# Random Iteration of Maps on a Cylinder and diffusive behavior

O. Castejón[*], V. Kaloshin[†]

June 13, 2015

**Abstract**

In this paper we propose a model of random compositions of maps of a cylinder, which in the simplified form is as follows: $(\theta, r) \in \mathbb{T} \times \mathbb{R} = \mathbb{A}$ and

$$f_{\pm 1} : \begin{pmatrix} \theta \\ r \end{pmatrix} \longmapsto \begin{pmatrix} \theta + r + \varepsilon u_{\pm 1}(\theta, r). \\ r + \varepsilon v_{\pm 1}(\theta, r). \end{pmatrix},$$

where $u_\pm$ and $v_\pm$ are smooth and $v_\pm$ are trigonometric polynomials in $\theta$ such that $\int v_\pm(\theta, r)\, d\theta = 0$ for each $r$. We study the random compositions

$$(\theta_n, r_n) = f_{\omega_{n-1}} \circ \cdots \circ f_{\omega_0}(\theta_0, r_0)$$

with $\omega_k \in \{-1, 1\}$ with equal probabilities. We show that under natural non-degeneracy hypothesis for $n \sim \varepsilon^{-2}$ the distributions of $r_n - r_0$ weakly converge to a diffusion process with explicitly computable drift and variance.

In the case of random iteration of the standard maps

$$f_{\pm 1} : \begin{pmatrix} \theta \\ r \end{pmatrix} \longmapsto \begin{pmatrix} \theta + r + \varepsilon v_{\pm 1}(\theta). \\ r + \varepsilon v_{\pm 1}(\theta) \end{pmatrix},$$

where $v_\pm$ are trigonometric polynomials such that $\int v_\pm(\theta)\, d\theta = 0$ we prove a vertical central limit theorem. Namely, for $n \sim \varepsilon^{-2}$ the distributions of $r_n - r_0$ weakly converge to a normal distribution $\mathcal{N}(0, \sigma^2)$ for $\sigma^2 = \frac{1}{4} \int (v_+(\theta) - v_-(\theta))^2\, d\theta$.

Such random models arise as a restrictions to a Normally Hyperbolic Invariant Lamination for a Hamiltonian flow of the generalized example of Arnold. We hope that this mechanism of stochasticity sheds some light on formation of diffusive behaviour at resonances of nearly integrable Hamiltonian systems.

[*]Universitat Politècnica de Catalunya, oriol.castejon@gmail.com

[†]University of Maryland at College Park, vadim.kaloshin@gmail.com

# Contents

# 1 Introduction

## 1.1 Motivation: Arnold diffusion and instabilities

By Arnold-Louiville theorem a completely integrable Hamiltonian system can be written in action-angle coordinates, namely, for action $p$ in an open set $U \subset \mathbb{R}^n$ and angle $\theta$ on an $n$-dimensional torus $\mathbb{T}^n$ there is a function $H_0(p)$ such that equations of motion have the form

$$\dot{\theta} = \omega(p), \quad \dot{p} = 0, \qquad \text{where } \ \omega(p) := \partial_p H_0(p).$$

The phase space is foliated by invariant $n$-dimensional tori $\{p = p_0\}$ with either periodic or quasi-periodic motions $\theta(t) = \theta_0 + t\,\omega(p_0) \pmod 1$. There are many different examples of integrable systems (see e.g. wikipedia).

It is natural to consider small Hamiltonian perturbations

$$H_\varepsilon(\theta, p) = H_0(p) + \varepsilon H_1(\theta, p), \qquad \theta \in \mathbb{T}^n, \ p \in U$$

where $\varepsilon$ is small. The new equations of motion become

$$\dot{\theta} = \omega(p) + \varepsilon \partial_p H_1, \quad \dot{p} = -\varepsilon \partial_\theta H_1.$$

In the sixties, Arnold [1] (see also [2, 3]) conjectured that *for a generic analytic perturbation there are orbits $(\theta, p)(t)$ for which the variation of the actions is of order one, i.e. $\|p(t) - p(0)\|$ that is bounded from below independently of $\varepsilon$ for all $\varepsilon$ sufficiently small.*

See [5, 10, 25, 26, 27] about recent progress proving this conjecture for convex Hamiltonians.

## 1.2 KAM stability

Obstructions to Arnold diffusion, and to any form of instability in general, are widely known, following the works of Kolmogorov, Arnold, and Moser called nowadays KAM theory. The fundamental result says that for a properly non-degenerate $H_0$ and for all sufficiently regular perturbations $\varepsilon H_1$, the system defined by $H_\varepsilon$ still has many invariant $n$-dimensional tori. These tori are small deformation of unperturbed tori and measure of the union of these invariant tori tends to the full measure as $\varepsilon$ goes to zero.

One consequence of KAM theory is that for $n = 2$ there are no instabilities. Indeed, generic energy surfaces $S_E = \{H_\varepsilon = E\}$ are 3-dimensional manifolds, KAM tori are 2-dimensional. Thus, KAM tori separate surfaces $S_E$ and prevent orbits from diffusing.

## 1.3 A priori unstable systems

As an interesting model [1] Arnold proposed to study the following example

$$H_\varepsilon(p,q,I,\varphi,t) = \frac{I^2}{2} + H_0(p,q) + \varepsilon H_1(p,q,I,\varphi,t) := \\ = \underbrace{\frac{I^2}{2}}_{harmonic\ oscillator} + \underbrace{\frac{p^2}{2} + (\cos q - 1)}_{pendulum} + \varepsilon H_1(p,q,I,\varphi,t), \tag{1}$$

where $q, \varphi, t \in \mathbb{T}$ are angles, $p, I \in \mathbb{R}$ are actions (see Fig. 3) and $H_1 = (\cos q - 1)(\cos\varphi + \cos t)$.

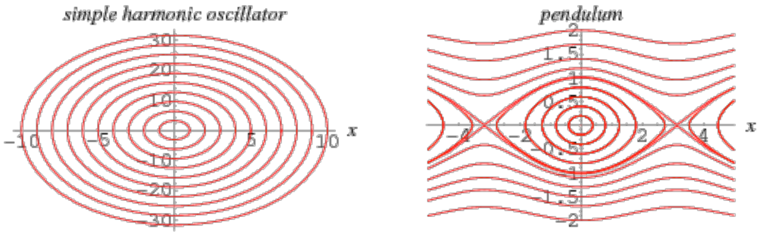


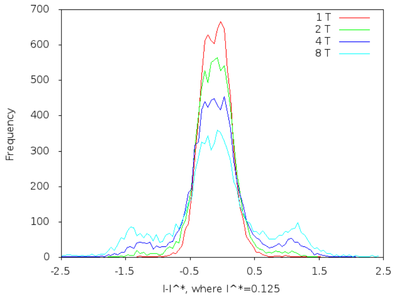

Figure 1: The rotor times the pendulum

For $\varepsilon = 0$ the system is a direct product of the harmonic oscillator $\ddot{\varphi} = 0$ and the pendulum $\ddot{q} = \sin q$. Instabilities occur when the $(p,q)$-component follows the separatrices $H_0(p,q) = 0$ and passes near the saddle $(p,q) = (0,0)$. Equations of motion for $H_\varepsilon$ have a (normally hyperbolic) invariant cylinder $\Lambda_\varepsilon$ which is $\mathcal{C}^1$ close to $\Lambda_0 = \{p = q = 0\}$. Systems having an invariant cylinder with a family of separatrix loops are called *an apriori unstable*. Since they were introduced by Arnold [1], they received a lot of attention both in mathematics and physics community see e.g. [4, 9, 10, 11, 13, 21, 40, 41].

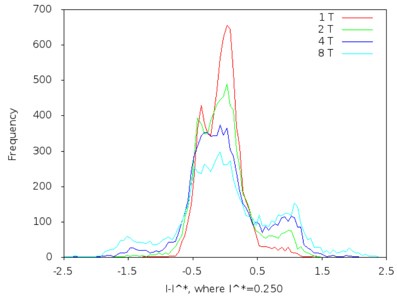
Chirikov [10] and his followers made extensive numerical studies for the Arnold example. It indicates that *the $I$-displacement behaves randomly, where randomness is due to choice of initial conditions near $H_0(p,q) = 0$.*

More exactly, integration of solutions whose "initial conditions" randomly chosen $\varepsilon$-close to $H_0(p,q) = 0$ and integrated over time $\sim -\varepsilon^{-2} \ln \varepsilon$-time. This leads to the $I$- displacement being of order of one and having some distribution. This coined the name for this phenomenon: *Arnold diffusion.*

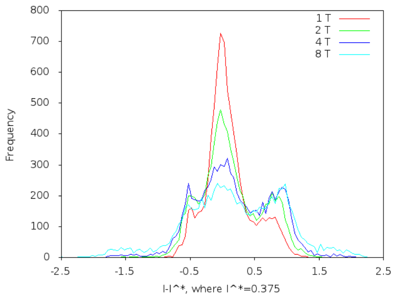
Let $\varepsilon = 0.01$ and $T = -\varepsilon^{-2} \ln \varepsilon$. On Fig. 1.3 we present several histograms plotting displacement of the $I$-component after time $T, 2T, 4T, 8T$ with 6 different groups of initial conditions, and histograms of $10^6$ points. In each group we start with a large set of initial conditions close to $p = q = 0$, $I = I^*$.[1] One of the distinct features is that only one distribution (a) is close symmetric, while in all others have a drift.



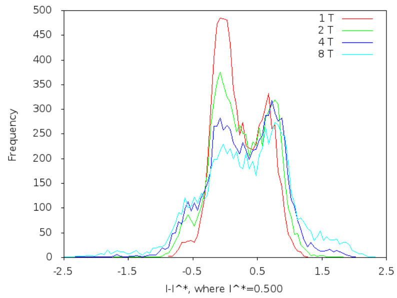
(a)



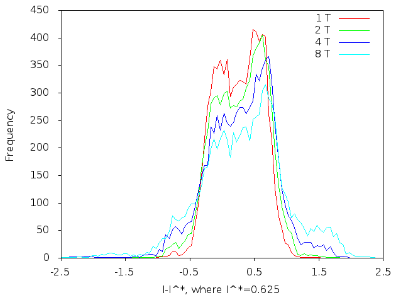
(b)



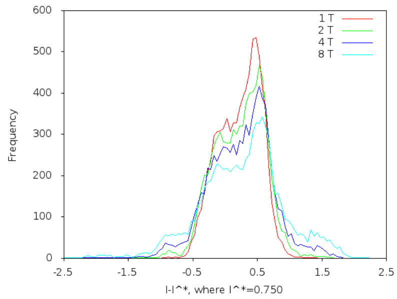
(c)



(d)



(e)



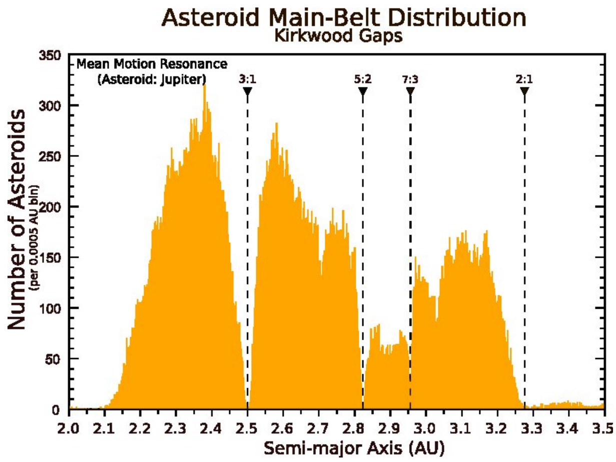
(f)

A similar stochastic behaviour was observed numerically in many other nearly integrable problems ([10] pg. 370, [16, 28], see also [37]). To give another illustrative example consider motion of asteroids in the asteroid belt.

[1]These histograms are part of the forthcoming paper of the second author with P. Roldan with extensive numerical analysis of dynamics of the Arnold's example.

## 1.4 Random fluctuations of eccentricity in Kirkwood gaps in the asteroid belt

The asteroid belt is located between orbits of Mars and Jupiter and has around one million asteroids of diameter of at least one kilometer. When astronoters build a histogram based on orbital period of asteroids there are well known gaps in distribution called *Kirkwood gaps* (see Figure below).




These gaps occur when ratio of of an asteroid and of Jupiter is a rational with small denominator: $1/3, 2/5, 3/7$. This correspond to so called *mean motion resonances for the three body problem.* Wisdom [42] made a numerical analysis of dynamics at mean motion resonance and observed *random fluctuations of eccentricity* of asteroids. As these fluctuations grow and eccentricity reaches a certain critical value an orbit of a hypothetic asteroid starts to cross the orbit of Mars. This eventually leads either to a collision of the asteroid with Mars or a close encounter. The latter changes the orbit so drastically that almost certainly it disappears from the asteroid belt. In [17] we only managed to prove existence of certain orbits whose eccentricity change by $O(1)$ for the restricted planar three body problem. Outside of these resonances one could argue that KAM theory provides stability [33].

## 1.5 Random iteration of cylinder maps

Consider the time one map of $H_\varepsilon$, denoted

$$F_\varepsilon : (p, q, I, \varphi) \to (p', q', I', \varphi').$$

It turns out that for initial conditions $\varepsilon$-close to $H_0(p, q) = 0$, except of a hypersurface, one can define a return map to an $O(\varepsilon)$-neighborhood of $(p, q) = 0$.

Often such a map is called *a separatrix map* and in the 2-dimensional case was introduced by physicists Filonenko-Zaslavskii [18]. In multidimensional setting such a map was defined and studied by Treschev [34, 39, 40, 41].

It turns starting near $(p,q) = 0$ and iterating $F_\varepsilon$ until the orbit comes back $(p,q) = 0$ leads to a family of maps of a cylinder

$$f_{\varepsilon,p,q} : (I,\varphi) \to (I',\varphi'), \qquad (I,\varphi) \in \mathbb{A} = \mathbb{R} \times \mathbb{T}$$

which are close to integrable. Since at $(p,q) = 0$ the $(p,q)$-component has a saddle, there is a sensitive dependence on initial condition in $(p,q)$ and returns do have some randomness in $(p,q)$. The precise nature of this randomness at the moment is not clear. There are several coexisting behaviours, including unstable diffusive, stable quasi-periodic, orbits can stick to KAM tori, and which one is dominant is to be understood. May be mechanism of capture into resonances [15] is also relevant in this setting.

In [22] we construct a normally hyperbolic lamination (NHL) for an open class of trigonometric perturbations of the form

$$H_1 = (\cos q - 1) P(\exp(i\varphi), \exp(it)).$$

Constructing unstable orbits along NHL is also discussed in [14]. In general, NHL give rise to a skew shift. For example, let $\Sigma = \{-1,1\}^{\mathbb{Z}}$ be the space of infinite sequences of $-1$'s and 1's and $\sigma : \Sigma \to \Sigma$ be the standard shift.

*Consider a skew product of cylinder maps*

$$F : \mathbb{A} \times \Sigma \to \mathbb{A} \times \Sigma, \qquad F(r,\theta;\omega) = (f_\omega(r,\theta), \sigma\omega),$$

*where each $f_\omega(r,\theta)$ is a nearly integrable cylinder maps, in the sense that it almost preserves the $r$-component* [2].

The goal of the present paper is to study a wide enough class of skew products so that they arise in Arnold's example with a trigonometric perturbation of the above type (see [22]).

Now we formalize our model and present the main result.

## 1.6 Diffusion processes and infinitesimal generators

In order to formalise the statement about diffusive behaviour we need to recall some basic probabilistic notions. Consider a Brownian motion $\{B_t,\, t \geq 0\}$.

[2] The reason we switch from the $(I,\varphi)$-coordinates on the cylinder to $(r,\theta)$ is because we perform a coordinate change.

A Brownian motion is a properly chosen limit of the standard random walk. A generalisation of a Brownian motion is *a diffusion process* or *an Ito diffusion.* To define it let $(\Omega, \Sigma, P)$ be a probability space. Let $R : [0, +\infty) \times \Omega \to \mathbb{R}$. It is called an Ito diffusion if it satisfies *a stochastic differential equation* of the form

$$\mathrm{d}R_t = b(R_t)\,\mathrm{d}t + \sigma(R_t)\,\mathrm{d}B_t, \tag{2}$$

where B is an Brownian motion, $b : \mathbb{R} \to \mathbb{R}$ and $\sigma : \mathbb{R} \to \mathbb{R}$ are Lipschitz functions called the drift and the variance respectively. For a point $r \in \mathbb{R}$, let $\mathbb{P}_r$ denote the law of $X$ given initial data $R_0 = r$, and let $\mathbb{E}_r$ denote expectation with respect to $\mathbb{P}_r$.

The *infinitesimal generator* of $R$ is the operator $A$, which is defined to act on suitable functions $f : \mathbb{R} \to \mathbb{R}$ by

$$Af(r) = \lim_{t \downarrow 0} \frac{\mathbb{E}_r[f(R_t)] - f(r)}{t}.$$

The set of all functions $f$ for which this limit exists at a point $r$ is denoted $D_A(r)$, while $D_A$ denotes the set of all $f$'s for which the limit exists for all $r \in \mathbb{R}$. One can show that any compactly-supported $\mathcal{C}^2$ function $f$ lies in $D_A$ and that

$$Af(r) = b(r)\frac{\partial f}{\partial r} + \frac{1}{2}\sigma^2(r)\frac{\partial^2 f}{\partial r \partial r}. \tag{3}$$

The distribution of a diffusion process is characterise by the drift $b(r)$ and the variance $\sigma^2(r)$.

# 2 The model and statement of the main result

Let $\varepsilon > 0$ be a small parameter and $l \geq 12$, $s \geq 0$ be integers. Denote by $\mathcal{O}_s(\varepsilon)$ a $\mathcal{C}^s$ function whose $\mathcal{C}^s$ norm is bounded by $C\varepsilon$ with $C$ independent of $\varepsilon$. Similar definition applies for a power of $\varepsilon$. As before $\Sigma$ denotes $\{0,1\}^{\mathbb{Z}}$ and $\omega = (\dots, \omega_0, \dots) \in \Sigma$.

Consider two nearly integrable maps:

$$\begin{array}{rcl} f_\omega : \mathbb{T} \times \mathbb{R} & \longrightarrow & \mathbb{T} \times \mathbb{R} \\ f_\omega : \begin{pmatrix} \theta \\ r \end{pmatrix} & \longmapsto & \begin{pmatrix} \theta + r + \varepsilon u_{\omega_0}(\theta, r) + \mathcal{O}_s(\varepsilon^{1+a}, \omega) \\ r + \varepsilon v_{\omega_0}(\theta, r) + \varepsilon^2 w_{\omega_0}(\theta, r) + \mathcal{O}_s(\varepsilon^{2+a}, \omega) \end{pmatrix}. \end{array} \tag{4}$$

for $\omega_0 \in \{-1, 1\}$, where $u_{\omega_0}$, $v_{\omega_0}$, and $w_{\omega_0}$ are bounded $\mathcal{C}^l$ functions, 1-periodic in $\theta$, $\mathcal{O}_s(\varepsilon^{1+a}, \omega)$ and $\mathcal{O}_s(\varepsilon^{2+a}, \omega)$ denote remainders depending on $\omega$ and uniformly $C^s$ bounded in $\omega$, and $a > 1/2$. Assume

$$\max |v_i(\theta, r)| \leq 1,$$

where maximum is taken over $i = -1, 1$ and all $(\theta, r) \in \mathbb{A}$, otherwise, renormalize $\varepsilon$.

We study random iterations of the maps $f_1$ and $f_{-1}$, such that at each step the probability of performing either map is 1/2. Importance of understanding iterations of several maps for problems of diffusion is well known (see e.g. [24, 33]).

Denote the expected potential and the difference of potentials by

$$\mathbb{E}u(\theta, r) := \frac{1}{2}(u_1(\theta, r) + u_{-1}(\theta, r)), \quad \mathbb{E}v(\theta, r) := \frac{1}{2}(v_1(\theta, r) + v_{-1}(\theta, r)),$$

$$u(\theta, r) := \frac{1}{2}(u_1(\theta, r) - u_{-1}(\theta, r)), \quad v(\theta, r) := \frac{1}{2}(v_1(\theta, r) - v_{-1}(\theta, r)).$$

Suppose the following assumptions hold:

**[H0]** (*zero average*) Let for each $r \in \mathbb{R}$ and $i = \pm 1$ we have $\int v_i(\theta, r)\, d\theta = 0$.

**[H1]** (*no common zeroes*) For each integer $n \in \mathbb{Z}$ potentials $v_1(\theta, n)$ and $v_{-1}(\theta, n)$ have no common zeroes and, equivalently, $f_1$ and $f_{-1}$ have no fixed points;

**[H2]** for each $r \in \mathbb{R}$ we have $\int_0^1 v^2(\theta, r) d\theta =: \sigma(r) \neq 0$;

**[H3]** The functions $v_i(\theta, r)$ are trigonometric polynomials in $\theta$, i.e. for some positive integer $d$ we have

$$v_i(\theta, r) = \sum_{k \in \mathbb{Z},\ |k| \leq d} v^{(k)}(r) \exp 2\pi i k \theta.$$

For $\omega \in \{-1, 1\}^{\mathbb{Z}}$ we can rewrite the maps $f_\omega$ in the following form:

$$f_\omega \begin{pmatrix} \theta \\ r \end{pmatrix} \longmapsto \begin{pmatrix} \theta + r + \varepsilon \mathbb{E}u(\theta, r) + \varepsilon \omega_0 u(\theta, r) + \mathcal{O}_s(\varepsilon^{1+a}, \omega) \\ r + \varepsilon \mathbb{E}v(\theta, r) + \varepsilon \omega_0 v(\theta, r) + \varepsilon^2 w_{\omega_0}(\theta, r) + \mathcal{O}_s(\varepsilon^{2+a}, \omega) \end{pmatrix}.$$

Let $n$ be positive integer and $\omega_k \in \{-1, 1\}$, $k = 0, \ldots, n-1$, be independent random variables with $\mathbb{P}\{\omega_k = \pm 1\} = 1/2$ and $\Omega_n = \{\omega_0, \ldots, \omega_{n-1}\}$. Given an initial condition $(\theta_0, r_0)$ we denote:

$$(\theta_n, r_n) := f^n_{\Omega_n}(\theta_0, r_0) = f_{\omega_{n-1}} \circ f_{\omega_{n-2}} \circ \cdots \circ f_{\omega_0}(\theta_0, r_0). \tag{5}$$

**[H4]** (*no common periodic orbits*) Suppose for any rational $r = p/q \in \mathbb{Q}$ with $p, q$ relatively prime, $1 \leq |q| \leq 2d$ and any $\theta \in \mathbb{T}$

$$\sum_{k=1}^{q} \left[ v_{-1}(\theta + \frac{k}{q}, r) - v_1(\theta + \frac{k}{q}, r) \right]^2 \neq 0.$$

This prohibits $f_1$ and $f_{-1}$ to have common periodic orbits of period $|q|$.

[**H5**] (*no degenerate periodic points*) Suppose for any rational $r = p/q \in \mathbb{Q}$ with $p, q$ relatively prime, $1 \leq |q| \leq d$, the function:

$$\mathbb{E}v_{p,q}(\theta, r) = \sum_{\substack{k \in \mathbb{Z} \\ 0<|kq|<d}} \mathbb{E}v^{kq}(r) e^{2\pi i k q \theta}$$

has distinct non-degenerate zeroes, where $\mathbb{E}v^j(r)$ denotes the $j$–th Fourier coefficient of $\mathbb{E}v(\theta, r)$.

A straightforward calculation shows that:

$$\begin{aligned}
\theta_n &= \theta_0 + n r_0 + \varepsilon \sum_{k=0}^{n-1} \left( \mathbb{E}u(\theta_k, r_k) + \mathbb{E}v(\theta_k, r_k) \right) \\
&\quad + \varepsilon \sum_{k=0}^{n-1} \omega_k \left( u(\theta_k, r_k) + v(\theta_k, r_k) \right) + \mathcal{O}_s(n \varepsilon^{1+a}) \\
r_n &= r_0 + \varepsilon \sum_{k=0}^{n-1} \mathbb{E}v(\theta_k, r_k) + \varepsilon \sum_{k=0}^{n-1} \omega_k v(\theta_k, r_k) + \mathcal{O}_s(n \varepsilon^{2+a})
\end{aligned} \tag{6}$$

Even though these maps might not be area-preserving, using normal forms we will simplify these maps significantly on a large domain of the cylinder.

**Theorem 2.1.** *Assume that in the notations above conditions* [**H0-H5**] *hold. Let $n_\varepsilon \varepsilon^2 \to s > 0$ as $\varepsilon \to 0$ for some $s > 0$. Then as $\varepsilon \to 0$ the distribution of $r_{n_\varepsilon} - r_0$ converges weakly to $R_s$, where $R_\bullet$ is a diffusion process of the form* (2), *with the drift and the variance*

$$b(R) = \int_0^1 E_2(\theta, R)\, d\theta, \qquad \sigma^2(R) = \int_0^1 v^2(\theta, R)\, d\theta.$$

*for some function $E_2$, defined in (11).*

- In the case that $u_{\pm 1} = v_{\pm 1}$ and they are independent of $r$ we have two area-preserving standard maps. In this case the assumptions become
  - [**H0**] $\int v_i(\theta) d\theta = 0$ for $i = \pm 1$;
  - [**H1**] $v_1$ and $v_{-1}$ have no common zeroes;
  - [**H2**] $v$ is not identically zero.
  - [**H3**] the functions $v_i$ are trigonometric polynomials;
  - [**H4**] the same condition as above without dependence on $r$;

– **[H5]** the same condition as above without dependence on $r$;

A good example is $u_1(\theta) = v_1(\theta) = \cos\theta$ and $u_{-1}(\theta) = v_{-1}(\theta) = \sin\theta$. In this case

$$\int_0^1 E_2(\theta, r) d\theta \equiv 0, \qquad \sigma^2 = \int_0^1 v^2(\theta)\, d\theta$$

and for $n \le \varepsilon^{-2}$ the distribution $r_n - r_0$ converges to the zero mean variance $\varepsilon n^2\sigma^2$ normal distribution, denoted $\mathcal{N}(0, \varepsilon n^2\sigma^2)$. More generally, we have the following "vertical central limit theorem":

**Theorem 2.2.** *Assume that in the notations above conditions* **[H0-H5]** *hold. Let $n_\varepsilon \varepsilon^2 \to s > 0$ as $\varepsilon \to 0$ for some $s > 0$. Then as $\varepsilon \to 0$ the distribution of $r_{n_\varepsilon} - r_0$ converges weakly to a normal random variable $\mathcal{N}(0, s^2\sigma^2)$.*

- Numerical experiments of Mockel [32] show that no common fixed points [H1] (resp. [H4]) is not neccessary for Theorem 2.1 to hold. One could probably replaced by a weaker non-degeneracy condition, e.g. that the linearisation of maps $f_{\pm 1}$ at the common fixed point (resp. periodic points) are different.

- In [35] Sauzin studies random iterations of the standard maps $(\theta, r) \to (\theta + r + \lambda\phi(\theta), r + \lambda\phi(\theta))$, where $\lambda$ is chosen randomly from $\{-1, 0, 1\}$ and proves the vertical central limit theorem; In [30, 36] Marco-Sauzin present examples of nearly integrable systems having a set of initial conditions exhibiting the vertical central limit theorem.

- In [29] Marco derives a sufficient condition for a skew-shift to be a step skew-shift.

- The condition [H3] that the functions $v_i$ are trigonometric polynomials in $\theta$ seems redundant too, however, removing it leads to considerable technical difficulties (see Section 3.2 and Remark 3.1). In short, for perturbations by a trigonometric polynomial there are finitely many resonant zones. This finiteness considerably simplifies the analysis.

- One can replace $\Sigma = \{0, 1\}^{\mathbb{Z}}$ with $\Sigma_N = \{0, 1, \ldots, N-1\}^{\mathbb{Z}}$, consider any finite number of maps of the form (4) and a transitive Markov chain with some transition probabilities. If conditions [H2–H4] are satisfied for the proper averages $\mathbb{E}v$ of $v$, then Theorem 2.1 holds.

# 3 Strategy of the proof

## 3.1 Strip decomposition

The main idea of the proof is to divide the cylinder $\mathbb{A}$ in strips $\mathbb{T} \times I^i_\beta$, where $I^j_\beta \subset \mathbb{R},\ j \in \mathbb{Z}$ are intervals of size $\varepsilon^\beta$, for any $0 < \beta \leq 1/5$. Then we will study how the random variable $r_n - r_0$ behaves in each strip. More precisely, decompose the process $r_n(\omega), n \in \mathbb{Z}_+$ into infinitely many time intervals defined by stopping times

$$0 < n_1 < n_2 < \dots, \tag{7}$$

where

- $r_{n_i}(\omega)$ is $\varepsilon$-close to the boundary between $I^j_\beta$ and $I^{j+1}_\beta$ for some $j \in \mathbb{Z}$
- $r_{n_{i+1}}(\omega)$ is $\varepsilon$-close to the other boundary of either $I^j_\beta$ or of $I^{j+1}_\beta$ and $n_{i+1} > n_i$ is the smallest integer with this property.

Since $\varepsilon \ll \varepsilon^\beta$, being $\varepsilon$-close to the boundary of $I^j_\beta$ with a negligible error means jump from $I^j_\beta$ to the neighbour interval $I^{j\pm1}_\beta$. In what follows for brevity we drop dependence of $r_n(\omega)$'s on $\omega$.

## 3.2 Subdivision of the cylinder into domains with different quantitative behaviour

Fix $b > 0$ such that $0 < \beta - 2b < 0.04$, small $\gamma > 0$, and $K_i := K_i(u_1, v_1, u_2, v_2)$, $i = 1, 2$, depending on functions $u_j, v_j,\ j = 1, 2$, such that $K_1 < K_2$ and all are independent of $\varepsilon$. Consider the $\varepsilon^\beta$-grid in $\mathbb{R}$. Denote by $I_\beta$ a segment whose end points are in the grid. We distinguish among three types of strips $I_\beta$. We will have strips of three types as well as transition zones from one to another. We define:

- **The Real Rational (RR) case:** A strip $I_\beta$ is called *real rational* if there exists a rational $p/q \in I_\beta$, with $\gcd(p, q) = 1$ and $|q| \leq d$. Clearly, there are just finitely many strips of this kind. However, this case is the most complicated one and requires a detailed study.
- **The Imaginary Rational (IR) case:** A strip $I_\beta$ is called *imaginary rational* if there exists a rational $p/q \in I_\beta$, with $\gcd(p, q) = 1$ with $d < |q| < \varepsilon^{-b}$.

The reason we call these strips are imaginary rational, because the leading term of the angular dynamics is a rational rotation, however, averaged systems appearing in the previous case are vanishing (see the next section).

We show that the imaginary rational strips occupy an $\mathcal{O}(\varepsilon^\rho)$-fraction of the cylinder (see Lm. A.1 in Sect. A). We can show that orbits spend small fraction of the total time in these strips and global behaviour is determined by behaviours in the complement, which we call totally irrational.

- **The Totally Irrational (TI) case:** A strip $I_\beta$ is called *totally irrational* if $r \in I_\beta$ and $|r - p/q| < \varepsilon^\beta$, with $\gcd(p, q) = 1$, then $|q| > \varepsilon^{-b}$.

  In this case, we show that there is a good "ergodization" and

$$\sum_{k=0}^{n-1} \omega_k v \left(\theta_0 + k\frac{p}{q}\right) \approx \sum_{k=0}^{n-1} \omega_k v \left(\theta_0 + k r_0^*\right).$$

  Loosely speaking, any $r_0^* \in I_\beta \cap (\mathbb{R} \setminus \mathbb{Q})$ can be treated as an irrational. These strips cover most of the cylinder and give the dominant contribution to the behaviour of $r_n - r_0$. Eventually it will lead to the desired weak convergence to a diffusion process (Theorem 2.1).

- **Transition zones, type I:** A zone is a transition zone if there is $p/q$ such that $\gcd(p, q) = 1$ and $|q| \leq d$ and it is defined by the corresponding annuli $K_1\varepsilon^{1/2} \leq |r - p/q| \leq K_2\varepsilon^{1/6}$.

  Analysis in these zones needs to be adapted as "influence" of real resonances is strong.

- **Transition zones, type II:** A zone is a transition zone if there is $p/q$ such that $\gcd(p, q) = 1$ and $|q| \leq d$ and it is defined by the corresponding annuli $K_2\varepsilon^{1/6} \leq |r - p/q| \leq \gamma$.

  Analysis in these zones requires an adjusted coordinates, otherwise, we still study the Totally Irrational and the Imaginary Rational strips inside of the type II Transition Zones.

**Remark 3.1.** Notice that finiteness of Real Rational strips follows from assumption [H3]. If the expected potential is not a trigonometric polynomial in $\theta$ this is not true.

## 3.3 The normal form

The first step is to find a normal form, so that the deterministic part of map (6) is as simple as possible. In short, we shall see that the deterministic system in both the Totally Irrational case and the Imaginary rational case are a small perturbation of the perfect twist map:

$$\begin{pmatrix} \theta \\ r \end{pmatrix} \longmapsto \begin{pmatrix} \theta + r \\ r \end{pmatrix}.$$

On the contrary, in the Real Rational case, the deterministic system will be close to a pendulum-like system:

$$\begin{pmatrix} \theta \\ r \end{pmatrix} \longmapsto \begin{pmatrix} \theta + r \\ r + \varepsilon E(\theta, r) \end{pmatrix},$$

for an "averaged" potential $E(\theta, r)$ (see e.g. Thm. 4.2, (12)). We note that this system has the following approximate first integral:

$$H(\theta, r) = \frac{r^2}{2} - \varepsilon \int_0^\theta E(s, r) ds, \tag{8}$$

so that indeed it is close to a pendulum-like system. This will lead to different qualitative behaviours when considering the random system. Inside the Real Rational strips as well as the transition zones we use $H$ as one of the coordinates.

The rigorous statement of these results about the normal forms is given in Theorem 4.2, Sect. 4.

### 3.4 Analysis of the Martingale problem in each kind of strip

The next step is to study the behaviour of the random system respectively in Totally Irrational, Imaginary Rational and Real Rational strips, as well as in the Transition Zones. This is done in Sections 5.1–5.6. More precisely, we use a discrete version of the scheme by Freidlin and Wentzell [20], giving a sufficient condition to have weak convergence to a diffusion process as $\varepsilon \to 0$ in terms of the associated Martingale problem (see Lemma C.1). Now using the results proved below we derive the main result — Theorem 2.1. This is done in two steps. First, we describe local behaviour in each strip and then we combine the information. Fix $s > 0$.

By the discrete version of Lemma C.1 is sufficient to prove that as $\varepsilon \to 0$ any time $n \leq s\varepsilon^{-2}$ and any $(\theta_0, r_0)$ we have

$$\mathbb{E}\Big(e^{-\lambda\varepsilon^2 n} f(r_n) + \varepsilon^2 \sum_{k=0}^{n-1} e^{-\lambda\varepsilon^2 k} \Big[\lambda f(r_k) - \Big(b(r_k) f'(r_k) + \frac{\sigma^2(r_k)}{2} f''(r_k)\Big)\Big]\Big) - f(r_0) \to 0, \tag{9}$$

We define Markov times $0 = n_0 < n_1 < n_2 < \cdots < n_{m-1} < n_m < n$ for some random $m = m(\omega)$ such that each $n_k$ is the stopping time as in (7). Almost surely $m(\omega)$ is finite. We decompose the above sum

$$\sum_{k=0}^{m} \quad \mathbb{E}\Bigg(e^{-\lambda\varepsilon^2 n_{k+1}} f(r_{n_{k+1}}) - e^{-\lambda\varepsilon^2 n_k} f(r_{n_k}) + \varepsilon^2 \sum_{s=n_k}^{n_{k+1}} e^{-\lambda\varepsilon^2 s}\left[\lambda f(r_s) - \left(b(r_s)f'(r_s) + \frac{\sigma^2(r_s)}{2} f''(r_s)\right)\right]\Bigg)$$

and show that it converges to $f(r_0)$.

### 3.4.1 A Totally Irrational Strip

Let the drift and the variance be

$$b(r) = \int_0^1 E_2(\theta, r)\, d\theta \quad \text{and} \quad \sigma^2(r) = \int_0^1 v^2(\theta, r)\, d\theta,$$

where the function $E_2$ is defined in (11). Let $r_0$ be $\varepsilon$-close to the boundary of two totally irrational strips and let $n_\beta$ be stopping of hitting $\varepsilon$-neighbourhoods of the adjacent boundaries. In Lemma 5.3 we prove that

$$\mathbb{E}\Bigg(e^{-\lambda\varepsilon^2 n_\beta} f(r_{n_\beta}) + \varepsilon^2 \sum_{k=0}^{n_\beta - 1} e^{-\lambda\varepsilon^2 k}\left[\lambda f(r_k) - \left(b(r_k)f'(r_k) + \frac{\sigma^2(r_k)}{2} f''(r_k)\right)\right]\Bigg) - f(r_0) = \mathcal{O}(\varepsilon^{2\beta+d}),$$

for some $d > 0$.

### 3.4.2 An Imaginary Rational Strip

Let the drift and the variance be

$$b_{IR}(\theta, r) = \frac{1}{q}\sum_{k=0}^{q-1} E_2(\theta + kr, r) \quad \text{and} \quad \sigma^2_{IR}(\theta, r) = \frac{1}{q}\sum_{k=0}^{q-1} v^2(\theta + kr, r).$$

Let $r_0$ be $\varepsilon$-close to the boundary of an imaginary rational strip and let $n_\beta$ be stopping of hitting $\varepsilon$-neighbourhoods of the adjacent boundaries. In Lemma 5.5

we prove that

$$\mathbb{E}\Bigg(e^{-\lambda\varepsilon^2 n_\beta} f(r_{n_\beta})+$$
$$\varepsilon^2 \sum_{k=0}^{n_\beta-1} e^{-\lambda\varepsilon^2 k}\left[\lambda f(r_k)-\left(b(\theta_k,r_k)f'(r_k)+\frac{\sigma^2(\theta_k,r_k)}{2}f''(r_k)\right)\right]\Bigg)$$
$$-f(r_0)=\mathcal{O}(\varepsilon^{2\beta+d}),$$

As one can see, the limiting process does not take place on a line, since the drift and diffusion coefficient depend also on the variable $\theta$.

Notice that the drift $b(\theta,r)$ and the variance $\sigma(\theta,r)$ both are $\theta$-dependent functions. In Section 3.5 we show that time spent in these strips is too small to affect the drift and the variance of the limiting process.

### 3.4.3 A Real Rational Strip

Let in the rescaled variable $r-p/q=R\sqrt{\varepsilon}$ the drift and the variance be

$$b_{RR}(\theta,R)=F(\theta,R), \qquad \sigma^2_{RR}(\theta,R)=(R\,p/q)^2\sum_{k=0}^{q-1} v^2(\theta+kR,R),$$

where $F$ is some function to be defined in (102). Consider the Real Rational case assuming that

$$|r-p/q|\le K_1\varepsilon^{1/2} \qquad \Longleftrightarrow \qquad |R|\le K_1$$

that is, that $r$ is close to the "pendulum" domain. In this case, we study the process $(\theta_{qn},H_n)$ with $H_n:=H^{p/q}(\theta_{qn},R_{qn})$, where $H^{p/q}(\theta,R)$ is an approximate first integral of the deterministic system (8). In the rescaled variables it has the form

$$H^{p/q}(\theta,R)=\frac{R^2}{2}-V^{p/q}(\theta),$$

where

$$V^{p/q}(\theta)=\int_0^\theta \mathbb{E}v_{p,q}(s,p/q)\,ds$$

for a properly defined averaged potential (see Thm. 4.2, (14)). In Lemma 5.11 we prove that, $H_n-H_0$ converges weakly to a diffusion process $R_t$ with $t=\varepsilon^2 n$.

Notice that the limiting process does not take place on a line. In this case it takes place on a graph, similarly as in [20]. More precisely, consider the level sets of the function $H^{p/q}(\theta,R)$. The critical points of the potential $V^{p/q}(\theta)$ give rise to critical points of the associated Hamiltonian system. Moreover, if the critical point is a local minimum of $V$, then it corresponds to a focus of the Hamiltonian system, while if it is a local maximum of $V^{p/q}$, then it corresponds to a saddle of the Hamiltonian system. Now, if for every value $H \in \mathbb{R}$ we identify all the points $(\theta, R)$ in the same connected component of the curve $\{H^{p/q}(\theta, R) = H\}$, we obtain a graph $\Gamma$ (see Figure 2 for an example). The interior vertices of this graph represent the saddle points of the underlying Hamiltonian system jointly with their separatrices, while the exterior vertices represent the focuses of the underlying Hamiltonian system. Finally, the edges of the graph represent the domains that have the separatrices as boundaries. The process $H_n$ takes places on this graph, and so it is a diffusion process on a graph.

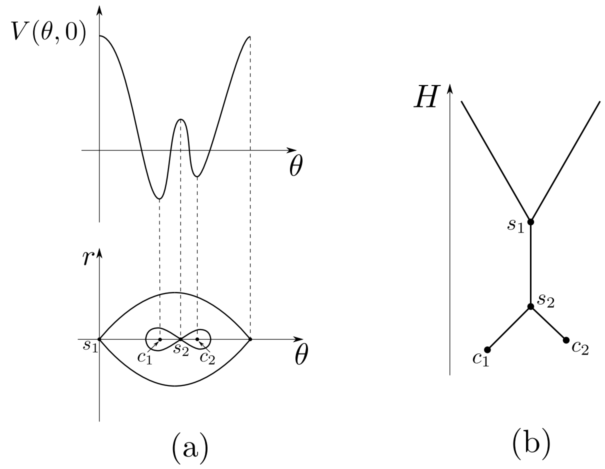



Figure 2: (a) A potential and the phase portrait of its corresponding Hamiltonian system. (b) The associated graph $\Gamma$.

### 3.4.4 A transition zone

Finally, in Lemma 5.14 we deal with the Transition Zones of Type I and Type II, that is the zones in the Real Rational strips such that $K_1 \leq |R| \leq K_2\varepsilon^{-1/3}$ and $K_2\varepsilon^{-1/3} \leq |R| \leq \gamma\varepsilon^{-1/2}$. In these strips we study the process $(\theta_{nq}, H_{nq}) = (\theta_{nq}, H(\theta_{qn}, R_{qn}))$. In this regime we fix small $\rho > 0$ and subdivide each zone in sub-strips

$$I_\rho(R_0) = \{H \in \mathbb{R} : |H - H_0| \leq |R_0|\, \varepsilon^{1/2-\rho}\}.$$

We prove that, inside each of one these sub-strips, as $\varepsilon \to 0$ the process $H_n - H_0$ converges weakly to a diffusion process $R_t$ with $t = \varepsilon^2 n$, zero drift and the variance:

$$\sigma^2_{TZ}(\theta, R) = |R|^2 \sum_{k=0}^{q-1} v^2(\theta + kR, R).$$

## 3.5 From the local diffusion in the rational strips to the global diffusion on the line

In this section we resolve the following problem. In order to combine all the previous results, which characterise the local behaviour of the process inside of infinitesimally small strips, to determine the global behaviour of the process in a $\mathcal{O}(1)-$strip. First, we prove that the Imaginary Rational and Real Rational strips cover a negligibly small part of any $\mathcal{O}(1)-$strip (see Section A). Then, one can argue that the process is determined by the process in the Totally Irrational strips.

Notice both the drift $b_{IR}(\theta, r)$ and the variance $b_{IR}(\theta, r)$ at any Imaginary Rational strip is given by $\theta$-dependent functions. Our main result (Theorem 2.1), however, is a diffusion process on a line. To prove that this dependence does not enter into the global diffusion process we show that the process spends infinitesimal amount of time inside of those strips as follows.

**Lemma 3.2.** *Let* $(\theta_k, r_k) = f^k_{\Omega_k}(\theta_0, r_0), k \geq 1$ *be a random orbit defined by (5) for some random sequence* $\{\omega_k\}_{k\in\mathbb{Z}_+}$*. Let* $n \leq \varepsilon^{-2}$ *and*

$$T_R(n) = \#\{0 \leq k \leq n : r_k \text{ belongs to either an Imaginary Rational or a Real Rational strip}\}.$$

*Then for any* $\rho > 0$ *and* $\varepsilon > 0$ *small enough*

$$\mathbb{P}\{T_R(n) \geq \rho n\} \leq \rho.$$

*Proof.* Define

$$b_{IR}(r) := \min_{\theta\in[0,1)} b_{IR}(\theta, r) \quad \text{and} \quad \sigma^2_{IR}(r) := \min_{\theta\in[0,1)} \sigma^2_{IR}(\theta, r)$$

Consider the process $R^{IR}_t$ with the drift $b_{IR}(r)$ and the variance $\sigma^2_{IR}(r)$. By definition this process spends more time in $I_\beta$ that the process with the drift $b_{IR}(\theta, r)$ and the variance $\sigma^2_{IR}(\theta, r)$. Moreover, it is a diffusion process on a line. Then, using a local time argument, it can be seen that the time spent on a given domain is proportional to the size of this domain up to a uniform constant. Hence, the time the original process spends in all the Imaginary Rational strips is infinitesimally small compared to the time it spends on the Totally Irrational ones. However, the time spent in the Imaginary Rational strip could be infinite and the argument would not be valid. This cannot happen, since if $r$ belongs to an Imaginary Rational strip one has that $\sigma^2(\theta, r) \neq 0$. Thus, it is enough to prove that for all imaginary rational $p/q$ one has $\sigma^2(\theta, p/q) \neq 0$. Indeed, if this is true, then for $|r - p/q| \leq \varepsilon^\beta$ and $\varepsilon$ is sufficiently small, one has that $\sigma^2(\theta, r) \neq 0$ by Lemma 3.3.

Finally, in the Real Rational case one can use a result from [19] that diffusion processes on a graph have well-defined local time. Thus, the time spent in all the Real Rational strips is infinitesimally small compared to the time spent in the Totally Irrational ones. Now one can have $\sigma^2_{RR} = 0$, but it happens just when $r = p/q$, which follows directly from assumption [**H2**]. In this case, one can see that $b_{RR}(\theta, r) \neq 0$, so that the process is non-degenerate and thus the fraction of time spend in the Real Rational strips is less than any ahead given fraction. □

**Lemma 3.3.** *$\sigma^2(\theta, p/q) \neq 0$ if $p/q$ is any Imaginary Rational.*

*Proof.* On the one hand, if $d < |q| \leq 2d$ this is ensured by hypothesis [**H4**]. On the other hand, if $|q| > 2d$ then $\sigma^2(\theta, p/q) = 0$ implies:

$$v(\theta + kp/q, p/q) = 0, \qquad k = 0, \cdots, q-1. \tag{10}$$

Now, since $v(\theta, p/q)$ is a trigonometric polynomial in $\theta$ of degree $d$, it can have at most $2d$ zeros, or else be identically equal to zero. The latter case cannot occur, since by assumption [**H2**] we know that

$$\int_0^1 v^2(\theta, r) \neq 0 \quad \text{for all} \quad r \in \mathbb{R},$$

so that $v(\theta, p/q) \not\equiv 0$. Thus, $v(\theta, p/q)$ has at most $2d$ zeros. Consequently equation (10) cannot be satisfied for all $k = 0, \cdots, q-1$, since $|q| > 2d$, so that $\sigma^2(\theta, p/q) \neq 0$. The same argument applies to the Transition Zones. □

Combining these facts one can apply the arguments from [20], sect. 8 and prove that the limiting diffusion process has the drift $b(r)$ and the variance $\sigma^2(r)$ corresponding to Totally Irrational strips.

## 3.6 Plan of the rest of the paper

In Section 4 we state and prove the normal form theorem for the expected cylinder map $\mathbb{E}f$. Main difference with a typical normal form is that we need to have not only the leading term in $\varepsilon$, but also $\varepsilon^2$-terms. The latter terms give information about the drift $b(r)$ (see (11)).

In Section 5.1 we analyse the Totally Irrational case and prove approximation for the expectation from Section 3.4.1.

In Section 5.2 we analyse the Imaginary Rational case and prove an analogous formula from Section 3.4.2.

In Section 5.3 we analyse the Real Rational case and prove an analogous formula from Section 3.4.3.

In Section 5.6 we study the Transition Zones and prove an analogous formula from Section 3.4.4.

In Section A we estimate measure of the complement to the Totally Irrational strips and the Transition Zones of type II.

In Section C we present several auxiliary lemmas used in the proof.

# 4 The Normal Form Theorem

In this section we shall prove the Normal Form Theorem, which will allow us to deal with the simplest possible deterministic system. To this end, we shall enunciate a technical lemma which we will need in the proof of the theorem. This is a simplified version (sufficient for our purposes) of Lemma 3.1 in [5].

**Lemma 4.1.** *Let $g(\theta, r) \in \mathcal{C}^l(\mathbb{T} \times B)$, where $B \subset \mathbb{R}$. Then:*

1. *If $l_0 \le l$ and $k \neq 0$, $\|g_k(r)e^{2\pi i k\theta}\|_{\mathcal{C}^{l_0}} \le |k|^{l_0 - l}\|g\|_{\mathcal{C}^l}$.*
2. *Let $g_k(r)$ be some functions that satisfy $\|\partial_{r^\alpha} g_k\|_{\mathcal{C}^0} \le M|k|^{-\alpha-2}$ for all $\alpha \le l_0$ and some $M > 0$. Then:*

$$\left\| \sum_{\substack{k\in\mathbb{Z} \\ 0<k\le d}} g_k(r)e^{2\pi i k\theta} \right\|_{\mathcal{C}^{l_0}} \le cM,$$

*for some constant $c$ depending on $l_0$.*

Let $\mathcal{R}$ be the finite set of resonances of maps (4), namely,

$$\mathcal{R} = \{p/q \in \mathbb{Q} \,:\, \gcd(p, q) = 1, |q| < d\}.$$

Denote by $\mathcal{O}(\varepsilon)$ a function whose $\mathcal{C}^0$-norm is bounded by $C\varepsilon$ for some $C$ independent of $\varepsilon$.

Define

$$E_2(\theta, r) = \mathbb{E}v(\theta, r)\,\partial_\theta S_1(\theta, r) + \mathbb{E}w(\theta, r), \quad b(r) = \int_0^1 E_2(\theta, r)d\theta, \tag{11}$$

where $S_1$ is a certain generating function defined in (21–22).

**Theorem 4.2.** *Consider the expected map* $\mathbb{E}f$ *of the map* (4)

$$\mathbb{E}f\begin{pmatrix}\theta\\ r\end{pmatrix}\longmapsto\begin{pmatrix}\theta+r+\varepsilon\mathbb{E}u(\theta,r)+\mathcal{O}_s(\varepsilon^{1+a})\\ r+\varepsilon\mathbb{E}v(\theta,r)+\varepsilon^2\mathbb{E}w(\theta,r)+\mathcal{O}_s(\varepsilon^{2+a})\end{pmatrix}.$$

*Assume that the functions* $\mathbb{E}u(\theta,r)$, $\mathbb{E}v(\theta,r)$ *and* $\mathbb{E}w(\theta,r)$ *are* $\mathcal{C}^l$, $l\geq 3$ *and* $\beta>0$ *small. Let* $0\leq s\leq l-2$. *Then there exists* $K>0$ *independent of* $\varepsilon$ *and a canonical change of variables:*

$$\begin{array}{rcl}\Phi:\mathbb{T}\times\mathbb{R} & \to & \mathbb{T}\times\mathbb{R},\\ (\tilde{\theta},\tilde{r}) & \mapsto & (\theta,r),\end{array}$$

*such that:*

- *If* $|\tilde{r}-p/q|\geq\beta$ *for all* $p/q\in\mathcal{R}$, *then:*

$$\Phi^{-1}\circ\mathbb{E}f\circ\Phi(\tilde{\theta},\tilde{r})=$$
$$\begin{pmatrix}\tilde{\theta}+\tilde{r}+\varepsilon\mathbb{E}u(\theta,r)+\varepsilon E_1(\theta,r)+\mathcal{O}_s(\varepsilon^{1+a})+\mathcal{O}_s(\varepsilon^2\beta^{-(2s+4)})\\ \tilde{r}+\varepsilon^2E_2(\tilde{\theta},\tilde{r})+\mathcal{O}_s(\varepsilon^{2+a})+\mathcal{O}_s(\varepsilon^3\beta^{-(3s+5)})\end{pmatrix},\tag{12}$$

  *where* $E_1$ *and* $E_2$ *are some* $\mathcal{C}^{l-1}$ *functions. There exists a constant* $K$ *such that for any* $0\leq s\leq l-1$ *one has:*

$$\|E_1\|_{\mathcal{C}^s}\leq K\|\mathbb{E}v\|_{\mathcal{C}^{s+1}},\qquad \|E_2\|_{\mathcal{C}^s}\leq K\beta^{-(2s+3)}.$$

  *Moreover,* $E_2$ *verifies:*

$$b(r):=\int_0^1E_2(\tilde{\theta},\tilde{r})d\tilde{\theta}=$$
$$\int_0^1\left[\partial_r\mathbb{E}v(\tilde{\theta},\tilde{r})\partial_\theta S_1(\tilde{\theta},\tilde{r})-\partial_\theta^2S_1(\tilde{\theta},\tilde{r})\left(\mathbb{E}u(\tilde{\theta},\tilde{r})-\mathbb{E}v(\tilde{\theta},\tilde{r})\right)\right]d\tilde{\theta}.\tag{13}$$

  *In particular,* $b(r)$ *satisfies* $\|b\|_{\mathcal{C}^0}\leq K$ *and in the area-preserving case (when* $\mathbb{E}u(\theta,r)=\mathbb{E}v(\theta,r)=\mathbb{E}v(\theta)$*),* $b(r)\equiv 0$.

- *If* $|\tilde{r}-p/q|\leq 2\beta$ *for a given* $p/q\in\mathcal{R}$, *then:*

$$\Phi^{-1}\circ\mathbb{E}f\circ\Phi(\tilde{\theta},\tilde{r})=\tag{14}$$
$$\begin{pmatrix}\tilde{\theta}+\tilde{r}+\varepsilon\left[\mathbb{E}u(\tilde{\theta},p/q)-\mathbb{E}v(\tilde{\theta},p/q)+\mathbb{E}v_{p,q}(\tilde{\theta},p/q)+E_3(\tilde{\theta})\right]+\mathcal{O}_s(\varepsilon^{1+a})+\mathcal{O}_s(\varepsilon^3\beta^{-(2s+4)})\\ \tilde{r}+\varepsilon\mathbb{E}v_{p,q}(\tilde{\theta},\tilde{r})+\varepsilon^2E_4(\tilde{\theta},\tilde{r})+\mathcal{O}_s(\varepsilon^{2+a})+\mathcal{O}_s(\varepsilon^3\beta^{-(3s+5)})\end{pmatrix},$$

  *where* $\mathbb{E}v_{p,q}$ *is the* $\mathcal{C}^l$ *function defined as:*

$$\mathbb{E}v_{p,q}(\tilde{\theta},\tilde{r})=\sum_{k\in\mathcal{R}_{p,q}}\mathbb{E}v^k(\tilde{r})e^{2\pi ik\tilde{\theta}},$$

*and* $E_3$ *is the* $\mathcal{C}^{l-1}$ *function:*

$$E_3(\tilde\theta) = -\sum_{k\notin\mathcal{R}_{p,q}} \frac{i(\mathbb{E}v^k)'(p/q)}{2\pi k} e^{2\pi i k\tilde\theta},$$

*where* $\mathcal{R}_{p,q} = \{k \in \mathbb{Z} \,:\, k \neq 0,\, |k| < d,\, kp/q \in \mathbb{Z}\}$.

*Moreover,* $E_4$ *is a* $\mathcal{C}^{l-1}$ *function and there exists a constant* $K$ *such that for all* $0 \le s \le l-1$ *one has:*

$$\|E_4\|_{\mathcal{C}^s} \le K\beta^{-(2s+3)}.$$

*Also,* $\Phi$ *is* $\mathcal{C}^2$*-close to the identity. More precisely, there exists a constant* $M$ *independent of* $\varepsilon$ *such that:*

$$\|\Phi - \mathrm{Id}\|_{\mathcal{C}^2} \le M\varepsilon. \tag{15}$$

*Proof.* For each $p/q \in \mathcal{R}$ we will perform a different change. Since the procedure is the same for all $p/q \in \mathcal{R}$, from now on we fix $p/q \in \mathcal{R}$. The procedure for the rest is analogous.

We will consider the canonical change defined implicitly by a given generating function $S(\theta,\tilde r) = \theta\tilde r + \varepsilon S_1(\theta,\tilde r)$, that is:

$$\begin{aligned} \tilde\theta &= \partial_{\tilde r} S(\theta,\tilde r) = \theta + \varepsilon\partial_{\tilde r} S_1(\theta,\tilde r) \\ r &= \partial_\theta S(\theta,\tilde r) = \tilde r + \varepsilon\partial_\theta S_1(\theta,\tilde r). \end{aligned}$$

We shall start by writing explicitly the first orders of the $\varepsilon$-series of $\Phi^{-1}\circ\mathbb{E}f\circ\Phi$. If $(\theta,r) = \Phi(\tilde\theta,\tilde r)$ is the change given by the generating function $S$, then one has:

$$\Phi(\tilde\theta,\tilde r) = \begin{pmatrix} \tilde\theta - \varepsilon\partial_{\tilde r}S_1(\tilde\theta,\tilde r) + \varepsilon^2\partial_\theta\partial_{\tilde r}S_1(\tilde\theta,\tilde r)\partial_{\tilde r}S_1(\tilde\theta,\tilde r) + \mathcal{O}_s(\varepsilon^3\|\partial_\theta^2\partial_{\tilde r}S_1(\partial_{\tilde r}S_1)^2\|_{\mathcal{C}^s}) \\ \tilde r + \varepsilon\partial_\theta S_1(\tilde\theta,\tilde r) - \varepsilon^2\partial_\theta^2 S_1(\tilde\theta,\tilde r)\partial_{\tilde r}S_1(\tilde\theta,\tilde r) + \mathcal{O}_s(\varepsilon^3\|\partial_\theta^3 S_1(\partial_{\tilde r}S_1)^2\|_{\mathcal{C}^s}) \end{pmatrix}. \tag{16}$$

Its inverse is given by:

$$\Phi^{-1}(\theta,r) = \begin{pmatrix} \theta + \varepsilon\partial_{\tilde r}S_1(\theta,r) - \varepsilon^2\partial_{\tilde r}^2 S_1(\theta,r)\partial_\theta S_1(\theta,r) + \mathcal{O}_s(\varepsilon^3\|\partial_{\tilde r}^3 S_1(\partial_\theta S_1)^2\|_{\mathcal{C}^s}) \\ r - \varepsilon\partial_\theta S_1(\theta,r) + \varepsilon^2\partial_\theta\partial_{\tilde r}S_1(\theta,r)\partial_\theta S_1(\theta,r) + \mathcal{O}_s(\varepsilon^3\|\partial_\theta\partial_{\tilde r}^2 S_1(\partial_\theta S_1)^2\|_{\mathcal{C}^s}) \end{pmatrix}. \tag{17}$$

Now, first we compute $\mathbb{E}f\circ\Phi(\tilde\theta,\tilde r)$. One can see that:

$$\mathbb{E}f\circ\Phi(\tilde\theta,\tilde r) = \begin{pmatrix} \tilde\theta + \tilde r + \varepsilon A_1 + \varepsilon^2 A_2 + \varepsilon^3 A_3 \\ \tilde r + \varepsilon B_1 + \varepsilon^2 B_2 + \varepsilon^3 B_3 \end{pmatrix},$$

where:

$$\begin{aligned}
A_1 &= \mathbb{E}u(\tilde{\theta},\tilde{r}) - \partial_{\tilde{r}} S_1(\tilde{\theta},\tilde{r}) + \partial_\theta S_1(\tilde{\theta},\tilde{r}) \\
A_2 &= -\partial_\theta \mathbb{E}u(\tilde{\theta},\tilde{r})\partial_{\tilde{r}} S_1(\tilde{\theta},\tilde{r}) + \partial_r \mathbb{E}u(\tilde{\theta},\tilde{r})\partial_\theta S_1(\tilde{\theta},\tilde{r}) + \partial_\theta\partial_{\tilde{r}} S_1(\tilde{\theta},\tilde{r})\partial_{\tilde{r}} S_1(\tilde{\theta},\tilde{r}) \\
&\quad -\partial_\theta^2 S_1(\tilde{\theta},\tilde{r})\partial_{\tilde{r}} S_1(\tilde{\theta},\tilde{r}), \\
A_3 &= \mathcal{O}_s(\|\partial_\theta^2\partial_{\tilde{r}} S_1(\partial_{\tilde{r}} S_1)^2\|_{\mathcal{C}^s}) + \mathcal{O}_s(\|\partial_\theta^3 S_1(\partial_{\tilde{r}} S_1)^2\|_{\mathcal{C}^s}) \\
&\quad +\mathcal{O}_s(\|\mathbb{E}u\|_{\mathcal{C}^{s+1}}\|\partial_\theta S_1\|_{\mathcal{C}^{s+1}}\|\partial_{\tilde{r}} S_1\|_{\mathcal{C}^s}) + \mathcal{O}_s(\|\mathbb{E}u\|_{\mathcal{C}^{s+2}}(\|\partial_\theta S_1\|_{\mathcal{C}^s} + \|\partial_{\tilde{r}} S_1\|_{\mathcal{C}^s})^2),
\end{aligned}$$

and:

$$\begin{aligned}
B_1 &= \mathbb{E}v(\tilde{\theta},\tilde{r}) + \partial_\theta S_1(\tilde{\theta},\tilde{r}), \\
B_2 &= -\partial_\theta \mathbb{E}v(\tilde{\theta},\tilde{r})\partial_{\tilde{r}} S_1(\tilde{\theta},\tilde{r}) + \partial_r \mathbb{E}v(\tilde{\theta},\tilde{r})\partial_\theta S_1(\tilde{\theta},\tilde{r}) \\
&\quad -\partial_\theta^2 S_1(\tilde{\theta},\tilde{r})\partial_{\tilde{r}} S_1(\tilde{\theta},\tilde{r}), \\
B_3 &= \mathcal{O}_s(\|\partial_\theta^3 S_1(\partial_{\tilde{r}} S_1)^2\|_{\mathcal{C}^s}) + \mathcal{O}_s(\|\mathbb{E}v\|_{\mathcal{C}^{s+1}}\|\partial_\theta S_1\|_{\mathcal{C}^{s+1}}\|\partial_{\tilde{r}} S_1\|_{\mathcal{C}^s}) \\
&\quad +\mathcal{O}_s(\|\mathbb{E}v\|_{\mathcal{C}^{s+2}}(\|\partial_\theta S_1\|_{\mathcal{C}^s} + \|\partial_{\tilde{r}} S_1\|_{\mathcal{C}^s})^2).
\end{aligned} \tag{18}$$

Then, using (17) one can see that:

$$\Phi^{-1}\circ\mathbb{E}f\circ\Phi(\tilde{\theta},\tilde{r}) = \begin{pmatrix} \tilde{\theta}+\tilde{r}+\varepsilon\hat{A}_1+\varepsilon^2\hat{A}_2 \\ \tilde{r}+\varepsilon\hat{B}_1+\varepsilon^2\hat{B}_2+\varepsilon^3\hat{B}_3 \end{pmatrix}, \tag{19}$$

where:

$$\begin{aligned}
\hat{A}_1 &= A_1 + \partial_{\tilde{r}} S_1(\tilde{\theta}+\tilde{r},\tilde{r}), \\
\hat{A}_2 &= A_2 + \varepsilon A_3 + \mathcal{O}_s(\|\partial_\theta\partial_{\tilde{r}} S_1 A_1\|_{\mathcal{C}^s}) + \mathcal{O}_s(\|\partial_{\tilde{r}}^2 S_1 B_1\|_{\mathcal{C}^s}) \\
&\quad +\mathcal{O}_s(\|\partial_{\tilde{r}}^2 S_1\partial_\theta S_1\|_{\mathcal{C}^s}),
\end{aligned}$$

and:

$$\begin{aligned}
\hat{B}_1 &= B_1 - \partial_\theta S_1(\tilde{\theta}+\tilde{r},\tilde{r}) \\
\hat{B}_2 &= B_2 - \partial_\theta^2 S_1(\tilde{\theta}+\tilde{r},\tilde{r})A_1 - \partial_{\tilde{r}}\partial_\theta S_1(\tilde{\theta}+\tilde{r},\tilde{r})B_1 \\
&\quad +\partial_\theta\partial_{\tilde{r}} S_1(\tilde{\theta}+\tilde{r},\tilde{r})\partial_\theta S_1(\tilde{\theta}+\tilde{r},\tilde{r}), \\
\hat{B}_3 &= B_3 + \mathcal{O}_s(\|\partial_\theta\partial_{\tilde{r}}^2 S_1(\partial_\theta S_1)^2\|_{\mathcal{C}^s}) \\
&\quad +\mathcal{O}_s(\|\partial_\theta^2 S_1(A_2+\varepsilon A_3)\|_{\mathcal{C}^s} + \|\partial_\theta\partial_{\tilde{r}} S_1 B_2\|_{\mathcal{C}^s}) \\
&\quad +\mathcal{O}_s(\|\partial_\theta^3 S_1 A_1^2\|_{\mathcal{C}^s} + \|\partial_\theta^2\partial_{\tilde{r}} S_1 A_1 B_1\|_{\mathcal{C}^s} + \|\partial_\theta\partial_{\tilde{r}}^2 S_1 B_1^2\|_{\mathcal{C}^s}) \\
&\quad +\mathcal{O}_s(\|\partial_\theta^2\partial_{\tilde{r}} S_1 A_1\partial_\theta S_1\|_{\mathcal{C}^s} + \|\partial_\theta\partial_{\tilde{r}}^2 S_1 B_1\partial_\theta S_1\|_{\mathcal{C}^s}) \\
&\quad +\mathcal{O}_s(\|\partial_\theta\partial_{\tilde{r}} S_1\partial_\theta^2 S_1 A_1\|_{\mathcal{C}^s} + \|(\partial_\theta\partial_{\tilde{r}} S_1)^2 B_1\|_{\mathcal{C}^s}).
\end{aligned} \tag{20}$$

Now that we know the terms of order $\varepsilon$ and $\varepsilon^2$ of $\Phi^{-1}\circ\mathbb{E}f\circ\Phi$, we shall proceed to find a suitable $S_1(\theta,\tilde{r})$ such that these terms are as simple as possible. More precisely, we want to simplify the second component of (19). Ideally we would like that $\hat{B}_1 = 0$. Namely, we want to solve the following equation whenever it is possible:

$$\partial_\theta S_1(\tilde{\theta}, \tilde{r}) + \mathbb{E}v(\tilde{\theta}, \tilde{r}) - \partial_\theta S_1(\tilde{\theta} + \tilde{r}, \tilde{r}) = 0.$$

One can easily find a solution of this equation by solving the corresponding equation for the Fourier coefficients. To that aim, we write $S_1$ and $\mathbb{E}v$ in their Fourier series:

$$S_1(\theta, \tilde{r}) = \sum_{k\in\mathbb{Z}} S_1^k(\tilde{r}) e^{2\pi i k\theta}, \tag{21}$$

$$\mathbb{E}v(\theta, r) = \sum_{\substack{k\in\mathbb{Z}\\ 0<|k|\le d}} \mathbb{E}v^k(r) e^{2\pi i k\theta}.$$

It is obvious that for $k > d$ and $k = 0$ we can take $S_1^k(\tilde{r}) = 0$. For $0 < k \le d$ we obtain the following homological equation for $S_1^k(\tilde{r})$:

$$2\pi i k S_1^k(\tilde{r}) \left(1 - e^{2\pi i k\tilde{r}}\right) + \mathbb{E}v^k(r) = 0. \tag{22}$$

Clearly, this equation cannot be solved if $e^{2\pi i k\tilde{r}} = 1$, i.e. if $k\tilde{r} \in \mathbb{Z}$. We note that there exists a constant $L$, independent of $\varepsilon$, $L < d^{-1}$, such that if $\tilde{r} \neq p/q$ satisfies:

$$0 < |\tilde{r} - p/q| \le L$$

then $k\tilde{r} \notin \mathbb{Z}$ for all $0 < k \le d$. Thus, restricting ourselves to the domain $|\tilde{r} - p/q| \le L$, we have that if $kp/q \notin \mathbb{Z}$ equation (22) always has a solution, and if $kp/q \in \mathbb{Z}$ this equation has a solution except at $\tilde{r} = p/q$. Moreover, in the case that the solution exists, it is equal to:

$$S_1^k(\tilde{r}) = \frac{i\mathbb{E}v^k(r)}{2\pi k\left(1 - e^{2\pi i k\tilde{r}}\right)}.$$

We will modify this solution slightly to make it well defined also at $\tilde{r} = p/q$. To this end, let us consider a $\mathcal{C}^\infty$ function $\mu(x)$ such that:

$$\mu(x) = \begin{cases} 1 & \text{if} \quad |x| \le 1, \\ 0 & \text{if} \quad |x| \ge 2, \end{cases}$$

and $0 < \mu(x) < 1$ if $x \in (1, 2)$. Then we define:

$$\mu_k(\tilde{r}) = \mu\left(\frac{1 - e^{2\pi i k\tilde{r}}}{2\pi k\beta}\right),$$

and take:

$$S_1^k(\tilde{r}) = \frac{i\mathbb{E}v^k(r)(1 - \mu_k(\tilde{r}))}{2\pi k(1 - e^{2\pi i k\tilde{r}})}. \tag{23}$$

We note that this function is well defined since the numerator is identically zero in a neighbourhood of $\tilde{r} = p/q$, the unique zero of the denominator (if it is a zero indeed, that is, if $k \in \mathcal{R}_{p,q}$). More precisely, we claim that:

$$\mu_k(\tilde{r}) = \begin{cases} 1 & \text{if} \quad k \in \mathcal{R}_{p,q} \text{ and } |\tilde{r} - p/q| \leq \beta/2, \\ 0 & \text{if} \quad k \in \mathcal{R}_{p,q} \text{ and } |\tilde{r} - p/q| \geq 3\beta, \\ 0 & \text{if} \quad k \notin \mathcal{R}_{p,q}. \end{cases} \tag{24}$$

Indeed if $k \in \mathcal{R}_{p,q}$ there exists a constant $M$ independent of $\tilde{r}$ and $\varepsilon$ such that:

$$\frac{1}{\beta}|\tilde{r} - p/q|(1 - M|\tilde{r} - p/q|) \leq \left|\frac{1 - e^{2\pi i k \tilde{r}}}{2\pi k \beta}\right| \leq \frac{1}{\beta}|\tilde{r} - p/q|(1 + M|\tilde{r} - p/q|).$$

Then, on the one hand, if $k \in \mathcal{R}_{p,q}$ and $|\tilde{r} - p/q| \leq \beta/2$ we have:

$$\left|\frac{1 - e^{2\pi i k \tilde{r}}}{2\pi k \beta}\right| \leq \frac{1}{2} + \frac{M}{4}\beta < 1,$$

for $\beta$ sufficiently small, and thus $\mu_k(\tilde{r}) = 1$. On the other hand, if $|\tilde{r} - p/q| \geq 3\beta$ then:

$$\left|\frac{1 - e^{2\pi i k \tilde{r}}}{2\pi k \beta}\right| \geq 3 - 9M\beta > 2,$$

for $\beta$ sufficiently small, and thus $\mu_k(\tilde{r}) = 0$. Finally, if $k \notin \mathcal{R}_{p,q}$ then:

$$\left|\frac{1 - e^{2\pi i k \tilde{r}}}{2\pi k \beta}\right| \geq \frac{M}{\beta} > 2$$

for $\beta$ sufficiently small and then we also have $\mu_k(\tilde{r}) = 0$.

Now we proceed to check that the first order terms of (19) take the form (12) if $|\tilde{r} - p/q| \geq 3\beta$ and (14) if $|\tilde{r} - p/q| \leq \beta/2$. On the one hand, by definitions (23) of the coefficients $S_1^k(\tilde{r})$ and (20) of $\hat{B}_1$, we have:

$$\hat{B}_1 = \sum_{0<|k|\leq d} \mu_k(\tilde{r}) \mathbb{E}v^k(\tilde{r}) e^{2\pi i k \tilde{\theta}}.$$

Then, recalling (24) we obtain:

$$\hat{B}_1 = \begin{cases} 0 & \text{if} \quad |\tilde{r} - p/q| \geq 3\beta \\ \displaystyle\sum_{k \in \mathcal{R}_{p,q}} \mathbb{E}v^k(\tilde{r}) e^{2\pi i k \tilde{\theta}} = \mathbb{E}v_{p,q}(\tilde{\theta}, \tilde{r}) & \text{if} \quad |\tilde{r} - p/q| \leq \beta/2. \end{cases} \tag{25}$$

where we have used the definition (15) of $\mathbb{E}v_{p,q}(\tilde{\theta}, \tilde{r})$. On the other hand, from the definition (23) of $S_1^k(\tilde{r})$ one can check that:

$$\begin{aligned} &-\partial_{\tilde{r}} S_1(\tilde{\theta}, \tilde{r}) + \partial_{\tilde{r}} S_1(\tilde{\theta} + \tilde{r}, \tilde{r}) \\ &= -\partial_\theta S_1(\tilde{\theta} + \tilde{r}, \tilde{r}) - \sum_{0<|k|<d} \frac{i(\mathbb{E}v^k)'(\tilde{r})(1 - \mu_k(\tilde{r})) + i\mathbb{E}v^k(\tilde{r})\mu_k'(\tilde{r})}{2\pi k} e^{2\pi i k \tilde{\theta}}. \end{aligned}$$

Recalling definitions (20) of $\hat{A}_1$ and (20) of $\hat{B}_1$, this implies that:

$$\hat{A}_1 = \mathbb{E}u(\tilde{\theta},\tilde{r}) - \mathbb{E}v(\tilde{\theta},\tilde{r}) + \hat{B}_1 \tag{26}$$

$$- \sum_{0<|k|<d} \frac{i(\mathbb{E}v^k)'(\tilde{r})(1-\mu_k(\tilde{r})) + i\mathbb{E}v^k(\tilde{r})\mu_k'(\tilde{r})}{2\pi k} e^{2\pi i k\tilde{\theta}}. \tag{27}$$

Then we use (25) and (24) again, noting that $\mu_k'(\tilde{r}) = 0$ in both regions $|\tilde{r}-p/q| \geq 3\beta$ and $|\tilde{r}-p/q| \leq \beta/2$. Moreover, we note that for $|\tilde{r}-p/q| \leq \beta/2$.

$$\mathbb{E}v_{p,q}(\tilde{\theta},\tilde{r}) = \mathbb{E}v_{p,q}(\tilde{\theta},p/q) + \mathcal{O}(\beta),$$

$$(\mathbb{E}v^k)'(\tilde{r}) = (\mathbb{E}v^k)'(p/q) + \mathcal{O}(\beta).$$

Define

$$E_1(\tilde{\theta},\tilde{r}) = -\sum_{0<|k|<d} \frac{i(\mathbb{E}v^k)'(\tilde{r})}{2\pi k} e^{2\pi i k\tilde{\theta}}. \tag{28}$$

Then the same holds for $\mathbb{E}u(\tilde{\theta},\tilde{r})$ and $\mathbb{E}v(\tilde{\theta},\tilde{r})$: recalling definition (15) of $E_3$, equation (26) yields:

$$\hat{A}_1 = \begin{cases} \mathbb{E}u(\tilde{\theta},\tilde{r}) - \mathbb{E}v(\tilde{\theta},\tilde{r}) + E_1(\tilde{\theta},\tilde{r}) & \text{if} \quad |\tilde{r}-p/q| \geq 3\beta, \\ \Delta\mathbb{E}(\tilde{\theta},p/q) + \mathbb{E}v_{p,q}(\tilde{\theta}) + E_3(\tilde{\theta}) + \mathcal{O}(\varepsilon^{1/6}) & \text{if} \quad |\tilde{r}-p/q| \leq \beta/2, \end{cases} \tag{29}$$

where $\mathbb{E}u(\tilde{\theta},p/q) - \mathbb{E}v(\tilde{\theta},p/q) = \Delta\mathbb{E}(\tilde{\theta},p/q)$. In conclusion, by (29) and (25) we obtain that the first order terms of (17) coincide with the first order terms of (12) and (14) in each region.

For the $\varepsilon^2-$terms we rename $\hat{B}_2$ in the following way:

$$E_2(\tilde{\theta},\tilde{r}) = \hat{B}_2|_{\{|\tilde{r}-p/q|\geq 3\beta\}}, \tag{30}$$

$$E_4(\tilde{\theta},\tilde{r}) = \hat{B}_2|_{\{|\tilde{r}-p/q|\leq \beta/2\}}. \tag{31}$$

Now we shall see that $E_2$ verifies (13). To avoid long notation, in the following we do not write explicitly that expressions $A_i$, $B_i$, $\hat{A}_i$ and $\hat{B}_i$ are restricted to the region $\{|\tilde{r}-p/q| \geq 3\beta\}$. We note that since in this region we have $\hat{B}_1 = 0$ by (25), recalling the definition (20) of $\hat{B}_1$ it is clear that $B_1 = \partial_\theta S_1(\tilde{\theta}+\tilde{r},\tilde{r})$. Hence, from definition (20) of $\hat{B}_2$ it is straightforward to see that:

$$\hat{B}_2 = B_2 - \partial_\theta^2 S_1(\tilde{\theta}+\tilde{r},\tilde{r})A_1. \tag{32}$$

Now we recall that $\hat{A}_1 = A_1 + \partial_{\tilde{r}} S_1(\tilde{\theta}+\tilde{r},\tilde{r})$. Then, using (29), for $|\tilde{r}-p/q| \geq 3\beta$ we obtain straightforwardly:

$$A_1 = \mathbb{E}u(\tilde{\theta},\tilde{r}) - \mathbb{E}v(\tilde{\theta},\tilde{r}) + E_1(\tilde{\theta},\tilde{r}) - \partial_{\tilde{r}} S_1(\tilde{\theta}+\tilde{r},\tilde{r}). \tag{33}$$

Using this and the definition (18) of $B_2$ in expression (32) one obtains:

$$\begin{aligned} E_2(\tilde{\theta},\tilde{r}) = \hat{B}_2|_{\{|\tilde{r}-p/q|\geq 3\beta\}} \;=\; & -\partial_\theta \mathbb{E}v(\tilde{\theta},\tilde{r})\partial_{\tilde{r}}S_1(\tilde{\theta},\tilde{r}) \\ & +\partial_r \mathbb{E}v(\tilde{\theta},\tilde{r})\partial_\theta S_1(\tilde{\theta},\tilde{r}) \\ & -\partial_\theta^2 S_1(\tilde{\theta},\tilde{r})\partial_{\tilde{r}}S_1(\tilde{\theta},\tilde{r}) \\ & -\partial_\theta^2 S_1(\tilde{\theta}+\tilde{r},\tilde{r})[\mathbb{E}u(\tilde{\theta},\tilde{r}) - \mathbb{E}v(\tilde{\theta},\tilde{r}) \\ & +E_1(\tilde{\theta},\tilde{r}) - \partial_{\tilde{r}}S_1(\tilde{\theta}+\tilde{r},\tilde{r})]. \end{aligned} \tag{34}$$

Now we show that this expression has the claimed average (13). On the one hand, it is clear that $\partial_\theta^2 S_1(\tilde{\theta},\tilde{r})\partial_{\tilde{r}}S_1(\tilde{\theta},\tilde{r})$ and $\partial_\theta^2 S_1(\tilde{\theta}+\tilde{r},\tilde{r})\partial_{\tilde{r}}S_1(\tilde{\theta}+\tilde{r},\tilde{r})$ have the same average, so:

$$\int_0^1 -\partial_\theta^2 S_1(\tilde{\theta},\tilde{r})\partial_{\tilde{r}}S_1(\tilde{\theta},\tilde{r}) + \partial_\theta^2 S_1(\tilde{\theta}+\tilde{r},\tilde{r})\partial_{\tilde{r}}S_1(\tilde{\theta}+\tilde{r},\tilde{r})d\tilde{\theta} = 0.$$

On the other hand, writing explicitly the zeroth Fourier coefficient of the product, one can see that:

$$\int_0^1 -\partial_\theta \mathbb{E}v(\tilde{\theta},\tilde{r})\partial_{\tilde{r}}S_1(\tilde{\theta},\tilde{r}) - \partial_\theta^2 S_1(\tilde{\theta}+\tilde{r},\tilde{r})E_1(\tilde{\theta},\tilde{r})d\tilde{\theta} = 0.$$

Thus, recalling (30) and using these two facts in equation (34) we obtain:

$$\int_0^1 E_2(\tilde{\theta},\tilde{r})d\tilde{\theta} =$$

$$\int_0^1 \left[\partial_r \mathbb{E}v(\tilde{\theta},\tilde{r})\partial_\theta S_1(\tilde{\theta},\tilde{r}) - \partial_\theta^2 S_1(\tilde{\theta},\tilde{r})\left(\mathbb{E}u(\tilde{\theta},\tilde{r}) - \mathbb{E}v(\tilde{\theta},\tilde{r})\right)\right] d\tilde{\theta},$$

so that (13) is proved.

We note that, from the definition (23) of the Fourier coefficients of $S_1$, it is clear that $S_1$ is $\mathcal{C}^l$ with respect to $r$. Since it just has a finite number of nonzero coefficients, it is analytic with respect to $\theta$. Then, from the definitions (30) of $E_2$ and (31) of $E_4$ and the expression (20) of $\hat{B}_2$, it is clear that both $E_2$ and $E_4$ are $\mathcal{C}^{l-1}$.

Finally we shall bound the $\mathcal{C}^0$-norms of the functions $E_2$, $b(r)$ and $E_4$ and also the error terms. To that aim, first let us bound the $\mathcal{C}^l$ norms of $S_1$ and its derivatives. We will use Lemma 4.1 and proceed similarly as in [5]. We note that:

1. If $\mu_k(\tilde{r}) \neq 1$ we have $|1 - e^{2\pi i k\tilde{r}}| > M\beta|k|$, and thus:

$$\left|\frac{1}{1-e^{2\pi i k\tilde{r}}}\right| < M^{-1}\beta^{-1}|k|^{-1}.$$

2. Then, using that $\|f\circ g\|_{\mathcal{C}^l} \leq C\|f_{|_{\mathrm{Im}(g)}}\|_{\mathcal{C}^l}\left(1+\|g\|^l_{\mathcal{C}^l}\right)$, we get that:

$$\left\|\frac{1}{1-e^{2\pi ik\tilde{r}}}\right\|_{\mathcal{C}^l} \leq M\beta^{-(l+1)}|k|^{-(l+1)},$$

for some constant $M$, not the same as item 1.

3. Using the rule for the norm of the composition again and the fact that $\|\mu\|_{\mathcal{C}^l}$ is bounded independently of $\beta$, we get:

$$\|\mu_k(\tilde{r})\|_{\mathcal{C}^l} \leq M\beta^{-l}|k|^{-l},$$

for some constant $M$, and the same bound is obtained for $\|1-\mu_k(\tilde{r})\|_{\mathcal{C}^l}$.

Using items 2 and 3 above and the fact that $\|\mathbb{E}v^k\|_{\mathcal{C}^l}$ are bounded, we get that:

$$\begin{aligned}\left\|\partial_{\tilde{r}^\alpha}\left[\frac{1-\mu_k(\tilde{r})i\mathbb{E}v^k(\tilde{r})}{2\pi k(1-e^{2\pi ik\tilde{r}})}\right]\right\|_{\mathcal{C}^0} &\leq M_1\sum_{\alpha_1+\alpha_2=\alpha}\frac{1}{2\pi|k|}\|1-\mu_k(\tilde{r})\|_{\mathcal{C}^{\alpha_1}}\left\|\frac{1}{1-e^{2\pi ik\tilde{r}}}\right\|_{\mathcal{C}^{\alpha_2}}\\ &\leq M_2\beta^{-(\alpha+1)}|k|^{-\alpha-2}.\end{aligned}$$

Then, by item 2 of Lemma 4.1, we obtain:

$$\|S_1\|_{\mathcal{C}^l} \leq M\beta^{-(l+1)}.$$

One can also see that $\|\partial_{\tilde{r}}S_1\|_{\mathcal{C}^l} \leq M\|S_1\|_{\mathcal{C}^{l+1}}$ and $\|\partial_\theta S_1\|_{\mathcal{C}^l} \leq M\|S_1\|_{\mathcal{C}^l}$. In general, one has:

$$\|\partial_\theta^n\partial_{\tilde{r}}^m S_1\|_{\mathcal{C}^l} \leq M\beta^{-(l+m+1)}. \tag{35}$$

Now, recalling definitions (30) of $E_2$ and (31) of $E_4$, and using either expression (34) or simply (20), bound (35) yields that for $0\leq s\leq l-1$ there exists some $K$ such that:

$$\|E_2\|_{\mathcal{C}^s} \leq K\beta^{-(2s+3)}, \qquad \|E_4\|_{\mathcal{C}^0} \leq K\beta^{-(2s+3)}.$$

To bound $b(r)$ we use again (35) and . Then from its definition (13) it is clear that for $0\leq s\leq l-1$ :

$$\|b\|_{\mathcal{C}^s} \leq K\beta^{-(s+1)}.$$

Similarly, and taking into account that for $n=1,2$ we have $\|\mathbb{E}u\|_{\mathcal{C}^{s+n}}\|, \|\mathbb{E}v\|_{\mathcal{C}^{s+n}}\|$ because $s\leq l-2$, one can easily bound the error terms in the equation for $\tilde{r}$:

$$\varepsilon^3\hat{B}_3 = \mathcal{O}_s(\varepsilon^3\beta^{-(3s+5)}), \tag{36}$$

and the error terms for the equation of $\tilde{\theta}$:

$$\varepsilon^2\hat{A}_2 = \mathcal{O}_s(\varepsilon^2\beta^{-(2s+4)}). \tag{37}$$

This finishes the proof for the normal forms (12) and (14) (in the latter case, we have to take into account the extra error term of order $\mathcal{O}(\varepsilon^{1+a})$ caused by the $\beta$ –error term in (29)).

To prove (15), we just need to recall (16) and use (35). Then one obtains:

$$\|\Phi - \text{Id}\|_{\mathcal{C}^2} \leq M'\varepsilon\|S_1\|_{\mathcal{C}^3}.$$

□

From now on we will consider that our deterministic system is in the normal form, and drop tildes.

# 5 Analysis of the Martingale problem in the strips of each type

## 5.1 The Totally Irrational case

First of all we note that in this case, as in the IR case, after performing the change to normal form, the $n$-th iteration of our map can be written as:

$$\begin{aligned} \theta_n &= \theta_0 + nr_0 + \mathcal{O}(n\varepsilon), \\ r_n &= r_0 + \varepsilon\sum_{k=0}^{n-1}\omega_k[v(\theta_k, r_k) + \varepsilon v_2(\theta_k, r_k)] + \varepsilon^2\sum_{k=0}^{n-1}E_2(\theta_k, r_k) + \mathcal{O}(n\varepsilon^{2+a}), \end{aligned} \tag{38}$$

where $v_2(\theta, r)$ is a given function which can be written explicitly in terms of $v(\theta, r)$ and $S_1(\theta, r)$.

Recall that $I_\beta$ is a totally irrational segment if $p/q \in I_\beta$, then $|q| > \varepsilon^{-b}$, where $0 < 2b < \beta$.

We recall that we define $b = (\beta - \rho)/2$ for a certain $0 < \rho < \beta$. In the following we shall assume that $\rho$ satisfies an extra condition, which will ensure that certain inequalities are satisfied. This inequalities involve the degree of differentiability of certain $\mathcal{C}^l$ functions. We assume that $l \geq 12$. Then we have that: $\frac{l-11}{l-1} > 0$, $\lim_{l\to\infty}\frac{l-11}{l-1} = 1$. Thus, there exists a constant $R > 0$ such that:

$$\frac{l-11}{l-1} > R > 0, \qquad \text{for all } l \geq 12. \tag{39}$$

Given $\beta$, satisfying:

$$0 < \beta \leq 1/5, \tag{40}$$

then we will take $\rho$ satisfying:

$$0 < \rho < R\beta. \tag{41}$$

**Lemma 5.1.** *Let $g$ be a $\mathcal{C}^l$ function, $l \geq 12$. Suppose $r^*$ satisfies the following condition if for some rational $p/q$ we have $|r^* - p/q| < \varepsilon^{\beta}$, then $|q| > \varepsilon^{-b}$. Then for any $A$ such that*

$$2\beta < A < (l-1)b - \beta, \quad 0 < \tau = 1 - 2A \leq \min\{A - 2\beta, (l-1)b - A - \beta, \beta\}$$

*and $\varepsilon$ small enough there is $N \leq \varepsilon^{-A}$ such that for some $K$ independent of $\varepsilon$ and any $\theta^*$ we have:*

$$\left| N \int_0^1 g(\theta, r^*) d\theta - \sum_{k=0}^{N-1} g(\theta^* + kr^*, r^*) \right| \leq K\varepsilon^{\tau+\beta}.$$

*In particular, one can choose any $0 < \beta \leq 1/5$, $A = 7\beta/3$, $\tau = A - 2\beta = \beta/3$, $b = \beta/3$.*

*Proof.* Denote $g_0(r) = \int_0^1 g(\theta, r) d\theta$. Expand $g(\theta, r)$ in its Fourier series, i.e.:

$$g(\theta, r) = g_0(r) + \sum_{m \in \mathbb{Z} \setminus \{0\}} g_m(r) e^{2\pi i m \theta}$$

for some $g_m(r) : \mathbb{R} \to \mathbb{C}$. Then we have:

$$\begin{aligned}
&\sum_{k=0}^{N-1} (g(\theta^* + kr^*, r^*) - g_0(r^*)) = \sum_{k=0}^{N-1} \sum_{m \in \mathbb{Z} \setminus \{0\}} g_m(r^*) e^{2\pi i m(\theta^* + kr^*, r^*))} \\
&= \sum_{k=0}^{N-1} \sum_{1 \leq |m| \leq [\varepsilon^{-b}]} g_m(r^*) e^{2\pi i m(\theta^* + kr^*)} + \sum_{k=0}^{N} \sum_{|m| \geq [\varepsilon^{-b}]} g_m(r^*) e^{2\pi i m(\theta^* + kr^*)} \\
&= \sum_{1 \leq |m| \leq [\varepsilon^{-b}]} g_m(r^*) e^{2\pi i m \theta^*} \sum_{k=0}^{N-1} e^{2\pi i m k r^*} + \sum_{k=0}^{N-1} \sum_{|m| \geq [\varepsilon^{-b}]} g_m(r^*) e^{2\pi i m(\theta^* + kr^*)} \\
&= \sum_{1 \leq |m| \leq [\varepsilon^{-b}]} g_m(r^*) e^{2\pi i m \theta^*} \frac{e^{2\pi i N m r^*} - 1}{e^{2\pi i m r^*} - 1} + \sum_{k=0}^{N-1} \sum_{|m| \geq [\varepsilon^{-b}]} g_m(r^*) e^{2\pi i m(\theta^* + kr^*)}.
\end{aligned} \tag{42}$$

To bound the first sum in (42) we distinguish into the following cases:

- If $r^*$ is rational $p/q$, we know that $|q| > \varepsilon^{-b}$.
  - If $|q| \leq \varepsilon^{-A}$, then pick $N = |q|$ and the first sum vanishes.
  - If $|q| > \varepsilon^{-A}$, then by definition of $r^*$ for any $s/m$ with $|m| < \varepsilon^{-b}$ we have or $|mr^* - s| > \varepsilon^{\beta}$. By the pigeon hole principle there exist integers $0 < N = \tilde{q} < \varepsilon^{-A}$ and $\tilde{p}$ such that $|\tilde{q} r^* - \tilde{p}| \leq 2\varepsilon^{A}$.

- If $r^*$ is irrational, consider a continuous fraction expansion $p_n/q_n \to r^*$ as $n \to \infty$. Choose $p'/q' = p_n/q_n$ with $n$ such that $q_{n+1} > \varepsilon^{-A}$. This implies that $|q'r^* - p'| < 1/|q_{n+1}| \le \varepsilon^A$. The same argument as above shows that for any $|m| < \varepsilon^{-b}$ we have $|mr^* - s| > \varepsilon^\beta$.

Let $N$ be as above. Then

$$\left| \sum_{1\le|m|\le[\varepsilon^{-b}]} g_m(r^*)\ e^{2\pi i m\theta^*}\ \frac{e^{2\pi i N m r^*}-1}{e^{2\pi i m r^*}-1} \right| \le 2\varepsilon^{A-\beta} \sum_{1\le|m|\le[\varepsilon^{-b}]} |g_m(r^*)|.$$

We point out that since $g(\theta, r)$ is $\mathcal{C}^l$, then its Fourier coefficients satisfy $|g_m(r^*)| \le C|m|^{-l},\ m \ne 0$. Thus we can bound the first sum in (42) by:

$$\left| \sum_{1\le|m|\le[\varepsilon^{-b}]} g_m(r^*)\ e^{2\pi i m\theta^*}\ \frac{e^{2\pi i N m r^*}-1}{e^{2\pi i m r^*}-1} \right| \tag{43}$$

$$\le \varepsilon^{A-\beta} \sum_{1\le|m|\le[\varepsilon^{-b}]} |g_m(r^*)| \ \le\ C\varepsilon^{A-\beta} \sum_{1\le|m|\le[\varepsilon^{-b}]} \frac{1}{m^2} \le K\varepsilon^{A-\beta}. \tag{44}$$

To bound the second sum we use again the bound for the Fourier coefficients $g_m(r^*)$:

$$\left| \sum_{k=0}^{N} \sum_{|m|\ge[\varepsilon^{-b}]} g_m(r^*) e^{2\pi i m(\theta+kr^*)} \right| \le N \sum_{|m|\ge[\varepsilon^{-b}]} \frac{1}{m^l} \le \\ KN\varepsilon^{(l-1)b} \le K\varepsilon^{(l-1)b-A}. \tag{45}$$

Clearly, taking $\tau = 1-2A \le \min\{A-2\beta, (l-1)b-A-\beta, \beta\}$, and substituting (43) and (45) in (42) we obtain the claim of the lemma. □

**Lemma 5.2.** *Let $\beta$ satisfy (40), and $b = (\beta-\rho)/2$ with $\rho$ satisfying (41). Let $n_\beta$ be an exit time of the process $(\theta_n, r_n)$ defined by (38) from some bounded domain $I_\beta$. Let $\delta > 0$ be small enough. Suppose that $n_\beta \ge \varepsilon^{-2(1-\beta)+\delta}$. For all $l \ge 8$ the following holds:*

1. *Given two $\mathcal{C}^l$ functions $h : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{T} \times \mathbb{R} \to \mathbb{R}$, there exists a constant $d > 0$ such that:*

$$\varepsilon^2 \sum_{k=0}^{n_\beta-1} e^{-\lambda\varepsilon^2 k} h(r_k)(g(\theta_k, r_k) - g_0(r_k)) = \mathcal{O}(\varepsilon^d),$$

*where $g_0(r) = \int_0^1 g(\theta, r) d\theta$.*

*2. If $n_\beta < \varepsilon^{-2(1-\beta)+\delta}$, then given a $\mathcal{C}^l$ function $g : \mathbb{T} \times \mathbb{R} \to \mathbb{R}$ and a collection of functions $h_k : \mathbb{R} \to \mathbb{R}$, with $\|h_k\|_{\mathcal{C}^0} \leq M$ and $\|h_{k+1} - h_k\|_{\mathcal{C}^0} \leq M\varepsilon^2$ for all $k$, there exists a constant $d > 0$ such that*

$$\varepsilon^2 \sum_{k=0}^{n_\beta - 1} h_k(r_k)(g(\theta_k, r_k) - g_0(r_k)) = \mathcal{O}(\varepsilon^{2\beta+d}),$$

*where $g_0(r) = \int_0^1 g(\theta, r)d\theta$.*

*Proof.* We shall prove both claims using Lemma 5.1. To that aim fix $2\beta < A < \min\{(l-1)b - \beta, (1-\beta)/2\}$. We note that (40) and (41) ensure that $2\beta < \min\{(l-1)b - \beta, (1-\beta)/2\}$, so there always exists such $A$.

Since we have $n_\beta \geq \varepsilon^{-(1-\beta)+\delta}$ and we have taken $A < (1-\beta)/2 < 1 - \beta - \delta$ (the second inequality being satisfied because $\delta$ is small), we have $\varepsilon^{-A} < n_\beta$. Choose $N < \varepsilon^{-A}$ from Lemma 5.1 and write $n_\beta = P_\beta N + Q_\beta$, for some integers $P_\beta$ and $0 \leq Q_\beta < N$. Then:

$$\begin{aligned} &\varepsilon^2 \left| \sum_{k=0}^{n_\beta-1} e^{-\lambda\varepsilon^2 k} h(r_k)(g(\theta_k, r_k) - g_0(r_k)) \right| \\ \leq\ &\varepsilon^2 \left| \sum_{k=0}^{P_\beta-1} \sum_{j=0}^{N-1} e^{-\lambda\varepsilon^2(kN+j)} h(r_{kN+j})(g(\theta_{kN+j}, r_{kN+j}) - g_0(r_{kN+j})) \right| \\ +\ &\varepsilon^2 \left| \sum_{j=0}^{Q_\beta-1} e^{-\lambda\varepsilon^2(P_\beta N+j)} h(r_{P_\beta N+j})(g(\theta_{P_\beta N+j}, r_{P_\beta N+j}) - g_0(r_{P_\beta N+j})) \right|. \end{aligned} \tag{46}$$

Let us prove item 1. We shall bound the two terms in the right hand side of (46) in a different way. Recall that in the normal form (12) we have

$$f_\omega \begin{pmatrix} \theta \\ r \end{pmatrix} \mapsto \begin{pmatrix} \theta + r + \varepsilon\mathbb{E}u(\theta, r) + \varepsilon\omega u(\theta, r) + \mathcal{O}_s(\varepsilon^{1+a}) \\ r + \varepsilon^2 E_2(\theta, r) + \mathcal{O}_s(\varepsilon^{2+a}). \end{pmatrix}. \tag{47}$$

On the one hand we have that for all $k \leq P_\beta$, and all $j \leq N$:

$$r_{kN+j} = r_{kN} + \mathcal{O}(N\varepsilon^2), \tag{48}$$

and:

$$\theta_{kN+j} = \theta_{kN} + jr_{kN} + \mathcal{O}(N^2\varepsilon). \tag{49}$$

Hence:

$$\begin{aligned} &e^{-\lambda\varepsilon^2(kN+j)} h(r_{kN+j})(g(\theta_{kN+j}, r_{kN+j}) - g_0(r_{kN+j})) \\ =\ &e^{-\lambda\varepsilon^2 kN} h(r_{kN})(g(\theta_{kN} + jr_{kN}, r_{kN}) - g_0(r_{kN})) + \mathcal{O}(e^{-\lambda\varepsilon^2 kN} N^2\varepsilon). \end{aligned}$$

Then:

$$\begin{aligned}&\varepsilon^2\left|\sum_{k=0}^{P_\beta-1}\sum_{j=0}^{N-1}e^{-\lambda\varepsilon^2(kN+j)}h(r_{kN+j})(g(\theta_{kN+j},r_{kN+j})-g_0(r_{kN+j}))\right|\\ \leq\ &\varepsilon^2\left|\sum_{k=0}^{P_\beta-1}e^{-\lambda\varepsilon^2kN}h(r_{kN})\sum_{j=0}^{N-1}(g(\theta_{kN}+jr_{kN},r_{kN})-g_0(r_{kN}))\right|\\ &+KN^3\varepsilon^3\sum_{k=0}^{P_\beta-1}e^{-\lambda\varepsilon^2kN}.\end{aligned}$$

Thus, using Lemma 5.1 we obtain:

$$\begin{aligned}&\varepsilon^2\left|\sum_{k=0}^{P_\beta-1}e^{-\lambda\varepsilon^2kN}h(r_{kN})\sum_{j=0}^{N-1}(g(\theta_{kN}+jr_{kN},r_{kN})-g_0(r_{kN}))\right|\\ \leq\ &K\varepsilon^{2+\tau+\beta}\sum_{k=0}^{P_\beta-1}e^{-\lambda\varepsilon^2kN}|h(r_{kN})|\leq\tilde{K}\varepsilon^{\tau+\beta},\end{aligned}$$

for some constants $K,\tilde{K}>0$. Moreover, we have:

$$KN^3\varepsilon^3\sum_{k=0}^{P_\beta-1}e^{-\lambda\varepsilon^2kN}\leq\tilde{K}\varepsilon^{1+\beta-2A}.$$

Thus:

$$\begin{aligned}&\varepsilon^2\left|\sum_{k=0}^{P_\beta-1}\sum_{j=0}^{N-1}e^{-\lambda\varepsilon^2(kN+j)}h(r_{kN+j})(g(\theta_{kN+j},r_{kN+j})-g_0(r_{kN+j}))\right|\\ &\qquad\leq K(\varepsilon^{\tau+\beta}+\varepsilon^{1+\beta-2A}).\end{aligned}\tag{50}$$

On the other hand we have:

$$\begin{aligned}&\varepsilon^2\left|\sum_{j=0}^{Q_\beta-1}e^{-\lambda\varepsilon^2(P_\beta N+j)}h(r_{P_\beta N+j})(g(\theta_{P_\beta N+j},r_{P_\beta N+j})-g_0(r_{P_\beta N+j}))\right|\\ &\leq\varepsilon^2K\sup_{(\theta,r)\in I_\beta}|h(r)(g(\theta,r)-g_0(r))|Q_\beta\leq\tilde{K}\varepsilon^{2-A}.\end{aligned}\tag{51}$$

In conclusion, using (50) and (51) in equation (46) we obtain:

$$\varepsilon^2\left|\sum_{k=0}^{n_\beta-1}e^{-\lambda\varepsilon^2k}h(r_k)(g(\theta_k,r_k)-g_0(r_k))\right|\leq K(\varepsilon^{2-A}+\varepsilon^{1+\beta-2A}+\varepsilon^{\tau+\beta}).$$

Denoting $d = \min\{2-A, 1+\beta-2A, \tau+\beta\}$, and letting $1-2A = \tau \geq \beta/3$. The proof of item 1 is complete.

Now let us prove item 2. The proof is very similar to item 1. We can use the same formula (46) substituting $e^{-\lambda\varepsilon^2 k}h(r_k)$ by $h_k(r_k)$. Again, using (48) and (49) we can write:

$$\begin{aligned} & h_{kN+j}(r_{kN+j})(g(\theta_{kN+j}, r_{kN+j}) - g_0(r_{kN+j})) \\ = \; & h_{kN}(r_{kN})(g(\theta_{kN} + jr_{kN}, r_{kN}) - g_0(r_{kN})) + \mathcal{O}(N^2\varepsilon). \end{aligned}$$

Then:

$$\begin{aligned} & \varepsilon^2 \left| \sum_{k=0}^{P_\beta - 1} \sum_{j=0}^{N-1} h_{kN+j}(r_{kN+j})(g(\theta_{kN+j}, r_{kN+j}) - g_0(r_{kN+j})) \right| \\ \leq \; & \varepsilon^2 \left| \sum_{k=0}^{P_\beta - 1} h_{kN}(r_{kN}) \sum_{j=0}^{N-1} (g(\theta_{kN} + jr_{kN}, r_{kN}) - g_0(r_{kN})) \right| + KN^3\varepsilon^3 P_\beta. \end{aligned}$$

Thus, using Lemma 5.1 we obtain:

$$\varepsilon^2 \left| \sum_{k=0}^{P_\beta - 1} h_{kN}(r_{kN}) \sum_{j=0}^{N-1} (g(\theta_{kN} + jr_{kN}, r_{kN}) - g_0(r_{kN})) \right| \leq$$

$$K\varepsilon^{2+\tau+\beta} P_\beta \leq K\varepsilon^{\tau+2\beta},$$

where we have used that $P_\beta \leq n_\beta \leq \varepsilon^{-2+2\beta-\delta} \leq \varepsilon^{-2+\beta}$. For the same reason we have $K\varepsilon^{3-2A} P_\beta \leq K\varepsilon^{1-2A+\beta}$.Thus:

$$\begin{aligned} & \varepsilon^2 \left| \sum_{k=0}^{P_\beta - 1} \sum_{j=0}^{N-1} h_{kN+j}(r_{kN+j})(g(\theta_{kN+j}, r_{kN+j}) - g_0(r_{kN+j})) \right| \leq \\ & \qquad K(\varepsilon^{\tau+2\beta} + \varepsilon^{1-2A+\beta}). \end{aligned} \tag{52}$$

On the other hand we have:

$$\begin{aligned} & \varepsilon^2 \left| \sum_{j=0}^{Q_\beta - 1} h_{P_\beta N+j}(r_{P_\beta N+j})(g(\theta_{P_\beta N+j}, r_{P_\beta N+j}) - g_0(r_{P_\beta N+j})) \right| \\ & \leq \varepsilon^2 K Q_\beta \leq \tilde{K}\varepsilon^{2-A}. \end{aligned} \tag{53}$$

In conclusion, using (52) and (53) in equation (46) we obtain:

$$\varepsilon^2 \left| \sum_{k=0}^{n_\beta - 1} e^{-\lambda\varepsilon^2 k} h_k(r_k)(g(\theta_k, r_k) - g_0(r_k)) \right| \leq K(\varepsilon^{2-A} + \varepsilon^{1-2A+\beta} + \varepsilon^{\tau+2\beta}).$$

Choosing $\tau = 1-2A > 0$ the last two terms are the same. In particular, $A < 1/2$ and $2 - A > 3/2$. Therefore, the first term is negligible. $\square$

Let $r_0$ belong to the TI case. Consider an interval $I_\beta = \{(\theta, r) \in \mathbb{T} \times \mathbb{R} : |r - r_0| \leq \varepsilon^\beta\}$, for some $0 < \beta \leq 1/5$. Denote $n_\beta \in \mathbb{N}$ the exit time from $I_\beta$, that is the first number such that $(\theta_{n_\beta}, r_{n_\beta}) \notin I_\beta$.

**Lemma 5.3.** *Let $\beta$ satisfy* (40), *and $b = (\beta - \rho)/2$ with $\rho$ satisfying* (41). *Take $f : \mathbb{R} \to \mathbb{R}$ be any $\mathcal{C}^l$ function with $l \geq 12$. Then there exists $d > 0$ such that for all $\lambda > 0$ one has:*

$$\mathbb{E}\Bigg( e^{-\lambda\varepsilon^2 n_\beta} f(r_{n_\beta}) + \varepsilon^2 \sum_{k=0}^{n_\beta - 1} e^{-\lambda\varepsilon^2 k}\left[\lambda f(r_k) - \left(b(r_k)f'(r_k) + \frac{\sigma^2(r_k)}{2} f''(r_k)\right)\right]\Bigg) - f(r_0) = \mathcal{O}(\varepsilon^{2\beta+d}),$$

*where for $E_2(\theta, r)$, defined in (11), we have*

$$b(r) = \int_0^1 E_2(\theta, r)d\theta, \qquad \sigma^2(r) = \int_0^1 v^2(\theta, r)d\theta.$$

*Proof.* Let us denote:

$$\eta = e^{-\lambda\varepsilon^2 n_\beta} f(r_{n_\beta}) + \varepsilon^2 \sum_{k=0}^{n_\beta - 1} e^{-\lambda\varepsilon^2 k}\left[\lambda f(r_k) - \left(b(r_k)f'(r_k) + \frac{\sigma^2(r_k)}{2} f''(r_k)\right)\right]. \tag{54}$$

First of all we shall use the law of total expectation. Fix a small enough $\delta > 0$. Then we have:

$$\begin{aligned}
\mathbb{E}(\eta) &= \mathbb{E}\left(\eta \,|\, \varepsilon^{-2(1-\beta)+\delta} \leq n_\beta \leq \varepsilon^{-2(1-\beta)-\delta}\right) \mathbb{P}\{\varepsilon^{-2(1-\beta)+\delta} \leq n_\beta \leq \varepsilon^{-2(1-\beta)-\delta}\} \\
&+ \mathbb{E}\left(\eta \,|\, n_\beta < \varepsilon^{-2(1-\beta)+\delta}\right) \mathbb{P}\{n_\beta < \varepsilon^{-2(1-\beta)+\delta}\} \\
&+ \mathbb{E}\left(\eta \,|\, n_\beta > \varepsilon^{-2(1-\beta)-\delta}\right) \mathbb{P}\{n_\beta > \varepsilon^{-2(1-\beta)-\delta}\}.
\end{aligned}$$

By Lemma C.2 for $\varepsilon$ sufficiently small and $c > 0$ independent of $\varepsilon$ we have

$$\mathbb{P}\{n_\beta < \varepsilon^{-2(1-\beta)+\delta}\} \leq \exp\left(-\frac{c}{\varepsilon^{2\delta}}\right). \tag{55}$$

Now we write:

$$e^{-\lambda\varepsilon^2 n_\beta} f(r_{n_\beta}) = f(r_0) + \sum_{k=0}^{n_\beta - 1}\left(e^{-\lambda\varepsilon^2(k+1)} f(r_{k+1}) - e^{-\lambda\varepsilon^2 k} f(r_k)\right).$$

Doing the Taylor expansion in each term inside the sum we get:

$$\begin{aligned} e^{-\lambda\varepsilon^2 n_\beta} f(r_{n_\beta}) &= f(r_0) + \sum_{k=0}^{n_\beta-1} \Big[ -\lambda\varepsilon^2 e^{-\lambda\varepsilon^2 k} f(r_k) + e^{-\lambda\varepsilon^2 k} f'(r_k)(r_{k+1}-r_k) \\ &\quad + \frac{1}{2} e^{-\lambda\varepsilon^2 k} f''(r_k)(r_{k+1}-r_k)^2 + \mathcal{O}(e^{-\lambda\varepsilon^2 k}\varepsilon^3) \Big]. \end{aligned}$$

Substituting this in (54) we get:

$$\begin{aligned} \eta &= f(r_0) + \sum_{k=0}^{n_\beta-1} \left[ e^{-\lambda\varepsilon^2 k} f'(r_k)(r_{k+1}-r_k) + \frac{1}{2} e^{-\lambda\varepsilon^2 k} f''(r_k)(r_{k+1}-r_k)^2 \right] \\ &\quad -\varepsilon^2 \sum_{k=0}^{n_\beta-1} e^{-\lambda\varepsilon^2 k} \left[ b(r_k) f'(r_k) + \frac{\sigma^2(r_k)}{2} f''(r_k) \right] + \sum_{k=0}^{n_\beta-1} \mathcal{O}(e^{-\lambda\varepsilon^2 k}\varepsilon^3). \quad (56) \end{aligned}$$

We note that using (38) we can write:

$$r_{k+1} - r_k = \varepsilon\omega_k [v(\theta_k, r_k) + \varepsilon v_2(\theta_k, r_k)] + \varepsilon^2 E_2(\theta_k, r_k) + \mathcal{O}(\varepsilon^{2+a}),$$

and also:

$$(r_{k+1} - r_k)^2 = \varepsilon^2 v^2(\theta_k, r_k) + \mathcal{O}(\varepsilon^3).$$

Thus we can rewrite (56) as:

$$\begin{aligned} \eta &= f(r_0) + \varepsilon \sum_{k=0}^{n_\beta-1} e^{-\lambda\varepsilon^2 k} f'(r_k)\omega_k \left[ v(\theta_k, r_k) + \varepsilon v_2(\theta_k, r_k) \right] \\ &\quad + \varepsilon^2 \sum_{k=0}^{n_\beta-1} e^{-\lambda\varepsilon^2 k} f'(r_k) \left[ E_2(\theta_k, r_k) - b(r_k) \right] \\ &\quad + \frac{\varepsilon^2}{2} \sum_{k=0}^{n_\beta-1} e^{-\lambda\varepsilon^2 k} f''(r_k) \left[ v^2(\theta_k, r_k) - \sigma^2(r_k) \right] \\ &\quad + \sum_{k=0}^{n_\beta-1} \mathcal{O}(e^{-\lambda\varepsilon^2 k}\varepsilon^{2+a}). \quad (57) \end{aligned}$$

Now we distinguish between the case $\varepsilon^{-2(1-\beta)+\delta} \leq n_\beta \leq \varepsilon^{-2(1-\beta)-\delta}$ and $n_\beta > \varepsilon^{-2(1-\beta)-\delta}$. Consider the former case. First, we show that the last term in (57) is $\mathcal{O}(\varepsilon^{\beta+d})$ for some $d > 0$. Indeed,

$$\left| \sum_{k=0}^{n_\beta-1} \mathcal{O}(e^{-\lambda\varepsilon^2 k}\varepsilon^{2+a}) \right| \leq K\varepsilon^{2+a} n_\beta \leq K\varepsilon^{2\beta+d}, \quad (58)$$

where $d = a - \delta > 0$ due to smallness of $\delta$, and $K$ is some positive constant. Now we use item 2 of Lemma 5.2 in (57) twice. First we take $h_k(r) = e^{-\lambda\varepsilon^2 k} f'(r)$ and $g(\theta, r) = E_2(\theta, r)$, and after we take $h_k(r) = e^{-\lambda\varepsilon^2 k} f''(r)$ and $g(\theta, r) = v^2(\theta, r)$. Then, recalling also (58), equation (57) for $\varepsilon^{-2(1-\beta)+\delta} \le n_\beta \le \varepsilon^{-2(1-\beta)-\delta}$ yields:

$$\eta = f(r_0) + \varepsilon \sum_{k=0}^{n_\beta - 1} e^{-\lambda\varepsilon^2 k} f'(r_k)\omega_k \left[v(\theta_k, r_k) + \varepsilon v_2(\theta_k, r_k)\right] + \mathcal{O}(\varepsilon^{2\beta+d}). \tag{59}$$

Now we focus on the case $n_\beta > \varepsilon^{-2(1-\beta)-\delta}$. The last term in (57) can be bounded by:

$$\left| \sum_{k=0}^{n_\beta - 1} \mathcal{O}(e^{-\lambda\varepsilon^2 k}\varepsilon^{2+a}) \right| \le K\varepsilon^{2+a} \sum_{k=0}^{n_\beta - 1} e^{-\lambda\varepsilon^2 k} = K\varepsilon^{2+a} \frac{1 - e^{-\lambda\varepsilon^2 n_\beta}}{1 - e^{-\lambda\varepsilon^2}} \le K_\lambda \varepsilon^a, \tag{60}$$

for some positive constants $K$ and $K_\lambda$. Similarly to the previous case, we use item 1 of Lemma 5.2 in (57) twice. First we take $h(r) = f'(r)$ and $g(\theta, r) = E_2(\theta, r)$, and after we take $h(r) = f''(r)$ and $g(\theta, r) = v^2(\theta, r)$. Using this and bound (60) in equation (57), we obtain the following bound for $n_\beta > \varepsilon^{-2(1-\beta)-\delta}$:

$$\eta = f(r_0) + \varepsilon \sum_{k=0}^{n_\beta - 1} e^{-\lambda\varepsilon^2 k} f'(r_k)\omega_k \left[v(\theta_k, r_k) + \varepsilon v_2(\theta_k, r_k)\right] + \mathcal{O}(\varepsilon^{d}). \tag{61}$$

Now we just need to note that since $\omega_k$ is independent of $r_k$ and $\theta_k$, we have for all $k \in \mathbb{N}$:

$$\begin{aligned}\mathbb{E}(\omega_k f'(r_k)[v(\theta_k, r_k) + \varepsilon v_2(\theta_k, r_k)]) &= \\ \mathbb{E}(\omega_k)\mathbb{E}(f'(r_k)[v(\theta_k, r_k) + \varepsilon v_2(\theta_k, r_k)]) &= 0,\end{aligned}$$

because $\mathbb{E}(\omega_k) = 0$. Thus, denoting $n_\varepsilon = [\varepsilon^{-2(1-\beta)+\delta}]$, if we take expectations in (59) and (61) and use the total expectation formula, it is clear that:

$$\begin{aligned}
& \mathbb{E}(\eta) - f(r_0) \\
= \ & \varepsilon \sum_{\substack{n\in\mathbb{N} \\ n \ge n_\varepsilon}} \mathbb{E}\left( \sum_{k=0}^{n-1} e^{-\lambda\varepsilon^2 k} f'(r_k)\omega_k \left[v(\theta_k, r_k) + \varepsilon v_2(\theta_k, r_k]\right) \right) \mathbb{P}\{n_\beta = n\} \\
+ \ & \mathcal{O}(\varepsilon^{2\beta+d})\mathbb{P}\{\varepsilon^{-2(1-\beta)+\delta} \le n_\beta \le \varepsilon^{-2(1-\beta)-\delta}\} + \mathcal{O}(\varepsilon^d)\mathbb{P}\{n_\beta > \varepsilon^{-2(1-\beta)-\delta}\} \\
= \ & \varepsilon \sum_{\substack{n\in\mathbb{N} \\ n \ge n_\varepsilon}} \left( \sum_{k=0}^{n-1} e^{-\lambda\varepsilon^2 k} \mathbb{E}(f'(r_k)\omega_k \left[v(\theta_k, r_k) + \varepsilon v_2(\theta_k, r_k)\right]) \right) \mathbb{P}\{n_\beta = n\} \qquad (62) \\
+ \ & \mathcal{O}(\varepsilon^{2\beta+d})\mathbb{P}\{\varepsilon^{-2(1-\beta)+\delta} \le n_\beta \le \varepsilon^{-2(1-\beta)-\delta}\} + \mathcal{O}(\varepsilon^d)\mathbb{P}\{n_\beta > \varepsilon^{-2(1-\beta)-\delta}\} \\
= \ & \mathcal{O}(\varepsilon^{2\beta+d})\mathbb{P}\{\varepsilon^{-2(1-\beta)+\delta} \le n_\beta \le \varepsilon^{-2(1-\beta)-\delta}\} + \mathcal{O}(\varepsilon^d)\mathbb{P}\{n_\beta > \varepsilon^{-2(1-\beta)-\delta}\}
\end{aligned}$$

By Lemma C.2 there exists a constant $a > 0$ such that:

$$\mathbb{P}\{n_\beta > \varepsilon^{-2(1-\beta)-\delta}\} = \mathcal{O}\left(e^{-\frac{a}{\varepsilon^\delta}}\right). \tag{63}$$

Clearly, if (63) is true then $\mathbb{P}\{n_\beta > \varepsilon^{-2(1-\beta)-\delta}\}$ is smaller than any order of $\varepsilon$ and then (62) finishes the proof of the lemma.

To prove (63), let us define $n_\delta := [\varepsilon^{-2(1-\beta)-\delta}]$. Define also $n_i := i[\varepsilon^{-2(1-\beta)}]$, $i = 0, \dots, [\varepsilon^{-\delta}]$. Clearly, if $n_\beta > \varepsilon^{-2(1-\beta)-\delta}$, then $|r_{n_{i+1}} - r_{n_i}| < 2\varepsilon^\beta$ for all $i$. In other words, we have that:

$$\begin{aligned} \mathbb{P}\{n_\beta > \varepsilon^{-2(1-\beta)-\delta}\} &\leq \mathbb{P}\{|r_{n_{i+1}} - r_{n_i}| < 2\varepsilon^\beta \text{ for all } i = 0, \dots, [\varepsilon^{-\delta}]\} \\ &= \prod_{i=0}^{[\varepsilon^{-\delta}]} \mathbb{P}\{|r_{n_{i+1}} - r_{n_i}| < 2\varepsilon^\beta\}, \end{aligned} \tag{64}$$

where in the last equality we have used that $r_{n_{i+1}} - r_{n_i}$ and $r_{n_{j+1}} - r_{n_j}$ are independent if $i \neq j$.

Now, take any $i$. Then:

$$r_{n_{i+1}} - r_{n_i} = \varepsilon \sum_{i=n_i}^{n_{i+1}-1} \omega_k v(\theta_k, r_k) + \mathcal{O}(\varepsilon^2(n_i - n_{i+1})).$$

Note that $\varepsilon^2(n_i - n_{i+1}) = \varepsilon^2[\varepsilon^{-2(1-\beta)}] \leq \varepsilon^{2\beta}$. Thus:

$$\varepsilon \left| \sum_{i=n_i}^{n_{i+1}-1} \omega_k v(\theta_k, r_k) \right| - \mathcal{O}(\varepsilon^{2\beta}) \leq |r_{n_{i+1}} - r_{n_i}|. \tag{65}$$

As a consequence, if $|r_{n_{i+1}} - r_{n_i}| \leq 2\varepsilon^\beta$ then $\varepsilon \left| \sum_{i=n_i}^{n_{i+1}-1} \omega_k v(\theta_k, r_k) \right| \leq 3\varepsilon^\beta$. Indeed, if this latter inequality does not hold, then:

$$\varepsilon \left| \sum_{i=n_i}^{n_{i+1}-1} \omega_k v(\theta_k, r_k) \right| - \mathcal{O}(\varepsilon^{2\beta}) > 3\varepsilon^\beta(1 - \mathcal{O}(\varepsilon^\beta)) \geq 2\varepsilon^\beta \geq |r_{n_{i+1}} - r_{n_i}|,$$

which is a contraditciton with (65). In other words:

$$\mathbb{P}\{|r_{n_{i+1}} - r_{n_i}| < 2\varepsilon^\beta\} \leq \mathbb{P}\left\{ \varepsilon \left| \sum_{i=n_i}^{n_{i+1}-1} \omega_k v(\theta_k, r_k) \right| \leq 3\varepsilon^\beta \right\}.$$

Now by Lemma C.2 that $(n_{i+1} - n_i)^{-1/2} \sum_{k=n_i}^{n_{i+1}-1} \omega_k v(\theta_k, r_k)$ converges in distribution to $\xi \sim \mathcal{N}(0, c^2)$ for some $c > 0$.

Thus, using that $n_{i+1} - n_i = [\varepsilon^{-2(1-\beta)}]$, as $\varepsilon \to 0$ we obtain:

$$\mathbb{P}\left\{\varepsilon\left|\sum_{i=n_i}^{n_{i+1}-1} \omega_k v(\theta_k, r_k)\right| \le 3\varepsilon^\beta\right\} = \mathbb{P}\{|\xi| \le 3\} + o(1) \le \rho < 1,$$

for some constant $\rho > 0$. Then, using this in (64) we get:

$$\mathbb{P}\{n_\beta > \varepsilon^{-2(1-\beta)-\delta}\} \le \rho^{1/\varepsilon^\delta}.$$

Defining $a = -\log\rho > 0$ (because $\rho < 1$) we obtain claim (63). □

## 5.2 The Imaginary Rational case

In this section we deal with the imaginary rational case. The ideas are basically the same as in the TI case. Recall that after performing the change to normal form the $n$-th iteration of our map can be written as:

$$\begin{aligned} \theta_n &= \theta_0 + nr_0 + \mathcal{O}(n\varepsilon), \\ r_n &= r_0 + \varepsilon \sum_{k=0}^{n-1} \omega_k[v(\theta_k, r_k) + \varepsilon v_2(\theta_k, r_k)] \\ &\quad + \varepsilon^2 \textstyle\sum_{k=0}^{n-1} E_2(\theta_k, r_k) + \mathcal{O}(n\varepsilon^{2+a}), \end{aligned} \tag{66}$$

where $v_2(\theta, r)$ is a given function which can be written explicitly in terms of $v(\theta, r)$ and $S_1(\theta, r)$.

We also recall that given an imaginary rational strip $I_\beta$ there exists a unique $r^* \in I_\beta$, with $r^* = p/q$ and $|q| < \varepsilon^{-b}$. Moreover, for all $r_0 \in I_\beta$ we have $|r_0 - r^*| \le \varepsilon^\beta$. Then by (66) for any $n \le n_\beta$ we have:

$$\begin{aligned} \theta_n &= \theta_0 + nr^* + \mathcal{O}(n\varepsilon^\beta), \\ r_n &= r^* + \mathcal{O}(\varepsilon^\beta). \end{aligned} \tag{67}$$

Define

$$g_0(\theta, r) = \frac{1}{q}\sum_{i=0}^{q-1} g(\theta + ir, r). \tag{68}$$

**Lemma 5.4.** *Let $\beta$ satisfy (40), and $b = (\beta - \rho)/2$ with $\rho$ satisfying (41). Let $n_\beta$ be an exit time of the process $(\theta_n, r_n)$ defined by (38) from some bounded domain $I_\beta$. For all $l \ge 1$ the following holds:*

1. *Given two $\mathcal{C}^l$ functions $h : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{T} \times \mathbb{R} \to \mathbb{R}$, there exists a constant $d > 0$ such that:*

$$\varepsilon^2 \sum_{k=0}^{n_\beta - 1} e^{-\lambda\varepsilon^2 k} h(r_k)(g(\theta_k, r_k) - g_0(\theta_k, r_k)) = \mathcal{O}(\varepsilon^d).$$

2. *If $n_\beta < \varepsilon^{-2+\beta}$, then given a $\mathcal{C}^l$ function $g : \mathbb{T}\times\mathbb{R}\to\mathbb{R}$ and a collection of functions $h_k : \mathbb{R}\to\mathbb{R}$, with $\|h_k\|_{\mathcal{C}^0} \le M$ and $\|h_{k+1}-h_k\|_{\mathcal{C}^0} \le M\varepsilon^2$ for all $k$, there exists a constant $d>0$ such that:*

$$\varepsilon^2 \sum_{k=0}^{n_\beta-1} h_k(r_k)(g(\theta_k,r_k)-g_0(\theta_k,r_k)) = \mathcal{O}(\varepsilon^{2\beta+d}).$$

*Proof.* Let us start with item 1. Write $n_\beta = P_\beta q + Q_\beta$, for some integers $P_\beta$ and $0 \le Q_\beta < q$. Then:

$$\begin{aligned}
& \varepsilon^2 \left| \sum_{k=0}^{n_\beta-1} e^{-\lambda\varepsilon^2 k} h(r_k)(g(\theta_k,r_k)-g_0(\theta_k,r_k)) \right| \\
\le\ & \varepsilon^2 \left| \sum_{k=0}^{P_\beta-1}\sum_{j=0}^{q-1} e^{-\lambda\varepsilon^2(kq+j)} h(r_{kq+j})(g(\theta_{kq+j},r_{kq+j})-g_0(\theta_{kq+j},r_{kq+j})) \right| \qquad (69)\\
+\ & \varepsilon^2 \left| \sum_{j=0}^{Q_\beta-1} e^{-\lambda\varepsilon^2(P_\beta q+j)} h(r_{P_\beta q+j})(g(\theta_{P_\beta q+j},r_{P_\beta q+j})-g_0(\theta_{P_\beta q+j},r_{P_\beta q+j})) \right|.
\end{aligned}$$

On the one hand, we note that by (67) and $j \le q < \varepsilon^{-b}$ we have

$$\theta_{kq+j} = \theta_{kq} + jr^* + \mathcal{O}(\varepsilon^{\beta-b}),$$

$$r_{kq+j} = r^* + \mathcal{O}(\varepsilon^{\beta}).$$

Then for all $k \le P_\beta$:

$$\begin{aligned}
& e^{-\lambda\varepsilon^2(kq+j)} h(r_{kq+j})(g(\theta_{kq+j},r_{kq+j})-g_0(\theta_{kq+j},r_{kq+j})) \\
=\ & e^{-\lambda\varepsilon^2 kq} h(r_{kq})(g(\theta_{kq}+jr^*,r^*)-g_0(\theta_{kq}+jr^*,r^*)) + \mathcal{O}(e^{-\lambda\varepsilon^2 kq}\varepsilon^{1-b}).
\end{aligned}$$

Then:

$$\begin{aligned}
& \varepsilon^2 \left| \sum_{k=0}^{P_\beta-1}\sum_{j=0}^{q-1} e^{-\lambda\varepsilon^2(kq+j)} h(r_{kq+j})(g(\theta_{kq+j},r_{kq+j})-g_0(\theta_{kq+j},r_{kq+j})) \right| \\
\le\ & \varepsilon^2 \left| \sum_{k=0}^{P_\beta-1} e^{-\lambda\varepsilon^2 kq} h(r_{kN}) \sum_{j=0}^{q-1}(g(\theta_{kq}+jr^*,r^*)-g_0(\theta_{kq}+jr^*,r^*)) \right| \\
& +K\varepsilon^{3-2b} \sum_{k=0}^{P_\beta-1} e^{-\lambda\varepsilon^2 kq}. \qquad (70)
\end{aligned}$$

Now, recalling that $r^* = p/q$, by the definition (68) of $g_0(\theta, r)$ for all $k < P_\beta$ we have:

$$\sum_{j=0}^{q-1}(g(\theta_{kq} + jr^*, r^*) - g_0(\theta_{kq} + jr^*, r^*)) = \sum_{j=0}^{q-1} g(\theta_{kq}, r^*) - qg_0(\theta_{kq}, r^*) = 0.$$

Moreover:

$$K\varepsilon^{3-2b} \sum_{k=0}^{P_\beta - 1} e^{-\lambda\varepsilon^2 kq} \leq K_\lambda \, \varepsilon^{1-2b}.$$

Using these estimates (70) yields:

$$\varepsilon^2 \left| \sum_{k=0}^{P_\beta - 1} \sum_{j=0}^{q-1} e^{-\lambda\varepsilon^2(kq+j)} h(r_{kq+j})(g(\theta_{kq+j}, r_{kq+j}) - g_0(\theta_{kq+j}, r_{kq+j})) \right| \leq K \, \varepsilon^{1-2b}. \tag{71}$$

Note that $1 - 2b = 1 - \beta + \rho > 0$, since $\beta < 1$ and $\beta - 2b = \rho > 0$.

On the other hand, we have:

$$\begin{aligned} &\varepsilon^2 \left| \sum_{j=0}^{Q_\beta - 1} e^{-\lambda\varepsilon^2(P_\beta N + j)} h(r_{P_\beta N+j})(g(_{P_\beta N+j}, r_{P_\beta N+j}) - g_0(r_{P_\beta N+j})) \right| \\ &\leq \varepsilon^2 K \sup_{(\theta,r)\in I_\beta} |h(r)(g(\theta, r) - g_0(\theta, r))| Q_\beta \leq \tilde{K}\varepsilon^{2-b}. \end{aligned} \tag{72}$$

Clearly, $2 - b > 0$. Substituting (71) and (72) in (69) yields item 1 of the Lemma.

Now let us consider item 2. If we take equation (70) and substitute $e^{-\lambda\varepsilon^2 k}h(r_k)$ by $h_k(r_k)$, we can write for all $k \leq P_\beta$:

$$\begin{aligned} &\varepsilon^2 \left| \sum_{k=0}^{P_\beta - 1} \sum_{j=0}^{q-1} h_{kq+j}(r_{kq+j})(g(\theta_{kq+j}, r_{kq+j}) - g_0(\theta_{kq+j}, r_{kq+j})) \right| \\ \leq \; &\varepsilon^2 \left| \sum_{k=0}^{P_\beta - 1} h_{kq}(r_{kq}) \sum_{j=0}^{q-1}(g(\theta_{kq} + jr^*, r^*) - g_0(\theta_{kq} + jr^*, r^*)) \right| + K\varepsilon^{3-2b}P_\beta. \end{aligned} \tag{73}$$

Again:

$$\sum_{j=0}^{q-1}(g(\theta_{kq} + jr^*, r^*) - g_0(\theta_{kq} + jr^*, r^*)) = 0,$$

and since $P_\beta < n_\beta < \varepsilon^{-2(1-\beta)-\delta}$ and $b = (\beta - \rho)/2$

$$K \, \varepsilon^{3-2b} P_\beta \leq K \, \varepsilon^{1+2\beta-2b} \leq K\varepsilon^{1+\rho+\beta}.$$

Then we have:

$$\varepsilon^2 \left| \sum_{k=0}^{P_\beta - 1} \sum_{j=0}^{q-1} e^{-\lambda\varepsilon^2(kq+j)} h(r_{kq+j})(g(\theta_{kq+j}, r_{kq+j}) - g_0(\theta_{kq+j}, r_{kq+j})) \right| \leq \\ K\,\varepsilon^{1+\rho+\beta}. \tag{74}$$

On the other hand we have:

$$\varepsilon^2 \left| \sum_{j=0}^{Q_\beta - 1} h_{P_\beta N+j}(r_{P_\beta N+j})(g(_{P_\beta N+j}, r_{P_\beta N+j}) - g_0(r_{P_\beta N+j})) \right| \\ \leq \varepsilon^2 K Q_\beta \leq \tilde{K}\varepsilon^{2-b}. \tag{75}$$

We note that $2 - b - 2\beta > 0$ for $b = (\beta - \rho)/2$ and $\beta < 4/5$ and that $1 - \beta > 0$ if $\beta < 1$. Thus, taking $\beta < 4/5$ bounds (74) and (75) prove item 2 of the Lemma with $d = \min\{2 - b - 2\beta, 1 - \beta, \rho\} > 0$. □

Let $r_0$ belong to the IR case. Consider an interval $I_\beta = \{(\theta, r) \in \mathbb{T} \times \mathbb{R} : |r - r_0| \leq \varepsilon^\beta\}$, for some $0 < \beta < 4/5$. Denote $n_\beta \in \mathbb{N}$ the exit time from $I_\beta$, that is the first number such that $(\theta_{n_\beta}, r_{n_\beta}) \notin I_\beta$.

**Lemma 5.5.** *Let $f : \mathbb{R} \to \mathbb{R}$ be any $\mathcal{C}^l$ function with $l \geq 3$. Then, there exists $d > 0$ such that for all $\lambda > 0$ one has:*

$$\mathbb{E}\Bigg(e^{-\lambda\varepsilon^2 n_\beta} f(r_{n_\beta}) + \\ \varepsilon^2 \sum_{k=0}^{n_\beta - 1} e^{-\lambda\varepsilon^2 k} \left[\lambda f(r_k) - \left(b(\theta_k, r_k) f'(r_k) + \frac{\sigma^2(\theta_k, r_k)}{2} f''(r_k)\right)\right]\Bigg) \\ - f(r_0) = \mathcal{O}(\varepsilon^{2\beta+d}),$$

*where:*

$$b(\theta, r) = \frac{1}{q}\sum_{i=0}^{q-1} E_2(\theta + ir, r), \qquad \sigma^2(\theta, r) = \frac{1}{q}\sum_{i=0}^{q-1} v^2(\theta + ir, r).$$

*Proof.* Let us fix any $0 < \delta < 1/6$. Again, denoting:

$$\eta = e^{-\lambda\varepsilon^2 n_\beta} f(r_{n_\beta}) + \\ \varepsilon^2 \sum_{k=0}^{n_\beta - 1} e^{-\lambda\varepsilon^2 k} \quad \left[\lambda f(r_k) - \left(b(\theta_k, r_k) f'(r_k) + \frac{\sigma^2(\theta_k, r_k)}{2} f''(r_k)\right)\right] \tag{76}$$

and using the law of total expectation, we have:

$$\mathbb{E}(\eta) = \mathbb{E}\left(\eta \,|\, \varepsilon^{-(1-\beta)+\delta} \le n_\beta \le \varepsilon^{-2(1-\beta)-\delta}\right) \mathbb{P}\{\varepsilon^{-2(1-\beta)+\delta} \le n_\beta \le \varepsilon^{-2(1-\beta)-\delta}\} \\ + \mathbb{E}\left(\eta \,|\, n_\beta < \varepsilon^{-(1-\beta)+\delta}\right) \mathbb{P}\{n_\beta < \varepsilon^{-2(1-\beta)+\delta}\} \\ + \mathbb{E}\left(\eta \,|\, n_\beta > \varepsilon^{-2(1-\beta)-\delta}\right) \mathbb{P}\{n_\beta > \varepsilon^{-2(1-\beta)-\delta}\}. \tag{77}$$

As in the proof of Lemma 5.3, for $\varepsilon$ sufficiently small by Lemma C.2 and some $C > 0$ independent of $\varepsilon$ we have

$$\mathbb{P}\{n_\beta < \varepsilon^{-2(1-\beta)+\delta}\} \le \exp\left(-\frac{C}{\varepsilon^{2\delta}}\right),$$

and, thus, (77) yields:

$$\mathbb{E}(\eta) = \mathbb{E}(\eta \,|\, \varepsilon^{-(1-\beta)+\delta} \le n_\beta \le \varepsilon^{-2(1-\beta)-\delta})\mathbb{P}\{\varepsilon^{-(1-\beta)+\delta} \le n_\beta \le \varepsilon^{-2(1-\beta)-\delta}\} \\ + \mathbb{E}\left(\eta \,|\, n_\beta > \varepsilon^{-2(1-\beta)-\delta}\right) \mathbb{P}\{n_\beta > \varepsilon^{-2(1-\beta)-\delta}\}. \tag{78}$$

Now, we can write:

$$e^{-\lambda\varepsilon^2 n_\beta} f(r_{n_\beta}) = f(r_0) + \sum_{k=0}^{n_\beta - 1} \Big[-\lambda\varepsilon^2 e^{-\lambda\varepsilon^2 k} f(r_k) + e^{-\lambda\varepsilon^2 k} f'(r_k)(r_{k+1} - r_k) \\ + \frac{1}{2} e^{-\lambda\varepsilon^2 k} f''(r_k)(r_{k+1} - r_k)^2 + \mathcal{O}(e^{-\lambda\varepsilon^2 k}\varepsilon^3)\Big],$$

and then (76) can be rewritten as:

$$\eta = f(r_0) + \\ \sum_{k=0}^{n_\beta - 1} \left[e^{-\lambda\varepsilon^2 k} f'(r_k)(r_{k+1} - r_k) + \frac{1}{2} e^{-\lambda\varepsilon^2 k} f''(r_k)(r_{k+1} - r_k)^2\right] \\ -\varepsilon^2 \sum_{k=0}^{n_\beta - 1} e^{-\lambda\varepsilon^2 k} \left[b(\theta_k, r_k) f'(r_k) + \frac{\sigma^2(\theta_k, r_k)}{2} f''(r_k)\right] + \sum_{k=0}^{n_\beta - 1} \mathcal{O}(e^{-\lambda\varepsilon^2 k}\varepsilon^3). \tag{79}$$

Now, using (66) we have:

$$r_{k+1} - r_k = \varepsilon\omega_k[v(\theta_k, r_k) + \varepsilon v_2(\theta_k, r_k)] + \varepsilon^2 E_2(\theta_k, r_k) + \mathcal{O}(\varepsilon^{2+a}),$$

and:

$$(r_{k+1} - r_k)^2 = \varepsilon^2 v^2(\theta_k, r_k) + \mathcal{O}(\varepsilon^3).$$

Thus, (79) writes out as:

$$\begin{aligned}\eta &= f(r_0)+\sum_{k=0}^{n_\beta-1} e^{-\lambda\varepsilon^2 k} f'(r_k)\varepsilon\omega_k\left[v(\theta_k,r_k)+\varepsilon v_2(\theta_k,r_k)\right]\\ &+\varepsilon^2\sum_{k=0}^{n_\beta-1} e^{-\lambda\varepsilon^2 k} f'(r_k)\left[E_2(\theta_k,r_k)-b(\theta_k,r_k)\right]\\ &+\frac{\varepsilon^2}{2}\sum_{k=0}^{n_\beta-1} e^{-\lambda\varepsilon^2 k} f''(r_k)\left[v^2(\theta_k,r_k)-\sigma^2(\theta_k,r_k)\right]\\ &+\sum_{k=0}^{n_\beta-1}\mathcal{O}(e^{-\lambda\varepsilon^2 k}\varepsilon^{2+a}). \end{aligned} \tag{80}$$

Now we distinguish between the cases $\varepsilon^{-2(1-\beta)+\delta}\le n_\beta\le\varepsilon^{-2(1-\beta)-\delta}$ and $n_\beta>\varepsilon^{-2(1-\beta)-\delta}$. We shall start assuming that $\varepsilon^{-2(1-\beta)+\delta}\le n_\beta\le\varepsilon^{-2(1-\beta)-\delta}$. As in the proof of Lemma 5.3 we have:

$$\left|\sum_{k=0}^{n_\beta-1}\mathcal{O}(e^{-\lambda\varepsilon^2 k}\varepsilon^{2+a})\right|\le K\varepsilon^{2+a}n_\beta\le K\varepsilon^{2\beta+a-\delta}. \tag{81}$$

We note that $1/6-\delta>0$ since we have taken $\delta<1/6$. Now we use item 2 of Lemma 5.4 in (80) twice, taking first $h_k(r)=e^{-\lambda\varepsilon^2 k}f'(r)$ and $g(\theta,r)=E_2(\theta,r)$, and after $h_k(r)=e^{-\lambda\varepsilon^2 k}f''(r)$ and $g(\theta,r)=v^2(\theta,r)$. Then, using also (81), equation (80) yields:

$$\eta=f(r_0)+\sum_{k=0}^{n_\beta-1} e^{-\lambda\varepsilon^2 k} f'(r_k)\varepsilon\omega_k\left[v(\theta_k,r_k)+\varepsilon v_2(\theta_k,r_k)\right]+\mathcal{O}(\varepsilon^{2\beta+d}), \tag{82}$$

for some suitable $d>0$.

Now turn to the case $n_\beta>\varepsilon^{-2(1-\beta)-\delta}$. The last term in (80) can be bounded by:

$$\left|\sum_{k=0}^{n_\beta-1}\mathcal{O}(e^{-\lambda\varepsilon^2 k}\varepsilon^{2+a})\right|\le K\varepsilon^{1+a}\sum_{k=0}^{n_\beta-1}e^{-\lambda\varepsilon^2 k}=K\varepsilon^{1+a}\frac{1-e^{-\lambda\varepsilon^2 n_\beta}}{1-e^{-\lambda\varepsilon^2}}\le K_\lambda\varepsilon^a, \tag{83}$$

for some positive constants $K$ and $K_\lambda$. Then, using item 1 of Lemma 5.4 twice (first with $h(r)=f'(r)$ and $g(\theta,r)=E_2(\theta,r)$, and later with $h(r)=f''(r)$ and $g(\theta,r)=v^2(\theta,r)$), we obtain the following bound for $n_\beta>\varepsilon^{-2(1-\beta)-\delta}$:

$$\eta=f(r_0)+\varepsilon\sum_{k=0}^{n_\beta-1} e^{-\lambda\varepsilon^2 k} f'(r_k)\omega_k\left[v(\theta_k,r_k)+\varepsilon v_2(\theta_k,r_k)\right]+\mathcal{O}(\varepsilon^d). \tag{84}$$

To finish the proof, we follow the same steps as in the proof of Lemma 5.3 and (78) yields:

$$\begin{aligned}
& \mathbb{E}(\eta) - f(r_0) \\
= & \sum_{n\in\mathbb{N}} \mathbb{E}\left(\sum_{k=0}^{n-1} e^{-\lambda\varepsilon^2 k} f'(r_k)\varepsilon\omega_k \left[v(\theta_k, r_k) + \varepsilon v_2(\theta_k, r_k)\right]\right) \mathbb{P}\{n_\beta = n\} \\
+ & \mathcal{O}(\varepsilon^{2\beta+d})\mathbb{P}\{\varepsilon^{-(1-\beta)+\delta} \leq n_\beta \leq \varepsilon^{-2(1-\beta)-\delta}\} + \mathcal{O}(\varepsilon^d)\mathbb{P}\{n_\beta < \varepsilon^{-(1-\beta)+\delta}\} \\
= & \mathcal{O}(\varepsilon^{2\beta+d}),
\end{aligned}$$

where by Lemma C.2 for some $C > 0$ independent of $\varepsilon$

$$\mathbb{P}\{n_\beta > \varepsilon^{-2(1-\beta)-\delta}\} = \exp\left(-\frac{C}{\varepsilon^{2\delta}}\right),$$

so that it is smaller than any power of $\varepsilon$. □

## 5.3 The Real Rational case

Here we study the system in the RR case, which is defined by:

$$|r - p/q| \leq C_1\varepsilon^\beta.$$

In this subsection we focus in the subdomain:

$$|r - p/q| \leq C_1\varepsilon^{1/2}.$$

The remaining part of the Real Rational strips are dealt with in Subsections 5.6 and 5.7.

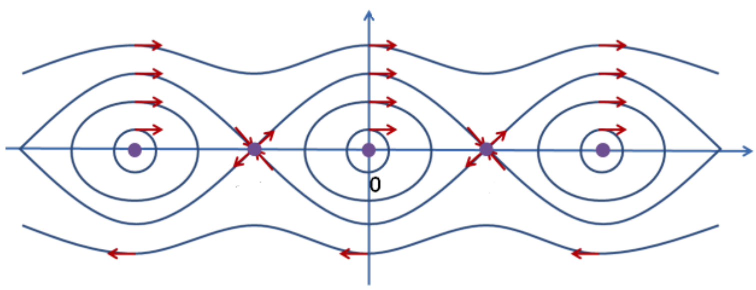



Figure 3: Level sets of the pendulum $p/q = 1/3$.

From the Normal Form Theorem, in the Real Rational strips the system takes the following form:

$$\begin{aligned} \theta_1 = &\quad \theta_0 + r_0 + \varepsilon\left[\mathbb{E}u(\theta_0, p/q) - \mathbb{E}v(\theta_0, p/q) + \mathbb{E}v_{p,q}(\theta_0, p/q) + E_3(\theta_0)\right] \\ + &\quad \varepsilon\omega_0 u(\theta_0, p/q) + \mathcal{O}(\varepsilon^{1+a}), \\ r_1 = &\quad r_0 + \varepsilon\mathbb{E}v_{p,q}(\theta_0, r_0) + \varepsilon\omega_0 v(\theta_0, r_0) \\ + &\quad \varepsilon^{3/2}\omega_0 v_2(\theta_0, r_0) + \varepsilon^{3/2}E_4(\theta_0, r_0) + \mathcal{O}(\varepsilon^{2+a}), \end{aligned} \tag{85}$$

where $v_2(\theta, r)$ can be written explicitly in terms of $v(\theta, r)$ and $S_1(\theta, r)$. The function $E_4$ is such that $\|E_4\|_{\mathcal{C}^0} \leq K$. We point out that it is a rescaled version of the function $E_4$ appearing in the Normal Form Theorem.

Recall also that:

$$\max\{\|\mathbb{E}u\|_{\mathcal{C}^0},\ \|\mathbb{E}v\|_{\mathcal{C}^0},\ \|\mathbb{E}v_{p,q}\|_{\mathcal{C}^0},\ \|E_3\|_{\mathcal{C}^0},\ \|u\|_{\mathcal{C}^0},\ \|v\|_{\mathcal{C}^0}\}\ \leq K.$$

Moreover, we have defined $v_2$ also in such a way that $\|v_2\|_{\mathcal{C}^0} \leq K$.

First, we switch to the resonant variable:

$$\hat{r} = r - p/q, \qquad 0 \leq |\hat{r}| \leq K_1\varepsilon^{1/2}.$$

With this new variable, system (85) writes out as:

$$\begin{aligned} \theta_1 = &\quad \theta_0 + p/q + \varepsilon\left[\mathbb{E}u(\theta_0, p/q) - \mathbb{E}v(\theta_0, p/q) + \mathbb{E}v_{p,q}(\theta_0, p/q) + E_3(\theta_0)\right] \\ + &\quad \hat{r}_0 + \varepsilon\omega_0 u(\theta_0, p/q) + \mathcal{O}(\varepsilon^{1+a}), \\ \hat{r}_1 = &\quad \hat{r}_0 + \varepsilon\hat{\mathbb{E}}v_{p,q}(\theta_0, \hat{r}_0) + \varepsilon^{3/2}\hat{E}_4(\theta_0, \hat{r}_0) + \varepsilon\omega_0\hat{v}(\theta_0, \hat{r}_0) \\ + &\quad \varepsilon^{3/2}\omega_0\hat{v}_2(\theta_0, \hat{r}_0) + \mathcal{O}(\varepsilon^{2+a}), \end{aligned} \tag{86}$$

where:

$$\begin{aligned} \hat{v}(\theta_0, \hat{r}_0) &= v(\theta_0, \hat{r}_0 + p/q), & \hat{v}_2(\theta_0, \hat{r}_0) &= v_2(\theta_0, \hat{r}_0 + p/q), \\ \hat{\mathbb{E}}v_{p,q}(\theta_0, \hat{r}_0) &= \mathbb{E}v_{p,q}(\theta_0, \hat{r}_0 + p/q), & \hat{E}_4(\theta_0, \hat{r}_0) &= E_4(\theta_0, \hat{r}_0 + p/q), \end{aligned}$$

From now on, we will abuse notation and drop all hats. We are interested in the

$q-$th iteration of map (86), which is given by:

$$
\begin{aligned}
\theta_q &= \theta_0 + qr_0 + \\
&+ \varepsilon \sum_{i=0}^{q-1} \left[\mathbb{E}u(\theta_i, p/q) - \mathbb{E}v(\theta_i, p/q) + (q-i)\mathbb{E}v_{p,q}(\theta_i, p/q) + E_3(\theta_i)\right] \\
&+ \varepsilon \sum_{i=0}^{q-1} \omega_i \left[u(\theta_i, p/q) + v(\theta_i, p/q)\right] + \mathcal{O}(\varepsilon^{1+a}), \\
r_q &= r_0 + \varepsilon \sum_{i=0}^{q-1} \left[\mathbb{E}v_{p,q}(\theta_i, r_i) + \varepsilon^{1/2} E_4(\theta_i, r_i)\right] \\
&+ \varepsilon \sum_{i=0}^{q-1} \omega_i \left[v(\theta_i, r_i) + \varepsilon^{1/2} v_2(\theta_i, r_i)\right] + \mathcal{O}(\varepsilon^{2+a}).
\end{aligned}
$$

Taking into account that $q$ is bounded for $0 \le i \le q$ we have:

$$
\theta_i = \theta_0 + i(p/q + r_0) + \mathcal{O}(\varepsilon), \qquad r_i = r_0 + \mathcal{O}(\varepsilon).
$$

Using this fact and that $|r| \le K_1 \varepsilon^{1/2}$ we can rewrite the last system as:

$$
\begin{aligned}
\theta_q &= \theta_0 + qr_0 + \varepsilon \mathbb{E}u^{(q)}(\theta_0) + \varepsilon u^{(q)}(\theta_0, \omega_0^q) + \mathcal{O}(\varepsilon^{3/2}), \\
r_q &= r_0 + \varepsilon \mathbb{E}v^{(q)}(\theta_0, r_0, \varepsilon) + \varepsilon v^{(q)}(\theta_0, r_0, \omega_0^q) + \varepsilon^{3/2} v_2^{(q)}(\theta_0, r_0, \omega_0^q) + \mathcal{O}(\varepsilon^2).
\end{aligned}
\tag{87}
$$

where $\omega_k^q = (\omega_{qk}, \ldots, \omega_{qk+q-1})$ and:

$$
\begin{aligned}
\mathbb{E}u^{(q)}(\theta) &= \sum_{i=0}^{q-1} \left[\mathbb{E}u(\theta + ip/q, p/q) - \mathbb{E}v(\theta + ip/q, p/q) + \right. \\
&+ \left. (q-i)\mathbb{E}v_{p,q}(\theta + ip/q, p/q) + E_3(\theta + ip/q)\right], \\
u^{(q)}(\theta, \omega_k^q) &= \sum_{i=0}^{q-1} (q-i)\omega_{qk+i} \left[u(\theta + ip/q, p/q) + v(\theta + ip/q, p/q)\right], \\
\mathbb{E}v^{(q)}(\theta, r, \varepsilon) &= \sum_{i=0}^{q-1} \left[\mathbb{E}v_{p,q}(\theta + i(p/q + r), r) + \varepsilon^{1/2} E_4(\theta + i(p/q + r), r)\right], \\
v^{(q)}(\theta, r, \omega_k^q) &= \sum_{i=0}^{q-1} \omega_{qk+i} v(\theta + i(p/q + r), r), \\
v_2^{(q)}(\theta, r, \omega_k^q) &= \sum_{i=0}^{q-1} \omega_{qk+i} v_2(\theta + i(p/q + r), r).
\end{aligned}
$$

Introduce a rescaled variable $r = R\sqrt{\varepsilon}$. Then $(\theta_1, R_1)$ are defined using system (87) with the corresponding rescaling. This can be rewritten in the following way, where we just keep the necessary $\varepsilon$-dependent terms:

$$\begin{aligned}
\theta_q &= \theta_0 + qR_0\sqrt{\varepsilon} + \varepsilon\mathbb{E}u^{(q)}(\theta_0) + \varepsilon u^{(q)}(\theta_0, \omega_0^q) + \mathcal{O}(\varepsilon^{3/2}), \\
R_q &= R_0 + \varepsilon^{1/2}\mathbb{E}v^{(q)}(\theta_0, R_0\sqrt{\varepsilon}, 0) + \varepsilon^{1/2}v^{(q)}(\theta_0, R_0\sqrt{\varepsilon}, \omega_0^q) \\
&\quad + \varepsilon v_2^{(q)}(\theta_0, 0, \omega_0^q) + \mathcal{O}(\varepsilon^{3/2}).
\end{aligned} \tag{88}$$

We note that the $n-$th iteration of map (88) is given by:

$$\begin{aligned}
\theta_{nq} &= \theta_0 + qnR_0\sqrt{\varepsilon} + \mathcal{O}(n\varepsilon), \\
R_{nq} &= R_0 + \varepsilon^{1/2}\sum_{k=0}^{n-1}\left[\mathbb{E}v^{(q)}(\theta_{qk}, R_{kq}\sqrt{\varepsilon}, 0) + v^{(q)}(\theta_{kq}, R_{kq}\sqrt{\varepsilon}, \omega_k^q)\right] + \mathcal{O}(n\varepsilon).
\end{aligned} \tag{89}$$

Moreover, using that for all $0 \leq k \leq n$ one has:

$$\theta_{kq} = \theta_0 + qkR_0\sqrt{\varepsilon} + \mathcal{O}(n\varepsilon)$$

and:

$$R_{kq} = R_0 + \mathcal{O}(n\varepsilon^{1/2})$$

system (89) can be written as:

$$\begin{aligned}
\theta_{nq} &= \theta_0 + qnR_0\sqrt{\varepsilon} + \mathcal{O}(n\varepsilon), \\
R_{nq} &= R_0 + \varepsilon^{1/2}\sum_{k=0}^{n-1}[\mathbb{E}v^{(q)}(\theta_0 + qkR_0\sqrt{\varepsilon}, R_0\sqrt{\varepsilon}, 0) \\
&\quad + v^{(q)}(\theta_0 + qkR_0\sqrt{\varepsilon}, R_0\sqrt{\varepsilon}, \omega_k^q)] + \mathcal{O}(n^2\varepsilon^{3/2}) + \mathcal{O}(n\varepsilon).
\end{aligned} \tag{90}$$

We shall subdivide the strip $I_{RR}$ in several regimes, which will be treated differently. Let $(\theta^*, 0) \in I_{RR}$ be such that:

$$\mathbb{E}v^{(q)}(\theta^*, 0, 0) = 0.$$

Fix some constants $C_1$ and $\gamma < 1/12$. We then define the following regimes:

- Regime 1:

$$D_1 = \{(\theta, R) \in I_{RR} : |\theta - \theta^*| \leq C_1\varepsilon^{\frac{1}{4}+\gamma}, |R| \leq C_1\varepsilon^{\frac{1}{4}+\gamma}\}.$$

- Regime 2:

$$D_2 = \{(\theta, R) \in I_{RR} : |\theta - \theta^*| > C_1\varepsilon^{\frac{1}{4}+\gamma}, |R| \leq C_1\varepsilon^{\frac{1}{4}+\gamma}\}.$$

- Regime 3:
$$D_3 = \{(\theta, R) \in I_{RR} \,:\, C_1\varepsilon^{\frac{1}{4}+\gamma} < |R| \le C_1\varepsilon^{\gamma}\}.$$
- Regime 4:
$$D_4 = \{(\theta, R) \in I_{RR} \,:\, C_1\varepsilon^{\gamma}\} < |R| \le C_1\}.$$

## 5.4 Regimes 1 and 2

We observe that by definition of $\theta^*$, there exists a constant $A > 0$ such that for all $(\theta, R) \in D_1$ one has:

$$\left|\mathbb{E}v^{(q)}(\theta, R\sqrt{\varepsilon}, 0)\right| \le A\varepsilon^{\frac{1}{4}+\gamma},$$

and for all $(\theta, R) \in D_2$:

$$\left|\mathbb{E}v^{(q)}(\theta, R\sqrt{\varepsilon}, 0)\right| \ge A\varepsilon^{\frac{1}{4}+\gamma}. \tag{91}$$

Moreover, we note $D_2$ has a finite number of connected components $D_2^j$, $j = 0, \dots, J$, and that, for fixed $j$, $\mathbb{E}v^{(q)}(\theta, R\sqrt{\varepsilon}, 0)$ does not change sign in $D_2^j$. Next lemmas give the exit time of regimes $D_1$ and $D_2^j$.

**Lemma 5.6.** *Let $(\theta_0, R_0) \in D_1$. Let $n^*$ denote the first exit time of the process $(\theta_{qn}, R_{qn})$ of this regime. Let $\delta > 0$ be a sufficiently small constant. Then, there exists a constant $b > 0$ such that:*

$$\mathbb{P}\{n^* > \varepsilon^{-1/2+2\gamma-\delta}\} \le e^{-\frac{b}{\varepsilon^{\delta}}}.$$

*Proof.* Let us denote $n_\gamma = [\varepsilon^{-1/2+2\gamma}]$, $n_i = in_\gamma$ and $n_\delta = [\varepsilon^{-\delta}]$. Then one has:

$$\begin{aligned}
\mathbb{P}\{n^* > \varepsilon^{-1/2+2\gamma-\delta}\} &\le \mathbb{P}\{|R_{n_{i+1}q} - R_{n_iq}| \le 2C_1\varepsilon^{1/4+\gamma} \mid \text{for all } i = 0, \dots, n_\delta\} \\
&\le \prod_{i=0}^{n_\delta} \mathbb{P}\{|R_{n_{i+1}q} - R_{n_iq}| \le 2C_1\varepsilon^{1/4+\gamma}\},
\end{aligned} \tag{92}$$

where in the last inequality we have used that $R_{n_{i+1}q} - R_{n_iq}$ and $R_{n_{j+1}q} - R_{n_jq}$ are independent for $i \ne j$. Let us assume that $(\theta_{kq}, R_{kq}) \in D_1$ for all $k = n_i, \dots, n_{i+1} - 1$. Since in $D_1$ one has $\left|\mathbb{E}v^{(q)}(\theta, R\sqrt{\varepsilon}, 0)\right| \le A\varepsilon^{\frac{1}{4}+\gamma}$, using (90) we can write:

$$\begin{aligned}
R_{n_{i+1}q} &= R_{n_iq} + \varepsilon^{1/2}\sum_{k=0}^{n_\gamma-1} v^{(q)}(\theta_{n_iq} + qkR_{n_iq}\sqrt{\varepsilon}, R_{n_iq}\sqrt{\varepsilon}, \omega^q_{n_i+k}) \\
&\quad + \mathcal{O}(n_\gamma^2\varepsilon^{3/2}) + \mathcal{O}(n_\gamma\varepsilon^{3/4+\gamma}).
\end{aligned}$$

Using that $n_\gamma = [\varepsilon^{-1/2+2\gamma}]$ yields:

$$R_{n_{i+1}q} = R_{n_iq} + \varepsilon^{1/2} \sum_{k=0}^{n_\gamma - 1} v^{(q)}(\theta_{n_iq} + qkR_{n_iq}\sqrt{\varepsilon}, R_{n_iq}\sqrt{\varepsilon}, \omega^q_{n_i+k}) + \mathcal{O}(\varepsilon^{1/4+3\gamma}). \quad (93)$$

Let us define:

$$\xi = \frac{1}{n_\gamma^{1/2}} \sum_{k=0}^{n_\gamma - 1} v^{(q)}(\theta_{n_iq} + qkR_{n_iq}\sqrt{\varepsilon}, R_{n_iq}\sqrt{\varepsilon}, \omega^q_{n_i+k}).$$

Then (93) yields:

$$\begin{aligned} \mathbb{P}\{|R_{n_{i+1}q} - R_{n_iq}| \le 2C_1\varepsilon^{1/4+\gamma}\} &\le \mathbb{P}\{|\varepsilon^{1/2}n_\gamma^{1/2}\xi + \mathcal{O}(\varepsilon^{1/4+3\gamma})| \le 2C_1\varepsilon^{1/4+\gamma}\} \\ &= \mathbb{P}\{|\xi + \mathcal{O}(\varepsilon^{2\gamma})| \le 2C_1\} \le \mathbb{P}\{|\xi| \le 3C_1\} \end{aligned}$$

We note that for $n_\gamma$ sufficiently large (i.e., $\varepsilon$ sufficiently small), the random variable $\xi$ converges in distribution to a normal random variable, that is $\xi \sim \mathcal{N}(0, \sigma^2)$, with $\sigma^2 \neq 0$. Then, there exists some constant $0 < \rho < 1$ such that:

$$\mathbb{P}\{|R_{n_{i+1}q} - R_{n_iq}| \le 2C_1\varepsilon^{1/4+\gamma}\} \le \mathbb{P}\{|\xi| \le 3C_1\} \le \rho.$$

This is valid for all $i = 0, \cdots, n_\delta$, so that using it in (92) we obtain:

$$\mathbb{P}\{n^* > \varepsilon^{-1/2+2\gamma-\delta}\} \le \rho^{[\varepsilon^{-\delta}]} \le \rho^{\varepsilon^{-\delta}},$$

and then the lemma is proved with $b = -\log\rho > 0$, since $\rho < 0$. □

**Lemma 5.7.** *Let $(\theta_0, R_0) \in D_1$. Let $n^*$ be the exit time of the process $(\theta_n, R_n)$ of $D_1$. Let $f : \mathbb{R} \to \mathbb{R}$ be any $\mathcal{C}^l$ function with $l \ge 3$. Then for all $\lambda > 0$ one has:*

$$\begin{aligned} &\mathbb{E}\Bigg( e^{-\lambda\varepsilon n^*} f(H_{n^*}) + \\ &\varepsilon \sum_{k=0}^{n^*-1} e^{-\lambda\varepsilon k} \left[ \lambda f(H_k) - b(\theta_{qk}, R_{qk}) f'(H_k) - \frac{\sigma^2(\theta_{qk}, R_{qk})}{2} f''(H_k) \right] \Bigg) \\ &\qquad\qquad -f(H_0) = \mathcal{O}(\varepsilon^{1+2\gamma-\delta}), \end{aligned}$$

*where:*

$$b(\theta, R) = F(\theta, R), \quad \sigma^2(\theta, R) = R^2_{qk} \sum_{i=0}^{q-1} v^2(\theta + ip/q, 0). \quad (94)$$

**Lemma 5.8.** *Let $j$ be fixed. Let $(\theta_0, R_0) \in D_2^j$. Let $n^*$ denote the first exit time of the process $(\theta_{qn}, R_{qn})$ of $D_2^j$. Let $\delta > 0$ be a sufficiently small constant. Then, there exist constants $M, b > 0$ such that:*

$$\mathbb{P}\{n^* > \varepsilon^{-1/2-2\gamma-\delta}\} \leq Me^{-\frac{b}{\varepsilon^{\delta/2}}}.$$

*Proof.* Recall that by (91) for all $(\theta, R) \in D_2$ we have:

$$|\mathbb{E}v^{(q)}(\theta, R, 0)| \geq A\varepsilon^{\frac{1}{4}+\gamma}, \tag{95}$$

for some constant $A > 0$. Denote $n_{\gamma,\delta} = [\varepsilon^{-1/2-2\gamma-\delta}]$. We note that:

$$\begin{aligned} & \mathbb{P}\{n^* > \varepsilon^{-1/2-2\gamma-\delta}\} \\ \leq\ & \mathbb{P}\{|R_{n_{\gamma,\delta}} - R_0| \leq 2C_1\varepsilon^{1/4+\gamma} \,|\, (\theta_{kq}, R_{kq}) \in D_2^j \text{ for all } k = 1, \ldots, n_{\gamma,\delta} - 1\}. \end{aligned}$$

Assume that $(\theta_{kq}, R_{kq}) \in D_2^j$ for $k = 1, \ldots, n_{\gamma,\delta} - 1$. Then, defining:

$$\xi = \frac{1}{n_{\gamma,\delta}^{1/2}} \sum_{k=0}^{n_{\gamma,\delta}-1} v^{(q)}(\theta_0 + qkR_0\sqrt{\varepsilon}, R_0\sqrt{\varepsilon}, \omega_k^q),$$

and using (90) we obtain:

$$R_{n_{\gamma,\delta}q} = R_0 + \varepsilon^{1/2} \sum_{k=0}^{n_{\gamma,\delta}-1} \mathbb{E}v^{(q)}(\theta qk, R_{qk}\sqrt{\varepsilon}, 0) + n_{\gamma,\delta}^{1/2}\xi + \mathcal{O}(n_{\gamma,\delta}^2\varepsilon^{3/2}) + \mathcal{O}(n_{\gamma,\delta}\varepsilon).$$

Substituting $n_{\gamma,\delta}$ by its value, we obtain:

$$R_{n_{\gamma,\delta}q} = R_0 + \varepsilon^{1/2} \sum_{k=0}^{n_{\gamma,\delta}-1} \mathbb{E}v^{(q)}(\theta qk, R_{qk}\sqrt{\varepsilon}, 0) + \varepsilon^{1/4-\gamma-\delta/2}\xi + \mathcal{O}(\varepsilon^{1/2-4\gamma-2\delta}).$$

Using the fact that $\mathbb{E}v^{(q)}(\theta, R, 0)$ does not change sign in $D_2^j$ and bound (95) we have:

$$\begin{aligned} \varepsilon^{1/2}\left|\sum_{k=0}^{n_{\gamma,\delta}-1} \mathbb{E}v^{(q)}(\theta_{qk}, R_{qk}\sqrt{\varepsilon}, 0)\right| &= \varepsilon^{1/2}\sum_{k=0}^{n_{\gamma,\delta}-1} |\mathbb{E}v^{(q)}(\theta_{qk}, R_{qk}\sqrt{\varepsilon}, 0)| \\ &\geq n_{\gamma,\delta}A\varepsilon^{3/4+\gamma} = \varepsilon^{1/4-\gamma-\delta}A. \end{aligned}$$

Then we can write:

$$|R_{qn_{\gamma,\delta}} - R_0| \geq \varepsilon^{1/4-\gamma-\delta}A - \left|\varepsilon^{1/4-\gamma-\delta/2}\xi + \mathcal{O}(\varepsilon^{1/2-4\gamma-2\delta})\right|.$$

Clearly, for sufficiently small $\varepsilon$ one has:

$$\varepsilon^{1/4-\gamma-\delta}A \geq 2C_1\varepsilon^{1/4+\gamma}.$$

If $|R_{n_{\gamma,\delta}q} - R_0| \leq \varepsilon^{1/4+\gamma}$, then necessarily one has:

$$\left|\varepsilon^{1/4-\gamma-\delta/2}\xi + \mathcal{O}(\varepsilon^{1/2-4\gamma-2\delta})\right| \geq K\varepsilon^{1/4-\gamma-\delta},$$

for some constant $K$. In other words:

$$\begin{aligned}
& \mathbb{P}\{|R_{qn_{\gamma,\delta}} - R_0| \leq 2C_1\varepsilon^{1/4+\gamma} \,|\, (\theta_{kq}, R_{kq}) \in D_2^j \text{ for all } k = 1, \ldots, n_{\gamma,\delta} - 1\} \\
\leq\ & \mathbb{P}\left\{\left|\varepsilon^{1/4-\gamma-\delta/2}\xi + \mathcal{O}(\varepsilon^{1/2-4\gamma-2\delta})\right| \geq \varepsilon^{1/4-\gamma-\delta}K\right\} \\
=\ & \mathbb{P}\left\{\left|\xi + \mathcal{O}(\varepsilon^{1/4-3\gamma-3\delta/2})\right| \geq K\varepsilon^{-\delta/2}\right\} \\
\leq\ & \mathbb{P}\left\{|\xi| \geq 2K\varepsilon^{-\delta/2}\right\},
\end{aligned}$$

where we have used that $1/4 - 3\gamma > 0$ because $\gamma < 1/12$. One can see that for $n_{\gamma,\delta}$ sufficiently large (i.e., $\varepsilon$ sufficiently small), $\xi \sim \mathcal{N}(0, \sigma^2)$. Moreover, since $\|v\|_{\mathcal{C}^0} \leq k$, one can see that $\sigma^2 \leq \tilde{K}$ for some constant $\tilde{K}$ independent of $\varepsilon$. Using that the tails of a normal random variable are exponentially small we then obtain:

$$\mathbb{P}\{|R_{qn_{\gamma,\delta}} - R_0| \leq 2C_1\varepsilon^{1/4+\gamma} \,|\, (\theta_{kq}, R_{kq}) \in D_2^j \text{ for all } k = 1, \ldots, n_{\gamma,\delta} - 1\} \leq Me^{-\frac{b}{\varepsilon^{\delta/2}}},$$

. □

**Lemma 5.9.** *Let $(\theta_0, R_0) \in D_2^j$. Let $n^*$ be the exit time of the process $(\theta_n, R_n)$ of $D_2$. Let $f : \mathbb{R} \to \mathbb{R}$ be any $\mathcal{C}^l$ function with $l \geq 3$. Then for all $\lambda > 0$ one has:*

$$\begin{aligned}
& \mathbb{E}\Bigg( e^{-\lambda\varepsilon n^*} f(H_{n^*}) + \\
& \varepsilon \sum_{k=0}^{n^*-1} e^{-\lambda\varepsilon k}\left[\lambda f(H_k) - b(\theta_{qk}, R_{qk})f'(H_k) - \frac{\sigma^2(\theta_{qk}, R_{qk})}{2} f''(H_k)\right]\Bigg) \\
& \qquad -f(H_0) = \mathcal{O}(\varepsilon^{1-2\gamma-\delta}),
\end{aligned}$$

*where:*

$$b(\theta, R) = F(\theta, R), \quad \sigma^2(\theta, R) = R_{qk}^2 \sum_{i=0}^{q-1} v^2(\theta + ip/q, 0). \tag{96}$$

## 5.5 Regimes 3 and 4

Finally we deal with $D_3$ and $D_4$. Both of them can be treated in a similar way. To study these regimes, we consider the Hamiltonian:

$$H(\theta, R) = \frac{R^2}{2} - \frac{1}{q}\int_0^\theta \mathbb{E}v^{(q)}(s, R\sqrt{\varepsilon}, 0)ds. \tag{97}$$

Let $H_n := H(\theta_{qn}, R_{qn})$. We study the process: $(\theta_{qn}, H_n) := (\theta_{qn}, H(\theta_{qn}, R_{qn}))$, where $(\theta_{qn}, R_{qn})$ is the process obtained iterating (88) $n$ times. One can see that:

$$\begin{aligned} H_1 = H_0 + & \sqrt{\varepsilon}R_0\, v^{(q)}(\theta_0, R_0\sqrt{\varepsilon}, \omega_0^q) \\ + & \qquad \varepsilon F(\theta_0, R_0) + \varepsilon G(\theta_0, R_0, \omega_0^q) + \mathcal{O}(\varepsilon^{3/2}), \end{aligned} \tag{98}$$

where $F$ and $G$ are:

$$\begin{aligned} F(\theta, R) \quad &= \quad -\frac{1}{q}\mathbb{E}v^{(q)}(\theta, 0, 0)\ \mathbb{E}u^{(q)}(\theta) \\ &- \quad \frac{1}{q}\mathbb{E}v^{(q)}(\theta, 0, 0)\int_0^\theta \partial_r \mathbb{E}v^{(q)}(s, 0, 0)ds \\ &- \quad \frac{q}{2}R^2\partial_\theta \mathbb{E}v^{(q)}(\theta, 0, 0) + \frac{1}{2}\left(\mathbb{E}v^{(q)}(\theta, 0, 0)\right)^2 \\ &+ \quad \frac{1}{2}\sum_{i=0}^{q-1} v^2(\theta + ip/q, 0), \\ G(\theta, R, \omega_k^q) \quad &= \quad -\frac{1}{q}\mathbb{E}v^{(q)}(\theta, 0, 0)\ u^{(q)}(\theta, \omega_k^q) \\ &- \quad \frac{1}{q}v^{(q)}(\theta, 0, \omega_k^q, 0)\int_0^\theta \partial_r \mathbb{E}v^{(q)}(s, 0, 0)ds \\ +\frac{1}{2}\sum_{\substack{i,j=0 \\ i\neq j}}^{q-1} \omega_{qk+i} \quad & \omega_{qk+j} \quad v(\theta + ip/q, 0)v(\theta + jp/q, 0) \\ &+ \quad \mathbb{E}v^{(q)}(\theta, 0, 0)v^{(q)}(\theta, 0, \omega_k^q) \\ &+ \quad Rv_2^{(q)}(\theta, 0, \omega_k^q). \end{aligned}$$

We note that since $|R| \le K_1$ we have:

$$\|F\|_{\mathcal{C}^0} \le K, \qquad \|G\|_{\mathcal{C}^0} \le K.$$

Moreover, one has that for all $k \ge 0$:

$$\mathbb{E}(G(\theta_{qk}, R_{qk}, \omega_k^q)) = 0.$$

In terms of the variable $H$, the $\{|R| \le K_1\}$ region can be written as:

$$I_{RR} := \{H \in \mathbb{R} \,:\, |H| \le K_3\}, \tag{99}$$

for some constant $K_3$. Now, given $a_1 > a_2 \ge 0$ we define:

$$D_{a_1 a_2} := \{(\theta, R) \in I_{RR} \,:\, C_1 \varepsilon^{a_1} < |R| \le C_1 \varepsilon^{a_2}\}.$$

Note that $D_3 = D_{a_1 a_2}$ with $a_1 = 1/4 + \gamma$ and $a_2 = \gamma$, and $D_4 = D_{a_1 a_2}$ with $a_1 = \gamma$, $a_2 = 0$.

**Lemma 5.10.** *Let $H^*$, $0 < a_1 < 1/2$, $0 \le a_2 < a_1$ and $0 < \alpha < 1/2 - a_1$ be some fixed constants. Let $(\theta_0, R_0) \in D_{a_1 a_2}$, and consider the strip:*

$$I = \{(\theta, R) \in D_{a_1 a_2} \,:\, |H(\theta, R) - H^*| \le C_1 \varepsilon^{1/2 + a_1 - \alpha}\}$$

*Let $n^*$ denote the first exit time of the process $(\theta_{qn}, R_{qn})$ of the strip $I$. Let $\delta > 0$ be a sufficiently small constant. Then, there exists a constant $b > 0$ such that:*

$$\mathbb{P}\{n^* > \varepsilon^{-2\alpha - \delta}\} \le e^{-\frac{b}{\varepsilon^\delta}}.$$

*Proof.* Let $n_\alpha = [\varepsilon^{-2\alpha}]$, $n_\delta = [\varepsilon^{-\delta}]$, and $n_i = i n_\alpha$. Clearly, one has:

$$\begin{aligned} \mathbb{P}\{n^* > \varepsilon^{-2\alpha - \delta}\} &\le \mathbb{P}\{|H_{n_{i+1}} - H_{n_i}| \le 2C_2 \varepsilon^{1/2 + a_1 - \alpha} \text{ for all i} = 0, \ldots, n_\delta - 1\} \\ &= \prod_{i=0}^{n_\delta} \mathbb{P}\{|H_{n_{i+1}} - H_{n_i}| \le 2C_2 \varepsilon^{1/2 + a_1 - \alpha}\}. \end{aligned} \tag{100}$$

By (98) one has that:

$$H_{n_{i+1}} = H_{n_i} + \varepsilon^{1/2} \sum_{k=0}^{n_\alpha - 1} R_{n_i + k}\, v^{(q)}(\theta_{n_i + k}, R_{n_i + k}\sqrt{\varepsilon}, \omega^q_{n_i + k}) + \mathcal{O}(n_\alpha \varepsilon).$$

Taking into account that:

$$\theta_{n_i + k} = \theta_{n_i} + kqR_{n_i}\sqrt{\varepsilon} + \mathcal{O}(n_\alpha \varepsilon),$$

$$R_{n_i + k} = R_{n_i} + \mathcal{O}(n_\alpha \varepsilon^{1/2}),$$

we can write:

$$H_{n_{i+1}} = H_{n_i} + \varepsilon^{1/2} \sum_{k=0}^{n_\alpha - 1} R_{n_i}\, v^{(q)}(\theta_{n_i} + kqR_{n_i}\sqrt{\varepsilon}, R_{n_i}\sqrt{\varepsilon}, \omega^q_{n_i + k}) + \mathcal{O}(n_\alpha^2 \varepsilon). \tag{101}$$

Let us define:

$$\xi = \frac{1}{n_\alpha^{1/2}} \sum_{k=0}^{n_\alpha - 1} \frac{R_{n_i}}{\varepsilon^{a_1}} v^{(q)}(\theta_{n_i} + kqR_{n_i}\sqrt{\varepsilon}, R_{n_i}\sqrt{\varepsilon}, \omega^q_{n_i+k}).$$

For $n_\alpha$ sufficiently large (i.e., for $\varepsilon$ sufficiently small), one has that $\xi$ converges in distribution to a normal random variable $\mathcal{N}(0, \sigma^2(\theta_{n_i}, R_{n_i}))$ with:

$$\sigma^2(\theta_{n_i}, R_{n_i}) = \frac{1}{n_\alpha} \sum_{k=0}^{n_\alpha - 1} \frac{R^2_{n_i}}{\varepsilon^{2a_1}} \sum_{j=0}^{q-1} v^2(\theta_{n_i} + j(p/q + \sqrt{\varepsilon}R_{n_i}), R_{n_i}).$$

Note that by assumption **[H4]** and the fact that $R_{n_i}/\varepsilon^{a_1} \geq C_1 > 0$ we have that $\sigma^2(\theta_{n_i}, R_{n_i}) \geq K > 0$ for some constant $K$. Then (101) yields:

$$\begin{aligned} H_{n_{i+1}} &= H_{n_i} + \varepsilon^{1/2+a_1} n_\alpha^{1/2} \xi + \mathcal{O}(n_\alpha^2 \varepsilon) \\ &= H_{n_i} + \varepsilon^{1/2+a_1-\alpha} \xi + \mathcal{O}(\varepsilon^{1-2\alpha}). \end{aligned}$$

Then:

$$\mathbb{P}\{|H_{n_{i+1}} - H_{n_i}| \leq 2C_2\varepsilon^{1/2+a_1-\alpha}\} = \mathbb{P}\{|\xi + \mathcal{O}(\varepsilon^{1/2-a_1-\alpha})| \leq 2C_2\} \leq \mathbb{P}\{|\xi| \leq 3C_2\},$$

where we have used that $1/2 - a_1 - \alpha > 0$. Since $\xi$ converges in distribution to $\mathcal{N}(0, \sigma^2(\theta_{n_i}, R_{n_1}))$ and $\sigma^2(\theta_{n_i}, R_{n_1}) \geq K > 0$, one has:

$$\mathbb{P}\{|H_{n_{i+1}} - H_{n_i}| \leq 2C_2\varepsilon^{1/2+a_1-\alpha}\} \leq \rho,$$

for some $0 < \rho < 1$. Using this in (100) one obtains the claim of the lemma with $b = -\log\rho > 0$. □

**Lemma 5.11.** *Let $H^*$, $0 < a_1 < 1/2$, $0 \leq a_2 < a_1$ and $0 < \alpha < 1/2 - a_1$ be some fixed constants. Let $(\theta_0, R_0) \in D_{a_1a_2}$, and consider the strip:*

$$I = \{(\theta, R) \in D_{a_1a_2} \, : \, |H(\theta, R) - H^*| \leq C_1\varepsilon^{1/2+a_1-\alpha}\}.$$

*Let $n^*$ be the exit time of the process $(\theta_n, R_n)$ of $I$. Let $f : \mathbb{R} \to \mathbb{R}$ be any $\mathcal{C}^l$ function with $l \geq 3$. Then for all $\lambda > 0$ one has:*

$$\begin{aligned} &\mathbb{E}\Bigg( e^{-\lambda\varepsilon n^*} f(H_{n^*}) + \\ &\varepsilon \sum_{k=0}^{n^*-1} e^{-\lambda\varepsilon k} \left[ \lambda f(H_k) - b(\theta_{qk}, R_{qk}) f'(H_k) - \frac{\sigma^2(\theta_{qk}, R_{qk})}{2} f''(H_k) \right] \Bigg) \\ &\qquad - f(H_0) = \mathcal{O}(\varepsilon^{3/2-2\alpha-\delta}), \end{aligned}$$

*where:*

$$b(\theta,R)=F(\theta,R),\quad \sigma^2(\theta,R)=R_{qk}^2\sum_{i=0}^{q-1}v^2(\theta+ip/q,0). \tag{102}$$

*Proof.* Denote:

$$\begin{aligned}\eta = {} & e^{-\lambda\varepsilon n^*}f(H_{n^*})+\\ & \varepsilon\sum_{k=0}^{n^*-1}e^{-\lambda\varepsilon k}\left[\lambda f(H_k)-b(\theta_{qk},R_{qk})f'(H_k)-\frac{\sigma^2(\theta_{qk},R_{qk})}{2}f''(H_k)\right].\end{aligned} \tag{103}$$

By the law of total expectation, one has:

$$\begin{aligned}\mathbb{E}(\eta) &= \mathbb{E}(\eta\,|\,n^*\le\varepsilon^{-2\alpha-\delta})\mathbb{P}\{n^*\le\varepsilon^{-2\alpha-\delta}\}\\ &\quad+\mathbb{E}(\eta\,|\,n^*>\varepsilon^{-2\alpha-\delta})\mathbb{P}\{n^*>\varepsilon^{-2\alpha-\delta}\}.\end{aligned}$$

Using Lemma 5.10, this writes out as:

$$\mathbb{E}(\eta)=\mathbb{E}(\eta\,|\,n^*\le\varepsilon^{-2\alpha-\delta})(1-e^{-\frac{b}{\varepsilon^\delta}})+\mathbb{E}(\eta\,|\,n^*>\varepsilon^{-2\alpha-\delta})e^{-\frac{b}{\varepsilon^\delta}}. \tag{104}$$

Now, writing:

$$e^{-\lambda\varepsilon n^*}f(H_{n^*})=f(H_0)+\sum_{k=0}^{n^*-1}\left[e^{-\lambda\varepsilon(k+1)}f(H_{k+1})-e^{-\lambda\varepsilon k}f(H_k)\right]$$

and expanding each term in the sum in its Taylor series, we can write:

$$\begin{aligned}e^{-\lambda\varepsilon n^*}f(H_{n^*}) &= f(H_0)+\sum_{k=0}^{n^*-1}\Big[-\lambda e^{-\lambda\varepsilon k}f(H_k)+e^{-\lambda\varepsilon k}f'(H_k)(H_{k+1}-H_k)\\ &\quad+\frac{1}{2}e^{-\lambda\varepsilon k}f''(H_k)(H_{k+1}-H_k)^2+\mathcal{O}(e^{-\lambda\varepsilon k}\varepsilon^{3/2})\Big],\end{aligned}$$

so that (103) writes out as:

$$\begin{aligned}\eta = {} & f(H_0)+\\ & \sum_{k=0}^{n^*-1}\left[e^{-\lambda\varepsilon k}f'(H_k)(H_{k+1}-H_k)+\frac{1}{2}e^{-\lambda\varepsilon k}f''(H_k)(H_{k+1}-H_k)^2\right]\\ & -\varepsilon\sum_{k=0}^{n^*-1}e^{-\lambda\varepsilon k}\left[b(\theta_{qk},R_{qk})f'(H_k)+\frac{\sigma^2(\theta_{qk},R_{qk})}{2}f''(H_k)\right]+\sum_{k=0}^{n^*-1}\mathcal{O}(e^{-\lambda\varepsilon k}\varepsilon^{3/2}).\end{aligned} \tag{105}$$

Now, using (98) it is clear that:

$$H_{k+1} - H_k = \sqrt{\varepsilon} R_{qk} v^{(q)}(\theta_{qk}, R_{qk}\sqrt{\varepsilon}, \omega_k^q) + \varepsilon F(\theta_{qk}, R_{qk}) + \varepsilon G(\theta_{qk}, R_{qk}, \omega_k^q) + \mathcal{O}(\varepsilon^{3/2}). \tag{106}$$

Moreover, we have:

$$\begin{aligned}(H_{k+1} - H_k)^2 &= \varepsilon R_{qk}^2 (v^{(q)}(\theta_{qk}, 0, \omega_k^q))^2 + \mathcal{O}(\varepsilon^{3/2}) \\ &= \varepsilon R_{qk}^2 \sum_{i=0}^{q-1} v^2(\theta_{qk} + ip/q, 0) + \varepsilon G_0(\theta_{qk}, R_{qk}, \omega_k^q) + \mathcal{O}(\varepsilon^{3/2}),\end{aligned} \tag{107}$$

where:

$$\begin{aligned}&G_0(\theta_{qk}, R_{qk}, \omega_k^q) \\ &= 2R_{qk}^2 \sum_{l=0}^{q-1} \sum_{j=l+1}^{q-1} \omega_{qk+l}\omega_{qk+j} v(\theta + lp/q, 0) v(\theta + jp/q, 0).\end{aligned}$$

We note that, since $R_{qk}$ and $\theta_{qk}$ are independent of $\omega_{qk+i}$ for all $i \geq 0$, one has $\mathbb{E}(G_0(\theta_{qk}, R_{qk}, \omega_k^q)) = 0$. Using (106) and (107), equation (105) writes out as:

$$\begin{aligned}\eta \;=\; & f(H_0) + \sqrt{\varepsilon} \sum_{k=0}^{n^*-1} e^{-\lambda\varepsilon k} f'(H_k) R_{qk} v^{(q)}(\theta_{qk}, \sqrt{\varepsilon} R_{qk}, \omega_k^q) \\ & + \varepsilon \sum_{k=0}^{n^*-1} e^{-\lambda\varepsilon k} f'(H_k) \left[F(\theta_{qk}, R_{qk}) - b(\theta_{qk}, R_{qk})\right] \\ & + \frac{\varepsilon}{2} \sum_{k=0}^{n^*-1} e^{-\lambda\varepsilon k} f''(H_k) \left[R_{qk}^2 \sum_{i=0}^{q-1} v^2(\theta_{qk} + ip/q, 0) - \sigma^2(\theta_{qk}, R_{qk})\right] \\ & + \varepsilon \sum_{k=0}^{n^*-1} e^{-\lambda\varepsilon k} \left[f'(H_k) G(\theta_{qk}, R_{qk}, \omega_k^q) + \frac{f''(H_k)}{2} G_0(\theta_{qk}, R_{qk}, \omega_k^q)\right] \\ & + \sum_{k=0}^{n^*-1} \mathcal{O}(e^{-\lambda\varepsilon k} \varepsilon^{3/2}).\end{aligned} \tag{108}$$

Now, by definition of $b(\theta, R)$ and $\sigma^2(\theta, R)$ it is clear that:

$$F(\theta_{qk}, R_{qk}) - b(\theta_{qk}, R_{qk}) \;=\; 0 \tag{109}$$

$$R_{qk}^2 \sum_{i=0}^{q-1} v^2(\theta_{qk} + ip/q, 0) - \sigma^2(\theta_{qk}, R_{qk}) \;=\; 0. \tag{110}$$

On the one hand, if $n^* \leq \varepsilon^{-2\alpha-\delta}$ the last term in (108) can be bounded by:

$$\left| \sum_{k=0}^{n^*-1} \mathcal{O}(e^{-\lambda\varepsilon k} \varepsilon^{3/2}) \right| \;\leq\; K\varepsilon^{3/2} n^* \leq K\varepsilon^{3/2-2\alpha-\delta}, \tag{111}$$

for some positive constant $K$. Then, using (109), (110) and (111) in equation (108) we obtain that for $n^* \leq \varepsilon^{-2\alpha-\delta}$ one has:

$$\begin{aligned} \eta &= f(H_0) + \sqrt{\varepsilon} \sum_{k=0}^{n^*-1} e^{-\lambda\varepsilon k} f'(H_k) R_{qk} v^{(q)}(\theta_{qk}, \sqrt{\varepsilon} R_{qk}, \omega_k^q) \\ &\quad + \varepsilon \sum_{k=0}^{n^*-1} e^{-\lambda\varepsilon k} \left[ f'(H_k) G(\theta_{qk}, R_{qk}, \omega_k^q) + \frac{f''(H_k)}{2} G_0(\theta_{qk}, R_{qk}, \omega_k^q) \right] \\ &\quad + \mathcal{O}(\varepsilon^{3/2-2\alpha-\delta}). \end{aligned} \tag{112}$$

On the other hand, if $n^* > \varepsilon^{-2\alpha-\delta}$ then:

$$\begin{aligned} \left| \sum_{k=0}^{n^*-1} \mathcal{O}(e^{-\lambda\varepsilon k} \varepsilon^{3/2}) \right| &\leq K\varepsilon^{3/2} \sum_{k=0}^{n^*-1} e^{-\lambda\varepsilon k} = K\varepsilon^{3/2} \frac{1 - e^{-\lambda\varepsilon n^*}}{1 - e^{-\lambda\varepsilon}} \\ &\leq K_\lambda \varepsilon^{1/2}, \end{aligned} \tag{113}$$

for some positive constants $K$, $K_\lambda$. Using (109), (110) and (113) in equation (108) we have that for $n^* > \varepsilon^{-2\alpha-\delta}$:

$$\begin{aligned} \eta &= f(H_0) + \sqrt{\varepsilon} \sum_{k=0}^{n^*-1} e^{-\lambda\varepsilon k} f'(H_k) R_{qk} v^{(q)}(\theta_{qk}, \sqrt{\varepsilon} R_{qk}, \omega_k^q) \\ &+ \varepsilon \sum_{k=0}^{n^*-1} e^{-\lambda\varepsilon k} \left[ f'(H_k) G(\theta_{qk}, R_{qk}, \omega_k^q) + \frac{f''(H_k)}{2} G_0(\theta_{qk}, R_{qk}, \omega_k^q) \right] + \mathcal{O}(\varepsilon^{1/2}). \end{aligned} \tag{114}$$

Thus, to finish the proof, we just need to use that:

$$\mathbb{E}\left(v^{(q)}(\theta_{qk}, \sqrt{\varepsilon} R_{qk}, \omega_k^q)\right) = \mathbb{E}\left(G_0(\theta_{qk}, R_{qk}, \omega_k^q)\right) = \mathbb{E}\left(G(\theta_{qk}, R_{qk}, \omega_k^q)\right) = 0,$$

and that $R_{qk}$ and $H_k$ are independent of $\omega_k^q$. Recalling formula (104) of $\mathbb{E}(\eta)$ and using these facts in (112) and (114), one obtains straightforwardly:

$$\mathbb{E}(\eta) - f(H_0) = \mathcal{O}(\varepsilon^{3/2-2\alpha-\delta})(1 - e^{-\frac{b}{\varepsilon^\delta}}) + \mathcal{O}(\varepsilon^{1/2}) e^{-\frac{b}{\varepsilon^\delta}} = \mathcal{O}(\varepsilon^{3/2-2\alpha-\delta}),$$

and the proof is finished. □

**Remark 5.12.** We note that there exist $H^*$ and $K_1$, $K_2$ such that:

$$D_{a_1a_2} = \{H \in \mathbb{R} \,:\, K_1\varepsilon^{2a_1} \leq |H(\theta, R) - H^*| \leq K_2\varepsilon^{2a_2}\}.$$

Thus there exist strips $I_j$ of the form:

$$I_j = \{(\theta, R) \in D_{a_1a_2} \,:\, |H(\theta, R) - H_j^*| \leq C_1\varepsilon^{1/2+a_1-\alpha}\}$$

such that, defining $N := [\varepsilon^{-1/2-a_1+\alpha+2a_2}]$, one has:

$$D_{a_1a_2} = \bigcup_{j=0}^{N} I_j.$$

Let us denote by $E$ the error terms given Lemma 5.10, that is $E = \mathcal{O}(\varepsilon^{3/2-2\alpha-\delta})$. Then the accumulated error along the domain $D_{a_1a_2}$ will be:

$$N^2E = \mathcal{O}(\varepsilon^{1/2-2a_1+4a_2-\delta}).$$

Tanking $\delta > 0$ sufficiently small, this error will be negligible if:

$$1/2 - 2a_1 + 4a_2 > 0.$$

We note that both $D_3$ and $D_4$ satisfy this condition, since in the first case one has $a_1 = 1/4 + \gamma$ and $a_2 = \gamma$, and:

$$1/2 - 2(1/4 + \gamma) + 4\gamma = 2\gamma > 0,$$

and in the second case one has $a_1 = \gamma$ and $a_2 = 0$, and:

$$1/2 - 2\gamma > 0,$$

since $\gamma < 1/12$.

## 5.6 Transition Zones type 1

Here we study the system in the RR case, in the subdomain:

$$C_1\varepsilon^{1/2} \leq |r - p/q| \leq C_2\varepsilon^{1/2-\tau},$$

for certain constants $C_1$ and $C_2$ and $\tau < 1/4$. Performing the same changes as in Section 5.3, namely $\hat{r} = r - p/q$ and $\hat{r} = R\sqrt{\varepsilon}$. Then, this region is defined by:

$$I_{TZ_1} := \{(\theta, R) \in \mathbb{T} \times \mathbb{R} \,:\, C_1 \leq |R| \leq C_2\varepsilon^{-\tau}\}.$$

As in Section 5.3, one has that:

$$\begin{aligned} \theta_{nq} &= \theta_0 + qnR_0\sqrt{\varepsilon} + \mathcal{O}(n\varepsilon), \\ R_{nq} &= R_0 + \varepsilon^{1/2}\sum_{k=0}^{n-1}[\mathbb{E}v^{(q)}(\theta_0 + qkR_0\sqrt{\varepsilon}, R_0\sqrt{\varepsilon}, 0) \\ &\quad +v^{(q)}(\theta_0 + qkR_0\sqrt{\varepsilon}, R_0\sqrt{\varepsilon}, \omega_k^q)] + \mathcal{O}(n^2\varepsilon^{3/2}) + \mathcal{O}(n\varepsilon). \end{aligned} \tag{115}$$

We consider the Hamiltonian $H$ defined in (97), which is:

$$H(\theta, R) = \frac{R^2}{2} - \frac{1}{q}\int_0^\theta \mathbb{E}v^{(q)}(s, R\sqrt{\varepsilon}, 0)ds.$$

Let $H_n := H(\theta_{qn}, R_{qn})$. Then, the process $(\theta_{qn}, H_n) := (\theta_{qn}, H(\theta_{qn}, R_{qn}))$, where $(\theta_{qn}, R_{qn})$ is the process obtained iterating (88) $n$ times, is defined in the same way as (98):

$$\begin{aligned} H_1 = H_0 + & \sqrt{\varepsilon}R_0\, v^{(q)}(\theta_0, R_0\sqrt{\varepsilon}, \omega_0^q) \\ + & \varepsilon F(\theta_0, R_0) + \varepsilon G(\theta_0, R_0, \omega_0^q) + \mathcal{O}(\varepsilon^{3/2}), \end{aligned} \tag{116}$$

where $F$ and $G$ are:

$$\begin{aligned} F(\theta, R) &= -\frac{1}{q}\mathbb{E}v^{(q)}(\theta, 0, 0)\ \mathbb{E}u^{(q)}(\theta) \\ &- \frac{1}{q}\mathbb{E}v^{(q)}(\theta, 0, 0)\int_0^\theta \partial_r \mathbb{E}v^{(q)}(s, 0, 0)ds \\ &- \frac{q}{2}R^2\partial_\theta \mathbb{E}v^{(q)}(\theta, 0, 0) + \frac{1}{2}\left(\mathbb{E}v^{(q)}(\theta, 0, 0)\right)^2 \\ &+ \frac{1}{2}\sum_{i=0}^{q-1} v^2(\theta + ip/q, 0), \\ G(\theta, R, \omega_k^q) &= -\frac{1}{q}\mathbb{E}v^{(q)}(\theta, 0, 0)\ u^{(q)}(\theta, \omega_k^q) \\ &- \frac{1}{q}v^{(q)}(\theta, 0, \omega_k^q, 0)\int_0^\theta \partial_r \mathbb{E}v^{(q)}(s, 0, 0)ds \\ +\frac{1}{2}\sum_{\substack{i,j=0 \\ i\neq j}}^{q-1} \omega_{qk+i} & \quad \omega_{qk+j} \quad v(\theta + ip/q, 0)v(\theta + jp/q, 0) \\ &+ \mathbb{E}v^{(q)}(\theta, 0, 0)v^{(q)}(\theta, 0, \omega_k^q) \\ &+ Rv_2^{(q)}(\theta, 0, \omega_k^q). \end{aligned}$$

Unlike Section (5.3), here we have $|R| \leq K_2\varepsilon^{-\tau}$ so that:

$$\|F\|_{\mathcal{C}^0} \leq K\varepsilon^{-2\tau}, \qquad \|G\|_{\mathcal{C}^0} \leq K\varepsilon^{-2\tau}.$$

Again, one has that for all $k \geq 0$:

$$\mathbb{E}(G(\theta_{qk}, R_{qk}, \omega_k^q)) = 0.$$

**Lemma 5.13.** *Let $R^* \in I_{TZ_1}$, $0 < \tau < 1/2$ be fixed constants. Define $H = H(R^*, 0)$, and let $\alpha$ be such that:*

$$0 < \alpha < 1/2 - \tau, \qquad 0 < \alpha < 1/6.$$

*Let $(\theta_0, R_0) \in I_{TZ_1}$, and consider the strip:*

$$I = \{(\theta, R) \in I_{TZ_1} \,:\, |H(\theta, R) - H^*| \leq C_2 |R^*| \varepsilon^{1/2-\alpha}\}.$$

*Let $n^*$ denote the first exit time of the process $(\theta_{qn}, R_{qn})$ of the strip $I$. Let $\delta > 0$ be a sufficiently small constant. Then, there exists a constant $b > 0$ such that:*

$$\mathbb{P}\{n^* > \varepsilon^{-2\alpha-\delta}\} \leq e^{-\frac{b}{\varepsilon^\delta}}.$$

*Proof.* Let $n_\alpha = [\varepsilon^{-2\alpha}]$, $n_\delta = [\varepsilon^{-\delta}]$, and $n_i = i n_\alpha$. Clearly, one has:

$$\begin{aligned}\mathbb{P}\{n^* > \varepsilon^{-2\alpha-\delta}\} &\leq \mathbb{P}\{|H_{n_{i+1}} - H_{n_i}| \leq 2C_2 |R^*| \varepsilon^{1/2-\alpha} \text{ for all i} = 0, \dots, n_\delta - 1\} \\ &= \prod_{i=0}^{n_\delta} \mathbb{P}\{|H_{n_{i+1}} - H_{n_i}| \leq 2C_2 |R^*| \varepsilon^{1/2-\alpha}\}. \qquad (117)\end{aligned}$$

Since $|R^*| \geq C_1 > 0$, one can easily see that there exist two constants $K_1, K_2 > 0$ such that:

$$K_1 |R^*(R - R^*)| \leq |H - H^*| \leq K_2 |R^*(R - R^*)|. \qquad (118)$$

This implies that there exists two constants $K_1, K_2 > 0$ (not necessarily the same) such that:

$$K_1 \leq |F(\theta, R)| \leq K_2 |R^*|^2, \qquad K_1 \leq |G(\theta, R)| \leq K_2 |R^*|. \qquad (119)$$

By (116) and (119) one has that:

$$H_{n_{i+1}} = H_{n_i} + \varepsilon^{1/2} \sum_{k=0}^{n_\alpha - 1} R_{n_i+k}\, v^{(q)}(\theta_{n_i+k}, R_{n_i+k}\sqrt{\varepsilon}, \omega^q_{n_i+k}) + \mathcal{O}(n_\alpha |R^*|^2 \varepsilon).$$

By (115) we have:

$$\begin{aligned}\theta_{n_i+k} &= \theta_{n_i} + kqR_{n_i}\sqrt{\varepsilon} + \mathcal{O}(n_\alpha \varepsilon), \\ R_{n_i+k} &= R_{n_i} + \mathcal{O}(n_\alpha \varepsilon^{1/2}),\end{aligned}$$

so that we can write:

$$H_{n_{i+1}} = H_{n_i} + \varepsilon^{1/2} \sum_{k=0}^{n_\alpha - 1} R_{n_i}\, v^{(q)}(\theta_{n_i} + kqR_{n_i}\sqrt{\varepsilon}, R_{n_i}\sqrt{\varepsilon}, \omega^q_{n_i+k}) + \mathcal{O}(n_\alpha^2 \varepsilon) + \mathcal{O}(n_\alpha |R^*|^2 \varepsilon). \qquad (120)$$

Let us define:

$$\xi = \frac{1}{n_\alpha^{1/2}} \sum_{k=0}^{n_\alpha-1} \frac{R_{n_i}}{R^*} v^{(q)}(\theta_{n_i} + kqR_{n_i}\sqrt{\varepsilon}, R_{n_i}\sqrt{\varepsilon}, \omega^q_{n_i+k}).$$

We note that for $R_{n_i} \in I$ one has:

$$0 < K_1 \leq \left| \frac{R_{n_i}}{R^*} \right| \leq K_2 \tag{121}$$

for some constants $K_1$, $K_2$. For $n_\alpha$ sufficiently large (i.e., for $\varepsilon$ sufficiently small), one has that $\xi$ converges in distribution to a normal random variable $\mathcal{N}(0, \sigma^2(\theta_{n_i}, R_{n_i}))$ with:

$$\sigma^2(\theta_{n_i}, R_{n_i}) = \frac{1}{n_\alpha} \sum_{k=0}^{n_\alpha-1} \frac{R^2_{n_i}}{(R^*)^2} \sum_{j=0}^{q-1} v^2(\theta_{n_i} + j(p/q + \sqrt{\varepsilon}R_{n_i}), R_{n_i}).$$

Note that by assumption [**H4**] and (121) we have that $\sigma^2(\theta_{n_i}, R_{n_i}) \geq K > 0$ for some constant $K$. Then (120) yields:

$$\begin{aligned} H_{n_{i+1}} &= H_{n_i} + \varepsilon^{1/2} R^* n_\alpha^{1/2} \xi + \mathcal{O}(n_\alpha^2 \varepsilon) + \mathcal{O}(n_\alpha |R^*|^2 \varepsilon) \\ &= H_{n_i} + \varepsilon^{1/2-\alpha} R^* \xi + \mathcal{O}(\varepsilon^{1-4\alpha}) + \mathcal{O}(|R^*|^2 \varepsilon^{1-2\alpha}). \end{aligned}$$

Then:

$$\begin{aligned} \mathbb{P}\{|H_{n_{i+1}} - H_{n_i}| &\leq 2C_2 R^* \varepsilon^{1/2-\alpha}\} \\ &= \mathbb{P}\{|\xi + \mathcal{O}(|R^*|^{-1}\varepsilon^{1/2-3\alpha}) + \mathcal{O}(|R^*|\varepsilon^{1/2-\alpha})| \leq 2C_2\} \\ &\leq \mathbb{P}\{|\xi| \leq 3C_2\}, \end{aligned}$$

where we have used that $|R^*|^{-1}\varepsilon^{1/2-3\alpha} << 1$ because $|R^*| > C_1 > 0$ and $1/2 - 3\alpha > 0$, and that $|R^*|\varepsilon^{1/2-\alpha} \leq \varepsilon^{1/2-\tau-\alpha} << 1$ because $1/2 - \tau - \alpha > 0$. Since $\xi$ converges in distribution to $\mathcal{N}(0, \sigma^2(\theta_{n_i}, R_{n_1}))$ and $\sigma^2(\theta_{n_i}, R_{n_1}) \geq K > 0$, one has:

$$\mathbb{P}\{|H_{n_{i+1}} - H_{n_i}| \leq 2C_2|R^*|\varepsilon^{1/2-\alpha}\} \leq \rho,$$

for some $0 < \rho < 1$. Using this in (100) one obtains the claim of the lemma with $b = -\log \rho > 0$. □

The proof of the following lemma is exactly the same as the proof of Lemma 5.11

**Lemma 5.14.** *Let $R^* \in I_{TZ_1}$, $0 < \tau < 1/2$ be fixed constants. Define $H = H(R^*, 0)$, and let $\alpha$ be such that:*

$$0 < \alpha < 1/2 - \tau, \qquad 0 < \alpha < 1/6.$$

*Let $(\theta_0, R_0) \in I_{TZ_1}$, and consider the strip:*

$$I = \{(\theta, R) \in I_{TZ_1} \, : \, |H(\theta, R) - H^*| \le C_2 |R^*| \varepsilon^{1/2-\alpha}\}.$$

*Let $n^*$ be the exit time of the process $(\theta_n, R_n)$ of $I$. Let $f : \mathbb{R} \to \mathbb{R}$ be any $\mathcal{C}^l$ function with $l \ge 3$. Then for all $\lambda > 0$ one has:*

$$\mathbb{E}\Bigg( e^{-\lambda \varepsilon n^*} f(H_{n^*}) + \varepsilon \sum_{k=0}^{n^*-1} e^{-\lambda \varepsilon k} \left[ \lambda f(H_k) - b(\theta_{qk}, R_{qk}) f'(H_k) - \frac{\sigma^2(\theta_{qk}, R_{qk})}{2} f''(H_k) \right] \Bigg) - f(H_0) = \mathcal{O}(\varepsilon^{3/2-2\alpha-\delta}),$$

*where:*

$$b(\theta, R) = F(\theta, R), \quad \sigma^2(\theta, R) = R_{qk}^2 \sum_{i=0}^{q-1} v^2(\theta + ip/q, 0). \tag{122}$$

**Remark 5.15.** Consider strips $I_j$ of the form:

$$I_j = \{(\theta, R) \in I_{TZ_1} \, : \, |H(\theta, R) - H_j^*| \le K |R^*| \varepsilon^{1/2-\alpha}\}.$$

Recalling (118) one has:

$$I_j = \{(\theta, R) \in I_{TZ_1} \, : \, |R - R_j^*| \le \tilde{K} \varepsilon^{1/2-\alpha}\}.$$

Then, defining $N := [\varepsilon^{-1/2+\alpha-\tau}]$, one can find suitable strips $I_j$ such that:

$$I_{TZ_1} = \bigcup_{j=0}^{N} I_j.$$

Let us denote by $E$ the error terms given Lemma 5.13, that is $E = \mathcal{O}(\varepsilon^{3/2-2\alpha-\delta})$. Then the accumulated error along $I_{TZ_1}$ will be:

$$N^2 E = \mathcal{O}(\varepsilon^{1/2-2\tau-\delta}) << 1,$$

since $\tau < 1/4$.

## 5.7 Transition Zones type 2

Here we study the system in the RR case, in the subdomain:

$$I_{TZ_2} := \{(\theta, r) \in \mathbb{T} \times \mathbb{R} \,:\, K_1 \varepsilon^{1/2-\tau} \le |r - p/q| \le K_2 \varepsilon^{\beta}\},$$

for certain constants $K_1$ and $K_2$. Similarly as in the Totally Irational case, after performing the change to normal form, the $n$-th iteration of our map can be written as:

$$\begin{aligned} \theta_n &= \theta_0 + n r_0 + \mathcal{O}(n\varepsilon), \\ r_n &= r_0 + \varepsilon \sum_{k=0}^{n-1} \omega_k [v(\theta_k, r_k) + \varepsilon v_2(\theta_k, r_k)] + \varepsilon^2 \sum_{k=0}^{n-1} E_2(\theta_k, r_k) + \mathcal{O}(n\varepsilon^{2+a}), \end{aligned} \tag{123}$$

where $v_2(\theta, r)$ is a given function which can be written explicitly in terms of $v(\theta, r)$ and $S_1(\theta, r)$. We consider substrips of the following form:

$$I = \{(\theta, r) \in I_{TZ_2} \,:\, |r - r^*| \le \varepsilon^{1/2-\alpha}\},$$

for some $0 < \alpha < 1/2$.

**Lemma 5.16.** *Let $r^*$ be fixed. Let $(\theta_0, r_0) \in I_{TZ_2}$, and consider the strip:*

$$I = \{(\theta, r) \in I_{TZ_2} \,:\, |r - r^*| \le \varepsilon^{1/2-\alpha}\},$$

*Let $n^*$ denote the first exit time of the process $(\theta_n, r_n)$ of the strip $I$. Let $\delta > 0$ be a sufficiently small constant. Then, there exists a constant $b > 0$ such that:*

$$\mathbb{P}\{n^* > \varepsilon^{-2\alpha-\delta}\} \le e^{-\frac{b}{\varepsilon^{\delta}}}.$$

**Lemma 5.17.** *Let $f : \mathbb{R} \to \mathbb{R}$ be any $\mathcal{C}^l$ function with $l \ge 3$. Then for all $\lambda > 0$ and $\delta > 0$ sufficiently small one has:*

$$\mathbb{E}\Bigg( e^{-\lambda\varepsilon^2 n^*} f(r_{n^*}) + \varepsilon^2 \sum_{k=0}^{n^*-1} e^{-\lambda\varepsilon^2 k} \left[ \lambda f(r_k) - \left( E_2(\theta_k, r_k) f'(r_k) + \frac{v^2(\theta_k, r_k)}{2} f''(r_k) \right) \right] \Bigg) - f(r_0) = \mathcal{O}(\varepsilon^{2+a-2\alpha-\delta}).$$

*Proof.* Let us denote:

$$\eta = e^{-\lambda\varepsilon^2 n^*} f(r_{n^*}) + \varepsilon^2 \sum_{k=0}^{n^*-1} e^{-\lambda\varepsilon^2 k} \left[ \lambda f(r_k) - \left( b(r_k) f'(r_k) + \frac{\sigma^2(r_k)}{2} f''(r_k) \right) \right]. \tag{124}$$

First of all we shall use the law of total expectation. Fix a small enough $\delta > 0$. Then we have:

$$\begin{aligned}\mathbb{E}\,(\eta) &= \mathbb{E}\left(\eta\,|\,n^* \leq \varepsilon^{-2\alpha-\delta}\right)\mathbb{P}\{n^* \leq \varepsilon^{-2\alpha-\delta}\} \\ &+ \mathbb{E}\left(\eta\,|\,n^* > \varepsilon^{-2\alpha-\delta}\right)\mathbb{P}\{n^* > \varepsilon^{-2\alpha-\delta}\}.\end{aligned}$$

Now we write:

$$e^{-\lambda\varepsilon^2 n^*} f(r_{n^*}) = f(r_0) + \sum_{k=0}^{n^*-1}\left(e^{-\lambda\varepsilon^2(k+1)} f(r_{k+1}) - e^{-\lambda\varepsilon^2 k} f(r_k)\right).$$

Doing the Taylor expansion in each term inside the sum we get:

$$\begin{aligned}e^{-\lambda\varepsilon^2 n^*} f(r_{n^*}) &= f(r_0) + \sum_{k=0}^{n^*-1}\Big[-\lambda\varepsilon^2 e^{-\lambda\varepsilon^2 k} f(r_k) + e^{-\lambda\varepsilon^2 k} f'(r_k)(r_{k+1} - r_k) \\ &\quad + \frac{1}{2} e^{-\lambda\varepsilon^2 k} f''(r_k)(r_{k+1} - r_k)^2 + \mathcal{O}(e^{-\lambda\varepsilon^2 k}\varepsilon^3)\Big].\end{aligned}$$

Substituting this in (124) we get:

$$\begin{aligned}\eta &= f(r_0) + \sum_{k=0}^{n^*-1}\left[e^{-\lambda\varepsilon^2 k} f'(r_k)(r_{k+1} - r_k) + \frac{1}{2} e^{-\lambda\varepsilon^2 k} f''(r_k)(r_{k+1} - r_k)^2\right] \\ &\quad -\varepsilon^2 \sum_{k=0}^{n^*-1} e^{-\lambda\varepsilon^2 k}\left[b(r_k) f'(r_k) + \frac{\sigma^2(r_k)}{2} f''(r_k)\right] + \sum_{k=0}^{n^*-1}\mathcal{O}(e^{-\lambda\varepsilon^2 k}\varepsilon^3). \qquad (125)\end{aligned}$$

We note that using (123) we can write:

$$r_{k+1} - r_k = \varepsilon\omega_k[v(\theta_k, r_k) + \varepsilon v_2(\theta_k, r_k)] + \varepsilon^2 E_2(\theta_k, r_k) + \mathcal{O}(\varepsilon^{2+a}),$$

and also:

$$(r_{k+1} - r_k)^2 = \varepsilon^2 v^2(\theta_k, r_k) + \mathcal{O}(\varepsilon^3).$$

Thus we can rewrite (125) as:

$$\begin{aligned}\eta &= f(r_0) + \varepsilon \sum_{k=0}^{n^*-1} e^{-\lambda\varepsilon^2 k} f'(r_k)\omega_k\left[v(\theta_k, r_k) + \varepsilon v_2(\theta_k, r_k)\right] \\ &\quad + \sum_{k=0}^{n^*-1}\mathcal{O}(e^{-\lambda\varepsilon^2 k}\varepsilon^{2+a}). \qquad (126)\end{aligned}$$

Now we distinguish between the case $n^* \leq \varepsilon^{-2\alpha-\delta}$ and $n^* > \varepsilon^{-2\alpha-\delta}$. Consider the former case. First, we show that the last term in (126) is $\mathcal{O}(\varepsilon^{2+a-2\alpha-\delta})$. Indeed,

$$\left|\sum_{k=0}^{n^*-1} \mathcal{O}(e^{-\lambda\varepsilon^2 k}\varepsilon^{2+a})\right| \leq K\varepsilon^{2+a}n^* \leq K\varepsilon^{2+a-2\alpha-\delta}, \tag{127}$$

where $K$ is some positive constant. Then, for $n^* \leq \varepsilon^{-2\alpha-\delta}$ we obtain:

$$\eta = f(r_0) + \varepsilon \sum_{k=0}^{n^*-1} e^{-\lambda\varepsilon^2 k} f'(r_k)\omega_k \left[v(\theta_k, r_k) + \varepsilon v_2(\theta_k, r_k)\right] + \mathcal{O}(\varepsilon^{2+a-2\alpha-\delta}). \tag{128}$$

Now we focus on the case $n^* > \varepsilon^{-2\alpha-\delta}$. The last term in (126) can be bounded by:

$$\left|\sum_{k=0}^{n^*-1} \mathcal{O}(e^{-\lambda\varepsilon^2 k}\varepsilon^{2+a})\right| \leq K\varepsilon^{2+a} \sum_{k=0}^{n^*-1} e^{-\lambda\varepsilon^2 k} = K\varepsilon^{2+a}\frac{1-e^{-\lambda\varepsilon^2 n^*}}{1-e^{-\lambda\varepsilon^2}} \leq K_\lambda \varepsilon^a, \tag{129}$$

for some positive constants $K$ and $K_\lambda$. Using this bound in equation (126), we obtain that for $n^* > \varepsilon^{-2\alpha-\delta}$:

$$\eta = f(r_0) + \varepsilon \sum_{k=0}^{n^*-1} e^{-\lambda\varepsilon^2 k} f'(r_k)\omega_k \left[v(\theta_k, r_k) + \varepsilon v_2(\theta_k, r_k)\right] + \mathcal{O}(\varepsilon^a). \tag{130}$$

Now we just need to note that since $\omega_k$ is independent of $r_k$ and $\theta_k$, we have for all $k \in \mathbb{N}$:

$$\begin{aligned} \mathbb{E}(\omega_k f'(r_k)[v(\theta_k, r_k) + \varepsilon v_2(\theta_k, r_k)]) &= \\ \mathbb{E}(\omega_k)\mathbb{E}(f'(r_k)[v(\theta_k, r_k) + \varepsilon v_2(\theta_k, r_k)]) &= 0, \end{aligned}$$

because $\mathbb{E}(\omega_k) = 0$. Thus, if we take expectations in (128) and (130) and use Lemma 5.16 it is clear that:

$$\mathbb{E}(\eta) - f(r_0) = \mathcal{O}(\varepsilon^{2+a-2\alpha-\delta})\mathbb{P}\{n^* \leq \varepsilon^{-2\alpha-\delta}\} + \mathcal{O}(\varepsilon^a)\mathbb{P}\{n^* > \varepsilon^{-2\alpha-\delta}\} = \mathcal{O}(\varepsilon^{2+a-2\alpha-\delta}).$$

□

## 5.8 An expextation lemma for a whole Real Rational Strip

Finally, we shall put all the information of the previous subsections toghether in ordrer to obtain an expectation lemma valid in the whole strip.

**Lemma 5.18.** *Let $\beta, \tau, \gamma > 0$ be such that:*

$$2\beta < \min\{1/2 - \tau, 1/2 - 2\gamma, 2\gamma\},$$

*and let $\delta > 0$ be a sufficiently small constant. Let $(\theta_0, R_0)$ be such that $|R_0| \leq C_2\varepsilon^\beta$. Let $n^*$ be the exit time of the process $(\theta_n, R_n)$ of this domain. Let $f : \mathbb{R} \to \mathbb{R}$ be any $\mathcal{C}^l$ function with $l \geq 3$. Then for all $\lambda > 0$ there exists $d > 0$ such that:*

$$\mathbb{E}\Bigg(e^{-\lambda\varepsilon n^*} f(H_{n^*}) + \varepsilon \sum_{k=0}^{n^*-1} e^{-\lambda\varepsilon k}\left[\lambda f(H_k) - b(\theta_{qk}, R_{qk}) f'(H_k) - \frac{\sigma^2(\theta_{qk}, R_{qk})}{2} f''(H_k)\right]\Bigg) - f(H_0) = \mathcal{O}(\varepsilon^{2\beta + d - \delta}),$$

*where:*

$$b(\theta, R) = F(\theta, R), \quad \sigma^2(\theta, R) = R_{qk}^2 \sum_{i=0}^{q-1} v^2(\theta + ip/q, 0). \tag{131}$$

*Proof.* We consider Markov times $0 < n_1 < n_2 < \cdots < n_m = n^*$ such that, for all $i$, $(\theta_{n_i}, R_{n_i})$ and $(\theta_{n_{i+1}}, R_{n_{i+1}})$ belong to different substrips of any of the domains defined in the previous subsections. We define:

$$\mathcal{N}_j = \{n_i \, : \, (\theta_k, R_k) \in D_j, \, k = n_i, \ldots, n_{i+1} - 1\},$$

$$\mathcal{N}_{TZ_1} = \{n_i \, : \, (\theta_k, R_k) \in I_{TZ_1}, \, k = n_i, \ldots, n_{i+1} - 1\},$$

$$\mathcal{N}_{TZ_2} = \{n_i \, : \, (\theta_k, R_k) \in I_{TZ_2}, \, k = n_i, \ldots, n_{i+1} - 1\}.$$

Then if we define:

$$\eta = e^{-\lambda\varepsilon n^*} f(H_{n^*}) + \varepsilon \sum_{k=0}^{n^*-1} e^{-\lambda\varepsilon k}\left[\lambda f(H_k) - b(\theta_{qk}, R_{qk}) f'(H_k) - \frac{\sigma^2(\theta_{qk}, R_{qk})}{2} f''(H_k)\right]\Bigg) - f(H_0)$$

and:

$$\eta_{n_i} = e^{-\lambda\varepsilon n_i} f(H_{n_i}) + \varepsilon \sum_{k=n_{i-1}}^{n_i-1} e^{-\lambda\varepsilon k}\left[\lambda f(H_k) - b(\theta_{qk}, R_{qk}) f'(H_k) - \frac{\sigma^2(\theta_{qk}, R_{qk})}{2} f''(H_k)\right]\Bigg) - f(H_{n_{i-1}}),$$

we have that:

$$\mathbb{E}(\eta) = \mathbb{E}\left(\sum_{i=0}^{m}\eta_{n_i}\right) = \mathbb{E}\left(\sum_{n_i\in\mathcal{N}_1}\eta_{n_i} + \sum_{n_i\in\mathcal{N}_2}\eta_{n_i} + \sum_{n_i\in\mathcal{N}_3}\eta_{n_i}\right.$$
$$\left. + \sum_{n_i\in\mathcal{N}_4}\eta_{n_i} + \sum_{n_i\in\mathcal{N}_{TZ_1}}\eta_{n_i} + \sum_{n_i\in\mathcal{N}_{TZ_2}}\eta_{n_i}\right).$$

Using the bounds given by the expectation lemmas, we obtain:

$$\begin{aligned}\mathbb{E}(\eta) &= \mathcal{O}(\mathbb{E}(|\mathcal{N}_1|)\varepsilon^{1+2\gamma-\delta}) + \mathcal{O}(\mathbb{E}(|\mathcal{N}_2|)\varepsilon^{1-2\gamma-\delta}) + \mathcal{O}(\mathbb{E}(|\mathcal{N}_3|)\varepsilon^{3/2-2\alpha-\delta}) \\ &\quad +\mathcal{O}(\mathbb{E}(|\mathcal{N}_4|)\varepsilon^{3/2-2\alpha-\delta}) + \mathcal{O}(\mathbb{E}(|\mathcal{N}_{TZ_1}|)\varepsilon^{3/2-2\alpha-\delta}) + \mathcal{O}(\mathbb{E}(|\mathcal{N}_{TZ_2}|)\varepsilon^{3/2-2\alpha-\delta}).\end{aligned}$$

One can easily see that, with probability exponentially small close to 1 as $\varepsilon \to 0$, one has:

$$\mathbb{E}(|\mathcal{N}_1|) \leq K, \qquad \mathbb{E}(|\mathcal{N}_2|) \leq K, \qquad \mathbb{E}(|\mathcal{N}_3|) \leq \varepsilon^{-1/2-2\gamma+2\alpha},$$
$$\mathbb{E}(|\mathcal{N}_4|) \leq \varepsilon^{-1-2\gamma+2\alpha}, \qquad \mathbb{E}(|\mathcal{N}_{TZ_1}|) \leq \varepsilon^{-1-2\tau+2\alpha}, \qquad \mathbb{E}(|\mathcal{N}_{TZ_2}|) \leq \varepsilon^{-1+2\beta+2\alpha}.$$

Then, using that $2\beta < \min\{1/2-\tau, 1/2-2\gamma, 2\gamma\}$ one obtains:

$$\mathbb{E}(\eta) = \mathcal{O}(\varepsilon^{2\beta+d} - \delta),$$

with $d = \min\{1/2-\tau, 1/2-2\gamma, 2\gamma\} - 2\beta > 0$. □

# A Measure of the domain covered by RR and IR intervals

In this section we show that, with the right choice of $b$, the measure of the the union of all strips of RR and IR type inside any compact set:

$$A_\beta = \cup_k I_\beta^k \subset \mathbb{T}\times B \qquad I_\beta^k \text{ strips of width } 2\varepsilon^\beta$$

goes to zero as $\varepsilon \to 0$.

In fact, we will do the proof for $A = [0,1]$. The general case is completely analogous. Let us consider:

$$\mathcal{R} = \{p/q \in \mathbb{Q} \,:\, p<q,\ \gcd(p,q)=1,\ q<\varepsilon^{-b}\} = \cup_{q=1}^{q_{\max}}\mathcal{R}_q \subset [0,1],$$

where $q_{\max} = [\varepsilon^{-b}]$ and:

$$\mathcal{R}_q = \{p/q \in \mathbb{Q} \,:\, p<q,\ \gcd(p,q)=1\}.$$

Finally we denote:

$$I_{\mathcal{R}} = \{I_\beta^k \subset [0,1] \,:\, \exists p/q \in \mathcal{R}\cap I_\beta\}.$$

**Lemma A.1.** *Let $\rho$ be fixed, $0 < \rho < \beta$, and define $b = (\beta - \rho)/2$. Then, for each $I_\beta$ such that there is at most one rational $p/q$ satisfying $|q| \leq \varepsilon^{-b}$ the union $I_{\mathcal{R}}$ has the Lebesgue measure $\mu(I_{\mathcal{R}}) \leq \varepsilon^\rho$ and, therefore, as $\varepsilon \to 0$:*

$$\mu(I_{\mathcal{R}}) \to 0,$$

*where $\mu$ denotes the Lebesgue measure.*

*Proof.* On the one hand, suppose that $p/q \in I_\beta$, $q \leq \varepsilon^{-b}$. Then, for all $p'/q' \in I_\beta$, with $p'$ and $q'$ relatively prime and $p'/q' \neq p/q$, we have:

$$\varepsilon^\beta \geq |p/q - p'/q'| \geq \frac{1}{qq'} \geq \frac{\varepsilon^b}{q'}.$$

Therefore:

$$q' \geq \varepsilon^{-\beta+b} = \varepsilon^{-b-3\rho/2} > \varepsilon^{-b},$$

so the first part of the claim is proved.

On the other hand we note that, if $q_1 \neq q_2$, then $\mathcal{R}_{q_1} \cap \mathcal{R}_{q_2} = \emptyset$. Moreover, it is clear that $\#\mathcal{R}_q \leq q - 1$ (and if $q$ is prime then $\#\mathcal{R}_q = q - 1$, so that the bound is optimal). Therefore we have:

$$\#\mathcal{R} \leq \sum_{q=1}^{q_{\max}} \#\mathcal{R}_q \leq \sum_{q=1}^{q_{\max}} q - 1 = \frac{q_{\max}^2}{2} < \varepsilon^{-2b}.$$

Since $\mu(I_\beta) = \varepsilon^\beta$, one has:

$$0 \leq \mu(I_{\mathcal{R}}) = \varepsilon^\beta \#\mathcal{R} < \varepsilon^\beta \varepsilon^{-2b} = \varepsilon^\rho,$$

so that the second claim of the lemma is also clear. □

# B A generalization of Theorem 2.1

Theorem 2.1 can be easily generalized in the following way.

**Theorem 2.1'.** *Let $\delta_1(\varepsilon)$ and $\delta_2(\varepsilon)$ be continuous functions such that for any $\gamma > 0$ one has:*

$$\frac{\varepsilon^{1+\gamma}}{\delta_i(\varepsilon)} \to 0,$$

*as $\varepsilon \to 0$. Then Theorem 2.1 applies to the random collection of maps $\tilde{f}_\omega$, where $\tilde{f}_\omega$ is the same as $f_\omega$ replacing the terms $\varepsilon \mathbb{E} u(\theta)$ and $\varepsilon \omega_0 v(\theta)$ by $\delta_1(\varepsilon) \mathbb{E} u(\theta)$ and $\delta_2(\varepsilon) \omega_0 v(\theta)$ respectively.*

**Remark B.1.** For instance, one can take $\delta_i(\varepsilon) = \varepsilon \log^n \varepsilon$ for any n.

# C Sufficient condition for weak convergence and auxiliary lemmas

In order to prove that the $r$-component exhibits a diffusion process we need to adapt several lemmas from Ch. 8 sec. 3 [20]. We recall some terminology and notations (see Ch. 1 sec. 1 [20] for more details).

In the notations of section 1.6 we have

**Lemma C.1.** *(see Lm. 3.1, [20]) Let $M$ be a metric space, $Y$ a continuous mapping $M \mapsto Y(M)$, $Y(M)$ being a complete separable metric space. Let $(X_t^\varepsilon, P_x^\varepsilon)$ be a family of Markov processes in $M$; suppose that the process $Y(X_t^\varepsilon)$ has continuous trajectories. Let $(y_t, P_y)$ be a Markov process with continuous paths in $Y(M)$ whose infinitesimal operator is $A$ with domain of definition $D_A$. Suppose that the space $C[0,\infty)$ of continuous functions on $[0,\infty)$ with values in $\Gamma$ is taken as the sample space, so that the distribution of the process in the space of continuous functions is simply $P_y$. Let $\Psi$ be a subset of the space $C(Y(M))$ such that for measures $\mu_1$, $\mu_2$ on $Y(M)$ the equality $\int f d\mu_1 = \int f d\mu_2$ for all $f \in \Psi$ implies $\mu_1 = \mu_2$. Let $D$ be a subset of $D_A$ such that for every $f \in \Psi$ and $\lambda > 0$ the equation $\lambda F - AF = f$ has a solution $F \in D$.*

*Suppose that for every $x \in M$ the family of distributions $Q_x^\varepsilon$ of $Y(X_\bullet^\varepsilon)$ in the space $C[0,\infty)$ corresponding to the probabilities $P_x^\varepsilon$ for all $\varepsilon$ is tight; and that for every compact $K \subset Y(M)$, for every $f \in D$ and every $\lambda > 0$,*

$$\mathbb{E}_x^\varepsilon \int_0^\infty \exp(-\lambda t)\left[\lambda f(Y(X_t^\varepsilon)) - Af(Y(X_t^\varepsilon))\right] dt \to f(Y(x))$$

*as $\varepsilon \to 0$ uniformly in $x \in Y^{-1}(K)$.*

*Then $Q_x^\varepsilon$ converges weakly as $\varepsilon \to 0$ to the probability measure $P_{Y(x)}$.*

In our case $Y(M)$ is the real line. We use a discrete version of this lemma in our proof.

Similarly, to Lemma 3.2 [20] one can show that the family of distributions $Q_x^\varepsilon$ (those of $Y(X_\bullet^\varepsilon)$ with respect to the probability measures $P_x^\varepsilon$ in the space $C[0,\infty)$) with small nonzero $\varepsilon$ is tight. Indeed, in our case speed of change of $I$ is bounded. Denote $H(X) = H(r,\theta) = r^2/2$. Then

- for every $T > 0$ and $\delta > 0$ there exists $H_0$ such that

$$\mathbb{P}_x^\varepsilon\{\max_{0<t<T} |H(X_t^\varepsilon)| > H_0\} < \delta.$$

- for every compact subset $K \subset \mathbb{A}$ and for every sufficiently small $\rho > 0$ there exists a constant $A_\rho$ such that for every $a \in K$ there exists a function $f_\rho^a(y)$

on $Y(\mathbb{A})$ such that $f_\rho^a(a) \equiv 1, f_\rho^a(y) \equiv 0$ for $\rho(y,a) \geq \rho$, $0 \leq f_\rho^a(y) \leq 1$ everywhere, and $f_\rho^a(Y(X_t^\varepsilon)) + A_\rho t$ is a submartingale for all $\varepsilon$ (see Stroock and Varadhan [38]).

In the proof we need an auxiliary lemmas. We study the random sums

$$S_n = \sum_{k=1}^{n} v_k \omega_k, \quad n \geq 1, \tag{132}$$

where $\{\omega_k\}_{k\geq 1}$ is a sequence of independent random variables with equal $\pm 1$ with equal probability 1/2 each and $\{v_k\}_{k\geq 1}$ is a sequence such that

$$\lim_{n\to\infty} \frac{\sum_{k=1}^n v_k^2}{n} = \sigma.$$

Here is a standard

**Lemma C.2.** $\{S_n/n^{1/2}\}_{n\geq 1}$ *converges in distribution to the normal dirtribution* $\mathcal{N}(0,\sigma^2)$.

Recall that a characteristic function of a random variable $X$ is a function $\phi_X : \mathbb{R} \to \mathbb{C}$ given by $\phi_X(t) = \mathbb{E}\exp(itX)$. Notice that it satisfies the following two properties:

- If $X, Y$ are independent random variables, then $\varphi_{X+Y} = \varphi_X \cdot \varphi_Y$.
- $\varphi_{aX}(t) = \varphi_X(at)$.

A sufficient condition to prove convergence in distribution is as follows.

**Theorem C.3** (Continuity theorem [6])**.** *Let* $\{X_n\}_{n\geq 1}, Y$ *be random variables. If* $\{\varphi_{X_n}(t)\}_{n\geq 1}$ *converges to* $\varphi_Y(t)$ *for every* $t \in \mathbb{R}$*, then* $\{X_n\}_{n\geq 1}$ *converges in distribution to* $Y$*.*

A direct calculation shows that

$$\lim_{n\to\infty} \log \phi_{S_n/\sqrt{n}}(t) = -\frac{t^2}{2\sigma^2} \qquad \text{for all} \quad t \in \mathbb{R}.$$

This way of proof was communicated to the authors by Yuri Lima.

**Acknowledgement:** The authors warmly thank Leonid Koralov for numerious envigorating discussions of various topics involving stochatic processes. Communications with Dmitry Dolgopyat, Yuri Bakhtin, Jinxin Xue were useful for the project and gladly acknowledged by the authors. The second author acknowledges partial support of the NSF grant DMS-1402164.

# References


[1] Arnold, V. I. Instabilities in dynamical systems with several degrees of freedom, Sov Math Dokl 5 (1964), 581–585;

[2] Arnold, V. I. Mathematical methods of classical mechanics, Graduate Texts in Mathematics, 60, Second Edition, Springer-Verlag, 1989.

[3] Arnold, V. I. Mathematical problems in classical physics. Trends and perspectives in applied mathematics, 1–20, Appl. Math. Sci., 100, Springer, New York, 1994.

[4] Bernard, P. The dynamics of pseudographs in convex Hamiltonian systems. *J. Amer. Math. Soc.*, 21(3):615–669, 2008.

[5] Bernard, P. Kaloshin, V. Zhang, K. Arnold diffusion in arbitrary degrees of freedom and 3-dimensional normally hyperbolic invariant cylinders, arXiv:1112.2773 [math.DS] 2011, 58pp, conditionally accepted to Acta Mathematica.

[6] Breiman, L. Probability Published by Society for Industrial and Applied Mathematics, 1992

[7] Brin, M. Stuck, G. Introduction to Dynamical Systems, Cambridge University Press, 2003.

[8] Cheng Ch.-Q., Arnold diffusion in nearly integrable Hamiltonian systems. arXiv: 1207.4016v2 9 Mar 2013, 127 pp;

[9] Cheng, Ch.-Q. Yan, J. Existence of diffusion orbits in a priori unstable Hamiltonian systems. *Journal of Differential Geometry*, 67 (2004), 457–517 & 82 (2009), 229–277;

[10] Chirikov. B. V. A universal instability of many-dimensional oscillator systems. Phys. Rep., 52(5): 264–379, 1979.

[11] Chirikov ,B.V. Vecheslavov, V.V. Theory of fast Arnold diffusion in many-frequency systems, J. Stat. Phys. 71(1/2): 243 (1993)

[12] de la Llave, R. Orbits of unbounded energy in perturbations of geodesic flows by periodic potentials. a simple construction preprint 70pp, 2005.

[13] Delshams, A. de la Llave, R. Seara, T. A geometric mechanism for diffusion in Hamiltonian systems overcoming the large gap problem: heuristics and rigorous verification on a model Mem. Amer. Math. Soc. 179 (2006), no. 844, pp.144

[14] de la Llave, Orbits of unbounded energy in perturbations of geodesic flows by periodic potentials. a simple construction preprint 70pp, 2005.

[15] Dolgopyat, D. Repulsion from resonance Memoires SMF, 128, 2012.

[16] Dumas, H. Laskar, J. Global Dynamics and Long-Time Stability in Hamiltonian via Numerical Frequency Analysis Phys Review Let. 70, no. 20, 1993, 2975–2979.

[17] Fejoz, J. Guardia, M. Kaloshin, V. Roldan, P. Kikrwood gaps and diffusion along mean motion resonance for the restricted planar three body problem, arXiv:1109.2892 [math.DS] 2013, to appear in Journal of the European Mathematical Society,

[18] Filonenko, N. Zaslavskii G. Stochastic instability of trapped particles and conditions of applicability of the quasi-linear approximation, Soviet Phys. JETP 27 (1968), 851–857.

[19] Freidlin, M. Sheu, S. Diffusion processes on graphs: stochastic differential equations, large deviation principle, Probability theory and related fields 116.2 (2000): 181–220;

[20] Freidlin, M. Wentzell, A. Random perturbations of dynamical systems, Grundlehren der Mathematischen Wissenschaften, Vol. 260, Springer, 2012, 258pp.

[21] Gidea, M. de la Llave. R Topological methods in the large gap problem. Discrete and Continuous Dynamical Systems, Vol. 14, 2006.

[22] Guardia, M. Kaloshin, V. Diffusive behavior for the generalized example of Arnold, in preparation

[23] Ibragimov, I. A. A note on the central limit theorems for dependent random variables, Theory of Probability and Its Applications, 1975.

[24] Kaloshin, V. Geometric proofs of Mather's accelerating and connecting theorems, Topics in Dynamics and Ergodic Theory (eds. S. Bezuglyi and S. Kolyada), London Mathematical Society, Lecture Notes Series, Cambridge University Press, 2003, 81—106

[25] Kaloshin, V. Zhang, K. A strong form of Arnold diffusion for two and a half degrees of freedom, arXiv:1212.1150 [math DS], 2012, 207pp,

[26] Kaloshin, V. Zhang, K. A strong form of Arnold diffusion for three and a half degrees of freedom, http://terpconnect.umd.edu/ vkaloshi/ 36pp,

[27] Kaloshin, V. Zhang, K. Dynamics of the dominant Hamiltonians, with application to Arnold diffusion arXiv:1212.1150 [math DS], 2012, 83pp,

[28] Laskar, J. Frequency analysis for multi-dimensional systems. Global dynamics and diffusion, Physica D, 67 (1993), 257–281, North-Holland;

[29] Marco, J.-P. Modèles pour les applications fibrées et les polysystèmes. (French) [Models for skew-products and polysystems] C. R. Math. Acad. Sci. Paris 346 (2008), no. 3-4, 203–208.

[30] Marco, J.-P. Sauzin, D. Wandering domains and random walks in Gevrey near integrable systems, Ergodic Theory & Dynamical Systems Vol. 24, 5 (volume spécial à la mémoire de M. Herman, édité par A. Fathi et J.-C. Yoccoz), 1619 – 1666, 2004.

[31] Moeckel, R. Transition tori in the five-body problem, J. Diff. Equations 129, 1996, 290–314.

[32] Moeckel, R. personal communications;

[33] Moser, J. Is the solar system stable? Math. Intelligencer, 1(2): 65–71, 1978/79.

[34] Piftankin G. Treshchev D. Separatrix maps in Hamiltonian systems, Russian Math. Surveys 62:2 219–322;

[35] Sauzin, D. Ergodicity and conservativity in the random iteration of standard maps, preprint 2006.

[36] Sauzin, D. Exemples de diffusion d'Arnold avec convergence vers un mouvement brownien, preprint 2006.

[37] Shatilov, D. Levichev, E. Simonov, E. and M. Zobov Application of frequency map analysis to beam-beam effects study in crab waist collision scheme Phys. Rev. ST Accel. Beams 14, January 2011

[38] Stroock, D.W. Varadhan, S.R.S. Multidimensional Diffusion Processes. Springer: Berlin, 1979.

[39] Treschev, D. Multidimensional Symplectic Separatrix Maps, J. Nonlinear Sciences 12 (2002), 27—58;

[40] Treschev, D. Evolution of slow variables in a priori unstable Hamiltonian systems Nonlinearity 17 (2004), no. 5, 1803–1841;

[41] Treschev, D Arnold diffusion far from strong resonances in multidimensional a priori unstable Hamiltonian systems Nonlinearity 25 (2012), no 9, 2717–2758.

[42] Wisdom, J. "The origin of the Kirkwood gaps - A mapping for asteroidal motion near the 3/1 commensurability". Astronomical Journal 87: 577–593, 1982.