AVERAGING FOR STOCHASTIC PERTURBATIONS OF INTEGRABLE SYSTEMS

GUAN HUANG, SERGEI KUKSIN, AND ANDREY PIATNITSKI

Abstract. We are concerned with averaging theorems for ε-small stochastic perturbations of integrable equations in $\mathbb{R}^d \times \mathbb{T}^n = \{(I, \varphi)\}$

$$\dot{I}(t) = 0, \quad \dot{\varphi}(t) = \theta(I), \tag{1}$$

and in $\mathbb{R}^{2n} = \{v = (\mathbf{v}_1, \dots, \mathbf{v}_n),\ \mathbf{v}_j \in \mathbb{R}^2\}$,

$$\dot{\mathbf{v}}_k(t) = W_k(I)\mathbf{v}_k^{\perp}, \quad k = 1, \dots, n, \tag{2}$$

where $I = (I_1, \dots, I_n)$ is the vector of actions, $I_j = \frac{1}{2}\|\mathbf{v}_j\|^2$. The vector-functions θ and W are locally Lipschitz and non-degenerate. Perturbations of these equations are assumed to be locally Lipschitz and such that some few first moments of the norms of their solutions are bounded uniformly in ε, for $0 \le t \le \varepsilon^{-1}T$. For I-components of solutions for perturbations of (1) we establish their convergence in law to solutions of the corresponding averaged I-equations, when $0 \le \tau := \varepsilon t \le T$ and $\varepsilon \to 0$. Then we show that if the system of averaged I-equations is mixing, then the convergence is uniform in the slow time $\tau = \varepsilon t \ge 0$.

Next using these results, for ε-perturbed equations (2) we construct well posed *effective stochastic equations* for $v(\tau) \in \mathbb{R}^{2n}$ (independent of ε) such that when $\varepsilon \to 0$, the actions of solutions for the perturbed equations with $t := \tau/\varepsilon$ converge in distribution to actions of solutions for the effective equations. Again, if the effective system is mixing, this convergence is uniform in the slow time $\tau \ge 0$.

We provide easy sufficient conditions on the perturbed equations which ensure that our results apply to their solutions.

1. Introduction

1.1. **Setting and problems.** The goal of this paper is to present an averaging theory for stochastic differential equations, obtained by diffusive perturbations of integrable deterministic differential equations. All equations in our work have locally Lipschitz coefficients.

In previous publication [12], written by two of us, an easier problem of stochastic perturbations of linear systems with imaginary spectrum was considered. This work may be regarded as a natural continuation of [12]. Both papers are based on the Khasminski approach to the averaging in stochastic systems and use some technical ideas, developed by the authors and A. Maiocchi in their work on stochastic PDEs [23, 22, 13, 11]. We consider two classes of problems as above. Firstly, given an integrable system in $\mathbb{R}^d \times \mathbb{T}^n$, [1]

$$\dot{I} = 0, \quad \dot{\varphi} = \theta(I), \qquad (I, \varphi) \in \mathbb{R}^d \times \mathbb{T}^n, \quad \mathbb{T}^n = \mathbb{R}^n / 2\pi\mathbb{Z}^n, \tag{1.1}$$

[1] If $n = d$ and $\theta(I) = \nabla h(I)$ for some C^1-function h on $\mathbb{R}^d$, then (1.1) is an integrable Hamiltonian system on the symplectic space $(\mathbb{R}^n \times \mathbb{T}^n, dI \wedge d\varphi)$ with the Hamiltonian function h. Otherwise (1.1) is integrable in some general sense.

we study the asymptotic properties, as $\varepsilon \to 0$, of solutions for perturbed systems of the form

$$\begin{aligned} dI &= \varepsilon P^I(I,\varphi)dt + \sqrt{\varepsilon}\,\Psi^I(I,\varphi)d\beta(t),\\ d\varphi &= \big[\theta(I) + \varepsilon P^\varphi(I,\varphi)\big]dt + \sqrt{\varepsilon}\,\Psi^\varphi(I,\varphi)d\beta(t), \end{aligned} \tag{1.2}$$

where $\beta(t)$ is the standard Wiener process in $\mathbb{R}^{d_1}$ and the matrices Ψ^I and Ψ^φ have corresponding dimensions. We also consider stochastic perturbations of integrable equations in $\mathbb{R}^{2n}$ [2]

$$\dot{\mathbf{v}}_k = W_k(I)\mathbf{v}_k^\perp, \qquad k = 1,2,\dots,n, \tag{1.3}$$

where $\mathbf{v}_k = (v_k, v_{-k})^t \in \mathbb{R}^2$, $\mathbf{v}_k^\perp = (-v_{-k}, v_k)^t$ and I_k is the k-th action $\frac{1}{2}\|\mathbf{v}_k\|^2$. Similar to the above we are interested in stochastic perturbations of the form

$$d\mathbf{v}_k = \big[W_k(I)\mathbf{v}_k^\perp + \varepsilon\mathbf{P}_k(v)\big]dt + \sqrt{\varepsilon}\sum_{j=1}^{n_1} B_{kj}(v)d\beta_j(t), \quad k = 1,\dots,n. \tag{1.4}$$

For convenience of language, below we call systems of equations like (1.3) and (1.4) just "equations".

Our goal is to study the asymptotic behaviour of solutions for systems (1.2) and (1.4) as $\varepsilon \to 0$ on time-interval of length $\sim \varepsilon^{-1}$, and the limiting behaviour of the solutions for $0 \le t \le \infty$ when $\varepsilon \to 0$, provided that suitable mixing assumptions hold. To study this behaviour it is convenient to pass to the slow time $\tau = \varepsilon t$, and this is what we do below in the main text. In the introduction we discuss our results using both the original fast time t and slow time τ.

1.2. **The results.** Systems (1.2) are examined in the first part of our work (Sections 2-4). The first main result of this analysis is Theorem 4.3 which describes the statistical behaviour of components $I^\varepsilon(t)$ of solutions for (1.2) on time-intervals of order ε^{-1}. There we assume that the mapping θ is nondegenerate in sense of Anosov (i.e. for almost all $I \in \mathbb{R}^d$ components of the vector $\theta(I)$ are rationally independent), that the diffusion in the I-equation in (1.2) is bounded non-degenerate, that the coefficients of equation (1.2) satisfy some mild regularity assumptions and that certain a-priori estimates for solutions of (1.2) hold uniformly in ε. The theorem states that the vector of I-components $I^\varepsilon(\varepsilon^{-1}\tau)$ of a solution for $0 \le \tau \le T$ converges in law to a solution $I^0(\tau)$ of an averaged stochastic differential equation, obtained by means of an appropriate averaging in φ of coefficients of the I-equation in (1.2). More precisely, the theorem states that

$$\mathcal{D}I^\varepsilon(\varepsilon^{-1}\tau) \rightharpoonup \mathcal{D}I^0(\tau), \tag{1.5}$$

where $\mathcal{D}$ signifies the distribution of a r.v. (random variable), the arrow stands for weak convergence of measures and $I^0(\tau)$ is a solution of a stochastic equation

$$dI(\tau) = \langle P^I\rangle(I)d\tau + \langle\!\langle \Psi^I\rangle\!\rangle(I)d\beta(\tau). \tag{1.6}$$

In (1.6) $\langle P^I\rangle$ is the averaging of P^I in angles and matrix $\langle\!\langle \Psi^I\rangle\!\rangle(I)$ is obtained from Ψ^I using the rules of stochastic averaging, see in Subsection 3.2.

In Section 5, assuming the mixing, we examine the asymptotic in time behaviour of components $I^\varepsilon(\tau)$. More precisely, Proposition 5.2 states that if in addition to the above-mentioned assumptions system (1.2) and the averaged equation (1.6) both are mixing with stationary measures ν^ε and μ^0 respectively,[3] then the I-projection of ν^ε weakly converges to μ^0 as $\varepsilon \to 0$. So for any solution $(I^\varepsilon, \varphi^\varepsilon)$ of (1.2) we have

$$\lim_{\varepsilon\to 0}\lim_{\tau\to\infty} \mathcal{D}I^\varepsilon(\tau) = \mu^0. \tag{1.7}$$

[2]If $W(I) = \nabla h(I)$ for some C^1-function h, then (1.3) is a Hamiltonian equation in $\mathbb{R}^{2n}$, integrable in the sense of Birkhoff.

[3]Concerning the mixing in SDEs and its basic properties which we use see [17, 12] and references in [12].

Convergence (1.5), valid for any finite time T, jointly with (1.7) suggest that, in fact, convergence (1.5) holds uniformly in $\tau \geq 0$. Later in Section 5 we show that indeed this is the case. Namely, let $\|\mu_1 - \mu_2\|_L^*$ be the dual-Lipschitz distance between measures μ_1, μ_2, see Definition 5.3.[4] In Theorem 5.5, assuming that equation (1.6) is mixing with a mild quantitative property of the rate of mixing,[5] we prove that

$$\lim_{\varepsilon \to 0} \sup_{\tau \geq 0} \|\mathcal{D}I^\varepsilon(\varepsilon^{-1}\tau) - \mathcal{D}I^0(\tau)\|_L^* = 0. \tag{1.8}$$

The behaviour of I-components of solutions for deterministically perturbed equation (1.2) with $\Psi^I = 0$ and $\Psi^\varphi = 0$ for typical initial data on time-intervals of order ε^{-1} is the subject of the classical averaging theory due to Anosov–Kosuga–Neishtadt, see [1, Sec. 6.1] and [25]. Moreover, due to Neishtadt and Bakhtin the set of atypical initial data for which the averaging fails has measure $\lesssim \varepsilon^\gamma$, where $\gamma > 0$ depends on "how non-degenerate mapping θ is". Theorem 4.3 is a natural counterpart of that theory for stochastic perturbations of integrable equations. But stochastic convergences (1.7) and (1.8), established in Proposition 5.2 and Theorem 5.5 and valid for any initial data, seem to have no deterministic analogy.

Equation (1.2) is a fast-slow stochastic system with fast variable φ and slow variable I. The averaging in such systems (which allows to approximatively describe the law of the slow variable I on time-integrals of order ε^{-1}) is a well developed topic of stochastic analysis, e.g. see papers [16, 29, 19], books [8, 24] and references therein. But in big majority of the corresponding works the fast motion of angles φ is given not by an ODE, but by a stochastic differential equation with I-depending coefficients and non-degenerate diffusion. Even if the fast motion of φ was given by a non-random ODE, then usually strong restrictions were imposed on ergodic properties of the latter (e.g. see [27, Section II.3]), and system (1.3) fails to meet them. Still, a local version of system (1.2) was considered in [9], where for processes $(I, \varphi)(t)$ whose I-components are defined in a bounded domain the authors proved convergence (1.5) till the time of exit of the vector $I(t)$ from the bounded domain above. However, the techniques used in [9] differ essentially from those in our proof. The exact statement of Theorem 4.3 and its proof, given in Sections 3-4, are crucially used in Sections 6-10 of our work, which form its main second part (discussed below). The uniform in time convergence (1.8) seems to be a completely new result.

Second part of the paper, made by Sections 6-10, is dedicated to equation (1.4), where we again assume that the frequency mapping W is Anosov–nondegenerate and that equation (1.4) satisfies some other restrictions, similar to those, imposed on system (1.2). It is convenient to rewrite the equation in terms of the action-angle variables $(I, \varphi) \in \mathbb{R}_+^n \times \mathbb{T}^n$, where $I_k = \frac{1}{2}\|\mathbf{v}_k\|^2$ and $\varphi_k = \arctan(v_k/v_{-k})$. Then the corresponding I-equation reads

$$dI_k = \varepsilon P_k^I(I, \varphi)dt + \sqrt{\varepsilon} B_{kj}^I(I, \varphi)d\beta_j(t), \quad k = 1, \dots, n, \tag{1.9}$$

where P^I and B^I are quadratic functions of $v, \mathbf{P}_k(v)$ and B_{kj}, while the φ-equation is

$$d\varphi_k(t) = W_k(I)dt + [\dots], \tag{1.10}$$

where $[\dots]$ stands for the drift and dispersion terms of order ε and $\sqrt{\varepsilon}$, respectively, which are singular when $I_k = 0$. See equations (6.10) and (6.11). We wish to apply the results from the first part to this system of (I, φ)-equations. This is not at all

[4]Also known as the Kantorovich–Rubinstein distance.

[5]and without assuming that system (1.2) is mixing.

straightforward since the coefficients of the (I, φ)-system have singularities at the locus

$$\aleph = \{v \in \mathbb{R}^{2n} : \mathbf{v}_j = 0 \text{ for some } j\},$$

and the dispersion in (1.9) vanishes on $\aleph$. This difficulty is overcome with the help of two additional groups of results. Firstly, in the rather technical Lemma 6.8 we establishe that trajectories of eq. (1.4) stay in the vicinity of locus $\aleph$ only a short time, uniformly in ε. This result allows us to prove in Theorem 6.7 that if we write solutions I^ε of (1.9) as $I^\varepsilon(\varepsilon^{-1}\tau)$, $0 \le \tau \le T$, then their laws are pre-compact in the space of measures on $C(0,T;\mathbb{R}^n)$ and every limiting point of this family is a weak solution $I(\tau)$ of the corresponding averaged I-equation (see equation (6.12)). Unfortunately the latter is an equation without uniqueness of a solution (see Remark 6.5), so we cannot conclude that the laws $\mathcal{D}I^\varepsilon(\varepsilon^{-1}\tau)$ converge to a limit when $\varepsilon \to 0$. To resolve this problem we use the second additional result. Namely, in Section 7 we return from the (I, φ)-system (1.9)–(1.10) back to the original equation (1.4), remove from them the fast terms $W_k \mathbf{v}_k^\perp dt$ and re-write the obtained equation in the slow time τ as

$$dv(\tau) = P(v)d\tau + B(v)d\beta(\tau),$$

where $P = (\mathbf{P}_1, \dots, \mathbf{P}_n)$, $B = (B_{kj})$ and $\beta = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_{n_1})$. Then we formally average this equation with respect to the natural action of the torus $\mathbb{T}^n$ on $\mathbb{R}^{2n}$ (see (7.1)), using the rules of stochastic averaging. Thus we get a stochastic equation

$$dv(\tau) = \langle P \rangle(v)d\tau + \langle\!\langle B \rangle\!\rangle(v)d\beta(\tau), \tag{1.11}$$

which we call *the effective equation*. The coefficients of (1.11) are locally Lipschitz, so its solution, if exists, is unique. The key step of our analysis of the averaging in equation (1.4) is Theorem 8.2, stating that the weak solution $I(\tau)$ of the averaged I-equation, which has been obtained as a limit in law of some sequence I^{ε_j} of solutions for (1.9), written in the slow time τ, can be lifted to a weak solution $v(\tau)$ of (1.11) (i.e. $I(\tau) = I(v(\tau))$). Since a solution $v(\tau)$ of (1.11) is unique, then as a consequence we get in Theorem 8.6 that the actions I^ε of solutions for equation (1.4), written in slow time τ, converge in law to a limit:

$$\mathcal{D}I^\varepsilon(\varepsilon^{-1}\tau) \rightharpoonup \mathcal{D}I(\tau) \quad \text{for } 0 \le \tau \le T \quad \text{as } \varepsilon \to 0, \tag{1.12}$$

where $I(\tau)$ is a weak solution of the averaged I-equation. The latter can be lifted to a unique weak solution $v(\tau)$ of (1.11).

Similarly to the uniform convergence (1.8), we show in Theorem 10.4 that if some few moments of solutions for equation (1.4) are bounded uniformly in time and in ε, and if the effective equation (1.11) is mixing, then convergence (1.12) is uniform in $\tau \ge 0$.

Proposition 10.9 provides an easy sufficient condition which implies that all the assumptions, required for the validity of results in Sections 6-10 are met. In Section 10.2 we discuss applications of our results to damped/driven Hamiltonian systems.

When we apply Theorem 8.6 (which ensures convergence (1.12)) to stochastic perturbations of "real" integrable Hamiltonian equations, we arrive at a difficulty due to the fact that often integrable Hamiltonian equations which appear in mechanics and physics can be put to the Birkhoff normal form (1.3) (with $W = \nabla h$) not on the whole $\mathbb{R}^{2n}$, but only locally.[6] Vey's theorem (see [7], [10, Section 2.3]) and references there) provides an instrumental sufficient condition which allows to write an integrable Hamiltonian equation in the form (1.3) in a neighbourhood of some point. Without lost of generality we assume that this point is the origin, and

[6] But see Example 6.3 in Section 6 for a class of equations (1.4) in $\mathbb{R}^{2n}$ which appear in the non-equilibrium statistical physics.

assume that the equation is written in the from (1.3) in the closure $\bar{\mathfrak{B}}_R$ of domain $\mathfrak{B}_R = \{v : \|I(v)\| < R\}$, for some $R > 0$. Accordingly, in Section 9 we discuss a local problem of examining equation (1.3) in $\bar{\mathfrak{B}}_R$. There for any initial data $v_0 \in \mathfrak{B}_R$ we consider the exit time θ_R^ε of a corresponding trajectory $v^\varepsilon(t)$ of (1.4) from $\mathfrak{B}_R$. We show that, firstly, θ_R^ε is a random variable of order ε^{-1} and, secondly, that for the trajectory v^ε, stopped at $t = \theta_R^\varepsilon$, a variation of Theorems 8.6 applies and implies that the action-vector $I(v^\varepsilon(t))$, written in the slow time $\tau = \varepsilon t$, for $\tau \le \varepsilon\theta_R^\varepsilon$ converges in distribution to a solution $I(\tau)$ of the averaged I-equation, stopped when $\|I\| = R$. Similar to Theorem 8.6, this solution $I(\tau)$ may be lifted to a solution $v(\tau)$ of the effective equation (1.11), stopped at $\partial\mathfrak{B}_R$.

Long time behaviour of deterministic perturbations of integrable systems (1.3) (in various parts of the phase-space $\mathbb{R}^{2n}$) is a classical problem of dynamical systems. If the initial data are allowed to be arbitrarily close to the locus $\aleph$, additional difficulty appears. In the case of Hamiltonian perturbations see [1, Section 6.3.7] for corresponding KAM theorems, and see [26] for a version of Nekhoroshev's theorem. It seems that for non-Hamiltonian perturbations of (1.3) no convenient averaging theorem, valid up to $\aleph$, is known. Similarly it seems that no systematical study of the averaging for stochastic perturbations of system (1.3) was performed before our work.

On the proofs. Our presentation and proofs are based on the approach to stochastic averaging, originated in the celebrated paper [16] by R. Khasminskii. In that we partially follow our previous works [11, 13, 23, 22], dedicated to the averaging in stochastic perturbations of linear (the first two paper) and non-linear (the last two) PDEs. The Khasminskii approach, as presented in our work, is a flexible form of stochastic averaging, applicable to study stochastic perturbations of integrable equations, linear and nonlinear, in finite and infinite dimensions. In particular, to study stochastic perturbations of linear and integrable PDEs. See [13, 11] and references in [11] for perturbations of linear PDEs, including "linear analogies" for them of Theorems 8.6 and 10.4. Also see [23] for an analogy of Theorem 4.3 for perturbations of the KdV equation by dissipation $\varepsilon\Delta$ and a white noise of order $\sqrt{\varepsilon}$, and see [22] for an analogy for that equation of Theorems 8.2 and 8.6. [7]

Notation. For a matrix A we denote by A^t its transposed and by $\mathbb{R}^n_+$ denote the set of vectors in $\mathbb{R}^n$ with non-negative components. For a Banach space E and $R > 0$, $B_R(E)$ stands for the open R-ball $\{e \in E : |e|_E < R\}$, and $\bar{B}_R(E)$ – for its closure $\{|e|_E \le R\}$; $C_b(E)$ stands for the space of bounded continuous function on E, and $C([0,T],E)$ – for the space of continuous curves $[0,T] \to E$, given the sup-norm. By $\mathcal{D}(\xi)$ we denote the law of a random variable ξ, by $\rightharpoonup$ – the weak convergence of measures, and by $\mathcal{P}(M)$ – the space of Borel probability measures on a metric space M. For a measurable mapping $F : M_1 \to M_2$ and $\mu \in \mathcal{P}(M_1)$ we denote by $F \circ \mu \in \mathcal{P}(M_2)$ the image of μ under F; i.e. $F \circ \mu(Q) = \mu(F^{-1}(Q))$.

For $v = (v_1, \dots, v_n) \in \mathbb{R}^n$ we set $\|v\|^2 = \sum_{k=1}^n |v_k|^2$ and $|v| = \sum_{k=1}^n |v_k|$. We denote $M_{n\times k}$ the space formed by $n\times k$ real matrices with the Hilbert-Schmidt norm $\|\cdot\|_{HS}$, i.e. the square root of the sum of the squares of all its elements. If $L = \mathbb{R}^d \times \mathbb{T}^n$, $n \ge 0$ (we set $\mathbb{T}^0 := \{0\}$) and $m \ge 0$, then $\mathrm{Lip}_m(L, E)$ is the collection of locally Lipschitz maps $F : L \to E$ such that for any $R > 0$ we have

$$(1 + |R|)^{-m} \sup_{\xi \in \bar{B}_R(\mathbb{R}^d)\times\mathbb{T}^n} |F(\xi)|_E =: \mathcal{C}^m(F) < \infty. \tag{1.13}$$

For a set Q we denote by $\mathbf{1}_Q$ its indicator function, by Q^c – its complement, and by $\mathcal{L}(Q)$ – its Lebesgue measure if $Q \subset \mathbb{R}^n$. For a function f, depending on angles

[7]Also see [10, Section 4.3] for a discussion of the results in [23, 22].

$\varphi \in \mathbb{T}^n$ (and maybe on some other variables) we denote

$$\langle f \rangle = (2\pi)^{-n} \int_{\mathbb{T}^n} f \, d\varphi. \tag{1.14}$$

Finally, for real numbers a and b, $a \vee b$ and $a \wedge b$ indicate their maximum and minimum.

2. Problem setup

Let $d_1 \in \mathbb{N}$, $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, $\beta(t)$, $t \geq 0$, be a standard d_1-dimensional Brownian motion defined on it, and $\{\mathcal{F}_t\}$ be the natural filtration, generated by the process $\beta(s)$, $0 \leqslant s \leqslant t$.

We start to examine a diffusive perturbation (1.2) of integrable systems (1.1). There $\varepsilon \in (0,1]$ is a small parameter, P^I is a d-dimensional vector function, θ and P^φ are n-dimensional functions, while $\Psi^I(\cdot)$ and $\Psi^\varphi(\cdot)$ are $d \times d_1$ - and $n \times d_1$-matrix functions. Our first goal is to study system (1.2) for $0 \leq t \lesssim \varepsilon^{-1}$. After passing to the slow time $\tau = \varepsilon t$, the system takes the form

$$\begin{aligned} dI^\varepsilon &= P^I(I^\varepsilon, \varphi^\varepsilon) d\tau + \Psi^I(I^\varepsilon, \varphi^\varepsilon) d\beta(\tau), \\ d\varphi^\varepsilon &= \left[\tfrac{1}{\varepsilon}\theta(I^\varepsilon) + P^\varphi(I^\varepsilon, \varphi^\varepsilon)\right] d\tau + \Psi^\varphi(I^\varepsilon, \varphi^\varepsilon) d\beta(\tau), \end{aligned} \tag{2.1}$$

where $0 \leq \tau \leq T$ for some fixed $T > 0$. It is equipped with an initial condition

$$I^\varepsilon(0) = I_0, \quad \varphi^\varepsilon(0) = \varphi_0. \tag{2.2}$$

Here $(I_0, \varphi_0) \in \mathbb{R}^d \times \mathbb{T}^n$ is either deterministic, or is a r.v., independent of the process β. We will mostly dwell on the first case since the second can be directly generalized from the first one (see Remark 4.4 and Amplification 8.8). Our goal is to examine the limiting behaviour of the distribution of a solution $(I^\varepsilon(\cdot), \varphi^\varepsilon(\cdot))$ as $\varepsilon \to 0$. In particular, to show that in the limit the law of $I^\varepsilon(\cdot)$ is a weak solution of a certain averaged equation, independent of φ and ε.

In what follows we always assume that the following conditions are fulfilled for system (2.1)–(2.2):

Assumption 2.1. (1) The Lebesgue measure of $I \in \mathbb{R}^d$ for which $\theta(I)$ is rationally dependent equals zero, that is $\mathcal{L}\big(\bigcup_{k \in \mathbb{Z}^n \setminus \{0\}} \{I \in \mathbb{R}^d : k \cdot \theta(I) = 0\}\big) = 0$.

(2) The matrix $a(I,\varphi) = \big(a_{ij}(I,\varphi)\big) := \Psi^I(I,\varphi)(\Psi^I)^t(I,\varphi)$ satisfies the uniform ellipticity condition, that is, there exists $\lambda > 0$ such that

$$\lambda |\xi|^2 \leq a(I,\varphi)\xi \cdot \xi \leq \lambda^{-1} |\xi|^2$$

for all $\xi \in \mathbb{R}^d$ and all $(I,\varphi) \in \mathbb{R}^d \times \mathbb{T}^n$.

(3) There exists $q > 0$ such that $\theta \in \mathrm{Lip}_q(\mathbb{R}^d, \mathbb{R}^n)$, $P^m \in \mathrm{Lip}_q(\mathbb{R}^d \times \mathbb{T}^n, \mathbb{R}^d)$ and $\Psi^m \in \mathrm{Lip}_q(\mathbb{R}^d \times \mathbb{T}^n, M_{d_m \times d_1})$, $m = I, \varphi$ and $d_I = d, d_\varphi = n$ (see (1.13)).

(4) There exists $T > 0$ such that for every $(I_0, \varphi_0) \in \mathbb{R}^d \times \mathbb{T}^n$, system (2.1)–(2.2) has a unique strong solution $\big(I^\varepsilon(\tau), \varphi^\varepsilon(\tau)\big) := \big(I^\varepsilon, \varphi^\varepsilon\big)(\tau; I_0, \varphi_0)$, $\tau \in [0,T]$, equal (I_0, φ_0) at $\tau = 0$. Moreover, there exists $q_0 > (q \vee 2)$ such that

$$\mathbf{E} \sup_{\tau \in [0,T]} |I^\varepsilon(\tau)|^{2q_0} \leq C(|I_0|, T) \qquad \forall \varepsilon \in (0,1], \tag{2.3}$$

where $C(\cdot)$ is a non-negative continuous function on $\mathbb{R}^2_+$ non-decreasing in both arguments.

Throughout the text except the last Section 10, the time $T > 0$ is fixed and the dependence on it usually is not indicated. The process $(I^\varepsilon(\tau), \varphi^\varepsilon(\tau))$, $\tau \in [0,T]$ is always understood as a unique strong solution of the system (2.1)–(2.2).

Remark 2.2. Item (1) in Assumption 2.1 is called *the Anosov condition*[8] and is rather mild. E.g. it clearly holds if the mapping θ is C^1-smooth, $n \geq d$ and the set of I's for which rank $\partial_I \theta(I) < d$ has zero measure. But it also may hold for systems (1.1) with $n < d$ (even for systems with $n = 1$). In particular, it holds for any n, d if the mapping θ is analytic and satisfied *Rüssmann's condition*: there exists $N \in \mathbb{N}$ such that for every $I \in \mathbb{R}^d$ the vectors

$$\frac{\partial^{|q|}\theta(I)}{\partial_1^{q_1}\dots\partial_d^{q_d}}, \quad q \in \mathbb{Z}_+^d, \quad |q| \leq N,$$

jointly span $\mathbb{R}^n$. Indeed, if this condition holds, then for any non-zero vector $s \in \mathbb{R}^n$ the function $\theta(I)\cdot s$ does not vanish identically (see [1, Section 6.3.2], item 7°). Then by analyticity the set in (1) has zero measure.

Since item (4) of Assumption 2.1 is formulated not in terms of coefficients of equations (2.1), we provide here a sufficient condition that ensures its validity.

Proposition 2.3. *Let all coefficients of equation* (2.1) *be globally Lipschitz continuous. Then for any $T > 0$ and $(I_0, \varphi_0) \in \mathbb{R}^d \times \mathbb{T}^n$ problem* (2.1)–(2.2) *has a unique solution, and inequality* (2.3) *holds for every $q_0 \in \mathbb{N}$.*

Proof. Under the above assumptions, for any (I_0, φ_0) system (2.1)-(2.2) has a unique solution.[9] For $q_0 \in \mathbb{N}$ we need to show that estimate (2.3) holds with a constant $C(q_0, I_0, T)$ that does not depend on ε. Let us fix an arbitrary $R > \|I_0\|^{2q_0}$ and introduce the stopping time $\tau_R^\varepsilon = \inf\{\tau > 0 : \|I^\varepsilon(\tau)\|^{2q_0} > R\}$. By Itô's formula, the process $\|I^\varepsilon(\tau \wedge \tau_R^\varepsilon)\|^{2q_0}$ satisfies the equation

$$\begin{aligned}
&\|I^\varepsilon(\tau\wedge\tau_R^\varepsilon)\|^{2q_0}\\
=&\|I_0\|^{2q_0} + 2q_0\int_0^{\tau\wedge\tau_R^\varepsilon}\|I^\varepsilon(s)\|^{2(q_0-1)}\Big(I^\varepsilon(s), P^I(I^\varepsilon(s),\varphi^\varepsilon(s))\Big)ds\\
&+\frac{1}{2}\int_0^{\tau\wedge\tau_R^\varepsilon}\mathrm{Trace}\big(\Psi^I(I^\varepsilon,\varphi^\varepsilon)\big)^t\nabla_I^2\|I^\varepsilon\|^{2q_0}\Psi^I(I^\varepsilon,\varphi^\varepsilon)ds\\
&+\int_0^{\tau\wedge\tau_R^\varepsilon}2q_0\|I^\varepsilon(s)\|^{2(q_0-1)}\Big(I^\varepsilon(s),\Psi^I(I^\varepsilon(s),\varphi^\varepsilon(s))d\beta(s)\Big).
\end{aligned} \tag{2.4}$$

Taking the expectation and using the global Lipschitz continuity of coefficients we conclude that

$$\mathbf{E}\|I^\varepsilon(\tau\wedge\tau_R^\varepsilon)\|^{2q_0} \leqslant C_1 + C_2\int_0^\tau \mathbf{E}\|I^\varepsilon(s\wedge\tau_R^\varepsilon)\|^{2q_0}ds$$

with constants C_1 and C_2 that depend continuously on q_0, $\|I_0\|$ and the global Lipschitz constants of the coefficients and do not depend on ε and on R. By Gronwall's lemma,

$$\mathbf{E}\|I^\varepsilon(\tau\wedge\tau_R^\varepsilon)\|^{2q_0} \leqslant C_1\exp(C_2\tau).$$

Since for any fixed τ the sequence $\tau\wedge\tau_R^\varepsilon$ a.s. converges to τ as $R\to\infty$, then by the Fatou lemma,

$$\mathbf{E}\|I^\varepsilon(\tau)\|^{2q_0} \leqslant C_1\exp(C_2\tau),\ \forall\tau\geqslant 0. \tag{2.5}$$

Now consider equation (2.4) without the stopping time τ_R^ε. For any $T > 0$, we have

$$\mathbf{E}\sup_{0\leqslant\tau\leqslant T}\|I^\varepsilon(\tau)\|^{2q_0} \leqslant \|I_0\|^{2q_0} + C\int_0^T\mathbf{E}\|I^\varepsilon(\tau)\|^{2q_0}d\tau + C\mathbf{E}\sup_{\tau\in[0,T]}M(\tau),$$

[8] "The set of slow variables I for which the motion of fast variable φ is not ergodic has zero measure," see [25], p. 12, assumption iii).

[9] To prove this we regard (2.1) as an equation on $\mathbb{R}^d\times\mathbb{R}^n$ with periodic in φ coefficients, evoke the usual theorem on stochastic equations in $\mathbb{R}^{d+n}$ with Lipschitz coefficients (see for instance [15, Theorem 2.9]) to get a solution for this equation, and next apply the projection $\mathbb{R}^d\times\mathbb{R}^n\to\mathbb{R}^d\times\mathbb{T}^n$ to obtain a solution in $\mathbb{R}^d\times\mathbb{T}^n$.

where $M(\tau)$ is the martingale term in (2.4), and the constant C does not depend on ε. Due to (2.5), the estimate (2.3) follows by applying Doob's inequality to $M(\tau)$. □

Assumption 2.1.(4) also holds for systems (1.2) where the coefficients are not globally Lipschitz, but items (2) and (3) of the assumption are valid and the vector field P is coercive. Cf. below Proposition 10.9, where this is discussed for system (1.4).

3. Tightness and averaged equation

In this section we first show the collection of the laws of the processes $I^\varepsilon(\tau), \tau \in [0,T]$ with $\varepsilon \in (0,1]$ is tight. Then we introduce the averaged equation and finally prove key technical statements.

Since $T > 0$ is fixed, then dependence of constants on it usually is not indicated.

3.1. **Tightness.** Notice that under condition (3) of Assumption 2.1 there exists an increasing function $\nu(M) : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\nu(M) \to \infty$ as $M \to \infty$, and

$$\begin{aligned}&\text{in the set } \{(I,\varphi) : |I| \le \nu(M)\} \text{ norms and Lipschitz constants}\\ &\qquad\text{of all coefficients in equations (2.1) do not exceed } M\end{aligned} \tag{3.1}$$

(the coefficient $\frac{1}{\varepsilon}\theta$ should be taken without the factor ε^{-1}). This a bit unusual way to write the locally Lipschitz property is convenient for the calculation below.

Lemma 3.1. *For any fixed (I_0, φ_0) the family of laws of processes $I^\varepsilon(\tau)$ ($\tau \in [0,T], \varepsilon \in (0,1]$) which are the I-components of solutions $(I^\varepsilon, \varphi^\varepsilon)(\tau; I_0, \varphi_0)$, is tight in the Banach space $C([0,T];\mathbb{R}^d)$.*

Proof. According to (4) of Assumption 2.1, for any $\delta > 0$ there exists $R = R(\delta) > 0$ such that $\mathbf{P}\{\sup_{\tau\in[0,T]} |I^\varepsilon(\tau)| > R\} < \delta$. Denoting by τ_R^ε the exit time $\tau_R^\varepsilon = \inf\{\tau > 0 : |I^\varepsilon(\tau)| > R\}$ and denoting by $(I_R^\varepsilon, \varphi_R^\varepsilon)$ the stopped process $(I^\varepsilon(\tau \wedge \tau_R^\varepsilon), \varphi(\tau \wedge \tau_R^\varepsilon))$, we have $\mathbf{P}\{(I^\varepsilon(\cdot), \varphi^\varepsilon(\cdot)) \ne (I_R^\varepsilon(\cdot), \varphi_R^\varepsilon(\cdot)) \text{ on } [0,T]\} < \delta$.

By (3) of Assumption 2.1 the functions P^I and Ψ^I are bounded in the ball $\{|I| \le R(\delta)\}$. Therefore for any moments $0 \le \tau_1 \le \tau_2 \le T$ we have

$$\begin{aligned}\mathbf{E}\big\{|I_R^\varepsilon(\tau_2) - I_R^\varepsilon(\tau_1)|^4\big\} &= \mathbf{E}\bigg\{\Big|\int\limits_{\tau_1\wedge\tau_R^\varepsilon}^{\tau_2\wedge\tau_R^\varepsilon} P^I(I^\varepsilon,\varphi^\varepsilon)ds + \int\limits_{\tau_1\wedge\tau_R^\varepsilon}^{\tau_2\wedge\tau_R^\varepsilon} \Psi^I(I^\varepsilon,\varphi^\varepsilon)d\beta(s)\Big|^4\bigg\}\\ &\le C_R(|\tau_2-\tau_1|^4 + |\tau_2-\tau_1|^2),\end{aligned}$$

and the assertion follows by a direct application of the Prokhorov theorem, cf. [12, Lemma 2.2]. □

3.2. **The averaged equation.** The vector field $P^I(I,\varphi)$ in the I-equation of (2.1) depends explicitly both on I-variables and φ-variables. Since we are interested in the evolution of the I-component of the system as $\varepsilon \to 0$, then we introduce in consideration the vector field $\langle P^I\rangle$, obtained by the averaging of P^I in angles φ,

$$\langle P^I\rangle(I) = \int_{\mathbb{T}^n} P^I(I,\varphi)\,d\varphi.$$

If an I is such that the vector $\theta(I)$ is rationally independent, then

$$\lim_{N\to\infty} \frac{1}{N}\int_0^N P^I\big(I, \varphi + t\theta(I)\big)\,dt = \langle P^I\rangle(I), \tag{3.2}$$

and the convergence is uniform in φ. Moreover, $\langle P^I\rangle \in \mathrm{Lip}_q(\mathbb{R}^d, \mathbb{R}^d)$ with the same q as for P^I, $\mathcal{C}^q(\langle P^I\rangle) = \mathcal{C}^q(P^I)$, and for $\langle P^I\rangle$ the function $\nu(N)$ (as in (3.1) is the same as for P^I; see [12, Section 3.2]. The convergence to the limit on the l.h.s. of

(3.2) is the faster, the more diophantine is the vector $\theta(I)$ (see [1, Section 6.1.5] for a related discussion).

Similarly we set

$$\langle a^I\rangle(I) = \int_{\mathbb{T}^n} \Psi^I(I,\varphi)\big(\Psi^I(I,\varphi)\big)^t\, d\varphi.$$

By (2) of Assumption 2.1,

$$\lambda|\xi|^2 \leqslant \langle a^I\rangle(I)\xi\cdot\xi \leqslant \lambda^{-1}|\xi|^2, \quad \forall \xi, I \in \mathbb{R}^d.$$

Let $\langle\!\langle \Psi^I\rangle\!\rangle(I)$ be the principal square root of $\langle a^I\rangle(I)$.[10] By the estimates above and [28, Theorem 5.2.1], $\langle\!\langle \Psi^I\rangle\!\rangle(\cdot)$ belongs to $\mathrm{Lip}_q(\mathbb{R}^d, M_{d\times d})$ and satisfies the uniform ellipticity condition as in (2) of Assumption 2.1 (with the same $\lambda > 0$).

We introduce the averaged equation for the limiting as $\varepsilon \to 0$ evolution of I-variables,

$$dI(\tau) = \langle P^I\rangle(I(\tau))d\tau + \langle\!\langle \Psi^I\rangle\!\rangle(I(\tau))dW(\tau), \quad I(0) = I_0, \tag{3.3}$$

where $W(\tau)$ is a standard d-dimensional Wiener process. As we discussed above, coefficients of (3.3) are locally Lipschitz, therefore for each $I_0 \in \mathbb{R}^d$ the equation has at most one solution $I(\tau; I_0)$. We will show in the next section that as $\varepsilon \to 0$, the I-component $I^\varepsilon(\tau)$ of a solution for (2.1) with $(I(0), \varphi(0)) = (I_0, \varphi_0)$, for any $\varphi_0 \in \mathbb{T}^n$ converges in law to solution $I(\tau; I_0)$ (in particular, the latter exists). The key technical results, needed to establish this convergence, are proved in the next subsection.

3.3. Main lemmas.

Lemma 3.2. *For the process $(I^\varepsilon(\tau), \varphi^\varepsilon(\tau))$, $\tau \in [0, T]$, we have*

$$\Upsilon^\varepsilon := \mathbf{E} \max_{0\le\tau\le T} \Big| \int_0^\tau P^I(I^\varepsilon(s), \varphi^\varepsilon(s))\, ds - \int_0^\tau \langle P^I\rangle(I^\varepsilon(s))\, ds\Big| \to 0 \quad \text{as } \varepsilon \to 0. \tag{3.4}$$

A proof of this relation relies on a number of auxiliary statements, given below.

Let us first define a family of sets $\mathcal{A}^\delta_{N,R}$, where $N, R > 0$ and $0 < \delta \le 1$, made by vectors I such that $\theta(I)$ has poor diophantine properties, and so the rate of convergence in (3.2) is slow:

$$\mathcal{A}^\delta_{N,R} = \Big\{ I \in \mathbb{R}^d \,:\, |I| < R,\ \max_{\varphi\in\mathbb{T}^n} \Big| \frac{1}{N}\int_0^N P^I(I, \varphi + t\theta(I))\, dt - \langle P^I\rangle(I)\Big| > \delta \Big\}. \tag{3.5}$$

We set $\mathcal{A}^\delta_N := \mathcal{A}^\delta_{N,\infty} = \cup_{R>0}\mathcal{A}^\delta_{N,R}$.

Lemma 3.3. *For any $\delta > 0$ and $R > 0$ we have*

$$\lim_{N\to\infty} \mathcal{L}(\mathcal{A}^\delta_{N,R}) = 0$$

Proof. For $N > 0$ denote $b_N(I) = \max_{\varphi\in\mathbb{T}^n} \Big| \frac{1}{N}\int_0^N P^I(I, \varphi + t\theta(I))dt - \langle P^I\rangle(I)\Big|$. If the vector $\theta(I)$ is non-resonant, then by (3.2) $b_N(I) \to 0$ as $N \to \infty$, So, by (1) of Assumption 2.1, $b_N(I) \to 0$ a.s., and the assertion follows since the a.s. convergence implies the convergence in measure. □

Lemma 3.4. *For any $R, N, \delta > 0$ the probability $\mathbf{P}\{I^\varepsilon(\tau) \in \mathcal{A}^\delta_{N,R}\}$ admits the upper bound*

$$\mathbf{E}\int_0^{T\wedge\tau^\varepsilon_R} \mathbf{1}_{\mathcal{A}^\delta_{N,R}}(I^\varepsilon(\tau))\, d\tau \le C(R)\big(\mathcal{L}(\mathcal{A}^\delta_{N,R})\big)^{\frac{1}{d}}, \tag{3.6}$$

where τ^ε_R is the exit time of I^ε from the ball $B_R = \{I \in \mathbb{R}^d : |I| < R\}$. The constant $C(R)$ does not depend on ε, N and δ.

[10] I.e. $\langle\!\langle \Psi^I\rangle\!\rangle(I)$ is a non-negative self-adjoint matrix such that $\big(\langle\!\langle \Psi^I\rangle\!\rangle\big)^2 = \langle a^I\rangle$.

Proof. Since the process $I^\varepsilon(\tau)$ satisfies Itô's equation $dI = P^I d\tau + \Psi^I d\beta(\tau)$, where the diffusion matrix $\Psi^I(\Psi^I)^t$ is uniformly non-degenerate by (2) of Assumption 2.1, then the assertion follows from [20, Theorem 2.2.4]. Indeed, by the latter result, choosing there $f(\tau,\cdot)$ to be the characteristic function of the set $\mathcal{A}^\delta_{N,R}$, we obtain

$$\mathbf{E}\int_0^{T\wedge\tau_R^\varepsilon} \mathbf{1}_{\mathcal{A}^\delta_{N,R}}(I^\varepsilon(\tau))\,d\tau \le C(R)\|\mathbf{1}_{\mathcal{A}^\delta_{N,R}}\|_{L^d(B_R)} = C(R)\big(\mathcal{L}(\mathcal{A}^\delta_{N,R})\big)^{\frac{1}{d}},$$

and the assertion of the lemma is proved. □

Due to (3.6),

$$\int_0^T \mathbf{P}\{I^\varepsilon(\tau)\in\mathcal{A}^\delta_N\}\,d\tau \le C(R)\big(\mathcal{L}(\mathcal{A}^\delta_{N,R})\big)^{\frac{1}{d}} + T\mathbf{P}\{\tau_R^\varepsilon < T\}. \tag{3.7}$$

Since $\{\omega : \tau_R^\varepsilon < T\}\subset\{\omega : \sup_{0\le\tau\le T}|I^\varepsilon(\tau)|\ge R\}$, then by (4) of Assumption 2.1 and Chebyshev's inequality we get that probability $\mathbf{P}\{\tau_R^\varepsilon < T\}$ tends to zero as $R\to\infty$, uniformly in $\varepsilon\in(0,1]$. Combining this fact with Lemma 3.3 to estimate the r.h.s. of (3.7) we conclude that for any $\delta>0$ there exists a positive function $N\mapsto\alpha^\delta_N$, $N>0$, converging to zero as $N\to\infty$, such that

$$\int_0^T \mathbf{P}\{I^\varepsilon(\tau)\in\mathcal{A}^\delta_N\}\,d\tau \le \alpha^\delta_N, \quad \forall\,\varepsilon\in(0,1]. \tag{3.8}$$

We are now in position to prove Lemma 3.2

Proof. [of Lemma 3.2] **Step 1:** Take $N>0$ and $\tau_0\in[0,\varepsilon N)$ which will be specified later and consider a partition of $[\tau_0,T)$ to intervals $[\tau_j,\tau_{j+1}) =: \Delta_j$, $0\le j\le j_N$, where $\tau_j = \tau_0 + j\varepsilon N$, j_N is the biggest j such that $\tau_j < T$, and $\tau_{j_N+1} := T$. We assume that $\varepsilon N \le \frac{1}{3}(1\wedge T)$ – below we deal with N's such that $\varepsilon N \ll 1$, so this assumption is not a restriction. Then

$$j_N \le 2/(\varepsilon N).$$

Next let us introduce the random variable

$$L^{\varepsilon,\delta}_{N,\tau_0}(\omega) = \#\big\{j\in[0,j_N] : I^\varepsilon(\tau_j)\in\mathcal{A}^\delta_N\big\},$$

which counts the moments τ_j's for which the frequency vector $\theta(I^\varepsilon(\tau_j))$ has poor diophantine properties. Since

$$L^{\varepsilon,\delta}_{N,\tau_0} = \sum_{j=0}^{j_N}\mathbf{1}_{\mathcal{A}^\delta_N}(I^\varepsilon(\tau_j)),$$

we obtain

$$\mathbf{E}L^{\varepsilon,\delta}_{N,\tau_0} = \sum_{j=0}^{j_N}\mathbf{P}\{I^\varepsilon(\tau_j)\in\mathcal{A}^\delta_N\}.$$

Integration of this equality in the variable τ_0 over interval $[0,\varepsilon N)$ yields

$$\int_0^{\varepsilon N}\mathbf{E}L^{\varepsilon,\delta}_{N,\tau_0}\,d\tau_0 = \sum_{j=0}^{j_N}\int_0^{\varepsilon N}\mathbf{P}\{I^\varepsilon(\tau_j(\tau_0))\in\mathcal{A}^\delta_N\}\,d\tau_0$$
$$= \int_0^T \mathbf{P}\{I^\varepsilon(\tau)\in\mathcal{A}^\delta_N\}\,d\tau \le \alpha^\delta_N,$$

where the last estimate follows from (3.8). Therefore there exists a *non-random* number $\tau_0^*\in[0,\varepsilon N)$ such that

$$\mathbf{E}L^{\varepsilon,\delta}_{N,\tau_0^*} \le \alpha^\delta_N(\varepsilon N)^{-1}. \tag{3.9}$$

From now on we fix this τ_0^* for the choice of τ_0 in the definition of the partition $\{\Delta_j\}$ of interval $[\tau_0,T)$. So from now on $\tau_j := \tau_0^* + j\varepsilon N$ for all j, and below we write $L^{\varepsilon,\delta}_{N,\tau_0}$ simply as $L^{\varepsilon,\delta}_N$.

Step 2: We define the first good event for our argument (there will be three of them) as a collection $\mathcal{E}_1$ of all ω such that $L_N^{\varepsilon,\delta}(\omega)$ is relatively small:

$$\mathcal{E}_1 = \big\{L_N^{\varepsilon,\delta} \le \ (\alpha_N^\delta)^{\frac{1}{2}}(\varepsilon N)^{-1}\big\}.$$

In view of (3.9) and Chebyshev's inequality, for the complement $\mathcal{E}_1^c = \Omega \smallsetminus \mathcal{E}_1$ we have

$$\mathbf{P}(\mathcal{E}_1^c) \le (\alpha_N^\delta)^{\frac{1}{2}}. \tag{3.10}$$

Due to (3) and (4) of Assumption 2.1, for any j the stochastic terms on the right-hand sides of (2.1) admit the following upper bounds:

$$\mathbf{E}\Big(\Big|\int_{\tau_j}^{\tau_{j+1}} \Psi^m(I^\varepsilon(s),\varphi^\varepsilon(s))\,d\beta(s)\Big|^2\Big) \le \mathbf{E}\int_{\tau_j}^{\tau_{j+1}} |\Psi^m(I^\varepsilon(s),\varphi^\varepsilon(s))|^2\,ds$$
$$\le C\int_{\tau_j}^{\tau_{j+1}} \mathbf{E}\big((|I^\varepsilon|+1)^{2q}\big)d\tau \le C_1\varepsilon N, \qquad m = I,\,\varphi.$$

Therefore, by Doob's inequality we have

$$\mathbf{E}\Big(\sup_{\tau\in\Delta_j}\Big|\int_{\tau_j}^{\tau} \Psi^m(I^\varepsilon(s),\varphi^\varepsilon(s))\,d\beta(s)\Big|^2\Big) \le C_2\varepsilon N, \qquad m = I,\,\varphi. \tag{3.11}$$

Similar, by (3) and (4) of Assumption 2.1 we have

$$\begin{aligned}\mathbf{E}\Big(\sup_{\tau\in\Delta_j}\Big|\int_{\tau_j}^{\tau} P^m(I^\varepsilon(s),\varphi^\varepsilon(s))\,ds\Big|^2\Big) &\le \varepsilon N\mathbf{E}\Big(\int_{\tau_j}^{\tau_{j+1}} \big|P^m(I^\varepsilon(s),\varphi^\varepsilon(s))\big|^2\,ds\Big)\\ &\le \varepsilon NC\int_{\tau_j}^{\tau_{j+1}} \mathbf{E}\big((|I^\varepsilon|+1)^{2q}\big)d\tau \le C_1(\varepsilon N)^2, \qquad m = I,\,\varphi.\end{aligned} \tag{3.12}$$

Step 3: For $j = 0,\dots,j_N$ we measure the size of the perturbative part in eq. (2.1) on interval Δ_j by the random variable

$$\zeta_j = \sum_{m=I,\varphi}\Big(\sup_{\tau\in\Delta_j}\Big|\int_{\tau_j}^{\tau} \Psi^m(I^\varepsilon(s),\varphi^\varepsilon(s))\,d\beta(s)\Big| + \sup_{\tau\in\Delta_j}\Big|\int_{\tau_j}^{\tau} P^m(I^\varepsilon(s),\varphi^\varepsilon(s))\,ds\Big|\Big),$$

and introduce another random variable which counts the number of large ζ_j's:

$$\widetilde{L}_N^{\varepsilon,\delta} = \#\big\{j : \zeta_j \ge \varepsilon^{\frac{1}{4}}\big\}.$$

We have $\mathbf{E}\zeta_j \le C_3(\varepsilon N)^{\frac{1}{2}}$. Indeed, $\mathbf{E}\zeta_j \le \big(\mathbf{E}\zeta_j^2\big)^{\frac{1}{2}}$ and due to (3.11) and (3.12)

$$\mathbf{E}\zeta_j^2 \le C\big(\varepsilon N + (\varepsilon N)^2\big) \le C\varepsilon N$$

(we recall that $\varepsilon N \le 1/3$). By Chebyshev's inequality $\mathbf{P}\{\zeta_j \ge \varepsilon^{\frac{1}{4}}\} \le C_3 N^{\frac{1}{2}}\varepsilon^{\frac{1}{4}}$, then

$$\mathbf{E}\widetilde{L}_N^{\varepsilon,\delta} = \mathbf{E}\Big(\sum_j \mathbf{1}_{\{\zeta_j\ge\varepsilon^{\frac{1}{4}}\}}\Big) = \sum_j \mathbf{E}\big(\mathbf{1}_{\{\zeta_j\ge\varepsilon^{\frac{1}{4}}\}}\big) = \sum_j \mathbf{P}\{\zeta_j \ge \varepsilon^{\frac{1}{4}}\} \le \frac{TC_3N^{\frac{1}{2}}\varepsilon^{\frac{1}{4}}}{\varepsilon N} = C_3TN^{-\frac{1}{2}}\varepsilon^{-\frac{3}{4}}.$$

Now we define the second good event for our argument as the set, where $\widetilde{L}_N^{\varepsilon,\delta}$ is not too big:

$$\mathcal{E}_2 = \big\{\omega\in\Omega : \widetilde{L}_N^{\varepsilon,\delta} \le \varepsilon^{-\frac{7}{8}}\big\}.$$

Again due to Chebyshev's inequality, the probability of its complement satisfies

$$\mathbf{P}(\mathcal{E}_2^c) \le C_3TN^{-\frac{1}{2}}\varepsilon^{\frac{1}{8}}. \tag{3.13}$$

Denote

$$\mathcal{M}_j := \{\omega\in\Omega : \zeta_j \le \varepsilon^{\frac{1}{4}} \quad \text{and} \quad \sup_{0\leqslant\tau\leqslant T}|I^\varepsilon(\tau)| \le \nu(M)\},$$

where $\nu(M)$ is defined in (3.1). In what follows we assume that $M \leqslant N$. Then for each $\omega \in \mathcal{M}_j$ on the interval Δ_j the curve $I(\tau)$ is close to $I(\tau_j)$:

$$\begin{aligned}&\sup_{\tau\in\Delta_j}\left|I^\varepsilon(\tau)-I^\varepsilon(\tau_j)\right|\\&\le \sup_{\tau\in\Delta_j}\Big|\int_{\tau_j}^{\tau}P^I(I^\varepsilon(s),\varphi^\varepsilon(s))\,ds\Big|+\sup_{\tau\in\Delta_j}\Big|\int_{\tau_j}^{\tau}\Psi^I(I^\varepsilon(s),\varphi^\varepsilon(s))\,d\beta(s)\Big|\le\varepsilon^{\frac14}.\end{aligned}\tag{3.14}$$

Due to the definition of $\nu(M)$, for $\omega\in\mathcal{M}_j$

$$\Big|\int_{\tau_j}^{\tau}(\theta(I^\varepsilon(s)-\theta(I^\varepsilon(\tau_j))\,ds\Big|\leqslant N\int_{\tau_j}^{\tau}\sup_{\tau\in\Delta_j}\left|I^\varepsilon(s)-I^\varepsilon(\tau_j)\right|ds\leqslant\varepsilon^{\frac54}N^2.$$

Therefore, for ω from the event $\mathcal{M}_j$, on the interval Δ_j the curve $\varphi(\tau)$ is close to its linear approximation:

$$\begin{aligned}&\sup_{\tau\in\Delta_j}\Big|\varphi^\varepsilon(\tau)-\varphi^\varepsilon(\tau_j)-\frac{\tau-\tau_j}{\varepsilon}\theta(I^\varepsilon(\tau_j))\Big|\\&\le\sup_{\tau\in\Delta_j}\Big|\int_{\tau_j}^{\tau}P^\varphi(I^\varepsilon(s),\varphi^\varepsilon(s))\,ds\Big|+\sup_{\tau\in\Delta_j}\Big|\frac1\varepsilon\int_{\tau_j}^{\tau}(\theta(I^\varepsilon(s)-\theta(I^\varepsilon(\tau_j))\,ds\Big|\\&\quad+\sup_{\tau\in\Delta_j}\Big|\int_{\tau_j}^{\tau}\Psi^\varphi(I^\varepsilon(s),\varphi^\varepsilon(s))\,d\beta(s)\Big|\le N^2\varepsilon^{\frac14}+2\varepsilon^{\frac14}.\end{aligned}\tag{3.15}$$

Step 4: Consider now a random collection of intervals $\Delta_j=[\tau_j,\tau_{j+1})$ such that

$$I^\varepsilon(\tau_j)\notin\mathcal{A}_N^\delta\ \text{(cf. event }\mathcal{E}_1),\ \ \zeta_j\le\varepsilon^{\frac14}\ \text{(cf. event }\mathcal{E}_2)\ \text{ and }\sup_{\tau\in\Delta_j}|I^\varepsilon(\tau)|\le\nu(M).\tag{3.16}$$

We call these intervals *typical*. Clearly,

$$\text{if }\Delta_j^\omega\text{ is typical, then }\omega\in\mathcal{M}_j.\tag{3.17}$$

For a typical interval Δ_j we will estimate the following quantity:

$$\begin{aligned}&\Big|\int_{\tau_j}^{\tau_{j+1}}\big(P^I(I^\varepsilon(s),\varphi^\varepsilon(s))-\langle P^I\rangle(I^\varepsilon(s))\big)\,ds\Big|\\&\le\Big|\int_{\tau_j}^{\tau_{j+1}}\big(P^I(I^\varepsilon(s),\varphi^\varepsilon(s))-P^I(I^\varepsilon(\tau_j),\varphi^\varepsilon(\tau_j)+\tfrac{s-\tau_j}{\varepsilon}\theta(I^\varepsilon(\tau_j)))\big)\,ds\Big|\\&+\Big|\int_{\tau_j}^{\tau_{j+1}}\big(P^I(I^\varepsilon(\tau_j),\varphi^\varepsilon(\tau_j)+\tfrac{s-\tau_j}{\varepsilon}\theta(I^\varepsilon(\tau_j)))-\langle P^I\rangle(I^\varepsilon(\tau_j))\big)\,ds\Big|\\&+\Big|\int_{\tau_j}^{\tau_{j+1}}\big(\langle P^I\rangle(I^\varepsilon(\tau_j))-\langle P^I\rangle(I^\varepsilon(s))\big)\,ds\Big|=J_1+J_2+J_3.\end{aligned}$$

Due to (3.14), (3.15) and (3.17),

$$J_1\le CN^2\varepsilon^{\frac54}+CN^3\varepsilon^{\frac54}=CN^3\varepsilon^{\frac54},\qquad J_3\le CN^2\varepsilon^{\frac54}.$$

Considering (3.5) and Lemma 3.3 we find that

$$\begin{aligned}J_2&=\Big|\int_{\tau_j}^{\tau_{j+1}}\big(P^I(I^\varepsilon(\tau_j),\varphi^\varepsilon(\tau_j)+\tfrac{s-\tau_j}{\varepsilon}\theta(I^\varepsilon(\tau_j)))-\langle P^I\rangle(I^\varepsilon(\tau_j))\big)\,ds\Big|\\&=\Big|\varepsilon N\frac1N\int_0^N\big(P^I(I^\varepsilon(\tau_j),\varphi^\varepsilon(\tau_j)+\xi\theta(I^\varepsilon(\tau_j)))-\langle P^I\rangle(I^\varepsilon(\tau_j))\big)\,d\xi\Big|\le\delta\varepsilon N.\end{aligned}$$

Thus for a typical interval Δ_j we have:

$$\Big|\int_{\tau_j}^{\tau_{j+1}}\big(P^I(I^\varepsilon(s),\varphi^\varepsilon(s))-\langle P^I\rangle(I^\varepsilon(s))\big)\,ds\Big|\le CN^3\varepsilon^{\frac54}+\delta\varepsilon N.\tag{3.18}$$

Step 5: We introduce the third good event

$$\mathcal{E}_3=\{\omega\in\Omega:\ \sup_{0\le\tau\le T}|I^\varepsilon(\tau)|\le\nu(M)\}$$

(corresponding to the third condition in the definition (3.16) of a typical interval). Due to (4) of Assumption 2.1,

$$\mathbf{P}(\mathcal{E}_3^c) \to 0 \quad \text{as } M \to \infty,$$

and the rate of convergence is independent of ε. Finally we set

$$\mathcal{E} = \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3.$$

Then by (3.10) and (3.13)

$$\mathbf{P}(\mathcal{E}^c) \le \mathbf{P}(\mathcal{E}_1^c) + \mathbf{P}(\mathcal{E}_2^c) + \mathbf{P}(\mathcal{E}_3^c) \le (\alpha_N^\delta)^{1/2} + C_3 T N^{-1/2} + \mathbf{P}(\mathcal{E}_3^c) =: \beta_{N,M}^\delta,$$

where $\beta_{N,M}^\delta \to 0$ as $N, M \to \infty$, for each $\delta > 0$. Note that although the sets $\mathcal{E}$ and $\mathcal{E}_j$ depend on ε, δ and N, the upper bound $\beta_{N,M}^\delta$ for $\mathbf{P}(\mathcal{E}^c)$ is independent of ε.

Denote

$$\mathcal{I}^\varepsilon := \sup_{0\le\tau\le T} \Big| \int_0^\tau \big(P^I(I^\varepsilon(s),\varphi^\varepsilon(s)) - \langle P^I\rangle(I^\varepsilon(s))\big)\,ds\Big|$$

and

$$\mathcal{I}_j^\varepsilon := \Big| \int_{\tau_j}^{\tau_{j+1}} \big(P^I(I^\varepsilon(s),\varphi^\varepsilon(s)) - \langle P^I\rangle(I^\varepsilon(s))\big)\,ds\Big|, \qquad 0 \le j \le j_N.$$

Apart from $\mathcal{I}^\varepsilon$ let us consider $\mathcal{I}_N^\varepsilon$, defined as follows:

$$\mathcal{I}_N^\varepsilon := \sup_{\tau=\tau_1,\dots,\tau_{j_N}} \Big| \int_{\tau_0}^\tau \big(P^I(I^\varepsilon(s),\varphi^\varepsilon(s)) - \langle P^I\rangle(I^\varepsilon(s))\big)\,ds\Big|.$$

Since $\mathcal{E} \subset \mathcal{E}_3$, then for $\omega \in \mathcal{E}$ the norms of vector fields $P^I(I^\varepsilon(s),\varphi^\varepsilon(s))$ and $\langle P^I\rangle(I^\varepsilon(s))$ are bounded by M (we recall (3.1)). Therefore,

$$|\mathcal{I}^\varepsilon - \mathcal{I}_N^\varepsilon| \le CN^2\varepsilon \quad \text{for} \quad \omega \in \mathcal{E}. \tag{3.19}$$

Conditions (3) and (4) of Assumption 2.1 imply that

$$\mathbf{E}(\mathcal{I}_N^\varepsilon)^2,\ \mathbf{E}(\mathcal{I}^\varepsilon)^2 \le CT^2 \quad \text{and} \quad \mathbf{E}(\mathcal{I}_j^\varepsilon)^2 \le C\varepsilon^2 N^3. \tag{3.20}$$

Due to (3.20) and (3.19),

$$\Upsilon^\varepsilon = \mathbf{E}\mathcal{I}^\varepsilon = \mathbf{E}\big(\mathbf{1}_{\mathcal{E}}\mathcal{I}^\varepsilon\big) + \mathbf{E}\big(\mathbf{1}_{\mathcal{E}^c}\mathcal{I}^\varepsilon\big) \le \mathbf{E}\big(\mathbf{1}_{\mathcal{E}}\mathcal{I}_N^\varepsilon\big) + CN^2\varepsilon + (CT^2)^{\frac12}(\beta_{N,M}^\delta)^{1/2}. \tag{3.21}$$

Step 6: Let us estimate $\mathbf{E}\big(\mathbf{1}_{\mathcal{E}}\mathcal{I}_N^\varepsilon\big)$. Denoting

$$\mathcal{J} = \mathcal{J}^\varepsilon = \{j : \Delta_j \ \text{ is typical}\}$$

we have that

$$\mathbf{E}\big(\mathbf{1}_{\mathcal{E}}\mathcal{I}_N^\varepsilon\big) \le \mathbf{E}\Big(\sum_{j\in\mathcal{J}} \mathcal{I}_j^\varepsilon\Big) + \mathbf{E}\Big(\mathbf{1}_{\mathcal{E}} \sum_{j\notin\mathcal{J}} \mathcal{I}_j^\varepsilon\Big) =: S_1 + S_2. \tag{3.22}$$

From the definitions of sets $\mathcal{E}_1, \mathcal{E}_2$ we see that for any $\omega \in \mathcal{E}$,

$$\#\mathcal{J}^c \le (\alpha_N^\delta)^{1/2}(\varepsilon N)^{-1} + \varepsilon^{-7/8}.$$

Since for each $\omega \in \mathcal{E}_3$ all $\mathcal{I}_j^\varepsilon$ are trivially bounded by $C\varepsilon NM$, then by the above estimate

$$S_2 \le CM\big((\alpha_N^\delta)^{1/2} + N\varepsilon^{1/8}\big).$$

Due to (3.18), for each ω

$$\sum_{j\in\mathcal{J}} \mathcal{I}_j^\varepsilon \le j_N\big(CN^3\varepsilon^{5/4} + \delta\varepsilon N\big) \le C_1\big(N^2\varepsilon^{1/4} + \delta\big).$$

So

$$S_1 \le C_1\big(N^2\varepsilon^{1/4} + \delta\big).$$

By (3.21), (3.22) and the estimates for S_1 and S_2,

$$\Upsilon^\varepsilon \le C'\beta_{N,M}^\delta + C_1\big(N^2\varepsilon^{1/4} + \delta\big) + CM\big((\alpha_N^\delta)^{1/2} + N\varepsilon^{1/8}\big) + CN^2\varepsilon.$$

Now for any $\delta > 0$ we choose $M = M(\delta)$ so big that $C'\beta^{\delta}_{N,M'} \leqslant \delta$ for all $M' \geqslant M$. Then for this M we choose sufficiently large N so that $CM(\alpha^{\delta}_{N})^{1/2} \le \delta$. For these M and N we have

$$\Upsilon^{\varepsilon} \le \delta + C_1 N^2 \varepsilon^{1/4} + C_1\delta + CN\varepsilon^{1/8} + CN^2\varepsilon.$$

Taking ε small enough we achieve that $\Upsilon^{\varepsilon} \le (C_1 + 2)\delta$. Since δ is an arbitrary positive number, then (3.4) follows. □

With exactly the same proof we also have

Lemma 3.5. *The following convergence holds as $\varepsilon \to 0$:*

$$\mathbf{E}\Big\{ \sup_{0\le\tau\le T} \Big| \int_0^{\tau} \big(\Psi^I(I^{\varepsilon}(s), \varphi^{\varepsilon}(s))(\Psi^I(I^{\varepsilon}(s), \varphi^{\varepsilon}(s)))^t - \langle a^I\rangle(I^{\varepsilon}(s))\big)\, ds\Big|\Big\} \to 0.$$

4. The averaging theorem

In the section we show that the limiting laws of the family $\{\mathcal{D}(I^{\varepsilon}(\cdot))\}$, as $\varepsilon \to 0$, are solutions of the martingale problem for the averaged equation (3.3), thus are weak solutions of the later. We begin with the corresponding definition. Let us introduce a natural filtered measurable space for the problem we consider

$$(\tilde{\Omega}, \mathcal{B}_T, \{\mathcal{B}_{\tau}, 0 \leqslant \tau \leqslant T\}), \tag{4.1}$$

where $\tilde{\Omega}$ is the Banach space $C([0,T];\mathbb{R}^d) = \{a(\tau), \tau \in [0,T]\}$, $\mathcal{B}_T$ is its Borel σ-algebra and a σ-algebra $\mathcal{B}_{\tau} \subset \mathcal{B}_T$ is generated by the restriction mapping $\tilde{\Omega} \ni a(\cdot) \mapsto a(\cdot)\mid_{[0,\tau]}$. Consider the process

$$N^I(\tau; a) \coloneqq a(\tau) - \int_0^{\tau} \langle P^I\rangle(a(s))ds,\ a \in \tilde{\Omega}, \tau \in [0,T],$$

and note that for any $0 \leqslant \tau \leqslant T$, $N^I(\tau;\cdot)$ is a $\mathcal{B}_{\tau}$-measurable continuous functional on $\tilde{\Omega}$.

Definition 4.1. (see [28, 15]). A measure Q on the space (4.1) is called a solution of the martingale problem for equation (3.3), if $a(0) = I_0$ Q-a.s. and

1) the process $\{N^I(\tau;a) \in \mathbb{R}^d, \tau \in [0,T]\}$ is a vector-martingale on the filtered space (4.1) with respect to the measure Q;

2) for any $k, j = 1\dots, d$ the process $N^I_k(\tau;a)N^I_j(\tau;a) - \int_0^{\tau}\langle a^I\rangle_{kj}(a(s))ds$ is a martingale on the space (4.1) with respect to the measure Q.

For each $\varepsilon > 0$ we define a probability measure Q^{ε} on $(\tilde{\Omega}, \mathcal{B}_T)$ as the law of $\{I^{\varepsilon}(\cdot)\}$ and denote by $\mathbf{E}^{Q^{\varepsilon}}$ the corresponding expectation. According to Lemma 3.1 the family $\{Q^{\varepsilon}\}$ is tight in $\mathcal{P}(C([0,T],\mathbb{R}^d)$. Take a sequence $\varepsilon_j \to 0$ such that

$$Q^{\varepsilon_j} \rightharpoonup Q^0,\ \text{ as } \varepsilon_j \to 0 \ \text{ in } \mathcal{P}(C([0,T],\mathbb{R}^d)). \tag{4.2}$$

Lemma 4.2. *The probability measure Q^0 above is a solution of the martingale problem for the averaged equation (3.3).*

Proof. For any $s \in [0,T]$, let $\Phi(\cdot)$ be a bounded continuous functional defined on $C([0,s];\mathbb{R}^d)$. Then, for any $\tau \in [0,T]$ such that $0 \le s \le \tau \le T$, we have

$$\begin{aligned}
&\mathbf{E}^{Q^0}\Big\{\big(N^I(\tau;a) - N^I(s;a)\big)\Phi(a([0,s]))\Big\}\\
&= \lim_{\varepsilon_j\to 0}\mathbf{E}^{Q^{\varepsilon_j}}\Big\{\big(N^I(\tau;a) - N^I(s;a)\big)\Phi(a([0,s]))\Big\}\\
&= \lim_{\varepsilon_j\to 0}\mathbf{E}\Big\{\Big(I^{\varepsilon_j}(\tau) - I^{\varepsilon_j}(s) - \int_s^\tau \langle P^I\rangle(I^{\varepsilon_j}(u))\,du\Big)\Phi(I^{\varepsilon_j}([0,s]))\Big\}\\
&= \lim_{\varepsilon_j\to 0}\mathbf{E}\Big\{\Big(\int_s^\tau \big(P^I(I^{\varepsilon_j}(u),\varphi^{\varepsilon_j}(u)) - \langle P^I\rangle(I^{\varepsilon_j}(u))\big)\,du\Big)\Phi(I^{\varepsilon_j}([0,s]))\Big\}\\
&+ \lim_{\varepsilon_j\to 0}\mathbf{E}\Big\{\Big(I^{\varepsilon_j}(\tau) - I^{\varepsilon_j}(s) - \int_s^\tau P^I(I^{\varepsilon_j}(u),\varphi^{\varepsilon_j}(u))\,du\Big)\Phi(I^{\varepsilon_j}([0,s]))\Big\},
\end{aligned}$$

where to get the first equality we have used the fact that the r.v.'s $\|I^\varepsilon\|^q_{C([0,T])}$ are uniformly integrable, which is guaranteed by (3) and (4) of Assumption 2.1 (cf. [12, Lemma 4.4]). The first limit on the r.h.s of the last equality vanishes due to Lemma 3.2 and the second one vanishes because $I^\varepsilon(\tau) - \int_0^\tau P^I(I^\varepsilon(u),\varphi^\varepsilon(u))\,du$ is a martingale. Therefore,

$$\mathbf{E}^{Q^0}\Big\{\Big(N^I(\tau;a) - N^I(s;a)\Big)\Phi(a([0,s]))\Big\} = 0,$$

for each Φ as above. So the process $N^I(\tau;a)$, $\tau \in [0,T]$, is a martingale with respect to the measure Q^0 and filtration $\{\mathcal{B}_\tau\}$.

Arguing in the same way and evoking Lemma 3.5 we conclude that the processes $N^I_k(\tau;a)N^I_j(\tau;a) - \int_0^\tau \langle a^I\rangle_{kj}(a(s))ds$, $\tau \in [0,T]$, $k,j = 1,\dots,d$, also are $(Q^0,\mathcal{B}_\tau)$-martingales. Hence, the assertion of the lemma is proved. □

As we have discussed before, since the drift term $\langle P^I\rangle$ and the dispersion matrix $\langle\!\langle \Psi^I\rangle\!\rangle$ in (3.3) are locally Lipschitz with respect I, then by the Yamada-Watanabe theorem (see [15, Section 5.3.D]) the just obtained solution of the martingale problem for (3.3) is unique. So, in particular, the limit Q^0 in (4.2) does not depend on the choice of ε_j, and the whole family Q^ε converges as $\varepsilon \to 0$ to the measure Q^0. Thus we have obtained the following result.

Theorem 4.3. *Under Assumption 2.1, for any $(I_0,\varphi_0) \in \mathbb{R}^d \times \mathbb{T}^n$ we have*

$$\mathcal{D}(I^\varepsilon(\cdot)) \rightharpoonup Q^0 \quad \text{as } \varepsilon \to 0 \quad \text{in } \mathcal{P}(C([0,T],\mathbb{R}^d)),$$

where Q^0 is the law of a weak solution $I^0(\tau)$ for problem (3.3). Moreover, for q_0 and $C(\cdot)$ as in (2.3) we have $\mathbf{E}\sup_{\tau\in[0,T]}|I^0(\tau)|^{2q_0} \leqslant C(|I_0|,T)$.

The last assertion follows directly from the Skorokhod representation theorem (see [2, Section 6] and Fatou's Lemma, cf. [12, Remark 4.8].

Remark 4.4. 1) It is straightforward to see that the statement of Theorem 4.3 remains true with the same proof if in (2.2) the initial data $(I_{0,\varepsilon},\varphi_{0,\varepsilon})$ depends on ε and converges to a limit (I_0,φ_0) as $\varepsilon \to 0$.

2) The result in Theorem 4.3 admits an immediate generalization to the case when the initial data (I_0,φ_0) is a random variable. Cf. Amplification 8.8.

3) A local in I version of Theorem 4.3 was earlier proved in [9] by another method.

5. Stationary solutions

The goal of this section is to characterize the asymptotic behaviour of stationary solutions of equations (2.1) and to find out their relation with a stationary solution of averaged equation (3.3). We recall that a solution $(I^\varepsilon(\tau),\varphi^\varepsilon(\tau))$, $\tau \geqslant 0$, of equation (2.1) is called stationary if there exists $\nu^\varepsilon \in \mathcal{P}(\mathbb{R}^d \times \mathbb{T}^n)$ such that $\mathcal{D}\big((I^\varepsilon(\tau), \varphi^\varepsilon(\tau))\big) = \nu^\varepsilon$ for all $\tau \geqslant 0$. Then the measure ν^ε is called a stationary measure for equation (2.1). A stationary solution and stationary measure of the averaged equation (3.3) are defined in the same way.

Throughout this section a strengthened version of Assumption 2.1 is imposed on the system (2.1). Namely, we assume that

Assumption 5.1. i) Items (1)-(3) of Assumption 2.1 hold true.

ii) For each $\varepsilon \in (0,1]$ and any $(I_0, \varphi_0) \in \mathbb{R}^d \times \mathbb{T}^n$ problem (2.1)-(2.2) has a unique strong solution $\big(I^\varepsilon, \varphi^\varepsilon\big)(\tau; I_0, \varphi_0)$, $\tau \in [0, +\infty)$, satisfying

$$\mathbf{E} \sup_{T' \leqslant \tau \leqslant T'+1} |I^\varepsilon(\tau; I_0, \varphi_0)|^{2q_0} \leqslant C_{q_0}(|I_0|), \tag{5.1}$$

for each $T' \geqslant 0$ and some number $q_0 > (q \vee 2)$.

iii) Equation (2.1) is mixing. So it has a stationary weak solution $\big(I^\varepsilon_{\rm st}(\tau), \varphi^\varepsilon_{\rm st}(\tau)\big)$ such that $\mathcal{D}\big(I^\varepsilon_{\rm st}(\tau), \varphi^\varepsilon_{\rm st}(\tau)\big) \equiv \nu^\varepsilon \in \mathcal{P}(\mathbb{R}^d \times \mathbb{T}^n)$, and

$$\mathcal{D}\big(I^\varepsilon(\tau; I_0, \varphi_0), \varphi^\varepsilon(\tau; I_0, \varphi_0)\big) \rightharpoonup \nu^\varepsilon \text{ in } \mathcal{P}(\mathbb{R}^d \times \mathbb{T}^n) \quad \text{as } \tau \to +\infty, \tag{5.2}$$

for each (I_0, φ_0).

Under Assumption 5.1 equation (2.1) defines in $\mathbb{R}^d \times \mathbb{T}^n$ a Markov process with the transition probability $\Sigma^\varepsilon_\tau(I, \varphi) \in \mathcal{P}(\mathbb{R}^d \times \mathbb{T}^n)$, $\tau \geq 0$, $(I, \varphi) \in \mathbb{R}^d \times \mathbb{T}^n$, where $\Sigma^\varepsilon_\tau(I, \varphi) = \mathcal{D}\big(I^\varepsilon(\tau; I, \varphi), \varphi^\varepsilon(\tau; I, \varphi)\big)$; e.g. see [15, Section 5.4.C].

As a consequence of (5.1) and (5.2) by a straightforward argument (e.g. see in [12] Lemma 5.3 and its proof), we obtain that the stationary solution $(I^\varepsilon_{\rm st}, \varphi^\varepsilon_{\rm st})$ satisfies the following estimate

$$\mathbf{E} \sup_{T' \leqslant \tau \leqslant T'+1} |I^\varepsilon_{\rm st}(\tau)|^{2q_0} \leqslant C_{q_0}(0). \tag{5.3}$$

Using (5.3) we derive from the first equation in (2.1) that for any $N \in \mathbb{N}$, the collection of measures $\{\mathcal{D}(I^\varepsilon_{\rm st}|_{[0,N]}), 0 < \varepsilon \leqslant 1\}$ is tight. Choosing for each N a sequence $\varepsilon^{(N)}_l \to 0$ such that

$$\mathcal{D}(I^{\varepsilon^{(N)}_l}_{\rm st}|_{[0,N]}) \rightharpoonup Q^0 \text{ in } \mathcal{P}\big(C([0,N], \mathbb{R}^d)\big)$$

and applying the diagonal procedure we conclude that for a subsequence $\{\varepsilon_l\}$ the relation $\mathcal{D}(I^{\varepsilon_l}_{\rm st}) \rightharpoonup Q^0$ holds in $\mathcal{P}(X)$, where X is the complete separable metric space $X = C([0,\infty), \mathbb{R}^d)$ with the distance

$$\operatorname{dist}(a_1, a_2) = \sum_{N=1}^{\infty} 2^{-N} \frac{\|a_1 - a_2\|_{C([0,N],\mathbb{R}^d)}}{1 + \|a_1 - a_2\|_{C([0,N],\mathbb{R}^d)}}, \qquad a_1, a_2 \in X\,.$$

Denote $\mu^\varepsilon_I(\tau) = \mathcal{D}(I^\varepsilon_{st}(\tau)) = I \circ \nu^\varepsilon$. Then $\mu^{\varepsilon_l}_I(0) \rightharpoonup \mu^0 := Q^0|_{\tau=0}$. Let $I^0(\tau)$ be a solution of equation (3.3) with an initial condition I_0, distributed as μ^0. Then, by Remark 4.4,

$$\mathcal{D}(I^{\varepsilon_l}_{\rm st}(\cdot)) \rightharpoonup \mathcal{D}(I^0(\cdot)) \quad \text{in } X, \tag{5.4}$$

and for any $\tau \geq 0$ we have $Q^0(\tau) = \mathcal{D}(I^0(\tau)) = \lim_{\varepsilon_l \to 0} \mathcal{D}I(^{\varepsilon_l}_{\rm st}(\tau)) = \lim_{\varepsilon_l \to 0} \mathcal{D}(I^{\varepsilon_l}_{\rm st}(0)) = \mu^0$. We obtained the following statement:

Proposition 5.2. *The process I^0 is a stationary weak solution of the averaged equation (3.3) and $\mathcal{D}(I^0(\tau)) \equiv \mu^0$, $\tau \in [0, \infty)$. In particular, any limiting point of the collection of measures $\{\mu^\varepsilon_I := \mathcal{D}(I^\varepsilon_{\rm st}(\tau))\}$ as $\varepsilon \to 0$ is a stationary measure of the averaged equation (3.3). If the latter is mixing, then its stationary measure is unique, and so convergence (5.4) holds as $\varepsilon \to 0$.*

5.1. **Uniform convergence in the averaging theorem.** To describe quantitively the weak convergence of measures in Theorem 4.3 we introduce the dual-Lipschitz distance.

Definition 5.3. Let M be a complete and separable metric space. For any two measures $\mu_1, \mu_2 \in \mathcal{P}(M)$ we define the dual-Lipschitz distance between them as

$$\|\mu_1 - \mu_2\|^*_{L,M} := \sup_{f \in C_b(M), |f|_L \leqslant 1} \Big| \langle f, \mu_1 \rangle - \langle f, \mu_2 \rangle \Big| = \sup_{f \in C_b(M), |f|_L \leqslant 1} \Big(\langle f, \mu_1 \rangle - \langle f, \mu_2 \rangle \Big) \leq 2,$$

where $|f|_L = |f|_{L,M} = \mathrm{Lip}\, f + \|f\|_{C(M)}$.

In this definition and below we denote

$$\langle f, \mu \rangle := \int_M f(m) \mu(dm). \tag{5.5}$$

The dual-Lipschitz distance converts $\mathcal{P}(M)$ to a complete metric space and induces on it a topology, equivalent to the weak convergence of measures, e.g. see [5, Section 1.3].

If equation (3.3) is mixing with a quantitive property as in the following assumption, then the convergence in Theorem 4.3 is uniform in time with respect to the dual-Lipschitz distance.

Assumption 5.4. 1) i) and (ii) of Assumption 5.1 hold true.

2) The averaged equation (3.3) is mixing with a unique stationary measure μ^0. Moreover, for each $M > 0$ and any $I_0, I_1 \in \bar{B}_M(\mathbb{R}^d)$ the laws of solutions for (3.3) with these initial data satisfy

$$\|\mathcal{D}(I^0(\tau; I_0)) - \mathcal{D}(I^0(\tau; I_1))\|^*_{L,\mathbb{R}^d} \leqslant \mathfrak{f}_M(\tau),$$

where $\mathfrak{f}_M$ is a positive continuous function of (M, τ) which goes to zero when $\tau \to \infty$ and is non-decreasing in M.

Theorem 5.5. *Under Assumption 5.4, in the setting of Theorem 4.3 for any initial data* $(I_0, \varphi_0) \in \mathbb{R}^d \times \mathbb{T}^n$ *we have*

$$\lim_{\varepsilon \to 0} \sup_{\tau \geqslant 0} \|\mathcal{D}(I^\varepsilon(\tau; I_0, \varphi_0)) - \mathcal{D}(I^0(\tau; I_0))\|^*_{L,\mathbb{R}^d} = 0.$$

Concerning the theorem's proof see Subsection 10.1. Assumptions 5.1 and 5.4 allow for an instrumental sufficient condition, cf. below Proposition 10.9.

6. Perturbations of integrable equations in $\mathbb{R}^{2n}$

In this section we study diffusive perturbations of integrable equations in $\mathbb{R}^{2n}$ in the framework of previous sections. By bold italic letters we denote various vectors in $\mathbb{R}^2$ (regarded as column-vectors).

Let us consider a perturbed integrable equation (1.4) for a vector $v = (\mathbf{v}_k,\ k = 1, \dots, n) \in \mathbb{R}^{2n}$ and write it in the slow time $\tau = \varepsilon t$:

$$\begin{cases} d\mathbf{v}_k = \varepsilon^{-1} W_k(I) \mathbf{v}_k^\perp d\tau + \mathbf{P}_k(v) d\tau + \sum_{j=1}^{n_1} B_{kj}(v) d\boldsymbol{\beta}_j(\tau), \quad k = 1, \dots, n, \\ v(0) = v_0 \in \mathbb{R}^{2n}. \end{cases} \tag{6.1}$$

Here $\mathbf{v}_k = \begin{pmatrix} v_k \\ v_{-k} \end{pmatrix} \in \mathbb{R}^2$, $\mathbf{v}_k^\perp = \begin{pmatrix} -v_{-k} \\ v_k \end{pmatrix}$, $B_{kj}(v)$ are 2×2 matrix functions, $I = (I_1, \dots, I_n)$ with $I_k = \frac{1}{2}\|\mathbf{v}_k\|^2$ and $\boldsymbol{\beta}_j(\tau) = \begin{pmatrix} \beta_j(\tau) \\ \beta_{-j}(\tau) \end{pmatrix}$ are independent standard Brownian motions in $\mathbb{R}^2$, defined on a filtered probability space

$$\big(\Omega, \mathcal{F}, \{\mathcal{F}_\tau\}, \mathbf{P}\big). \tag{6.2}$$

The unperturbed part (1.3) of equation (6.1) is integrable, and the functions $I_k(v), k = 1,\dots,n$, are its integrals of motion. If $W(I) = \nabla h(I)$ for some C^1-function h, then equation (1.3) is Hamiltonian with the Hamiltonian function $h(I(v))$, and then it is is called integrable in the sense of Birkhoff.

As in previous sections, the initial data v_0 can be deterministic or random, and a solution of problem (6.1) will be denoted $v^\varepsilon(\tau;v_0) = (\mathbf{v}_k^\varepsilon(\tau;v_0), k = 1,\dots n)$, or simply $v(\tau)$. We will focus on the deterministic case, always assuming that equation (6.1) satisfies the following assumption.

Assumption 6.1. (1) The Lebesgue measure of $I \in \mathbb{R}^n_+$ for which components of the vector $W(I) =: (W_k(I), k = 1,\dots,n)$ are rationally dependent is equal to zero. That is, $\mathcal{L}\big(\cup_{k\in\mathbb{Z}^n\setminus\{0\}}\{I \in \mathbb{R}^n_+ : W(I)\cdot k = 0\}\big) = 0$.

(2) The $2n\times 2n$ diffusion matrix $S(v) = B(v)B(v)^t$, where $B(v) = (B_{kj}(v))$, satisfies the uniform ellipticity condition. That is, there exists $\lambda > 0$ such that

$$\lambda\|\xi\|^2 \leqslant S(v)\xi\cdot\xi \leqslant \lambda^{-1}\|\xi\|^2, \quad \forall v,\xi \in \mathbb{R}^{2n}. \tag{6.3}$$

(3) There exists $q > 0$ such that $W(I) \in \mathrm{Lip}_q(\mathbb{R}^n,\mathbb{R}^n)$, $P(v) := (\mathbf{P}_k(v), k = 1,\dots,n) \in \mathrm{Lip}_q(\mathbb{R}^{2n},\mathbb{R}^{2n})$ and $B(v) \in \mathrm{Lip}_q(\mathbb{R}^{2n}, M_{2n\times 2n_1})$ (we recall (1.13)).

(4) There exists $T > 0$ such that for every $v_0 \in \mathbb{R}^{2n}$ equation (6.1) has a unique strong solution $v^\varepsilon(\tau;v_0)$, $\tau \in [0,T]$, equal v_0 at $\tau = 0$. Moreover, there exists $q_0 > (q\vee 4)$ such that

$$\mathbf{E}\sup_{\tau\in[0,T]}\|v^\varepsilon(\tau;v_0)\|^{2q_0} \leqslant C_{q_0}(\|v_0\|,T), \quad \forall\varepsilon\in(0,1], \tag{6.4}$$

where $C_{q_0}(\cdot)$ is a non-negative continuous function on $\mathbb{R}^2_+$, non-decreasing in both arguments.

Remark 6.2. 1) In the assumption above, (1) is Anosov's condition (see Remark 2.2), and it holds, in particular, if $W(I)$ is a constant vector with rationally independent coefficients. Equations (6.1) with constant frequency vectors W and without assuming that its components are rationally independent are examined in [12], and for the case $W =$const the results, given below in Sections 6-10, are special cases of more general theorems in that work. But equations (6.1) with non-constant frequency vectors $W(I)$ are significantly more complicated then those with $W =$const.

2) Item (4) holds if assumptions (2), (3) are valid and if the coefficients of equation (6.1) are globally Lipschitz (cf. Proposition 2.3), or if the vector field P is coercive, see below Proposition 10.9.

Example 6.3. In statistical physics they often examine stochastic perturbations of chains of nonlinear oscillators

$$\ddot{q}_k = -Q(q_k), \qquad k = 1,\dots,n, \tag{6.5}$$

where Q is a polynomial $Q^0(q) = \alpha q + \beta q^3$, $\alpha,\beta > 0$, or more generally, is a smooth function of the form

$$Q(q) = Q^0(q) + O(q^4). \tag{6.6}$$

System (6.5) may be re-written in the Hamiltonian form as

$$\dot{q}_k = (\partial/\partial p_k)H, \quad \dot{p}_k = -(\partial/\partial q_k)H, \qquad k = 1,\dots,n, \tag{6.7}$$

where $H = \sum\big(\frac12 p_k^2 + \int_0^{q_k} Q(l)dl\big)$. E.g. see [6] and [3, Section 4]. Under a suitable diagonal symplectomorphism

$$\mathbb{R}^{2n}\to\mathbb{R}^{2n}, \quad (q,p)\mapsto v = (\mathbf{v}_1,\dots,\mathbf{v}_n), \;\; \mathbf{v}_j = \mathbf{F}(q_j,p_j),$$

system (6.7) may be re-written in the form (1.3), where $W_k = \omega(I_k)$, $k = 1,\dots,n$, and ω is a smooth function. See below Appendix B concerning canonical transformations $\mathbf{F} : \mathbb{R}^2 \to \mathbb{R}^2$ with the required properties. In the just cited papers the vector field $P = (\mathbf{P}_1,\dots,\mathbf{P}_n)$ in (6.1) is of the type "next neighbour interaction", i.e. $\mathbf{P}_k = \mathbf{P}_k(\mathbf{v}_{k-1},\mathbf{v}_k,\mathbf{v}_{k+1})$. Our results apply to corresponding equations (6.1) with non-degenerate dispersion matrices B if assumptions (1), (3) and (4) hold. The first two are easy to verify, while for a sufficient condition for (4) see Proposition 10.9. Below in Section 10.2 we discuss a class of damped/driven Hamiltonian systems which includes the equations we are now discussing if there the vector field P is Hamiltonian with added damping and the stochastic term is a diagonal random force.

Usually the systems in works of physicists correspond to equations (6.1) with degenerate diffusion such that only the 2×2-matrices B_{11} and B_{nn} are non-zero. To treat them the methods of our work should be developed further. Still we mention that stochastic perturbations of linear equations (6.5) with $Q(q) = \alpha q$, $\alpha > 0$, and with degenerate (or non-degenerate) diffusion may be examined using the results of [12]. □

Let us consider the action-angles mapping $\mathbb{R}^{2n} \to \mathbb{R}^n_+ \times \mathbb{T}^n$, $v \mapsto (I,\varphi)(v)$, where

$$\begin{aligned} I(v) &= (I_k(v), k = 1,\dots,n), \qquad I_k(v) = \frac{1}{2}\|\mathbf{v}_k\|^2,\\ \varphi(v) &= (\varphi_k(v), k = 1,\dots,n), \qquad \varphi_k(v) = \mathrm{Arg}(\mathbf{v}_k) = \arctan\frac{v_k}{v_{-k}} \end{aligned} \tag{6.8}$$

and $\varphi_k := 0$ if $\mathbf{v}_k = 0$. Then

$$\mathbf{v}_k = \sqrt{2I_k}\,(\cos\varphi_k, \sin\varphi_k), \qquad k = 1,\dots,n. \tag{6.9}$$

By Itô's formula, if $v(\tau)$ is a solution of (6.1), then the equations for the actions $I_k(v)$ read

$$dI_k = \mathbf{v}_k^t\mathbf{P}_k(v)d\tau + \frac{1}{2}\sum_{j=1}^{n_1}\|B_{kj}(v)\|^2_{HS}d\tau + \sum_{j=1}^{n_1}\mathbf{v}_k^tB_{kj}(v)d\boldsymbol{\beta}_j(\tau), \quad k = 1,\dots,n, \tag{6.10}$$

where $\|\cdot\|_{HS}$ is the Hilbert-Schmidt norm. Equations for $\varphi_k(v)$, $k = 1,\dots,n$, hold if all $\mathbf{v}_k$ are non-zero and read

$$d\varphi_k(\tau) = \varepsilon^{-1}W_k(I)d\tau + \Phi^1_k(v)d\tau + \sum_{j=1}^{n_1}\Phi^2_{kj}(v)d\boldsymbol{\beta}_j(\tau), \quad k = 1,\dots,n, \tag{6.11}$$

where

$$\Phi^1_k(v) = \Big(\nabla_{\mathbf{v}_k}\arctan\big(\frac{v_k}{v_{-k}}\big)\Big)\cdot\mathbf{P}_k(v) + \frac{1}{2}\sum_{j=1}^{n_1}\mathrm{Trace}\Big(B_{kj}(v)\Big(\nabla^2_{\mathbf{v}_k}\arctan\big(\frac{v_k}{v_{-k}}\big)\Big)B_{kj}(v)^t\Big),$$

and $\Phi^2_{kj}(v) = \Big(\nabla_{\mathbf{v}_k}\arctan\big(\frac{v_k}{v_{-k}}\big)\Big)\cdot B_{kj}(v)$.

Remark 6.4. Note that in view of (6.9), for I_k near zero the r.h.s of (6.10) is a Hölder-$\frac{1}{2}$ function of I_k and the r.h.s. of (6.11) has a strong singularity when I_k vanishes. Moreover, the dispersion part of (6.10) vanishes with I_k. Hence, system (6.10)+(6.11) is singular and degenerated at the set $\cup_{k=1}^n\{(I,\varphi) : I_k = 0\}$.

As in previous sections, we introduce the averaged equations for $I_k(\tau)$, $k = 1,\dots,n$, as

$$\begin{aligned} dI_k(\tau) &= F_k(I)d\tau + \sum_{j=1}^n K_{kj}(I)d\beta_j(\tau),\\ F_k(I) &= \langle\mathbf{v}_k^t\mathbf{P}_k\rangle(I) + \tfrac{1}{2}\langle\sum_j\|B_{kj}\|^2_{HS}\rangle(I), \end{aligned} \tag{6.12}$$

with the initial condition

$$I(0) = I_0 = I(v_0). \tag{6.13}$$

The brackets $\langle\cdot\rangle$ signify the averaging in φ, see (1.14), and the dispersion matrix $K(I) = \big(K_{kj}(I)\big)_{1\leqslant k,j\leqslant n}$ is chosen to be the principal symmetric square root of the averaged diffusion matrix $S(I) = (S_{km}(I))_{1\leqslant k,m\leqslant n}$ of equation (6.10),

$$S_{km} := \Big\langle \sum_{j=1}^{n_1} \mathbf{v}_k^t B_{kj} B_{mj}^t \mathbf{v}_m \Big\rangle(I),\ k,m = 1,\dots,n. \tag{6.14}$$

So $K = K^t \geq 0$ and

$$\sum_{j=1}^{n} K_{kj}(I) K_{mj}(I) = S_{km}(I),\ k,m = 1,\dots,n.$$

Remark 6.5. Under (3) of Assumption 6.1, the drift and dispersion terms of (6.12) are only Hölder-$\frac{1}{2}$ smooth with respect to I_k (and Lipschitz with respect to $\sqrt{I_k}$), $k = 1,\dots,n$. Moreover, the dispersion term vanishes at $I_k = 0$. If we strengthen the assumption by assuming that the vector-field P and the dispersion matrix B are C^2-smooth, then by a straightforward application of Whitney's theorem (see in [31] Theorem 1 and the last remark of the paper), the drift term in (6.12) will be C^1-smooth in I. However, this will not improve the fact that the dispersion term vanishes with I_k (and may have there a square-root singularity, see [12, Proposition 6.2] for an example). Thus the well-posedness of equation (6.12) is a delicate question.

Remark 6.6. Dispersion matrix K of equation (6.12) should not necessarily be neither symmetric nor square, and if we replace it by another (possibly non-square) matrix $\bar K$ such that $\bar K \bar K^t = S$, we would get a new equation with the same set of weak solutions. See [28, Section 5.3] and [15, Section 5.4.B]. This fact concerning equation (6.12) and other SDEs is systematically used below.

The following analogue of Theorem 4.3 holds for solutions of equation (6.1).

Theorem 6.7. *Under Assumption 6.1, for any $v_0 \in \mathbb{R}^{2n}$ the collection of laws of the processes $\{I(v^\varepsilon(\tau;v_0)),\tau\in[0,T]\}$, $0<\varepsilon\leq 1$, is tight in $\mathcal{P}(C([0,T],\mathbb{R}^n_+))$. If we take any sequence $\varepsilon_j\to 0$ such that $\mathcal{D}(I(v^{\varepsilon_j}(\cdot;v_0)) \rightharpoonup Q^0 \in \mathcal{P}(C([0,T],\mathbb{R}^n_+))$, then Q^0 is the law of a weak solution $I^0(\tau)$, $\tau\in[0,T]$, of the averaged equation (6.12), equal $I_0 = I(v_0)$ at $\tau = 0$. Moreover,*

$$\mathbf{E} \sup_{\tau\in[0,T]} |I^0(\tau)|^{q_0} \leqslant C_{q_0}(\|v_0\|,T), \tag{6.15}$$

and for $k = 1,\dots,n$,

$$\mathbf{E}\int_0^T \mathbf{1}_{\{I_k\in[0,\delta]\}}(I^0(\tau))d\tau \to 0 \ \text{as}\ \delta\to 0. \tag{6.16}$$

As we discussed in Remark 6.5, the uniqueness of a solution to the averaged equation (6.12), (6.13) is a delicate issue. Therefore in the above theorem we state the convergence only for a subsequence $\varepsilon_j \to 0$. In Section 7 we construct an *effective equation* in v-variables such that for its solution $v(\cdot)$ with the same initial data v_0 the process of actions $I(v(\cdot))$ is exactly the weak solution $I^0(\tau)$ from the theorem above. This equation has locally Lipschitz coefficients, so its solution is unique, and thus the convergence in Theorem 6.7 holds as $\varepsilon\to 0$.

Let us denote

$$I^\varepsilon(\tau) = I(v^\varepsilon(\tau;v_0)) \ \text{and}\ \varphi^\varepsilon(\tau) = \varphi(v^\varepsilon(\tau;v_0)),\ \tau\in[0,T].$$

Denote also $\mathcal{S}_\delta = \{I\in\mathbb{R}^n_+ : \min_{1\leq j\leq n} I_j \leqslant \delta\}$ and $\bar A_{N,\delta} = A_N^\delta \smallsetminus \mathcal{S}_\delta$, where A_N^δ is defined as in (3.5), but with P^I replaced by the drift term in (6.10). The following lemma is a key technical result for a proof of Theorem 6.7. It states that with high probability the process $I^\varepsilon(\tau)$, $\tau \in [0,T]$ stay away from the locus $\mathcal{S}_0 = \cup_{k=1}^n \{I \in \mathbb{R}_+^n : I_k = 0\}$, for most of the time, uniformly in ε.

Lemma 6.8. *Under Assumption 6.1, there exist a function $\kappa(\delta)$ and a function $\alpha(\delta,N) : \mathbb{R}_+ \times \mathbb{R}_+ \mapsto \mathbb{R}_+$ such that $\kappa(\delta) \to 0$ as $\delta \to 0$, $\alpha(\delta,N) \to 0$ as $N \to \infty$ for each $\delta > 0$, and*

$$\mathbf{E}\int_0^T \mathbf{1}_{\bar{A}_{N,\delta}}(I^\varepsilon(\tau))d\tau \leqslant \alpha(\delta,N), \tag{6.17}$$

$$\mathbf{E}\int_0^T \mathbf{1}_{\mathcal{S}_\delta}(I^\varepsilon(\tau))\,d\tau \le \kappa(\delta) \tag{6.18}$$

for any $\delta > 0$, uniformly in $0 < \varepsilon \le 1$.

A proof of (6.17) follows from the same argument as in the demonstration of Lemma 3.4 since the diffusion in (6.10) is non-degenerate outside $\mathcal{S}_\delta$. That of (6.18) is rather technical due to the degeneracy at the locus $\mathcal{S}_0$ and presence of the ε^{-1}-term in (6.1). The proof of this inequality is based on an argument, similar to that used in [23, 22] for an infinite-dimensional stochastic equation. It relies on introducing a family of auxiliary processes which are suitable rotations of a solution $v^\varepsilon(\tau;v_0)$. They are constructed as Itô processes such that their actions coincide with I^ε's, while presence of the rotations allows to remove terms of order ε^{-1} from the equations for these processes. Detail of the proof of (6.18) is provided in Appendix A.

With the help of Lemma 6.8 the proof of Theorem 4.3 can be adapted to demonstrate Theorem 6.7. We only give here a sketch of the corresponding argument, emphasizing the differences and leaving detail to the reader.

Sketch of the proof of Theorem 6.7: We introduce the starting time τ_0, number N and intervals Δ_j in the same way as at Step 1 in the proof of Lemma 3.2. In particular, $\Delta_j = [\tau_j, \tau_{j+1}) = [\tau_0 + j\varepsilon N, \tau_0 + (j+1)\varepsilon N)$. By the argument from the proof of Lemma 3.1 and Chebyshev's inequality, for any $\delta > 0$ and $N > 0$ it holds that

$$\mathbf{P}\Big\{ \max_{0 \leqslant \tau' < \tau'' \leqslant \tau' + \varepsilon N \leqslant T} |I^\varepsilon(\tau') - I^\varepsilon(\tau'')| \geqslant \tfrac{\delta}{2} \Big\} \longrightarrow 0, \quad \text{as } \varepsilon \to 0. \tag{6.19}$$

By Prokhorov's theorem this relation implies the tightness of the family $\{I^\varepsilon(\cdot)\}$ in the space $C([0,T];\mathbb{R}^n)$. So any sequence $\varepsilon_j \to 0$ has a subsequence such that along it the laws $\mathcal{L}(I^\varepsilon(\cdot))$ converge weakly in $C([0,T];\mathbb{R}^n)$ to a limit probability measure Q^0. From Lemma 6.8 and convergence (6.19) we derive that for any $\delta > 0$ and $N > 0$

$$\lim_{\varepsilon \to 0} \mathbf{E}\, \frac{\#\big\{j : \{I^\varepsilon(\tau) : \tau \in \Delta_j\} \cap \mathcal{S}_{\frac{\delta}{2}} \ne \varnothing\big\}}{T/(\varepsilon N)} = 0.$$

For this proof we call an interval Δ_j *typical*, if in addition to the properties, listed in (3.16), the curve $\{I^\varepsilon(\tau) : \tau \in \Delta_j\}$ does not intersect the set $\mathcal{S}_{\frac{\delta}{2}}$. Then direct analogies of Lemmas 3.2 and 3.5 hold for the process $(I^\varepsilon(\tau), \varphi^\varepsilon(\tau))$ due to the argument, used in Section 3 and enriched by Lemma 6.8. Next, arguing in the same way as in Section 4, we show that the limiting measure Q^0 is a solution of the martingale problem for equation (6.12), satisfying $Q^0\{I(0) = I_0\} = 1$. Relation (6.16) follows from (6.17).

7. The effective equation

In this section we construct an effective equation for (6.1) with small ε. This is a v-equation such that under the mapping $v \mapsto I(v)$ its weak solutions go to weak solutions of (6.12). Thus by Theorem 6.7 the equation controls the behaviour of actions $I^k(v^\varepsilon(\tau;v_0))$ as $\varepsilon\to 0$. The construction of the effective equation is a finite-dimensional modification of the infinite-dimensional construction, used in [22] for purposes of averaging a stochastic PDE with analytic nonlinearity.

To get the effective equation we firstly remove from equation (6.1) the fast rotating terms $\varepsilon^{-1}W_k\mathbf{v}_k^\perp$, and then average the resulting equation with respect to the action of the n-torus on $\mathbb{R}^{2n}$, using the rules of the stochastic averaging, similar to how we earlier got the averaged I-equation (3.3) from the I-equation in (2.1). The action of $\mathbb{T}^n=\{\theta\}$ on $\mathbb{R}^{2n}$ is given by the block-diagonal matrix

$$\Phi_\theta=\operatorname{diag}\{\Phi_\theta^1,\dots,\Phi_\theta^n\},\quad \Phi_\theta^k=\begin{pmatrix}\cos\theta_k & -\sin\theta_k\\ \sin\theta_k & \cos\theta_k\end{pmatrix},\quad 1\leqslant k\leqslant n. \tag{7.1}$$

The drift in the effective equation is the averaging $\langle P\rangle=(\langle\mathbf{P}\rangle_1,\dots,\langle\mathbf{P}\rangle_n)$ of the vector field P in (6.1) with respect to the action Φ_θ. For convenience of future calculation we abbreviate $\langle P\rangle(v)=R(v)=(\mathbf{R}_1(v),\dots\mathbf{R}_n)(v)$. Then

$$\mathbf{R}_k(v):=\langle\mathbf{P}\rangle_k(v)=\int_{\mathbb{T}^n}\Phi_{-\theta}^k\mathbf{P}_k(\Phi_\theta v)d\theta,\quad k=1,\dots,n$$

(cf. [10, Section 3]).

To obtain the dispersion matrix $\langle\!\langle B\rangle\!\rangle(v)$ of the effective equation we start with $2n\times 2n$ matrix $X(v)=\big(X_{km}(v)\big)_{1\leqslant k,m\leqslant n}$, formed by 2×2-blocks $X_{km}(v)$,

$$X_{km}(v):=\sum_{j=1}^{n_1}\int_{\mathbb{T}^n}\Phi_{-\theta}^kB_{kj}(\Phi_\theta v)\big(B_{mj}(\Phi_\theta v)\big)^t\Phi_\theta^m d\theta.$$

That is, $X(v)=\int_{\mathbb{T}^n}\Phi_{-\theta}B(\Phi_\theta v)(B(\Phi_\theta v))^t\Phi_\theta d\theta$, where we denoted by B the block matrix $B=\big(B_{kj}\big)$. The dispersion matrix in question $\langle\!\langle B\rangle\!\rangle(v)=\big(\langle\!\langle B_{kj}\rangle\!\rangle(v)\big)_{1\leqslant k,j\leqslant n}$ is the principal square root of $X(v)$ (see the 10-th footnote). So

$$\sum_{j=1}^{n}\langle\!\langle B_{kj}\rangle\!\rangle(v)\langle\!\langle B_{mj}\rangle\!\rangle^t(v)=\sum_{j=1}^{n_1}\int_{\mathbb{T}^n}\Phi_{-\theta}^kB_{kj}(\Phi_\theta v)\big(B_{mj}(\Phi_\theta v)\big)^t\Phi_\theta^m d\theta \tag{7.2}$$

for $1\leqslant k,m\leqslant n$.

Then the effective equation for (6.1) is the following one:

$$d\mathbf{v}_k(\tau)=\mathbf{R}_k(v)d\tau+\sum_{j=1}^{n}\langle\!\langle B_{kj}\rangle\!\rangle(v)d\boldsymbol{\beta}_j(\tau),\quad k=1,\dots,n, \tag{7.3}$$

or

$$dv(\tau)=R(v)d\tau+\langle\!\langle B\rangle\!\rangle(v)d\beta(\tau),$$

where $\beta(\tau)=(\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_n)(\tau)$ is a standard Wiener process in $\mathbb{R}^{2n}$, defined on the space (6.2).

Proposition 7.1. *Under (2) and (3) of Assumption 6.1,*

i) the vector-function $R(v)$ and matrix functions $X(v)$ and $\langle\!\langle B\rangle\!\rangle(v)$ are locally Lipschitz in v;

ii) for any $\tilde\theta\in\mathbb{T}^n$, $R(\Phi_{-\tilde\theta}v)=\Phi_{-\tilde\theta}R(v)$, while $X(\Phi_{-\tilde\theta}v)=\Phi_{-\tilde\theta}X(v)\Phi_{\tilde\theta}$ and $\langle\!\langle B\rangle\!\rangle(\Phi_{-\tilde\theta}v)=\Phi_{-\tilde\theta}\langle\!\langle B\rangle\!\rangle(v)\Phi_{\tilde\theta}$.

Proof. i) The local Lipschitz continuity of $R(v)$ and $X(v)$ follow from the relations which define them. Since the operator $BB^t(v)$ is uniformly elliptic (see (6.3)), then $X(v)$ is uniformly elliptic as well, so the Lipschitz continuity of $\langle\!\langle B\rangle\!\rangle(v)=(X(v))^{1/2}$ is a consequence of [28, Lemma 5.2.1 and Theorem 5.2.2] and the Lipschitz continuity of $X(v)$.

ii) The relations for $R(v)$ and $X(v)$ are direct consequences of their definitions, and the relation for $\langle\!\langle B\rangle\!\rangle$ follows from that for X. □

Since the coefficients of equation (7.3) are locally Lipschitz, then its strong solution, if exists, is unique. So by the Yamada-Watanbe theorem (see in [15]) we have

Proposition 7.2. *If $v^1(\tau)$ and $v^2(\tau)$, $0 \le \tau \le T$, are weak solutions of equation (7.3) such that $\mathcal{D}(v^1(0)) = \mathcal{D}(v^2(0))$, then $\mathcal{D}(v^1(\cdot)) = \mathcal{D}(v^2(\cdot))$, and a strong solution with the initial data $v^1(0)$ exists for $0 \le \tau \le T$.*

Since for a matrix Q we have $\|Q\|^2_{HS} = \mathrm{tr}\, QQ^t$, then taking the trace of relation (7.2) with $k = m$ we get that $\sum_{j=1}^n \|\langle\!\langle B_{kj}\rangle\!\rangle(v)\|^2_{HS} = \sum_{j=1}^{n_1}\int_{\mathbb{T}^n}\|\Phi^k_\theta B_{kj}(\Phi_\theta v)\|^2_{HS}$. Using this equality we write Itô's formula for the actions $I_k(v(\tau))$, $k = 1,\dots,n$, of a solution $v(\tau)$ for (7.3), as

$$dI_k = \mathbf{v}_k^t\mathbf{R}_k(v)d\tau + \frac{1}{2}\Big(\sum_{j=1}^n\int_{\mathbb{T}^n}\|\Phi^k_\theta B_{kj}(\Phi_\theta v)\|^2_{HS}d\theta\Big)d\tau + \sum_{j=1}^n \mathbf{v}_k^t\langle\!\langle B_{kj}\rangle\!\rangle(v)d\boldsymbol{\beta}_j(\tau). \tag{7.4}$$

The first term of the drift in this equation may be re-written as

$$\mathbf{v}_k^t\mathbf{R}_k(v) = \mathbf{v}_k^t\int_{\mathbb{T}^n}\Phi^k_{-\theta}\mathbf{P}_k(\Phi_\theta v)d\theta = \int_{\mathbb{T}^n}(\mathbf{v}_k^t\mathbf{P}_k)(\Phi_\theta v)d\theta = \langle\mathbf{v}_k^t\mathbf{P}_k\rangle(I).$$

Since $\|\Phi^k_\theta B_{kj}(\Phi_\theta v)\|^2_{HS} = \|B_{kj}(\Phi_\theta v)\|^2_{HS}$, then in the second term of the drift we have $\int_{\mathbb{T}^n}\|\Phi^k_\theta B_{kj}(\Phi_\theta v)\|^2_{HS}d\theta = \langle\|B_{kj}\|^2_{HS}\rangle(I)$. Therefore the drift in (7.4) is $F_k(I)$, i.e. is the same as that in (6.12).

Using once again (7.2) we see that the diffusion matrix in (7.4) is $\bar S = (\bar S_{km})$ with

$$\begin{aligned}\bar S_{km} &:= \sum_j \mathbf{v}_k^t\langle\!\langle B_{kj}\rangle\!\rangle(v)(\langle\!\langle B_{mj}\rangle\!\rangle(v))^t\mathbf{v}_m\\ &= \sum_j\int_{\mathbb{T}^n}(\Phi^k_\theta\mathbf{v}_k)^tB_{kj}(\Phi_\theta v)(B_{mj}(\Phi_\theta v))^t\Phi^m_\theta\mathbf{v}_m d\theta\\ &= \Big\langle\sum_j\mathbf{v}_k^tB_{kj}(v)B^t_{mj}(v)\mathbf{v}_m\Big\rangle(I) = S_{km}(I),\end{aligned}$$

where $S_{km}(I)$ is as in (6.14). We conclude that the diffusion matrices of equations (7.4) and (6.12) also coincide. Hence,

Proposition 7.3. *Let the process $v(\tau) \in \mathbb{R}^{2n}$, $0 \leqslant \tau \leqslant T$ be a weak solution of (7.3).*

i) Then the process $\big(I(v(\tau)), v(\tau)\big) \in \mathbb{R}^n_+ \times \mathbb{R}^{2n}$, $0 \leqslant \tau \leqslant T$ is a strong solution of system (7.4)+(7.3), driven by the set of Brownian motions, corresponding to the weak solution v.

ii) The drift and diffusion matrix in (7.4) are functions of the actions $\{I_k\}$ and are the same as in equation (6.12). So the process $I(v(\tau))$, $0 \leqslant \tau \leqslant T$ is a weak solution of (6.12).

A disadvantage of system (7.4)+(7.3) is that dispersion matrix in (7.4) depends both on the actions I and angles φ, despite the corresponding diffusion matrix depends only on the actions. To fix this let us denote by $(\sqrt{2I(v)}, 0)$ the $2n$-vector with components $(\sqrt{2I_i(v)}, 0)^t$, $i = 1,\dots,n$. Then $(\sqrt{2I(v)}, 0) = \Phi_{-\varphi(v)}v$, so by Proposition 7.1.ii),

$$\begin{aligned}\mathbf{v}_k^t\langle\!\langle B_{kj}\rangle\!\rangle(v)d\boldsymbol{\beta}_j(\tau) &= (\sqrt{2I_k(v)}, 0)\Phi^k_{-\varphi(v)}\langle\!\langle B_{kj}\rangle\!\rangle(\Phi_{\varphi(v)}(\sqrt{2I(v)}, 0))d\boldsymbol{\beta}_j(\tau)\\ &= (\sqrt{I_k(v)}, 0)\langle\!\langle B_{kj}\rangle\!\rangle((\sqrt{2I(v)}, 0))\Phi^j_{-\varphi(v)}d\boldsymbol{\beta}_j(\tau)\\ &= M_{kj}(I)d\tilde{\boldsymbol{\beta}}_j(\tau).\end{aligned}$$

Here $d\tilde{\boldsymbol{\beta}}_j(\tau) = \Phi^j_{-\varphi(v)}d\boldsymbol{\beta}_j(\tau)$, $j = 1,\dots,n$, are differentials of independent standard Brownian motions in $\mathbb{R}^2$ and $M_{kj}(I)$ is the 2-vector $(\sqrt{I_k}, 0)\langle\!\langle B_{kj}\rangle\!\rangle\big((\sqrt{I}, 0)\big)$. Then equation (7.4) can be re-written as

$$dI_k = F_k(I)d\tau + \sum_{j=1}^{n} \big(M_{kj}(I), d\tilde{\boldsymbol\beta}_j(\tau)\big),\ k = 1,\dots,n. \tag{7.5}$$

When driven by the set of Brownian motions $\{\tilde{\boldsymbol\beta}_j(\tau)\}$, the effective equation (7.3) reads

$$d\mathbf{v}_k = \mathbf{R}_k(v)d\tau + \sum_{j=1}^{n} \tilde{B}_{kj}(v)d\tilde{\boldsymbol\beta}_j(\tau),\ k = 1,\dots,n, \tag{7.6}$$

where $\tilde{B}_{kj}(v) = \langle\!\langle B_{kj}\rangle\!\rangle(v)\Phi^j_{\varphi(v)}$. The system (7.5)+(7.6) is just the system (7.4)+(7.3), written using another standard Wiener process in $\mathbb{R}^{2n}$, so the two systems have the same sets of weak solutions. In difference with equation (7.3), dispersion matrix $\tilde{B}(v)$ in (7.6) is not locally Lipschitz. But for any $N \geq 1$ and $\delta > 0$ it is Lipschitz in domain $\{v : \|v\| \leq N, \|\mathbf{v}_j\| > \delta\ \forall j\}$.

The drift and diffusion in equation (7.5) are the same as in (7.4), so by Proposition 7.3 ii) they are the same as in equation (6.12). Thus equations (7.5) and (6.12) have the same set of weak solutions. We have established

Lemma 7.4. *Systems (7.5)+(7.6) and (7.4)+ (7.3) have the same set of weak solutions. So do equations (7.5) and (6.12).*

8. Lifting of solutions

In this section we prove that a weak solution $I(\tau)$ of the averaged equation (6.12), constructed in Theorem 6.7, is distributed as $I(v(\tau))$, where $v(\tau)$ is some weak solution of the effective equation (7.3). That is, $I(\tau)$ can be lifted to a weak solution of (7.3). We follow a strategy from [22], where such a lifting is constructed for an infinite dimensional equation. Since we work with weak solutions of the equations, then using Lemma 7.4 we replace averaged equation (6.12) by equation (7.5), and effective equation (7.3) – by equation (7.6). If $v(\tau)$ solves (7.6), then by Itô's formula the equation for the vector of actions is (7.5), while equations for φ_k's read

$$d\varphi_k(\tau) = R^\varphi_k(v)d\tau + \sum_{j=1}^{n} \mathfrak{R}^\varphi_{kj}(v)d\tilde{\boldsymbol\beta}_j(\tau), \quad k = 1,\dots,n, \tag{8.1}$$

where

$$R^\varphi_k(v) = \Big(\nabla_{\mathbf{v}_k}\arctan\big(\frac{v_k}{v_{-k}}\big)\Big)\cdot\mathbf{R}_k(v) + \frac{1}{2}\sum_{j=1}^{n}\mathrm{Trace}\Big(\tilde{B}_{kj}(v)\Big(\nabla^2_{\mathbf{v}_k}\arctan\big(\frac{v_k}{v_{-k}}\big)\Big)\big(\tilde{B}_{kj}(v)\big)^t\Big),$$

and $\mathfrak{R}^\varphi_{kj}(v) = \Big(\nabla_{\mathbf{v}_k}\arctan\big(\frac{v_k}{v_{-k}}\big)\Big)\cdot\tilde{B}_{kj}(v)$.

For any $\theta = (\theta_1,\dots,\theta_n) \in \mathbb{T}^n$ and any vector $I = (I_1,\dots,I_n) \in \mathbb{R}^n_+$ we set

$$V_\theta(I) = (\mathbf{V}_{\theta_1}(I_1),\dots,\mathbf{V}_{\theta_n}(I_n))), \tag{8.2}$$

where

$$\mathbf{V}_\alpha(J) = (\sqrt{2J}\cos\alpha, \sqrt{2J}\sin\alpha)^t \in \mathbb{R}^2.$$

Then $\varphi(V_\theta(I)) = \theta$, $I(V_\theta(I)) = I$ and the mapping $\mathbb{R}^n_+ \times \mathbb{T}^n : (I,\varphi) \mapsto V_\varphi(I)$ is a left-inverse for the mapping $v \mapsto (I,\varphi)$. For any $I = (I_1,\dots,I_n) \in \mathbb{R}^n_+$ let us denote

$$[I] = \min_{1\leqslant k\leqslant n}\{I_k\}. \tag{8.3}$$

Then for a $\delta > 0$ the mapping $v \mapsto (I,\varphi)$ defines a diffeomorphism of domain $\mathbb{R}^{2n}_\delta := \{v \in \mathbb{R}^{2n} : [I(v)] > \delta\}$ and domain $\{I \in \mathbb{R}^n_+ : [I] > \delta\} \times \mathbb{T}^n$. Therefore we have

Lemma 8.1. *For any $\delta > 0$, on domain $\mathbb{R}^{2n}_\delta$ equation (7.6) is equivalent to system (7.5)+(8.1) in the following sense: let $\tau_1 \leqslant \tau_2$ be two stopping times with respect to the natural filtration. Then*

i) if for $\tau_1 \le \tau \le \tau_2$ a process $v(\tau)$ lies in $\mathbb{R}^{2n}_\delta$ and is a weak solution of (7.6), then for such τ's $(I(v), \varphi(v))$ is a weak solution of (7.5)+(8.1);

ii) if for $\tau_1 \le \tau \le \tau_2$ a process $(I(\tau), \varphi(\tau))$ satisfies $[I(\tau)] > \delta$ and is a weak solution of (7.5)+(8.1), where $v(\tau) = V_{\varphi(\tau)}(I(\tau))$, then for such τ's the process $v(\tau) = V_{\varphi(\tau)}(I(\tau))$ is a weak solution of (7.6).

The following statement is the main result of this section.

Theorem 8.2. *If the process $I^0(\tau) = (I^0_k(\tau), k = 1, \dots, n) \in \mathbb{R}^n_+$, $0 \leqslant \tau \leqslant T$, is a weak solution of equation (6.12), constructed in Theorem 6.7 by taking the limit along a sequence $\varepsilon_i \to 0$, then, for any vector $\theta \in \mathbb{T}^n$, there exists a weak solution $v^\theta(\tau) \in \mathbb{R}^{2n}$, $0 \leqslant \tau \leqslant T$ of the effective equation (7.3) such that*

(i) the law of $I(v^\theta(\cdot))$ coincides with that of $I^0(\cdot)$;

(ii) $v^\theta(0) = V_\theta(I_0)$ a.s.

The properties of solutions of equation (7.3) are important for analysis in the following sections. They are the subject of Theorem 8.6 below.

We begin with explaining the key ideas of the theorem's proof.

Strategy of the proof: By Lemma 7.4 we may regard $I^0(\tau)$ as a weak solution of equation (7.5). Since the uniqueness of a solution $v(\tau)$ is claimed in Proposition 7.2, then only its existence and properties (i), (ii) should be established. For any $\delta > 0$, we will divide $[0, T)$ into a finite or countable set of random closed intervals (see Figure 1) Λ_j, $j \geqslant 0$, and Δ_j, $j \geqslant 1$, such that

(1) $\Lambda_0 \leqslant \Delta_1 \leqslant \Lambda_1 \leqslant \Delta_2 \leqslant \dots$,

(2) $[I^0] \geqslant \delta$ on each Λ_j, and $[I^0] \leqslant 2\delta$ on each Δ_j.

For definiteness we assume that $[I_0] \ge \delta$.

Next we iteratively construct on these intervals a process $v^\delta(\tau)$ such that

$$I(v^\delta(\tau)) \equiv I^0(\tau), \quad a.e. \tag{8.4}$$

Suppose that we already know v^δ at the left end point of some Λ_j. To construct $v^\delta(\tau)$ on Λ_j we note that since on every Λ_j we have $[I(v^\delta(\cdot))] = [I^0(\cdot)] \geqslant \delta$, then by Lemma 8.1 there equation (7.6) is equivalent to (7.5)+(8.1). As $I(\tau) = I^0(\tau)$ is known, it remains to solve (8.1), regarded as a stochastic equation with progressively measurable coefficients for n-vector $\varphi(\tau)$. Since the initial value of φ is given on the left end point of Λ_j, such a solution φ is uniquely determined. Then for $\tau \in \Lambda_j$ we set $v^\delta(\tau) := V_{\varphi(\tau)}(I^0(\tau))$. By Lemma 8.1 this v^δ solves (7.6) weakly on Λ_j. Clearly, (8.4) is satisfied.

On the next interval Δ_{j+1} we have $[I^0] \leqslant 2\delta$. We want to extend $v^\delta(\tau)$ to Δ_{j+1}, keeping the property (8.4), so that when eventually v^δ is constructed on all Λ_j's and Δ_k's, we may obtain the desired weak solution of (7.6) by taking a limit as $\delta \to 0$. Such a task turns out to be not easy since by (8.4) we have $\mathbf{v}_k = \big(\sqrt{2I_k(\tau)}\cos\varphi_k(\tau), \sqrt{2I_k(\tau)}\sin\varphi_k(\tau)\big)$ with some phase function φ_k, and so on Δ_{j+1} a-priori $|\dot{\mathbf{v}}_k| \sim \delta^{-1/2}$. Hence, a naive extension may fail to guarantee the existence of a limit as $\delta \to 0$. To construct a right lifting of $I^0(\tau)$ when it is small, we use the fact the $I^0(\tau)$ is a limit of the process $I^{\varepsilon_i}(\tau) := I(v^{\varepsilon_i}(\tau))$, where v^{ε_i} solves equation (6.1) with $\varepsilon = \varepsilon_i$. The process $v^{\varepsilon_i}(\tau)$ is a lifting of $I^{\varepsilon_i}(\tau)$ which is singular as $\varepsilon_i \to 0$. In Appendix A for each $\varepsilon_i > 0$ a modified process $\bar{v}^{\varepsilon_i}(\tau)$ is constructed such that $I(\bar{v}^{\varepsilon_i}(\tau)) = I(v^{\varepsilon_i}(\tau))$ and $|\frac{d}{d\tau}\bar{v}^{\varepsilon_i}| \sim 1$ as $\varepsilon_i \to 0$. A limit in law of processes $\bar{v}^{\varepsilon_i}(\cdot)$ as $\varepsilon_i \to 0$ provides a right lifting of I^0 on interval Δ_{j+1}. We thus extend the process $v^\delta(\tau)$ to Δ_{j+1} in such a way that (8.4) holds and $|\dot{v}^\delta| \sim 1$.

By iterating the two constructions above we obtain a process $v^\delta(\tau)$ which solves (7.6) for $\tau \in \cup\Lambda_j$ and satisfies good estimates on the complementary set $\cup\Delta_j$. By Theorem 6.7 the Lebesgue measure of $\cup\Delta_j$ becomes small with δ. This allows us to show that any limit distribution of the process $v^\delta(\cdot)$ as $\delta \to 0$ gives a weak solution $v(\cdot)$ for (7.6).

To run this construction we need good upper bounds for the numbers of intervals Λ_j and Δ_j, which could be large when the norm of the process $I^0(\tau)$ is large. To get the bounds we begin the proof by introducing the stoping times $\tau_N = \inf_\tau\{|I^0(\tau)| = N/2\}$ and replacing $I^0(\tau)$ by a trivial modification for $\tau \geqslant \tau_N$. Then we first obtain a weak solution $v_N(\tau)$, corresponding to the modified process $I^0_N(\tau)$, and next take the limit as $N \to \infty$ to get a real weak solution $v(\tau)$ of (7.6).

Proof. Let us introduce a natural filtered measurable space $\{\Omega, \mathcal{B}, \{\mathcal{B}_\tau\}, 0 \leqslant \tau \leqslant T\}$ for the problem we consider, where Ω is the Banach space

$$\Omega = \Omega_I \times \Omega_v := C([0,T], \mathbb{R}^n_+) \times C([0,T], \mathbb{R}^{2n}), \tag{8.5}$$

$\mathcal{B}$ is its Borel σ-algebra and $\mathcal{B}_\tau$ is the σ-algebra generated by the set of random variables $\{r(s) : 0 \leqslant s \leqslant \tau \text{ and } r(\cdot) \in \Omega\}$. Denote by π_I and π_v the natural projections $\pi_I : \Omega \to \Omega_I$ and $\pi_v : \Omega \to \Omega_v$. We will prove the theorem by constructing a probability measure $\mathbf{Q}$ on Ω such that $\pi_I \circ \mathbf{Q} = \mathcal{D}(I^0(\cdot))$, $\pi_v \circ \mathbf{Q}$ is the distribution of a weak solution of (7.6), and $\mathbf{Q}$-a.s. for $(I'(\cdot), v'(\cdot)) \in \Omega$ we have $I(v') = I'$. This will be achieved in four steps.

Step 1. Redefine the equations for large amplitudes.

For any $N \in \mathbb{N}$ consider the stopping time

$$\tau_N = \inf\{\tau \in [0,T] \big| \|v(\tau)\|^2 = 2|I(v(\tau))| = N\},$$

(here and in similar situations below $\tau_N = T$ if the set is empty). For $\tau \geqslant \tau_N$ and each $\varepsilon > 0$ we redefine equation (6.1) to the trivial equation

$$d\mathbf{v}_k = d\boldsymbol{\beta}_k(\tau), \ k = 1, \dots, n, \tag{8.6}$$

and redefine accordingly equations (6.10), (7.5) (7.6) and (8.1). We denote the new equations as $(6.1)_N$, $(6.10)_N$ $(7.5)_N$, $(7.6)_N$ and $(8.1)_N$. So if $v^\varepsilon_N(\tau)$ is a solution of $(6.1)_N$, then $I^\varepsilon_N(\tau) := I(v^\varepsilon_N(\tau))$ satisfies $(6.10)_N$. That is, for $\tau \leqslant \tau_N$, it satisfies (6.10), while for $\tau \geqslant \tau_N$ it is a solution of the equation

$$dI_k = \tfrac{1}{2}d\tau + (v_k d\beta_k + v_{-k} d\beta_{-k}) = \tfrac{1}{2}d\tau + \sqrt{2I_k}\, dw_k(\tau), \quad k = 1, \dots, n,$$

where $w_k(\tau)$ is the Wiener process $\int_0^\tau (\cos\varphi_k d\beta_k(\tau) + \sin\varphi_k d\beta_{-k}(\tau))$. So $(6.10)_N$ is the equation

$$dI_k = \mathbf{1}_{\tau\leqslant\tau_N} \cdot \langle \text{r.h.s of } (6.10)\rangle + \mathbf{1}_{\tau\geqslant\tau_N}\big(\tfrac{1}{2}d\tau + \sqrt{2I_k}\, dw_k(\tau)\big), \quad k = 1, \dots, n.$$

Accordingly, the averaged equation $(7.5)_N$ reads

$$dI_k = \mathbf{1}_{\tau\leqslant\tau_N}\Big(F_k(I)d\tau + \sum_{j=1}^n \big(M_{kj}(I), d\tilde{\boldsymbol{\beta}}_j(\tau)\big)\Big) + \mathbf{1}_{\tau\geqslant\tau_N}\Big(\tfrac{1}{2}d\tau + \sqrt{2I_k}d\tilde{\beta}_k(\tau)\Big), \ k = 1, \dots, n.$$

Here $\tilde{\boldsymbol{\beta}}_j(\tau)$, $j = 1, \dots, n$, are independent standard Brownian motions in $\mathbb{R}^2$.

For the sequence $\varepsilon_i \to 0$, where we have the convergence $\mathcal{D}(I^{\varepsilon_i}(\cdot)) \rightharpoonup \mathcal{D}(I^0(\cdot))$, choosing a suitable subsequence (if neccessary) we achieve that also $\mathcal{D}(I^{\varepsilon_i}_N(\cdot)) \rightharpoonup \mathcal{D}(I_N(\cdot))$ for some process $I_N(\tau)$, for each $N \in \mathbb{N}$.

Lemma 8.3. *For every $N \in \mathbb{N}$,*

i) the process $I_N(\tau)$, $0 \leqslant \tau \leqslant T$, is a weak solution of $(7.5)_N$ such that $\mathcal{D}(I_N) = \mathcal{D}(I^0)$ for $\tau \leqslant \tau_N$ and $\mathcal{D}(I_N(\cdot)) \rightharpoonup \mathcal{D}(I^0(\cdot))$ as $N \to \infty$.

ii) the statement in Lemma 8.1 holds true with $(7.6)_N$ and $(7.5)_N$+$(8.1)_N$.

Proof. i) The first part of the statement follows from the same argument as in the proof of Theorem 6.7. Recalling that, for each $\varepsilon > 0$, $\mathcal{D}(I_N^\varepsilon(\cdot)) = \mathcal{D}(I^\varepsilon(\cdot)) =: \mathbf{Q}^\varepsilon$ for $\tau \leqslant \tau_N$ and passing to the limit as $\varepsilon_i \to 0$ we get the second assertion of the lemma. As $\mathbf{Q}^\varepsilon\{\tau_N < T\} \leqslant CN^{-1}$ uniformly in ε, then the last assertion also follows.

ii) The assertion follows by the same argument as that used to prove Lemma 8.1. □

Now we fix any $N \in \mathbb{N}$. Our goal is to construct for each $\delta > 0$ a measure $\mathbf{Q}_\delta^N$ on Ω such that for its natural process $(I_\delta^N(\tau), v_\delta^N(\tau))$, $\tau \in [0,T]$ we have

(1) $\mathcal{D}(I_\delta^N(\cdot)) = \mathcal{D}(I_N(\cdot))$;
(2) $I(v_\delta^N(\cdot)) \equiv I_\delta^N(\cdot)$, $\mathbf{Q}_\delta^N$-a.s;
(3) the process $v_N^\delta(\tau)$ is an Itô process with bounded (in terms of N) drift term and diffusion matrix. Moreover, it solves $(7.6)_N$ for τ outside a small random set, where $[I(\tau)] \lesssim \delta$ (see (8.3)).

Next we will prove the assertion of Theorem 8.2 by taking the limits $\delta \to 0$ and $N \to \infty$.

Step 2: Construction of the measure $\mathbf{Q}_\delta := \mathbf{Q}_\delta^N$ for every $\delta > 0$ and N fixed.

We start with finding an auxiliary Itô process $\bar{w}(\tau)$ which covers a version of the process $I_N(\tau)$ (but has no relation with the effective equation (7.6)). Since N is fixed, then below in the step the index N usually is dropped.

Lemma 8.4. *There exists a continuous $\mathcal{B}_\tau$-adapted continuous process $(\bar{I}(\tau), \bar{w}(\tau)) \in \mathbb{R}_+^n \times \mathbb{R}^{2n}$, $0 \leqslant \tau \leqslant T$, such that $\mathcal{D}(\bar{I}(\cdot)) = \mathcal{D}(I_N(\cdot))$, $\bar{I}(\cdot) = I(\bar{w}(\cdot))$ a.s., and $\bar{w}(\tau)$ is an Itô process in $\mathbb{R}^{2n}$ of the form*

$$\begin{aligned} d\bar{w}(\tau) &= B(\tau)d\tau + a(\tau)d\beta(\tau), \quad \tau \in [0,T], \\ |B(\tau)| &\leqslant C, \quad C^{-1}\mathbb{I} \leqslant a(\tau)\big(a(\tau)\big)^t \leqslant C\mathbb{I}, \end{aligned} \tag{8.7}$$

where the constant C depends only on N, and $\beta(\tau)$ is a standard Brownian motion in $\mathbb{R}^{2n}$.

Proof. By the construction in Appendix A (see there (A.4)), for each $\delta > 0$ there exists an Itô process $\bar{w}^\delta(\tau) \in \mathbb{R}^n$, $0 \leqslant \tau \leqslant T$, such that

$$d\bar{w}^\delta(\tau) = B^\delta(\tau)d\tau + a^\delta(\tau)d\beta(\tau), \quad \tau \in [0,T].$$

Here

$$|B^\delta(\tau)| \leqslant C, \quad C^{-1}\mathbb{I} \leqslant a^\delta(\tau)\big(a(\tau)\big)^t \leqslant C\mathbb{I},$$

with a uniform in δ constant $C > 0$ depending only on N and $I(\bar{w}^\delta(\cdot)) = I(v^\varepsilon(\cdot))$ a.s., where v^ε solves equation $(6.1)_N$.

Since the family of the processes $\Big\{\big(I(\bar{w}^\delta(\cdot)), \bar{w}^\delta(\cdot)\big)\Big\}_{\delta\in(0,1]}$ is tight in $C([0,T];\mathbb{R}_+^n\times\mathbb{R}^{2n})$, then taking if necessary a subsequence $\delta_i \to 0$ we achieve that the process $\big(I(\bar{w}^{\delta_i}(\cdot)), \bar{w}^{\delta_i}(\cdot)\big)$ weakly converges in law in $C([0,T];\mathbb{R}_+^n\times\mathbb{R}^{2n})$ to a limit process $(\bar{I}(\tau), \bar{w}(\tau))$, $0 \leqslant \tau \leqslant T$. Using the same arguments as in **Step 3** of Appendix A we conclude that $\bar{w}(\tau)$ admits representation (8.7) (cf. equation (A.5)). So $(\bar{I}(\tau), \bar{w}(\tau))$ is a desired process. □

Now let $(\bar{I}(\tau), \bar{w}(\tau))$, $0 \leqslant \tau \leqslant T$, be a continuous process as in Lemma 8.4. Fix any $\delta > 0$. For the process $\bar{I}(\tau)$, $0 \leqslant \tau \leqslant T$, we define stopping times $\tau_j^\pm \leqslant T$ such that $\cdots < \tau_j^- < \tau_j^+ < \tau_{j+1}^- < \dots$, similarly to the stopping times in Appendix A. Namely,

(1) If $[\bar{I}(0)] \leqslant \delta$, then $\tau_1^- = 0$; otherwise $\tau_0^+ = 0$.
(2) If τ_j^- is defined, then τ_j^+ is the first moment after τ_j^- when $[\bar{I}(\tau)] \geqslant 2\delta$ (if this never happens, then we set $\tau_j^+ = T$; similar in the item below).

(3) If τ_j^+ is defined, then τ_{j+1}^- is the first moment after τ_j^+ when $[\bar I(\tau)] \leqslant \delta$.

We denote $\Delta_j = [\tau_j^-, \tau_j^+]$, $\Lambda_j = [\tau_j^+, \tau_{j+1}^-]$ and set

$$\Delta^\delta = \cup \Delta_j, \quad \Lambda^\delta = \cup \Lambda_j.$$

See Figure 1.

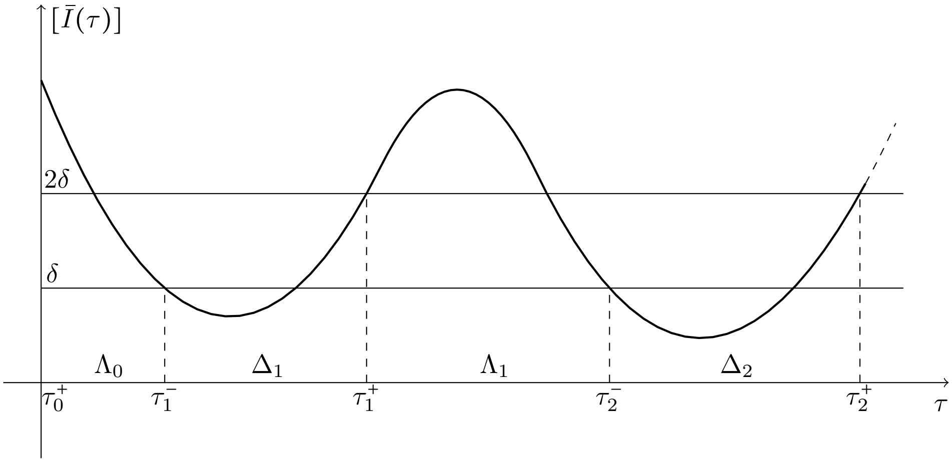

FIGURE 1. A typical behaviour of the stopping times $\tau_j^\pm$

By Theorem 6.7,

$$\mathbf{E}\,\mathcal{L}(\Delta^\delta) = o_\delta(1), \quad \text{so } \mathbf{E}\,\mathcal{L}(\Lambda^\delta) = T - o_\delta(1), \tag{8.8}$$

where $\mathcal{L}$ is the Lebesgue measure and $o_\delta(1)$ goes to zero with δ. Due to the truncation (8.6) for $\|v\| \geqslant N^{1/2}$, there exists $c(\delta, N) > 0$ depending only on N and δ such that for each j,

$$\mathbf{E}\mathcal{L}(\Lambda_j) \geqslant c(\delta, N) > 0. \tag{8.9}$$

For $j = 0, 1, \dots$ we will iteratively construct on segments $[0, \tau_j^\pm]$ continuous process $(I(\cdot), v(\cdot))$ (defined on the space Ω as in (8.5)) such that $I(\tau) = I(v(\tau))$ a.s. and $\mathcal{D}(I(\cdot)) = \mathcal{D}(I_N(\cdot))$. Moreover, on each segment $\Lambda_l \subset [0, \tau_j^\pm]$ the process $v(\tau)$, $\tau \in \Lambda_l$, will be a weak solution of $(7.6)_N$. Next we will obtain a desirable measure $\mathbf{Q}_\delta$ as a limit of the laws of these processes as $j \to \infty$.

For the sake of definiteness we assume that $[I(0)] > \delta$, so $0 = \tau_0^+$. With suitably chosen $\mathcal{B}_\tau$-adapted Brownian motions $\{\tilde{\boldsymbol\beta}_j, j = 1, \dots, n\}$, we can assume that the process $\bar I(\tau)$, $\tau \in [0, T]$ is a strong solution of $(7.5)_N$.

a) Let $\tau \in \Lambda_0$. Substituting in $(8.1)_N$ $v = V_{\varphi(\tau)}(\bar I(\tau))$, we get for the n-vector $\varphi(\tau)$ an equation, denoted by (S_φ), with $\mathcal{B}_\tau$-adapted coefficients and driven by $\{\tilde{\boldsymbol\beta}_j\}$ (cf. equation (8.1)). Moreover, the drift term and the dispersion matrix are uniformly Lipschitz continuous in φ, where the Lipschitz constants may depend on δ and N. Hence, for any $\theta \in \mathbb{T}^n$ there exists a unique solution $\varphi(\tau)$, $\tau \in \Lambda_0$, of (S_φ) with $\varphi(0) = \theta$. Then by Lemma 8.3 the process $\tilde v^1(\tau) = V_{\varphi(\tau)}(\bar I(\tau))$, $\tau \in \Lambda_0$, is a weak solution of $(7.6)_N$. Obviously $I(\tilde v^1) \equiv \bar I$ and $\tilde v^1(0) = V_\theta(I_0)$.

b) Let $\tau \geq \tau_1^-$. For $(\xi_1, \xi_2) \in (\mathbb{R}^2 \smallsetminus \{0\}) \times (\mathbb{R}^2 \smallsetminus \{0\})$ we denote by $U(\xi_1, \xi_2) \in \mathrm{SO}(2)$ the unique rotation of $\mathbb{R}^2$ that maps $\frac{\xi_2}{|\xi_2|}$ to $\frac{\xi_1}{|\xi_1|}$ (cf. Appendix A). Next for any two vectors $v^j = (\mathbf{v}_1^j, \dots, \mathbf{v}_n^j) \in (\mathbb{R}^2 \smallsetminus \{0\})^n$, $j = 1, 2$, we set

$$\mathcal{U}(v^1, v^2) := \mathrm{diag}\{U(\mathbf{v}_1^1, \mathbf{v}_1^2), \dots, U(\mathbf{v}_n^1, \mathbf{v}_n^2)\}.$$

Consider the process $\hat v^1(\tau) := \mathcal{U}\Big(\tilde v^1(\tau_1^-), \bar w(\tau_1^-)\Big)\bar w(\tau)$, $\tau \geq \tau_1^-$. This still is an Itô process of the form as in (8.7) with the same constant C. Moreover,

$$\hat v^1(\tau_1^-) = \tilde v^1(\tau_1^-),$$

and $I(\hat v^1(\cdot)) = \bar I(\cdot)$ a.s. Now consider the continuous process

$$\big(\bar I^1(\tau), \bar v^1(\tau)\big) := \big(\bar I(\tau), \mathbf{1}_{\tau\leqslant\tau_1^-}\tilde v^1(\tau) + \mathbf{1}_{\tau>\tau_1^-}\hat v^1(\tau)\big),\ \tau \in [0,T],$$

and denote $\mathbf{Q}_{1,\delta} := \mathcal{D}\Big(\big(\bar I^1(\cdot), \bar v^1(\cdot)\big)\Big) \in \mathcal{P}(\Omega)$. Clearly, we have $\pi_I \circ \mathbf{Q}_{1,\delta} = \mathcal{D}(\bar I^1(\cdot)) = \mathcal{D}(I_N(\cdot))$, and $I(\bar v^1(\cdot)) = \bar I^1(\cdot)$, $\mathbf{Q}_{1,\delta}$-a.s. Furthermore the process $\bar v^1(\cdot)$ solves $(7.6)_N$ weakly on random interval $[0,\tau_1^+]$, while on Δ_1 and on the whole interval $[\tau_1^+, T]$ it is an Itô process of the form (8.7).

c) For $\tau \in \Lambda_1$, by the same method as in a) we construct a weak solution $\tilde v^2(\tau)$, $\tau \in \Lambda_1$, of equation $(7.6)_N$, equal $\bar v^1(\tau_1^+)$ at $\tau = \tau_1^+$, a.s. Consider the continuous process

$$\big(\bar I^2(\tau), \bar v^2(\tau)\big) := \big(\bar I(\tau), \mathbf{1}_{\tau\leqslant\tau_1^+}\bar v^1(\tau) + \mathbf{1}_{\tau_1^+<\tau\leqslant\tau_2^-}\tilde v^2(\tau) + \mathbf{1}_{\tau\geqslant\tau_2^-}\hat v^2(\tau)\big),\ \tau \in [0,T],$$

where $\hat v^2(\cdot) := \mathcal{U}\Big(\tilde v^2(\tau_2^-), \bar w(\tau_2^-)\Big)\bar w(\cdot)$. Denote $\mathbf{Q}_{2,\delta} = \mathcal{D}\Big(\big(\bar I^2(\cdot), \bar v^2(\cdot)\big)\Big)$. As in b), $\pi_I \circ \mathbf{Q}_{2,\delta} = \mathcal{D}(\bar I^2(\cdot)) = \mathcal{D}(I_N(\cdot))$ and $I(\bar v^2(\cdot)) = \bar I^2(\cdot)$, $\mathbf{Q}_{2,\delta}$-a.s. The process $\bar v^2(\cdot)$ solves $(7.6)_N$ on random intervals Λ_j, $j = 0,1$ and is an Itô process of the form (8.7) on $[0,T] \smallsetminus \cup_{j=0}^1 \Lambda_j$.

d) Iteratively we construct on the space Ω measures $\mathbf{Q}_{j,\delta}$, $j \in \mathbb{N}$. Due to (8.9) we know that a.s. the sequence $\tau_j^\pm$ stabilizes at $\tau = T$ after a finite (random) number of steps. Accordingly, as $j \to \infty$, the sequence of measure $\mathbf{Q}_{j,\delta}$ converges to a limiting measure $\mathbf{Q}_\delta = \mathbf{Q}_\delta^N$ on Ω.

Let $(I_\delta(\tau), v_\delta(\tau))$, $\tau \in [0,T]$ be the natural process of the measure $\mathbf{Q}_\delta$. We then have

i) $\mathcal{D}(I_\delta(\cdot)) = \mathcal{D}(I_N(\cdot))$;

ii) $I(v_\delta(\cdot)) = I_\delta(\cdot)$ $\mathbf{Q}_\delta$-a.s.;

iii) for $\tau \in \Lambda^\delta$ the process v_δ is a weak solution of $(7.6)_N$, while for $\tau \in \Delta^\delta$ $v_\delta(\tau)$ is an Itô process as in (8.7), where C does not depend on N.

Step 3. Limit $\delta \to 0$.

From the construction we know that the set of measures $\{\mathbf{Q}_\delta^N, 0 \leqslant \delta \leqslant 1\}$ is tight. Let $\mathbf{Q}^N$ be any limiting measure as $\delta \to 0$. Then

$$\mathbf{Q}_{\delta_j}^N \rightharpoonup \mathbf{Q}^N \quad \text{as} \quad \delta_j \to 0, \tag{8.10}$$

for some sequence $\{\delta_j\}$. Since $\mathcal{D}(I_\delta(\cdot)) = \mathcal{D}(\bar I(\cdot)) = \mathcal{D}(I_N(\cdot))\ \ \forall \delta > 0$, then $\pi_I \circ \mathbf{Q}^N = \mathcal{D}(I_N(\cdot))$. For the projection of $\mathbf{Q}^N$ to Ω_v we have

Lemma 8.5. *The measure $\mathbf{P}^N := \pi_v \circ \mathbf{Q}^N$ is a solution of the martingale problem for equation $(7.6)_N$.*

Proof. Let us denote the drift terms and dispersion terms of $(7.6)_N$ as $\mathbf{R}_k^N(\tau, v)$ and $\tilde B_{kj}^N(\tau, v)$, respectively. Consider the natural process $v_\delta = (\mathbf{v}_{1,\delta}, \cdots, \mathbf{v}_{n,\delta})$ of the measure $\mathbf{P}_\delta^N := \pi_v \circ \mathbf{Q}_\delta^N$ on Ω_v. It satisfies the system of Itô equations,

$$\begin{aligned} d\mathbf{v}_k(\tau) &= \Big(\mathbf{1}_{\tau\in\Lambda^\delta}\mathbf{R}_k^N(\tau, v) + \mathbf{1}_{\tau\in\Delta^\delta}B_k(\tau)\Big)d\tau \\ &\quad + \mathbf{1}_{\tau\in\Lambda^\delta}\sum_{j=1}^n \tilde B_{kj}^N(\tau, v)d\boldsymbol{\beta}_j(\tau) + \mathbf{1}_{\tau\in\Delta^\delta}\sum_{j=1}^n a_{kj}(\tau)d\mathbf{w}_j(\tau) \qquad k = 1,\dots,n. \\ &=: \mathbf{A}_k^\delta(\tau)d\tau + \sum_{j=1}^n \Big(G_{kj}^\delta(\tau, v)d\boldsymbol{\beta}_j(\tau) + C_{kj}^\delta(\tau)d\mathbf{w}_j(\tau)\Big), \end{aligned} \tag{8.11}$$

Here $B(\tau)$ and $a(\tau)$ are the drift and dispersion, corresponding to the Itô process v_δ on Δ^δ (see item d)iii) of Step 2), so dispersion matrices $G^\delta_{kj}(\tau)$ and $C^\delta_{kj}(\tau)$ are supported by non-intersecting unions of random time-intervals Λ^δ and Δ^δ. Furthermore, for any $\delta > 0$ and $k, m = 1, \dots, n$, we have

i) the process $\gamma^\delta_k(\tau) = \mathbf{v}_k(\tau) - \int_0^\tau \mathbf{A}^\delta_k(s)ds \in \mathbb{R}^2$ is a $\mathbf{P}^N_\delta$-martingale;

ii) the process $\Gamma^\delta_{km} = \gamma^\delta_k(\tau)\big(\gamma^\delta_m(\tau)\big)^t - \frac{1}{2}\int_0^\tau \big(X^\delta_{km}(s) + Y^\delta_{km}(s)\big)ds$, where

$$X^\delta_{km}(s) = \sum_{j=1}^n G^\delta_{kj}(s)\big(G^\delta_{mj}(s)\big)^t, \qquad Y^\delta_{km}(s) = \sum_{j=1}^n C^\delta_{kj}(s)\big(C^\delta_{mj}(s)\big)^t,$$

is a $\mathbf{P}^N_\delta$-martingale.

Note that for any $\delta > 0$, by (8.8), we have

$$\begin{aligned} &\mathbf{E}^{\mathbf{P}^N_\delta} \sup_{0\leqslant\tau\leqslant T} \Big| \int_0^\tau (\mathbf{A}^\delta_k(s) - \mathbf{R}^N_k(s, v(s))ds \Big| \\ &\leqslant \mathbf{E}^{\mathbf{P}^N_\delta} \int_{\Delta^\delta} \Big(|\mathbf{R}^N_k(s, v(s))| + |B_k(s)| \Big) ds \leqslant C(N) o_\delta(1). \end{aligned} \tag{8.12}$$

By (8.10), $\mathbf{P}^N_{\delta_j} \rightharpoonup \mathbf{P}^N := \pi_v(Q^N)$ as $\delta_j \to 0$.

(1) We first show that the process $\gamma^0_k(\tau) = \mathbf{v}_k(\tau) - \int_0^\tau \mathbf{R}^N_k(s, v(s))ds$ is a $\mathbf{P}^N$-martingale. Let us take any $0 \leqslant \tau_1 \leqslant \tau_2 \leqslant T$ and $\Psi \in C_b(\Omega_v)$ such that $\Psi(\xi(\cdot))$ depends only on $\xi(\tau)$ with $\tau \in [0, \tau_1]$. We have to show that

$$\mathbf{E}^{\mathbf{P}^N} \Big(\big(\gamma^0_k(\tau_2) - \gamma^0_k(\tau_1)\big)\Psi(\xi) \Big) = 0. \tag{8.13}$$

The l.h.s equals

$$\begin{aligned} &\lim_{\delta_j\to 0} \mathbf{E}^{\mathbf{P}^N_{\delta_j}} \big(\gamma^0_k(\tau_2) - \gamma^0_k(\tau_1)\Psi(\xi) \big) \\ &= \lim_{\delta_j\to 0} \mathbf{E}^{\mathbf{P}^N_{\delta_j}} \Big(\Psi(\xi)\big(\mathbf{v}_k(\tau_2) - \mathbf{v}_k(\tau_1) - \int_{\tau_1}^{\tau_2} \mathbf{R}^N_k(s, v(s))ds\big) \Big) \\ &= \lim_{\delta_j\to 0} \mathbf{E}^{\mathbf{P}^N_{\delta_j}} \Big(\Psi(\xi)\big(\int_{\tau_1}^{\tau_2} (\mathbf{A}^\delta_k(s) - \mathbf{R}^N_k(v(s)))ds \big) \Big) = 0, \end{aligned}$$

where in the second equality we use the fact that γ^δ_k is a $\mathbf{P}^N_\delta$-martingale and in the third we use (8.12). So (8.13) is established, and so $\gamma^0_k(\tau)$ is a $\mathbf{P}^N$-martingale.

(2) We then show the process $\Gamma^0_{km}(s) = (\gamma^0_k(\tau))(\gamma^0_m(\tau))^t - \frac{1}{2}\int_0^\tau X^0_{km}(s)ds$ is a $\mathbf{P}^N$-martingale, where $X^0_{km}(s) = \sum_{j=1}^n \tilde{B}^N_{kj}(s, v(s))\big(\tilde{B}^N_{mj}(s, v(s))\big)^t$. Note that by the definition, for $\delta > 0$ we have

$$\mathbf{E} \sup_{0\leqslant\tau\leqslant T} \Big| \int_0^\tau (X^0_{km}(s) - X^\delta_{km}(s) - Y^\delta_{km}(s))ds \Big| \leqslant C(N)\mathbf{E}^{\mathbf{P}^N_\delta} \mathbf{1}_{s\in\Delta^\delta} \leqslant C(N)o_\delta(1).$$

Then Γ^0_{km} is a $\mathbf{P}^N$-martingale due to the same reasoning as in (1). This finishes the proof of the lemma. □

Step 4. Limit $N \to \infty$.

By the construction and estimate (6.15) the set of measures $\{\mathbf{Q}^N, N \in \mathbb{N}\}$ is tight. Consider any limiting measure $\mathbf{Q}$ for this family as $N \to \infty$. Repeating in a simpler way the proof of Lemma 8.5, we find that $\pi_v \circ \mathbf{Q}$ solves the martingale problem of (7.6). Therefore, it is a weak solution for (7.6) with $\pi_v \circ \mathbf{Q}(v(0) = V_\theta(I)) = 1$ and $I \circ \pi_v \circ \mathbf{Q} = \mathcal{D}(I^0(\cdot))$. Hence the assertion of Theorem 8.2 is established. □

Theorem 8.2, Proposition 7.2 and Theorem 6.7 jointly imply the following.

Theorem 8.6. *Under Assumption 6.1,* *i) For any $v_0 \in \mathbb{R}^{2n}$ effective equation (7.3) has a unique strong solution $v(\tau; v_0)$, $\tau \in [0, T]$, equal v_0 at $\tau = 0$. It satisfies*

$$\mathbb{E} \sup_{0 \leqslant \tau \leqslant T} \|v(\tau; v_0)\|^{2q_0} \leqslant C_{q_0}(|v_0|, T) < +\infty, \tag{8.14}$$

where $C_{q_0}(\cdot, \cdot)$ is the same as in Assumption 6.1.

ii) For any $v_0 \in \mathbb{R}^{2n}$, solution $v^\varepsilon(\tau; v_0)$ of equation (6.1) with $v^\varepsilon(0; v_0) = v_0$ satisfies

$$\mathcal{D}\big(I(v^\varepsilon(\cdot; v_0))\big) \rightharpoonup \mathcal{D}\big(I(v(\cdot; v_0))\big) \;\; in \;\; \mathcal{P}(C[0, T], \mathbb{R}^n_+) \;\; as \; \varepsilon \to 0. \tag{8.15}$$

Moreover, the process $I^0(\tau) := I(v(\cdot; v_0))$, $\tau \in [0, T]$ is a weak solution of (6.12), equal $I_0 = I(v_0)$ at $\tau = 0$.

Remark 8.7. A straightforward analysis of the proof shows that it goes through without changes if $v^\varepsilon(\tau; v_{\varepsilon 0})$ solves (6.1) with an initial data $v_{\varepsilon 0}$ which converges to v_0 as $\varepsilon \to 0$. In this case still

$$\mathcal{D}\big(I(v^\varepsilon(\cdot; v_{\varepsilon 0}))\big) \rightharpoonup \mathcal{D}\big(I(v(\cdot; v_0))\big) \;\; \text{in} \;\; \mathcal{P}(C[0, T], \mathbb{R}^n_+) \;\; \text{as } \varepsilon \to 0. \tag{8.16}$$

The result in Theorem 8.6 admits an immediate generalization to the case when the initial data v_0 in (8.15) is a random variable:

Amplification 8.8. Let v_0 be a random variable independent of the Wiener processes $\boldsymbol{\beta}_j(\tau)$, $j = 1, \dots, n_1$. Then still convergence (8.15) holds.

Proof. Let v^ε be a weak solution of (6.1) with $v^\varepsilon(0) = v_0$. Let $(\Omega', \mathcal{F}', \mathbf{P}')$ be some probability space and $\xi_0^{\omega'}$ be a r.v. on Ω', distributed as v_0. Then $v^{\varepsilon\omega}(\tau; \xi_0^{\omega'})$ is a weak solution of (6.1), defined on the probability space $\Omega' \times \Omega = \{(\omega', \omega)\}$. Take f to be a bounded continuous function on $C([0, T], \mathbb{R}^n_+)$. Then by the theorem above, for each $\omega' \in \Omega'$

$$\lim_{\varepsilon \to 0} \mathbf{E}^\Omega f\Big(I\big(v^{\varepsilon\omega}(\cdot; \xi_0^{\omega'})\big)\Big) = \mathbf{E}^\Omega f\Big(I\big(v^{\omega}(\cdot; \xi_0^{\omega'})\big)\Big),$$

where $v^\omega(\cdot; \xi_0^{\omega'})\big)$ is a weak solution of (7.3) with $v^\omega(0) = \xi_0^{\omega'}$. Since f is bounded, then by the Lebesgue dominated convergence theorem we have

$$\begin{aligned}\lim_{\varepsilon \to 0} \mathbf{E} f(I(v^\varepsilon(\cdot; v_0))) &= \lim_{\varepsilon \to 0} \mathbf{E}^{\Omega'} \mathbf{E}^\Omega f(I(v^{\varepsilon\omega}(\cdot; \xi_0^{\omega'}))) \\ &= \mathbf{E}^{\Omega'} \mathbf{E}^\Omega f(I(v^{\omega}(\cdot; \xi_0^{\omega'}))) = \mathbf{E} f(I(v(\cdot; v_0))).\end{aligned}$$

This implies the required convergence (8.15). □

Proposition 8.9. *1) A weak solution v^θ, $\theta \in \mathbb{T}^n$, as in Theorem 8.2 may be chosen to be $v^\theta(\tau) = \Phi_\theta v(\tau; v_0)$, where $v(\tau; v_0)$ is the strong solution from Theorem 8.6.*

2) More generally, if $\tilde{\tau}$ is a non-negative constant and $\theta \in \mathbb{T}^n$ is a r.v., measurable with respect to $\mathcal{F}_{\tilde{\tau}}$ (see (6.2)), then the process $\hat{v}(\tau) = \Phi_\theta v(\tau; v_0)$, $\tau \geq \tilde{\tau}$, is a weak solution of equation (7.3).

Proof. It suffices to prove 2) since it implies 1) if we choose $\tilde{\tau} = 0$. Substituting in (7.3) $v(\tau) = \Phi_{-\theta} \hat{v}(\tau)$ we get that

$$d\hat{v}(\tau) = \Phi_\theta R\big(\Phi_{-\theta} \hat{v}(\tau)\big) d\tau + \Phi_\theta \langle\!\langle B \rangle\!\rangle \big(\Phi_{-\theta} \hat{v}(\tau)\big) d\beta(\tau).$$

Or, using Proposition 7.1.ii), that

$$d\hat{v}(\tau) = R(\hat{v}(\tau)) d\tau + \langle\!\langle B \rangle\!\rangle (\hat{v}(\tau)) \Phi_\theta d\beta(\tau).$$

Since the r.v. θ is $\mathcal{F}_{\tilde{\tau}}$-measurable, then the process $t \mapsto \Phi_\theta\big(\beta(t + \tilde{\tau}) - \beta(\tilde{\tau})\big)$ is a standard Wiener process in $\mathbb{R}^{2n}$. Thus, $\hat{v}(\tau)$, $\tau \geq \tilde{\tau}$, is a weak solution of (7.3). □

9. Equations in bounded domains

In this section we consider problem (6.1) in the set

$$\mathfrak{B} = \{v \in \mathbb{R}^{2n} : I(v) \in B\},$$

where $B = B_R(\mathbb{R}^n)$ for some $R > 0$. We assume that all coefficients in (6.1) are defined and Lipschitz continuous on the set $\bar{\mathfrak{B}}$ (so $W(I)$ is defined and Lipschitz continuous on $\bar{B}$). We also assume that conditions of items (1)–(2) of Assumption 6.1 are fulfilled on $\bar{\mathfrak{B}}$. Let us consider the effective equation (7.3) as in Section 7 on set $\bar{\mathfrak{B}}$ (note that to calculate coefficients of the equation on $\bar{\mathfrak{B}}$ it suffices to know P, B and W only where they are defined).

Let $v^\varepsilon(\tau)$ be a solution of (6.1) with the v_0 as above. Denote by τ_R^ε its exit time from domain $\mathfrak{B}$. Then $\tau_R^\varepsilon = \inf\{\tau > 0 : I^\varepsilon(v(\tau)) \in \partial B\}$, where $I^\varepsilon(\tau) := I^\varepsilon(v(\tau))$ satisfies (6.10). Similarly, let τ_R^{h} stands for the exit time from $\mathfrak{B}$ of a solution $v(\tau)$ of equation (7.3), equal v_0 at $\tau = 0$. Again, it equals the exit time of $I(v(\tau))$ from B.

Theorem 9.1. *Under the above assumptions, for any $T > 0$ and any $v_0 \in \mathfrak{B}$, as $\varepsilon \to 0$ the family of processes $\{I^\varepsilon(\cdot \wedge \tau_R^\varepsilon)\}$ converges in law, weakly in the space $(C([0,T];\mathbb{R}^n), \mathcal{B}_T)$, to a weak solution $I^{\mathrm{h}}(\cdot \wedge \tau_R^{\mathrm{h}})$ of problem (6.12), (6.13). The latter solution is obtained as the action-vector for a unique weak solution of the effective equation (7.3), equal v_0 at $\tau = 0$ and stopped at $\partial\mathfrak{B}$.*

Proof. The required statement is essentially a consequence of Theorems 6.7 and 8.2. Indeed, by Lemma 5.2 in [12] coefficients $\mathbf{P}_k(v)$ and $B_{kj}(v)$ can be extended from the set $\bar{\mathfrak{B}}$ to the whole space $\mathbb{R}^{2n}$ in such a way that the extensions are bounded, Lipschitz continuous, and the extended matrix $B(v)B(v)^t$ is positive definite. Using the same lemma we also extend $W(I)$ to a Lipschitz continuous vector-function $\widetilde{W}$ on $\mathbb{R}^d$ with compact support.

Consider the function Γ on $\mathbb{R}^n$,

$$\Gamma(x) = \begin{cases} \exp\left(-\frac{1}{|x|-R}\right), & |x| > R, \\ 0, & |x| \le R. \end{cases}$$

It is smooth, globally Lipschitz, and is flat on ∂B. For $\alpha \in \mathbb{R}$ define vector-functions W_α on $\mathbb{R}^n$ as $W_\alpha(x) = \alpha\nabla\Gamma(x) + \widetilde{W}(x)$. All of them are continuations of W to $\mathbb{R}^n$.

Lemma 9.2. *There is at most countable number of α's for which the components of W_α are rationally dependent on a set of positive measure.*

Proof. Since $W_\alpha = W$ on $\bar{B}_R$, then in view of Assumption 6.1.(1) it suffices to examine intersections of the sets of rational dependence of components of W_α with $\mathbb{R}^n \smallsetminus B$.

Assume that there exists two distinct α_1, α_2 and a non-zero vector $m \in \mathbb{Z}^n$ such that

$$\mathcal{L}\big(\{x \in \mathbb{R}^n \smallsetminus B : m \cdot W_{\alpha_1} = 0 \text{ and } m \cdot W_{\alpha_2} = 0\}\big) > 0.$$

Then $m \cdot \nabla\Gamma(x) = 0$ on a set of positive measure in $\mathbb{R}^n \smallsetminus B$. But $\Gamma(x)$ may be written as $f(|x|^2)$, where $f(r)$ is a smooth function, vanishing for $r \le R^2$. Then $\nabla\Gamma(x) = 2xf'(|x|^2)$ and we see that the set under discussion has zero measure since $f'(r) > 0$ for $r > R^2$. Therefore for each $\delta > 0$ and any non-zero vector $m \in \mathbb{Z}^n$ the number of α's for which $\mathcal{L}\big(\{x \in \mathbb{R}^d \smallsetminus B : m \cdot \nabla W_\alpha = 0\}\big) > \delta$ is at most countable. This implies the assertion. □

Denote $X_T = C([0,T], \mathbb{R}^n)$, take any number α_0, different from the countable family in the lemma above, and choose W_{α_0} for the extentson of W. Then by

Proposition 2.3 Assumption 6.1.(4) holds, and so Theorems 6.7, 8.2 and 8.6 apply to the obtained stochastic equation in $\mathbb{R}^{2n}$. Thus for any $T>0$ the corresponding process I^ε converges in law in X_T as $\varepsilon\to 0$ to a solution of the averaged equation (6.12) and may be lifted to a solution of the corresponding effective equation. The initial condition remains unchanged.

Let $\tau_R(I)=\min\big(T,\inf\{\tau\in(0,+\infty):I(\tau)\in\partial B\}\big)$. For an arbitrary bounded continuous functional $\mathcal{R}$ on X_T, consider there another functional

$$\mathcal{R}_B(I)=\mathcal{R}\big(I(\tau\wedge\tau_R(I))\big),\quad I\in X_T.$$

It is not continuous. However, the following statement ensures that it is almost surely continuous with respect to the measure on X_T, generated by the limit process $I^{\rm h}$, constructed in Theorem 8.6 (and called there I^0).

Lemma 9.3. *Under the standing assumptions, let $v(\tau)$ be any solution of the corresponding effective equation in $\mathbb{R}^{2n}$ and $I^h(\tau)=I(v(\tau))$. Then*

$$\mathbf{P}\big(\inf\{\tau>0:I^{\rm h}(\tau)\in\partial B\}=\inf\{\tau>0:I^{\rm h}(\tau)\notin\bar B\}\big)=1.$$

Proof. The desired inequality is an immediate consequence of the fact that the diffusion coefficient of the process $v(\tau)$ at ∂B does not degenerate in the direction of a normal vector to ∂B. □

Combining this with the statements of Theorems 6.7, 8.2 and [2, Theorem 5.2] we conclude that the law of $I^\varepsilon(\tau\wedge\tau_R^\varepsilon)$ converges to that of $I^{\rm h}(\tau\wedge\tau_R^{\rm h})$. □

It should be noted that the distribution of τ_R^ε does not concentrate in the vicinity of zero, as $\varepsilon\to 0$. So Theorem 9.1 with high probability describes the behaviour of solutions for (6.1) on t-time intervals of order ε^{-1}. More precisely, we have

Proposition 9.4. *Let $\|v_0\|<R-r$ with $0<r<R$. Then there exists a constant $C=C(R)$ such that, for all $T\in[0,1]$,*

$$\mathbf{P}\{\tau_R^\varepsilon<T\}\leqslant\frac{C(R)(T+\sqrt{T})}{r^2},\quad\forall\varepsilon\in(0,1].$$

Proof. As was explained in the beginning of the proof of Theorem 9.1, we may assume that the coefficients $\mathbf{P}_k(v)$ and $B_{kj}(v)$ are extended from the set $\bar{\mathfrak{B}}$ to the whole space $\mathbb{R}^{2n}$ in such a way that the extensions are bounded, uniformly Lipschitz continuous, and the extended matrix $B(v)B(v)^t$ is positive definite. The function $W(I)$ also admits an extension to $\mathbb{R}^d$, the extended function is bounded and uniformly Lipschitz continuous.

Then a solution of (6.1) is well defined for all $\tau>0$, and by the Itô formula we have

$$d\|v\|^2(\tau)=2\sum_{k=1}^n\mathbf{v}_k^t\mathbf{P}_k(v)d\tau+\sum_{k=1}^n\sum_{j=1}^{n_1}\|B_{kj}(v)\|_{HS}^2d\tau+2\sum_{k=1}^n\sum_{j=1}^{n_1}\mathbf{v}_k^tB_{kj}(v)d\boldsymbol{\beta}_j(\tau),\quad(9.1)$$

Therefore,

$$\begin{aligned}&\mathbf{E}\Big(\sup_{0\leqslant\tau\leqslant T}\big|\|v\|^2(\tau)-\|v_0\|^2\big|\Big)\\ \leqslant\ &\mathbf{E}\Big(\sup_{0\leqslant\tau\leqslant T}\Big|\textstyle\int_0^\tau\big[2\sum_{k=1}^n\mathbf{v}_k^t(s)\mathbf{P}_k(v(s))+\sum_{k=1}^n\sum_{j=1}^{n_1}\|B_{kj}(v)\|_{HS}^2(s)\big]ds\Big|\Big)\quad(9.2)\\ &+\mathbf{E}\Big(\sup_{0\leqslant\tau\leqslant T}\Big|\textstyle\int_0^\tau2\sum_{k=1}^n\sum_{j=1}^{n_1}\mathbf{v}_k^t(s)B_{kj}(v(s))d\boldsymbol{\beta}_j(s)\Big|\Big)\end{aligned}$$

According to (2.3),

$$\mathbf{E}\sup_{0\leqslant\tau\leqslant1}\|v\|^2(\tau)\le C_0(R)\qquad\forall\varepsilon\in(0,1].\quad(9.3)$$

Considering the boundedness of $\mathbf{P}_k$ and B_{kj} we conclude that

$$\mathbf{E}\Big(\sup_{0\leqslant\tau\leqslant T}\Big|\int_0^\tau\big[2\sum_{k=1}^n\mathbf{v}_k^t(s)\mathbf{P}_k(v(s))+\sum_{k=1}^n\sum_{j=1}^{n_1}\|B_{kj}(v)\|_{HS}^2(s)\big]ds\Big|\Big)\leqslant C_1(R)T \quad (9.4)$$

for all $T\in[0,1]$. The second term on the right-hand side of (9.2) can be estimated with the help of the BDG inequality as follows:

$$\mathbf{E}\Big(\sup_{0\leqslant\tau\leqslant T}\Big|\int_0^\tau 2\sum_{k=1}^n\sum_{j=1}^{n_1}\mathbf{v}_k^t(s)B_{kj}(v(s))d\boldsymbol{\beta}_j(s)\Big|\Big)\leqslant C\Big(\mathbf{E}\int_0^T\|v\|^2(s)ds\Big)^{\frac12}\leqslant C_2(R)\sqrt{T}$$

Combining the last inequality with (9.2) and (9.3) yields

$$\mathbf{E}\big(\sup_{0\leqslant\tau\leqslant T}\big|\|v\|^2(\tau)-\|v_0\|^2\big|\big)\leqslant C(R)(T+\sqrt{T}) \quad (9.5)$$

for all $T\in[0,1]$. Since $\{\omega\in\Omega:\tau_R\leqslant T\}\subset\{\omega:\sup_{0\leqslant\tau\leqslant T}\big|\|v\|^2(\tau)-\|v_0\|^2\big|\geqslant 2Rr-r^2\}$, from (9.5) by the Chebyshef inequality we obtain

$$\mathbf{P}\{\tau_R\leqslant T\}\leqslant C(R)\frac{T+\sqrt{T}}{2Rr-r^2}\leqslant C(R)\frac{T+\sqrt{T}}{r^2}.$$

□

10. Mixing and uniform convergence

In this section we establish the uniform in time convergence in distribution of the actions of solutions for equation (6.1) to those for solutions of effective equation (7.3), with respect to the dual-Lipschitz metric (see Definition 5.3) in the space of probability measures. The proof uses the approach developed in [11, 12], where a similar result was obtained in the easier case when the frequency vector W in equation (6.1) is constant (cf. Remark 6.2.1)).

Proposition 10.1. *Under the assumption of Amplification 8.8 let the r.v. v_0 be such that $\|v_0\|\le R$ a.s., for some $R>0$. Then the rate of convergence in (8.15) with respect to the dual-Lipschitz distance depends only on R.*

Proof. The proof of Amplification 8.8 shows that it suffices to verify that for a non-random initial vector $v_0\in\bar B_R(\mathbb{R}^{2n})$ the rate of convergence in (8.15) depends only on R. Assume the opposite. Then there exist a $\delta>0$, a sequence $\varepsilon_j\to0$ and vectors $v_j\in\bar B_R(\mathbb{R}^{2n})$ such that

$$\|\mathcal{D}\big(I(v^{\varepsilon_j}(\cdot;v_j))\big)-\mathcal{D}\big(I(v^0(\cdot;v_j))\big)\|^*_{L,C([0,T],\mathbb{R}^n_+)}\geqslant\delta. \quad (10.1)$$

By (6.4) and (8.14), using the same argument as in the proof of Lemma 3.1, we know that the two sets of probability measures $\{\mathcal{D}\big(I(v^{\varepsilon_j}(\cdot;v_j))\big)\}$ and $\{\mathcal{D}\big(v^0(\cdot;v_j)\big)\}$ are tight, respectively in $\mathcal{P}(C([0,T],\mathbb{R}^n_+))$ and $\mathcal{P}(C([0,T],\mathbb{R}^{2n}))$. Therefore, there exists a sequence $k_j\to\infty$ such that $\varepsilon_{k_j}\to0$, $v_{k_j}\to v_0$,

$$\mathcal{D}(I(v^{\varepsilon_{k_j}}(\cdot;v_{k_j})))\rightharpoonup Q_0^I\ \text{ in }\ \mathcal{P}(C([0,T],\mathbb{R}^n_+)),$$

and

$$\mathcal{D}(v^0(\cdot;v_{k_j}))\rightharpoonup Q_0^v\ \text{ in }\ \mathcal{P}(C([0,T],\mathbb{R}^{2n})).$$

Then due to (10.1),

$$\|Q_0^I-I\circ Q_0^v\|^*_{L,C([0,T],\mathbb{R}^n_+)}\geqslant\delta. \quad (10.2)$$

Since in the well-posed eq. (7.3) the drift and dispersion are locally Lipschitz and its solutions satisfy estimates (8.14), then the law $\mathcal{D}(v^0(\cdot;v'))$ is continuous with respect to the law of the initial condition v'. Therefore the limiting measure Q_0^v is the unique weak solution of the effective equation (7.3) with initial condition $v^0(0)=v_0$. By (8.16) the measure Q_0^I equals $I\circ Q_0^v$. This contradicts (10.2) and proves the assertion. □

In this section, we make the following assumption concerning system (6.1) and the corresponding effective equation (7.3).

Assumption 10.2. The first three items (1)-(3) of Assumption 6.1 hold, and

(4) For any $v_0 \in \mathbb{R}^{2n}$ a unique strong solution $v^\varepsilon(\tau; v_0)$ of (6.1) is such that for some $q_0 > (q \vee 4)$ we have

$$\mathbf{E} \sup_{T' \leqslant \tau \leqslant T'+1} \|v^\varepsilon(\tau; v_0)\|^{2q_0} \leqslant C_{q_0}(\|v_0\|), \tag{10.3}$$

for every $T' \geqslant 0$ and $\varepsilon \in (0, 1]$, where C_{q_0} is a continuous non-decreasing function.

(5) Effective equation (7.3) is mixing with a stationary measure μ^0 and a (strong) stationary solution $v^{st}(\tau)$, $\tau \geq 0$.

(6) For any its solution $v(\tau)$, $\tau \geq 0$, such that $\mathcal{D}(v(0)) =: \mu$ and $\langle \|v\|^{2q_0}, \mu(dz)\rangle = \mathbf{E}\|v(0)\|^{2q_0} \leq M'$ for some $M' > 0$ (we recall notation (5.5)) we have

$$\|\mathcal{D}(v(\tau)) - \mu^0\|^*_{L,\mathbb{R}^{2n}} \leqslant g_{M'}(\tau, d) \quad \forall \tau \geqslant 0, \text{ if } \|\mu - \mu^0\|^*_{L,\mathbb{R}^{2n}} \leqslant d \leqslant 2. \tag{10.4}$$

Here the function

$$g : \mathbb{R}_+ \times [0, 2] \times \mathbb{R}_+ \ni (\tau, d, M) \mapsto g_M(\tau, d),$$

is continuous, vanishes with d, converges to zero when $\tau \to \infty$ and is such that for each fixed $M \geq 0$ the function $(\tau, d) \mapsto g_M(\tau, d)$ is uniformly continuous in d, for $(\tau, d) \in [0, \infty) \times [0, 2]$. [11]

We emphasize that we assume the mixing only for the effective equation (7.3), but not for the original equation (6.1). Since Assumption 10.2 implies Assumption 6.1, then the assertions of Section 8 with any $T > 0$ hold for solutions of equation (6.1) which we analyse in this section.

Assumption (6) above may seem rather restrictive. But it is not, as shows the next result:

Proposition 10.3. *If we keep all the conditions in Assumption 10.2 except (6) and assume that for each $M > 0$ and any $v^1, v^2 \in \bar{B}_M(\mathbb{R}^{2n})$ we have*

$$\|\mathcal{D}(v(\tau; v^1)) - \mathcal{D}(v(\tau; v^2))\|^*_{L,\mathbb{R}^{2n}} \leq \mathfrak{g}_M(\tau), \tag{10.5}$$

where $\mathfrak{g}$ is a non-negative continuous function of $(M, \tau) \in \mathbb{R}^2_+$ which goes to zero when $\tau \to \infty$ and is non-decreasing in M, then (6) holds with a suitable function g.

For a proof of the proposition we refer the reader to [12, Section 7.1].

Note that (10.5) holds (with $\mathfrak{g}$ replaced by $2\mathfrak{g}$) if

$$\|\mathcal{D}v(\tau; v) - \mu^0\|^*_{L,\mathbb{R}^{2n}} \leq \mathfrak{g}_M(\tau) \quad \forall\, v \in \bar{B}_M(\mathbb{R}^{2n}).$$

Usually a proof of mixing for eq. (7.3) in fact establishes the estimate above. So, given assumptions (1)-(5), condition (6) is a rather mild restriction.

Theorem 10.4. *Under Assumption 10.2, for any $v_0 \in \mathbb{R}^{2n}$*

$$\lim_{\varepsilon \to 0} \sup_{\tau \geqslant 0} \|I \circ \mathcal{D}(v^\varepsilon(\tau; v_0)) - I \circ \mathcal{D}(v^0(\tau; v_0))\|^*_{L,\mathbb{R}^n_+} = 0,$$

where $v^\varepsilon(\tau; v_0)$ and $v^0(\tau; v_0)$ solve respectively (6.1) and (7.3) with the same initial condition v_0.

---

[11] So g_M extends to a continuous function on $[0, \infty] \times [0, 2]$ which vanishes when $\tau = \infty$ or $d = 0$.

Proof. Below we abbreviate $\|\cdot\|^*_{L,\mathbb{R}^n_+}$ to $\|\cdot\|^*_L$. Since v_0 is fixed, we also abbreviate $v^\varepsilon(\tau;v_0)$ to $v^\varepsilon(\tau)$. Due to (10.3)

$$\mathbf{E}\|v^\varepsilon(\tau)\|^{2q_0} \leqslant C_{q_0}(\|v_0\|) =: M^* \quad \forall\, \tau \geq 0. \tag{10.6}$$

By (10.6) and (8.15) also

$$\mathbf{E}\|v^0(\tau;v_0)\|^{2q_0} = \langle \|v\|^{2q_0}, \mathcal{D}(v^0(\tau;v_0))\rangle \leq M^* \quad \forall\, \tau \geq 0. \tag{10.7}$$

Since $\mathcal{D}(v^0(\tau;0)) \rightharpoonup \mu^0$ as $\tau \to \infty$, then from the estimate above with $v_0 = 0$ we get that

$$\langle \|v\|^{2q_0}, \mu^0\rangle \leq C_{q_0}(0) \leq M^*. \tag{10.8}$$

The constants in estimates below depend on M^*, but usually this dependence is not indicated. For any $T \geq 0$ we denote by $v^0_T(\tau)$, $\tau \geq 0$, a weak solution of effective equation (7.3) such that

$$\mathcal{D}(v^0_T(0)) = \mathcal{D}(v^\varepsilon(T)). \tag{10.9}$$

Thus $v^0_T(\tau)$ depends on ε, and $v^0_0(\tau) = v^0(\tau;v_0)$.

Below in this proof by $\kappa^j(\cdot)$, $j = 1,\dots,5$, we denote various monotonically increasing continuous functions $\mathbb{R}_+ \to \mathbb{R}_+$, vanishing at zero and positive outside it. These functions depend only on the constants in Assumption 10.2 and functions ν in (3.1), C in (10.3) and g in (10.4). A function κ^j, $j \geq 2$, may be constructed in terms of functions κ^l with $l < j$.

It is straightforward to see that the proof of Lemma 6.8 implies that the process $I^\varepsilon(\tau) = I(v^\varepsilon(\tau;v_0))$ satisfies the estimate below (recall that $[I] = \min_{1\leqslant j\leqslant n} I_j$)

$$\int_{T'}^{T'+1} \mathbf{P}\big(\{[I^\varepsilon(\tau)] < \gamma\}\big) \leqslant \kappa^1(\gamma), \qquad \forall T' \geqslant 0, \quad \forall \gamma, \varepsilon \in (0,1], \tag{10.10}$$

for some fixed function $\kappa^1(\cdot)$, and that the action-vector of the stationary solution $v^{st}(\tau)$ of equation (7.3) also meet this estimate. Then

$$\mathbf{P}\big(\{[I(v^{st}(\tau))] < \gamma\}\big) < \kappa^1(\gamma), \ \forall \tau \geqslant 0, \ \forall \gamma \in (0,1]. \tag{10.11}$$

We will say that a moment of time $\tau \geqslant 0$ is $(\gamma,\varepsilon)-$*typical*, where $\varepsilon, \gamma \in (0,1]$, if $\mathbf{P}\big(\{[I^\varepsilon(\tau)] < \gamma\}\big) \leqslant \kappa^1(\gamma)$. In view of (10.10), we have

Lemma 10.5. *For each $\gamma, \varepsilon \in (0,1]$, every interval $[T', T'+1]$, $T' \geqslant 0$, contains an (γ,ε)-typical moment of time $\tau = \tau(T',\gamma,\varepsilon)$.*

We continue with two technical lemmas, needed to prove principal Lemma 10.8. By (10.6) and (10.8) and Chebyshev's inequality, for any $\tau \geq 0$ and $R > 0$,

$$\langle \mathcal{D}v^\varepsilon(\tau), B_R\rangle,\ \langle \mu^0, B_R\rangle \geq 1 - R^{-2q_0}M^* =: 1 - \kappa^2(R^{-1}). \tag{10.12}$$

Lemma 10.6. *For any $\varepsilon \in (0,1]$ and $\tau \geq 0$,*

$$\|\mathcal{D}I(v^\varepsilon(\tau)) - I \circ \mu^0\|^*_L \leq \kappa^3\big(\|\mathcal{D}v^\varepsilon(\tau) - \mu^0\|^*_{L,\mathbb{R}^{2n}}\big) \tag{10.13}$$

(see Definition 5.3 for the distance $\|\cdot\|^*_L$ and the norm $|\cdot|_L$, used below).

Proof. Let us abbreviate $\mathcal{D}v^\varepsilon(\tau) =: m$. For $R \geqslant 2$, consider a function $G_R(t)$ on $\mathbb{R}_+$ as in Figure 2.

Then $|G_R| \leqslant 1$ and $\operatorname{Lip} G_R \leqslant 1$. For any $f \in C_b(\mathbb{R}^n_+)$, $|f|_L \leqslant 1$ we have

$$\begin{aligned}\langle f, I\circ m\rangle - \langle f, I\circ\mu^0\rangle &= \langle (fG_R)\circ I, m\rangle - \langle (fG_R)\circ I, \mu^0\rangle \\ &\quad + \langle (f(1-G_R))\circ I, m\rangle - \langle (f(1-G_R))\circ I, \mu^0\rangle \\ &\leqslant 2R\|m - \mu^0\|^*_{L,\mathbb{R}^{2n}} + 2\kappa^2\big(\tfrac{1}{R-1}\big),\end{aligned}$$

since on the ball B_R we have $|I| \leqslant R^2$ and $\operatorname{Lip} I \leqslant 2R$. Minimizing the r.h.s in $R \geqslant 2$, we get (10.13). ∎

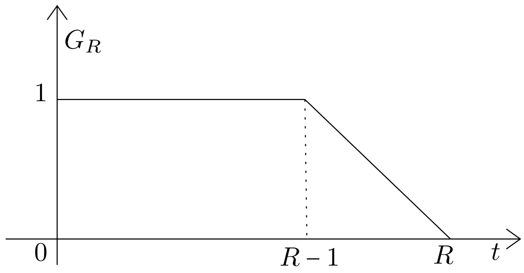

FIGURE 2.

In the lemma below Ω is the original probability space and Ω^1 is the segment $[0,1]$ with the Borel sigma-algebra $\mathcal{F}^1$ and the Lebesgue measure (i.e., another probability space, independent from Ω).

Lemma 10.7. *Let $\varepsilon\in(0,1]$. For any $\bar T\ge 0$ consider solution $v^\varepsilon(\tau)$ on the interval $J=[\bar T,\bar T+1]$ and denote*

$$\delta:=\sup_{\tau\in J}\|\mathcal{D}(I(v^\varepsilon(\tau))-I\circ\mu^0\|_L^*.$$

Then there exists a function[12] *$\kappa^5(\delta)$ and for each $\delta>0$ exists $r=r(\bar T,\delta,\varepsilon)\in J$ and a r.v. $\theta=\theta(\bar T,\delta,\varepsilon)\in\mathbb{T}^n$, such that*

$$\hat v^\varepsilon(r):=\Phi_\theta\tilde v^\varepsilon(r),\qquad \mathcal{D}\tilde v^\varepsilon(r)=\mathcal{D}v^\varepsilon(r)$$

satisfies

$$\|\mathcal{D}\hat v^\varepsilon(r)-\mu^0\|^*_{L,\mathbb{R}^{2n}}\le\kappa^5(\delta). \tag{10.14}$$

The random vector $\tilde v^\varepsilon(r)$ and the random variable $\theta\in\mathbb{T}^n$ are defined on the extended probability space $\Omega\times\Omega_1$ and are measurable with respect to sigma-algebra $\mathcal{F}_r\times\mathcal{F}^1$, where $\mathcal{F}_r$ is the sigma-algebra of the past (see (6.2)) and $\mathcal{F}^1$ is the sigma-algebra on the probability space Ω^1.

Proof. For some $\tau\in J$ let us denote $I(v^\varepsilon(\tau))=I^\varepsilon_\tau$, $I(v^{st}(\tau))=I^{st}_\tau$ and abbreviate

$$v^\varepsilon(\tau)=:v^\varepsilon,\ \ v^{st}(\tau)=v^{st},\ \ \mu^\varepsilon=\mathcal{D}(v^\varepsilon).$$

By Lemma 10.5 and (10.11) for any $\gamma\in(0,1]$ there exists $\tau=\tau(\bar T,\gamma,\varepsilon)\in J$ such that

$$\mathbf{P}([I^\varepsilon_\tau]<\gamma),\ \mathbf{P}([I^{st}_\tau]<\gamma)\le\kappa^1(\gamma). \tag{10.15}$$

On $\mathcal{P}(\mathbb{R}^n_+)$ consider the Kantorovich distance

$$\|\mu-\nu\|_K=\sup_{Lip(f)\le 1}\big(\langle f,\mu\rangle-\langle f,\nu\rangle\big)\le\infty,\quad \mu,\nu\in\mathcal{P}(\mathbb{R}^n_+).$$

Clearly $\|\mu-\nu\|_K\ge\|\mu-\nu\|^*_{L,\mathbb{R}^n_+}$, and due to estimates (10.6) and (10.8) also

$$\|\mathcal{D}I^\varepsilon_\tau-\mathcal{D}I^{st}_\tau\|_K\le\kappa^4(\|\mathcal{D}I^\varepsilon_\tau-\mathcal{D}I^{st}_\tau\|^*_{L,\mathbb{R}^n_+})\le\kappa^4(\delta), \tag{10.16}$$

for some function κ^4. See [4, Section 11.4] and [30, Chapter 7]. By the Kantorovich–Rubinstein theorem (see [5, 30, 4]) there exist r.v.'s $\tilde I^\varepsilon_\tau,\tilde I^{st}_\tau$ such that $\mathcal{D}\tilde I^\varepsilon_\tau=\mathcal{D}I^\varepsilon_\tau$, $\mathcal{D}\tilde I^{st}_\tau=\mathcal{D}I^{st}_\tau$ and

$$\mathbf{E}|\tilde I^\varepsilon_\tau-\tilde I^{st}_\tau|=\|\mathcal{D}I^\varepsilon_\tau-\mathcal{D}I^{st}_\tau\|_K\le\kappa^4(\delta). \tag{10.17}$$

Now consider the operator of projecting vectors to their actions:

$$I:\mathbb{R}^{2n}\to\mathbb{R}^n_+,\quad v\mapsto I(v).$$

[12] It depends on functions κ^1 and κ^4. The latter appears below in the lemma's proof and depends only on the constants in Assumption 10.2.

Then $I\circ\mu^\varepsilon = \mathcal{D}(I^\varepsilon_\tau) =: \nu^\varepsilon$. On the extended probability space $\Omega\times\Omega^1 = \{(\omega,\omega_1)\}$ exists a r.v. $\tilde v^\varepsilon(\omega,\omega_1)\in\mathbb{R}^{2n}$ such that

$$\mathcal{D}\tilde v^\varepsilon = \mu^\varepsilon,\quad I\big(\tilde v^\varepsilon(\omega,\omega_1)\big) = \tilde I^\varepsilon_\tau(\omega),\ \ a.s.$$

Indeed, by the regular conditional probability theorem (see [5, Sec. 10.2], [15, Sec. 5.3.C]) the measure μ^ε may be disintegrated as

$$\mu^\varepsilon(dv) = \mu_I(dv)\nu^\varepsilon(dI),$$

where μ_I is a measure on $\mathbb{R}^{2n}$ which is a measurable function of $I\in\mathbb{R}^n_+$ and is such that $\mu_{I'}(I^{-1}(I')) = 1$ for every $I'\in\mathbb{R}^n_+$. Then there exists a measurable mapping $\eta:\mathbb{R}^n_+\times\Omega_1\to\mathbb{R}^{2n}$, $(I,\omega_1)\mapsto\eta^{\omega_1}(I)$, such that $\mu_I\equiv\mathcal{D}\eta^{\cdot}(I)$, see [21, Theorem 1.2.28]. Now it remains to set $\tilde v^\varepsilon(\omega,\omega_1) = \eta^{\omega_1}(\tilde I^\varepsilon_\tau(\omega))$. Indeed, for any function $f\in C_b(\mathbb{R}^{2n})$ we have

$$\mathbf{E}^\omega\mathbf{E}^{\omega_1} f(\eta^{\omega_1}(\tilde I^\varepsilon_\tau(\omega))) = \int\nu^\varepsilon(dI)\Big(\int f(v)\mu_I(dv)\Big).$$

Similar, on $\Omega\times\Omega^1$ exists a r.v. $\tilde v^{st}$ such that

$$\mathcal{D}\tilde v^{st} = \mu^0,\quad I\big(\tilde v^{st}(\omega,\omega_1)\big) = \tilde I^{st}_\tau(\omega),\ \ a.s.$$

Let us denote $\tilde\varphi^\varepsilon_\tau = \varphi(\tilde v^\varepsilon)$ and $\tilde\varphi^{st}_\tau = \varphi(\tilde v^{st})$. Then $\tilde v^\varepsilon = V_{\tilde\varphi^\varepsilon_\tau}(\tilde I^\varepsilon_\tau)$ and $\tilde v^{st} = V_{\tilde\varphi^{st}_\tau}(\tilde I^{st}_\tau)$. Let us set

$$\hat v^\varepsilon = V_{\tilde\varphi^{st}_\tau}(\tilde I^\varepsilon) = \Phi_{\tilde\varphi^{st}_\tau-\tilde\varphi^\varepsilon_\tau}(\tilde v^\varepsilon).$$

Then $\varphi(\hat v^\varepsilon) = \varphi(\tilde v^{st})$, so

$$\|\hat v^\varepsilon - \tilde v^{st}\| \le \tfrac{1}{2\sqrt{\gamma}}|\tilde I^\varepsilon_\tau - \tilde I^{st}_\tau|\quad\text{if}\quad [\tilde I^\varepsilon_\tau],\ [\tilde I^{st}_\tau]\ge\gamma. \tag{10.18}$$

Now we estimate the distance between $\mathcal{D}(\hat v^\varepsilon)$ and $\mathcal{D}\tilde v^{st} = \mu^0$. To do that we take any function $f\in C_b(\mathbb{R}^{2n})$, $|f|_L\le 1$, and consider

$$X := \mathbf{E}\big(f(\hat v^\varepsilon) - f(\tilde v^{st})\big).$$

Introducing the events

$$Q^1_\gamma = \{[I(\hat v^\varepsilon)]<\gamma\},\ Q^2_\gamma = \{[I(\tilde v^{st})]<\gamma\},\quad Q_\gamma = Q^1_\gamma\cup Q^2_\gamma$$

we write X as

$$X = X_1 + X_2,\quad X_1 = \mathbf{E}\big(f(\hat v^\varepsilon) - f(v^{st})\big)\mathbf{1}_{Q_\gamma},\ \ X_2 = \mathbf{E}\big(f(\hat v^\varepsilon) - f(v^{st})\big)\mathbf{1}_{Q^c_\gamma}.$$

Since in view of (10.15)

$$\mathbf{P}(Q^1_\gamma),\ \mathbf{P}(Q^2_\gamma)\le 2\kappa^1(\gamma) \tag{10.19}$$

and $|f|\le 1$, then $X_1\le 4\kappa^1(\gamma)$. As Lip$f\le 1$, then using (10.18) and (10.17) we get that

$$X_2\le\mathbf{E}\big|\hat v^\varepsilon - \tilde v^{st}\big|\mathbf{1}_{Q^c_\gamma}\le\tfrac{1}{2\sqrt{\gamma}}\mathbf{E}|\tilde I^\varepsilon_\tau - \tilde I^{st}_\tau|\le\tfrac{\kappa^4(\delta)}{2\sqrt{\gamma}}.$$

Re-denoting $\hat v^\varepsilon$ to $\hat v^\varepsilon(\tau)$ and $\tilde v^{st}$ to $\tilde v^{st}(\tau)$ we get from the estimates on X_1, X_2 that

$$\mathbf{E}\big(f(\hat v^\varepsilon(\tau)) - f(v^{st}(\tau))\big)\leqslant 8\kappa^1(\gamma) + \tfrac{1}{2\sqrt{\gamma}}\kappa^4(\delta).$$

Minimizing in $\gamma\in(0,1]$ we achieve that the r.h.s. is less then $\kappa^5(\delta)$ for some $\gamma(\bar T,\delta,\varepsilon)$, that is for some $\tau(\gamma(\bar T,\delta,\varepsilon),\varepsilon) =: r(\bar T,\delta,\varepsilon)$. Since f is any continuous function with $|f|_L\le 1$, then

$$\|\mathcal{D}\hat v^\varepsilon(r(\bar T,\delta,\varepsilon)) - \mathcal{D}v^{st}(r(\bar T,\delta,\varepsilon))\|^*_{L,\mathbb{R}^{2n}}\le\kappa^5(\delta).$$

This relation and the formula for $\hat v^\varepsilon(\tau)$ prove the lemma. □

Now we state and prove a key lemma for the proof of the theorem. Below functions κ^3 and κ^5 are as in Lemmas 10.6 and 10.7, and we recall notation (10.9).

Lemma 10.8. (1) *For any $T>0$ and $\delta>0$ there exists $\varepsilon_1=\varepsilon_1(\delta,T)>0$ such that if $\varepsilon\le\varepsilon_1$, then*

$$\sup_{\tau\in[0,T]}\|I\circ\mathcal{D}(v^\varepsilon(\bar T+\tau))-I\circ\mathcal{D}(v^0_{\bar T}(\tau))\|^*_{L,\mathbb{R}^n_+}\leqslant\delta/2\quad\forall\,\bar T\geqslant 0. \tag{10.20}$$

(2) *For any $\delta>0$, choose a function $T^*=T^*(\delta)\ge 0$ such that $\kappa^3(g_{M^*}(T,2))\leqslant\delta/2$ for any $T\geqslant T^*(\delta)$. Then there exists $\varepsilon_2=\varepsilon_2(\delta)\in(0,1]$ with the following property:*
assume that a non-random $T'=T'(\delta,\varepsilon)\ge 0$ is such that for every $0<\varepsilon\le\varepsilon_2$ it holds that

$$\|I\circ\mathcal{D}(v^\varepsilon(T'))-I\circ\mu^0\|^*_L\leqslant\delta \tag{10.21}$$

and

$$\|\mathcal{D}\hat v^\varepsilon-\mu^0\|^*_{L,\mathbb{R}^{2n}}\le\kappa^5(\delta) \tag{10.22}$$

for $\hat v^\varepsilon=\hat v^\varepsilon(T')=\Phi_\theta(v^\varepsilon(T'))$, where θ is some r.v. (depending on δ and ε), measurable with respect to $\mathcal{F}_{T'}$. Then for $0<\varepsilon\le\varepsilon_2$ we have

$$\sup_{\theta\in[0,1]}\|I\circ\mathcal{D}(v^\varepsilon(T'+T^*+\theta))-I\circ\mu^0\|^*_L\leqslant\delta, \tag{10.23}$$

and

$$\sup_{\tau\in[T',T'+T^*+1]}\|I\circ\mathcal{D}(v^\varepsilon(\tau))-I\circ\mu^0\|^*_L\leqslant\frac{\delta}{2}+\kappa^3\Big(\max_{\theta\in[0,T^*+1]}g_{M^*}(\theta,\kappa^5(\delta))\Big). \tag{10.24}$$

Proof. For a measure $\nu\in\mathcal{P}(\mathbb{R}^{2n})$ we denote by $v^\varepsilon(\tau;\nu)$ a weak solution of eq. (6.1) such that $\mathcal{D}(v^\varepsilon(0))=\nu$, and define $v^0(\tau;\nu)$ similarly. Since eq. (6.1) defines a Markov process in $\mathbb{R}^{2n}$ (e.g. see [15, Section 5.4.C] and [17, Section 3.3]), then

$$I\circ\mathcal{D}(v^\varepsilon(\tau;\nu))=\int_{\mathbb{R}^{2n}}I\circ\mathcal{D}(v^\varepsilon(\tau;v))\,\nu(dv),$$

and a similar relation holds for $I\circ\mathcal{D}(v^0(\tau;\nu))$.

(1) Denote $\nu^\varepsilon=\mathcal{D}(v^\varepsilon(\bar T))$. Then

$$\mathcal{D}(v^\varepsilon(\bar T+\tau))=\mathcal{D}(v^\varepsilon(\tau;\nu^\varepsilon)),\quad \mathcal{D}(v^0_{\bar T}(\tau))=\mathcal{D}(v^0(\tau;\nu^\varepsilon)). \tag{10.25}$$

By (10.12), for any $\delta>0$ there exists $K_\delta>0$ such that for each ε, $\nu^\varepsilon(\mathbb{R}^{2n}\smallsetminus\bar B_{K_\delta})\le\delta/8$, where $\bar B_{K_\delta}:=\bar B_{K_\delta}(\mathbb{R}^{2n})$. So

$$\nu^\varepsilon=A^\varepsilon\nu^\varepsilon_\delta+\bar A^\varepsilon\bar\nu^\varepsilon_\delta,\quad A^\varepsilon=\nu^\varepsilon(\bar B_{K_\delta}),\ \bar A^\varepsilon=\nu^\varepsilon(\mathbb{R}^{2n}\smallsetminus\bar B_{K_\delta}),$$

where ν^ε_δ and $\bar\nu^\varepsilon_\delta$ are the conditional probabilities $\nu^\varepsilon(\cdot\mid\bar B_{K_\delta})$ and $\nu^\varepsilon(\cdot\mid\mathbb{R}^{2n}\smallsetminus\bar B_{K_\delta})$. Accordingly,

$$\mathcal{D}(v^\kappa(\tau;\nu^\varepsilon))=A^\varepsilon\mathcal{D}(v^\kappa(\tau;\nu^\varepsilon_\delta))+\bar A^\varepsilon\mathcal{D}(v^\kappa(\tau;\bar\nu^\varepsilon_\delta)),$$

where $\kappa=\varepsilon$ or $\kappa=0$. Therefore,

$$\begin{aligned}&\|I\circ\mathcal{D}(v^\varepsilon(\tau;\nu^\varepsilon))-I\circ\mathcal{D}(v^0(\tau;\nu^\varepsilon))\|^*_L\\&\le A^\varepsilon\|I\circ\mathcal{D}(v^\varepsilon(\tau;\nu^\varepsilon_\delta))-I\circ\mathcal{D}(v^0(\tau;\nu^\varepsilon_\delta))\|^*_L+\bar A^\varepsilon\|I\circ\mathcal{D}(v^\varepsilon(\tau;\bar\nu^\varepsilon_\delta))-I\circ\mathcal{D}(v^0(\tau;\bar\nu^\varepsilon_\delta))\|^*_L.\end{aligned}$$

The second term on the r.h.s obviously is bounded by $2\bar A^\varepsilon\leqslant\frac{\delta}{4}$. While by Proposition 10.1, there exists $\varepsilon_1>0$, depending only on K_δ and T, such that for $0\le\tau\le T$ and $\varepsilon\in(0,\varepsilon_1]$ the first term in the r.h.s. is $\leqslant\frac{\delta}{4}$. Due to (10.25) this proves the first assertion.

(2) Let us choose $\varepsilon_2(\delta) := \varepsilon_1(T^*(\delta)+1,\delta)$. We have

$$\begin{aligned}&\sup_{\tau\in[0,1]}\|I\circ\mathcal{D}(v^\varepsilon(T'+T^*+\tau))-I\circ\mu^0\|_L^*\\&\leqslant\sup_{\tau\in[0,1]}\|I\circ\mathcal{D}(v^\varepsilon(T'+T^*+\tau)-I\circ\mathcal{D}(v^0_{T'}(T^*+\tau))\|_L^*\\&+\sup_{\tau\in[0,1]}\|I\circ\mathcal{D}(v^0_{T'}(T^*+\tau))-I\circ\mu^0\|_L^*.\end{aligned}\tag{10.26}$$

By (10.20) and the choice of ε_2, the first term in the r.h.s is less than $\frac{\delta}{2}$. Let us examine the second one. By Proposition 8.9,

$$\mathcal{D}\big(I(v^0(T^*+\tau;\hat v^\varepsilon))\big)=\mathcal{D}\big(I(v^0_{T'}(T^*+\tau))\big),\quad\forall\tau\in[0,1].$$

Thus the second term in the r.h.s of (10.26) equals $\sup_{\tau\in[0,1]}\|I\circ\mathcal{D}(v^0(T^*+\tau;\hat v^\varepsilon))-I\circ\mu^0\|_L^*$. Since $\|\hat v^\varepsilon\|=\|v^\varepsilon(T')\|$, then by (10.21) and (10.4)

$$\|\mathcal{D}(v^0(T^*+\tau;\hat v^\varepsilon))-\mu^0\|^*_{L,\mathbb{R}^{2n}}\leqslant g_{M^*}(T^*+\tau,\kappa^5(\delta)).$$

So in view of Lemma 10.6 the second term in the r.h.s of (10.26) is bounded by $\sup_{\tau\in[0,1]}\kappa^3(g_{M^*}(T^*+\theta,\kappa^5(\delta)))$, which is $\leqslant\delta/2$ due to the definition of $T^*(\delta)$. This proves (10.23).

Similarly,

$$\begin{aligned}\sup_{\tau\in[T',T'+T^*+1]}&\|I\circ\mathcal{D}(v^\varepsilon(\tau))-I\circ\mu^0\|_L^*\\&\leqslant\sup_{\theta\in[0,T^*+1]}\|I\circ\mathcal{D}(v^\varepsilon(T'+\theta))-I\circ\mathcal{D}(v^0(\theta;\hat v^\varepsilon))\|_L^*\\&+\sup_{\theta\in[0,T^*+1]}\|I\circ\mathcal{D}(v^0(\theta;\hat v^\varepsilon))-I\circ\mu^0\|_L^*\end{aligned}$$

By (10.20) and the definition of ε_2, the first term in the r.h.s is less than $\delta/2$, while by (10.4) and Lemma (10.6), the second term is bounded by $\kappa^3(\lambda)$, where $\lambda=\max_{\theta\in[0,T^*+1]}g_{M^*}(\theta,\kappa^5(\delta))$. Thus we proved (10.24). ☐

Now we are ready to prove the theorem. Let us fix arbitrary $\delta>0$ and take some $0<\delta_1\le\delta/4$. Below in the proof the functions ε_1, ε_2 and T^* are as in Lemma 10.8. We will abbreviate $T^*(\delta_1)=:T^*$, $\varepsilon_2(\delta_1)=:\varepsilon_2$ and will always assume that

$$0<\varepsilon\le\varepsilon_2.$$

i) By the definition of T^*, (10.4) and (10.6),

$$\begin{aligned}\|\mathcal{D}\big(v^0_{\bar T}(\tau)\big)-\mu^0\|^*_{L,\mathbb{R}^{2n}}\le g_{M^*}(\tau,2)\quad\forall\,\bar T\ge0,\\ \kappa^3(g_{M^*}(\tau,2))\le\delta_1/2\quad\forall\,\tau\ge T^*.\end{aligned}\tag{10.27}$$

ii) By (10.20)

$$\sup_{0\le\tau\le T^*+1}\|I\circ\mathcal{D}\big(v^\varepsilon(\tau)\big)-I\circ\mathcal{D}\big(v^0(\tau;v_0)\big)\|_L^*\le\tfrac{\delta_1}{2}.\tag{10.28}$$

From (10.27), (10.28) and Lemma 10.6 we have

$$\sup_{\tau\in[0,1]}\|I\circ\mathcal{D}\big(v^\varepsilon(T^*+\tau)\big)-I\circ\mu^0\|_L^*<\delta_1.\tag{10.29}$$

iii) By (10.28) and Lemma 10.7 there exists $T_1'=T_1'(T^*,\delta_1,\varepsilon)\in[T^*,T^*+1]$ such that apart from (10.29), there exists a r.v. $\theta_1\in\mathbb{T}^n$, measurable with respect to $\mathcal{F}_{T_1'}$, and such that $\hat v_1^\varepsilon(T_1')=\Phi_{\theta_1}v^\varepsilon(T_1')$ satisfies

$$\|\mathcal{D}\hat v_1^\varepsilon(T_1')-\mu^0\|^*_{L,\mathbb{R}^{2n}}\le\kappa^5(\delta_1),\tag{10.30}$$

Considering $v^\varepsilon(T_1'+\tau)$, $\tau\ge0$, we get that (10.29) holds with T^* replaced by $T_1'+T^*$, and an analogy of (10.30) holds for T_1' replaced with some $T_2'\in[T_1'+T^*,T_1'+T^*+1]$.

Iterating this argument we construct a sequence T'_N, $N = 1, \dots,$ such that $T'_{N+1} \in [T'_N + T_*, T'_N + T_* + 1]$,

$$\|I \circ \mathcal{D}\big(v^\varepsilon(T'_N)\big) - I \circ \mu^0\|^*_L \le \delta_1 \quad \forall\, N,$$

and

$$\|\mathcal{D}\hat{v}^\varepsilon_N(T'_N) - \mu^0\|^*_{L,\mathbb{R}^{2n}} \le \kappa^5(\delta_1) \quad \forall\, N,$$

for $\hat{v}^\varepsilon_N(T'_N) = \Phi_{\theta_N} v^\varepsilon(T'_N)$ with a suitable $\mathcal{F}_{T'_N}$-measurable θ_N. So by (10.24)

$$\sup_{\tau \in [T'_N, T'_{N+1}]} \|I \circ \mathcal{D}\big(v^\varepsilon(\tau)\big) - I \circ \mu^0\|^*_L \le \frac{\delta_1}{2} + \kappa^3\Big(\max_{\theta \in [0, T^*+1]} g_{M^*}\big(\theta, \kappa^5(\delta_1)\big)\Big), \quad (10.31)$$

for every N.

iv) Finally, by (10.28) if $\tau \le T'_1$ and by (10.27) with $\bar{T} = 0$ jointly with (10.31) if $\tau \ge T'_1$, we have that

$$\|I \circ \mathcal{D}\big(v^\varepsilon(\tau)\big) - I \circ \mathcal{D}\big(v^0(\tau; v_0)\big)\|^*_L \le \delta_1 + \kappa^3\Big(\max_{\theta \in [0, T^*+1]} g_{M^*}\big(\theta, \kappa^5(\delta_1)\big)\Big), \qquad \forall\, \tau \ge 0.$$

By the assumption, imposed in (6) of Assumption 10.2 on function g_M, $g_M(t,d)$ is uniformly continuous in d and vanishes at $d = 0$. Recall that κ^3 and κ^5 are both monotonically increasing continuous functions, vanishing at 0. Therefore we have that there exists $\delta^* > 0$, which we may assume to be $\leqslant \delta/2$, such that if $\delta_1 \le \delta^*$, then $\kappa^3(g_{M^*}(\theta, \kappa^5(\delta_1))) \leqslant \delta/2$ for all $\theta \geqslant 0$. Then by the estimate above,

$$\|I \circ \mathcal{D}\big(v^\varepsilon(\tau)\big) - I \circ \mathcal{D}\big(v^0(\tau; v_0)\big)\|^*_L \le \delta, \quad \forall \tau \geqslant 0 \quad \text{if} \quad \varepsilon \le \varepsilon_2\big(\delta^*(\delta)\big) > 0,$$

for every positive δ. This proves the theorem's assertion. □

We end this section with a sufficient condition for the validity of (4)-(6) in Assumption 10.2.

Proposition 10.9. *Assume that (2)-(3) of Assumption 6.1 hold true and the field $P(v)$ in (6.1) is coercive. That is, there exist $\alpha_1 > 0$ and $\alpha_2 \geqslant 0$ such that*

$$(P(v), v) \leqslant -\alpha_1\|v\| + \alpha_2, \ \forall v \in \mathbb{R}^{2n}, \qquad (10.32)$$

where $(v, w) = \sum_{j=1}^n \mathbf{v}_j \cdot \mathbf{w}_j$ is the inner product on $\mathbb{R}^{2n}$. Then (4)-(6) in Assumption 10.2 hold true.

Proof. For the drift term in (6.1) $b(v) = (\varepsilon^{-1} W_k(I)\mathbf{v}_k^\perp + \mathbf{P}_k(v), k = 1, \dots, n)$, by (10.32), we have

$$(b(v), v) = (P(v), v) \leqslant -\alpha_1\|v\| + \alpha_2, \ \forall v \in \mathbb{R}^{2n}.$$

For the drift term R in effective equation (7.3), by (10.32) and the definition or $R = \langle P \rangle$ we have

$$(R(v), v) = \sum_{k=1}^n \int_{\mathbb{T}^n} \mathbf{P}_k(\Phi_\theta v) \cdot \Phi_\theta^k \mathbf{v}_k d\theta = \int_{\mathbb{T}^n} (P(\Phi_\theta v), \Phi_\theta v) d\theta \leqslant -\alpha_1\|v\| + \alpha_2, \ \forall v \in \mathbb{R}^{2n}.$$

By the definition of the diffusion matrix of equation (7.3) we know that the uniform ellipticity condition as in (2) of Assumption 6.1 holds for the effective equation (7.3). Then the assertion of proposition directly follows from [12, Proposition 9.3]. □

Remark 10.10. The assumption in Proposition 10.9 also ensures the mixing in equation (6.1) for each $\varepsilon \in (0, 1]$, see in [12]. In this case, for corresponding stationary measures μ^ε, the measures $I \circ \mu^\varepsilon$ converge weakly to $I \circ \mu^0$ as $\varepsilon \to 0$.

10.1. **On the proof of Theorem 5.5.** Due to Proposition 10.3 the laws of solutions $I^0(\tau)$ for equation (3.3) obey estimates (10.4) (where $v(\tau)$ is replaced by $I^0(\tau)$). Now, denoting $v^\varepsilon(\tau) := (I^\varepsilon(\tau), \varphi^\varepsilon(\tau))$ we may repeat for $I(v^\varepsilon(\tau)) = I^\varepsilon(\tau)$ the proof of Theorem 10.4, given above in Section 10, with $I \circ v^0(\tau)$ replaced by $I^0(\tau)$. In fact, a proof of Theorem 5.5 is simpler than that of Theorem 10.4 since in the former theorem the mapping $v^\varepsilon \mapsto (I^\varepsilon, \varphi^\varepsilon)$ is a trivial isomorphism, while in the setting of the latter theorem this is the non-linear action-angle mapping (6.8), which is singular when some $\mathbf{v}_j^\varepsilon$ vanishes. Accordingly, to prove an analogy of the key Lemma 10.8 for solutions of equation (2.1) the technical argument, contained in Lemmas 10.5, 10.6 and 10.7 becomes redundant. We skip details of an exact realisation of this sketch.

10.2. **Damped/driven Hamiltonian systems.** As an example let us consider system (6.1), where the drift $P(v)$ is a damped Hamiltonian field and the diffusion is a diagonal additive random force:

$$d\mathbf{v}_k = \varepsilon^{-1} W_k(I)\mathbf{v}_k^\perp d\tau - \nu_k \mathbf{v}_k d\tau + J\nabla_{\mathbf{v}_k} h(v) d\tau + \gamma_k d\boldsymbol{\beta}_k(\tau), \quad k = 1, \dots, n, \tag{10.33}$$

$v(0) = v_0$. Here Hamiltonian $h(v)$ is C^2-smooth, $J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, ν_k's are nonnegative real numbers and γ_k's are nonzero real numbers. Results in Sections 6-10 apply to this system if Assumption 10.2 holds. There items (1)-(3) are easy to meet, while items (4)-(6) hold e.g. if (1)-(3) are satisfied, all ν_k's are positive and Hamiltonian $h(v)$ commutes with $\|v\|^2$. Indeed, then the scalar product $(J\nabla h(v), v)$ vanishes, so the drift in the equation is coercive, and items (4)-(6) hold in view of Proposition 10.9.

To write down the effective equation we note that its dispersion is the same as in eq. (10.33) since now matrix $X(v)$ as in Section 7 is $\operatorname{diag}\{\gamma_k^2\}$. The drift $R(v)$ is an averaging of the drift $-\{\nu_k \mathbf{v}_k\} + J\nabla h(v)$. It is easy to see that the averaging does not change the first term. The averaging of the second one is $J\nabla\langle h\rangle(v)$, where $\langle h\rangle$ is the averaged Hamiltonian, $\langle h\rangle(v) = \int_{\mathbb{T}^n} h(\Phi_\theta v) d\theta$ (see Proposition 3.5 in [12]). So the effective equation reads

$$d\mathbf{v}_k = -\nu_k \mathbf{v}_k d\tau + J\nabla_{\mathbf{v}_k}\langle h\rangle(v) d\tau + \gamma_k d\boldsymbol{\beta}_k(\tau), \quad k = 1, \dots, n. \tag{10.34}$$

Obviously the function $\langle h\rangle(v)$ depends only on the actions i.e. $\langle h\rangle = h^0(I_1, \dots, I_n)$, where h^0 is a continuous function, and $h^0(, \dots, I_j, \dots) = h^0(, \dots, -I_j, \dots)$ for each j. Since h is C^2-smooth, then the averaged Hamiltonian $\langle h\rangle$ is C^2-smooth in v as well. From here and Whitney's theorem (see in [31] Theorem 1 with $s = 1$ and the remark, concluding that paper) we derive that the function $h^0(I)$ is C^1-smooth. So in effective equation (10.34)

$$J\nabla_{\mathbf{v}_k}\langle h\ \rangle(v) = (\partial h^0(I)/\partial I_k)\mathbf{v}_k^\perp, \quad h^0 \in C^1. \tag{10.35}$$

In particular, applying Ito's formula to an action $I_k = \frac{1}{2}|\mathbf{v}_k|^2$, where $v(\tau)$ solves (10.34), we get that

$$\frac{d}{d\tau}\mathbf{E}\tfrac{1}{2}|\mathbf{v}_k|^2(\tau) = -2\nu_k \mathbf{E}\tfrac{1}{2}|\mathbf{v}_k|^2(\tau) + \gamma_k^2. \tag{10.36}$$

So the stationary measure μ^0 for (10.34) is such that $\int \frac{1}{2}|\mathbf{v}_k|^2 \mu^0(dv) = \gamma_k^2/(2\nu_k)$ for each k, and we get from Theorem 10.4 that for any v_0 solution $v^\varepsilon(\tau)$ of (10.33) satisfies

$$\lim_{\tau\to\infty, \varepsilon\to 0} \mathbf{E}\tfrac{1}{2}|\mathbf{v}_k^\varepsilon|^2(\tau) = \frac{\gamma_k^2}{2\nu_k}, \quad \forall k.$$

If all ν_k's vanish (but still γ_k's are nonzero and $h(v)$ commutes with $\|v\|^2$), then Assumption 10.2 does not hold since (10.3) fails. Nonetheless, Assumption 6.1 remains valid (which can be readily verified by the same argument as in the proof of Proposition 2.3), and the effective equation has the form (10.34), (10.35) with $\nu_k \equiv 0$. So we derive from (10.36) with $\nu_k = 0$ that on any finite time-interval $[0,T]$ the averaged actions of solutions $v^\varepsilon(\tau)$ for (10.33) approximately have linear growth with τ:

$$\mathbf{E}\tfrac{1}{2}|\mathbf{v}_k^\varepsilon|^2(\tau) = \tfrac{1}{2}|\mathbf{v}_{0k}|^2 + \gamma_k^2\tau + o_{\epsilon\to 0}(1), \quad 0 \le \tau \le T, \quad \forall\, k,$$

where the term $o_{\epsilon\to 0}(1) \to 0$ as $\epsilon \to 0$.

Appendix A. Proof of Lemma 6.8

We fix $\varepsilon \in (0,1]$ and do not indicate the dependence on it. Relation (6.18) is already established. A proof of (6.17) goes in 3 steps.

Step 1: Constructing for a fixed k and any $\delta \in (0,1]$ an Itô process $\bar{\mathbf{v}}_k^\delta(\tau)$, $\tau \in [0,T]$, such that $|\bar{\mathbf{v}}_k^\delta| \equiv |\bar{\mathbf{v}}_k|$, and if $\|\bar{\mathbf{v}}_k^\delta\| \geqslant \delta$ then no $\frac{1}{\varepsilon}$-term explicitly appear in the drift term.

Denote by $U = U(\zeta_1,\zeta_2) : (\mathbb{R}^2 \smallsetminus \{0\}) \times (\mathbb{R}^2 \smallsetminus \{0\}) \mapsto SO(2)$ the unique rotation of $\mathbb{R}^2$ that maps $\frac{\zeta_2}{|\zeta_2|}$ to $\frac{\zeta_1}{|\zeta_1|}$. Then $U(\zeta_2,\zeta_1) = (U(\zeta_1,\zeta_2))^{-1} = (U(\zeta_1,\zeta_2))^t$.

Let $v(\tau) = (\mathbf{v}_k(\tau), k = 1,\dots,n)$ be a solution of equation (6.1). We introduce the vector-functions

$$\bar{\mathbf{P}}_k(\bar{\mathbf{v}}_k, v) = U(\bar{\mathbf{v}}_k,\mathbf{v}_k)\mathbf{P}_k(v), \quad \bar{B}_{kj}(\bar{\mathbf{v}}_k, v) = U(\bar{\mathbf{v}}_k,\mathbf{v}_k)B_{kj}(v),$$

where $k = 1,\dots,n$, $j = 1,\dots,n_1$. We fix some k and consider the following stochastic equation for $\bar{\mathbf{v}}_k(\tau) \in \mathbb{R}^2$:

$$d\bar{\mathbf{v}}_k = \bar{\mathbf{P}}_k(\bar{\mathbf{v}}_k, v(\tau))d\tau + \sum_{j=1}^{n_1} \bar{B}_{kj}(\bar{\mathbf{v}}_k, v(\tau))d\boldsymbol{\beta}_j(\tau). \tag{A.1}$$

Its coefficients are well defined for all non-zero $\mathbf{v}_k$ and $\bar{\mathbf{v}}_k$. The equation (A.1), given some initial data, has a unique solution as long as $\|\mathbf{v}_k\|, \|\bar{\mathbf{v}}_k\| \geqslant \delta$ for any fixed $\delta > 0$.

For an arbitrary $\delta \in (0,\frac{1}{2})$ we define the stopping times $\tau_j^\pm$ as follows: $\tau_0^+ = 0$,

$$\tau_j^- = \inf\Big\{s \ge \tau_{j-1}^+ \,:\, \min_{1\le k\le n}\|\mathbf{v}_k(s)\| \le \delta \text{ or } \|v(s)\| \ge \delta^{-1}\Big\}, \quad j \ge 1,$$

$$\tau_j^+ = \inf\Big\{s \ge \tau_j^- \,:\, \min_{1\le k\le n}\|\mathbf{v}_k(s)\| \ge 2\delta \text{ and } \|v(s)\| \le (2\delta)^{-1}\Big\}, \quad j \ge 1.$$

Note that $\tau_0^+ \le \tau_1^-$ and $\tau_j^- < \tau_j^+ < \tau_{j+1}^-$ for $j \ge 1$. See again Fig. 1, where now the line is the graph of the function $\|\mathbf{v}_k(\tau)\|$.

Since on each interval Λ_j the norm of solution $v(\tau)$ of (6.1) is bonded by δ^{-1}, then Λ_j cannot be too short. So the sequence $\tau_j^\pm$ stabilizes at T after a finite random number of steps.

Now we construct a continuous process $\bar{\mathbf{v}}_k^\delta(\tau)$, $\tau \in [0,T]$. We set $\bar{\mathbf{v}}_k^\delta(\tau_0^+) = \mathbf{v}_k(\tau_0^+)$. For any $j \geqslant 0$ we define $\bar{\mathbf{v}}_k^\delta$ on the segment $\Lambda_j := [\tau_j^+, \tau_{j+1}^-]$ as a solution of equation (A.1), while on the complementary segments $\Delta_r = [\tau_r^-, \tau_r^+]$ we set

$$\bar{\mathbf{v}}_k^\delta(s) = U\big(\bar{\mathbf{v}}_k(\tau_r^-), \mathbf{v}_k(\tau_r^-)\big)\mathbf{v}_k(s). \tag{A.2}$$

Lemma A.1. *If $\|\bar{\mathbf{v}}_k^\delta(\tau_j^+)\| = \|\mathbf{v}_k(\tau_j^+)\|$, then $\|\bar{\mathbf{v}}_k^\delta(s)\| = \|\mathbf{v}_k(s)\|$ for all $s \in \Lambda_j$, a.s.*

Proof. Denote $I_k^\delta = \frac{1}{2}\|\bar{\mathbf{v}}_k^\delta\|^2$. By Itô's formula, on the segment Λ_{j-1},

$$dI_k^\delta = \big(\bar{\mathbf{v}}_k^\delta, \bar{\mathbf{P}}_k(\bar{\mathbf{v}}_k^\delta, v(\tau))\big)d\tau + \sum_{i=1}^{n_1}\Big(\frac{1}{2}\|\bar{B}_{ki}(\bar{\mathbf{v}}_k^\delta, v(\tau))\|_{HS}^2 d\tau + \big(\bar{\mathbf{v}}_k^\delta, \bar{B}_{ki}(\bar{\mathbf{v}}_k^\delta, v)\big)d\boldsymbol{\beta}_i(\tau)\Big),$$

and $I_k = \frac{1}{2}\|\mathbf{v}_k\|^2$ satisfies

$$dI_k = \big(\mathbf{v}_k, \mathbf{P}_k(v)\big)d\tau + \sum_{i=1}^{n_1}\Big(\frac{1}{2}\|B_{ki}(v(\tau))\|_{HS}^2 d\tau + \big(\mathbf{v}_k, B_{ki}(v)\big)d\boldsymbol{\beta}_i(\tau)\Big).$$

By construction we have the following relations for the drift and diffusion term of these two equations:

$$\big(\bar{\mathbf{v}}_k^\delta, \bar{\mathbf{P}}_k(\bar{\mathbf{v}}_k^\delta, v)\big) + \frac{1}{2}\sum_{i=1}^{n_1}\|\bar{B}_{ki}(\bar{\mathbf{v}}_k^\delta, v)\|_{HS}^2 = \frac{\|\bar{\mathbf{v}}_k^\delta\|}{\|\mathbf{v}_k\|}\big(\mathbf{v}_k, \mathbf{P}_k(v)\big) + \frac{1}{2}\sum_{i=1}^{n_1}\|B_{ki}(v)\|_{HS}^2,$$

$$\big(\bar{\mathbf{v}}_k^\delta, \bar{B}_{ki}(\bar{\mathbf{v}}_k^\delta, v)\big) = \frac{\|\bar{\mathbf{v}}_k^\delta\|}{\|\mathbf{v}_k\|}\big(\mathbf{v}_k, B_{ki}(v)\big).$$

For the squared difference $(I_k - I_k^\delta)^2$ we have,

$$\begin{aligned} d(I_k - I_k^\delta)^2 = \Big(&2(I_k - I_k^\delta)\frac{|\mathbf{v}_k| - |\bar{\mathbf{v}}_k^\delta|}{|\mathbf{v}_k|}\big(\mathbf{v}_k, \mathbf{P}_k(\mathbf{v})\big) \\ &+ \frac{(|\mathbf{v}_k| - |\bar{\mathbf{v}}_k^\delta|)^2}{|\mathbf{v}_k|^2}\sum_{i=1}^{n_1}\|(\mathbf{v}_k, B_{ki}(v))\|^2\Big)d\tau + d\mathcal{M}_\delta^\delta \end{aligned} \tag{A.3}$$

with $\mathcal{M}_\tau^\delta$ being a square integrable martingale. Since $I_k - I_k^\delta = \frac{1}{2}\big(\|\mathbf{v}_k\| - \|\bar{\mathbf{v}}_k^\delta\|\big)\big(\|\mathbf{v}_k\| + \|\bar{\mathbf{v}}_k^\delta\|\big)$, letting $\mathcal{J}_k^\delta(\tau) = (I_k - I_k^\delta)^2((\tau \vee \tau_j^+) \wedge \tau_{j+1}^-)$ and taking the expectation in (A.3) we have

$$\mathbf{E}\mathcal{J}_k^\delta(\tau) \le \mathbf{E}J_k^\delta(0) + c(\delta)\int_0^\tau \mathbf{E}\mathcal{J}_k^\delta(s)\,ds.$$

Since $J_k^\delta(\tau_j^+) = 0$, then by Gronwall's lemma, $\mathcal{J}_k^\delta(\tau) = 0$ for $\tau \in \Lambda_j$. The assertion of the lemma is proved. ☐

By Lemma A.1 we have $\|\mathbf{v}_k(s)\| = \|\bar{\mathbf{v}}_k^\delta(s)\|$ for all $s \in \Lambda_0$. By (A.2) $\|\mathbf{v}_k(s)\| = \|\bar{\mathbf{v}}_k^\delta(s)\|$ for $s \in \Delta_j$. Iterating this procedure we conclude that $\|\mathbf{v}_k(s)\| = \|\bar{\mathbf{v}}_k^\delta(s)\|$ on the whole interval $[0, T]$.

We define

$$\widehat{\mathbf{P}}_k(\bar{\mathbf{v}}_k^\delta, v, s) = \begin{cases} \bar{\mathbf{P}}_k(\bar{\mathbf{v}}_k^\delta, v), & \text{if } s \in \bigcup_j \Lambda_j, \\ U_j\big[\frac{1}{\varepsilon}W_k(I)\mathbf{v}_k^\perp + \mathbf{P}_k(v)\big], & \text{if } s \in \bigcup_j \Delta_j, \end{cases}$$

where $U_j = U(\bar{\mathbf{v}}_k(\tau_j^-), (\mathbf{v}_k(\tau_j^-)))$, and

$$\widehat{B}_{ki}(\bar{\mathbf{v}}_k^\delta, \mathbf{v}, s) = \begin{cases} \bar{B}_{ki}(\bar{\mathbf{v}}_k^\delta, v), & \text{if } s \in \bigcup_j \Lambda_j, \\ U_jB_{ki}(v(s)), & \text{if } s \in \bigcup_j \Delta_j. \end{cases}$$

Then $\bar{\mathbf{v}}_k^\delta$ satisfies the equation

$$\bar{\mathbf{v}}_k^\delta(\tau) = \mathbf{v}_k(0) + \int_0^\tau \widehat{\mathbf{P}}_k(\bar{\mathbf{v}}_k^\delta(s), \mathbf{v}(s), s)\,ds + \int_0^\tau \sum_{i=1}^{n_1}\widehat{B}_{ki}(\bar{\mathbf{v}}_k^\delta(s), \mathbf{v}(s), s)\,d\boldsymbol{\beta}_i(s). \tag{A.4}$$

Notice that under Assumption 6.1 the diffusion coefficient in this Ito equation does not degenerate.

Step 2: Truncation at a level $\|v\| = R$.

Define the stopping time $\tau_R = \inf\{\tau \in [0,T] : \|v\| \geqslant R\}$. We define the processes $\mathbf{v}_k^R$ equal to $\mathbf{v}_k$ for $\tau \in [0, \tau_R]$ and satisfying the trivial equation $d\mathbf{v}_k^R(\tau) = d\boldsymbol{\beta}(\tau)$ for $\tau \in [\tau_R, T]$. We also set $\bar{\mathbf{v}}_k^{\delta,R}$ to be equal to $\bar{\mathbf{v}}_k^\delta$ for $\tau \in [0, \tau_R]$ and for $\tau > \tau_R$ equal to a solution of the equation $d\bar{\mathbf{v}}_k^{\delta,R} = U(\mathbf{v}_k(\tau_R), \bar{\mathbf{v}}_k^{\delta,R}(\tau_R))d\boldsymbol{\beta}(\tau)$. Clearly, $\|\bar{\mathbf{v}}_k^{\delta,R}\| \equiv \|\mathbf{v}_k^R\|$. By (4) of Assumption 6.1 we have $\mathbf{P}\{\mathbf{v}_k^R(\tau) \neq \mathbf{v}_k(\tau) \text{ for some } \tau \in [0,T]\} \to 0$ as $R \to \infty$. Therefore, it is sufficient to prove the lemma for $\mathbf{v}_k$ replaced by $\mathbf{v}_k^R$ with arbitrary $R > 0$.

Step 3: Taking a limit as $\delta \to 0$.

Now fix $R > 0$. We denote $\bar{\mathbf{v}}_k^{\delta,R}(\cdot)$ still as $\bar{\mathbf{v}}_k^\delta(\cdot)$.

Argue as in Lemma 3.1 we find that the family of processes $\bar{\mathbf{v}}_k^{\delta,R}(\cdot)$, $\delta \in (0,1]$ is tight in $C([0,T],\mathbb{R}^2)$. Therefore, for a subsequence $\delta_j \to 0$, $\mathcal{D}(\bar{\mathbf{v}}_k^{\delta_j}(\cdot))$ converges to some $\mathcal{Q}_k^0 \in \mathcal{P}(C([0,T],\mathbb{R}^2))$. Consider the processes

$$\mathcal{Y}_k^\delta(\tau) := \int_0^\tau \widehat{\mathbf{P}}_k(\bar{\mathbf{v}}_k^\delta(s), \mathbf{v}(s), s)\, ds, \quad \mathcal{M}_k^\delta(\tau) := \sum_{i=1}^{n_1} \int_0^\tau \widehat{B}_{ki}(\bar{\mathbf{v}}_k^\delta(s), v(s), s)\, d\boldsymbol{\beta}_i(s),$$

with an obvious change for $\tau \geqslant \tau_R$. The sequence of pairs $(\mathcal{Y}_k^{\delta_j}(\tau), \mathcal{M}_k^{\delta_j}(\tau))$ is tight in $C(0,T;\mathbb{R}^4)$. If $(\mathcal{Y}_k^0(\cdot), \mathcal{M}_k^0(\cdot))$ is a limiting in law process as $\delta_j \to 0$, then $\mathcal{D}\mathbf{v}_k(\cdot) = \mathcal{Q}_k^0$, where

$$\bar{\mathbf{v}}_k(\tau) = \mathcal{Y}_k^0(\tau) + \mathcal{M}_k^0(\tau), \quad \tau \in [0,T].$$

Denote $C_{R,k}^P = \sup\{|\mathbf{P}_k(v)|, \|v\| \leqslant R\}$ and $C_{R,k}^W = \sup\{|W_k(I(v))|, \|v\| \leqslant R\}$. Then for any $0 \le \tau' < \tau'' \le T$ we have

$$|\mathcal{Y}_k^\delta(\tau'') - \mathcal{Y}_k^\delta(\tau')| \le C_{R,k}^P|\tau'' - \tau'| + C_{R,k}^W \varepsilon^{-1}\Big|\bigcup_j [\tau_j^-, \tau_j^+] \cap [\tau', \tau'']\Big|.$$

From the definition of $\tau_j^\pm$ and τ_j^+ it follows that

$$\mathbf{E}\Big|\bigcup_j \Delta_j \cap [\tau', \tau'']\Big| \le \mathbf{E}\Big|\bigcup_j \Delta_j \cap [0,T]\Big| \le \mathbf{E}\int_0^T \mathbf{1}_{\{|\mathbf{v}_k(\tau)| \le 2\delta\}}\, d\tau.$$

By Theorem 2.2.4 in [20] for each $\varepsilon > 0$ the term on the right-hand side of this inequality tends to zero as $\delta \to 0$. Therefore, $\mathbf{E}\Big|\bigcup_j \Delta_j \cap [\tau', \tau'']\Big| \to 0$. Using the fact that the set $\{\varphi \in C(0,T;\mathbb{R}^2)\,;\, |\varphi(\tau') - \varphi(\tau'')| \le 2C_{R,k}^P|\tau' - \tau''|\}$ is closed we derive from the convergence $\mathcal{D}(\mathcal{Y}_k^{\delta_j}(\cdot)) \rightharpoonup \mathcal{D}(\mathcal{Y}_k^0(\cdot))$ that

$$\mathbf{P}\big\{|\mathcal{Y}_k^0(\tau') - \mathcal{Y}_k^0(\tau'')| \le C_{P,k}|\tau' - \tau''|\big\} = 1.$$

So $\mathcal{Y}_k^0(\tau) = \int_0^\tau \rho_k^0(s)\, ds$ with $|\rho(s)| \le C_{R,k}^P$.

The processes $\mathcal{M}_k^\delta$ are continuous square integrable martingales with respect to the natural filtration. Since their second moment are bounded uniformly in δ, then the limit process $\mathcal{M}_k^0$ is also a square integrable martingale. Denoting by $\langle \mathcal{M}_k^\delta \rangle(\tau)$ the quadratic characteristic of $\mathcal{M}_k^\delta(\tau)$, from Corollary VI.6.7 in [14] we deduce that $\langle \mathcal{M}_k^\delta \rangle(\tau) \to \langle \mathcal{M}_k^0 \rangle(\tau)$ as $\delta \to 0$.

Under Assumption 6.1 the quadratic characteristic of $\mathcal{M}_k^\delta$ satisfies the estimates

$$C_m(\tau'' - \tau')|\zeta|^2 \le \big([\langle \mathcal{M}_k^\delta \rangle(\tau'') - \langle \mathcal{M}_k^\delta \rangle(\tau')]\zeta, \zeta\big) \le C_m^{-1}(\tau'' - \tau')|\zeta|^2, \quad \zeta \in \mathbb{R}^2,$$

for some constant $C_m > 0$. Then the quadratic characteristic of $\mathcal{M}_k^0$ also meets these estimates. Thus there exists a progressively measurable random matrix function $\sigma_k(s)$ with values in the space of symmetric 2×2 matrices such that $C_m|\zeta|^2 \le (\sigma_k(t)\zeta, \zeta) \le C_m^{-1}|\zeta|^2$ for all $\zeta \in \mathbb{R}^2$, and $\mathcal{M}_k^0(\tau) = \int_0^\tau \sigma^{\frac{1}{2}}(s) dB_s$, where B_s is a standard Wiener process in $\mathbb{R}^2$. Therefore, the process $\bar{\mathbf{v}}_k$ admits the following representation:

$$\bar{\mathbf{v}}_k(\tau) = \bar{\mathbf{v}}_k(0) + \int_0^\tau \rho_k^0(s)\, ds + \int_0^\tau \sigma_k^{\frac{1}{2}}(s) dB_s \tag{A.5}$$

The desired statement is now an immediate consequence of Theorem 2.2.4 in [20].

Appendix B. On Birkhoff integrability of Hamiltonian systems with one degree of freedom

Here we prove a global version of the well known fact that 1d Hamiltonian systems are integrable (e.g. see Example 3 in [1, Section 5.3]). Consider the plane $\mathbb{R}^2 = \{\mathbf{x} = (x,y)\}$, equipped with the standard area-form $dx \wedge dy$. By B_r we denote the open disc $\{\|\mathbf{x}\| < r\}$, $r > 0$, and set $I(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|^2 = \frac{1}{2}(x^2 + y^2)$.

Theorem B.1. *Assume that $H \in C^\infty(\mathbb{R}^2)$ satisfies the following:*
i) $dH(0) = 0$ and $d^2H(0)$ is positively definite;
ii) for each $\mathbf{x} \neq 0$, $dH(\mathbf{x}) \neq 0$;
iii) for each $a \in H(\mathbb{R}^2 \setminus \{0\})$ the level set $M_a = \{\mathbf{x} \in \mathbb{R}^2 : H(\mathbf{x}) = a\}$ is a connected loop.

Then there exist a smooth canonical change of coordinates (SCCC) $\Psi : \mathbb{R}^2 \to \mathbb{R}^2$, $\Psi(0) = 0$ and a smooth function h, $h'(0) \neq 0$, such that $H(\mathbf{x}) = h\big(I(\Psi(\mathbf{x}))\big)$.

Proof. Step 1: By Vey's Theorem (see [7] and see a 1d version of the theorem in [18, Appendix D]), there exist $\delta \in [0, 1/2]$ and a SCCC $Q_\delta : B_\delta \to \mathbb{R}^2$, $(p,q) \mapsto (x,y)$, such that $Q_\delta(0) = 0$ and $H \circ Q_\delta(p,q) = f(\frac{p^2+q^2}{2})$, where f is a smooth function, satisfying $f'(0) \neq 0$.

Step 2: Now we construct a SCCC Q defined on the whole plane, such that for some $0 < \delta' < \delta$ we have $Q|_{B_{\delta'}} = Q_\delta|_{B_{\delta'}}$ and $Q|_{\mathbb{R}^2 \setminus B_1} = L$, where L is the linear symplectic transformation $L = dQ(0) = dQ_\delta(0)$.

Let $\bar{Q}_\delta = L^{-1} \circ Q_\delta$, $(p,q) \mapsto (x,y)$. Then

$$d\bar{Q}_\delta(0) = Id \quad \text{and} \quad \partial_p x(p,q) > 0, \tag{B.1}$$

if $\|(p,q)\|$ is small. Then, decreasing δ if needed, we achieve that the transformation $\bar{Q}_\delta$ admits a smooth generating function $S(x,q)$, so $\bar{Q}_\delta(p,q) = (x,y)$ if and only if $p = \partial_q S, y = \partial_x S$ (e.g. see [1, Section 1.3]). Since $d\bar{Q}_\delta(0) = \mathrm{id}$, then $S(x,q) = xq + o(\|(x,q)\|^2)$. Now we extend $S(x,q)$ from a small neighbourhood of the origin to the whole (x,q)-plane in such a way that $S(x,q) = xq$ for $\|(x,q)\| \geqslant 1$, keeping the condition $\frac{\partial^2 S}{\partial x \partial q} > 0$ for $\|(x,q)\| \leqslant 1$. The extended S is as a generating function of a SCCC $Q' : \mathbb{R}^2 \to \mathbb{R}^2$, $(p,q) \mapsto (x,y)$. Then $Q'|_{\mathbb{R}^2 \setminus B_1} = Id$ and $Q'|_{B_{\delta'}} = \bar{Q}_\delta|_{B_{\delta'}}$ for a small enough $\delta' < \delta$. The required SCCC is obtained as $Q = L \circ Q'$.

Step 3: Denote $H_1 := H \circ Q(p,q)$. Clearly, conditions i)-iii) stay true for H_1 and in $B_{\delta'}$,

$$H_1(p,q) = f(\tfrac{p^2+q^2}{2}). \tag{B.2}$$

Let $a_0 \in H_1(\mathbb{R}^2 \setminus \{0\})$. By the Loiuville–Arnold theorem in a small neighbourhood of the curve $M_{a_0} = \{H_1 = a_0\}$ exists a SCCC Q_1: $(p,q) \mapsto (I, \varphi) \in \mathbb{R} \times \mathbb{T}$, $\mathbb{T} = \mathbb{R}/2\pi$, such that $dp \wedge dq = dI \wedge d\varphi$ and $H_1(p,q) = h(I(p,q))$ for a smooth function h. Moreover, $M_a \cong \{I = I(a)\} \times \mathbb{T}$ and $I(a) = \frac{1}{2\pi}\oint_{M_a} pdq$. By Green's formula,

$$I(a) = \frac{1}{2\pi}\int_{M_a} pdq = \frac{1}{2\pi}\iint_{D(M_a)} dp \wedge dp, \tag{B.3}$$

where $D(M_a)$ is the domain enclosed by M_a. Hence for each $a \in H_1(\mathbb{R}^2 \setminus \{0\})$, $2\pi I(a)$ is the area enclosed by M_a. So the action variable $I(p,q) = I(a(p,q))$ is well defined globally on $\mathbb{R}^2 \setminus \{0\}$.

The angle variable $\varphi \in \mathbb{T}$ is defined modulo a shift $\varphi \mapsto \varphi + \zeta(I)$, where ζ is any smooth function. To specify its choice we find a smooth curve $l_0 \subset \mathbb{R}^2$ from the origin to infinity, such that $l_0 \cap B_{\delta'} \subset \{p > 0, q = 0\}$ and l_0 insects each level set M_a in exactly one point. Setting $\varphi|_{l_0} = 0$ we get a uniquely defined smooth angle variable φ on $\mathbb{R}^2 \setminus \{0\}$. The constructed variables (I, φ) define an action-angle transformation $\mathbb{R}^2 \setminus \{0\} \to \mathbb{R}_{>0} \times \mathbb{T}$, $(p,q) \mapsto (I, \varphi)$.

Now let $G(p,q) = \sqrt{2I}(\cos\varphi, \sin\varphi)$. We then have a SCCC $G\colon \mathbb{R}^2 \setminus \{0\} \circlearrowleft$ such that $H_1(p,q) = h(I(G(p,q)))$. From (B.2), (B.3) and the normalisation $\varphi(p,0) = 0$ if $0 < p < \delta'$ we have that $G(p,q) = (p,q)$ on $B_{\delta'} \setminus \{0\}$. Therefore G extends $Q\mid_{B_{\delta'}}$ to a SCCC of $\mathbb{R}^2$.

We define the wanted SCCC as $\Psi = G \circ Q^{-1}$ and the assertion of the theorem follows. □

Acknowledgement We thank Anatoli Neishtadt for discussions of the deterministic averaging. This research was supported by the Ministry of Science and Higher Education of the Russian Federation (megagrant No. 075-15-2022-1115). GH was also supported by National Natural Science Foundation of China (No. 20221300605).

References

[1] V. I. Arnold, V. V. Kozlov, and A. I. Neishtadt. *Mathematical Aspects of Classical and Celestial Mechanics*. Springer-Verlag, New York–Berlin–Heidelberg, third edition, 2006.
[2] P. Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, New York, 1999.
[3] F. Bonetto, J.L. Lebowitz, and L. Rey-Bellet. Fourier's law: A challenge to theorists. In *Mathematical Physics*, pages 128–150. Imp. Coll. Press, London, 2000.
[4] A. Boritchev and S. B. Kuksin. *One-dimensional turbulence and the stochastic Burgers equation*. Mathematical Surveys and Monographs. AMS Publications, Providence, 2021.
[5] R. M. Dudley. *Real Analysis and Probability*. Cambridge studies in advanced mathematics. Cambridge University Press, 2002.
[6] J.-P. Eckmann, C.-A. Pillet, and L. Rey-Bellet. Non-equilibrium statistical mechanics of anharmonic chains coupled to two heat baths at different temperatures. *Comm. Math. Phys.*, 201:657–697, 1999.
[7] L.H. Eliasson. Normal forms for hamiltonian systems with Poisson commuting integrals-elliptic case. *Comment. Math. Helv.*, 65:4–35, 1990.
[8] M. Freidlin and A. Wentzell. *Random Perturbations of Dynamical Systems*. Springer-Verlag, New York, 2nd edition, 1998.
[9] M. I. Freidlin and A. D. Wentzel. Averaging principle for stochastic perturbations of multi-frequency systems. *Stochastics and Dynamics*, 3:393–408, 2003.
[10] G. Huang and S. Kuksin. KdV equation under periodic boundary conditions and its perturbations. *Nonlinearity*, 27:1–28, 2014.
[11] G. Huang and S. Kuksin. On averaging and mixing for stochastic PDEs. *J. Dyn. Diff. Eq.*, 2022. doi 10.1007/s10884-022-10202-w.
[12] G. Huang and S. Kuksin. Averaging and mixing for stochastic perturbations of linear conservative systems. *Russian Mathematical Surveys, to appear*, 2023.
[13] G. Huang, S. B. Kuksin, and A. Maiocchi. Time-averaging for weakly nonlinear CGL equations with arbitrary potential. *Fields Inst. Comm.*, 75:323–349, 2015.
[14] J. Jacod and A. N. Shiryaev. *Limit Theorems for Stochastic Processes*. Springer-Verlag, New York–Berlin–Heidelberg, 1987.
[15] I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer-Verlag, New York–Berlin–Heidelberg, 2002.
[16] R. Khasminski. On the avaraging principle for Ito stochastic differential equations. *Kybernetika*, 4:260–279, 1968. (in Russian).
[17] R. Khasminskii. *Stochastic Stability of Differential Equations*. Springer-Verlag, New York–Berlin–Heidelberg, second edition, 2012.
[18] B. Khesin, S. B. Kuksin, and D. Peralta-Salas. Global, local and dense non-mixing of the 3d euler equation. *ARMA*, 238:1087–1112, 2020.
[19] Yu. Kifer. Stochastic versions of Anosov's and Neistadt's theorems on averaging. *Stochastics and Dynamics*, 1:1–21, 2001.
[20] N. V. Krylov. *Controlled Diffusion Processes*. Springer-Verlag, New York–Berlin–Heidelberg, 1980.
[21] S. Kuksin and A. Shirikyan. *Mathematics of Two-Dimensional Turbulence*. Cambridge University Press, Cambridge, 2012.
[22] S. B. Kuksin. Damped-driven KdV and effective equations for long-time behaviour of its solutions. *GAFA*, 20:1431–1463, 2010.
[23] S. B. Kuksin and A. Piatnitski. Khasminsii–Whitham averaging for randomly perturbed KdV equation. *J. Math. Pures Appl.*, 89:400–428, 2008.

[24] S.-J. Liu and M. Krstic. *Stochastic Averaging and Stochastic Extremum Seeking*. Springer, 2012.
[25] P. Lochak and C. Meunier. *Multiphase Averaging for Classical Systems*. Springer-Verlag, New York–Berlin–Heidelberg, 1988.
[26] L. Niederman. Nonlinear stability around an elliptic equilibrium point in an Hamiltonian system. *Nonlinearity*, 11:1465–1479, 1998.
[27] A. V. Skorokhod. *Asymptotic Methods in the Theory of Stochastic Differential Equations*, volume 78 of *Translations of Mathematical Monographs*. AMS Publications, Providence, 1989.
[28] D. Stroock and S. R. S. Varadhan. *Multidimensional Diffusion Processes*. Springer-Verlag, New York–Berlin–Heidelberg, 1979.
[29] A. Yu. Veretennikov. On the averaging principle for systems of stochastic differential equations. *Math. USSR Sbornik*, 69:271–284, 1991.
[30] C. Villani. *Optimal Transport*. Springer-Verlag, Berlin, 2009.
[31] H. Whitney. Differentiable even functions. *Duke Math. J.*, 10:159–160, 1943. Also see in H. Whitney, Collected Papers vol. 1, Birkhäuser, 1992, pp. 309-310.

School of Mathematics and Statistics, Beijing Institute of Technology, Beijing, China and Peoples' Friendship University of Russia (RUDN University), Moscow, Russia
Email address: huangguan@bit.edu.cn

Université Paris Cité and Sorbonne Université, CNRS, IMJ-PRG, F-75013 Paris, France, and Peoples' Friendship University of Russia, Moscow, Russia, and Steklov Mathematical Institute of Russian Academy of Sciences, Moscow, Russia
Email address: sergei.kuksin@imj-prg.fr

The Arctic University of Norway, campus Narvik, Norway, and Institute for Information Transmission Problems of RAS, Moscow, Russia, and Peoples' Friendship University of Russia (RUDN University), Moscow, Russia
Email address: apiatnitski@gmail.com