HAL open science

AVERAGING AND MIXING FOR STOCHASTIC PERTURBATIONS OF LINEAR CONSERVATIVE SYSTEMS

Guan Huang, Sergei Kuksin

To cite this version:

Guan Huang, Sergei Kuksin. AVERAGING AND MIXING FOR STOCHASTIC PERTURBATIONS OF LINEAR CONSERVATIVE SYSTEMS. Russian Mathematical Surveys, 2023, 78 (4), pp.585-633. 10.4213/rm10081e . hal-04466883

HAL Id: hal-04466883

https://hal.science/hal-04466883v1

Submitted on 19 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

AVERAGING AND MIXING FOR STOCHASTIC PERTURBATIONS OF LINEAR CONSERVATIVE SYSTEMS

GUAN HUANG AND SERGEI KUKSIN

Dedicated to the memory of M.I.Vishik
on the occasion of his 100-th birthday.

Abstract. We study stochastic perturbations of linear systems of the form

$$dv(t) + Av(t)\,dt = \varepsilon P(v(t))dt + \sqrt{\varepsilon}\,\mathcal{B}(v(t))dW(t), \quad v \in \mathbb{R}^D, \qquad (*)$$

where A is a linear operator with non-zero imaginary spectrum. It is assumed that the vector field $P(v)$ and the matrix-function $\mathcal{B}(v)$ are locally Lipschitz with at most a polynomial growth at infinity, that the equation is well posed and first few moments of norms of solutions $v(t)$ are bounded uniformly in ε. We use the Khasminski approach to stochastic averaging to show that as $\varepsilon \to 0$, a solution $v(t)$, written in the interaction representation in terms of operator A, for $0 \le t \le \mathrm{Const}\,\varepsilon^{-1}$ converges in distribution to a solution of an effective equation. The latter is obtained from $(*)$ by means of certain averaging. Assuming that eq. $(*)$ and/or the effective equation are mixing, we examine this convergence further.

Contents

1. Introduction 1
2. Linear systems and their perturbations 4
3. Averaging of vector fields with respect to the frequency vector 8
4. Effective equation and the averaging theorem 11
5. Stationary solutions and mixing 20
6. The non-resonant case 23
7. Uniform in time convergence. 25
8. Averaging for systems with general noises 31
9. A sufficient condition for Assumptions 2.1, 5.1 and 7.1 32
Appendix A. Proof of Proposition 9.3 34
Appendix B. Representation of martingales 36
Appendix C. Itô's formula for complex processes 36
Appendix D. Projections to convex sets 38
Acknowledgement 38
References 38

1. Introduction

1.1. **The setting and problems.** The goal of this paper is to present an averaging theory for perturbations of conservative linear differential equations by locally Lipschitz nonlinearities and stochastic terms. Namely, we will examine stochastic equations

$$dv(t) + Av(t)\,dt = \varepsilon P(v(t))dt + \sqrt{\varepsilon}\,\mathcal{B}(v(t))dW(t), \qquad v \in \mathbb{R}^D, \tag{1.1}$$

where $0 < \varepsilon \le 1$, A is a linear operator with non-zero pure imaginary eigenvalues $\{i\lambda_j\}$ (so the dimension D is even), P is a locally Lipschitz vector field on $\mathbb{R}^D$, $W(t)$ is the standard Wiener process in $\mathbb{R}^N$ and $\mathcal{B}(v)$ is an $D \times N$-matrix. We wish to study for small ε the behaviour of solutions for eq. (1.1) on time-intervals of order ε^{-1}, and under some additional restriction on the equation examine the limiting as $\varepsilon \to 0$ behaviour of solutions uniformly in time.

1.2. **Our results and their deterministic analogies.** We tried to make our work "reader-friendly" and accessible to people with just a limited knowledge of stochastic calculus. To achieve that in the main part of the paper we restrict ourselves to the case of equations with additive noise $\sqrt{\varepsilon}\,\mathcal{B}\,dW(t)$ and exploit there a technical convenience: we introduce in $\mathbb{R}^D$ a complex structure, re-writing the phase-space $\mathbb{R}^D$ as $\mathbb{C}^{D/2}$ (recall that D is even), in such a way that in the corresponding complex coordinates the operator A is diagonal, $A = \mathrm{diag}\{i\lambda_j\}$. General equations (1.1) are discussed in Section 8, where they are treated in parallel with earlier considered equations with additive noise.

As it is custom in the classical deterministic Krylov–Bogolyubov (K-B) averaging (e.g. see [4], [1] and [13]), to study solutions $v(t) \in \mathbb{C}^{D/2}$ we write them in the interaction representation which preserves the norms of the complex components $v_j(\tau)$, but amends their angles. See below substitution (2.8). The first principal result of the work is given by Theorem 4.7, where we assume uniform in ε and in $t \le C\varepsilon^{-1}$ bounds on a first few moments of norms of solutions. The theorem states that as $\varepsilon \to 0$, for $t \le C\varepsilon^{-1}$ solutions $v(t)$, written in terms of the interaction representation, weakly converge in distribution to solutions of an additional *effective equation*. The latter is obtained from eq. (1.1) by means of certain averaging of vector field P in terms of the spectrum $\{i\lambda_j\}$ and in many cases may be written down explicitly. Proof of Theorem 4.7, given in Section 4, is obtained by a synthesis of the K-B method (as it is presented e.g. in [13]) and the Khasminski approach to the stochastic averaging [16]; it may serve as an introduction to the latter. The number of works on the stochastic averaging is immense, see Section 1.3 for some references. We were not able to find there the result of Theorem 4.7, but do not insist on its novelty (and certainly related statements may be found in the literature).

In Section 5 we suppose that the bounds on the moments of norms of solutions, mentioned above, are uniform in time, and that eq. (1.1) is mixing. So its solutions, as time goes to infinity, converge in distribution to a unique stationary measure (which is a Borel measure in $\mathbb{R}^D = \mathbb{C}^{D/2}$). In Theorem 5.5, postulating that the effective equation as well is mixing, we prove that when $\varepsilon \to 0$ the stationary measure for eq. (1.1) converges to that for the effective equation. Note that this convergence holds without passing to the interaction representation.

In short Section 6 we discuss non-resonant systems (1.1) (where the frequencies $\{\lambda_j\}$ are rationally independent). In particular, we show that then the actions $I_j(v(t))$ of solutions $v(t)$ (see below (1.2)), as $\varepsilon \to 0$, converge in distribution to solutions of a system of stochastic equations, depending only on actions. The convergence hold on time-intervals $0 \le t \le C\varepsilon^{-1}$.

In Section 7 we keep the assumption on the norms of solutions from Section 5. Supposing that the effective equation is mixing (and without assuming that for the original equation (1.1)) we prove there Theorem 7.4. It states that the convergence as in the principal Theorem 4.7 is uniform for $t \ge 0$ (and not only for $t \le C\varepsilon^{-1}$).

In Proposition 9.4 we present a simple sufficient condition on eq. (1.1), based on results from [17], which insures that Theorems 4.7, 5.5 and 7.4 apply to it.

In Section 8 we pass to general equations (1.1), where the dispersion matrix $\mathcal{B}$ depends on v. Assuming the same estimates on solutions as in Section 4 we show that Theorem 4.7 remains valid if either the matrix $\mathcal{B}(v)$ is non-degenerate, or it is a C^2-smooth function of

v. Theorems 5.5 and 7.4 also remains true for general systems (1.1), but we do not discuss this, hoping that the corresponding modifications of the proofs should be clear after reading Section 8.

A deterministic analogy of our results, which deals with eq. (1.1) with $W = 0$ and describes the behaviour of its solutions on time-intervals of order ε^{-1} in the interaction representation in comparison with solutions for a corresponding effective equation, is given by the K-B averaging (see [4, 1, 13]).[1] Theorem 5.5 has no analogy for deterministic systems, but Theorem 7.4 has it. Namely, for the K-B averaging it is known that if the effective equation has a globally asymptotically stable equilibrium, then the convergence of solutions for eq. $(1.1)_{W=0}$, written in the interaction representation, to solutions of the effective equation, is uniform in time. This result is known in folklore as the second K-B theorem and may be found in [6].

The K-B method and Khasminski approach to averaging which we exploit, are flexible tools. They are applicable to various stochastic systems in finite and infinite dimensions, including stochastic PDEs, and the specific realization of the two methods which we use now is inspired by our previous work on averaging for stochastic PDEs. See [12, 19] for an analogy for SPDEs of Theorem 4.7, [12] – for an analogy of Theorem 5.5 and [11] – for an analogy of Theorem 7.4 (also see [7] for more results and references on averaging for SPDEs).

1.3. **Relation with classical stochastic averaging.** The averaging in stochastic systems is a well developed topic, usually dealing with fast-slow stochastic systems, e.g. see publications [16], [10, Section 7], [23, Section II.3], [21], [25], [18] and references there. To explain relation of that theory with our work let us write eq. (1.1) in the complex form $v(t) \in \mathbb{C}^n$, $n = D/2$ (when the operator A is diagonal) and then pass to the slow time $\tau = \varepsilon t$ and action-angle coordinates $(I, \varphi) = (I_1, \dots, I_n; \varphi_1, \dots, \varphi_n) \in \mathbb{R}^n_+ \times \mathbb{T}^n$, where $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$, $\mathbb{T}^n = \mathbb{R}^n/(2\pi\mathbb{Z}^n)$ and

$$I_k(v) = \tfrac{1}{2}|v_k|^2 = \tfrac{1}{2}v_k\bar{v}_k, \quad \varphi_k(v) = \operatorname{Arg} v_k \in \mathbb{S}^1 = \mathbb{R}/2\pi\mathbb{Z}, \quad k = 1, \dots, n \tag{1.2}$$

(if $v_k = 0$, we set $\varphi_k(v) = 0 \in \mathbb{S}^1$). In these coordinates equation (1.1) takes the form

$$\begin{cases} dI(\tau) = P^I(I, \varphi)d\tau + \Psi^I(I, \varphi)d\beta(\tau), \\ d\varphi(\tau) + \varepsilon^{-1}\Lambda d\tau = P^\varphi(I, \varphi)d\tau + \Psi^\varphi(I, \varphi)d\beta(\tau). \end{cases} \tag{1.3}$$

Here $\beta = (\beta_1, \dots, \beta_N)$, where $\{\beta_l\}$ are independent standard real Wiener processes, and the coefficients of the system are given by Itô's formula. This is a fast-slow system with the slow variable I and fast variable φ. The stochastic averaging treats systems like (1.3), usually adding to the fast part of the φ-equation a non-degenerate stochastic term of order $\varepsilon^{-1/2}$. The (first) goal of a system's analysis usually is to prove that on time-intervals $0 \leq \tau \leq T$ distributions of the I-components of solutions converge as $\varepsilon \to 0$ to distributions of solutions for a suitably averaged I-equation. After that other goals may be pursued.[2]

Unfortunately the stochastic averaging does not apply directly to systems (1.3), coming from equations (1.1), since then the coefficients of the φ-equation have singularities when some I_k vanishes, and since the fast φ-equation is rather degenerate if the vector Λ is resonant. Instead we borrow from the theory the Khasminski method [16] of stochastic averaging

[1] Theorem 4.7 also applies to such equations, but then its assertion becomes unnatural.

[2] E.g. one may study deviation of the I-components of solutions from the averaged dynamics (see [10, 18]) or, under stronger restrictions on the system, may examine behaviour of solutions on longer time intervals (see paper [15] and works, descending from it).

and apply it to eq. (1.1), written in the interaction representation, thus arriving at the assertion of Theorem 4.7. Averaging theorem for stationary solutions of eq. (1.3) and for the corresponding stationary measures are known in the stochastic averaging, but (naturally) they control the limiting behaviour only of the I-components of the stationary solutions and measures, while our Theorem 5.5 describes a limit of the whole stationary measure. It seems that no analogy of Theorem 7.4 is known in the stochastic averaging.

At the origin of this paper are lecture notes for an online course which SK was teaching in the Shandong University (PRC) in the autumn term of the year 2020.

Notation. 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. For any $0 < \alpha \leqslant 1$ and $u \in C([0,T],E)$,

$$\|u\|_\alpha = \sup_{0\leqslant\tau<\tau'\leqslant T} \frac{|u(\tau') - u(\tau)|_E}{|\tau' - \tau|^\alpha} + \sup_{\tau\in[0,T]} |u(\tau)|_E \le \infty. \tag{1.4}$$

This is a norm in the Hölder space $C^\alpha([0,T],E)$. We denote the standard C^m-norm for C^m-smooth functions on E as $|\cdot|_{C^m(E)}$. 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 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. $F \circ \mu(Q) = \mu(F^{-1}(Q))$.

If $m \ge 0$ and L is $\mathbb{R}^n$ or $\mathbb{C}^n$, then $\mathrm{Lip}_m(L,E)$ is the collection of maps $F : L \to E$ such that for any $R \geqslant 1$ we have

$$(1+|R|)^{-m}\big(\mathrm{Lip}(F|_{\bar{B}_R(L)}) + \sup_{v\in\bar{B}_R(L)} |F(v)|_E\big) =: \mathcal{C}^m(F) < \infty, \tag{1.5}$$

where $\mathrm{Lip}(f)$ is the Lipschitz constant of a mapping f (note that, in particular, $|F(v)|_E \le \mathcal{C}^m(F)(1+|v|_L)^m$ for any $v \in L$). For a complex matrix $A = (A_{ij})$, $A^* = (A^*_{ji})$ stands for its Hermitian conjugate: $A^*_{ij} = \bar{A}_{ji}$ (so for a real matrix B, B^* is the transposed matrix). For a set Q we denote by $\mathbf{1}_Q$ its indicator function, and by Q^c – its complement. Finally, $\mathbb{R}_+$ ($\mathbb{Z}_+$) stands for the set of nonnegative real numbers (nonnegative integers), and for real numbers a and b, $a \vee b$ and $a \wedge b$ indicate their maximum and minimum.

2. Linear systems and their perturbations

In this section we give the setting of the problem and specify assumptions on the operator A, vector field P and noise $\sqrt{\varepsilon}\,\mathcal{B}(v)dW$ in eq. (1.1). To simplify presentation and explain better the ideas, in the main part of the text we assume that the noise is additive, i.e. $\mathcal{B}$ is a constant matrix (possibly degenerate). We will discuss general equations (1.1) in Section 8.

2.1. Assumptions on A and $W(t)$.

We assume that the unperturbed linear system

$$(d/dt)v + Av = 0, \qquad v \in \mathbb{R}^D, \tag{2.1}$$

is such that all its trajectories are bounded as $t \to \pm\infty$. Then the eigenvalues of A are pure imaginary, go in pairs $\pm i\lambda_j$ and A has no Jordan cells. We also assume that A is invertible. So

(1) eigenvalues of A are of the form $\pm i\lambda_j$, where $\mathbb{R} \ni \lambda_j \ne 0$;
(2) in the Jordan normal form of A there are no Jordan cells.

By these assumptions $D = 2n$, and in $\mathbb{R}^{2n}$ there exists a base $\{\mathbf{e}_1^+, \mathbf{e}_1^-, \dots, \mathbf{e}_n^+, \mathbf{e}_n^-\}$ in which the linear operator A takes the block-diagonal form:

$$A = \begin{pmatrix} 0 & -\lambda_1 & & & \\ \lambda_1 & 0 & & 0 & \\ & & \ddots & & \\ & 0 & & 0 & -\lambda_n \\ & & & \lambda_n & 0 \end{pmatrix}$$

We denote by $(x_1, y_1, \dots, x_n, y_n)$ the coordinates, corresponding to this base, and for $j = 1, \dots, n$ set $z_j = x_j + iy_j$. Then $\mathbb{R}^{2n}$ becomes the space of complex vectors $(z_1, \dots, z_n)$, i.e. $\mathbb{R}^{2n} \simeq \mathbb{C}^n$. In the complex coordinates the standard inner product in $\mathbb{R}^{2n}$ reads

$$\langle z, z' \rangle = \mathrm{Re} \sum_{j=1}^{n} z_j \bar{z}_j', \quad z,\ z' \in \mathbb{C}^n. \tag{2.2}$$

Let us denote by

$$\Lambda = (\lambda_1, \dots, \lambda_n) \in (\mathbb{R} \setminus \{0\})^n$$

the *frequency vector* of the linear system (2.1). Then in the complex coordinates z the operator A reads

$$Az = \mathrm{diag}\{i\Lambda\} z,$$

where $\mathrm{diag}\{i\Lambda\}$ is the diagonal operator, sending a vector $(z_1, \dots, z_n)$ to $(i\lambda_1 z_1, \dots, i\lambda_n z_n)$. Therefore in $\mathbb{R}^{2n}$, written as the complex space $\mathbb{C}^n$, linear equation (2.1) takes the diagonal form,

$$(d/dt) v_k + i\lambda_k v_k = 0, \qquad 1 \leqslant k \leqslant n.$$

Below we examine the perturbed eq. (1.1), using these complex coordinates.

We next discuss the random process $W(t)$, written in the complex coordinates. The standard *complex Wiener process* has the form

$$\beta^c(t) = \beta^+(t) + i\beta^-(t) \in \mathbb{C}, \tag{2.3}$$

where $\beta^+(t)$ and $\beta^-(t)$ are independent standard (real) Wiener processes, defined on some probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Then $\bar{\beta}^c(t) = \beta^+(t) - i\beta^-(t)$, and any Wiener process $W(t) \in \mathbb{C}^n$ may be conveniently written in the complex from as

$$W_k = \sum_{l=0}^{n_1} \Psi^1_{kl} \beta^c_l + \sum_{l=1}^{n_1} \Psi^2_{kl} \bar{\beta}^c_l, \quad k = 1, \dots, n, \tag{2.4}$$

where $\Psi^1 = (\Psi^1_{kl})$ and $\Psi^2 = (\Psi^2_{kl})$ are complex $n \times n_1$ matrices and $\{\beta^c_l\}$ are independent standard complex Wiener processes. Again, in order to simplify presentation below we suppose that the noise in eq. (1.1) is of the form

$$W(t) = \sum_{l=1}^{n_1} \Psi_{kl} \beta^c_l(t), \quad k = 1, \dots, n.$$

We *do not* assume that the matrix Ψ is non-degenerate (in particular, it may be zero). Then the perturbed equation (1.1) in the complex coordinates reads as

$$dv_k + i\lambda_k v_k dt = \varepsilon P_k(v) dt + \sqrt{\varepsilon} \sum \Psi_{kl} d\beta^c_l(t), \quad k = 1, \dots, n\,, \tag{2.5}$$

where $v = (v_1, \dots, v_n) \in \mathbb{C}^n$ and $0 < \varepsilon \leq 1$.

The results, obtained below for eq. (2.5), remain true for general equations (1.1) on the price of heavier calculation. The corresponding argument is sketched in Section 8.

2.2. **Assumptions on P and on the perturbed equation.** Our first goal is to study equation (2.5) with $0 < \varepsilon \ll 1$ on a time interval $0 \le t \le \varepsilon^{-1}T$, where $T > 0$ is a fixed constant. Introducing the slow time

$$\tau = \varepsilon t$$

we write the equation as

$$dv_k(\tau) + i\varepsilon^{-1}\lambda_k v_k d\tau = P_k(v)d\tau + \sum_{l=1}^{n_1} \Psi_{kl} d\tilde{\beta}_l^c(\tau), \qquad k = 1, \dots, n, \ \ 0 \le \tau \le T. \tag{2.6}$$

Here $\{\tilde{\beta}_l^c(\tau), l = 1, \dots, n_1\}$ is another set of independent standard complex Wiener processes, which we now re-denote back to $\{\beta_l^c(\tau), l = 1, \dots, n_1\}$. We stress that the equation above is nothing but the original eq. (1.1), where its linear part (2.1) is conservative and non-degenerate in the sense of conditions (1) and (2), written in the complex coordinates and slow time. So all results below concerning eq. (2.6) (and eq. (2.11)) may be reformulated for eq. (1.1) on the price of heavier notation.

Everywhere below in this paper we assume the following assumption concerning the well-posedness of eq. (2.6):

Assumption 2.1. (1) The drift $P(v) = (P_1(v), \dots, P_n(v))$ is a locally Lipschitz vector field, belonging to $\mathrm{Lip}_{m_0}(\mathbb{C}^n, \mathbb{C}^n)$ for some $m_0 \ge 0$ (see (1.5)).

(2) For any $v_0 \in \mathbb{C}^n$, equation (2.6) has a unique strong solution $v^\varepsilon(\tau; v_0)$, $\tau \in [0, T]$, equal v_0 at $\tau = 0$. Moreover, there exists $m_0' > (m_0 \vee 1)$ such that

$$\mathbf{E} \sup_{0 \leqslant \tau \leqslant T} |v^\varepsilon(\tau; v_0)|^{2m_0'} \leqslant C_{m_0'}(|v_0|, T) < \infty \qquad \forall\, 0 < \varepsilon \le 1, \tag{2.7}$$

where $C_{m_0'}(\cdot)$ is a non-negative continuous function on $\mathbb{R}_+^2$, non-decreasing in both arguments.

Our proofs easily generalize to the case when the vector field P is locally Lipschitz, satisfying $|P(v)| \le C(1 + |v|)^{m_0}$ for all v and some $C > 0$, $m_0 \ge 0$.[3] In this case the argument remains essentially the same (but become a bit longer), and the constants in estimates depend not only on m_0, but also on the locally Lipschitz constant of P, which is the function $R \mapsto \mathrm{Lip}\, P|_{\bar{B}_R(\mathbb{C}^n)}$.

Below $T > 0$ is fixed and the dependence of constants on T usually is not indicated. Solutions of eq. (2.6) are assumed to be strong, unless otherwise stated. As usual, strong solutions are understood in the sense of an integral equation. I.e., $v^\varepsilon(\tau; v_0) = v(\tau)$, $0 \le \tau \le T$, is a strong solution, equal v_0 at $\tau = 0$, if a.s.

$$v_k(\tau) + \int_0^\tau \big(i\varepsilon^{-1}\lambda_k v_k(s) - P_k(v(s))\big)ds = v_{0k} + \sum_{l=1}^{n_1} \Psi_{kl}\beta_l^c(\tau), \quad k = 1, \dots, n,$$

for $0 \le \tau \le T$.

2.3. **Interaction representation.** Now in eq. (2.6) we pass to the *interaction representation*, which means that we substitute

$$v_k(\tau) = e^{-i\tau\varepsilon^{-1}\lambda_k} a_k(\tau), \quad k = 1, \dots, n. \tag{2.8}$$

Then $v_k(0) = a_k(0)$ and we obtain the following equations for variables $a_k(\tau)$:

$$da_k(\tau) = e^{i\tau\varepsilon^{-1}\lambda_k} P_k(v) + e^{i\tau\varepsilon^{-1}\lambda_k} \sum_{l=1}^{n_1} \Psi_{kl} d\beta_l^c(\tau), \quad k = 1, \dots, n. \tag{2.9}$$

[3] See [13] for averaging in deterministic perturbation of eq. (2.1) by locally Lipschitz vector fields.

Actions $I_k = |a_k|^2/2$ for solutions of (2.9) are the same as the actions for solutions of (2.6). Compare to (2.6), in eq. (2.9) we have removed from the drift the large term $\varepsilon^{-1}\mathrm{diag}(i\Lambda)v$ on the price that now coefficients of the system are fast oscillating functions of τ.

To rewrite conveniently the equations above we introduce the rotation operators Φ_w: for any real vector $w = (w_1, \dots, w_n) \in \mathbb{R}^n$ we denote

$$\Phi_w : \mathbb{C}^n \to \mathbb{C}^n, \quad \Phi_w = \mathrm{diag}\{e^{iw_1}, \dots, e^{iw_n}\}. \tag{2.10}$$

Then $(\Phi_w)^{-1} = \Phi_{-w}$, $\Phi_{w_1} \circ \Phi_{w_2} = \Phi_{w_1+w_2}$, $\Phi_0 = \mathrm{id}$, each Φ_w is an unitary transformation, so $\Phi_w^* = \Phi_w^{-1}$. Moreover,

$$|(\Phi_w z)_j| = |z_j|, \qquad \forall z, w, j.$$

In terms of operators Φ we write $v(\tau)$ as $\Phi_{\tau\varepsilon^{-1}\Lambda}a(\tau)$, and write system (2.9) as

$$da(\tau) = \Phi_{\tau\varepsilon^{-1}\Lambda}P(\Phi_{-\tau\varepsilon^{-1}\Lambda}a(\tau))d\tau + \Phi_{\tau\varepsilon^{-1}\Lambda}\Psi d\beta^c(\tau), \quad a(\tau) \in \mathbb{C}^n, \tag{2.11}$$

where $\beta^c(\tau) = (\beta_1^c(\tau), \dots, \beta_{n_1}^c(\tau))$. This is the equation which we are going to study for small ε for $0 \leqslant \tau \leqslant T$ under the initial condition

$$a(0) = v(0) = v_0. \tag{2.12}$$

The solution $a^\varepsilon(\tau; v_0) = \Phi_{-\tau\varepsilon^{-1}\Lambda}v^\varepsilon(\tau; v_0)$ of (2.11), (2.12) also satisfies estimate (2.7), for each $\varepsilon \in (0, 1]$.

We recall that a C^1-diffeomorphism G of $\mathbb{C}^n$ transforms a vector filed V to the field G_*V, where $(G_*V)(v) = dG(u)\big(V(u)\big)$ for $u = G^{-1}v$. In particular,

$$\big((\Phi_{\tau\varepsilon^{-1}\Lambda})_*P\big)(v) = \Phi_{\tau\varepsilon^{-1}\lambda_k} \circ P(\Phi_{-\varepsilon\tau\Lambda}v).$$

So equation (2.11) may be written as

$$da(\tau) = \big((\Phi_{\tau\varepsilon^{-1}\Lambda})_*P\big)(a(\tau))d\tau + \Phi_{\tau\varepsilon^{-1}\Lambda}\Psi d\beta^c(\tau).$$

2.4. **The compactness.** For $0 < \varepsilon \leqslant 1$ we denote by $a^\varepsilon(\tau; v_0)$ a solution of equation (2.11), equals v_0 at $\tau = 0$. Then

$$a^\varepsilon(\tau; v_0) = \Phi_{\tau\varepsilon^{-1}\lambda_k}v^\varepsilon(\tau; v_0).$$

A unique solution $v^\varepsilon(\tau; v_0)$ of (2.6) exists by Assumption 2.1, so a solution $a^\varepsilon(\tau; v_0)$ as well exists and is unique. Our goal is to examine its law

$$Q_\varepsilon := \mathcal{D}(a^\varepsilon(\cdot; v_0)) \in \mathcal{P}(C([0, T], \mathbb{C}^n))$$

as $\varepsilon \to 0$. While v_0 is fixed we will usually write $a^\varepsilon(\tau; v_0)$ as $a^\varepsilon(\tau)$.

Lemma 2.2. *Under Assumption 2.1, the set of probability measures Q_ε, $\varepsilon \in (0, 1]\}$ is pre-compact in the space $\mathcal{P}(C([0, T], \mathbb{C}^n))$ with respect to the weak topology.*

Proof. Let us denote the random force in equation (2.11) by $d\zeta^\varepsilon(\tau) := \Phi_{\tau\varepsilon^{-1}\Lambda}\Psi d\beta^c(\tau)$, where $\zeta^\varepsilon(\tau) = (\zeta_l^\varepsilon(\tau), l = 1, \dots, n_1)$. For any k we have

$$\zeta_k^\varepsilon(\tau) = \int_0^\tau d\zeta_k^\varepsilon = \int_0^\tau e^{is\varepsilon^{-1}\lambda_k}\sum_{l=1}^{n_1}\Psi_{kl}d\beta_l^c(s).$$

So $\zeta^\varepsilon(\tau)$ is a stochastic integral of a non-random vector function. Hence, it is a Gaussian random process with zero mean value, and its increments over non-intersecting time-intervals are independent. For each k,

$$\mathbf{E}|\zeta_k^\varepsilon(\tau)|^2 = \int_0^\tau 2\sum_{l=1}^{n_1}|\Psi_{kl}|^2 ds =: 2C_k^\zeta\tau, \qquad C_k^\zeta = \sum_{l=1}^{n_1}|\Psi_{kl}|^2 \geq 0,$$

and $\mathbf{E}\zeta_k^\varepsilon(\tau)\zeta_j^\varepsilon(\tau) = \mathbf{E}\bar{\zeta}_k^\varepsilon(\tau)\bar{\zeta}_j^\varepsilon(\tau) = 0$. Therefore $\zeta_k^\varepsilon(\tau) = C_k^\zeta \beta_k^c(\tau)$, where by Lévy's theorem (see [14, p. 157]) $\beta_k^c(\tau)$ is a standard complex Wiener process. However, the processes ζ_j^ε and ζ_k^ε with $j \neq k$ are not necessarily independent.

By the basic properties of Wiener process, a.s. the curve $[0,T] \ni \tau \mapsto \zeta^\varepsilon(\omega,\tau) \in \mathbb{C}^n$ is Hölder-continuous with exponent $\frac{1}{3}$, and since C_k^ζ does not depend on ε, then abbreviating $(C^{1/3}([0,T],\mathbb{C}^n)$ to $C^{1/3}$ we have

$$\mathbf{P}\big(\zeta^\varepsilon(\cdot) \in \bar{B}_R(C^{1/3})\big) \to 1 \ \text{ as } \ R \to \infty,$$

uniformly in ε. Let us write equation (2.11) as

$$da^\varepsilon(\tau) = V^\varepsilon(\tau)d\tau + d\zeta^\varepsilon(\tau).$$

By Assumption 2.1 and since $|a^\varepsilon(\tau)| \equiv |v^\varepsilon(\tau)|$ we have that

$$\mathbf{E} \sup_{\tau\in[0,T]} |V^\varepsilon(\tau)| \leqslant \mathcal{C}^{m_0}(P)\mathbf{E}(1 + \sup_{\tau\in[0,T]} |v^\varepsilon(\tau)|)^{m_0} \leqslant C(|v_0|) < \infty.$$

Therefore, by Chebyshev's inequality,

$$\mathbf{P}(\sup_{0\leqslant\tau\leqslant T} |V^\varepsilon(\tau)| > R) \leqslant C(|v_0|)R^{-1},$$

uniformly in $\varepsilon \in (0,1]$. Since

$$a^\varepsilon(\tau) = v_0 + \int_0^\tau V^\varepsilon(s)ds + \zeta^\varepsilon(\tau),$$

then we get from the above that

$$\mathbf{P}\big(\|a^\varepsilon(\cdot)\|_{1/3} > R\big) \to 0 \quad \text{as} \quad R \to \infty, \tag{2.13}$$

uniformly in $\varepsilon \in (0,1]$. By the Ascoli-Arzela theorem the sets $\bar{B}_R(C^{1/3})$ are compact in $C([0,T];\mathbb{C}^n)$, and in view of (2.13) for any $\delta > 0$ there exists R_δ such that

$$Q_\varepsilon\big(\bar{B}_{R_\delta}(C^{1/3})\big) \geqslant 1 - \delta, \quad \forall \varepsilon > 0.$$

So by Prokhorov's theorem the set of measures $\{Q_\varepsilon, 0 < \varepsilon \leqslant 1\}$ is pre-compact in $\mathcal{P}(C[0,T],\mathbb{C}^n)$. □

By this lemma, for any sequence $\varepsilon_l \to 0$ there exists a subsequence $\varepsilon_l' \to 0$ and a measure $Q_0 \in \mathcal{P}(C([0,T],\mathbb{C}^n))$ such that

$$Q_{\varepsilon_l'} \rightharpoonup Q_0 \quad \text{as} \ \ \varepsilon_l' \to 0. \tag{2.14}$$

3. Averaging of vector fields with respect to the frequency vector

For a vector field $\tilde{P} \in \mathrm{Lip}_{m_0}(\mathbb{C}^n,\mathbb{C}^n)$ we denote

$$Y_{\tilde{P}}(a;t) = \big((\Phi_{t\Lambda})_*\tilde{P}\big)(a) = \Phi_{t\Lambda} \circ \tilde{P}(\Phi_{-t\Lambda}a), \quad a \in \mathbb{C}^n, \ t \in \mathbb{R},$$

and for $T' > 0$ define *partial averaging* $\langle\!\langle \tilde{P} \rangle\!\rangle^{T'}$ of the vector field $\tilde{P}$ with respect to the frequency vector Λ as follows:

$$\langle\!\langle \tilde{P} \rangle\!\rangle^{T'}(a) = \frac{1}{T'}\int_0^{T'} Y_{\tilde{P}}(a;t)dt = \frac{1}{T'}\int_0^{T'} \Phi_{t\Lambda} \circ \tilde{P}(\Phi_{-t\Lambda}a)dt. \tag{3.1}$$

Lemma 3.1. *For any $T' > 0$, $\langle\!\langle \tilde{P} \rangle\!\rangle^{T'}(a) \in Lip_{m_0}(\mathbb{C}^n,\mathbb{C}^n)$ and $\mathcal{C}^{m_0}(\langle\!\langle \tilde{P} \rangle\!\rangle^{T'}) \leq \mathcal{C}^{m_0}(\tilde{P})$ (see (1.5)).*

Proof. If $a \in \bar{B}_R(\mathbb{C}^n)$, then $\Phi_{-t\Lambda}a \in \bar{B}_R(\mathbb{C}^n)$ for each t. So $|Y_{\tilde{P}}(a;t)| = |(\Phi_{t\Lambda})_*\tilde{P}(a)| = |\tilde{P}(\Phi_{-t\Lambda}a))|$, and from here

$$|\langle\!\langle \tilde{P}\rangle\!\rangle^{T'}(a)| \leqslant \sup_{0\le t\le T'} |Y_{\tilde{P}}(a;t)| \leqslant \mathcal{C}^{m_0}(\tilde{P})(1+R)^{m_0}.$$

Similarly, for any $a_1, a_2 \in \bar{B}_R(\mathbb{C}^n)$,

$$|Y_{\tilde{P}}(a_1;t) - Y_{\tilde{P}}(a_2;t)| = |\tilde{P}(\Phi_{-t\Lambda}a_1) - \tilde{P}(\Phi_{-t\Lambda}a_2)| \leqslant \mathcal{C}^{m_0}(\tilde{P})(1+R)^{m_0}|a_2 - a_1| \quad \forall t \geqslant 0,$$

so that

$$|\langle\!\langle \tilde{P}\rangle\!\rangle^{T'}(a_1) - \langle\!\langle \tilde{P}\rangle\!\rangle^{T'}(a_1)| \leqslant \mathcal{C}^{m_0}(\tilde{P})(1+R)^{m_0}|a_1 - a_1|.$$

This proves the assertion. □

Now we define *averaging* of the vector field $\tilde{P}$ with respect to the frequency vector Λ as

$$\langle\!\langle \tilde{P}\rangle\!\rangle(a) = \lim_{T'\to\infty} \langle\!\langle \tilde{P}\rangle\!\rangle^{T'}(a) = \lim_{T'\to\infty} \frac{1}{T'}\int_0^{T'} (\Phi_{t\Lambda})_*\tilde{P}(a)dt. \tag{3.2}$$

Lemma 3.2. *(1) The limit (3.2) exists for any a. Moreover, $\langle\!\langle \tilde{P}\rangle\!\rangle$ belongs to $Lip_{m_0}(\mathbb{C}^n, \mathbb{C}^n)$ and $\mathcal{C}^{m_0}(\langle\!\langle \tilde{P}\rangle\!\rangle) \le \mathcal{C}^{m_0}(\tilde{P})$.*

(2) If $a \in \bar{B}_R(\mathbb{C}^n)$, then the rate of convergence in (3.2) depends not on a, but only on R.

This is the main lemma of deterministic averaging for vector fields. For its proof see [13, Lemma 3.1].[4]

The averaged vector field $\langle\!\langle \tilde{P}\rangle\!\rangle$ is invariant with respect to transformations $\Phi_{\theta\Lambda}$:

Lemma 3.3. *For all $a \in \mathbb{C}^n$ and $\theta \in \mathbb{R}$,*

$$\big(\Phi_{\theta\Lambda}\big)_*\langle\!\langle \tilde{P}\rangle\!\rangle(a) \equiv \Phi_{\theta\Lambda} \circ \langle\!\langle \tilde{P}\rangle\!\rangle \circ \Phi_{-\theta\Lambda}(a) = \langle\!\langle \tilde{P}\rangle\!\rangle(a).$$

Proof. For definiteness let $\theta > 0$. For any $T' > 0$ we have:

$$\langle\!\langle \tilde{P}\rangle\!\rangle^{T'}(\Phi_{-\theta\Lambda}(a)) = \frac{1}{T'}\int_0^{T'} \Phi_{t\Lambda} \circ \tilde{P}(\Phi_{-t\Lambda} \circ \Phi_{-\theta\Lambda}(a))dt = \frac{1}{T'}\int_0^{T'} \Phi_{t\Lambda} \circ \tilde{P}(\Phi_{-(t+\theta)\Lambda}a)dt.$$

Since $\Phi_{t\Lambda} = \Phi_{-\theta\Lambda} \circ \Phi_{(t+\theta)\Lambda}$, then this equals

$$\frac{1}{T'}\Phi_{-\theta\Lambda}\Big(\int_0^{T'} \Phi_{(t+\theta)\Lambda} \circ \tilde{P}(\Phi_{-(t+\theta)\Lambda}a)dt\Big) = \Phi_{-\theta\Lambda} \circ \langle\!\langle \tilde{P}\rangle\!\rangle^{T'}(a) + O\Big(\frac{1}{T'}\Big).$$

Passing to the limit as $T' \to \infty$ we obtain the assertion. □

The statement below asserts that the averaged vector filed $\langle\!\langle P\rangle\!\rangle$ is as least as smooth as P.

Proposition 3.4. *If $P \in C^m(\mathbb{C}^n)$ for some $m \in \mathbb{N}$, then $\langle\!\langle P\rangle\!\rangle \in C^m(\mathbb{C}^n)$ and $|\langle\!\langle P\rangle\!\rangle|_{C^m(\bar{B}_R)} \leqslant |P|_{C^m(\bar{B}_R)}$, $\forall R > 0$.*

Proof. We first fix any $R > 0$. Then there exists a sequence of polynomial vector fields $\{P_{R,j}, j \in \mathbb{N}\}$ (cf. below item 3) in Section 3.1) such that $|P_{R,j} - P|_{C^m(\bar{B}_R)} \to 0$ as $j \to \infty$. An easy calculation shows that

$$|\langle\!\langle P_{R,j}\rangle\!\rangle^T - \langle\!\langle P_{R,j}\rangle\!\rangle|_{C^m(\bar{B}_R)} \to 0 \quad \text{as } T \to \infty, \tag{3.3}$$

[4] In fact, if $\tilde{P}$ is any locally Lipschitz vector field, then $\langle\!\langle \tilde{P}\rangle\!\rangle$ is well defined and locally Lipschitz with the same locally Lipschitz constant (see discussion after Assumption 2.1) as $\tilde{P}$. See in [13].

for each j. Since the transformations $\Phi_{t\Lambda}$ are unitary, then differentiating the integral in (3.1) in a we get that

$$|\langle\!\langle \tilde{P} \rangle\!\rangle^T|_{C^m(\bar{B}_R)} \leqslant |\tilde{P}|_{C^m(\bar{B}_R)} \quad \forall\, T > 0, \tag{3.4}$$

for any C^m-vector field $\tilde{P}$. Therefore

$$|\langle\!\langle P_{R,j} \rangle\!\rangle^T - \langle\!\langle P \rangle\!\rangle^T|_{C^m(\bar{B}_R)} \leq |P_{R,j} - P|_{C^m(\bar{B}_R)} =: \kappa_j \to 0 \quad \text{as } j \to \infty, \quad \forall\, T > 0. \tag{3.5}$$

So

$$|\langle\!\langle P_{R,j} \rangle\!\rangle^T - \langle\!\langle P_{R,k} \rangle\!\rangle^T|_{C^m(\bar{B}_R)} \leq 2\kappa_{j\wedge k} \quad \forall\, T > 0.$$

From this estimate and (3.3) we find that $|\langle\!\langle P_{R,j} \rangle\!\rangle - \langle\!\langle P_{R,k} \rangle\!\rangle|_{C^m(\bar{B}_R)} \leq 2\kappa_{j\wedge k}$. Thus $\{\langle\!\langle P_{R,j} \rangle\!\rangle\}$ is a Cauchy sequence in $C^m(\bar{B}_R)$. So it C^m-converges to a limiting field $\langle\!\langle P_{R,\infty} \rangle\!\rangle$. As $P_{R,j}$ converges to P in $C^m(\bar{B}_R)$, then using again (3.4) we find that $|\langle\!\langle P_{R,\infty} \rangle\!\rangle|_{C^m(\bar{B}_R)} \leq |P|_{C^m(\bar{B}_R)}$. But by Lemma 3.2 $\langle\!\langle P_{R,\infty} \rangle\!\rangle$ must equal $\langle\!\langle P \rangle\!\rangle$. Since $R > 0$ is arbitrary, the assertion of the proposition follows. □

Finally we note that if a vector field P is Hamiltonian, then its averaging $\langle\!\langle P \rangle\!\rangle$ also is. Looking ahead we state the corresponding result here, despite the averaging of functions $\langle\cdot\rangle$ is defined below in Section 3.2.

Proposition 3.5. *If a locally Lipschitz vector field P is Hamiltonian, i.e. $P(z) = i\frac{\partial}{\partial \bar{z}} H(z)$ for some C^1-function H, then $\langle\!\langle P \rangle\!\rangle$ also is Hamiltonian and $\langle\!\langle P \rangle\!\rangle = i\frac{\partial}{\partial \bar{z}}\langle H \rangle$.*

For a proof see [13, Theorem 5.2].

3.1. **Calculating the averagings. 1)** The frequency vector $\Lambda = (\lambda_1, \dots, \lambda_n)$ is called *completely resonant* if its components λ_j are proportional to some $\lambda > 0$. I.e. if $\lambda_j/\lambda \in \mathbb{Z}$ for all j. In this case all trajectories of the original linear system (2.1) are periodic, the operator $\Phi_{t\Lambda}$ is $2\pi/\lambda$-periodic in t and so

$$\langle\!\langle \tilde{P} \rangle\!\rangle(a) = \langle\!\langle \tilde{P} \rangle\!\rangle^{2\pi/\lambda}(a) = \frac{\lambda}{2\pi} \int_0^{2\pi/\lambda} (\Phi_{t\Lambda})_* \tilde{P}(a) dt. \tag{3.6}$$

Completely resonant linear systems (2.1) and their perturbations (1.1) often occur in applications. In particular – in non-equilibrium statistical physics. There the dimension $D = 2n$ is large, all λ_j's are equal, and the Wiener process $W(t)$ in (1.1) may be very degenerate (it may have only two non-zero components). E.g. see [9], where more references may be found.

2) Let us consider the case, opposite to the above and assume that frequency vector Λ is *non-resonant*:

$$\sum_{i=1}^n m_j \lambda_j \neq 0 \qquad \forall (m_1, \dots, m_n) \in \mathbb{Z}^n \setminus \{0\} \tag{3.7}$$

(that is, the real numbers λ_j are rationally independent). Then

$$\langle\!\langle \tilde{P} \rangle\!\rangle(a) = \frac{1}{(2\pi)^n} \int_{\mathbb{T}^n} (\Phi_w)_* \tilde{P}(a)\, dw, \qquad \mathbb{T}^n = \mathbb{R}^n/(2\pi\mathbb{Z}^n). \tag{3.8}$$

Indeed, if $\tilde{P}$ is a polynomial vector field, then (3.8) easily follows from (3.2) by a direct component-wise calculation. The general case is a consequence of this result since any vector field may be approximated by polynomial fields. Details are left to the reader (cf. Lemma 3.5 in [13], where $\tilde{P}^{res}$ equals to the r.h.s. of (3.8) if the vector Λ is non-resonant).

The r.h.s. of (3.8) obviously is invariant with respect to all rotations $\Phi_{w'}$, so it depends not on the vector a, but only on the corresponding torus

$$\{z \in \mathbb{C}^n : I_j(z) = I_j(a) \quad \forall\, j\} \tag{3.9}$$

(see (1.2)) to which a belongs,

$$\big(\Phi_w\big)_* \langle\!\langle \tilde{P} \rangle\!\rangle(a) \equiv \langle\!\langle \tilde{P} \rangle\!\rangle(a) \qquad \forall w \in \mathbb{C}^n \quad \text{if } \Lambda \text{ is non-resonant.} \tag{3.10}$$

See below Section 6 for discussing equations (1.1) with non-resonant vectors Λ.

3) If the field $\tilde{P}$ in (3.2) is polynomial, i.e.

$$\tilde{P}_j(a) = \sum_{|\alpha|,|\beta| \le N} C_j^{\alpha,\beta} a^\alpha \bar{a}^\beta, \qquad j = 1, \dots, n, \tag{3.11}$$

for some $N \in \mathbb{N}$, where $\alpha, \beta \in \mathbb{Z}_+^n$, $a^\alpha = \prod a_j^{\alpha_j}$ and $|\alpha| = \sum |\alpha_j|$, then $\langle\!\langle \tilde{P} \rangle\!\rangle = \tilde{P}^{res}$. Here $\tilde{P}^{res}$ is a polynomial vector field such that for each j, $\tilde{P}_j^{res}(a)$ is given by the r.h.s. of (3.11), where the summation is taken over all $|\alpha|, |\beta| \le N$, satisfying $\Lambda \cdot (\alpha - \beta) = \lambda_j$. This easily follows from an explicit calculation of the integral in (3.1) (see [13, Lemma 3.5]).

3.2. **Averaging of functions.** Similarly to definition (3.2), for a locally Lipschitz function $f \in \mathrm{Lip}_m(\mathbb{C}^n, \mathbb{C})$, $m \ge 0$, we define its averaging with respect to a frequency vector Λ as

$$\langle f \rangle(a) = \lim_{T' \to \infty} \frac{1}{T'} \int_0^{T'} f(\Phi_{-t\Lambda} a) dt, \quad a \in \mathbb{C}^n. \tag{3.12}$$

Then by the same argument as above we obtain:

Lemma 3.6. *If $f \in Lip_m(\mathbb{C}^n, \mathbb{C})$, then*

1. *The limit (3.12) exists for every a, and for $a \in \bar{B}_R(\mathbb{C}^n)$ the rate of convergence in (3.12) depends not on a, but only on R.*
2. *$\langle f \rangle \in Lip_m(\mathbb{C}^n; \mathbb{C})$ and $\mathcal{C}^m(\langle f \rangle) \le \mathcal{C}^m(f)$.*
3. *If f is C^m-smooth for some $m \in \mathbb{N}$, then $\langle f \rangle$ also is, and the C^m-norm of the latter is bounded by the C^m-norm of the former.*
4. *The function $\langle f \rangle$ commutes with the operators $\Phi_{\theta\Lambda}$, $\theta \in \mathbb{R}$, in the sense that $\langle f \circ \Phi_{\theta\Lambda} \rangle = \langle f \rangle \circ \Phi_{\theta\Lambda} = \langle f \rangle$.*

If the vector Λ is non-resonant, then similar to (3.8) we have

$$\langle f \rangle(a) = \frac{1}{(2\pi)^n} \int_{\mathbb{T}^n} f(\Phi_{-w} a)\, dw. \tag{3.13}$$

The r.h.s. of (3.13) is the *averaging of function f in angles*. It is constant on the tori (3.9).

4. Effective equation and the averaging theorem

In this section we show that the limiting measure Q_0 in (2.14) is independent of the choice of the sequence $\varepsilon_l' \to 0$, so $\mathcal{D}(a^\varepsilon) \rightharpoonup Q_0$ as $\varepsilon \to 0$, and represent Q_0 as the law of a solution of an auxiliary *effective equation*. The drift in this equation is the averaged drift in eq. (2.6). Now we construct its dispersion.

The diffusion matrix for equation (2.11) is the $n \times n$ complex matrix

$$\mathcal{A}^\varepsilon(\tau) = (\Phi_{\tau\varepsilon^{-1}\Lambda}\Psi) \cdot (\Phi_{\tau\varepsilon^{-1}\Lambda}\Psi)^*.$$

Denoting

$$\Phi_{\tau\varepsilon^{-1}\Lambda}\Psi = \big(e^{i\tau\varepsilon^{-1}\lambda_l}\Psi_{lj}\big) =: \big(\psi_{lj}^\varepsilon(\tau)\big) = \psi^\varepsilon(\tau), \tag{4.1}$$

we have

$$\mathcal{A}^{\varepsilon}_{kj}(\tau) = \sum_{l=1}^{n_1} \psi^{\varepsilon}_{kl}(\tau)\bar{\psi}^{\varepsilon}_{jl}(\tau) = e^{i\tau\varepsilon^{-1}(\lambda_k-\lambda_j)} \sum_{l=1}^{n_1} \Psi_{kl}\bar{\Psi}_{jl}.$$

So for any $\tau > 0$,

$$\frac{1}{\tau}\int_0^{\tau} \mathcal{A}^{\varepsilon}_{kj}(s)ds = \big(\sum_{l=1}^{n_1} \Psi_{kl}\bar{\Psi}_{jl}\big)\frac{1}{\tau}\int_0^{\tau} e^{is\varepsilon^{-1}(\lambda_k-\lambda_j)}ds,$$

and we immediately see that

$$\frac{1}{\tau}\int_0^{\tau} \mathcal{A}^{\varepsilon}_{kj}(\tau)d\tau \to A_{kj} \quad \text{as} \quad \varepsilon \to 0, \tag{4.2}$$

where

$$A_{kj} = \begin{cases} \sum_{l=1}^{n_1} \Psi_{kl}\bar{\Psi}_{jl}, & \text{if } \lambda_k = \lambda_j, \\ 0, & \text{otherwise.} \end{cases} \tag{4.3}$$

Clearly, $A_{kj} = \bar{A}_{jk}$, so A is a Hermitian matrix. If $\lambda_k \neq \lambda_j$ for $k \neq j$, then

$$A = \operatorname{diag}\{b_1, \dots, b_n\}, \quad b_k = \sum_{l=1}^{n_1} |\Psi_{kl}|^2. \tag{4.4}$$

For any vector $\xi \in \mathbb{C}^n$ we get from (4.2) that $\langle A\xi, \xi\rangle \geqslant 0$ since obviously $\langle \mathcal{A}^{\varepsilon}(\tau)\xi, \xi\rangle = |\psi^{\varepsilon}(\tau)\xi|^2 \geqslant 0$ for each ε. Therefore A is a nonnegative Hermitian matrix, and there exists another nonnegative Hermitian matrix B (called the principal square root of A) such that $BB^* = B^2 = A$. The matrix B is non-degenerate if Ψ is.

Example 4.1. If Ψ is a diagonal matrix diag $\{\psi_1, \dots, \psi_n\}$, $\psi_j \in \mathbb{R}$, then $\mathcal{A}^{\varepsilon}(\tau) = |\Psi|^2$. In this case $A = |\Psi|^2$ and $B = |\Psi|$ =diag $\{|\psi_1|, \dots, |\psi_n|\}$.

In fact, it is not needed that the matrix B is square and Hermitian, and the argument below remains true if we take for B any complex $n \times N$-matrix (with any $N \in \mathbb{N}$), satisfying the equation

$$BB^* = A.$$

Now we define the *effective equation* for eq. (2.11) as follows:

$$da_k - \langle\!\langle P\rangle\!\rangle_k(a)d\tau = \sum_{l=1}^{n} B_{kl}d\beta^c_l, \quad k = 1, \dots, n. \tag{4.5}$$

Here the matrix B is as above and $\langle\!\langle P\rangle\!\rangle$ is the resonant averaging of the vector field P. We will usually consider this equation with the same initial condition as equations (2.6) and (2.11):

$$a(0) = v_0. \tag{4.6}$$

Since the vector field $\langle\!\langle P\rangle\!\rangle$ is locally Lipschitz and the dispersion matrix B is constant, then a strong solution of eq. (4.5), (4.6), if exists, is unique.

Note that the effective dispersion B in (4.5) is a square root of an explicit matrix, and due to item 3) of Section 3.1 if the vector field $P(v)$ is polynomial, then the effective drift $\langle\!\langle P\rangle\!\rangle(a)$ also is given by an explicit formula.

Proposition 4.2. *The limiting probability measure Q_0 in (2.14) is a weak solution of effective equation (4.5), (4.6).*

We recall that a measure $Q \in \mathcal{P}(C([0,T],\mathbb{C}^n))$ is a *weak solution* of equation (4.5), (4.6) if $Q = \mathcal{D}(\tilde{a})$, where the random process $\tilde{a}(\tau)$, $0 \le \tau \le T$, is a weak solution of (4.5), (4.6). [5]

A proof of this result is preceded by a number of lemmas. Till the end of this section we assume Assumption 2.1. As in Section 3 we denote

$$Y(a;\tau\varepsilon^{-1}) := (\Phi_{\tau\varepsilon^{-1}\Lambda})_* P(a). \tag{4.7}$$

Then equation (2.11) for a^ε reads as

$$da^\varepsilon(\tau) - Y(a^\varepsilon,\tau\varepsilon^{-1})d\tau = \Phi_{\tau\varepsilon^{-1}\Lambda}\Psi d\beta^c(\tau). \tag{4.8}$$

Denote

$$\tilde{y}(a,\tau\varepsilon^{-1}) = Y(a,\tau\varepsilon^{-1}) - \langle\!\langle P\rangle\!\rangle(a) = (\Phi_{\tau\varepsilon^{-1}\Lambda})_* P(a) - \langle\!\langle P\rangle\!\rangle(a). \tag{4.9}$$

The following key lemma shows that integrals of $\tilde{y}(a^\varepsilon,\tau\varepsilon^{-1})$ with respect to τ become small with ε, uniformly in the segment of integrating.

Lemma 4.3. *For a solution $a^\varepsilon(\tau)$ of equation (2.11), (2.12) we have*

$$\mathbf{E}\max_{0\leqslant\tau\leqslant T}\Big|\int_0^\tau \tilde{y}(a^\varepsilon(s),s\varepsilon^{-1})ds\Big| \to 0 \quad as \quad \varepsilon\to 0.$$

The lemma is proved in Subsection 4.1.

Now let us introduce a natural filtered measurable space for the problem we consider,

$$(\tilde{\Omega},\mathcal{B},\{\mathcal{B}_\tau, 0\leqslant\tau\leqslant T\}), \tag{4.10}$$

where $\tilde{\Omega}$ is the Banach space $C([0,T],\mathbb{C}^n) = \{a := a(\cdot)\}$, $\mathcal{B}$ is its Borel σ-algebra and $\mathcal{B}_\tau$ is the sigma-algebra, generated by the r.v.'s $\{a(s) : 0\leqslant s\leqslant\tau)\}$. Consider the process on $\tilde{\Omega}$, defined by the l.h.s. of (4.5)

$$N^{\langle\!\langle P\rangle\!\rangle}(\tau;a) = a(\tau) - \int_0^\tau \langle\!\langle P\rangle\!\rangle(a(s))ds, \qquad a\in\tilde{\Omega},\ \tau\in[0,T]. \tag{4.11}$$

Note that for any $0\le\tau\le T$, $N^{\langle\!\langle P\rangle\!\rangle}(\tau;\cdot)$ is a $\mathcal{B}_\tau$-measurable continuous functional on $C([0,T];\mathbb{C}^n)$.

Lemma 4.4. *The random process $N^{\langle\!\langle P\rangle\!\rangle}(\tau;a)$ is a martingale on the space (4.10) with respect to the limiting measure Q_0 in (2.14).*

Proof. Let us fix a $\tau\in[0,T]$ and take a $\mathcal{B}_\tau$-measurable function $f^\tau\in C_b(\tilde{\Omega})$. We will show that

$$\mathbf{E}^{Q_0}\Big(N^{\langle\!\langle P\rangle\!\rangle}(t;a)f^\tau(a)\Big) = \mathbf{E}^{Q_0}\Big(N^{\langle\!\langle P\rangle\!\rangle}(\tau;a)f^\tau(a)\Big) \quad \text{for any } \tau\leqslant t\leqslant T, \tag{4.12}$$

which would imply the assertion. To establish this let us first consider the process

$$N^{Y,\varepsilon}(\tau;a^\varepsilon) := a^\varepsilon(\tau) - \int_0^\tau Y(a^\varepsilon,s\varepsilon^{-1})ds,$$

who is a martingale in view of (4.8). As

$$N^{Y,\varepsilon}(\tau;a^\varepsilon) - N^{\langle\!\langle P\rangle\!\rangle}(\tau;a^\varepsilon) = \int_0^\tau \big[\langle\!\langle P\rangle\!\rangle(a^\varepsilon(s)) - Y(a^\varepsilon(s),s\varepsilon^{-1})\big]ds,$$

then by Lemma 4.3,

$$\max_{0\le\tau\le T}\mathbf{E}\big|N^{Y,\varepsilon}(\tau;a^\varepsilon) - N^{\langle\!\langle P\rangle\!\rangle}(\tau;a^\varepsilon)\big| = o_\varepsilon(1). \tag{4.13}$$

[5] Concerning weak solutions of SDEs e.g. see e.g. [14, Section 5.3].

Here and below in this proof $o_\varepsilon(1)$ stands for a quantity which goes to zero with ε. Since $N^{Y,\varepsilon}$ is a martingale, then by this relation

$$\mathbf{E}\big(N^{\langle\!\langle P\rangle\!\rangle}(t;a^\varepsilon)f^\tau(a^\varepsilon)\big)+o_\varepsilon(1)=\mathbf{E}\big(N^{Y,\varepsilon}(t;a^\varepsilon)f^\tau(a^\varepsilon)\big)$$
$$=\mathbf{E}\big(N^{Y,\varepsilon}(\tau;a^\varepsilon)f^\tau(a^\varepsilon)\big)=\mathbf{E}\big(N^{\langle\!\langle P\rangle\!\rangle}(\tau;a^\varepsilon)f^\tau(a^\varepsilon)\big)+o_\varepsilon(1).$$

So

$$\mathbf{E}^{Q_\varepsilon}\big(N^{\langle\!\langle P\rangle\!\rangle}(t;a)f^\tau(a)-N^{\langle\!\langle P\rangle\!\rangle}(\tau;a)f^\tau(a)\big)=o_\varepsilon(1). \tag{4.14}$$

To get (4.12) we will pass in this relation to a limit as $\varepsilon\to 0$. For this end for $M>0$ consider the function

$$G_M(t)=\begin{cases} t, & \text{if } |t|\le M,\\ M\operatorname{sgn} t, & \text{otherwise.}\end{cases}$$

Since by Assumption 2.1 and Lemma 3.2

$$\mathbf{E}^{Q_\varepsilon}\big(\sup_{\tau\in[0,T]}|N^{\langle\!\langle P\rangle\!\rangle}(\tau;a)|^2\big)\leqslant\mathbf{E}^{Q_\varepsilon}\big(C_P(1+\sup_{\tau\in[0,T]}|a(\tau)|^{2(m_0\vee 1)})\big)\leqslant C_{P,m_0}(|v_0|),$$

then for any ε we have

$$\mathbf{E}^{Q_\varepsilon}\big|(1-G_M)\circ\big(N^{\langle\!\langle P\rangle\!\rangle}(t;a)f^\tau(a)-N^{\langle\!\langle P\rangle\!\rangle}(\tau;a)f^\tau(a)\big)\big|\le CM^{-1}. \tag{4.15}$$

As $Q_{\varepsilon'_l}\rightharpoonup Q_0$, then by Fatou lemma this estimate stays true for $\varepsilon=0$.

Relations (4.14) and (4.15) show that

$$\mathbf{E}^{Q_\varepsilon}\big(G_M\circ(N^{\langle\!\langle P\rangle\!\rangle}(t;a)f^\tau(a)-N^{\langle\!\langle P\rangle\!\rangle}(\tau;a)f^\tau(a))\big)=o_\varepsilon(1)+o_{M^{-1}}(1).$$

From this and convergence (2.14) we derive the relation

$$\mathbf{E}^{Q_0}\big(G_M\circ(N^{\langle\!\langle P\rangle\!\rangle}(t;a)f^\tau(a)-N^{\langle\!\langle P\rangle\!\rangle}(\tau;a)f^\tau(a))\big)=o_{M^{-1}}(1),$$

which jointly with estimate $(4.15)_{\varepsilon=0}$ imply (4.12) when we send M to ∞. The lemma is proved. □

Definition 4.5. A measure Q on the space (4.10) is called a solution of the martingale problem for effective equation (4.5) with the initial condition (4.6) if $a(0)=v_0$ Q–a.s. and

1) the process $\{N^{\langle\!\langle P\rangle\!\rangle}(\tau;a)\in\mathbb{C}^n,\ \tau\in[0,T]\}$ (see (4.11)) is a vector–martingale on the filtered space (4.10) with respect to the measure Q;

2) for any $k,j=1,\dots,n$ the process

$$N_k^{\langle\!\langle P\rangle\!\rangle}(\tau;a)\overline{N_j^{\langle\!\langle P\rangle\!\rangle}}(\tau;a)-2\int_0^\tau(BB^*)_{kj}ds,\quad \tau\in[0,T] \tag{4.16}$$

(where $BB^*=A$) is a martingale on the space (4.10) with respect to the measure Q, as well as the process $N_k^{\langle\!\langle P\rangle\!\rangle}(\tau;a)N_j^{\langle\!\langle P\rangle\!\rangle}(\tau;a)$.[6]

This is a classical definition, written in complex coordinates. See [24] and [14, Section 5.4], where we profited from [14, Remark 4.12] and the result of [14, Problem 4.13] since by Lemma 3.2 the vector field $\langle\!\langle P\rangle\!\rangle$ in (4.5) is locally Lipschitz.

We have

Lemma 4.6. *The limiting measure Q_0 in (2.14) is a solution of the martingale problem for effective equation (4.5), (4.6).*

[6] That is, $\langle N_k^{\langle\!\langle P\rangle\!\rangle}(\tau;a),\overline{N_j^{\langle\!\langle P\rangle\!\rangle}}(\tau;a)\rangle(\tau)=2\int_0^\tau(BB^*)_{kj}$ and $\langle N_k^{\langle\!\langle P\rangle\!\rangle}(\tau;a),N_j^{\langle\!\langle P\rangle\!\rangle}(\tau;a)\rangle(\tau)=0$. See Appendix B.

Proof. Since condition 1) in Definition 4.5 has been verified in Lemma 4.4, it remains to check condition 2). For the second term in (4.16) we have, as $\varepsilon \to 0$,

$$\int_0^\tau \big(\psi^\varepsilon(s)(\psi^\varepsilon(s))^*\big)_{kj} ds = \int_0^\tau e^{i\varepsilon^{-1}(\lambda_k-\lambda_j)s}(\psi\psi^*)_{kj} ds \to \tau A_{kj}, \tag{4.17}$$

where the matrix (A_{kj}) is given by (4.3). Let us pass to the first term. Since by (2.11) and (4.1)

$$N^{Y,\varepsilon}(\tau) = v_0 + \int_0^\tau \psi^\varepsilon(s) d\beta^c(s), \qquad \psi^\varepsilon_{lj}(s) = e^{is\varepsilon^{-1}\lambda_l}\Psi_{lj},$$

then by the complex Itô formula (see Appendix C) and Assumption 2.1, for any $k, j \in \{1, \dots, n\}$ the process

$$N^{Y,\varepsilon}_k(\tau)\overline{N^{Y,\varepsilon}_j}(\tau) - 2\int_0^\tau (\psi^\varepsilon(s)(\psi^\varepsilon(s))^*)_{kj} ds, \tag{4.18}$$

is a martingale. As when verifying 1), we will compare (4.16) with (4.18). To do this let us consider

$$\begin{aligned} N^{\langle\!\langle P\rangle\!\rangle}_k(\tau; a^\varepsilon)\overline{N^{\langle\!\langle P\rangle\!\rangle}_j}(\tau; a^\varepsilon) &- N^{Y,\varepsilon}_k(\tau; a^\varepsilon)\overline{N^{Y,\varepsilon}_j}(\tau; a^\varepsilon) \\ &= \Big(a^\varepsilon_k(\tau) - \int_0^\tau \langle\!\langle P\rangle\!\rangle_k(a^\varepsilon(s))\, ds\Big)\Big(\bar a^\varepsilon_j(\tau) - \int_0^\tau \langle\!\langle \bar P\rangle\!\rangle_j(a^\varepsilon(s))\, ds\Big) \\ &- \Big(a^\varepsilon_k(\tau) - \int_0^\tau Y_k(a^\varepsilon(s), s\varepsilon^{-1})\, ds\Big)\Big(\bar a^\varepsilon_j(\tau) - \int_0^\tau \bar Y_j(a^\varepsilon(s), s\varepsilon^{-1})\, ds\Big) =: M_{kj}(a^\varepsilon, \tau). \end{aligned}$$

Closely repeating the proof of (4.13) we get that

$$\sup_{0\le\tau\le T} \mathbf{E}\big|M_{kj}(a^\varepsilon; \tau)\big| = o_\varepsilon(1) \quad \text{as } \varepsilon \to 0.$$

Since (4.18) is a martingale, then this relation and (4.17) imply that (4.16) is a martingale due to the same reasoning by which relations (4.13) and the fact that $N^{Y,\varepsilon}(\tau; a^\varepsilon)$ is a martingale imply that $N^{\langle\!\langle P\rangle\!\rangle}(\tau; a)$ also is one. To pass to a limit as $\varepsilon \to 0$ the proof uses that the random variables like $N^{Y,\varepsilon}_k(\tau; a^\varepsilon)\overline{N^{Y,\varepsilon}_j}(\tau; a^\varepsilon)$ are integrable uniformly in $\varepsilon > 0$ due to Assumption 2.1, where $m_0' > m_0$.

Similarly for any k and j, the process $N^{\langle\!\langle P\rangle\!\rangle}_k(\tau)N^{\langle\!\langle P\rangle\!\rangle}_j(\tau)$ also is a martingale. The lemma's assertion is established. □

Now we can prove Proposition 4.2:

Proof. [of Proposition 4.2] It is well known that a solution of the martingale problem for a stochastic differential equation is its weak solution. Instead of referring to a corresponding theorem (see [24] or [14, Section 5.4]), again following [16], we give a short direct proof, based on another strong result from the stochastic calculus. By Lemma 4.6 and the martingale representation theorem for complex processes (see Appendix B), we know that there exists an extension $(\hat\Omega, \hat{\mathcal{B}}, \hat{\mathbf{P}})$ of the probability space $(\tilde\Omega, \mathcal{B}, Q_0)$ and on it exist standard independent complex Wiener processes $\beta^c_1(\tau), \dots, \beta^c_n(\tau)$ such that

$$da_j(\tau) - \langle\!\langle P\rangle\!\rangle_j(a) d\tau = \sum_{l=1}^n B_{jl} d\beta^c_l(\tau), \quad j = 1, \dots, n,$$

where the dispersion B is a nonnegative Hermitian matrix, satisfying $BB^* = A$. Therefore the measure Q_0 is a weak solution of effective equation (4.5). We thus proved the assertion of the proposition. □

By Lemma 3.2, in effective equation (4.5) the drift term $\langle\!\langle P\rangle\!\rangle$ is locally Lipschitz. So its strong solution (if exists) is unique. By Proposition 4.2 the measure Q_0 is a weak solution of eq. (4.5). Hence, by the Yamada-Watanabe theorem [14, Section 5.3.D], [24, Chapter 8] a strong solution for the effective equation exists, and its weak solution is unique. Therefore the limit $Q_0 = \lim_{\varepsilon_l\to 0} Q_{\varepsilon_l}$ does not depend on the sequence $\varepsilon_l \to 0$. So the convergence holds as $\varepsilon \to 0$ and we thus have established

Theorem 4.7. *For any $v_0 \in \mathbb{C}^n$ the solution $a^\varepsilon(\tau; v_0)$ of problem (2.11), (2.12) satisfies*

$$\mathcal{D}(a^\varepsilon(\cdot; v_0)) \rightharpoonup Q_0 \text{ in } \mathcal{P}(C([0,T],\mathbb{C}^n)) \quad \text{as } \varepsilon \to 0, \tag{4.19}$$

where Q_0 is the law of a unique weak solution $a^0(\tau; v_0)$ of effective equation (4.5), (4.6).

Remark 4.8. 1) A straightforward analysis of the theorem's proof shows that it goes without changes if $a^\varepsilon(\tau)$ solves eq. (2.11) with an initial data $v_{\varepsilon 0}$ which converges to v_0 when $\varepsilon \to 0$. So

$$\mathcal{D}(a^\varepsilon(\cdot; v_{\varepsilon 0})) \rightharpoonup Q_0 \text{ in } \mathcal{P}(C([0,T],\mathbb{C}^n)) \quad \text{as } \varepsilon \to 0, \text{ if } v_{\varepsilon 0} \to v_0 \text{ when } \varepsilon \to 0. \tag{4.20}$$

2) Denoting as before $\mathcal{D}(a^\varepsilon(\cdot; v_0)) = Q_\varepsilon \in \mathcal{P}(C([0,T];\mathbb{C}^n))$ we use the Skorokhod representation theorem (see [2, Section 6]) to find a sequence $\varepsilon_j \to 0$ and processes $\xi_j(\tau)$, $0 \le \tau \le T$, $j = 0, 1, \dots$, such that $\mathcal{D}(\xi_0) = Q_0$, $\mathcal{D}(\xi_j) = Q_{\varepsilon_j}$, and $\xi_j \to \xi_0$ in $C([0,T];\mathbb{C}^n)$, a.s. Then (2.7) and Fatou's lemma imply that

$$\mathbf{E}\|a^0\|^{2m_0'}_{C([0,T];\mathbb{C}^n)} = \mathbf{E}^{Q_0}\|a\|^{2m_0'}_{C([0,T];\mathbb{C}^n)} = \mathbf{E}\|\xi_0\|^{2m_0'}_{C([0,T];\mathbb{C}^n)} \le C_{m_0'}(|v_0|, T). \tag{4.21}$$

The result of Theorem 4.7 admits an immediate generalisation to the case when the initial data v_0 in (2.12) is a random variable:

Amplification 4.9. *Let v_0 be a r.v., independent from the Wiener process $\beta^c(\tau)$. Then still the convergence (4.19) holds.*

Proof. It suffices to establish (4.19) when a^ε is a weak solution of the problem. Now, let $(\Omega', \mathcal{F}', \mathbf{P}')$ be another probability space and $\xi_0^{\omega'}$ be a r.v. on Ω', distributed as v_0. Then $a^{\varepsilon\omega}(\tau; \xi_0^{\omega'})$ is a weak solution of (2.11), (2.12), defined on the probability space $\Omega' \times \Omega$. Take f to be a bounded continuous function on $C([0,T],\mathbb{C}^n)$. Then by the theorem above, for each $\omega' \in \Omega'$

$$\lim_{\varepsilon\to 0} \mathbf{E}^\Omega f(a^{\varepsilon\omega}(\cdot; \xi_0^{\omega'})) = \mathbf{E}^\Omega f(a^{0\omega}(\cdot; \xi_0^{\omega'})).$$

Since f is bounded, then by the Lebesgue dominated convergence theorem we have

$$\lim_{\varepsilon\to 0} \mathbf{E} f(a^\varepsilon(\cdot; v_0)) = \lim_{\varepsilon\to 0} \mathbf{E}^{\Omega'}\mathbf{E}^\Omega f(a^{\varepsilon\omega}(\cdot; \xi_0^{\omega'})) = \mathbf{E}^{\Omega'}\mathbf{E}^\Omega f(a^{0\omega}(\cdot; \xi_0^{\omega'})) = \mathbf{E} f(a^0(\cdot; v_0)).$$

This implies the required convergence (4.19). □

The convergence, stated in the last amplification, holds uniformly in the class of random initial data v_0, a.s. bounded by a fixed constant. To state the result we have to introduce a distance in the space of measures.

Definition 4.10. Let M be a Polish (i.e. 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

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

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

In the definition and below we denote

$$\langle f, \mu \rangle := \int_M f(m)\mu(dm). \tag{4.22}$$

Example 4.11. Consider the Polish spaces $C([0,T],\mathbb{C}^n)$ and $\mathbb{C}^n$ and the mappings

$$\Pi_t : C([0,T],\mathbb{C}^n) \to \mathbb{C}^n, \quad a(\cdot) \mapsto a(t), \qquad 0 \le t \le T.$$

Noting that $|f \circ \Pi_t|_{L,\mathbb{C}^n} \le |f|_{L,C([0,T],\mathbb{C}^n)}$ for each t we get that

$$\|\Pi_t \circ \mu_1 - \Pi_t \circ \mu_2\|^*_{L,\mathbb{C}^n} \le \|\mu_1 - \mu_2\|^*_{L,C([0,T],\mathbb{C}^n)} \tag{4.23}$$

for all $\mu_1, \mu_2 \in \mathcal{P}(C([0,T],\mathbb{C}^n))$ and all $0 \le t \le T$ (where $\Pi_t \circ \mu_j \in \mathcal{P}(\mathbb{C}^n)$ stands for the image of μ_j under the mapping Π_t).

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 [8, Section 11.3] and [5, Section 1.7]).

Proposition 4.12. *Under the assumption of Amplification 4.9 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 (4.19) with respect to the dual-Lipschitz distance depends only on R.*

Proof. The proof of Amplification 4.9 shows that it suffices to verify that for a non-random initial data $v_0 \in \bar{B}_R(\mathbb{C}^n)$ the rate of convergence in (4.19) depends only on R. Assume the opposite. Then there exist a $\delta > 0$, a sequence $\varepsilon_j \to 0$ and vectors $v_j \in \bar{B}_R(\mathbb{C}^n)$ such that

$$\|\mathcal{D}(a^{\varepsilon_j}(\cdot; v_j)) - \mathcal{D}(a^0(\cdot; v_j))\|^*_{L,C([0,T],\mathbb{C}^n)} \geqslant \delta. \tag{4.24}$$

By the same argument as in the proof of Lemma 2.2, we know that the two sets of probability measures $\{\mathcal{D}(a^{\varepsilon_j}(\cdot; v_j))\}$ and $\{\mathcal{D}(a^0(\cdot; v_j))\}$ are pre-compact in $C([0,T],\mathbb{C}^n)$. Therefore, there exists a sequence $k_j \to \infty$ such that $\varepsilon_{k_j} \to 0$, $v_{k_j} \to v_0$ and

$$\mathcal{D}(a^{\varepsilon_{k_j}}(\cdot; v_{k_j})) \rightharpoonup \tilde{Q}_0, \quad \mathcal{D}(a^0(\cdot; v_{k_j})) \rightharpoonup Q_0 \text{ in } \mathcal{P}(C([0,T],\mathbb{C}^n)).$$

Then

$$\|\tilde{Q}_0 - Q_0\|^*_{L,C([0,T],\mathbb{C}^n)} \geqslant \delta. \tag{4.25}$$

Since in the well-posed eq. (4.5) the drift and dispersion are locally Lipschitz, then the law $\mathcal{D}(a^0(\cdot; v'))$ is continuous with respect to the initial condition v'.[7] Therefore Q_0 is the unique weak solution of the effective equation (4.5) with initial condition $a^0(0) = v_0$. By (4.20) the measure $\tilde{Q}_0$ also is a weak solution of the problem (4.5), (4.6). Hence, $Q_0 = \tilde{Q}_0$. This contradicts (4.25) and proves the assertion. □

We continue with an obvious application of Theorem 4.7 to solutions $v^\varepsilon(\tau; v_0)$ of the original eq. (2.6). Consider the action-mapping $(z_1, \dots, z_n) \mapsto (I_1, \dots, I_n) =: I$ (see (1.2)). Since the interaction representation (2.8) does not change the actions, then we have from the theorem that

Corollary 4.13. *For any v_0, as $\varepsilon \to 0$*

$$\mathcal{D}(I(v^\varepsilon(\cdot; v_0))) \rightharpoonup I \circ \mathcal{D}(a(\cdot; v_0)) \quad \text{in} \quad \mathcal{P}(C([0,T],\mathbb{R}^n_+)), \tag{4.26}$$

where $a(\cdot; v_0)$ is a unique weak solution of effective equation (4.5), (4.6).

[7] This is well known, and can be easily proved using the estimate in Remark 4.8.2).

Example 4.14. If the drift P in (2.6) is globally Lipschitz, that is, $\mathrm{Lip}(P) \leqslant M$ for some $M > 0$, then it is not hard to see that Assumption 2.1 holds, so Theorem 4.7 and Corollary 4.13 apply. A more interesting example is discussed below in Section 9.

4.1. **Proof of Lemma 4.3.** In this subsection we denote by $\mathcal{H}_k(r; c_1, \dots)$, $k = 1, 2, \dots$, non-negative functions of $r > 0$ which go to zero with r and depend on parameters $c_1, \dots$ (dependence of $\mathcal{H}_k$'s on T and P is not indicated). Also, for an event Q we denote $\mathbf{E}_Q f(\xi) = \mathbf{E}\mathbf{1}_Q f(\xi)$.

For any $M_1 \geqslant 1$ we set

$$\mathcal{E}^1 = \mathcal{E}^{1\varepsilon}_{M_1} = \{\omega \in \Omega : \sup_{0 \leqslant \tau \leqslant T} |a^\varepsilon(\tau)| \leqslant M_1\}.$$

By Assumption 2.1 and Chebyshev's inequality,

$$\mathbf{P}(\Omega \smallsetminus \mathcal{E}^1) \leqslant \mathcal{H}_1({M_1}^{-1}; |v_0|).$$

Recalling that $\tilde{y}$ was defined in (4.9), by Lemma 3.2 we have that

$$|\tilde{y}(a^\varepsilon(s), s\varepsilon^{-1})| \leqslant |Y(a^\varepsilon(s), s\varepsilon^{-1})| + |\langle\!\langle P \rangle\!\rangle(a^\varepsilon(s))| \leqslant 2\mathcal{C}^{m_0}(P)|a^\varepsilon(s)|^{m_0}.$$

So, abbreviating $\tilde{y}(a^\varepsilon(s), s\varepsilon^{-1})$ to $\tilde{y}(s)$, in view of (2.7) we have:

$$\begin{aligned}
\mathbf{E}_{\Omega\smallsetminus\mathcal{E}^1} \max_{\tau\in[0,T]} |\int_0^\tau \tilde{y}(s)ds| &\leqslant \int_0^T \mathbf{E}\big(\mathbf{1}_{\Omega\smallsetminus\mathcal{E}^1}|\tilde{y}(s)|\big)ds \\
&\leqslant 2\mathcal{C}^{m_0}(P)\big(P(\Omega\smallsetminus\mathcal{E}^1)\big)^{1/2}(\int_0^T \mathbf{E}|a^\varepsilon(s)|^{2m_0}ds)^{1/2} \\
&\leqslant 2\mathcal{C}^{m_0}(P)(\mathcal{H}_1({M_1}^{-1}; |v_0|))^{1/2} =: \mathcal{H}_2({M_1}^{-1}; |v_0|).
\end{aligned}$$

Now we should estimate $\mathbf{E}_{\mathcal{E}^1} \max_{\tau\in[0,T]} |\int_0^\tau \tilde{y}(s)ds|$. For an $M_2 \geqslant 1$ consider the event

$$\mathcal{E}^2 = \mathcal{E}^{2\varepsilon}_{M_2} = \{\omega \in \Omega : \|a^\varepsilon\|_{1/3} \leqslant M_2\}$$

(see (1.4)). Then by (2.13)

$$\mathbf{P}(\Omega \smallsetminus \mathcal{E}^2) \leqslant \mathcal{H}_3(M_2^{-1}; |v_0|).$$

Therefore,

$$\mathbf{E}_{\Omega\smallsetminus\mathcal{E}^2} \max_{\tau\in[0,T]} |\int_0^\tau \tilde{y}(s)ds| \leqslant \big(\mathbf{P}(\Omega\smallsetminus\mathcal{E}^2)\big)^{1/2}(C_P \int_0^T \mathbf{E}|a^\varepsilon(s)|^{2m_0}ds)^{1/2} \leqslant \mathcal{H}_4(M_2^{-1}; |v_0|)\,.$$

It remains to bound $\mathbf{E}_{\mathcal{E}^1\cap\mathcal{E}^2} \max_{\tau\in[0,T]} \big|\int_0^\tau \tilde{y}(s)ds\big|$.

We set

$$N = [T/\sqrt{\varepsilon}] + 1, \qquad L = T/N.$$

Then $C^{-1}\sqrt{\varepsilon} \leqslant L \leqslant C\sqrt{\varepsilon}$ and $c^{-1}/\sqrt{\varepsilon} \leqslant N \leqslant c/\sqrt{\varepsilon}$ for some constants C and c. We consider a partition of interval $[0, T]$ by points $\tau_l = lL$, $l = 0, \dots, N$, and denote

$$\eta_l = \int_{\tau_l}^{\tau_{l+1}} \tilde{y}(s)ds, \quad l = 0, \dots, N-1.$$

For any $\tau \in [0, T]$ let us find $l = l(\tau)$ such that $\tau \in [\tau_l, \tau_{l+1}]$. Then

$$|\int_0^\tau \tilde{y}(s)ds| \leqslant |\eta_1| + \dots + |\eta_l| + |\int_{\tau_l}^\tau \tilde{y}(s)ds|.$$

If $\omega \in \mathcal{E}^1$, then $|\int_{\tau_l}^\tau \tilde{y}(s)ds| \leqslant 2\mathcal{C}^{m_0}(P){M_1}^{m_0}L$. Therefore

$$\mathbf{E}_{\mathcal{E}^1\cap\mathcal{E}^2} \max_{0\leqslant\tau\leqslant T} \Big|\int_0^\tau \tilde{y}(s)ds\Big| \leqslant 2\mathcal{C}^{m_0}(P){M_1}^{m_0}L + \sum_{l=0}^{N-1} \mathbf{E}_{\mathcal{E}^1\cap\mathcal{E}^2}|\eta_l|, \tag{4.27}$$

and it remains to estimate the integrals $\mathbf{E}_{\mathcal{E}^1\cap\mathcal{E}^2}|\eta_l|$ for $l=0,\dots,N-1$. Observe that

$$|\eta_l| \leqslant \Big|\int_{\tau_l}^{\tau_{l+1}} [\tilde{y}(a^\varepsilon(s), s\varepsilon^{-1}) - \tilde{y}(a^\varepsilon(\tau_l), s\varepsilon^{-1})]ds\Big| + \Big|\int_{\tau_l}^{\tau_{l+1}} \tilde{y}(a^\varepsilon(\tau_l), s\varepsilon^{-1})ds\Big| =: |U_l^1| + |U_l^2|.$$

Since $\tilde{y}(a^\varepsilon, \tau\varepsilon^{-1}) = (\Phi_{\tau\varepsilon^{-1}\Lambda})_* P(a^\varepsilon) - \langle\!\langle P\rangle\!\rangle(a^\varepsilon)$ and $P, \langle\!\langle P\rangle\!\rangle \in \mathrm{Lip}_{m_0}(\mathbb{C}^n, \mathbb{C}^n)$, then for $\omega \in \mathcal{E}^1\cap\mathcal{E}^2$ the integrand in U_l^1 is bounded by

$$2\mathcal{C}^{m_0}(P){M_1}^{m_0} \sup_{\tau_l\leqslant s\leqslant\tau_{l+1}} |a^\varepsilon(s) - a^\varepsilon(\tau_l)| \leqslant 2\mathcal{C}^{m_0}(P){M_1}^{m_0}M_2L^{1/3}.$$

So

$$|U_l^1| \leqslant 2\mathcal{C}^{m_0}(P){M_1}^{m_0}M_2L^{4/3}.$$

Now consider integral U_l^2. By the definition of $\tilde{y}(a^\varepsilon, \tau\varepsilon^{-1})$, we have

$$U_l^2 = \int_{\tau_l}^{\tau_{l+1}} Y(a^\varepsilon(\tau_l), s\varepsilon^{-1})ds - L\langle\!\langle P\rangle\!\rangle(a^\varepsilon(\tau_l)\big) =: Z^1 + Z^2.$$

For integral Z^1, making the change of variable $s = \tau_l + \varepsilon x$ with $x \in [0, \frac{L}{\varepsilon}]$, we have

$$Z^1 = \varepsilon\int_0^{L/\varepsilon} Y(a^\varepsilon(\tau_l), \tau_l\varepsilon^{-1} + x)dx.$$

Since

$$Y(a^\varepsilon(\tau_l), \tau_l\varepsilon^{-1} + x) = \Phi_{\tau_l\varepsilon^{-1}\Lambda}\circ\Phi_{x\Lambda}P\big(\Phi_{-x\Lambda}(\Phi_{-\tau_l\varepsilon^{-1}\Lambda}a^\varepsilon(\tau_l))\big),$$

then

$$Z^1 = L\Phi_{\tau_l\varepsilon^{-1}\Lambda}\Big(\frac{\varepsilon}{L}\int_0^{L/\varepsilon}\Phi_{x\Lambda}P\big(\Phi_{-x\Lambda}\big(\Phi_{-\tau_l\varepsilon^{-1}\Lambda}a^\varepsilon(\tau_l)\big)\big)dx\Big) = L\Phi_{\tau_l\varepsilon^{-1}\Lambda}\langle\!\langle P\rangle\!\rangle^{L/\varepsilon}\big(\Phi_{-\tau_l\varepsilon^{-1}\Lambda}a^\varepsilon(\tau_l)\big)$$

(see (3.1)). As $L/\varepsilon \sim \varepsilon^{-1/2} \gg 1$ and $\big|\Phi_{-\tau_l\varepsilon^{-1}\Lambda}a^\varepsilon(\tau_l)\big| = |a^\varepsilon(\tau_l)| \leqslant M_1$ for $\omega \in \mathcal{E}^1\cap\mathcal{E}^2$, then by Lemma 3.2 the partial averaging $\langle\!\langle P\rangle\!\rangle^{L/\varepsilon}$ is close to the complete averaging $\langle\!\langle P\rangle\!\rangle$. Thus

$$\begin{aligned}\big|Z^1 - L\Phi_{\tau_l\varepsilon^{-1}\Lambda}\langle\!\langle P\rangle\!\rangle\big(\Phi_{-\tau_l\varepsilon^{-1}\Lambda}(a^\varepsilon(\tau_l))\big)\big| &= \big|Z^1 - L(\Phi_{\tau_l\varepsilon^{-1}\Lambda})_*\langle\!\langle P\rangle\!\rangle(a^\varepsilon(\tau_l))\big| \\ &= \big|Z^1 - L\langle\!\langle P\rangle\!\rangle(a^\varepsilon(\tau_l))\big| \leq L\mathcal{H}_5(\sqrt{\varepsilon}; M_1, |v_0|),\end{aligned}$$

where we used Lemma 3.3 to get the second equality. Since $Z^2 = -L\langle\!\langle P\rangle\!\rangle(a^\varepsilon(\tau_l))$, then

$$|U_l^2| = |Z^1 + Z^2| \leq L\mathcal{H}_5(\sqrt{\varepsilon}; M_1, |v_0|).$$

We then have obtained that

$$\mathbf{E}_{\mathcal{E}^1\cap\mathcal{E}^2}|\eta_l| \leqslant 2\mathcal{C}^{m_0}(P){M_1}^{m_0}M_2L^{4/3} + L\mathcal{H}_5(\sqrt{\varepsilon}; M_1, |v_0|).$$

Together with (4.27) it gives us that

$$\mathbf{E}_{\mathcal{E}^1\cap\mathcal{E}^2}\Big(\max_{0\leqslant\tau\leqslant T}\Big|\int_0^\tau \tilde{y}(s)ds\Big|\Big) \leqslant 2\mathcal{C}^{m_0}(P){M_1}^{m_0}L + 2\mathcal{C}^{m_0}(P){M_1}^{m_0}M_2L^{1/3} + \mathcal{H}_5(\sqrt{\varepsilon}; M_1, |v_0|).$$

Therefore

$$\begin{aligned}\mathbf{E}\max_{0\leqslant\tau\leqslant T}\Big|\int_0^\tau \tilde{y}(s)ds\Big| &\leqslant \mathcal{H}_2({M_1}^{-1}; |v_0|) + \mathcal{H}_4(M_2^{-1}; |v_0|) \\ &\quad + 2\mathcal{C}^{m_0}(P){M_1}^{m_0}(M_2+1)\varepsilon^{1/6} + \mathcal{H}_5(\sqrt{\varepsilon}; M_1, |v_0|).\end{aligned}$$

Now for any $\delta > 0$, we perform the following procedure:

(1) choose M_1 so large that $\mathcal{H}_2({M_1}^{-1}; |v_0|) \leqslant \delta$;
(2) choose M_2 so large that $\mathcal{H}_4(M_2^{-1}; |v_0|) \leqslant \delta$;
(3) finally, choose $\varepsilon_\delta > 0$ so small that

$$2\mathcal{C}^{m_0}(P){M_1}^{m_0}(M_2+1)\varepsilon^{1/3} + \mathcal{H}_5(\sqrt{\varepsilon}; M_1, |v_0|) \leqslant \delta \quad \text{if} \quad 0 < \varepsilon \leqslant \varepsilon_\delta.$$

We have seen that for any $\delta > 0$, $\mathbf{E}\max_{0\leqslant\tau\leqslant T}\left|\int_0^\tau \tilde{y}(a^\varepsilon(s), s\varepsilon^{-1})ds\right| \le 3\delta$ if $\varepsilon \le \varepsilon_\delta$. So

$$\mathbf{E}\max_{0\leqslant\tau\leqslant T}\Big|\int_0^\tau [Y(a^\varepsilon(s), s\varepsilon^{-1}) - \langle\!\langle P\rangle\!\rangle(a^\varepsilon(s))]ds\Big| \to 0 \quad \text{as} \quad \varepsilon \to 0,$$

which completes the proof of Lemma 4.3.

5. Stationary solutions and mixing

In this section we study relation between stationary solutions of equation (2.6) and of the effective equation. We recall that a solution $a(\tau)$, $\tau \geqslant 0$, of equation (2.6) (or of effective equation (4.5)) is stationary if $\mathcal{D}(a(\tau)) \equiv \mu$ for all $\tau \geqslant 0$ and some measure $\mu \in \mathcal{P}(\mathbb{C}^n)$, called a *stationary measure* for the corresponding equation.

Through this section we assume that equation (2.6) satisfies the following strengthening of Assumption 2.1:

Assumption 5.1. (1) The drift $P(v)$ is a locally Lipschitz vector filed, belonging to $\mathrm{Lip}_{m_0}(\mathbb{C}^n, \mathbb{C}^n)$ for some $m_0 \in \mathbb{N}$.

(2) For any $v_0 \in \mathbb{C}^n$ equation (2.6) has a unique strong solution $v^\varepsilon(\tau; v_0)$, $\tau \ge 0$, equal v_0 at $\tau = 0$. There exists $m_0' > (m_0 \vee 1)$ such that

$$\mathbf{E}\sup_{T'\leqslant\tau\leqslant T'+1}|v^\varepsilon(\tau; v_0)|^{2m_0'} \leqslant C_{m_0'}(|v_0|) \tag{5.1}$$

for every $T' \geqslant 0$ and $\varepsilon \in (0, 1]$, where $C_{m_0'}(\cdot)$ is a continuous non-decreasing function.

(3) Equation (2.6) is *mixing*. That is, it has a stationary solution $v^\varepsilon_{st}(\tau)$, $\mathcal{D}(v^\varepsilon_{st}(\tau)) \equiv \mu^\varepsilon \in \mathcal{P}(\mathbb{C}^n)$, and

$$\mathcal{D}(v^\varepsilon(\tau; v_0)) \rightharpoonup \mu^\varepsilon \text{ in } \mathcal{P}(\mathbb{C}^n) \quad \text{as } \tau \to +\infty, \tag{5.2}$$

for every v_0.

Under Assumption 5.1 equation (2.6) defines in $\mathbb{C}^n$ a mixing Markov process with the transition probability $\Sigma_\tau(v) \in \mathcal{P}(\mathbb{C}^n)$, $\tau \ge 0$, $v \in \mathbb{C}^n$, where $\Sigma_\tau(v) = \mathcal{D}v^\varepsilon(\tau; v)$; e.g. see [14, Section 5.4.C]. Let us denote by X the complete separable metric space $X = C([0, \infty), \mathbb{C}^n)$ with the distance

$$\mathrm{dist}(a_1, a_2) = \sum_{N=1}^{\infty} 2^{-N}\frac{\|a_1 - a_2\|_{C([0,N],\mathbb{C}^n)}}{1 + \|a_1 - a_2\|_{C([0,N],\mathbb{C}^n)}}, \qquad a_1, a_2 \in X\,, \tag{5.3}$$

and consider on X continuous function $g(a) = \sup_{0\le t\le 1}|a(t)|^{2m_0'}$. Denoting $\mu^\tau_\varepsilon = \mathcal{D}(v^\varepsilon(\tau; 0))$ we have by the Markov property that

$$\mathbf{E}\sup_{T'\leqslant\tau\leqslant T'+1}|v^\varepsilon(\tau; 0)|^m = \int_{\mathbb{C}^n}\mathbf{E}g(v^\varepsilon(\cdot; v_0))\mu^{T'}_\varepsilon(dv_0), \tag{5.4}$$

and

$$\mathbf{E}\sup_{0\leqslant\tau\leqslant 1}|v^\varepsilon_{st}(\tau; 0)|^m = \int_{\mathbb{C}^n}\mathbf{E}g(v^\varepsilon(\cdot; v_0))\mu^\varepsilon(dv_0). \tag{5.5}$$

The l.h.s. of (5.4) is estimated in (5.1). To estimate (5.5) we will pass in the r.h.s. of (5.4) to the limit as $T' \to \infty$, using (5.2). To do that we start with a lemma

Lemma 5.2. *Let $n_1, n_2 \in \mathbb{N}$, $\mathcal{B} \subset \mathbb{R}^{n_1}$ is a closed convex set which contains more than one point, and $F : \mathcal{B} \to \mathbb{R}^{n_2}$ is a Lipschitz mapping. Then F may be extended to a map $\tilde{F} : \mathbb{R}^{n_1} \to \mathbb{R}^{n_2}$ in such a way that*

a) $\mathrm{Lip}\,\tilde{F} = \mathrm{Lip}\,F$,

b) $\tilde{F}(\mathbb{R}^{n_1}) = F(\mathcal{B})$.

Proof. Let $\Pi:\mathbb{R}^{n_1}\to\mathcal{B}$ be the projection, sending any point of $\mathbb{R}^{n_1}$ to a nearest point in $\mathcal{B}$. Then $\operatorname{Lip}\Pi=1$, see Appendix D, and obviously $\tilde F=F\circ\Pi$ is a required extension of F. $\square$

Since $\mathcal{C}^{m_0}(P)=:C_*<\infty$, then for any $M\in\mathbb{N}$ the restriction of P to $\bar B_M(\mathbb{C}^n)$ and its Lipschitz constant both are bounded by $(1+M)^{m_0}C_*$. In view of the lemma above we may extend $P|_{\bar B_M(\mathbb{C}^n)}$ to a Lipschitz mapping $P^M:\mathbb{C}^n\to\mathbb{C}^n$ such that

$$\operatorname{Lip}(P^M)\le(1+M)^{m_0}C_*,\quad \sup|P^M(v)|\le(1+M)^{m_0}C_*.$$

Considering for a solution $v(\tau)$ of equation (2.6) the stopping time $\tau_M=\inf\{t\ge0:|v(t)|\ge M\}$ and denoting by $v^{\varepsilon M}$ the stopped solution $v^{\varepsilon M}(\tau;v_0)=v^\varepsilon(\tau\wedge\tau_M;v_0)$, we note that the process $v^{\varepsilon M}$ will not change if in (2.6) we replace $P(v)$ by $P^M(v)$. So $v^{\varepsilon M}(\tau;v_0)$ is a stopped solution of a stochastic equation with Lipschitz coefficients, and thus a.s. the curve $v^{\varepsilon M\omega}(\cdot;v_0)\in X$ continuously depends on v_0, for each $M\in\mathbb{N}$. As

$$g(v^{\varepsilon M})\le g(v^\varepsilon)\qquad\text{a.s.},$$

then in view of (5.2) and (5.1), for every M and any $N>0$,

$$\int\mathbf{E}(N\wedge g)(v^{\varepsilon M}(\cdot;v))\,\mu^\varepsilon(dv)=\lim_{T'\to\infty}\int\mathbf{E}(N\wedge g)(v^{\varepsilon M}(\cdot;v))\,\mu_\varepsilon^{T'}(dv)\le C_m(0)$$

(to get the last inequality from (5.1) we used the Markov property). Passing in the l.h.s. to the limit as $N\to\infty$ using the monotone convergence theorem we see that

$$\int\mathbf{E}g(v^{\varepsilon M}(\cdot;v))\,\mu^\varepsilon(dv)\le C_m(0). \tag{5.6}$$

Since for every v a.s. $g(v^{\varepsilon M}(\cdot;v))\nearrow g(v^\varepsilon(\cdot;v))\le\infty$ as $M\to\infty$, then using again the latter theorem we derive from (5.6) that

$$\int\mathbf{E}g(v^\varepsilon(\cdot;v))\,\mu^\varepsilon(dv)\le C_m(0).$$

Evoking (5.5) we get

Lemma 5.3. *The stationary solution $v^\varepsilon_{st}(\tau)$ satisfies estimate (5.1) with $C_{m_0'}(|v_0|)$ replaced by $C_{m_0'}(0)$.*

Let us consider the interaction representation for v^ε_{st}, $v^\varepsilon_{st}(\tau)=\Phi_{-\tau\varepsilon^{-1}\Lambda}a^\varepsilon(\tau)$ (note that a^ε is not a stationary process!). Then $a^\varepsilon(\tau)$ satisfies equation (2.11), so for any $N\in\mathbb{N}$ the collection of measures $\{\mathcal{D}(a^\varepsilon|_{[0,N]}),0<\varepsilon\leqslant1\}$ is tight in view of (5.1) (for the same reason as in Section 2.4). We choose a sequence $\varepsilon_l\to0$ (depending on N) such that

$$\mathcal{D}(a^{\varepsilon_l}|_{[0,N]})\rightharpoonup Q_0\text{ in }\mathcal{P}\big(C([0,N],\mathbb{C}^n)\big).$$

Applying the diagonal process and replacing $\{\varepsilon_l\}$ by a subsequence, which we still denote $\{\varepsilon_l\}$, we achieve that $\mathcal{D}a^{\varepsilon_l}\rightharpoonup Q_0$ in $\mathcal{P}(X)$ (see (5.3)).

Since $a^\varepsilon(0)=v^\varepsilon_{st}(0)$, then

$$\mu^{\varepsilon_l}\rightharpoonup\mu^0:=Q_0|_{\tau=0}.$$

Let $a^0(\tau)$ be a process in $\mathbb{C}^n$ such that $\mathcal{D}(a^0)=Q_0$. Then

$$\mathcal{D}(a^{\varepsilon_l}(\tau))\rightharpoonup\mathcal{D}(a^0(\tau))\quad\forall\,0\le\tau<\infty. \tag{5.7}$$

In particular, $\mathcal{D}(a^0(0))=\mu^0$.

Proposition 5.4. *(1) The limiting process a^0 is a stationary weak solution of effective equation (4.5) and $\mathcal{D}(a^0(\tau)) \equiv \mu^0$, $\tau \in [0,\infty)$. In particular, the limiting points of the collection of stationary measures $\{\mu^\varepsilon, \varepsilon \in (0,1]\}$ as $\varepsilon \to 0$ are stationary measures of the effective equation.*

(2) Any limiting measure μ^0 is invariant under operators $\Phi_{\theta\Lambda}$, $\theta \in \mathbb{R}$. So $\mathcal{D}(\Phi_{\theta\Lambda} a^0(\tau)) = \mu^0$ for all $\theta \in \mathbb{R}$ and $\tau \in [0,\infty)$.

Proof. (1) Using Lemma 5.3 and repeating the argument in the proof of Proposition 4.2 we obtain that a^0 is a weak solution of the effective equation. It remains to prove its stationarity.

Take any bounded Lipschitz function f on $\mathbb{C}^n$ and consider

$$\mathbf{E}\int_0^1 f(v_{st}^{\varepsilon_l}(\tau))d\tau = \mathbf{E}\int_0^1 f(\Phi_{-\tau\varepsilon_l^{-1}\Lambda} a^{\varepsilon_l}(\tau))d\tau.$$

Using the same argument as in the proof of Lemma 4.3 (but applied to averaging of functions rather than that of vector-fields), we obtain that

$$\begin{aligned}\mathbf{E}\int_0^1 f\big(v_{st}^{\varepsilon_l}(\tau)\big)d\tau - \mathbf{E}\int_0^1 \langle f\rangle\big(a^{\varepsilon_l}(\tau)\big)d\tau \\ = \mathbf{E}\int_0^1 \big(f\big(\Phi_{-\tau\varepsilon_l^{-1}\Lambda} a^{\varepsilon_l}(\tau)\big) - \langle f\rangle(a^{\varepsilon_l}(\tau))\big)d\tau \to 0 \quad \text{as } \varepsilon_l \to 0.\end{aligned} \tag{5.8}$$

By Lemma 3.6,

$$\langle f\rangle\big(a^{\varepsilon_l}(\tau)\big) = \langle f\rangle\big(\Phi_{\tau\varepsilon_l^{-1}\Lambda} v_{st}^{\varepsilon_l}(\tau)\big) = \langle f\rangle\big(v_{st}^{\varepsilon_l}(\tau)\big)$$

for every τ. Since the process $v_{st}^{\varepsilon_l}(\tau)$ is stationary, then

$$\mathbf{E} f(v_{st}^{\varepsilon_l}(\tau)) = \text{Const} \quad \text{and} \quad \mathbf{E}\langle f\rangle\big(a^{\varepsilon_l}(\tau)\big) = \mathbf{E}\langle f\rangle(v_{st}^{\varepsilon_l}(\tau)) = \text{Const}'.$$

So we get from (5.8) that

$$\mathbf{E} f\big(v_{st}^{\varepsilon_l}(\tau)\big) - \mathbf{E}\langle f\rangle\big(a^{\varepsilon_l}(\tau)\big) \to 0 \ \text{ as } \varepsilon_l \to 0 \quad \forall\, \tau. \tag{5.9}$$

For any τ let us consider $\tilde{f}_\tau = f \circ \Phi_{\tau\varepsilon_l^{-1}\Lambda}$. Then $f(a^{\varepsilon_l}(\tau)) = \tilde{f}_\tau(v_{st}^{\varepsilon_l}(\tau))$. Since $\langle f\rangle = \langle \tilde{f}_\tau\rangle$ by Lemma 3.6.(4), then applying (5.9) to $\tilde{f}_\tau$ we get:

$$\lim_{\varepsilon_l\to 0} \mathbf{E} f(a^{\varepsilon_l}(\tau)) = \lim_{\varepsilon_l\to 0} \mathbf{E}\tilde{f}_\tau(v_{st}^{\varepsilon_l}(\tau)) = \lim_{\varepsilon_l\to 0} \mathbf{E}\langle \tilde{f}_\tau\rangle(a^{\varepsilon_l}(\tau)) = \lim_{\varepsilon_l\to 0} \mathbf{E}\langle f\rangle(a^{\varepsilon_l}(\tau)).$$

From this relation, (5.9) and (5.7) we find that $\mathbf{E} f(a^0(\tau)) = \int f(v)\,\mu^0(dv)$ for each τ and every f as above. This implies the first assertion of the lemma.

(2) Passing to the limit in (5.9) using (5.7) we have

$$\int f(v)\mu^0(dv) = \mathbf{E}\langle f\rangle(a^0(\tau)) \quad \forall\, \tau.$$

Using this relation with $f := f \circ \Phi_{\theta\Lambda}$ and next with $f := f$ we get that

$$\int f \circ \Phi_{\theta\Lambda}(v)\mu^0(dv) = \mathbf{E}\langle f \circ \Phi_{\theta\Lambda}\rangle(a^0(\tau)) = \mathbf{E}\langle f\rangle(a^0(\tau)) = \int f(v)\mu^0(dv),$$

for any $\theta \in \mathbb{R}$ and any τ, for every bounded Lipschitz function f. This implies the second assertion. □

If effective equation (4.5) is mixing, then it has a unique stationary measure. In this case the measure μ^0 in Proposition 5.4 does not depend on a choice of the sequence $\varepsilon_l \to 0$, and so $\mu^\varepsilon \rightharpoonup \mu^0$ as $\varepsilon \to 0$. Therefore, we have

Theorem 5.5. *If in addition to Assumption 5.1 we assume that the effective equation is mixing and μ^0 is its unique stationary measure, then*

$$\mu^\varepsilon \rightharpoonup \mu^0 \text{ in } \mathcal{P}(\mathbb{C}^n) \quad \text{as} \quad \varepsilon \to 0.$$

Moreover, the measure μ^0 is invariant for all operators $\Phi_{\theta\Lambda}$, and the law of the stationary solution of equation (2.6), written in the interaction presentation, converges to the law of the stationary solution of effective equation (4.5).

We recall that Theorem 4.7 and Corollary 4.13 only assure that on finite time-intervals $\tau \in [0,T]$ the actions of solutions for eq. (2.6), as $\varepsilon \to 0$, converge in law to the actions of solutions for the effective equation with the same initial data. In difference, when $\varepsilon \to 0$ *all* of the stationary measure for eq. (2.6) converges to that for the effective equation. This important fact was first observed in [9] for a special class of equations (2.6).

Corollary 5.6. *Under the assumption of Theorem 5.5 , for any $v_0 \in \mathbb{C}^n$ we have*

$$\lim_{\varepsilon\to 0}\lim_{\tau\to\infty} \mathcal{D}(v^\varepsilon(\tau;v_0)) = \mu^0.$$

Proof. Since $\lim_{\tau\to\infty}\mathcal{D}(v^\varepsilon(\tau)) = \mu^\varepsilon$, then the result follows from Theorem 5.5. □

Remark 5.7. Let us decomplexify $\mathbb{C}^n$ to $\mathbb{R}^{2n}$ and write the effective equation in the real coordinates $\{x = (x_1,\dots,x_{2n})\}$:

$$dx_j(\tau) - \langle\!\langle P\rangle\!\rangle_j(x)d\tau = \sum_{l=1}^{2n}\mathcal{B}_{jl}dW_l(\tau),\ j = 1,\dots,2n,$$

where W_l are independent standard real Wiener processes. Then the stationary measure $\mu^0 \in \mathcal{P}(\mathbb{R}^{2n})$ satisfies the stationary Fokker-Plank equation

$$\frac{1}{2}\sum_{l=1}^{2n}\sum_{j=1}^{2n}\frac{\partial^2}{\partial x_l\partial x_j}(\mathcal{B}_{lj}\mu^0) = \sum_{l=1}^{2n}\frac{\partial}{\partial x_l}(\langle\!\langle P\rangle\!\rangle_l(x)\mu^0)$$

in the sense of distributions. If the dispersion matrix Ψ is non-degenerate, then the diffusion $\mathcal{B}$ also is, and since the drift $\langle\!\langle P\rangle\!\rangle(x)$ is locally Lipschitz, then by the standard theory of the Fokker-Plank equation $\mu^0 = \varphi(x)dx$, where $\varphi \in C^1(\mathbb{R}^{2n})$.[8]

6. The non-resonant case

Assume that the frequency vector $\Lambda = (\lambda_1,\dots,\lambda_n)$ is non-resonant (see (3.7)). In Section 3.1.2) we saw that in this case vector field $\langle\!\langle P\rangle\!\rangle$ may be calculated via the averaging (3.8) and commutes with all rotations Φ_w, $w \in \mathbb{R}^n$. For any $j \in \{1,\dots,n\}$ let us denote $w^{j,t} := (0,\dots 0,t,0,\dots,0)$ (only the j-th entry is non-zero). Consider $\langle\!\langle P\rangle\!\rangle_1(z)$ and write it as $\langle\!\langle P\rangle\!\rangle_1(z) = z_1R_1(z_1,\dots,z_n)$, for some complex function R_1. Since for $w = w^{1,t}$ we have $\Phi_w(z) = (e^{it}z_1,z_2,\dots,z_n)$, then now the first component in relation (3.10) reads $e^{-it}z_1R_1(e^{-it}z,z_1,\dots,z_n) = e^{-it}z_1R_1(z_1,\dots,z_n)$, for every t. So

$$R_1(e^{-it}z_1,z_2,\dots,z_n) \equiv R_1(z_1,\dots,z_n)$$

and $R_1(z_1,\dots,z_n)$ depends not on z_1, but only on $|z_1|$. Similarly we verify that $R_1(z_1,\dots,z_n)$ depends only on $|z_2|,\dots,|z_n|$. Therefore $\langle\!\langle P\rangle\!\rangle_1(z) = z_1R_1(|z_1|,\dots,|z_n|)$. Same is true for any $\langle\!\langle P\rangle\!\rangle_j(z)$. We then obtain the following statement:

[8] E.g. firstly Theorem 1.6.8 from [3] implies that $\mu^0 = \varphi(x)dx$, where φ is a Hölder function, and then by the usual elliptic regularity $\varphi \in C^1$.

Proposition 6.1. *If Assumption 2.1 holds and the frequency vector Λ is non-resonant, then $\langle\langle P\rangle\rangle$ satisfies (3.8), and*

(1) $\langle\langle P\rangle\rangle_j(a) = a_j R_j(|a_1|,\dots,|a_n|)$, $j = 1,\dots,n$.

(2) The effective equation reads

$$da_j(\tau) - a_j R_j(|a_1|,\dots,|a_n|)d\tau = b_j d\beta_j^c(\tau), \quad j = 1,\dots,n, \tag{6.1}$$

where $b_j = (\sum_{l=1}^n |\Psi_{jl}|^2)^{1/2}$ (and $a \mapsto (a_1R_1,\dots,a_nR_n)$ is a locally Lipschitz vector-field).

(3) If $a(\tau)$ is a solution of (6.1), then the vector of its actions $I(\tau) = (I_1,\dots,I_n)(\tau) \in \mathbb{R}^n_+$ is a weak solution of equation

$$dI_j(\tau) - 2I_j(\Re R_j)\big(\sqrt{2I_1},\dots,\sqrt{2I_n}\big)d\tau - b_j^2 d\tau = b_j\sqrt{2I_j}\,dW_j(\tau), \quad I_j(0) = \tfrac{1}{2}|v_{0j}|^2, \tag{6.2}$$

$j = 1,\dots,n$, where $\{W_j\}$ are independent standard real Wiener processes.

(4) If in addition the assumptions of Theorem 5.5 are met and the matrix Ψ is non-degenerate, then the stationary measure μ^0 reads $d\mu^0 = p(I)dI\,d\varphi$, where p is a continuous function on $\mathbb{R}^n_+$, C^1-smooth outside the boundary $\partial\mathbb{R}^n_+$.

Proof. (1) is proved above, and (2) follows from it and (4.4).

(3) Writing the diffusion in effective equation (4.5) as in eq. (6.1) and applying Itô's formula (see Appendix C) to $I_j = \frac{1}{2}|a_j|^2$ we get that

$$dI_j(\tau) - \tfrac{1}{2}\big(\bar a_j\langle\langle P\rangle\rangle_j + a_j\overline{\langle\langle P\rangle\rangle_j}\big)d\tau - b_j^2 d\tau = b_j\langle a_j(\tau), d\beta_j(\tau)\rangle =: b_j|a_j|d\xi_j(\tau), \tag{6.3}$$

where $d\xi_j(\tau) = \langle a_j/|a_j|, d\beta_j(\tau)\rangle$ (see (2.2)) and for $a_j = 0$ we define $a_j/|a_j|$ as 1. Using (1) we see that the l.h.s. in (6.3) is the same as that in (6.2). Since $\big|a_j/|a_j|\big|(\tau) \equiv 1$ for each j, then by Lévy's theorem (e.g. see [14, p. 157]), $\xi(\tau) = (\xi_1,\dots,\xi_n)(\tau)$ is a standard n-dimensional Wiener process and (3) follows.

(4) By Theorem 5.5 the stationary measure μ^0 is invariant for all operators $\Phi_{\theta\Lambda}$, $\theta \in \mathbb{R}$. Since the curve $\theta \mapsto \theta\Lambda \in \mathbb{T}^n$ is dense in $\mathbb{T}^n$, then μ^0 is invariant for all operators Φ_w, $w \in \mathbb{T}^n$. As the matrix Ψ is non-degenerate, then by Remark 5.7 we have $d\mu^0 = \tilde p(z)dz$, where $\tilde p$ is a C^1-function (dz stands for the volume element in $\mathbb{C}^n \simeq \mathbb{R}^{2n}$). Let us write $z_j = \sqrt{2I_j}e^{i\varphi_j}$. Then $d\mu^0 = p(I,\varphi)dz$. In the coordinates (I,φ) the operators Φ_w reads as $(I,\varphi) \mapsto (I,\varphi+w)$. Since μ^0 is invariant for all of them, then p does not depend on φ. So $d\mu^0 = p(I)dz = p(I)\ dI\,d\varphi$ and (4) holds. □

By assertion (3) of the proposition, in the non-resonant case eq. (6.2) describes asymptotic as $\varepsilon \to 0$ behaviour of actions I_j of solutions for eq. (2.6). But how regular is that equation? Let us denote by $r_j = |a_j| = \sqrt{2I_j}$, $1 \le j \le n$, the radii of components of a vector $a \in \mathbb{C}^n$, consider the smooth polar-coordinates mapping $\mathbb{R}^n_+ \times \mathbb{T}^n \to \mathbb{C}^n$, $(r,\varphi) \mapsto (r_1e^{i\varphi_1},\dots,r_ne^{i\varphi_n})$, and extend it to a mapping

$$\Phi : \mathbb{R}^n \times \mathbb{T}^n \to \mathbb{C}^n,$$

defined by the same formula. A j-th component of the drift in eq. (6.2), written in the form (6.3) without the Itô term $b_j^2 d\tau$, is $\Re\big(a_j\overline{\langle\langle P\rangle\rangle_j(a)}\big)$. Due to (3.8), in the polar coordinates we write this as

$$\frac{1}{(2\pi)^n}\int_{\mathbb{T}^n}\Re\Big(r_je^{i\varphi_j}\overline{e^{iw_j}P_j(r,\varphi-w)}\Big)dw = \frac{1}{(2\pi)^n}\int_{\mathbb{T}^n}\Re\big(r_je^{i\theta_j}\bar P_j(r,\theta)\big)d\theta =: F_j(r), \quad r \in \mathbb{R}^n,$$

where $F_j(r)$ is a continuous function, vanishing with r_j. Since the integrand in the second integral will not change if for some $l = 1,\dots,n$ we replace r_l by $-r_l$ and replace θ_l by $\theta_l + \pi$,

then $F_j(r)$ is even in each its argument r_l. So it may be written as

$$F_j(r_1,\dots,r_n) = f_j(r_1^2,\dots,r_n^2), \qquad f_j \in C(\mathbb{R}^n_+),$$

where $f_j(x_1,\dots,x_n)$ vanishes with x_j. Now assume that the vector field P is C^2-smooth. In this case the integrand in the integral for F_j is C^2-smooth in $(r,\theta) \in \mathbb{R}^n \times \mathbb{T}^n$ and F_j is a C^2-smooth function of r. Then by a result of H. Whitney (see in [27] Theorem 1 with $s = 1$ and the remark, concluding the paper), f_j extends to $\mathbb{R}^n$ in such a way that $f_j(x)$ is C^1-smooth in each its argument x_l and $(\partial/\partial x_l) f_j(x)$ is continuous on $\mathbb{R}^n$. So f_j is C^1-smooth. Since $r_j^2 = 2I_j$, then we have established the following result.

Proposition 6.2. *If the frequency vector Λ is non-resonant and P is C^2 smooth, then equation (6.2) may be written as*

$$dI_j(\tau) - G_j(I_1,\dots,I_n)d\tau - b_j^2 d\tau = b_j\sqrt{2I_j}\, dW_j(\tau), \quad 1 \le j \le n, \tag{6.4}$$

where G is a C^1-smooth vector-field such that $G_j(I)$ vanishes with I_j, for each j.

We stress that although due to the square-root singularity in the dispersion the averaged I-equation (6.4) is a stochastic equation without uniqueness of a solution, still the limiting law $\mathcal{D}(I(\cdot))$ for actions of solutions for eq. (2.6) is uniquely defined by Corollary 4.13.

7. Uniform in time convergence.

In this section we study the uniform in time convergence in distribution of solutions for eq. (2.11) to those for effective equation (4.5), with respect to the dual-Lipschitz metric (see Definition 4.10). These results are finite-dimensional versions of those, obtained in [11] for stochastic PDEs. Throughout the section we assume

Assumption 7.1. The first two items of Assumption 5.1 hold, and

(3′) Effective equation (4.5) is mixing with a stationary measure μ^0. For any its solution $a(\tau)$, $\tau \ge 0$, such that $\mathcal{D}(a(0)) =: \mu$ and $\langle |z|^{2m_0'}, \mu(dz)\rangle = \mathbf{E}|a(0)|^{2m_0'} \le M'$ for some $M' > 0$ (we recall notation (4.22)) we have

$$\|\mathcal{D}(a(\tau)) - \mu^0\|^*_{L,\mathbb{C}^n} \leqslant g_{M'}(\tau,d) \qquad \forall \tau \geqslant 0, \ \text{ if } \|\mu - \mu^0\|^*_{L,\mathbb{C}^n} \leqslant d \leqslant 2. \tag{7.1}$$

Here the function $g : \mathbb{R}^3_+ \to \mathbb{R}_+$, $(\tau, d, M) \mapsto g_M(\tau,d)$, is continuous, vanishes with d, converges to zero when $\tau \to \infty$ and is such that for each fixed $M \ge 0$ the function $(\tau,d) \mapsto g_M(\tau,d)$ is uniformly continuous in d for $(\tau,d) \in [0,\infty) \times [0,2]$. [9]

We emphasize that now we assume mixing for the effective equation, but not for the original equation (2.6). Since Assumptions 7.1 imply Assumptions 2.1, then the assertions of Section 4 with any $T > 0$ hold to solutions of equations (2.11) which we analyse in this section.

Proposition 7.2. *Assume that the first two items of Assumption 5.1 hold, eq. (4.5) is mixing with a stationary measure μ^0, and for each $M > 0$ and any $v^1, v^2 \in \bar{B}_M(\mathbb{C}^n)$ we have*

$$\|\mathcal{D}a(\tau; v^1) - \mathcal{D}a(\tau; v^2)\|^*_{L,\mathbb{C}^n} \le \mathfrak{g}_M(\tau), \tag{7.2}$$

where $\mathfrak{g}$ is a continuous function of (M,τ) which goes to zero when $\tau \to \infty$ and is a non-decreasing function of M. Then condition (3′) holds with some function g.

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

The proposition is proved below in Subsection 7.1.

Note that (7.2) holds (with $\mathfrak{g}$ replaced by $2\mathfrak{g}$) if

$$\|\mathcal{D}a(\tau;v^1)-\mu^0\|^*_{L,\mathbb{C}^n}\le \mathfrak{g}_M(\tau)\quad \forall\, v^1\in \bar B_M(\mathbb{C}^n). \tag{7.3}$$

Usually a proof of mixing for eq. (4.5) in fact establishes (7.3). So condition $(3')$ is a rather mild restriction.

Example 7.3. If the assumption of Proposition 9.4 below is fulfilled, then (7.2) is satisfied since in this case (7.3) holds with $\mathfrak{g}_M(\tau)=\bar V(M)e^{-c\tau}$. Here $c>0$ is a constant and $\bar V(M)=\max\{V(x): x\in \bar B_M(\mathbb{C}^n)\}$, where $V(x)$ is the Lyapunov function as in Proposition 9.3. See e.g. [22, Theorem 2.5] and [20, Section 3.3].

Theorem 7.4. *Under Assumption 7.1, for any $v_0\in\mathbb{C}^n$*

$$\lim_{\varepsilon\to 0}\sup_{\tau\geqslant 0}\|\mathcal{D}(a^\varepsilon(\tau;v_0))-\mathcal{D}(a^0(\tau;v_0))\|^*_{L,\mathbb{C}^n}=0,$$

where $a^\varepsilon(\tau;v_0)$ and $a^0(\tau;v_0)$ solve respectively (2.11) and (4.5) with the same initial condition $a^\varepsilon(0;v_0)=a^0(0;v_0)=v_0$.

Proof. Since v_0 is fixed, we abbreviate $a^\varepsilon(\tau;v_0)$ to $a^\varepsilon(\tau)$. Due to (5.1)

$$\mathbf{E}|a^\varepsilon(\tau)|^{2m_0'}\leqslant C_{m_0'}(|v_0|)=:M^*\quad \forall\,\tau\ge 0. \tag{7.4}$$

By (7.4) and (4.19) [10]

$$\mathbf{E}|a^0(\tau;v_0)|^{2m_0'}=\langle |a|^{2m_0'},\mathcal{D}a^0(\tau;v_0)\rangle\le M^*\quad \forall\,\tau\ge 0. \tag{7.5}$$

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

$$\langle |a|^{2m_0'},\mu^0\rangle\le C_{m_0'}(0)=:C_{m_0'}. \tag{7.6}$$

For later usage we note that since to derive estimates (7.5) and (7.6) we only used Assumptions 5.1.(1), 5.1.(2) and the fact that eq. (4.5) is mixing, then the two estimates hold under the assumptions of Proposition 7.2.

The constants in estimates below depend on M^*, but usually this dependence is not indicated. For any $T\ge 0$ we denote by $a^0_T(\tau)$ a weak solution of effective equation (4.5) such that

$$\mathcal{D}a^0_T(0)=\mathcal{D}a^\varepsilon(T).$$

Note that $a^0_T(\tau)$ depends on ε and that $a^0_0(\tau)=a^0(\tau;v_0)$.

Lemma 7.5. *Take any $\delta>0$. Then*

(1) for any $T>0$ there exists $\varepsilon_1=\varepsilon_1(\delta,T)>0$ such that if $\varepsilon\le\varepsilon_1$, then

$$\sup_{\tau\in[0,T]}\|\mathcal{D}(a^\varepsilon(T'+\tau))-\mathcal{D}(a^0_{T'}(\tau))\|^*_{L,\mathbb{C}^n}\leqslant\delta/2\quad\forall\,T'\geqslant 0. \tag{7.7}$$

(2) Let us choose a $T^*=T^*(\delta)>0$, satisfying $g_{M^*}(T,2)\le\delta/4$ for any $T\ge T^*$. Then there exists $\varepsilon_2=\varepsilon_2(\delta)>0$ such that if $\varepsilon\le\varepsilon_2$ and $\|\mathcal{D}(a^\varepsilon(T'))-\mu^0\|^*_{L,\mathbb{C}^n}\leqslant\delta$ for some $T'\ge 0$, then also

$$\|\mathcal{D}(a^\varepsilon(T'+T^*))-\mu^0\|^*_{L,\mathbb{C}^n}\leqslant\delta, \tag{7.8}$$

[10] Indeed, for any $N>0$ the estimate with $|a|^{2m_0'}$ replaced by $|a|^{2m_0'}\wedge N$ follows from the convergence $\mathcal{D}a^\varepsilon(\tau;v_0)\rightharpoonup\mathcal{D}a^0(\tau;v_0)$. Then the required estimate follows from Fatou's lemma as $N\to\infty$.

and

$$\sup_{\tau\in[T',T'+T^*]} \|\mathcal{D}(a^\varepsilon(\tau)) - \mu^0\|^*_{L,\mathbb{C}^n} \leqslant \tfrac{\delta}{2} + \sup_{\tau\geq 0} g_{M^*}(\tau,\delta). \tag{7.9}$$

Below we abbreviate $\|\cdot\|^*_{L,\mathbb{C}^n}$ to $\|\cdot\|^*_L$, for a measure $\nu\in\mathcal{P}(\mathbb{C}^n)$ denote by $a^\varepsilon(\tau;\nu)$ a weak solution of eq. (2.11) such that $\mathcal{D}(a^\varepsilon(0)) = \nu$, and define $a^0(\tau;\nu)$ similarly. Since eq. (2.11) defines a Markov process in $\mathbb{C}^n$ (e.g. see [14, Section 5.4.C] and [17, Section 3.3]), then

$$\mathcal{D}a^\varepsilon(\tau;\nu) = \int_{\mathbb{C}^n} \mathcal{D}a^\varepsilon(\tau;v)\,\nu(dv),$$

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

Proof. Denote $\nu^\varepsilon = \mathcal{D}(a^\varepsilon(T'))$. Then

$$\mathcal{D}(a^\varepsilon(T'+\tau)) = \mathcal{D}(a^\varepsilon(\tau;\nu^\varepsilon)), \quad \mathcal{D}(a^0_{T'}(\tau)) = \mathcal{D}(a^0(\tau;\nu^\varepsilon)). \tag{7.10}$$

By (7.4), for any $\delta>0$ there exists $K_\delta>0$ such that for each ε, $\nu^\varepsilon(\mathbb{C}^n\smallsetminus\bar B_{K_\delta})\leq\delta/8$, where $\bar B_{K_\delta} := \bar B_{K_\delta}(\mathbb{C}^n)$. So

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

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

$$\mathcal{D}(a^\kappa(\tau;\nu^\varepsilon)) = A^\varepsilon\mathcal{D}(a^\kappa(\tau;\nu^\varepsilon_\delta)) + \bar A^\varepsilon\mathcal{D}(a^\kappa(\tau;\bar\nu^\varepsilon_\delta)), \tag{7.11}$$

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

$$\|\mathcal{D}(a^\varepsilon(\tau;\nu^\varepsilon))-\mathcal{D}(a^0(\tau;\nu^\varepsilon))\|^*_L \leq A^\varepsilon\|\mathcal{D}(a^\varepsilon(\tau;\nu^\varepsilon_\delta))-\mathcal{D}(a^0(\tau;\nu^\varepsilon_\delta))\|^*_L+\bar A^\varepsilon\|\mathcal{D}(a^\varepsilon(\tau;\bar\nu^\varepsilon_\delta))-\mathcal{D}(a^0(\tau;\bar\nu^\varepsilon_\delta))\|^*_L.$$

The second term on the r.h.s obviously is bounded by $2\bar A^\varepsilon\leqslant\frac{\delta}{4}$. While by Proposition 4.12 and (4.23) there exists $\varepsilon_1>0$, depending only on K_δ and T, such that for $0\leq\tau\leq T$ and $\varepsilon\in(0,\varepsilon_1]$ the first term in the r.h.s. is $\leqslant\frac{\delta}{4}$. Due to (7.10) this proves the first assertion.

To prove the second assertion let us choose $\varepsilon_2=\varepsilon_1(\delta/2,T^*(\delta))$. Then we get from (7.7), (7.4), (7.1) and the definition of T^* that for $\varepsilon\leq\varepsilon_2$,

$$\|\mathcal{D}(a^\varepsilon(T'+T^*)) - \mu^0\|^*_L \leqslant \|\mathcal{D}(a^\varepsilon(T'+T^*)) - \mathcal{D}(a^0_{T'}(T^*))\|^*_L + \|\mathcal{D}(a^0_{T'}(T^*)) - \mu^0\|^*_L \leqslant \delta.$$

This proves (7.8). Next, in view of (7.7) and (7.1), (7.5),

$$\sup_{\theta\in[0,T^*]} \|\mathcal{D}(a^\varepsilon(T'+\theta)) - \mu^0\|^*_L \leqslant \sup_{\theta\in[0,T^*]} \|\mathcal{D}(a^\varepsilon(T'+\theta)) - \mathcal{D}(a^0_{T'}(\theta))\|^*_L$$
$$+ \sup_{\theta\in[0,T^*]} \|\mathcal{D}(a^0_{T'}(\theta)) - \mu^0\|^*_L \leqslant \frac{\delta}{2} + \max_{\theta\in[0,T^*]} g_{M^*}(\theta,\delta).$$

This implies (7.9). □

Now we continue to prove the theorem. Let us fix arbitrary $\delta>0$ and take some $0<\delta_1\leq\delta/4$. Below in the proof the functions ε_1, ε_2 and T^* are as in Lemma 7.5.

i) By the definition of T^*, (7.1) and (7.4),

$$\|\mathcal{D}\big(a^0_{T'}(\tau)\big) - \mu^0\|^*_L \leq \delta_1 \quad \forall\,\tau\geq T^*(\delta_1), \tag{7.12}$$

for any $T'\geq 0$. We will abbreviate $T^*(\delta_1)$ to T^*.

ii) By (7.7), if $\varepsilon\leq\varepsilon_1=\varepsilon_1(\delta_1,T^*)>0$, then

$$\sup_{0\leq\tau\leq T^*} \|\mathcal{D}\big(a^\varepsilon(\tau)\big) - \mathcal{D}\big(a^0(\tau;v_0)\big)\|^*_L \leq \tfrac{\delta_1}{2}. \tag{7.13}$$

In particular, in view of (7.12) with $T' = 0$,

$$\|\mathcal{D}\big(a^\varepsilon(T^*)\big) - \mu^0\|_L^* < 2\delta_1. \tag{7.14}$$

iii) By (7.14) and (7.8) with $\delta := 2\delta_1$ and with $T' = nT^*$, $n = 1, 2, \dots$ we get inductively that

$$\|\mathcal{D}\big(a^\varepsilon(nT^*)\big) - \mu^0\|_L^* \le 2\delta_1 \quad \forall\, n \in \mathbb{N}, \tag{7.15}$$

if $\varepsilon \le \varepsilon_2 = \varepsilon_2(2\delta_1)$.

iv) Now by (7.15) and (7.9) with $\delta := 2\delta_1$, for any $n \in \mathbb{N}$ and $0 \le \theta \le T^*$,

$$\|\mathcal{D}\big(a^\varepsilon(nT^* + \theta)\big) - \mu^0\|_L^* \le \delta_1 + \sup_{\theta \ge 0} g_{M^*}(\theta, 2\delta_1), \tag{7.16}$$

if $\varepsilon \le \varepsilon_2(2\delta_1)$.

v) Finally, if $\varepsilon \le \varepsilon_\#(\delta_1) = \min\big(\varepsilon_1(\delta_1, T^*), \varepsilon_2(2\delta_1)\big)$, then by (7.13) if $\tau \le T^*$ and by (7.12)+(7.16) if $\tau \ge T^*$ we have that

$$\|\mathcal{D}\big(a^\varepsilon(\tau)\big) - \mathcal{D}\big(a^0(\tau; v_0)\big)\|_L^* \le 2\delta_1 + \sup_{\theta \ge 0} g_{M^*}(\theta, 2\delta_1) \qquad \forall\, \tau \ge 0.$$

By the assumption, imposed in $(3')$ on function g_M, $g_M(t, d)$ is uniformly continuous in d and vanishes at $d = 0$. So there exists $\delta^* = \delta^*(\delta)$, which we may assume to be $\le \delta/4$, such that if $\delta_1 = \delta^*$, then $g_{M^*}(\theta, 2\delta_1) \le \delta/2$ for every $\theta \geqslant 0$. Then by the estimate above,

$$\|\mathcal{D}\big(a^\varepsilon(\tau)\big) - \mathcal{D}\big(a^0(\tau; v_0)\big)\|_L^* \le \delta \quad \text{if} \quad \varepsilon \le \varepsilon_*(\delta) := \varepsilon_\#\big(\delta^*(\delta)\big) > 0,$$

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

Since the interaction representation does not change actions, then for the action variables of solutions for the original equations (2.6) we have

Corollary 7.6. *Under the assumptions of Theorem 7.4 the actions of a solution $v^\varepsilon(\tau; v_0)$ for eq. (2.6) which equals v_0 at $\tau = 0$ satisfy*

$$\lim_{\varepsilon \to 0} \sup_{\tau \geqslant 0} \|\mathcal{D}(I\big(v^\varepsilon(\tau; v_0)\big)) - \mathcal{D}((I\big(a^0(\tau; v_0)\big)\big)\|_{L, \mathbb{C}^n}^* = 0.$$

In [9, Theorem 2.9] the assertion of the corollary is proved for a class of systems (2.6). The proof in [9] is based on the observation that the mixing rate in the corresponding equation (2.6) is uniform in $0 < \varepsilon \le 1$. This is a delicate property which is harder to establish than $(3')$ in Assumption 7.1. We also note that Theorem 7.4 immediately implies that if equations (2.11) are mixing with stationary measures μ^ε, then $\mu^\varepsilon \rightharpoonup \mu^0$. Cf. Theorem 5.5.

7.1. **Proof of Proposition 7.2.** In this subsection we write solutions $a^0(\tau; v)$ of effective equation (4.5) as $a(\tau; v)$. We will prove the assertion of Proposition 7.2 in four steps.

i) At this step, for any non-random $v^1, v^2 \in \bar{B}_M(\mathbb{C}^n)$ we denote by $a_j(\tau) := a(\tau; v^j)$, $j = 1, 2$, and examine the distance $\|\mathcal{D}(a_1(\tau)) - \mathcal{D}(a_2(\tau))\|_L^*$ as a function of τ and $|v^1 - v^2|$. Let us set $w(\tau) = a_1(\tau) - a_2(\tau)$ and assume that $|v^1 - v^2| \le \bar{d}$ for some $\bar{d} \ge 0$. Then we have

$$\dot{w} = \langle\!\langle P \rangle\!\rangle(a_1) - \langle\!\langle P \rangle\!\rangle(a_2).$$

Since by Lemma 3.2 and Assumption 2.1.(1)

$$|\langle\!\langle P \rangle\!\rangle(a_1(\tau)) - \langle\!\langle P \rangle\!\rangle(a_2(\tau))| \le C|w(\tau)|\, X(\tau), \quad X(\tau) = 1 + |a_1(\tau)|^{m_0} \vee |a_2(\tau)|^{m_0},$$

then $(d/d\tau)|w|^2 \le 2CX(\tau)|w|^2$ with $|w(0)| \le \bar{d}$. So

$$|w(\tau)| \le \bar{d} \exp\big(C \int_0^\tau X(l)\, dl\big). \tag{7.17}$$

Denote $Y(T) = \sup_{0\le\tau\le T}|X(\tau)|$. By (5.1) estimate (2.7) holds with $C_{m_0'}(|v_0|,T) = C_{m_0'}(M)(T+1)$. So by Remark 4.8.2) $\mathbf{E}Y(T) \le (C_{m_0'}(M)+1)(T+1)$ (since $m_0' > (m_0\vee 1)$).

For $K>0$ denote by $\Omega_K(T)$ the event $\{Y(T)\ge K\}$. Then $\mathbf{P}(\Omega_K(T)) \le (C_{m_0'}(M)+1)(T+1)K^{-1}$, and $\int_0^\tau X(l)\,dl \le \tau K$ for $\omega\notin\Omega_K(T)$. From here and (7.17) we see that if f is such that $|f|\le 1$ and $\operatorname{Lip} f\le 1$, then

$$\begin{aligned}\mathbf{E}\big(f(a_1(\tau))-f(a_2(\tau))\big) &\le 2\mathbf{P}(\Omega_K(\tau)) + \bar d\exp\big(C\tau K\big)\\ &= 2(C_{m_0'}(M)+1)(\tau+1)K^{-1} + \bar d\exp\big(C\tau K\big)\quad \forall\, K>0.\end{aligned}\tag{7.18}$$

Let us denote by $g^1_M(\tau,\bar d)$ the function in the r.h.s. with $K = \ln\ln\big(\bar d^{-1}\vee 3\big)$. This is a continuous function of $(\tau,\bar d,M)\in\mathbb{R}^3_+$, vanishing when $\bar d=0$. Due to (7.2) and (7.18),

$$\begin{aligned}\|\mathcal{D}(a(\tau;v^1))-\mathcal{D}(a(\tau;v^2))\|^*_L &= \|\mathcal{D}(a_1(\tau))-\mathcal{D}(a_2(\tau))\|^*_L\\ &\le \mathfrak{g}_M(\tau)\wedge g^1_M(\tau,\bar d)\wedge 2 =: g^2_M(\tau,\bar d)\quad \text{if } |v^1-v^2|\le\bar d.\end{aligned}\tag{7.19}$$

The function g^2_M is continuous in the variables $(\tau,\bar d,M)$, vanishes with $\bar d$ and goes to zero when $\tau\to\infty$ since $\mathfrak{g}_M(\tau)$ does.

ii) At this step we consider a solution $a^0(\tau;\mu) =: a(\tau;\mu)$ of (4.5) with $\mathcal{D}(a(0))=\mu$ as in Assumption 7.1.(3′) and examine the l.h.s of (7.1) as a function of τ. For any $M>0$ consider the conditional probabilities $\mu_M = \mathbf{P}(\cdot\mid \bar B_M(\mathbb{C}^n))$ and $\bar\mu_M = \mathbf{P}(\cdot\mid \mathbb{C}^n\setminus\bar B_M(\mathbb{C}^n))$. Then

$$\mathcal{D}(a(\tau;\mu)) = A_M\mathcal{D}(a(\tau;\mu_M)) + \bar A_M\mathcal{D}(a(\tau;\bar\mu_M)),\tag{7.20}$$

where $A_M = \mu(\bar B_M(\mathbb{C}^n))$ and $\bar A_M = \mu(\mathbb{C}^n\setminus\bar B_M(\mathbb{C}^n))$ (cf. (7.11)). As $\mathbf{E}|a(0)|^{2m_0'}\le M'$, then $\bar A_M = \mathbf{P}\{a(0)>M\}\leqslant M'/M^{2m_0'}$. Since eq. (4.5) is assumed to be mixing, then $\|\mathcal{D}(a(\tau;0))-\mu^0\|^*_L\le\bar g(\tau)$, where $\bar g\ge 0$ is a continuous function, going to 0 as $\tau\to\infty$. So in view of (7.2),

$$\|\mathcal{D}(a(\tau;v))-\mu^0\|^*_L\le\mathfrak{g}_M(\tau)+\bar g(\tau) =: \tilde g_M(\tau),\qquad \forall\, v\in\bar B_M(\mathbb{C}^n).$$

From here,

$$\|\mathcal{D}(a(\tau;\mu_M))-\mu^0\|^*_L = \|\int\big[\mathcal{D}(a(\tau;v))\big]\mu_M(dv)-\mu^0\|^*_L\le\int\|\mathcal{D}(a(\tau;v))-\mu^0\|^*_L\mu_M(dv)\le\tilde g_M(\tau).$$

Therefore due to (7.20),

$$\begin{aligned}\|\mathcal{D}(a(\tau;\mu))-\mu^0\|^*_L &\leqslant A_M\|\mathcal{D}(a(\tau;\mu_M))-\mu^0\|^*_L + \bar A_M\|\mathcal{D}(a(\tau;\bar\mu_M))-\mu^0\|^*_L\\ &\le \|\mathcal{D}(a(\tau;\mu_M))-\mu^0\|^*_L + 2\bar A_M\le\tilde g_M(\tau)+2\frac{M'}{M^{2m_0'}}\quad \text{for any } M>0 \text{ and } \tau\ge 0.\end{aligned}$$

Let $M_1(\tau)>0$ be a continuous non-decreasing function, growing to infinity with τ, and such that $\tilde g_{M_1(\tau)}(\tau)\to 0$ as $\tau\to\infty$ (it exists since $\tilde g_M(\tau)$ is a continuous function of (M,τ), going to 0 as $\tau\to\infty$ for each fixed M). Then

$$\|\mathcal{D}(a(\tau;\mu))-\mu^0\|^*_L\leqslant 2\frac{M'}{M_1(\tau)^{2m_0'}}+\tilde g_{M_1(\tau)}(\tau) =: \hat g_{M'}(\tau).\tag{7.21}$$

Obviously $\hat g_{M'}(\tau)\ge 0$ is a continuous function on $\mathbb{R}^2_+$, converging to 0 as $\tau\to\infty$.

iii) Now we examine the l.h.s of (7.1) as a function of τ and d. Recall that the Kantorovich distance between measures ν_1,ν_2 on $\mathbb{C}^n$ is

$$\|\nu_1-\nu_2\|_K = \sup_{\operatorname{Lip} f\le 1}\langle f,\nu_1\rangle-\langle f,\nu_2\rangle\le\infty.$$

Obviously $\|\nu_1-\nu_2\|_L^* \le \|\nu_1-\nu_2\|_K$. Due to (7.6) and the assumption on μ, the $2m_0'$-moments of μ and μ^0 are bounded by $M'\vee C_{m_0'}$. Thus

$$\|\mu-\mu^0\|_K \le \tilde C(M'\vee C_{m_0'})^{\gamma_1} d^{\gamma_2} := D, \quad \gamma_1=\tfrac{1}{2m_0'},\ \gamma_2=\tfrac{2m_0'-1}{2m_0'}, \tag{7.22}$$

see [5, Section 11.4] and [26, Chapter 7]. So by the Kantorovich–Rubinstein theorem (see [26, 5]) there exist r.v.'s ξ and ξ_0, defined on a new probability space $(\Omega',\mathcal{F}',\mathbf{P}')$, such that $\mathcal{D}(\xi)=\mu$, $\mathcal{D}(\xi_0)=\mu^0$ and

$$\mathbf{E}\,|\xi_1-\xi_0| = \|\mu-\mu^0\|_K. \tag{7.23}$$

Then using (7.19) and denoting by $a_{st}(\tau)$ a stationary solution of eq. (4.5), $\mathcal{D}(a_{st}(\tau))\equiv\mu^0$, we have:

$$\begin{aligned}\|\mathcal{D}(a(\tau))-\mu^0\|_L^* &= \|\mathcal{D}(a(\tau;a(0)))-\mathcal{D}(a_{st}(\tau))\|_L^* \le \mathbf{E}^{\omega'}\|\mathcal{D}(a(\tau;\xi^{\omega'}))-\mathcal{D}(a(\tau;\xi_0^{\omega'}))\|_L^*\\ &\le \mathbf{E}^{\omega'} g_{\bar M}^2(\tau,|\xi^{\omega'}-\xi_0^{\omega'}|), \qquad \bar M=\bar M^{\omega'}=|\xi^{\omega'}|\vee|\xi_0^{\omega'}|.\end{aligned}$$

As $\mathbf{E}^{\omega'}\bar M^{2m_0'}\le 2(M'\vee C_{m_0'})$ by (7.5) and the assumption on μ, then denoting $Q_K'=\{\bar M\ge K\}\subset\Omega'$, for any $K>0$ we have

$$\mathbf{P}^{\omega'}(Q_K')\le 2K^{-2m_0'}(M'\vee C_{m_0'}).$$

Since $g_M^2\le 2$ and for $\omega'\notin Q_K'$ we have $|\xi^{\omega'}|,|\xi_0^{\omega'}|\le K$, then

$$\|\mathcal{D}(a(\tau))-\mu^0\|_L^*\le 4K^{-2m_0'}(M'\vee C_{m_0'})+\mathbf{E}^{\omega'}g_K^2(\tau,|\xi^{\omega'}-\xi_0^{\omega'}|).$$

Now let $\Omega_r'=\{|\xi^{\omega'}-\xi_0^{\omega'}|\ge r\}$. Then by (7.23) and (7.22), $\mathbf{P}^{\omega'}\Omega_r'\le Dr^{-1}$. So

$$\|\mathcal{D}(a(\tau))-\mu^0\|_L^*\le 4K^{-2m_0'}(M'\vee C_{m_0'})+2Dr^{-1}+g_K^2(\tau,r) \qquad \forall\,\tau\ge 0,\ \forall\,K,r>0. \tag{7.24}$$

iv) End of the proof. Let $g_0(s)$ be a positive continuous function on $\mathbb{R}_+$ such that $g_0(s)\to\infty$ as $s\to+\infty$ and $|C_{m_0'}\big(g_0(s)\big)(\ln\ln s)^{-1/2}|\leqslant 2C_{m_0'}(0)$ for $s\geqslant 3$. Taking $r=D^{1/2}$ and choosing in the r.h.s. of (7.24) $K=g_0(r^{-1})$, we denote the r.h.s as $g_{M'}^3(\tau,r)$ (so we substitute in (7.24) $D=r^2$ and $K=g_0(r^{-1})$). By (7.24) and the definition of g_M^2 (see (7.18) and (7.19)), we have

$$\begin{aligned}g_{M'}^3(\tau,r)&\leqslant 4(g_0(r^{-1}))^{-2m_0'}(M'\vee C_{m_0'})+2r\\ &\quad+2(C_{m_0'}(g(r^{-1}))+1)\big(\ln\ln(r^{-1}\vee 3)\big)^{-1}+r\exp\big(C\tau\ln\ln(r^{-1}\vee 3)\big).\end{aligned}$$

By the choice of g_0, as $r\to 0$ the first, the second and the fourth terms converge 0. The third term is $\leqslant 4(C_{m_0'}(0)+1)(\ln\ln(r^{-1}))^{-1/2}$ for $r\leqslant\frac13$, so it also converges to zero with r. Hence $g_{M'}^3(\tau,r)$ defines a continuous function on $\mathbb{R}_+^3$, vanishing with r. Using the expression for D in (7.22) let us write $r=D^{1/2}$ as $r=R_{M'}(d)$, where R is a continuous function $\mathbb{R}_+^2\to\mathbb{R}_+$, non-decreasing in d and vanishing with d. Setting $g_{M'}^4(\tau,d)=g_{M'}^3(\tau,R_{M'}(d\wedge 2))$ we get from the above that

$$\|\mathcal{D}(a(\tau))-\mu^0\|_L^*\le g_{M'}^4(\tau,d) \quad \text{if } \|\mu-\mu^0\|_L^*\le d\le 2.$$

Finally, evoking (7.21) we arrive at (7.1) with $g=g^5$, where

$$g_{M'}^5(\tau,d)=g_{M'}^4(\tau,d)\wedge\hat g_{M'}(\tau)\wedge 2.$$

The function g^5 is continuous, vanishes with d and converges to zero as $\tau\to\infty$. For any fixed $M'>0$ this convergence is uniform in d due to the term $\hat g_{M'}(\tau)$. So for a fixed $M'>0$ the function $(\tau,d)\mapsto g_{M'}^5(\tau,d)$ extends to a continuous function on the compact set $[0,\infty]\times[0,2]$, where it vanishes when $\tau=\infty$. Thus $g_{M'}^5$ is uniformly continuous in d, and the assertion of the proposition is proved.

8. Averaging for systems with general noises

In this section we sketch a proof of Theorem 4.7 for equations (1.1) with general stochastic terms $\sqrt{\varepsilon}\,\mathcal{B}(v)\,dW$. The proof follows the argument in Section 4 with an extra difficulty which appears in the case of equations with non-additive degenerate noises.

Let us consider v-equation (2.6) with a general noise (possibly non-additive) and decomplexify it by writing the components $v_k(\tau)$ as $(\tilde{v}_{2k-1}(\tau), \tilde{v}_{2k}(\tau)) \in \mathbb{R}^2$, $k = 1, \dots, n$. Now a solution $v(\tau)$ is a vector in $\mathbb{R}^{2n}$ and the equation reads

$$dv(\tau) + \varepsilon^{-1} A v(\tau)\,d\tau = P(v(\tau))\,d\tau + \mathcal{B}(v(\tau))\,d\beta(\tau), \qquad v(0) = v_0 \in \mathbb{R}^{2n}. \tag{8.1}$$

Here A is the block-diagonal matrix as in Section 2, $\mathcal{B}(v)$ is a real $2n \times n_2$-matrix and $\beta(\tau) = (\beta_1(\tau), \dots, \beta_{n_2}(\tau))$, where $\{\beta_j(\tau)\}$ are independent standard real Wiener processes. Note that in the real coordinates in $\mathbb{R}^{2n} \simeq \mathbb{C}^n$ the operator Φ_w in (2.10) with some vector $w \in \mathbb{R}^n$ is given by a block-diagonal matrix, where for $j = 1, \dots, n$ its j-th diagonal block is the 2×2-matrix of rotation by the angle w_j.

In this section we assume that

Assumption 8.1. The dirft $P \in \mathrm{Lip}_{m_0}(\mathbb{R}^{2n}, \mathbb{R}^{2n})$, the matrix-function $\mathcal{B}(v)$ belongs to $\mathrm{Lip}_{m_0}\big(\mathbb{R}^{2n}, M(2n \times n_2)\big)$, equation (8.1) is well-posed and its solutions satisfy (2.7).

Passing to the interaction representation $v(\tau) = \Phi_{\tau\varepsilon^{-1}\Lambda} a(\tau)$ we rewrite the equation as

$$da(\tau) = \Phi_{\tau\varepsilon^{-1}\Lambda} P(v(\tau))\,d\tau + \Phi_{\tau\varepsilon^{-1}\Lambda}\mathcal{B}(v(\tau))\,d\beta(\tau), \qquad a(0) = v_0. \tag{8.2}$$

As in Section 4 we will see that as $\varepsilon \to 0$, the asymptotic behaviour of distributions of the equation's solutions is described by an effective equation. As before, the effective drift is $\langle\!\langle P \rangle\!\rangle(a)$. To calculate the effective dispersion, as in the proof of Lemma 4.6 we consider the martingale

$$N^{Y,\varepsilon} := a^\varepsilon(\tau) - \int_0^\tau Y(a^\varepsilon(s), s\varepsilon^{-1})ds = v_0 + \int_0^\tau \mathcal{B}^\Lambda(a^\varepsilon(s); s\varepsilon^{-1})d\beta(s),$$

where Y is defined in (4.7) and $\mathcal{B}^\Lambda(a;t) = \Phi_{t\Lambda}\mathcal{B}(\Phi_{-t\Lambda}a)$. By Itô's formula, for $i, j = 1, \dots, n$ the process

$$N_i^{Y,\varepsilon}(\tau)N_j^{Y,\varepsilon}(\tau) - \int_0^\tau \mathcal{A}_{ij}^\Lambda(a^\varepsilon(s); s\varepsilon^{-1})ds, \qquad (\mathcal{A}_{ij}^\Lambda(a;t)) = \mathcal{B}^\Lambda(a;t)\mathcal{B}^{\Lambda *}(a;t),$$

where $\mathcal{B}^{\Lambda *}$ is the transpose of $\mathcal{B}^\Lambda$, also is a martingale. By a straightforward analogy of Lemma 3.2, the limit

$$\mathcal{A}^0(a) := \lim_{T\to\infty} \frac{1}{T}\int_0^T \mathcal{A}^\Lambda(a;t)dt$$

exists and belongs to $\mathrm{Lip}_{2m_0}(\mathbb{R}^{2n}, M(2n \times 2n))$. Then, by analogy with Lemma 4.3,

$$\mathbf{E}\Big|\int_0^\tau \mathcal{A}^\Lambda(a^\varepsilon(s); s\varepsilon^{-1})ds - \int_0^\tau \mathcal{A}^0(a^\varepsilon(s))ds\Big| \to 0, \quad \text{as } \varepsilon \to 0,$$

for any $\tau \geqslant 0$. From here, as in Section 4, we conclude that now for the effective diffusion we should take $\mathcal{A}^0(a)$, which is a non-negative symmetric matrix. Denoting by $\mathcal{B}^0(a) = \mathcal{A}^0(a)^{1/2}$ its principal square root, as in Section 4 we verify that any limiting measure Q_0 as in (2.14) is a solution for the martingale problem for the effective equation

$$da(\tau) - \langle\!\langle P \rangle\!\rangle(a(\tau))d\tau = \mathcal{B}^0(a(\tau))d\beta(\tau), \qquad a(0) = v_0, \tag{8.3}$$

and so is a weak solution of the equation. If the noise in equation (8.1) is additive, then $\mathcal{B}^0$ is a constant matrix, eq. (8.3) has a unique solution and Theorem 4.7 with (modified) effective

equation (8.3) remains true for solutions of eq. (8.2). In particular, the theorem applies to equation (2.5) with general additive random forces (2.4) (but then the effective dispersion matrix is given by a more complicated formula than in Section 4).

Similarly, if the diffusion in (8.1) is non-degenerate in the sense that

$$|\mathcal{B}(v)\mathcal{B}^*(v)\xi| \geqslant \alpha|\xi|, \qquad \forall\, v,\ \forall\, \xi \in \mathbb{R}^{2n}, \tag{8.4}$$

for some $\alpha > 0$, then the matrix $\mathcal{B}^\Lambda(a,\tau)$ also satisfies (8.4) for all a and τ, i.e. $\langle \mathcal{A}^\Lambda(a; s\varepsilon^{-1})\xi, \xi\rangle \geqslant \alpha|\xi|^2$. Thus $\mathcal{A}^0(a) \geqslant \alpha\mathbb{I}$, and so $\mathcal{B}^0(a) = \mathcal{A}^0(a)^{1/2}$ is a locally Lipschitz matrix function of a (e.g. see [24, Theorem 5.2.2]). So again eq. (8.3) has a unique solution and Theorem 4.7 remains true for eq. (8.1) (with the effective equation of the form (8.3)).

To treat equations (8.1) with degenerate non-additive noises we write the matrix $\mathcal{A}^0(a)$ as

$$\mathcal{A}^0(a) = \lim_{T\to\infty} \frac{1}{T}\int_0^T \big(\Phi_{t\Lambda}\mathcal{B}(\Phi_{-t\Lambda}a)\big)\cdot\big(\Phi_{t\Lambda}\mathcal{B}(\Phi_{-t\Lambda}a)\big)^* dt.$$

By the same reason as in Proposition 3.4 we find that

$$|\mathcal{A}^0|_{C^2(B_R)} \leqslant C|\mathcal{B}|^2_{C^2(B_R)}, \quad \forall R > 0.$$

Now using [24, Theorem 5.2.3], we get that

$$\mathrm{Lip}(\mathcal{B}^0(a)|_{\bar{B}_R}) \leqslant C|\mathcal{A}^0|^{1/2}_{C^2(B_{R+1})} \leqslant C_1|\mathcal{B}|_{C^2(B_{R+1})}, \quad \forall R > 0. \tag{8.5}$$

So the matrix-function $\mathcal{B}^0(a)$ is locally Lipschitz continuous, eq. (8.3) has a unique solution and the assertion of Theorem 4.7 remains true for eq. (8.1). We have obtained

Theorem 8.2. *Suppose that Assumption 8.1 holds and for the matrix function $\mathcal{B}(v)$ in (8.1) one of the following three options is true:*

(1) it is v-independent;
(2) it satisfies the non-degeneracy condition (8.4);
(3) it is a C^2-smooth matrix-function of v.

Then for any $v_0 \in \mathbb{R}^{2n}$ a solution $a^\varepsilon(\tau; v_0)$ of eq. (8.2) satisfies

$$\mathcal{D}(a^\varepsilon(\cdot; v_0)) \rightharpoonup Q_0 \text{ in } \mathcal{P}(C([0,T],\mathbb{C}^n)) \quad \text{as } \varepsilon \to 0,$$

where Q_0 is the law of a unique weak solution of the effective equation (8.3).

An obvious analogy of Corollary 4.13 holds for solutions of eq. (8.1).

9. A sufficient condition for Assumptions 2.1, 5.1 and 7.1

In this section we derive a condition which implies Assumptions 2.1, 5.1 and 7.1. So when it is met, all theorems in Sections 4, 5 and 7 apply to eq. (2.6).

Consider a stochastic differential equation on $\mathbb{R}^l$,

$$dx = b(x)d\tau + \sigma(x)d\beta(\tau), \quad x \in \mathbb{R}^l,\ \tau \geqslant 0, \tag{9.1}$$

where $\sigma(x)$ is an $l\times k$-matrix and $\beta(\tau)$ is a standard Wiener processes in $\mathbb{R}^k$. We assume

Assumption 9.1. The drift $b(x)$ and dispersion $\sigma(x)$ are locally Lipschitz in x, and $\mathcal{C}^m(b)$, $\mathcal{C}^m(\sigma) \leq C < \infty$ for some $m \geq 0$.

The diffusion $a(x) = \sigma(x)\cdot\sigma^t(x)$ is a nonnegative symmetric $l\times l$-matrix. Consider the differential operator

$$\mathscr{L}\big(v(x)\big) = \sum_{j=1}^l b_j(x)\frac{\partial v}{\partial x_j} + \frac{1}{2}\sum_{i=1}^l\sum_{j=1}^l a_{ij}(x)\frac{\partial^2 v}{\partial x_i \partial x_j}.$$

We have the following result concerning the well-posedness of eq. (9.1) from [17, Theorem 3.5]:

Theorem 9.2. *Assume Assumption 9.1 and suppose that there exists a nonnegative function $V(x) \in C^2(\mathbb{R}^l)$ such that for some constant $c > 0$ we have*

$$\mathscr{L}(V(x)) \leqslant cV(x) \quad \forall \tau \geqslant 0,\ x \in \mathbb{R}^l,$$

and

$$\inf_{|x|>R} V(x) \to \infty \quad \text{as } R \to \infty.$$

Then for any $x_0 \in \mathbb{R}^l$ equation (9.1) has a unique strong solution $X(\tau)$ with the initial condition $X(0) = x_0$. Furthermore, the process $X(\tau)$ satisfies

$$\mathbf{E}V(X(\tau)) \leqslant e^{c\tau}V(x_0) \quad \forall \tau \geqslant 0.$$

The function V is called *a Lyapunov function* for equation (9.1). In its terms a sufficient condition for the mixing in equation (9.1) is given by the following statement:

Proposition 9.3. *Assume that in addition to Assumption 9.1 we have*

(1) the drift b satisfies

$$\langle b(x), x\rangle \leqslant -\alpha_1|x| + \alpha_2 \qquad \forall x \in \mathbb{R}^l, \tag{9.2}$$

for some constants $\alpha_1 > 0$ and $\alpha_2 \geqslant 0$, where $\langle\cdot,\cdot\rangle$ is the standard inner product in $\mathbb{R}^l$.

(2) The diffusion matrix $a(x) = \sigma(x)\sigma(x)^t$ is uniformly non-degenerate, that is

$$\gamma_2\mathbb{I} \leqslant a(x) \leqslant \gamma_1\mathbb{I} \quad \forall x \in \mathbb{R}^l, \tag{9.3}$$

for some $\gamma_1 \geqslant \gamma_2 > 0$.

Then for any $c' > 0$ equation (9.1) has a smooth Lyapunov function $V(x)$, equal $\exp(c'|x|)$ for $|x| \geq 1$, estimate (5.1) holds true for its solutions for every $m \in \mathbb{N}$, and the equation is mixing.

In Appendix A it is shown how to derive the proposition from abstract results in [17]. Moreover, it can be proved that under the assumptions of the proposition the equation is exponentially mixing and (7.3) holds, see Example 7.3.

Let us decomplexify $\mathbb{C}^n$ to $\mathbb{R}^{2n}$ and identify equation (2.6) with a real equation (9.1), where $l = 2n$ (and $x = v$). Then $b(v) \cong (b_j(v) = (-i\varepsilon^{-1}\lambda_j v_j + P_j(v),\ j = 1,\dots,n)$, where $b_j \in \mathbb{C} \cong \mathbb{R}^2 \subset \mathbb{R}^{2n}$. Since in complex terms the real inner product reads as $\langle v, w\rangle = Re\sum v_j\bar{w}_j$, then

$$\langle b(v), v\rangle = \langle P(v), v\rangle.$$

So for equation (2.6) condition (9.2) is equivalent to

$$\langle P(v), v\rangle \leqslant -\alpha_1|v| + \alpha_2, \quad \forall v \in \mathbb{C}^n, \tag{9.4}$$

for some constants $\alpha_1 > 0$ and $\alpha_2 \geqslant 0$.

Now consider effective equation (4.5). Since in eq. (2.11) the drift is

$$Y(a, \tau\varepsilon^{-1}) = (\Phi_{\tau\varepsilon^{-1}\Lambda})_* P(a),$$

then under the assumption (9.4) we have

$$\langle Y(a, \tau\varepsilon^{-1}), a\rangle = \langle P(\Phi_{\tau\varepsilon^{-1}\Lambda}a), \Phi_{\tau\varepsilon^{-1}\Lambda}a\rangle \leqslant -\alpha_1|\Phi_{-\tau\varepsilon^{-1}\Lambda}a| + \alpha_2 = -\alpha_1|a| + \alpha_2,$$

for all ε. Therefore $\langle\!\langle P\rangle\!\rangle$ satisfies

$$\langle\langle\!\langle P\rangle\!\rangle(a), a\rangle = \lim_{T\to\infty}\frac{1}{T}\int_0^T \langle Y(a, \tau\varepsilon^{-1}), a\rangle d\tau \leqslant -\alpha_1|a| + \alpha_2.$$

We saw that assumption (9.4) implies the validity of condition (9.2) also for the effective equation.

As it was pointed out, if the dispersion matrix Ψ is non-degenerate, then the dispersion B in the effective equation is non-degenerate as well. The corresponding diffusion matrix also is non-degenerate and condition (9.3) holds for it. Thus we obtained the following statement:

Proposition 9.4. *If in equation (2.6) the dispersion matrix Ψ is non-degenerate, the drift $P \in Lip_{m_0}$ for some $m_0 \in \mathbb{N}$ and (9.4) holds for some constants $\alpha_1 > 0$ and $\alpha_2 \geqslant 0$, then the assumption of Theorem 5.5 holds true, and so does the assumption of Theorems 4.7 and 7.4.*

Appendix A. Proof of Proposition 9.3

By condition (9.3) the diffusion a is uniformly bounded. So there exist constants $k_1, k_2 > 0$ such that

$$\mathrm{Tr}(a(x)) \leqslant k_1, \ \|a(x)\| \leqslant k_2 \quad \forall x \in \mathbb{R}^l. \tag{A.1}$$

Take $V(x) = e^{c' f(x)}$, where $c' > 0$ is a positive constant and $f(x)$ is a non-negative smooth function, equal $|x|$ for $|x| \geq 1$, and such that its first and second derivatives are bounded by 3. Then

$$\frac{\partial V(x)}{\partial x_j} = c' V(x) \partial_{x_j} f(x), \quad \frac{\partial^2 V(x)}{\partial x_j \partial x_j} = c' V(x) \partial_{x_i x_j} f(x) + c'^2 V(x) \partial_{x_i} f(x) \partial_{x_j} f(x).$$

Therefore we have that

$$\mathscr{L}\big(V(x)\big) = c' V(x) \mathcal{K}(c', x),$$

where

$$\mathcal{K}(c', x) = \sum_{j=1}^{l} b_j(x) \partial_{x_j} f(x) + \frac{1}{2} \sum_{ij} a_{ij}(x) \partial_{x_i x_j} f(x) + \frac{1}{2} c' \sum_{ij} a_{ij}(x) \partial_{x_i} f(x) \partial_{x_j} f(x).$$

From the condition (9.2) and (A.1) it is evident that

$$\begin{cases} |\mathcal{K}(c', x)| \leqslant (c'+1)C, & \text{if } |x| < 1, \\ \mathcal{K}(c', x) \leqslant (-\alpha_1 + \frac{\alpha_2}{|x|} + \frac{C}{|x|} + c'C), & \text{if } |x| \geqslant 1, \end{cases} \tag{A.2}$$

where $C > 0$ is a constant depending on k_1, k_2 and $\sup_{|x| \leqslant 1} |b(x)|$. Then we achieve that

$$\mathscr{L}\big(V(x)\big) \leqslant c V(x) \quad \forall x \in \mathbb{R}^l,$$

where $c = c'(\alpha_2 + (c'+1)C)$. Clearly $\inf_{|x|>R} V(x) \to \infty$ as $R \to \infty$. So $V(x)$ is a Lyapunov function for equation (9.1). Then by Theorem 9.2 for any $x_0 \in \mathbb{R}^l$ the equation has a unique solution $X(\tau) = X(\tau; x_0)$, equal x_0 at $\tau = 0$, which satisfies

$$\mathbf{E} e^{c' f(X(\tau))} \leqslant e^{c\tau} e^{c' f(x_0)} \quad \forall \tau \geqslant 0.$$

Let us apply Itô's formula to the process $F\big(X(\tau)\big) = e^{\eta' f(X(\tau))}$, where $0 < \eta' \leqslant \frac{1}{2} c'$ is a constant to be determined later. Then

$$\begin{aligned} dF\big(X\big) &= \mathscr{L}\big(F(X)\big) d\tau + \eta' F(X) \langle \nabla f(x), \sigma^t(X) dW \rangle \\ &= \eta' F(X) \mathcal{K}(\eta', X) d\tau + \eta' F(X) \langle \nabla f(x), \sigma^t(X) dW \rangle. \end{aligned}$$

By (A.2), choosing $\eta' = \min\{\frac{\alpha_1}{4C}, \frac{c'}{2}\}$, we have

$$F(X) \mathcal{K}(\eta', X) \leq -\frac{\alpha_1}{2} F(X) + C_0(\alpha_1, \eta', k_1, k_2)$$

uniformly in X. Then

$$dF(X) \leqslant \big(-\frac{\alpha_1}{2}\eta' F(X) + C_0\big)d\tau + \eta' F(X)\langle \nabla f(x), \sigma^t(X)dW\rangle, \tag{A.3}$$

where the constant $C_0 > 0$ depends on k_1, k_2, α_1, η' and α_2. Taking expectation and applying Gronwall's lemma we obtain that

$$\mathbf{E}e^{\eta' f(X(\tau))} \leqslant e^{-\frac{\alpha_1}{2}\eta'\tau} e^{\eta' f(x_0)} + C_1, \quad \tau \geqslant 0, \tag{A.4}$$

where $C_1 > 0$ depends on the same parameters as C_0.

Now let us take any $T \geqslant 0$ and for $\tau \in [T, T+1]$ consider relation (A.3), where $F(X)$ is replaced by $\tilde{F}(X) = e^{\tilde{\eta} f(X)}$ with $0 < \tilde{\eta} \leq \frac{1}{2}\eta'$, and integrate it from T to τ:

$$\begin{aligned} \tilde{F}(X(\tau)) &\leq \tilde{F}(X(T)) + C_0 + \tilde{\eta}\int_T^\tau \tilde{F}(X)\langle \nabla f(x), \sigma^t(s,X)dW\rangle \\ &=: \tilde{F}(X(T)) + C_0 + \mathcal{M}(\tau). \end{aligned} \tag{A.5}$$

Due to estimate (A.4), $\mathcal{M}(\tau)$ is a continuous square-integrable martingale. Therefore by Doob's inequality,

$$\mathbf{E}\sup_{T\leq\tau\leq T+1}|\mathcal{M}(\tau)|^2 \leq 4\mathbf{E}|\mathcal{M}(T+1)|^2 \leq C\int_T^{T+1}\mathbf{E}\tilde{F}^2(X(s))ds \leq C\int_T^{T+1}\mathbf{E}F(X(s))ds \leq C',$$

where C' depends on k_1, k_2, $\alpha_1, \eta', \alpha_2$ and $|x_0|$. From here, (A.5) and (A.4) it follows that

$$\mathbf{E}\sup_{T\leq\tau\leq T+1} e^{\tilde{\eta} f(X(\tau))} \leq C'',$$

where C'' depends on the same parameters as C'. This bound implies that solutions $X(\tau)$ satisfy estimate (5.1) in Assumption 5.1 for every $m \geq 0$.

To prove the proposition it remains to show that under the imposed assumptions equation (9.1) is mixing. Due to [17, Theorem 4.3] we just need to verify that there exists an absorbing ball $B_R = \{|x| \leqslant R\}$ such that for any compact set $K \subset \mathbb{R}^l \setminus B_R$

$$\sup_{x_0\in K}\mathbf{E}\tau(x_0) < \infty, \tag{A.6}$$

where $\tau(x_0)$ is the hitting time for B_R of trajectory $X(\tau; x_0)$. Indeed, let $x_0 \in K \subset \mathbb{R}^l \setminus B_R$ for some $R > 0$ to be determined later. We set $\tau_M := \min\{\tau(x_0), M\}$, $M > 0$. Applying Itô's formula to the process $F(\tau, X(\tau)) = e^{\frac{1}{4}\eta'\alpha_1\tau}|X(\tau)|^2$ and using (A.4) we find that

$$dF(\tau, X(\tau)) = \Big(\frac{\eta'\alpha_1}{4}F(\tau, X(\tau)) + \mathscr{L}\big(F(\tau, X(\tau))\big)\Big)d\tau + d\mathcal{M}(\tau),$$

where $\mathcal{M}(\tau)$ is a corresponding stochastic integral. By (A.1), (A.4) and (9.2), we have

$$\mathbf{E}e^{\frac{\eta'\alpha_1}{4}\tau_M}|X(\tau_M)|^2 + \mathbf{E}\int_0^{\tau_M} e^{\frac{\eta'\alpha_1 s}{4}}(2\alpha_1|X(\tau)| - C_3)ds \leqslant |x_0|^2 + 2e^{\eta' f(x_0)} =: \gamma(x_0),$$

where $C_3 > 0$ depends on α_1, α_2, k_1 and k_2. Since $|X(s)| \geqslant R$ for $0 \leqslant s \leqslant \tau_M$, then

$$\mathbf{E}\Big(C_3\int_0^{\tau_M} e^{\frac{\eta'\alpha_1 s}{4}}ds\Big) \leqslant \gamma(x_0),$$

if $R \geq C_3/\alpha_1$. Therefore $\mathbf{E}\tau_M \leqslant \gamma(x_0)/C_3$. Letting $M \to \infty$ we verify (A.6) for $R \geq C_3/\alpha_1$. This completes the proof of Proposition 9.3.

Appendix B. Representation of martingales

Let $\{M_k(t),\ t \in [0,T]\}$, $k = 1,\dots,d$, be continuous square-integrable martingales on a filtered probability space $(\Omega, \mathcal{F}, \mathbf{P}, \{\mathcal{F}_t\})$. We recall that their brackets (or their cross-variational process) is an $\{\mathcal{F}_t\}$-adapted continuous matrix-valued process of bounded variation $\langle M_k, M_j\rangle(t), 1 \le k, j \le d$, a.s. vanishing at $t = 0$, and such that for all k, j the process $M_k(t)M_j(t) - \langle M_k, M_j\rangle(t)$ is an $\{\mathcal{F}_t\}$-martingale; see [14], Definition 1.5.5 and Theorem 1.5.13.

Theorem B.1 ([14], Theorem 3.4.2)**.** *Let $(M_k(t), 1 \le k \le d)$ be a vector of martingale as above. Then there exists an extension $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbf{P}}, \tilde{\mathcal{F}}_t)$ of the probability space on which are defined independent standard Wiener processes $W_1(t), \dots, W_d(t)$, and a measurable adapted matrix $X = (X_{kj}(t))_{k,j=1,\dots,d}$, $t \in [0,T]$ such that $\mathbf{E}\int_0^T \|X(s)\|^2 ds < \infty$, and $\tilde{\mathbf{P}}$-a.s. we have the representations*

$$M_k(t) - M_k(0) = \sum_{j=1}^{d} \int_0^t X_{kj}(s) dW_j(s), \quad 1 \leqslant k \leqslant d,\ t \in [0,T],$$

and

$$\langle M_k, M_j\rangle(t) = \sum_{l=1}^{d} \int_0^t X_{kl}(s) X_{jl}(s) ds, \quad 1 \leqslant k, j \leqslant d,\ t \in [0,T].$$

Now let $(N_1(t), \dots, N_d(t)) \in \mathbb{C}^d$ be a vector of complex continuous square-integrable martingales. Then $N_j(t) = N_j^+(t) + iN_j^-(t)$, where $\big(N_1^+(t), N_1^-(t), \dots, N_d^+(t), N_d^-(t)\big) \in \mathbb{R}^{2d}$ is a vector of real continuous martingales. The brackets $\langle N_i, N_j\rangle$ and $\langle N_i, \bar{N}_j\rangle$ are defined by linearity. For example, $\langle N_i, N_j\rangle = \langle N_i^+, N_j^+\rangle - \langle N_i^-, N_j^-\rangle + i\langle N_i,^+ N_j^-\rangle + i\langle N_i,^- N_j^+\rangle$.[11] Equivalently $\langle N_i, N_j\rangle$ may be be defined as the only adapted continuous complex process of bounded variation, vanishing at zero, and such that $N_iN_j - \langle N_i, N_j\rangle$ is a martingale. The brackets $\langle N_i, \bar{N}_j\rangle$ may be defined similarly. The result above implies a representation theorem for complex continuous martingales. Below we give its special case, relevant for our work.

Corollary B.2. *Suppose that the brackets $\langle N_i, N_j\rangle(t)$ and $\langle \bar{N}_i, \bar{N}_j\rangle(t)$ all vanish, while the brackets $\langle N_i, \bar{N}_j\rangle(t)$, $1 \le i, j \le d$, a.s. are absolutely continuous complex processes. Then there exist an adapted process $\Psi(t)$, valued in $d \times d$ complex matrices, satisfying $\mathbf{E}\int_0^T \|\Psi(t)\|^2 dt < \infty$, and independent standard complex Wiener processes $\beta_1^c(t), \dots, \beta_d^c(t)$, all defined on an extension of the original probability space, such that a.s.*

$$N_j(t) - N_j(0) = \sum_{k=1}^{d} \int_0^t \Psi_{jk}(s) d\beta_k^c(s) \quad \forall\, 0 \le t \le T,\ j = 1, \dots, d.$$

Moreover, $\langle N_i, N_j\rangle(t) \equiv 0$ and

$$\langle N_i, \bar{N}_j\rangle(t) = 2\int_0^t (\Psi\Psi^*)_{ij}(s) ds, \quad 1 \le i, j \le d.$$

Appendix C. Itô's formula for complex processes

Consider a complex Itô process $v(t) \in \mathbb{C}^n$, defined on a filtered probability space:

$$dv(t) = g(t)\, dt + M^1(t)\, dB(t) + M^2(t)\, d\bar{B}(t)\,. \tag{C.1}$$

[11] There is no need to define the brackets $\langle \bar{N}_i, \bar{N}_j\rangle$ and $\langle \bar{N}_i, N_j\rangle$ since these are just the processes, complex-conjugated to $\langle N_i, N_j\rangle$ and $\langle N_i, \bar{N}_j\rangle$, respectively.

Here $v(t)$ and $g(t)$ are adapted processes in $\mathbb{C}^n$, M^1 and M^2 are adapted processes in the space of complex $n\times N$–matrices, and $B(t)=(\beta_1^c(t),\dots,\beta_N^c(t))$, $\bar B(t)=(\bar\beta_1^c(t),\dots,\bar\beta_N^c(t))$, where $\{\beta_j^c\}$ are independent standard complex Wiener processes. We recall that for a C^1–function f on $\mathbb{C}^n$, $\partial f/\partial z_j=\frac12\big(\partial f/\partial x_j-i\partial f/\partial y_j\big)$ and $\partial f/\partial\bar z_j=\frac12\big(\partial f/\partial x_j+i\partial f/\partial y_j\big)$. If f is a polynomial of z_j and $\bar z_j$, then $\partial f/\partial z_j$ and $\partial f/\partial\bar z_j$ may be calculated as if z_j and $\bar z_j$ are independent variables.

The processes g,M^1,M^2 and the function $f(t,v)$ in the theorem below are assumed to satisfy the usual conditions, needed for the applicability of Itô's formula (e.g. see in [14]), which we do not repeat here.

Theorem C.1. *Let $f(t,v)$ be a C^2–smooth complex function. Then*

$$\begin{aligned}df(t,v(t))&=\Big[\partial f/\partial t+d_vf(t,v)g+d_{\bar v}f(t,v)\bar g\\&+\mathrm{Trace}\Big[\big(M^1(M^2)^t+M^2(M^1)^t\big)\frac{\partial^2 f}{\partial v\partial v}+\big(\bar M^1(\bar M^2)^t+\bar M^2(\bar M^1)^t\big)\frac{\partial^2 f}{\partial\bar v\partial\bar v}\\&+2\big(M^1(\bar M^1)^t+M^2(\bar M^2)^t\big)\frac{\partial^2 f}{\partial v\partial\bar v}\Big]\Big]dt\\&+d_vf(M^1dB+M^2d\bar B)+d_{\bar v}f(\bar M^1d\bar B+\bar M^2dB)\,.\end{aligned}\tag{C.2}$$

Here $d_vf(t,v)g=\sum\frac{\partial f}{\partial v_j}g_j$, $d_{\bar v}f(t,v)\bar g=\sum\frac{\partial f}{\partial\bar v_j}\bar g_j$, $\frac{\partial^2 f}{\partial v\partial v}$ stands for the matrix with entries $\frac{\partial^2 f}{\partial v_j\partial v_k}$, etc. If the function f is real valued, then $d_{\bar v}f(v)=\overline{d_vf(v)}$, and the Itô term, given by the second and third lines of (C.2), reeds

$$2Re\,\mathrm{Trace}\Big[(M^1(M^2)^t+M^2(M^1)^t)\frac{\partial^2 f}{\partial v\partial v}+(M^1(\bar M^1)^t+M^2(\bar M^2)^t)\frac{\partial^2 f}{\partial v\partial\bar v}\Big].$$

To prove the result one may re-write $v(t)$ as an Itô process in R^{2d} in terms of the real Wiener processes $\mathrm{Re}\,W_j(t)$, $\mathrm{Im}\,W_j(t)$, apply the usual Itô's formula to $f(t,v(t))$ and then re-write the result back in terms of the complex Wiener processes. Corresponding straightforward calculation is rather heavy and it is not easy to make it without mistake. Below we suggest a better way to get the formula.

Proof. The linear part of the formula (C.2), given by its first and forth lines, follows from the real case by linearity. It remains to prove that the Itô term has the form, given by the second and third lines. From the real formula we see that the Itô term is an expression, linear in $\partial^2 f/\partial v\partial v$, $\partial^2 f/\partial\bar v\partial\bar v$ and $\partial^2 f/\partial v\partial\bar v$, with the coefficients, quadratic in the matrices M^1 and M^2. So it may be written as

$$\Big[\,\mathrm{Trace}\big(Q^1\frac{\partial^2 f}{\partial v\partial v}\big)+\mathrm{Trace}\big(Q^2\frac{\partial^2 f}{\partial\bar v\partial\bar v}\big)+\mathrm{Trace}\big(Q^3\frac{\partial^2 f}{\partial v\partial\bar v}\big)\Big]dt,\tag{C.3}$$

where Q^j are complex $n\times n$-matrices, quadratic in M^1,M^2. We should show that they have the form, specified in (C.2). To do that we note that since the processes β_j^c are independent and have the form (2.3), then for all j,l their brackets have the following form:

$$\langle\beta_j^c,\beta_l^c\rangle=\langle\bar\beta_j^c,\bar\beta_l^c\rangle=0,\quad\langle\beta_j^c,\bar\beta_l^c\rangle=\langle\bar\beta_j^c,\beta_l^c\rangle=2\delta_{j,l}t.\tag{C.4}$$

Now let in (C.1) $g=0,v(0)=0$ and M^1,M^2 are constant matrices. Then

$$v(t)=M^1B(t)+M^2\bar B(t).$$

Taking $f(v) = v_{i_1} v_{i_2}$ and using (C.4) we see that

$$f(v(t)) = \Big(\sum_j M^1_{i_1 j} B_j(t) + \sum_j M^2_{i_1 j} \bar{B}_j(t)\Big) \cdot \Big(\sum_j M^1_{i_2 j} B_j(t) + \sum_j M^2_{i_2 j} \bar{B}_j(t)\Big)$$
$$= \Big[\big(M^1 (M^2)^t\big)_{i_1 i_2} + \big(M^2 (M^1)^t\big)_{i_1 i_2}\Big] 2\,t + \text{ a martingale.}$$

Since due to (C.3) the linear in t part should equal $(Q_{i_1 i_2} + Q_{i_2 i_1})t$, then $Q^1 = M^1 (M^2)^t + M^2 (M^1)^t$. Similar, considering $f(v) = \bar{v}_{i_1} \bar{v}_{i_2}$ we find that

$$Q^2 = \bar{M}^1 (\bar{M}^2)^t + \bar{M}^2 (\bar{M}^1)^t,$$

while letting $f(v) = v_{i_1} \bar{v}_{i_2}$ leads to $2\big(M^1 (\bar{M}^1)^t + M^2 (\bar{M}^2)^t\big)_{i_1 i_2} = Q^3_{i_1 i_2}$; so

$$Q^3 = 2(M^1 (\bar{M}^1)^t + M^2 (\bar{M}^2)^t).$$

This completes the proof of (C.2). The second assertion of the theorem follows by a straightforward calculation. □

Appendix D. Projections to convex sets

Lemma D.1. *Let $\mathcal{B}$ be a closed convex subset of a Hilbert space X of finite or infinite dimension. Let $\mathcal{B}$ contains at least two points and let $\Pi : X \to \mathcal{B}$ be the projection, sending any point of X to a nearest point of $\mathcal{B}$. Then Lip $\Pi = 1$.*

Proof. Let $A, B \in X$ and let $a = \Pi A \in \mathcal{B}, b = \Pi B \in \mathcal{B}$. If $A, B \in \mathcal{B}$ then $a = A$ and $B = b$. So $\operatorname{Lip} \Pi \geq 1$ and it remains to show that

$$\|a - b\| \leq \|A - B\| \qquad \forall\, A \neq B.$$

If $a = b$ the assertion is trivial. Otherwise consider the vectors $\xi = b - a$, $l^a = A - a$, $l^b = B - b$, and introduce in X an orthonormal basis $(e_1, e_2, \dots)$ such that $e_1 = \xi/|\xi|$. Then $\xi = (\xi_1, \xi_2, \dots)$, where $\xi_1 = |\xi|$ and $\xi_j = 0$ for $j \geq 2$. Since a is a point in $[a, b] \subset X$, the closest to A, then $l^a_1 = l^a \cdot e_1 \leq 0$. Similar $l^b_1 \geq 0$. Thus

$$\|B - A\| = \|\xi + l^b - l^a\| \geq |\xi_1 + l^b_1 - l^a_1| \geq \xi_1 = \|b - a\|,$$

and the assertion is proved. □

Note that an analogy of the lemma's statement for a Banach space X in general is wrong.

Acknowledgement

GH was supported by National Natural Science Foundation of China (No. 20221300605). Both authors were supported by the Ministry of Science and Higher Education of the Russian Federation (megagrant No. 075-15-2022-1115).

References

[1] V. I. Arnold, V. V. Kozlov, and A. I. Neishtadt. *Mathematical Aspects of Classical and Celestial Mechanics*. Springer, third edition, 2006.
[2] P. Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, New York, 1999.
[3] V.I. Bogachev, N. V. Krylov, M. Rockner, and S.V. Shaposhnikov. *Fokker-Plank-Kolmogorov equations*. AMS Publications, 2015.
[4] N. N. Bogoliubov and Y. A. Mitropolsky. *Asymptotic Methods in the Theory of Non-linear Oscillations*. Gordon and Breach, New York, 1961.
[5] A. Boritchev and S. B. Kuksin. *One-dimensional turbulence and the stochastic Burgers equation*. Mathematical Surveys and Monographs. AMS Publications, Providence, 2021.

[6] V. Burd. *The Averaging Method on Infinite Time-Interval and Some Problems of the Theory of Oscillations*. Yaroslavl University Press, Yaroslavl, 2013. (in Russian).
[7] J. Duan and W. Wang. *Effctive Dynamics of Stochastic Partial Differential Equations*. Elsevier, 2014.
[8] R. M. Dudley. *Real Analysis and Probability*. Cambridge studies in advanced mathematics. Cambridge University Press, 2002.
[9] A. Dymov. Nonequilibrium statistical mechanics of weakly stochastically perturbed system of oscillators. *Ann. Henri Poincare*, 17:1825–1882, 2016.
[10] M. Freidlin and A. Wentzell. *Random Perturbations of Dynamical Systems*. Springer-Verlag, New York, 2nd edition, 1998.
[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, S. B. Kuksin, and A. Maiocchi. Time-averaging for weakly nonlinear CGL equations with arbitrary potential. *Fields Inst. Comm.*, 75:323–349, 2015.
[13] W. Jian, S.B. Kuksin, and Y. Wu. The Bogolyubov-Krylov averaging. *Russ. Math. Surveys*, 75:427–444, 2018.
[14] I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, 2002.
[15] R. Khasminski. On stochastic processes defined by differential equations with a small parameter. *Th. Probab. Appl.*, 11:211–228, 1966.
[16] R. Khasminski. On the averaging principle for Ito stochastic differential equations. *Kybernetika*, 4:260–279, 1968. (in Russian).
[17] R. Khasminskii. *Stochastic Stability of Differential Equations*. Springer, second edition, 2012.
[18] Yu. Kifer. *Large Deviations and Adiabatic Transitions for Dynamical Systems and Markov Processes in Fully Coupled Averaging*, volume 944 of *Memoirs of the American Mathematical Society*. AMS Publications, Providence, 2009.
[19] S. B. Kuksin and A. Maiocchi. Resonant averaging for small-amplitude solutions of stochastic nonlinear Schrödinger equations. *Proc. A of the Royal Soc. Edinburgh*, 147:357–394, 2017.
[20] A. Kulik. *Ergodic Behaviour of Markov Processes*. De Gruyter, Berlin, 2018.
[21] S.-J. Liu and M. Krstic. *Stochastic Averaging and Stochastic Extremum Seeking*. Springer, 2012.
[22] J. Mattingly, A. Stuart, and D. Higham. Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stochastic processes and their applications*, 101(2):185–232, 2002.
[23] A. V. Skorokhod. *Asymptotic Methods in the Theory of Stochastic Differential Equations*, volume 78 of *Translations of Mathematical Monographs*. AMS Publications, Providence, 1989.
[24] D. Stroock and S. R. S. Varadhan. *Multidimensional Diffusion Processes*. Springer, 1979.
[25] A. Yu. Veretennikov. On the averaging principle for systems of stochastic differential equations. *Math. USSR Sbornik*, 69:271–284, 1991.
[26] C. Villani. *Optimal Transport*. Springer, 2009.
[27] 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,& Peoples' Friendship University of Russia (RUDN University), Moscow, Russia
E-mail address: huangguan@tsinghua.edu.cn

Université Paris-Diderot (Paris 7), UFR de Mathématiques - Batiment Sophie Germain, 5 rue Thomas Mann, 75205 Paris, & Peoples' Friendship University of Russia (RUDN University), Moscow, Russia
E-mail address: sergei.kuksin@imj-prg.fr