{\varepsilon} > 0$ so that, if $\varepsilon \in (0, \bar{\varepsilon})$, for any two standard pairs ℓ^0, ℓ^1 located at the same trapping set:*

$$\mathrm{d}_W(F_{\varepsilon *}^n \mu_{\ell^0}, F_{\varepsilon *}^n \mu_{\ell^1}) \leq C_\# \exp(-c_\# m_{\mathrm{B}} \cdot n\varepsilon / \log \varepsilon^{-1})$$

Proof. Our main task is essentially a bookkeeping problem: as we push forward a standard couple we will produce matched pairs (hopefully more and more of them), prestandard pairs that cannot be used for anything as yet, and standard pairs that have recovered and are ready to reenter in the dating business. To keep track of all these objects some notation is needed. Let $\gamma > 0$ small and $r \in \mathbb{N}$ large enough to be specified later; define $\mathcal{R}_{\mathrm{C}}$ by requiring that

$$\mathcal{K}_{\mathrm{C}} = \lfloor \mathcal{R}_{\mathrm{C}} \log \varepsilon^{-1} \rfloor = 2 \lfloor r \log \varepsilon^{-1} \rfloor > \lfloor \mathcal{R}_{\mathrm{B}} \log \varepsilon^{-1} \rfloor,$$

where $\mathcal{R}_{\mathrm{B}}$ is the constant appearing in Lemma 9.6.

To fix ideas we assume that ℓ^0 and ℓ^1 both located at $\mathcal{T}_{\epsilon,i}$; we will now inductively define:

- for $q \geq 0$, a sequence $(\underline{\mathfrak{L}}^{[q]*})_q$ of couplings of N_{S}-prestandard families located at $\mathcal{T}_{\epsilon,i}$ and a corresponding sequence of weights $(M_{[q]*})_q$
- for $q \geq 1$, a sequence $(\underline{\mathfrak{L}}^{[q]})_q$ of $C_\# \varepsilon^{1+\tau}$-matched standard couplings located at $\mathcal{T}_{\epsilon,i}$ and a corresponding sequence of weights $(M_{[q]})_q$.

The reader should think of such families as a bookkeeping device to account for the dynamics after $q \mathcal{K}_{\mathrm{C}} N_{\mathrm{S}}$ iterates. Roughly speaking $\underline{\mathfrak{L}}^{[q]}$ are the standard pairs that we are able to couple at time $q \mathcal{K}_{\mathrm{C}} T_{\mathrm{S}}$. At later times some of this standard pairs break up (this is recorded by the random variables $U_{[k]}$ defined in Lemma 9.3) or lose some mass (in form of, possibly very short, prestandard pairs) while some standard pairs never had a chance to couple. The family $\underline{\mathfrak{L}}^{[q]*}$ contains all the standard pairs that are available to try a new coupling in the time interval $[q \mathcal{K}_{\mathrm{C}} T_{\mathrm{S}}, (q+1) \mathcal{K}_{\mathrm{C}} T_{\mathrm{S}}]$. The reason why such a scheme is going to converge is that, as time goes on, less and less mass uncouples (see Lemma 9.3), while it is always possible to couple a fix percentage of the uncoupled mass (see Lemma 9.6).

[38] Remark that $n_{\mathrm{Z},i}$ does not depend on ϵ, provided ϵ has been chosen small enough.

Let us now describe the induction step: at step q, we inductively assume that $M_{[q]*}$ and $\underline{\mathfrak{L}}^{[q]*}$ are defined, together with $M_{[s]}$ and $\underline{\mathfrak{L}}^{[s]}$ for $0 < s \leq q$ and construct $M_{[q+1]}$, $\underline{\mathfrak{L}}^{[q+1]}$, $M_{[q+1]*}$ and $\underline{\mathfrak{L}}^{[q+1]*}$.

For the base step, let $M_{[0]*} = 1$ and $\underline{\mathfrak{L}}^{[0]*} = \underline{\ell}$.

Next, consider $q > 0$. Let $\tau > 0$ be the constant given by Lemma 9.3. By our inductive assumptions $\underline{\mathfrak{L}}^{[q]*}$ is a couple of N_{S}-prestandard families; let $\underline{\mathfrak{L}}^{[q]*}_{N_{\mathrm{S}}}$ be a standard pushforward of $\underline{\mathfrak{L}}^{[q]*}$; then, we can apply Lemma 9.6 to each couple of standard pairs in $\underline{\mathfrak{L}}^{[q]*}_{N_{\mathrm{S}}}$ with $\mathcal{R} = (\mathcal{K}_{\mathrm{C}} - 1)/\log\varepsilon^{-1} \geq \mathcal{R}_{\mathrm{B}}$ and obtain

$$\left[F_\varepsilon^{\mathcal{K}_{\mathrm{C}}N_{\mathrm{S}}}\underline{\mathfrak{L}}^{[q]*}\right] \ni m_{\mathrm{B}}\left(\underline{\mathfrak{L}}^{[q]*}\right)^{\mathrm{B}} + (1-m_{\mathrm{B}})\left(\underline{\mathfrak{L}}^{[q]*}\right)^{\mathrm{R}}.$$

We define $\underline{\mathfrak{L}}^{[q+1]} = \left(\underline{\mathfrak{L}}^{[q]*}\right)^{\mathrm{B}}$ and $M_{[q+1]} = M_{[q]*}m_{\mathrm{B}}$. Observe that, by construction, $\underline{\mathfrak{L}}^{[q+1]}$ is a $\varepsilon^{1+\tau}$-matched standard coupling located at $\mathcal{T}_{\epsilon,i}$. Let us denote with $\underline{\mathfrak{L}}^{[q+1]}_{[k]}$ the sequence of couplings and with $U^{[q+1]}_{[k]}$ the sequence of random variables which we obtain by applying Lemma 9.3 to each standard coupling in $\underline{\mathfrak{L}}^{[q+1]}$.

Note that, according to Lemma 9.3, a certain number of pairs will break up as times goes by. The variables $U^{[q]}_{[k]}$ keep track of when such breakups occurred. Moreover, recall that Lemma 9.3 asserts that if a standard couple in $\underline{\mathfrak{L}}^{[q]}_{[k]}$ broke up at time s (i.e. $U^{[q]}_{[k]} = s$), then it will recover at time sN_{S}. Hence the couple in the family $\underline{\mathfrak{L}}^{[q]}_{[k]}$ that broke up at the step $\mathcal{O}(k/2)$ have recovered (that is, are standard), thus available for starting again a coupling procedure.

Then, we define the coupling $\underline{\mathfrak{L}}^{[q+1]*}$ so that

$$\begin{aligned} M_{[q+1]*}\underline{\mathfrak{L}}^{[q+1]*} &= M_{[q]*}(1-m_{\mathrm{B}})\left(\underline{\mathfrak{L}}^{[q]*}\right)^{\mathrm{R}} + \\ &\quad + \sum_{s=1}^{q} M_{[s]}\nu^{[s]}_{[Y_s^q\mathcal{K}_{\mathrm{C}}]}\left(U^{[s]}_{[Y_s^q\mathcal{K}_{\mathrm{C}}]} \in [(Y_s^q-1)\mathcal{K}_{\mathrm{C}}/2, Y_s^q\mathcal{K}_{\mathrm{C}}/2)\right) \\ &\qquad \times \underline{\mathfrak{L}}^{[s]}_{[Y_s^q\mathcal{K}_{\mathrm{C}}]}\Big|\left\{U^{[s]}_{[Y_s^q\mathcal{K}_{\mathrm{C}}]} \in [(Y_s^q-1)\mathcal{K}_{\mathrm{C}}/2, Y_s^q\mathcal{K}_{\mathrm{C}}/2)\right\} \end{aligned}$$

where $Y_s^q = q - s + 1$. In the above expression, the first term accounts for standard pairs which did not come close enough during the current step and we could not start coupling. The second terms account for standard pairs which we coupled in some previous step, broke up and recovered some time between the beginning and the end of the current step. Correspondingly we let

$$M_{[q+1]*} = M_{[q]*}(1-m_{\mathrm{B}})\sum_{s=1}^{q} M_{[s]}\left(M_{\mathrm{C}_{(Y_s^q-1)\mathcal{K}_{\mathrm{C}}/2}} - M_{\mathrm{C}_{Y_s^q\mathcal{K}_{\mathrm{C}}/2}}\right). \tag{9.8}$$

Observe that by definition $\underline{\mathfrak{L}}^{[q+1]*}$ is located at $\mathcal{T}_{\epsilon,i}$. Now that we defined the auxiliary sequences of couplings, we claim that

$$\frac{M_{[q+1]*}}{M_{[q]*}} \in [\vartheta_*, \vartheta], \tag{9.9}$$

where $\vartheta = 1 - \frac{1}{2}m_{\mathrm{B}}$ and $\vartheta_* = 1 - m_{\mathrm{B}}$; observe that both ϑ and ϑ_* increase with r by Lemma 9.6(c). In fact by (9.8) and the definition of $M_{[s]}$, we have:

$$\frac{M_{[q+1]*}}{M_{[q]*}} = (1-m_{\mathrm{B}}) + m_{\mathrm{B}}\sum_{s=0}^{q-1}\frac{M_{[s]*}}{M_{[q]*}}\left(M_{\mathrm{C}_{(Y_s^q-2)\mathcal{K}_{\mathrm{C}}/2}} - M_{\mathrm{C}_{(Y_s^q-1)\mathcal{K}_{\mathrm{C}}/2}}\right);$$

the above immediately implies the lower bound $\frac{M_{[q+1]*}}{M_{[q]*}} \geq \vartheta_*$, since every term of the sum is positive. In order to prove the upper bound observe that, by the lower bound and the above equation:

$$\frac{M_{[q+1]*}}{M_{[q]*}} \leq (1 - m_{\mathrm{B}}) + \vartheta_*^{-1} m_{\mathrm{B}} \gamma \left[\sum_{k=0}^{q-1} \vartheta_*^{-k} \exp(-c_\# r k) \right]$$

where we used (9.7); observe that by choosing γ small and r large we can make the second term arbitrarily small, from which we conclude that (9.9) holds.

Let us now fix $n > 0$; let $k = \lfloor n/N_{\mathrm{S}} \rfloor$, $q = \lfloor k/\mathcal{K}_{\mathrm{C}} \rfloor$ and, for $0 \leq s \leq q$ define $\varkappa_s = k - s\mathcal{K}_{\mathrm{C}}$. Let us first construct a coupling $\underline{\mathfrak{L}}_{kN_{\mathrm{S}}} \in \left[F_\varepsilon^{kN_{\mathrm{S}}} \underline{\ell} \right]$ given by:

$$\begin{aligned}
\underline{\mathfrak{L}}_{kN_{\mathrm{S}}} := & \sum_{s=1}^{q} M_{[s]} \nu_{[\varkappa_s]}^{[s]} (U_{[\varkappa_s]}^{[s]} = \infty) \cdot \underline{\mathfrak{L}}_{[\varkappa_s]}^{[s]} | \{ U_{[\varkappa_s]}^{[s]} = \infty \} \\
& + \sum_{s=1}^{q} M_{[s]} \nu_{[\varkappa_s]}^{[s]} (U_{[\varkappa_s]}^{[s]} \in [Y_s^{q-1} \mathcal{K}_{\mathrm{C}}/2, \varkappa_s)) \cdot \underline{\mathfrak{L}}_{[\varkappa_s]}^{[s]} | \{ U_{[\varkappa_s]}^{[s]} \in [Y_s^{q-1} \mathcal{K}_{\mathrm{C}}/2, \varkappa_s) \} \\
& + M_{[q]*} \underline{\mathfrak{L}}_{\varkappa_q N_{\mathrm{S}}}^{[q]*},
\end{aligned}$$

where, $\underline{\mathfrak{L}}_n^{[q]*}$ is an arbitrary n-pushforward of $\underline{\mathfrak{L}}^{[q]*}$. In the above expression, the first term accounts for pairs which we coupled at earlier steps and have not broken up yet; the second term accounts for all pairs which we coupled at any of the previous steps, broke up and have not recovered yet. The third and last term accounts for N_{S}-prestandard pairs which were uncoupled but recovered by the beginning of step q and will try to get coupled in this step.

For pairs belonging to the first term we can use Lemma 9.3(a) and obtain that every pair in $\underline{\mathfrak{L}}_{[\varkappa_s]}^{[s]} | \{ U_{[\varkappa]}^{[s]} = \infty \}$ is $C_\# \varepsilon^{1+\tau/2} \exp(-c_\# \varkappa_s)$-matched. For pairs belonging to the families appearing in the remaining two terms we do not have any estimate on the Wasserstein distance , therefore we can only bound it with 1.

Thus, we can use Corollary 9.2 and conclude that:

$$\begin{aligned}
\mathrm{d_W}(F_{\varepsilon*}^n \mu_{\ell^0}, F_{\varepsilon*}^n \mu_{\ell^1}) & \leq \mathrm{d_W}(F_{\varepsilon*}^{n-kN_{\mathrm{S}}} \mu_{\mathfrak{L}_{kN_{\mathrm{S}}}^0}, F_{\varepsilon*}^{n-kN_{\mathrm{S}}} \mu_{\mathfrak{L}_{kN_{\mathrm{S}}}^1}) \\
& \leq C_\# \sum_{s=1}^{q} M_{[s]} \nu_{[\varkappa_s]}^{[s]} (U_{[\varkappa_s]}^{[s]} = \infty) \varepsilon^{\tau/2} \exp(-c_\# \varkappa_s) \\
& \quad + C_\# \sum_{s=1}^{q} M_{[s]} \nu_{[\varkappa_s]}^{[s]} (U_{[\varkappa_s]}^{[s]} \in [Y_s^{q-1} \mathcal{K}_{\mathrm{C}}/2, \bar{m}_s)) + C_\# M_{[q]*} \\
& = \mathrm{I} + \mathrm{II} + \mathrm{III}.
\end{aligned}$$

Let us estimate term I: by our estimate for $M_{[s]*}$ we gather

$$\begin{aligned}
\mathrm{I} & \leq C_\# m_{\mathrm{B}} \varepsilon^{\tau/2} \sum_{s=1}^{q} \vartheta^s \exp(-c_\# \varkappa_s) \\
& \leq C_\# m_{\mathrm{B}} \varepsilon^{\tau/2} \sum_{s=1}^{q} \vartheta^s \exp(-c_\# (q-s) \mathcal{K}_{\mathrm{C}}) \leq C_\# m_{\mathrm{B}} \varepsilon^{\tau/2} \vartheta^q.
\end{aligned}$$

This proves exponential decay for term I. Similarly, for term II: using (9.7) we obtain

$$\mathrm{II} \leq C_\# m_{\mathrm{B}} \gamma \sum_{s=1}^{q} \vartheta^s \exp(-c_\# (q-s) r) \leq C_\# m_{\mathrm{B}} \gamma \vartheta^q,$$

by choosing r sufficiently large. We already proved, just after (9.8), exponential decay for term III, (i.e. $M_{[q]*} \leq \vartheta^q$). The proof then readily follows by collecting all above estimates. □

9.4. **Proof of the Main Theorem.** Our Main Theorem is a direct consequence of Lemma 9.7 and the definition of Wasserstein distance (see (8.1)); but first we owe to the reader the proof of the the following

Lemma 9.8. *Let μ be an SRB measure and let $B(\mu)$ denote its ergodic basin (see Remark 2.6); then*

(a) *μ is a weak limit of standard families.*
(b) *if $\mathrm{Leb}(B(\mu) \cap \mathcal{T}_{\epsilon,i}) > 0$, then μ is a weak limit of standard families that are located at $\mathcal{T}_{\epsilon,i}$*

Proof. For ease of notation, let $B = B(\mu)$; by Fubini's Theorem there exists a standard pair $\ell = (\mathbb{G}, \rho)$ (e.g. horizontal and with constant density) which intersects B and so that $\mu_\ell(B) > 0$; let us denote by $\mu_{\ell,B}$ the normalized restriction of μ_ℓ to B, i.e. for any test function Φ we let $\mu_{\ell,B}(\Phi) = \mu_\ell(B)^{-1} \cdot \mu_\ell(\mathbb{1}_B \cdot \Phi)$. Observe that by definition of B:

$$\frac{1}{n}\sum_{k=0}^{n-1} F^k_{\varepsilon*}\mu_{\ell,B} \to \mu \text{ weakly as } n \to \infty.$$

Fix $\varrho > 0$ be arbitrarily small; since the set $\mathbb{G}^{-1}(B) \subset [a, b]$ is measurable, it can be approximated with a finite number of disjoint intervals up to error ϱ. We conclude that there exist $N > 0$ and an N-prestandard family $\mathfrak{L}_B$ so that $\|\mu_{\mathfrak{L}_B} - \mu_{\ell,B}\|_{\mathrm{TV}} < \varrho$, where $\|\cdot\|_{\mathrm{TV}}$ denotes the total variation norm. Hence, for any n:

$$\left\|\frac{1}{n}\sum_{k=0}^{n-1} F^k_{\varepsilon*}\mu_{\mathfrak{L}_B} - \frac{1}{n}\sum_{k=0}^{n-1} F^k_{\varepsilon*}\mu_{\ell,B}\right\|_{\mathrm{TV}} < \varrho.$$

Moreover, observe that for any n

$$\left\|\frac{1}{n}\sum_{k=0}^{n-1} F^k_{\varepsilon*}\mu_{\mathfrak{L}_B} - \frac{1}{n-N}\sum_{k=N}^{n-1} F^k_{\varepsilon*}\mu_{\mathfrak{L}_B}\right\|_{\mathrm{TV}} < 2\frac{N}{n-N}.$$

Since $\frac{1}{N-n}\sum_{k=N}^{n-1} F^k_{\varepsilon*}\mu_{\mathfrak{L}_B}$ can be decomposed, by definition, in a standard family, the proof of (a) follows choosing n sufficiently large.

The proof of (b) also follows from the same argument, since our assumption guarantees that we can choose ℓ to be located at $\mathcal{T}_{\epsilon,i}$; by (6.16) the standard family $\frac{1}{N-n}\sum_{k=N}^{n-1} F^k_{\varepsilon*}\mu_{\mathfrak{L}_B}$ is located at to $\mathcal{T}_{\epsilon,i}$, which proves (b). □

We now proceed to the proof of the Main Theorem of Section 2, which we will now state, as promised, in the following stronger version. Let us denote with $n_{\mathcal{T}} \leq n_{\mathrm{Z}}$ the number of disjoint non-empty trapping sets $\mathcal{T}_{\epsilon,i}$ and, for any i, recall that we denote by $n_{\mathrm{Z},i}$ the number of sinks $\theta_{j,-}$ that are contained in $\mathcal{T}_{\epsilon,i}$ (recall also Footnote 38).

Theorem 9.9. *Assume that (A1), (A2) and (A3) hold and let $\theta_{i,-}$ be a recurrent sink. Then there exists a unique SRB measure $\mu_{\varepsilon,i}$ so that $\operatorname{supp}\mu_{\varepsilon,i} \subset \{\theta \in \mathcal{T}_{\epsilon,i}\}$ (in particular, if $\mathcal{T}_{\epsilon,i} = \mathcal{T}_{\epsilon,j}$ then $\mu_{\varepsilon,i} = \mu_{\varepsilon,j}$). The measure $\mu_{\varepsilon,i}$ enjoys exponential decay of correlation for Hölder observables in the following sense. There exist $C_1, C_2, C_3, C_4 > 0$ (independent of ε) so that for any $\alpha \in (0, 3]$, $\beta \in (0, 1]$ and any two functions $A \in \mathcal{C}^\alpha(\{\theta \in \mathcal{T}_{\epsilon,i}\})$, $B \in \mathcal{C}^\beta(\mathbb{T}^2)$:*

$$|\mathrm{Leb}(A \cdot B \circ F^n_\varepsilon) - \mathrm{Leb}(A)\mu_{\varepsilon,i}(B)| \leq C_1 \sup_\theta \|A(\cdot, \theta)\|_{\mathcal{C}^\alpha} \sup_x \|B(x, \cdot)\|_{\mathcal{C}^\beta} e^{-\alpha\beta c_{\varepsilon,i} n},$$

where

$$c_{\varepsilon,i} = \begin{cases} C_2\varepsilon/\log\varepsilon^{-1} & \text{if } n_{\mathrm{Z},i}=1, \\ C_3\exp(-C_4\varepsilon^{-1}) & \text{otherwise.} \end{cases} \tag{9.10}$$

Our Main Theorem stated in Section 2 then follows as a corollary:

Corollary 9.10. *Under assumptions (A1), (A2), (A3) and (A5), provided $\varepsilon > 0$ is sufficiently small, F_ε admits exactly $n_{\mathcal{T}}$ SRB measures.*

Under assumptions (A1), (A2), (A3) and (A4), there exists a unique SRB measure μ_ε for F_ε; moreover μ_ε enjoys exponential decay of correlations as stated in the Main Theorem of Section 2.

Proof. If (A4) holds, then (see Remark 6.15) $\mathcal{T}_{\epsilon,1} = \mathbb{T}$ and thus $n_{\mathrm{Z}} = n_{\mathrm{Z},1}$. Then, Theorem 9.9, immediately implies existence and uniqueness of the SRB measure μ_ε for F_ε and that μ_ε enjoys the required properties.

On the other hand, if (A5) holds, we want to prove that there cannot be any other SRB measure than the ones found by Theorem 9.9. We can argue as follows: let μ be an SRB measure; as in the proof of Lemma 9.8, there exists a standard pair ℓ so that $\mu_\ell(B(\mu)) > 0$; by (6.15), we gather that, for some $n > 0$ and $i \in \{1, \cdots, n_{\mathrm{Z}}\}$, $\mathrm{Leb}\,(B(\mu) \cap F_\varepsilon^{-n}\{\theta \in \mathcal{T}_{\epsilon,i}\}) > 0$. Since by definition $B(\mu)$ is a F_ε-invariant set, we gather $F^n_{\varepsilon*}\mathrm{Leb}\,(B(\mu) \cap \{\theta \in \mathcal{T}_{\epsilon,i}\}) > 0$, but since F_ε is a local diffeomorphism, $F^n_{\varepsilon*}\mathrm{Leb}$ is absolutely continuous with respect to the Lebesgue measure. Consequently, we have $\mathrm{Leb}\,(B(\mu) \cap \{\theta \in \mathcal{T}_{\epsilon,i}\}) > 0$, hence $\mu = \mu_{\varepsilon,i}$ by Lemma 9.8(b). We thus conclude that F_ε admits exactly $n_{\mathcal{T}}$ SRB measures. $\square$

Proof of Theorem 9.9. Let $\theta_{i,-}$ be a recurrent sink and ℓ be a standard pair located at $\mathcal{T}_{\epsilon,i}$. First, we prove that the sequence $F^n_{\varepsilon*}\mu_\ell$ weakly converges to a SRB measure $\mu_{\varepsilon,i}$ which is independent of ℓ. In fact, Remark 9.5 implies that if $n > \lfloor T_{\mathrm{F}}\varepsilon^{-1}\rfloor$, the measure $F^n_\varepsilon\mu_\ell$ can be decomposed in a standard family which is located at $\mathcal{T}_{\epsilon,i}$. Then, for any $n > \lfloor T_{\mathrm{F}}\varepsilon^{-1}\rfloor$, $m > 0$ and Hölder observable $B \in \mathcal{C}^\beta(\mathbb{T}^2,\mathbb{R})$ (where $\beta \in (0,1]$), Lemma 9.7 implies:

$$\begin{aligned} |F^{n+m}_{\varepsilon*}\mu_\ell(B) - F^n_{\varepsilon*}\mu_\ell(B)| &\leq \int_{\mathcal{A}_m} d\nu(\alpha)\left|F^n_{\varepsilon*}\mu_{\ell_m(\alpha)}(B) - F^n_{\varepsilon*}\mu_\ell(B)\right| \\ &\leq C_\#\exp(-\beta c_\varepsilon(n - \lfloor T_{\mathrm{F}}\varepsilon^{-1}\rfloor))\|B\|_{x,\beta} \\ &\leq C_\#\exp(-\beta c_\varepsilon n)\|B\|_{x,\beta}, \end{aligned} \tag{9.11}$$

where $\mathfrak{L}_m = (\ell_m, \mathcal{A}_m)$ is a standard m-pushforward of ℓ, $\|B\|_{x,\beta} = \sup_x \|B(x,\cdot)\|_{\mathcal{C}^\beta}$ and c_ε satisfies (9.10) by Lemma 9.6(c).

In particular, $F^n_{\varepsilon*}\mu_\ell(B)$ is a Cauchy sequence and, choosing B Lipschitz, this implies that the sequence of probability measures $F^n_{\varepsilon*}\mu_\ell$ has a unique weak accumulation point. Lemma 9.7 also implies that the sequence $F^n_{\varepsilon*}\mu_{\ell'}$ has the same weak accumulation point for any ℓ' which is located at $\mathcal{T}_{\epsilon,i}$ and that convergence is exponentially fast. Let us denote by $\mu_{\varepsilon,i}$ this accumulation point; by construction it is F_ε-invariant.

We now show that $\mu_{\varepsilon,i}$ is indeed a SRB measure in the sense of Remark 2.6: consider a measurable partition $\{I_\xi\}_{\xi\in\Xi}$ of $\mathbb{T} \times \mathcal{T}_{\epsilon,i}$ in horizontal segments[39] of length between $\delta/2$ and δ with indices in some measure space Ξ. That is, we let $I_\xi = [a_\xi, b_\xi] \times \{y_\xi\}$ for some $a_\xi, b_\xi, y_\xi \in \mathbb{T}$ with $\delta/2 \leq b_\xi - a_\xi \leq \delta$. Let $\mathrm{Leb}_i = \mathrm{Leb}_{\{\theta\in\mathcal{T}_{\epsilon,i}\}}$ be the restriction of Lebesgue measure to $\{\theta \in \mathcal{T}_{\epsilon,i}\}$ normalized to be a probability measure. Then by definition $\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}F^k_{\varepsilon*}\mathrm{Leb}_i$ is a convex combination of SRB measures whose ergodic basin intersects $\{\theta \in \mathcal{T}_{\epsilon,i}\}$ in a positive Lebesgue measure set. By Lemma 9.8(b) and our previous argument, we conclude

[39] Notice that by Lemma 6.13(d) $\mathcal{T}_{\epsilon,i}$ contains a neighborhood of $\theta_{i,-}$.

that any such measure has to be equal to $\mu_{\varepsilon,i}$; we conclude that $\mu_{\varepsilon,i}$ is itself an SRB measure. By invariance of $\mu_{\varepsilon,i}$ and since it can be approximated by standard families located at $\mathcal{T}_{\epsilon,i}$, we conclude using (6.16), that $\operatorname{supp}\mu_{\varepsilon,i} \subset \{\theta \in \mathcal{T}_{\epsilon,i}\}$. Moreover, by our construction, it is clear that if $\mathcal{T}_{\epsilon,i} = \mathcal{T}_{\epsilon,j}$, then $\mu_{\varepsilon,i} = \mu_{\varepsilon,j}$.

In order to conclude, we need to check that we have exponential decay of correlations for Hölder observables. To start, let us first assume $A \in \mathcal{C}^3(\{\theta \in \mathcal{T}_{\epsilon,i}\})$ and consider the measurable partition $\{I_\xi\}_{\xi\in\Xi}$ introduced above. Then we can write:

$$\operatorname{Leb}_i(A \cdot B \circ F_\varepsilon^n) = \int_\Xi \nu(d\xi) \int_{I_\xi} A(x, y_\xi) B \circ F_\varepsilon^n(x, y_\xi) dx, \tag{9.12}$$

where ν is the natural factor measure on Ξ. Next, we set $d\hat\nu = \left[\int_{I_\xi} A(x, y_\xi)dx\right] d\nu$ and $\hat A_\xi(x) = A(x, y_\xi)\left[\int_{I_\xi} A(x, y_\xi)dx\right]^{-1}$. In particular, $\int_\Xi \hat\nu(d\xi) = \operatorname{Leb}_i(A)$. Then, by definition, $\ell_\xi = (\mathbb{G}_\xi, \hat A_\xi)$ with $\mathbb{G}_\xi(x) = (x, y_\xi)$, is a standard pair provided $\min A \geq 1$ and $\|A(x,\cdot)\|_{\mathcal{C}^3} \leq C$ for some appropriate constant $C > 0$ (see Section 5.1.1 to recall definitions and notations). Thus we can write

$$\operatorname{Leb}_i(A \cdot B \circ F_\varepsilon^n) = \int_\Xi \hat\nu(d\xi)\mu_{\ell_\xi}(B \circ F_\varepsilon^n) = \int_\Xi \hat\nu(d\xi) F_{\varepsilon *}^n \mu_{\ell_\xi}(B).$$

since $|F_{\varepsilon*}^n\mu_{\ell_\xi}(B) - \mu_{\varepsilon,i}(B)| < C_\# \exp(-\beta c_\varepsilon n)\|B\|_{x,\beta}$, we obtain exponential decay of correlations, provided that A satisfies the additional properties listed above.

Let us now consider the case of a general A. Obviously it suffices to have an estimate for cA, where c is some small constant. But then we can write

$$cA = \{c(A + \|A\|_{L^\infty}) + 1\} - \{c\|A\|_{L^\infty} + 1\}$$

which, for $c \leq (C-1)(2\|A(x,\cdot)\|_{\mathcal{C}^3})^{-1}$, is the difference of two functions both satisfying the hypotheses above. Thus, for all $A \in \mathcal{C}^3$ and $B \in \mathcal{C}^\beta$, we have

$$|\operatorname{Leb}_i(A \cdot B \circ F_\varepsilon^n) - \operatorname{Leb}_i(A)\mu_\varepsilon(B)| \leq C_1\|A\|_{\theta,3}\|B\|_{x,\beta} e^{-\beta c_\varepsilon n}. \tag{9.13}$$

To conclude, let us consider the case $A \in \mathcal{C}^\alpha$, $\alpha < 3$; for arbitrary $\varrho > 0$ let $A_\varrho \in \mathcal{C}^3$ such that $\|A - A_\varrho\|_{\theta,0} \leq \varrho^\alpha\|A\|_{\theta,\alpha}$, and $\|A_\varrho\|_{\theta,3} \leq C_\#\varrho^{-3+\alpha}\|A\|_{\theta,\alpha}$.[40] Then, by equations (9.12), (9.11) and (9.13), we have

$$\begin{aligned}
|\operatorname{Leb}_i(A \cdot B \circ F_\varepsilon^n) - \operatorname{Leb}_i(A)\mu_\varepsilon(B)| \leq & |\operatorname{Leb}_i(A_\varrho \cdot B_\varrho \circ F_\varepsilon^n) - \operatorname{Leb}_i(A_\varrho)\mu_\varepsilon(B_\varrho)| \\
& + C_\#\varrho^\alpha\|A\|_{\theta,\alpha}\|B\|_{x,\beta} \\
\leq & (C_\#\varrho^{-3+\alpha}e^{-\beta c_\varepsilon n} + C_\#\varrho^\alpha)\|A\|_{\theta,\beta}\|B\|_{x,\beta}.
\end{aligned}$$

Optimizing ϱ, as a function of n, we obtain $\varrho = e^{-\beta c_\varepsilon n/3}$, which yields the wanted result (absorbing the factor 3 in the constants C_2 and C_3). □

Remark 9.11. *Once again (see Remark 2.9) Theorem 9.9 is stated for Lebesgue measure just for simplicity. In fact it holds for any initial measure that can be obtained as weak limit of standard families located at $\mathcal{T}_{\epsilon,i}$. In particular, we have exponential decay of correlations for initial conditions distributed according to the SRB measures $\mu_{\varepsilon,i}$ themselves. Only, in this case our estimate (9.10) for the decay rate is quite possibly not optimal (e.g. Remark 2.9 when when $n_{\mathrm{Z},i} > 1$ and the discussion at the end of Subsection 3.1 for when $n_{\mathrm{Z},i} = 1$).*

[40] Such approximate functions can be obtained by standard mollification.

10. Coupling: Proofs

This is the most probabilistic part of the paper: it is then natural to adopt a more probabilistic notation. As we have painstakingly explained on which spaces the various relevant random variables live and how their laws are defined, from now on we will simply use $\mathbb{P}$ and $\mathbb{E}$ for designating, respectively, their probability and expectation, unless some ambiguity might arise.

We start with an easy corollary of Lemma 7.2 and Lemma 9.1 with $N = N_{\mathrm{S}}$ which ensures that a $\Delta\varepsilon$-matched coupling which is supported on $\mathbb{H}$ will geometrically decrease its Wasserstein distance after time N_{S} except in an event of exponentially small probability.

Corollary 10.1. *For any $\bar{\Delta} > 0$ there exists $\bar{\varepsilon} > 0$ so that the following holds. For any $\varepsilon \in (0, \bar{\varepsilon})$, $\Delta \in (0, \bar{\Delta})$ and $\underline{\ell}$ a $\Delta\varepsilon$-matched standard couple so that $\theta^*_{\underline{\ell}^0} \in \mathbb{H}$; let $\underline{\mathfrak{L}}^{\mathrm{C}}_{N_{\mathrm{S}}}$ be the family obtained by applying Lemma 9.1 with $N = N_{\mathrm{S}}$ to the couple $\underline{\ell}$. Then:*

$$\mathbb{P}\big(\theta^*_{\underline{\ell}^{\mathrm{C},0}_{N_{\mathrm{S}}}(\cdot)} \in \mathbb{H}, \underline{\ell}^{\mathrm{C}}_{N_{\mathrm{S}}}(\cdot) \textit{ is } \mathrm{d}_W(\underline{\ell}^0, \underline{\ell}^1)\exp(-T_{\mathrm{S}}/2)\textit{-matched}\big) \geq 1 - C_\# \exp(-c_\#\varepsilon^{-1}).$$

Proof. Let us apply Lemma 7.2 to $\underline{\ell}^{\mathrm{C},0}_0$; by Lemma 9.1(a) with $N = N_{\mathrm{S}}$ we then obtain:

$$\mathbb{P}(\{\theta_{N_{\mathrm{S}}} \in \hat{\mathbb{H}}, \zeta_{N_{\mathrm{S}}} \leq -9T_{\mathrm{S}}/16\}) \geq 1 - C_\# \exp(-c_\#\varepsilon^{-1}).$$

Since $\underline{\mathfrak{L}}^{\mathrm{C},0}_{N_{\mathrm{S}}}$ is a N_{S}-pushforward of $\underline{\ell}^{\mathrm{C},0}_0$, we can define the subset

$$\mathcal{A}'^{\mathrm{C}}_{N_{\mathrm{S}}} = \boldsymbol{\alpha}_{N_{\mathrm{S}}}(\{\theta_{N_{\mathrm{S}}} \in \hat{H}_k, \zeta_{N_{\mathrm{S}}} < -9T_{\mathrm{S}}/16\}).$$

Standard distortion estimates then imply that for any $\alpha \in \mathcal{A}'^{\mathrm{C}}_{N_{\mathrm{S}}}$ and $p, q \in U_\alpha$, we have $|\theta_{N_{\mathrm{S}}}(q) - \theta_{N_{\mathrm{S}}}(p)| \leq C_\#\varepsilon$ and $\zeta_{N_{\mathrm{S}}}(q) \leq \zeta_{N_{\mathrm{S}}}(p) + C_\#\varepsilon$. Lemma 9.1(b), with $N = N_{\mathrm{S}}$, (8.3) and recalling that T_{S} is assumed to be large (in particular we can assume $T_{\mathrm{S}} > 1$), concludes the proof of the corollary. □

10.1. **Proof of Lemma 9.3.**

We will define the sequence $\underline{\mathfrak{L}}_{[k]}$ and the random variables $U_{[k]}$ by an inductive construction in which we also introduce an auxiliary sequence of random variables $\mathcal{G}_{[k]} : \mathcal{A}_{[k]} \to \mathbb{R}$. In particular, such random variables will satisfy the following assumptions: let $\Delta_k = \exp(-kT_{\mathrm{S}}/4)$

- **(i)** if $U_{[k]}(\alpha) = \infty$, the couple $\underline{\ell}_{[k]}(\alpha)$ is $\Delta_k \exp(-\mathcal{G}_{[k]}(\alpha)T_{\mathrm{S}})\varepsilon^{1+\tau/2}$-matched;
- **(ii)** for any $l < k$, we have $\nu_{[k]}(U_{[k]} = l) = \nu_{[k-1]}(U_{[k-1]} = l)$ and $\underline{\mathfrak{L}}_{[k]}|\{U_{[k]} = l\} \in \big[F^{N_{\mathrm{S}}}_\varepsilon \underline{\mathfrak{L}}_{[k-1]}|\{U_{[k-1]} = l\}\big]$; finally, $\underline{\mathfrak{L}}_{[l]}|\{U_{[l]} = l-1\}$ is a coupling of lN_{S}-prestandard families.
- **(iii)** $\mathcal{G}_{[k]} \geq -2\Psi$, for all $k \in \mathbb{N}$, where Ψ is defined in (4.8). In addition, $\mathcal{G}_{[k]} \geq 0$ for all $k \leq C_\#\tau \log \varepsilon^{-1}$.

Properties (i) and (iii) above trivially imply item (a) of our statement, provided ε is small enough, while (ii) corresponds exactly to item (b). Intuitively, the variable $\mathcal{G}_{[k]}$ is a measure of the closeness of a couple of standard pairs, at iterate k, compared with our minimal expectation expressed by Δ_k. If $\mathcal{G}_{[k]}$ becomes negative, then it means that the couple has failed to get as close as we like in such a drastic manner that we give up on it and break it up. Let us specify our inductive construction.

For the base step, we define $\underline{\mathfrak{L}}_{[0]} = \underline{\ell}$, $\mathcal{G}_{[0]} = \frac{\bar{k}}{2T_{\mathrm{S}}} \log \varepsilon^{-1}$, $\bar{k} \leq \tau/2 + \frac{T_{\mathrm{S}}}{\log \varepsilon^{-1}}$, and $U_{[0]} = \infty$.

Next, we assume that $\underline{\mathfrak{L}}_{[l]}$, $U_{[l]}$ and $\mathcal{G}_{[l]}$ are already defined for $0 \leq l \leq k$ and proceed to define $\underline{\mathfrak{L}}_{[k+1]}$, $U_{[k+1]}$ and $\mathcal{G}_{[k+1]}$. For each $\alpha \in \mathcal{A}_{[k]}$ we will define a family $\underline{\mathfrak{L}}_{N_{\mathrm{S}}}(\alpha) \in \big[F^{N_{\mathrm{S}}}_\varepsilon \underline{\ell}_{[k]}(\alpha)\big]$ and for each $\alpha' \in \mathcal{A}(\alpha)$ we will define $\mathcal{G}_{[k+1]}(\alpha')$

and $U_{[k+1]}(\alpha')$. We then define $\underline{\mathfrak{L}}_{[k+1]} \in \left[F_\varepsilon^{N_{\rm S}}\underline{\mathfrak{L}}_{[k]}\right]$ by considering the convex combination

$$\underline{\mathfrak{L}}_{[k+1]} = \sum_{\alpha\in\mathcal{A}_{[k]}} \nu_{[k]}(\{\alpha\})\underline{\mathfrak{L}}_{N_{\rm S}}(\alpha).$$

The random variables $\mathcal{G}_{[k+1]}$ and $U_{[k+1]}$ are thus naturally defined on $\mathcal{A}_{[k+1]}$.[41]

We proceed to define $\underline{\mathfrak{L}}_{N_{\rm S}}(\alpha)$ for $\alpha \in \mathcal{A}_{[k]}$. There are several possibilities:

- $U_{[k]}(\alpha) = \infty$ and $\mathcal{G}_{[k]}(\alpha) \geq 0$: by inductive assumption **(i)**, the couple $\underline{\ell}_{[k]}(\alpha)$ is $\Delta_k\varepsilon^{1+\tau/2}$-matched; we can thus apply Lemma 9.1 with $N = N_{\rm S}$ to $\underline{\ell}_{[k]}(\alpha)$ with $\Delta = \Delta_k\varepsilon^{\tau/2}$ and define $\underline{\mathfrak{L}}_{N_{\rm S}}(\alpha) = m_{\rm C}\underline{\mathfrak{L}}^{\rm C}_{N_{\rm S}}(\alpha)+(1-m_{\rm C})\underline{\mathfrak{L}}^{\rm U}_{N_{\rm S}}(\alpha)$.

  If $\alpha' \in \mathcal{A}^{\rm U}(\alpha)$, we let $U_{[k+1]}(\alpha') = k$ and $\mathcal{G}_{[k+1]}(\alpha') = \mathcal{G}_{[k]}(\alpha) + 1/4$. Observe, *en passant*, that by Lemma 9.1(c) with $N = N_{\rm S}$ the couple $\underline{\ell}(\alpha')$ is $N_{\rm S} + C_\#(k + \log\varepsilon^{-1})$-prestandard. Choosing ε sufficiently small we can ensure that $\underline{\ell}(\alpha')$ is indeed $(k + 1)N_{\rm S}$-prestandard.

  If, on the other hand $\alpha' \in \mathcal{A}^{\rm C}(\alpha)$, we let $U_{[k+1]}(\alpha') = \infty$ and define $\mathcal{G}_{[k+1]}$ as:

$$\mathcal{G}_{[k+1]}(\alpha') = \begin{cases}\mathcal{G}_{[k]}(\alpha) + \frac14 & \text{if } \underline{\ell}_{[k+1]}(\alpha') \text{ is } \Delta_k(\alpha,\tau)\text{-matched}\\ \mathcal{G}_{[k]}(\alpha) - 2\Psi & \text{otherwise;}\end{cases}$$

  where $\Delta_k(\alpha,\tau) = \Delta_k\exp(-(\mathcal{G}_{[k]}(\alpha) + \frac12)T_{\rm S})\varepsilon^{1+\frac{\tau}{2}}$.

Remark 10.2. *Note that, by our assumptions, $\Delta_0(\alpha,\tau) \geq \varepsilon^{1+\frac{3\tau}{4}}$. Thus the second option above can only occur if $k \geq C_\#\tau\log\varepsilon^{-1}$.*

- $U_{[k]}(\alpha) = \infty$ and $\mathcal{G}_{[k]}(\alpha) < 0$: we declare the couple to break up and let $\underline{\mathfrak{L}}_{N_{\rm S}}(\alpha)$ be an arbitrary $N_{\rm S}$-pushforwards of $\underline{\ell}_{[k]}(\alpha)$. Also, for any $\alpha' \in \mathcal{A}(\alpha)$ we let $U_{[k+1]}(\alpha') = k$ and $\mathcal{G}_{[k+1]}(\alpha') = \mathcal{G}_{[k]}(\alpha) + 1/4$.
- if $U_{[k]}(\alpha) < \infty$, we let $\underline{\mathfrak{L}}_{N_{\rm S}}(\alpha)$ be an arbitrary $N_{\rm S}$-pushforward of $\underline{\ell}_{[k]}(\alpha)$. Also, for any $\alpha' \in \mathcal{A}(\alpha)$ we let $U_{[k+1]}(\alpha') = U_{[k]}(\alpha)$ and $\mathcal{G}_{[k+1]}(\alpha') = \mathcal{G}_{[k]}(\alpha) + 1/4$.

Inductive assumptions **(i)**, **(ii)** and **(iii)** then immediately follow from the above definitions and by Remark 10.2 using (4.8). As noticed earlier, they imply items (a) and (b). We are now left to show item (c): in order to do so, first observe that by definition and Corollary 10.1 we have

$$\mathbb{P}(\mathcal{G}_{[k+1]} - \mathcal{G}_{[k]} = -2\Psi\big|\theta^*_{\underline{\ell}^0_{[k]}} \in \mathbb{H}) \leq C_\#\exp(-c_\#\varepsilon^{-1}). \tag{10.1}$$

We now use the above inequality to prove a preliminary result:

Sub-lemma 10.3. *For any $\hat\gamma > 0$, there exists $\vartheta \in (0,1)$ such that*

$$\mathbb{P}\left(\inf_{0\leq j\leq\mathcal{K}_{\rm A}}\mathcal{G}_{[k+j]} < 0\right) \leq \hat\gamma\vartheta^{k/\log\varepsilon^{-1}},$$

where recall $\mathcal{K}_{\rm A} = \lfloor\mathcal{R}_{\rm A}\log\varepsilon^{-1}\rfloor$ with $\mathcal{R}_{\rm A}$ defined in Lemma 7.5.

Proof. Let us fix $p \in \mathbb{N}$ sufficiently large to be specified later; for $j \geq 0$, we define auxiliary random variables:

$$X_{[j]} = \begin{cases}1 & \text{if } \mathcal{G}_{[(j+1)p\mathcal{K}_{\rm A}]} \geq \mathcal{G}_{[jp\mathcal{K}_{\rm A}]} + \mathcal{K}_{\rm A}\\ -1 & \text{otherwise.}\end{cases}$$

[41] Note that there exists a natural measure-preserving immersion $\mathbf{i} : \mathcal{A}_{[\mathbf{k+1}]} \to \mathcal{A}_{[\mathbf{k}]}$, thus one can always see $U_{[k]}$ as a random variable on $\mathcal{A}_{[k+1]}$ and similarly for the other random variables. It is thus possible to view all the relevant random variables on the same natural probability space (given by the last time at which we are interested). We will use this implicitly in the following.

Then we claim that if p is sufficiently large, there exists $\beta' < \beta$ (where β is defined in Lemma 7.5) so that:[42]

$$\mathbb{P}(X_{[j+1]} = -1 \big| \boldsymbol{\alpha}_{[(j+1)p\mathcal{K}_{\mathrm{A}}]}) \leq C_{\#}\varepsilon^{\beta'}, \tag{10.2}$$

provided ε is small enough.

Remark 10.4. *Observe that, provided that $p > 4$, if $U_{[jp\mathcal{K}_{\mathrm{A}}]} < \infty$ (i.e. a breakup already happened earlier than step $jp\mathcal{K}_{\mathrm{A}}$), then we automatically have $X_{[j]} = 1$ and thus (10.2) trivially holds. This is indeed the reason to define $\mathcal{G}_{[jp\mathcal{K}_{\mathrm{A}}+1]} = \mathcal{G}_{[jp\mathcal{K}_{\mathrm{A}}]} + 1/4$ after a breakup.*

Estimate (10.2) suffices to conclude the proof of our sub-lemma: in fact observe that conditioning on the random variable $\boldsymbol{\alpha}_{[(j+1)p\mathcal{K}_{\mathrm{A}}]}$ (defined in Remark 5.5) is finer than conditioning on $X_{[0]} \cdots X_{[j]}$. By definition of conditional probability it follows

$$\mathbb{P}(X_{[j+1]} = -1 \big| X_{[0]} \cdots X_{[j]}) \leq C_{\#}\varepsilon^{\beta'}.$$

Observe that, by construction, for any $0 \leq s \leq p\mathcal{K}_{\mathrm{A}}$, $\mathcal{G}_{[jp\mathcal{K}_{\mathrm{A}}+s]} - \mathcal{G}_{[jp\mathcal{K}_{\mathrm{A}}]} \geq -2s\Psi$. Hence, we conclude that

$$\mathcal{G}_{[kp\mathcal{K}_{\mathrm{A}}]} - \mathcal{G}_{[0]} \geq (1/2 - p\Psi)\mathcal{K}_{\mathrm{A}} k + (1/2 + p\Psi)\mathcal{K}_{\mathrm{A}} \sum_{l=0}^{k-1} X_{[l]}. \tag{10.3}$$

Choose $c \in (0, 1 - C_{\#}\varepsilon^{\beta'})$ so that $(1/2 - p\Psi) + c(1/2 + p\Psi) > 1/2$. Thus Lemmata A.1 and A.2 imply that there exists $\vartheta \in (0,1)$ and $a > 0$ such that

$$\mathbb{P}\left(\sum_{l=0}^{k-1} X_{[l]} \leq ck - a\right) \leq \hat{\gamma}\vartheta^k.$$

Thus, provided that we choose $\bar{k}$ sufficiently large (relative to a), (10.3) implies that

$$\mathbb{P}(\mathcal{G}_{[kp\mathcal{K}_{\mathrm{A}}]} < k\mathcal{K}_{\mathrm{A}}/2) \leq \hat{\gamma}\vartheta^k$$

which, by Remark 10.2, would conclude the proof of our sub-lemma.

To really conclude, we are left with the proof of (10.2). Notice that

$$\mathbb{P}(X_{[j+1]} = 1 \big| \boldsymbol{\alpha}_{[(j+1)p\mathcal{K}_{\mathrm{A}}]}) \geq \mathbb{P}(X_{[j+1]} = 1 \big| \boldsymbol{\alpha}_{[(j+1)p\mathcal{K}_{\mathrm{A}}]}; A_{\mathbb{H}})\mathbb{P}(A_{\mathbb{H}}),$$

where we have introduced the event $A_{\mathbb{H}} = \{\theta^{*,0}_{[(j+1)p\mathcal{K}_{\mathrm{A}}+r]} \in \mathbb{H}\ \forall r : \mathcal{K}_{\mathrm{A}} \leq r \leq p\mathcal{K}_{\mathrm{A}}\}$, where $\theta^{*,0}_n$ denotes the average θ with respect to the marginal of the the first component of the standard coupling. By Lemma 7.5 we have

$$\mathbb{P}(A_{\mathbb{H}}) \geq 1 - (p-1)\mathcal{K}_{\mathrm{A}}\varepsilon^{\beta}.$$

On the other hand, by iterating $(p-1)\mathcal{K}_{\mathrm{A}}$ times (10.1) we obtain that

$$\mathbb{P}\left(\mathcal{G}_{[(j+1)p\mathcal{K}_{\mathrm{A}}]} - \mathcal{G}_{[(jp+1)\mathcal{K}_{\mathrm{A}}]} \geq \frac{(p-1)}{4}\mathcal{K}_{\mathrm{A}} \middle| \boldsymbol{\alpha}_{[(j+1)p\mathcal{K}_{\mathrm{A}}]}; A_{\mathbb{H}}\right) \geq 1 - \frac{(p-1)\mathcal{K}_{\mathrm{A}}}{\exp(c_{\#}\varepsilon^{-1})}.$$

Thus, with overwhelming probability,

$$\mathcal{G}_{[(j+1)p\mathcal{K}_{\mathrm{A}}]} - \mathcal{G}_{[jp\mathcal{K}_{\mathrm{A}}]} \geq \frac{(p-1)}{4}\mathcal{K}_{\mathrm{A}} - 2\Psi\mathcal{K}_{\mathrm{A}} \geq \mathcal{K}_{\mathrm{A}},$$

provided $p > 4(1+2\Psi)+1$. That is to say that $X_{[j+1]} = 1$, which proves (10.2). □

[42] The conditioning means simply that we specify the standard pair to which the process belongs at the iteration step $(j+1)p\mathcal{K}_{\mathrm{A}} N_{\mathrm{S}}$.

We can now prove item (c): by our inductive construction, Lemma 9.1(a), with $N = N_{\mathrm{S}}$, and Sub-lemma 10.3 we have:

$$\begin{aligned}\mathbb{P}(U_{[k+1]} = \infty) &= \mathbb{P}(U_{[k]} = \infty, \mathcal{G}_{[k]} \geq 0) m_{\mathrm{C}}(\Delta_k \varepsilon^{\tau/2}) \\ &\geq \mathbb{P}(U_{[k-\mathcal{K}_{\mathrm{A}}]} = \infty, \inf_{j \leq \mathcal{K}_{\mathrm{A}}} \mathcal{G}_{[k-j]} \geq 0) \prod_{j=0}^{\mathcal{K}_{\mathrm{A}}-1} m_{\mathrm{C}}(\Delta_{k-j}\varepsilon^{\tau/2}) \\ &\geq \mathbb{P}(U_{[k-\mathcal{K}_{\mathrm{A}}]} = \infty) \prod_{j=0}^{\mathcal{K}_{\mathrm{A}}-1} m_{\mathrm{C}}(\Delta_{k-j}\varepsilon^{\tau/2}) - \hat{\gamma}\vartheta^{k/\log \varepsilon^{-1}}.\end{aligned}$$

The above inequality implies, for ε small enough,

$$\mathbb{P}(U_{[k]} = \infty) \geq \prod_{j=0}^{k-1} m_{\mathrm{C}}(\Delta_j \varepsilon^{\tau/2}) - C_\# \hat{\gamma} \geq \exp(-c_\# \hat{\gamma}).$$

Finally, for $j > k$, again by our construction, Lemma 9.1 with $N = N_{\mathrm{S}}$ and Sub-Lemma 10.3,

$$\begin{aligned}\mathbb{P}(U_{[k]} = \infty) - \mathbb{P}(U_{[j]} = \infty) &\leq \left(1 - \prod_{l=k}^{j-1} m_{\mathrm{C}}(\Delta_l \varepsilon^{\tau/2})\right) + C_\# \hat{\gamma}\vartheta^{k/\log \varepsilon^{-1}} \\ &\leq C_\# \exp(-c_\# k)\varepsilon^{\tau} + C_\# \hat{\gamma}\vartheta^{k/\log \varepsilon^{-1}}\end{aligned}$$

provided we choose ε to be small enough. The two inequalities above prove (9.6) and (9.7) and conclude the proof of our Lemma. $\square$

10.2. **Proof of Lemma 9.6.** First, we prove the following

Sub-lemma 10.5. *Let ℓ be a standard pair located at $\mathcal{T}_{\epsilon,i}$; there exists $p'_{\mathrm{B}} > 0$ so that:*

$$\mu_\ell(\theta_{\mathcal{K}_{\mathrm{A}} N_{\mathrm{S}}} \in \hat{H}_i) > p'_{\mathrm{B}}, \tag{10.4}$$

where, recall, $\mathcal{K}_{\mathrm{A}} = \lfloor \mathcal{R}_{\mathrm{A}} \log \varepsilon^{-1} \rfloor$ and $\mathcal{R}_{\mathrm{A}}$ is the constant obtained in Lemma 7.5. Moreover, if $n_{\mathrm{Z},i} = 1$, p'_{B} can be chosen to be uniform in ε; otherwise $p'_{\mathrm{B}} = C_\# \exp(-c_\# \varepsilon^{-1})$.

Proof. If $n_{\mathrm{Z},i} = 1$, then $\hat{\mathbb{H}} \cap \mathcal{T}_{\epsilon,i} = \hat{H}_i$ and the statement immediately follows by Lemma 7.5 and forward invariance of trapping sets (6.16), which proves (10.4) for any $p'_{\mathrm{B}} < 1 - \varepsilon^{\beta}$.

Assume now that $n_{\mathrm{Z}} > 1$: Lemma 6.13(a) guarantees the existence of an ϵ-admissible $(\theta^*_\ell, \theta_{i,-})$-path of length bounded by $T_{\mathcal{T}}$; Theorem 6.3 then implies that

$$\mu_\ell(\theta_{\lfloor T_{\mathcal{T}} \varepsilon^{-1} \rfloor} \in \hat{H}_i) > \exp(-c_\# \varepsilon^{-1}).$$

We can then conclude by using Corollary 7.3, which proves (10.4) for $p'_{\mathrm{B}} = e^{-c_\# \varepsilon^{-1}}$. $\square$

Let us now conclude the proof of Lemma 9.6: let $C > 0$ be the constant given by Lemma 7.4 and let $J \subset \mathbb{T}^1$ be the interval $B(\theta_{i,-}, C\sqrt{\varepsilon})$. Subdivide J into $\lfloor \varepsilon^{-1/2} \rfloor$ subintervals $\{I_j\}$ of equal length $C_\# \varepsilon$. By Theorem 6.8 we can choose $T > 0$ sufficiently large such that for any standard pair ℓ located at J and for any j:

$$\mu_\ell(\theta_{\lfloor T\varepsilon^{-1} \rfloor} \in I_j) > p''_{\mathrm{B}} \varepsilon^{1/2}.$$

where $p''_{\mathrm{B}} > 0$ is uniform in ε and independent of ℓ. Thus, combining the above observation with Lemma 7.4, we conclude that if ℓ is a standard pair with $\theta^*_\ell \in H_i$, and we let $\mathcal{K} = \lfloor(\mathcal{R}_\mathrm{D} + 1)\log \varepsilon^{-1}\rfloor$, where $\mathcal{R}_\mathrm{D}$ is the constant found in Lemma 7.4; then, for all j,

$$\mu_\ell(\theta_{\mathcal{K}N_\mathrm{S}} \in I_j) > \frac{1}{2}p''_\mathrm{B}\varepsilon^{1/2}. \tag{10.5}$$

Hence, together with Sub-Lemma 10.5, we proved that if ℓ^0 and ℓ^1 are any two standard pairs located at the same trapping set $\mathcal{T}_{\epsilon,i}$, the probability that their $(\mathcal{K}_\mathrm{A} + \mathcal{K})N_\mathrm{S}$-image have θ-coordinates which are $C_\#\varepsilon$-close is at least $\frac{1}{2}p'_\mathrm{B}p''_\mathrm{B}$.

We now need to find pairs that are actually $\Delta\varepsilon$-matched for some $\Delta > 0$; this task can be accomplished by the following argument. Let $I \subset \mathbb{T}^1$ be a fixed interval of length δ. Since our maps are uniformly expanding in the x direction, there exist $M > 0$ and $p \in (0, 1)$ so that, given any standard pair ℓ, we can construct an M-pushforward of ℓ so that one of the standard pairs lies above the interval I and this standard pair has probability larger than p. Moreover, by Remark 5.8, we can assume that this ℓ has a flat density, decreasing p by a factor $2/3$; the leftover pairs will be $\mathcal{O}(1)$-prestandard. We can then construct the canonical coupling of all pairs which lie above I and the independent coupling of all other pairs.

We thus proved that if $\mathcal{R} > \mathcal{R}_\mathrm{A} + \mathcal{R}_\mathrm{D} + 1$, then there exists a coupling

$$\left[F_\varepsilon^{\lfloor\mathcal{R}\log\varepsilon^{-1}\rfloor N_\mathrm{S}}\underline{\ell}\right] \ni m'_\mathrm{B}\widetilde{\underline{\mathfrak{L}}}^\mathrm{B} + (1 - m'_\mathrm{B})\widetilde{\underline{\mathfrak{L}}}^\mathrm{R},$$

where $m'_\mathrm{B} = \frac{1}{2}pp'_\mathrm{B}p''_\mathrm{B}$ and $\widetilde{\underline{\mathfrak{L}}}^\mathrm{B}$ is a $\Delta\varepsilon$-matched standard coupling whose components are supported on $\mathbb{T} \times H_i$ and $\widetilde{\underline{\mathfrak{L}}}^\mathrm{R}$ is a $\mathcal{O}(1)$-prestandard coupling. In order to conclude the proof of our statement we need to obtain couplings which are $\varepsilon^{1+\tau}$-matched: to do so it suffices to apply iteratively Lemma 9.1 with $N = N_\mathrm{S}$ to pairs in $\widetilde{\underline{\mathfrak{L}}}^\mathrm{B}$. Using Corollary 10.1 (as we did in the proof of Sub-lemma 10.3) we conclude that there exists C' so that a substantial portion of the mass of a $(C'\tau\log\varepsilon^{-1})N_\mathrm{S}$-pushforward of $\widetilde{\underline{\mathfrak{L}}}^\mathrm{B}$ will be $\varepsilon^{1+\tau}$-matched and the leftover pairs will be $\mathcal{O}(\tau\log\varepsilon^{-1})$-prestandard, which concludes our proof choosing $\mathcal{R}_\mathrm{B} = \mathcal{R}_\mathrm{A} + \mathcal{R}_\mathrm{D} + 1 + C'\tau$. □

11. Conclusions and open problems

In this work we have discussed the case in which the dynamics of the fast variable is given by a one dimensional expanding map. In this setting we proved exponential decay of correlation for an open set of partially hyperbolic endomorphisms of the two-torus $\mathbb{T}^2$. To keep the exposition as terse as possible, we did not investigate in detail the adiabatic, metastable, regime. This can be done similarly to [21, 30] and is postponed to future work.

Another natural issue, already pointed out in Section 2, is the necessity of hypothesis (A2). In our scheme of proof it is certainly needed. Nevertheless, we provided an example in Section 3.4 that does not satisfy (A2) and yet numerical computations seems to show that it behaves similarly to the examples for which (A2) is satisfied [45]. This suggests that our understanding of the possible mechanisms of convergence to equilibrium is partial at best, and that further thought is much needed.

Next, observe that assumption (A1) is substantial: the set of ω such that $\{\theta : \bar{\omega}(\theta) = 0\} = \emptyset$ is open. If $\bar{\omega}$ has no zeros, then the averaged motion is a rotation, with no sinks or sources; the main mechanism to establish a coupling argument would then be the diffusion centered on the rotation. Note however that this would require a time scale ε^{-2} to bring any two standard pairs close enough to couple them, [16]. This situation is of considerable interest in non-Equilibrium Statistical Mechanics when the dynamics is Hamiltonian and the slow variables are the energies of nearby, weakly interacting, systems, see [19]. In this case we conjecture that, generically, the system should be mixing and the correlations should decay exponentially with rate which would be, at best, ε^2. However, to prove such a result stands as a substantial challenge in the field.

Finally, it would be very interesting to prove analogous results for the case in which the fast variable evolves according to a more general hyperbolic system and when the slow variable is higher dimensional. The first generalization could prove rather difficult when trying to extend, e.g., the needed results of our paper [9] to the case of flows or systems with discontinuities. The second does not pose any particular problem as far as the results in [9] are concerned. The difficulties come instead from the fact that in higher dimension a generic dynamics has many different types of ω-limit sets (not just sinks or the whole space, as it is in one dimension) and these possibilities give rise to situations to which the ideas put forward in the present paper may not easily apply.

Appendix A. Random walks

We start by recalling a well known fact about one dimensional random walks (it can be obtained, e.g., from Cramer's Theorem).

Lemma A.1. *Let $\xi_k \in \{-1,1\}$ be a sequence of i.i.d. random variables with distribution $\mathbb{P}(\xi_i = 1) = p$ for $p \in (0,1)$. Let $\boldsymbol{\Xi}_0 = 0$ and for $n > 0$, define: $\boldsymbol{\Xi}_n = \sum_{j=1}^n \xi_j$. For any $c < 2p-1$ there exist $\vartheta, \varrho \in (0,1)$ such that, for any $k \in \mathbb{N}$ and $a \in \mathbb{R}$:*

$$\mathbb{P}(\boldsymbol{\Xi}_k \leq kc - a) \leq \varrho^a \vartheta^k.$$

Next, we introduce an useful comparison argument:

Lemma A.2. *Let $\xi_k \in \{-1,1\}$ be a sequence of independent random variables and let $\eta_k \in \{-1,0,1\}$ be a random process such that*

$$\mathbb{P}(\eta_{k+1} = 1 | \eta_1 \cdots \eta_k) \geq \mathbb{P}(\xi_{k+1} = 1).$$

For $n > 0$ define the random variables

$$\boldsymbol{\Xi}_n = \sum_{j=1}^n \xi_j \qquad\qquad \mathbf{H}_n = \sum_{j=1}^n \eta_j$$

where $N > 0$ is some fixed natural number (if $n = 0$ we let them all equal to 0); then for each $n \in \mathbb{N}$ and $L \in \mathbb{Z}$:

$$\mathbb{P}(\mathbf{H}_k \leq L) \leq \mathbb{P}(\boldsymbol{\Xi}_k \leq L). \tag{A.1}$$

In particular, if $\tau_{\boldsymbol{\Xi}}$ is the hitting time $\tau = \inf\{k : \boldsymbol{\Xi}_k \geq L\}$ and $\tau_{\mathbf{H}} = \inf\{k : \mathbf{H}_k \geq L\}$ we have, for any $s > 0$:

$$\mathbb{P}(\tau_{\mathbf{H}} > s) \leq \mathbb{P}(\tau_{\boldsymbol{\Xi}} > s)$$

Proof (see [17, Proposition 2.4]*).* The proof amounts to design a suitable coupling (ξ_k^*, η_k^*) of the random variables ξ_k and η_k. Let us introduce an auxiliary sequence U_k of independent random variables uniformly distributed on $[0,1]$ and define the random variables

$$\xi_k^* = \begin{cases} +1 & \text{if } U_k < \mathbb{P}(\xi_k \geq 1) \\ -1 & \text{otherwise} \end{cases}$$

and

$$\eta_k^* = \begin{cases} +1 & \text{if } U_k < \mathbb{P}\left(\eta_k = 1 | \eta_1 = \eta_1^*, \cdots, \eta_{k-1} = \eta_{k-1}^*\right) \\ -1 & \text{if } U_k \geq 1 - \mathbb{P}\left(\eta_k = -1 | \eta_1 = \eta_1^*, \cdots, \eta_{k-1} = \eta_{k-1}^*\right) \\ 0 & \text{otherwise.} \end{cases}$$

We then define

$$\Xi_n^* = \sum_{j=1}^n \xi_j^* \qquad\qquad \mathbf{H}_n^* = \sum_{j=1}^n \eta_j^*$$

Clearly ξ_k^* (resp. η_k^*) has the same distribution of ξ_k (resp. η_k) and consequently Ξ_k^* (resp. $\mathbf{H}_k^*$) has the same distribution of Ξ_k (resp. $\mathbf{H}_k$). Moreover, $\xi_k^* \leq \eta_k^*$ by design which in turn implies that $\Xi_k^* \leq \mathbf{H}_k^*$. This concludes the proof of our lemma. □

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(2):351–398, 2000.

[2] J. F. Alves, S. Luzzatto, and V. Pinheiro. Markov structures and decay of correlations for non-uniformly expanding dynamical systems. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 22(6):817–839, 2005.

[3] C. Bonatti and M. Viana. SRB measures for partially hyperbolic systems whose central direction is mostly contracting. *Israel J. Math.*, 115:157–193, 2000.

[4] N. Chernov and R. Markarian. *Chaotic billiards*, volume 127 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2006.

[5] N. I. Chernov. Markov approximations and decay of correlations for Anosov flows. *Ann. of Math. (2)*, 147(2):269–324, 1998.

[6] F. H. Clarke, Y. S. Ledyaev, R. J. Stern, and P. R. Wolenski. *Nonsmooth analysis and control theory*, volume 178 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1998.

[7] A. A. de Castro Júnior. Backward inducing and exponential decay of correlations for partially hyperbolic attractors. *Israel J. Math.*, 130:29–75, 2002.

[8] A. A. de Castro Júnior. Fast mixing for attractors with a mostly contracting central direction. *Ergodic Theory Dynam. Systems*, 24(1):17–44, 2004.

[9] J. De Simoi and C. Liverani. Limit theorems for fast-slow partially hyperbolic systems. Preprint arXiv:1408.5453.

[10] J.-D. Deuschel and D. W. Stroock. *Large deviations*, volume 137 of *Pure and Applied Mathematics*. Academic Press, Inc., Boston, MA, 1989.

[11] D. Dolgopyat. On decay of correlations in Anosov flows. *Ann. of Math. (2)*, 147(2):357–390, 1998.

[12] D. Dolgopyat. On dynamics of mostly contracting diffeomorphisms. *Comm. Math. Phys.*, 213(1):181–201, 2000.

[13] D. Dolgopyat. On mixing properties of compact group extensions of hyperbolic systems. *Israel J. Math.*, 130:157–205, 2002.

[14] D. Dolgopyat. Limit theorems for partially hyperbolic systems. *Trans. Amer. Math. Soc.*, 356(4):1637–1689 (electronic), 2004.

[15] D. Dolgopyat. On differentiability of SRB states for partially hyperbolic systems. *Invent. Math.*, 155(2):389–449, 2004.

[16] D. Dolgopyat. Averaging and invariant measures. *Mosc. Math. J.*, 5(3):537–576, 742, 2005.

[17] D. Dolgopyat. Bouncing balls in non-linear potentials. *Discrete Contin. Dyn. Syst.*, 22(1-2):165–182, 2008.

[18] D. Dolgopyat. Repulsion from resonances. *Mém. Soc. Math. Fr. (N.S.)*, (128):vi+119, 2012.

[19] D. Dolgopyat and C. Liverani. Energy transfer in a fast-slow Hamiltonian system. *Comm. Math. Phys.*, 308(1):201–225, 2011.

[20] M. Field, I. Melbourne, and A. Török. Stability of mixing and rapid mixing for hyperbolic flows. *Ann. of Math. (2)*, 166(1):269–291, 2007.

[21] M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Heidelberg, third edition, 2012. Translated from the 1979 Russian original by Joseph Szücs.

[22] S. Gouëzel. Decay of correlations for nonuniformly expanding systems. *Bull. Soc. Math. France*, 134(1):1–31, 2006.

[23] S. Gouëzel and C. Liverani. Banach spaces adapted to Anosov systems. *Ergodic Theory and Dynamical Systems*, 26(1):189–217, 2006.

[24] M. Grayson, C. Pugh, and M. Shub. Stably ergodic diffeomorphisms. *Ann. of Math. (2)*, 140(2):295–329, 1994.

[25] O. Jenkinson. Ergodic optimization. *Discrete Contin. Dyn. Syst.*, 15(1):197–224, 2006.

[26] O. Jenkinson and I. D. Morris. Lyapunov optimizing measures for C^1 expanding maps of the circle. *Ergodic Theory Dynam. Systems*, 28(6):1849–1860, 2008.

[27] Y. Kifer. *Random perturbations of dynamical systems*, volume 16 of *Progress in Probability and Statistics*. Birkhäuser Boston Inc., Boston, MA, 1988.
[28] Y. Kifer. Averaging in dynamical systems and large deviations. *Invent. Math.*, 110(2):337–370, 1992.
[29] Y. Kifer. Averaging principle for fully coupled dynamical systems and large deviations. *Ergodic Theory Dynam. Systems*, 24(3):847–871, 2004.
[30] Y. Kifer. Large deviations and adiabatic transitions for dynamical systems and Markov processes in fully coupled averaging. *Mem. Amer. Math. Soc.*, 201(944):viii+129, 2009.
[31] N. S. Krylov. *Works on the foundations of statistical physics*. Princeton University Press, Princeton, N.J., 1979. Translated from the Russian by A. B. Migdal, Ya. G. Sinai [Ja. G. Sinaĭ] and Yu. L. Zeeman [Ju. L. Zeeman], With a preface by A. S. Wightman, With a biography of Krylov by V. A. Fock [V. A. Fok], With an introductory article "The views of N. S. Krylov on the foundations of statistical physics" by Migdal and Fok, With a supplementary article "Development of Krylov's ideas" by Sinaĭ, Princeton Series in Physics.
[32] C. Liverani. On contact Anosov flows. *Ann. of Math. (2)*, 159(3):1275–1312, 2004.
[33] C. Liverani. The 2009 Michael Brin Prize in dynamical systems. *J. Mod. Dyn.*, 4(2):i–ii, 2010.
[34] J. Y. Marcelo Viana. Geometric and measure-theoretical structures of maps with mostly contracting center. *arXiv:1410.6308*, 2014.
[35] I. Melbourne. Rapid decay of correlations for nonuniformly hyperbolic flows. *Trans. Amer. Math. Soc.*, 359(5):2421–2441 (electronic), 2007.
[36] E. Olivieri and M. E. Vares. *Large deviations and metastability*, volume 100 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2005.
[37] K. R. Parthasarathy. On the category of ergodic measures. *Illinois J. Math.*, 5:648–656, 1961.
[38] Y. Pesin. On the work of Dolgopyat on partial and nonuniform hyperbolicity. *J. Mod. Dyn.*, 4(2):227–241, 2010.
[39] C. Pugh and M. Shub. Stably ergodic dynamical systems and partial hyperbolicity. *J. Complexity*, 13(1):125–179, 1997.
[40] D. Ruelle. A mechanical model for Fourier's law of heat conduction. *Comm. Math. Phys.*, 311(3):755–768, 2012.
[41] M. Shub and A. Wilkinson. Pathological foliations and removable zero exponents. *Invent. Math.*, 139(3):495–508, 2000.
[42] M. Tsujii. Physical measures for partially hyperbolic surface endomorphisms. *Acta Math.*, 194(1):37–132, 2005.
[43] M. Tsujii. Quasi-compactness of transfer operators for contact Anosov flows. *Nonlinearity*, 23(7):1495–1545, 2010.
[44] C. Villani. Hypocoercivity. *Mem. Amer. Math. Soc.*, 202(950):iv+141, 2009.
[45] D. Volk. Private communication.
[46] L.-S. Young. Recurrence times and rates of mixing. *Israel J. Math.*, 110:153–188, 1999.
[47] L.-S. Young. What are SRB measures, and which dynamical systems have them? *J. Statist. Phys.*, 108(5-6):733–754, 2002. Dedicated to David Ruelle and Yasha Sinai on the occasion of their 65th birthdays.

Jacopo De Simoi, Department of Mathematics, University of Toronto, 40 St George St. Toronto, ON M5S 2E4
E-mail address: jacopods@math.utoronto.ca
URL: http://www.math.utoronto.ca/jacopods

Carlangelo Liverani, Dipartimento di Matematica, Università di Roma (Tor Vergata), Via della Ricerca Scientifica, 00133 Roma, Italy.
E-mail address: liverani@mat.uniroma2.it
URL: http://www.mat.uniroma2.it/~liverani