Available online at www.sciencedirect.com

ScienceDirect

J. Math. Pures Appl. 89 (2008) 400–428

JOURNAL DE MATHÉMATIQUES PURES ET APPLIQUÉES

www.elsevier.com/locate/matpur

# Khasminskii–Whitham averaging for randomly perturbed KdV equation

Sergei B. Kuksin [a,b,*], Andrey L. Piatnitski [c,d]

[a] C.M.L.S., École polytechnique, 91128 Palaiseau, France
[b] Department of Mathematics, Heriot-Watt University, Edinburgh EH14 4AS, UK
[c] Narvik University College, HiN, PB 385, 8505 Narvik, Norway
[d] Lebedev Physical Institute RAS, Leninski prospect 53, 119991 Moscow, Russia

Received 25 September 2007

Available online 16 January 2008

**Abstract**

We consider the damped-driven KdV equation:

$$\dot{u} - \nu u_{xx} + u_{xxx} - 6uu_x = \sqrt{\nu}\,\eta(t,x), \quad x \in S^1, \qquad \int u\,\mathrm{d}x \equiv \int \eta\,\mathrm{d}x \equiv 0,$$

where $0 < \nu \leqslant 1$ and the random process $\eta$ is smooth in $x$ and white in $t$. For any periodic function $u(x)$ let $I = (I_1, I_2, \ldots)$ be the vector, formed by the KdV integrals of motion, calculated for the potential $u(x)$. We prove that if $u(t,x)$ is a solution of the equation above, then for $0 \leqslant t \lesssim \nu^{-1}$ and $\nu \to 0$ the vector $I(t) = (I_1(u(t,\cdot)), I_2(u(t,\cdot)), \ldots)$ satisfies the (Whitham) averaged equation.


**Résumé**

Nous considérons l'équation Kdv avec amortissement :

$$\dot{u} - \nu u_{xx} + u_{xxx} - 6uu_x = \sqrt{\nu}\,\eta(t,x), \quad x \in S^1, \qquad \int u\,\mathrm{d}x \equiv \int \eta\,\mathrm{d}x \equiv 0,$$

où $0 < \nu \leqslant 1$ et le processus aléatoire $\eta$ est régulier en $x$ et blanc en $t$. Pour toute fonction périodique $u(x)$, soit $I = (I_1, I_2, \ldots)$ un vecteur, de composantes les intégrales KdV du mouvement correspondant au potentiel $u(x)$. Nous démontrons que si $u(t,x)$ est une solution de l'équation ci-dessus, alors pour $0 \leqslant t \lesssim \nu^{-1}$ et $\nu \to 0$, le vecteur $I(t) = (I_1(u(t,\cdot)), I_2(u(t,\cdot)), \ldots)$ vérifie l'équation moyenne de Whitham.


*Keywords:* KdV equation; Random process; Whitham averaged equation

* Corresponding author at: C.M.L.S., École polytechnique, 91128 Palaiseau, France.
*E-mail addresses:* kuksin@math.polytechnique.fr (S.B. Kuksin), andrey@sci.lebedev.ru (A.L. Piatnitski).

0021-7824/$ – see front matter 
doi:10.1016/j.matpur.2007.12.003

## 0. Introduction

It is well known since the pioneer works of Novikov and Lax that the KdV equation,

$$\dot{u} + u_{xxx} - 6uu_x = 0, \tag{0.1}$$

defines an integrable infinite-dimensional Hamiltonian system in a space $H^p$ of $2\pi$-periodic Sobolev functions of order $p \geqslant 0$ with zero meanvalue. It means that KdV has infinitely many integrals of motion $I_1, I_2, \dots$, which are non-negative analytic functions on $H^p$, and for any non-negative sequence $I = (I_1, I_2, \dots)$ the set $T_I = \{u\colon\ I_j(u) = I_j\ \forall j\}$ is an analytic torus in $H^p$ of dimension $|J(I)| \leqslant \infty$, where $J$ is the set $J = \{j\colon\ I_j > 0\}$. Each torus $T_I$ carries an analytic cyclic coordinate $\varphi = \{\varphi_j,\ j \in J(I)\}$, and in the coordinates $(I, \varphi)$ the KdV-dynamics takes the integrable form:

$$\dot{I} = 0, \qquad \dot{\varphi} = W(I). \tag{0.2}$$

The *frequency vector* $W$ analytically depends on $I$. See [17,11] and Section 2 below.

Importance of these remarkable features of KdV is jeopardised by the fact that KdV arises in physics only as an approximation for ‘real’ equations, and it is still unclear up to what extend the integrability property persists in the ‘real’ equations, or how it can be used to study them.

The persistence problem turned out to be difficult, and the progress in its study is slow. In particular, it was established that small Hamiltonian perturbations of KdV do not destroy majority of time-quasiperiodic solutions, corresponding to (0.2) with $|J(I)| < \infty$ (see [14,11]), but it is unknown how these perturbations affect the almost-periodic solutions ($|J(I)| = \infty$), and whether solutions of the perturbed equations are stable in the sense of Nekhoroshev.

Probably it is even more important to understand the behaviour of solutions for KdV, perturbed by non-Hamiltonian terms (e.g., to understand how small dissipation affects the equation). The first step here should be to study how a $\nu$-perturbation affects the dynamics (0.2) on time-intervals of order $\nu^{-1}$. For perturbations of finite-dimensional integrable systems this question is addressed by the classical averaging theory, originated by Laplace and Lagrange. During more than 200 years of its history this theory was much developed, and good understanding of the involved phenomena was achieved, e.g. see in [1]. In particular, it is known that for a perturbed finite-dimensional integrable system,

$$\dot{I} = \nu f(I, \varphi), \qquad \dot{\varphi} = W(I) + \nu g(I, \varphi), \quad \nu \ll 1, \tag{0.3}$$

where $I \in \mathbb{R}^n$, $\varphi \in \mathbb{T}^n$, on time-intervals of order $\nu^{-1}$ the action $I(t)$ may be well approximated by solutions of the *averaged equation*:

$$\dot{I} = \nu \langle f \rangle (I), \qquad \langle f \rangle (I) = \int_{\mathbb{T}^n} f(I, \varphi)\, \mathrm{d}\varphi, \tag{0.4}$$

provided that the initial data $(I(0), \varphi(0))$ are typical. This assertion is known as the *averaging principle*.

The behaviour of solutions of infinite-dimensional systems (0.3) on time-intervals of order $\gtrsim \nu^{-1}$ is poorly understood. Still applied mathematicians believe that the averaging principle holds, and use (0.4) to study solutions of (0.3) with $n = \infty$. In particular, if (0.3) is a perturbed KdV equation, written in the variables $(I, \varphi)$, then (0.4) is often called the *Whitham equation* (corresponding to the perturbed KdV). The approximation for $I(t)$ in (0.3) with $0 \leqslant t \leqslant \nu^{-1}$ by $I(t)$, satisfying (0.4), is called the *Whitham averaging principle* since in [19] the averaging is systematically used in similar situations. In so far the Whitham averaging for the perturbed KdV equation under periodic boundary conditions was not rigorously justified. Instead mathematicians, working in this field, either postulate the averaging principle and study the averaged equations (e.g., see [5] and [3]), or postulate that the solution regularly—in certain sense—depends on the small parameter and show that this assumption implies the Whitham principle, see [12].

The main goal of this paper is to justify the Whitham averaging for randomly perturbed equations.

Let us start with random perturbations of the integrable system (0.2) with $I \in \mathbb{R}^n$, $\varphi \in \mathbb{T}^n$, where $n < \infty$. Introducing the fast time $\tau = \nu t$ we write the perturbed system as the Ito equation:

$$\begin{aligned} \mathrm{d}I &= F\, \mathrm{d}\tau + \sigma\, \mathrm{d}\beta_\tau, \\ \mathrm{d}\varphi &= \big(\nu^{-1} W(I) + G\big)\, \mathrm{d}\tau + g\, \mathrm{d}\beta_\tau. \end{aligned} \tag{0.5}$$

Here $F$, $G$, $\sigma$ and $g$ depend on $(I,\varphi)$, $\beta_\tau$ is a vector-valued Brownian motion and $\sigma$, $g$ are matrices. It was claimed in [9][1] and proved in [7] that (under certain assumptions, where the main one is non-degeneracy of the diffusion $\sigma$ and of the frequency-map $W$) when $\nu\to 0$, the solution $I(\tau)$ converges in distribution to a solution of the averaged equation,

$$\mathrm{d}I = \langle F\rangle(I)\,\mathrm{d}\tau + \langle\sigma\rangle(I)\,\mathrm{d}\beta, \tag{0.6}$$

where $\langle F\rangle$ is defined as in (0.4) and the matrix $\langle\sigma\rangle(I)$ is a symmetric square root of the matrix $\int_{T^n}\sigma\sigma^t\,\mathrm{d}\varphi$.

Now let us consider a randomly perturbed ('damped-driven') KdV equation:

$$\dot u - \nu u_{xx} + u_{xxx} - 6uu_x = \sqrt{\nu}\,\eta(t,x). \tag{0.7}$$

As before, $x\in S^1$ and $\int u\,\mathrm{d}x \equiv \int \eta\,\mathrm{d}x \equiv 0$. The force $\eta$ is a Gaussian random field, white in time $t$:

$$\eta = \frac{\partial}{\partial t}\sum_{s\in\mathbb{Z}_0} b_s\beta_s(t)e_s(x),$$

where $\mathbb{Z}_0=\mathbb{Z}\setminus\{0\}$, $\beta_s(t)$ are standard independent Wiener processes, and $\{e_s,\ s\in\mathbb{Z}_0\}$ is the usual trigonometric basis:

$$e_s(x)=\begin{cases}\cos sx, & s>0,\\ \sin sx, & s<0.\end{cases} \tag{0.8}$$

Concerning the real constants $b_s$ we assume that

$$b_s\leqslant C_m|s|^{-m}\quad \forall m,s \tag{0.9}$$

with some constants $C_m$ (so $\eta(t,x)$ is smooth in $x$), and

$$b_s\neq 0\quad \forall s. \tag{0.10}$$

The factor $\sqrt{\nu}$ in front of the force $\eta(t,x)$ is natural since under this scaling solutions of (0.7) remains of order 1 as $t\to\infty$ and $\nu\to 0$. Eq. (0.7) defines a Markov process in the function space $H^p$. Due to (0.10) it has a unique stationary measure. Let $u^\nu(t,\ x)$, $t\geqslant 0$, be a corresponding stationary in time solution for (0.7); or let $u^\nu$ be a solution, satisfying,

$$u^\nu(0,x)=u_0(x), \tag{0.11}$$

where $u_0(x)$ is a non-random smooth function. In Section 1 we prove that all moments of all Sobolev norms $\|u^\nu(t,\cdot)\|_m$ are bounded uniformly in $\nu>0$ and $t\geqslant 0$. Let us write $u^\nu(\tau)$ as $(I^\nu(\tau),\varphi^\nu(\tau))$. These processes satisfy the infinite-dimensional equation (0.5), so by the just mentioned estimates the processes $\{I^\nu(\cdot),\ 0<\nu\leqslant 1\}$ form a tight family, and along suitable sequences $\nu_j\to 0$ we have a weak convergence in distribution,

$$I^{\nu_j}(\cdot)\to I^0(\cdot), \tag{0.12}$$

where, according to the type of the solutions $u^\nu(\tau)$, the limiting process $I^0(\tau)$ is either stationary in $\tau$, or satisfies $I^0(0)=I(u_0(\cdot))$.

The main results of this work are the following two theorems, proved in Section 6:

**Theorem A.** *The limiting process $I^0(\tau)$ satisfies the Whitham equation* (0.6)*, corresponding to the perturbed KdV equation* (0.7)*. It is non-degenerate in the sense that for any $\tau>0$ and each $k\geqslant 1$ we have $\mathbf{P}\{I^0_k(\tau)=0\}=0$.*

**Theorem B.** *If the processes $u^\nu(\tau)$ are stationary in $\tau$, then for any $\tau\geqslant 0$ the law of the pair $(I^{\nu_j}(\tau),\varphi^{\nu_j}(\tau))$ converges to the product measure $q^0\times \mathrm{d}\varphi$, where $q^0$ is the law of $I^0(0)$ and $\mathrm{d}\varphi$ is the Haar measure on $\mathbb{T}^\infty$.*

[1] The main theorem of [9] deals with the situation when the unperturbed system is a stochastic equation with a non-degenerate diffusion for $\varphi$, but in its last section it is claimed that the ideas of the proof also apply to (0.5).

The proof is based on the scheme, suggested by Khasminskii in [9], see also [6] and [18]. It uses the estimates from Section 1 and more sophisticated estimates, obtained in Sections 4 and 5. Namely, we use crucially Lemma 4.3 (Section 4) and Lemma 5.2 (Section 5). In the former coupling arguments are evoked to prove that for any $k$ probability of the event $\{I_k^\nu(t) < \delta\}$ goes to zero with $\delta$, uniformly in $\nu$ and $t$. This is important since (0.5) is an equation for $I$ in the octant $\{I \mid I_j > 0\ \forall j\}$ which degenerates at the boundary $\{I \mid I_j = 0 \text{ for some } j\}$. In the latter we examine the random process $W^m(\tau) = W^m(I^\nu(\tau))$, where $W^m$ is the vector, formed by the first $m$ components of the frequency vector $W$. Exploiting Krylov's results from [13] we estimate the density against the Lebesgue measure of the law of the averaged vector $s^{-1}\int_0^s W^m(I^\nu(\tau))\,d\tau$, $s \sim 1$. We use this estimate to show that with probability close to one the components of the vector $W^m(\tau)$ are non-commensurable, so the fast motion $(d/d\tau)\varphi^m = \nu^{-1}W^m(\tau)$ is ergodic on the torus $\mathbb{T}^m \subset \mathbb{T}^\infty$, for any $m$. This is a crucial step of the proof of Theorem A. Our proof of Lemma 5.2 is 'hard' in the sense that it uses heavily the analyticity of the frequency map $W(I)$.

The arguments above are applied to the perturbed KdV equation, written in the Birkhoff normal form (Eq. (2.1) in Section 2). They apply as well to perturbations of other Birkhoff-integrable equations if their solutions satisfy good a priori estimates uniformly in the small parameter, and the corresponding transformation to the Birkhoff coordinates is smooth and is polynomially bounded at infinity. In the KdV case which we consider, half of the required bounds on the transformation is established in the recent paper [10]. We are certain that the remaining half can be obtained similarly, but do not prove them in this work, see Theorem 2.3 in Section 2.

The Whitham equation (0.6), corresponding to the perturbed KdV (0.7), is a complicated infinite-dimensional stochastic differential equation. Theorem A implies that for any smooth initial data $I(0)$ it has a weak solution, but we do not know if this solution is unique. We point out that, firstly, if (0.6) has a unique solution and the process $u^\nu(\tau)$ satisfy (0.11), then the law of the limiting process $I^0$ is independent of the sequence $\{\nu_j\}$, and the convergence (0.12) holds for $\nu \to 0$. Secondly, if (0.6) has a unique stationary measure, then a similar assertion holds for stationary solutions $u^\nu(\tau)$.

**The inviscid limit.** Let us consider the stationary solutions of Eq. (0.7) in the original time $t$. The a priori estimates from Section 1 imply that this family is tight in $C([0,T]; H^p)$ for any $T > 0$ and any $p > 0$. Therefore, along sequences $\nu_j \to 0$, we have convergence in distribution,

$$u^{\nu_j}(\cdot) \to u^0(\cdot) \tag{0.13}$$

(the limiting process $u^0(t)$ a priori depends of the sequence $\{\nu_j\}$). The arguments, applied in Section 10 of [15] to the randomly perturbed Navier–Stokes equation (0.14) also apply to (0.7). They show that a.e. realisation of the limiting process $u^0(t,x)$ is a smooth solution of the KdV equation (0.1). In particular, the law $\mu^0$ of the random variable $u^0(0,\cdot) \in H^p$ is an invariant measure for the dynamical system which KdV defines in $H^p$. But KdV has infinitely many integrals of motion; so it has a lot of invariant measures. How to distinguish among them the measure $\mu^0$? Noting that $u^\nu(t)_{t=0} = u^\nu(\tau)_{\tau=0}$, we apply Theorem B to get that the isomorphism $u(\cdot) \mapsto (I,\varphi)$ transforms $\mu^0$ to the measure $q^0 \times d\varphi$. In particular, if (0.6) has a unique stationary measure, then the measure $\mu^0$ is uniquely defined, and the convergence (0.13) holds for $\nu \to 0$.

This discussion shows that in difference with the deterministic situation, averaged randomly perturbed equations describe not only behaviour of solutions for a pre-limiting equation on time-intervals of order $\nu^{-1}$, but also its asymptotic in time properties. Indeed, under the double limit 'first $t \to \infty$, next $\nu \to 0$', the distribution of any solution converges to a measure, simply expressed in terms of a stationary measure of the averaged equation.

**The Eulerian limit.** The perturbed KdV equation (0.7) is a reasonable model for the randomly perturbed 2D NSE:

$$\begin{gathered} \dot u - \nu\Delta u + (u\cdot\nabla)u + \nabla p = \sqrt{\nu}\,\eta(t,x), \quad x \in \mathbb{T}^2, \\ \operatorname{div} u = 0, \qquad \int u\,dx \equiv \int \eta\,dx \equiv 0, \end{gathered} \tag{0.14}$$

obtained by replacing in (0.14) the 2D Euler equation $(0.14)_{\nu=0}$ (which is a Hamiltonian PDE with infinitely many integrals of motion) by KdV. Under restrictions on the random force $\eta(t,x)$, similar to those imposed on the force in (0.7), Eq. (0.14) (interpreted as a Markov process in the space of divergence-free vector fields $u(x)$), has a unique stationary measure, see in [15]. Let $(u^\nu(t), p^\nu(t))$ be the corresponding stationary solution. Then, along sequences $\nu_j \to 0$, the convergence in distribution holds:

$$\big(u^{\nu_j}(\cdot), p^{\nu_j}(\cdot)\big) \to \big(u^0(\cdot), p^0(\cdot)\big), \tag{0.15}$$

where the limiting process $(u^0, p^0)$ is stationary in time, is sufficiently smooth in $t$ and $x$, and a.e. its realisation satisfies the free Euler equation $(0.14)_{\nu=0}$. Accordingly, the law $\mu^0$ of $u^0(0)$ is an invariant measure for the dynamical system, which the Euler equation defines in the space of divergence-free vector fields. To study the measure $\mu^0$ (in fact, the set of measures $\mu^0$, since it is possible that now the limit depends on the sequence $\{\nu_j\}$), is an important problem in (mathematical) 2D turbulence. The problem, addressed in this work, may be considered as its model.

**Agreements.** Analyticity of maps $B_1 \to B_2$ between Banach spaces $B_1$ and $B_2$, which are the real parts of complex spaces $B_1^c$ and $B_2^c$, is understood in the sense of Fréchet. All analytic maps which we consider possess the following additional property: for any $R$ a map analytically extends to a complex $(\delta_R > 0)$-neighbourhood of the ball $\{|u|_{B_1} < R\}$ in $B_1^c$. When two random variables are equal almost sure, we usually drop the specification "a.s.".

**Notations.** $\chi_A$ stands for the indicator function of a set $A$ (equal 1 in $A$ and equal 0 outside $A$). By $\varkappa(t)$ we denote various functions of $t$ such that $\varkappa(t) \to 0$ when $t \to \infty$, and by $\varkappa_\infty(t)$ denote functions $\varkappa(t)$ such that $\varkappa(t) = o(t^{-N})$ for each $N$. We write $\varkappa(t) = \varkappa(t; R)$ to indicate that $\varkappa(t)$ depends on a parameter $R$. For a measurable set $Q \subset \mathbb{R}^n$ we denote by $|Q|$ its Lebesgue measure.

## 1. The equation and its solutions

We denote by $H$ the Hilbert space,

$$H = \left\{u \in L_2\big(S^1\big)\colon \int u \, dx = 0\right\}$$

with the scalar product $\langle u, v\rangle = \frac{1}{\pi}\int_0^{2\pi} u(x)v(x)\,dx$. Then $\{e_s,\ s \in \mathbb{Z}_0\}$ (see (0.8)) is its Hilbert basis. We set $H^m$ to be the $m$th Sobolev space, formed by functions with zero mean-value, and given the norm $\|u\|_m = \langle \frac{\partial^m u}{\partial x^m}, \frac{\partial^m u}{\partial x^m}\rangle^{1/2}$.

We write the KdV equation as

$$\dot u + V(u) = 0, \quad V(u) = u_{xxx} - 6uu_x, \tag{1.1}$$

and rewrite Eq. (0.7) as

$$\dot u - \nu u_{xx} + V(u) = \sqrt{\nu}\,\eta(t,x). \tag{1.2}$$

It is well known that a dissipative nonlinear equation in one space-dimension with a white in time right-hand side has a unique strong solution if the equation's solutions satisfy sufficiently strong a priori estimates. In Appendix A we show that any smooth solution of (0.7) with a deterministic initial data,

$$u(0) = u_0, \tag{1.3}$$

where $u_0 \in H^m$, $m \geqslant 1$, satisfies the following estimates:

$$\mathbf{E}e^{\sigma\|u(t)\|_0^2} \leqslant \max\big(\mathbf{E}e^{\sigma\|u(0)\|_0^2}, 2e^{2\sigma B_0}\big), \tag{1.4}$$

$$\mathbf{E}\big\|u(t)\big\|_m^2 \leqslant \max\big(4\mathbf{E}\big\|u(0)\big\|_m^2, C_m'\big), \tag{1.5}$$

$$\mathbf{E}\big\|u(t)\big\|_m^k \leqslant C\big(\|u_0\|_{mk}, B_{m+1}, m, k\big). \tag{1.6}$$

Here $t \geqslant 0$, $k \in \mathbb{N}$ and $\sigma \leqslant (2\max b_s^2)^{-1}$.

Accordingly, we have the following result:

**Theorem 1.1.** *For any deterministic $u_0 \in H^m$, $m \geqslant 1$, the problem* (0.7), (1.3) *has a unique solution $u(t,x)$. It satisfies estimates* (1.4)–(1.6).

Due to assumption (0.10), Eq. (0.7) has a unique stationary measure $\mu_\nu$ and any solution converges to $\mu_\nu$ in distribution. For the randomly forced 2D NSE equation this result now is well known (e.g., see in [15]). The proofs for Eq. (0.7) are simpler and we do not discuss them.

Let $u_\nu^0(t,x)$ be a solution of (0.7), (1.3) with $u_0 = 0$. Since $\mathcal{D}(u_\nu^0(t)) \rightharpoonup \mu_\nu$, then Theorem 1.1 and the Fatou lemma imply:

**Theorem 1.2.** *The unique stationary measure $\mu_\nu$ satisfies the estimates*:

$$\int_H e^{\sigma \|u\|_0^2} \mu_\nu(\mathrm{d}u) \leqslant C_\sigma < \infty \quad \forall \sigma \leqslant \left(2 \max b_s^2\right)^{-1},$$

$$\int_H \|u\|_m^k \mu_\nu(\mathrm{d}u) \leqslant C_{m,k} < \infty \quad \forall m, k.$$

## 2. Preliminaries on the KdV equation

In this section we discuss integrability of the KdV equation (1.1).

For $r \geqslant 0$ let us denote by $h^r$ an abstract Hilbert space with the basis $\{f_j, j = \pm 1, \pm 2, \ldots\}$ and the norm $|\cdot|_r$, where

$$|v|_r^2 = \sum_{j \geqslant 1} j^{1+2r} \left(v_j^2 + v_{-j}^2\right) \quad \text{for } v = \sum_{j \in \mathbb{Z}_0} v_j f_j.$$

We denote $\mathbf{v}_j = \binom{v_j}{v_{-j}}$, and identify a vector $v = \sum v_j f_j \in h^r$ with the sequence $(\mathbf{v}_1, \mathbf{v}_2, \ldots)$.

**Theorem 2.1.** *(See [11].) There exist an analytic diffeomorphism $\Psi : H \to h^0$ and an analytic functional $K$ on $h^0$ of the form,*

$$K\left(\sum v_j f_j\right) = \tilde{K}(I_1, I_2, \ldots), \quad I_j = \frac{1}{2}\left(v_j^2 + v_{-j}^2\right),$$

*with the following properties*:

(1) *$\Psi$ defines, for any $m \in \mathbb{N}$, an analytic diffeomorphism $\Psi : H^m \to h^m$;*
(2) $\mathrm{d}\Psi(0)$ *is the map $H^m \ni \sum u_s e_s \mapsto \sum |s|^{-1/2} v_s f_s \in h^m$;*
(3) *a curve $u(t) \in C^1(0, T; H)$ is a solution of* (1.1) *if and only if $v(\tau) = \Psi(u(t))$ satisfies the equations*:

$$\dot{v}_j = -\operatorname{sign}(j)\, v_{-j} W_{|j|}(I_1, I_2, \ldots), \quad j \in \mathbb{Z}_0, \tag{2.1}$$

*where $W_l = \frac{\partial \tilde{K}}{\partial I_l}$ for $l = 1, 2, \ldots$.*

**Corollary 2.2.** *If $u(t)$ is a solution of* (1.1) *and $\Psi(u) = v = \sum v_s f_s$, then*

$$I_k(t) = \frac{1}{2}\left(v_k^2 + v_{-k}^2\right)(t) = \text{const} \quad \forall k = 1, 2, \ldots. \tag{2.2}$$

If $v \in h^r$, then the vector $I = (I_1, I_2, \ldots)$ belongs to the space

$$h_I^r = \left\{ I\colon |I|_{h_I^r} = 2 \sum j^{1+2r} |I_j| < \infty \right\}.$$

In fact, $I \in h_{I+}^r$, where

$$h_{I+}^r = \left\{ I \in h_I^r\colon I_j \geqslant 0 \; \forall j \right\}.$$

**Amplification.** The function $\tilde{K}$ in Theorem 2.1 is analytic in $h_{I+}^0$. That is, it analytically extends to the vicinity of this set in the space $h_I^0$.

The quantities $I_1, I_2, \dots$ are called *the actions*. Each vector $\mathbf{v}_j$ can be characterised by the action $I_j$ and *the angle*:

$$\varphi_j = \arctan \frac{v_{-j}}{v_j}.$$

We will write $v = (I, \varphi)$, where $\varphi = (\varphi_1, \varphi_2, \dots)$. The vector $\varphi = (\varphi_1, \varphi_2, \dots)$ belongs to the torus $\mathbb{T}^\infty$. We provide the latter with the Tikhonov topology, so it becomes a compact set.

The functions $u \to v_k(u)$, $k \in \mathbb{Z}_0$, form a coordinate system on $H$. They are called the *Birkhoff coordinates*, and the system of Eqs. (2.1)—the *Birkhoff normal form* for the KdV equation. The normal forms is a classical tool to study finite-dimensional Hamiltonian systems and their perturbations locally in the vicinity of an equilibrium (see [16, §30]). For all important finite-dimensional systems the normal forms do not exist globally. In contrast, Theorem 2.1 shows that the KdV equation is an infinite-dimensional Hamiltonian system which admits a normal form globally in the whole space $H$. To take all advantages of this normal form we will need some information about asymptotic properties of the transformation $\Psi(u)$ when $u \to \infty$:

**Theorem 2.3.** *For $m = 0, 1, \dots$ there are polynomials $P_m$ and $Q_m$ such that*

$$\left|\mathrm{d}^j \Psi(u)\right|_m \leqslant P_m\big(\|u\|_m\big), \quad j = 0, 1, 2,$$

*and*

$$\left\|\mathrm{d}^j \Psi^{-1}(v)\right\|_m \leqslant Q_m\big(|v|_m\big), \quad j = 0, 1,$$

*for all $u, v$ and all $m \geqslant 0$. Here for $j \geqslant 1$ $|\mathrm{d}^j \Psi|_m$ is the norm of the corresponding poly-linear map from $H^m$ to $h^m$, and similar with $\|\mathrm{d}^j \Psi^{-1}\|_m$.*

**Proof.** The estimates for the norms $|\Psi(u)|_m$ and $\|\Psi^{-1}(v)\|_m$ follows from Theorem 2.1 in [10].[2]

We do not prove here the estimate for $\mathrm{d}^j \Psi(u)$ with $j = 1, 2$. We are certain that modern spectral techniques (e.g., see [10,2]) allow to establish them, but we think that this paper is not a proper place for a corresponding rather technical research. □

**Remark.** We do not use the fact that the coordinate system $v = (\mathbf{v}_1, \mathbf{v}_2, \dots)$ is symplectic, but only that it puts the KdV equation to the form (2.1). Therefore we may replace $v$ by another smooth coordinate system $v' = (\mathbf{v}'_1, \mathbf{v}'_2, \dots)$ such that $I'_j = I_j$ for all $j$ and $\varphi'_j = \varphi_j + \Phi_j(I_1, I_2, \dots)$. Non-symplectic coordinate systems are easier to construct, and it is possible that a proof of Theorem 2.3 simplifies if we replace there $v$ by a suitable system $v'$.

For a function $f$ on a Hilbert space $H$ we write $f \in \mathrm{Lip}_{\mathrm{lock}}(H)$ if

$$\big|f(u_1) - f(u_2)\big| \leqslant P(R)\|u_1 - u_2\| \quad \text{if } \|u_1\|, \|u_2\| \leqslant R, \tag{2.3}$$

where $P$ is a continuous function (depending on $f$). Clearly the set of functions $\mathrm{Lip}_{\mathrm{lock}}(H)$ is an algebra. Due to the Cauchy inequality any analytic function on $H$ belongs to $\mathrm{Lip}_{\mathrm{lock}}(H)$ (see *Agreements*). In particular,

$$W_l \in \mathrm{Lip}_{\mathrm{lock}}\big(h^I_r\big) \quad \text{for } l \in \mathbb{N},\ r \geqslant 0. \tag{2.4}$$

## 3. Equation (0.7) in the Birkhoff coordinates

For $k = 1, 2, \dots$ we denote:

$$\Psi_k : H^m \to \mathbb{R}^2, \qquad \Psi_k(u) = \mathbf{v}_k,$$

where $\Psi(u) = v = (\mathbf{v}_1, \mathbf{v}_2, \dots)$. Let $u(t) = u^\nu(t)$ be a solution of (0.7), which either is a stationary solution, or satisfies (1.3) with a $\nu$-independent non-random $u_0$. Applying Ito's formula to the map $\Psi_k$ we get:

[2] Note that the quantity, denoted there $\|J\|_{p-1/2}$, equals $|v|_{p+1}$ up to a constant factor, and $Q_{2p}$ satisfies the estimates $Q_{2p} \leqslant R_{1p}(\|u\|_{p+1})$ and $\|u\|_{p+1} \leqslant R_{2p}(Q_{2p})$, where $R_{1p}$ and $R_{2p}$ are some polynomials.

$$\mathrm{d}\mathbf{v}_k=\bigg(\mathrm{d}\Psi_k(u)\big(\nu u_{xx}+V(u)\big)+\frac{1}{2}\nu\sum_{j\in\mathbb{Z}_0}b_j^2\,\mathrm{d}^2\Psi_k(u)[e_j,e_j]\bigg)\,\mathrm{d}t+\sqrt{\nu}\,\mathrm{d}\Psi_k(u)\bigg(\sum_{j\in\mathbb{Z}_0}b_je_j\,\mathrm{d}\beta^j\bigg). \tag{3.1}$$

Let us denote:

$$\mathrm{d}\Psi_k(u)\Big(\sum b_je_j\,\mathrm{d}\beta^j\Big)=B_k(u)\,\mathrm{d}\beta=\sum_j B_{kj}(u)\,\mathrm{d}\beta^j,\quad B_{kj}\in\mathbb{R}^2\ \forall k,j.$$

Then the diffusion term in (3.1) may be written as $\sqrt{\nu}\,B_k(u)\,\mathrm{d}\beta$.

Since $I_k=\frac{1}{2}|\Psi_k|^2$ is an integral of motion (see (2.2)), then application of Ito's formula to the functional $\frac{1}{2}|\mathbf{v}_k|^2=I_k$ and Eq. (3.1) results in

$$\mathrm{d}I_k=\nu\bigg(\big(\mathrm{d}\Psi_k(u)u_{xx},\mathbf{v}_k\big)+\frac{1}{2}\bigg(\sum_j b_j^2\,\mathrm{d}^2\Psi_k(u)[e_j,e_j],\mathbf{v}_k\bigg)+\frac{1}{2}\sum_j b_j^2\big|\mathrm{d}\Psi_k(u)e_j\big|^2\bigg)\,\mathrm{d}t+\sqrt{\nu}\,\big(B_k(u)\,\mathrm{d}\beta,\mathbf{v}_k\big) \tag{3.2}$$

(here and below $(\cdot,\cdot)$ indicates the scalar product in $\mathbb{R}^2$). Note that in difference with (3.1), Eq. (3.2) 'depends only on the slow time' in the sense that all terms in its right-hand side have a factor $\nu$ or $\sqrt{\nu}$.

Let us consider the infinite-dimensional Ito process with components (3.2), $k\geqslant 1$. The corresponding diffusion is $\sqrt{\nu}\sigma\,\mathrm{d}\beta$, where $\sigma=(\sigma_{kj}(u),\,k\in\mathbb{N},\,j\in\mathbb{Z}_0)$ and

$$\sigma_{kj}=\big(B_{kj}(u),\mathbf{v}_k\big)=b_j\big(\mathrm{d}\Psi_k(u)e_j,\Psi_k(u)\big).$$

Consider the *diffusion matrix* $a$,

$$a(u)=\sigma(u)\sigma^t(u),\qquad a_{k_1k_2}=\sum_{j\in\mathbb{Z}_0}\sigma_{k_1j}\sigma_{k_2j}. \tag{3.3}$$

**Lemma 3.1.** *For any $u\in H$ the sums in* (3.3) *converge. The matrix $a$ is symmetric and defines a bounded linear operator in $l^2$. If $a\xi=0$ for some $\xi\in l^2$, then $\xi_k\neq 0$ only if $\mathbf{v}_k=0$, where $v=\Psi(u)$. In particular, if $\mathbf{v}_k\neq 0\ \forall k$, then* $\operatorname{Ker} a=\{0\}$. *Moreover, if $|v_j|\geqslant\delta$ for $1\leqslant j\leqslant m$, then for any $\xi\in\mathbb{R}^m\times\{0\}\subset\mathbb{R}^\infty$ we have*:

$$\big\langle a(u)\xi,\xi\big\rangle_{l_2}=\big|\sigma^t(u)\xi\big|_{l_2}^2\geqslant C|\xi|_{l_2}^2, \tag{3.4}$$

*where $C$ depends on $\delta$, $m$, $|v|_1$ and the sequence $\{b_j\}$.*

**Proof.** Using (0.9) and Theorem 1.1 we get that $|\sigma_{kj}|\leqslant C|j|^{-1}\eta_k$, where $\eta\in l^2$. Therefore $\sigma$ defines a bounded linear operator $H\to l^2$ and $\sigma^t$ defines a bounded operator $l^2\to H$. So $a=\sigma\sigma^t$ is a bounded operator in $l^2$ and its matrix is well defined. Let us take any vector $\xi$. Then $(a\xi,\xi)_{l^2}=\langle\sigma^t\xi,\sigma^t\xi\rangle$, where

$$\big(\sigma^t\xi\big)_j=\sum_k b_j\big(\mathrm{d}\Psi_k(u)e_j,\mathbf{v}_k\big)\xi_k=b_j\Big\langle e_j,\mathrm{d}\Psi(u)^*\Big(\bigoplus\xi_k\mathbf{v}_k\Big)\Big\rangle. \tag{3.5}$$

Hence, $\xi\in\operatorname{Ker} a$ if and only if $\mathrm{d}\Psi(u)^*(\bigoplus\xi_k\mathbf{v}_k)=0$. Since $\mathrm{d}\Psi(u)$ is an isomorphism, then in this case $\xi_k\mathbf{v}_k=0$ for each $k$, and the assertion follows.

To prove (3.4) we abbreviate $\bigoplus\xi_k\mathbf{v}_k=\xi_v$ and denote $\mathrm{d}\Psi(u)^*\xi_v=\eta$. Then $\sigma^t(u)\xi=\operatorname{diag}\{b_j\}\eta$ (see (3.5)). Due to the first assertion of Theorem 2.3,

$$\|\eta\|_1^2\leqslant C_1\big(|v|_1\big)|\xi_v|_1^2\leqslant C_1\big(|v|_1\big)|\xi|_{l_2}^2C_2\big(|v|_0\big)m^3.$$

So

$$\sum_{k=N+1}^{\infty}\eta_k^2\leqslant N^{-2}C_1C_2|\xi|_{l_2}^2m^3,$$

for any $N$. Since $(\mathrm{d}\Psi(u)^*)^{-1}=(\mathrm{d}\Psi(u)^{-1})^*$, then the second assertion of the theorem implies that

$$\sum_{k=1}^{\infty}\eta_k^2=\|\eta\|_0^2\geqslant C_0\big(|v|_0\big)|\xi_v|_0^2\geqslant C_0\big(|v|_0\big)|\xi|_{l_2}^2\delta^2.$$

Choosing $N=[(2C_1C_2C_0^{-1}\delta^{-2}m^3)^{1/2}]+1$ we get that $\sum_{k=1}^N\eta_k^2\geqslant\frac{1}{2}C_0\delta^2|\xi|_{l_2}^2$. Accordingly,

$$\left|\sigma^t(u)\xi\right|_{l_2}^2\geqslant C'\sum_{k=1}^N\eta_k^2\geqslant\frac{1}{2}C'C_0\delta^2|\xi|_{l_2}^2,$$

where $C'$ depends on the sequence $\{b_j\}$ and $N$. □

We see that the infinite-dimensional Ito process $(3.2)_{k\in\mathbb{N}}$, defined for $I\in h_{I+}^0$, has non-degenerate diffusion outside the boundary $\partial h_{I+}^0=\{I\colon\ I_j=0 \text{ for some } j\geqslant 0\}$.

By applying Ito's formula to the $k$th angle $\varphi_k=\arctan(\frac{v_{-k}}{v_k})$ $(k\geqslant 1)$ and using (2.1) we obtain:

$$\begin{aligned}\mathrm{d}\varphi_k=\Bigg[&W_k(I)+\nu|\mathbf{v}_k|^{-2}\big(\mathrm{d}\Psi_k(u)u_{xx},\mathbf{v}_k^\perp\big)+\nu|\mathbf{v}_k|^{-2}\Bigg(\sum_{j=1}^\infty b_j^2\,\mathrm{d}^2\Psi_k[e_j,e_j],\mathbf{v}_k^\perp\Bigg)\\&-\nu|\mathbf{v}_k|^{-2}\sum_{j\in\mathbb{Z}_0}\big((B_{kj},\mathbf{v}_k)\big(B_{kj},\mathbf{v}_k^\perp\big)\big)\Bigg]\mathrm{d}t+\sqrt{\nu}|\mathbf{v}_k|^{-2}\big(B_k(u),\mathbf{v}_k^\perp\big)\,\mathrm{d}\beta,\end{aligned}$$

where $\mathbf{v}_k^\perp=\binom{-v_{-k}}{v_k}$. Denote for brevity the drift and diffusion coefficients in the above equation by $W_k(I)+\nu G_k(v)$ and $\sqrt{\nu}\,g_k^j(v)$, respectively. Denoting similarly the drift coefficients in (3.2) by $\nu F_k(v)$ we rewrite the equation for the pair $(I_k,\varphi_k)$ $(k\geqslant 1)$ as

$$\begin{aligned}\mathrm{d}I_k(t)&=\nu F_k(v)\,\mathrm{d}t+\sqrt{\nu}\,\sigma_k(v)\,\mathrm{d}\beta_t,\\\mathrm{d}\varphi_k(t)&=\big[W_k(I)+\nu G_k(v)\big]\,\mathrm{d}t+\sqrt{\nu}\,g_k(v)\,\mathrm{d}\beta_t.\end{aligned}\tag{3.6}$$

Introducing the fast time,

$$\tau=\nu t,$$

we rewrite the system (3.6) as

$$\begin{aligned}\mathrm{d}I_k(\tau)&=F_k(v)\,\mathrm{d}\tau+\sigma_k(v)\,\mathrm{d}\beta_\tau,\\\mathrm{d}\varphi_k(\tau)&=\left[\frac{1}{\nu}W_k(I)+G_k(v)\right]\mathrm{d}\tau+g_k(v)\,\mathrm{d}\beta_\tau.\end{aligned}\tag{3.7}$$

Here $\beta=(\beta_j,\ j\in\mathbb{Z}_0)$, where $\beta_j(\tau)$ are new standard independent Wiener processes.

In the lemma below $P_k$ and $P_{kN}$ are some polynomials.

**Lemma 3.2.** *For $k\in\mathbb{N}$, $j\in\mathbb{Z}_0$ we have*:

(i) *the function $F_k$ is analytic in each space $h^r$, $r\geqslant 2$ (so $F_k\in\mathrm{Lip}_{\mathrm{lock}}(h^r)$), and has a polynomial growth as $|v|_k\to\infty$;*
(ii) *the function $\sigma_{kj}(v)$ is analytic in $h^r$, $r\geqslant 0$, and for any $N\geqslant 1$ satisfies $|\sigma_{kj}(v)|\leqslant j^{-N}P_{kN}(|v|_r)$ $\forall v\in h^r$;*
(iii) *for any $r\geqslant 2$, $\delta>0$ and $N\geqslant 1$ the functions $G_k(v)\chi_{\{I_k>\delta\}}$ and $g_{kj}(v)\chi_{\{I_k>\delta\}}$ are bounded, respectively, by $\delta^{-1}P_k(|v|_r)$ and $\delta^{-1}j^{-N}P_{kN}(|v|_r)$.*

**Proof.** The assertions concerning the functions $F_k$ and $G_k$ follow from Theorem 2.3 since the set of analytical functions with polynomial growth at infinity is an algebra. To get the assertions about $\sigma_k$ and $g_k$ we also use (0.9). □

## 4. More estimates

In this section and in the following Sections 5–6 we consider solutions of Eq. (3.6), written in the form (3.7), which either are stationary in time, or satisfy the $\nu$-independent initial condition (1.3), where for simplicity $u_0$ is smooth and non-random,

$$u_0\in H^\infty=\bigcap_m H^m.$$

First we derive for these solutions additional estimates, uniform in $\nu$.

**Lemma 4.1.** *For any $\nu > 0$, $T > 0$ and $m, N \in \mathbb{N}$ the process $I(\tau)$ satisfies the estimate*:

$$\mathbf{E} \sup_{0 \leqslant \tau \leqslant T} \left|I(\tau)\right|_{h_I^m}^N = \mathbf{E} \sup_{0 \leqslant \tau \leqslant T} \left|v(\tau)\right|_m^{2N} \leqslant C(N, m, T). \tag{4.1}$$

**Proof.** For the sake of definiteness we consider a stationary solution $v(\tau) = \{v_k^\nu(\tau)\}$. Cauchy problem (3.7), (1.3) can be considered in the same way. Applying Ito's formula to the expression $k^m I_k^N$ gives:

$$\mathrm{d}\big(k^m I_k^N\big) = k^m \left( \left( N I_k^{N-1} F_k(v) + \frac{1}{2} N(N-1) I_k^{N-2} \sum_{j=1}^\infty \big(B_{kj}(v), \mathbf{v}_k\big)^2 \right) \mathrm{d}\tau + N I_k^{N-1} \sigma_k(v)\, \mathrm{d}\beta_\tau \right).$$

Therefore,

$$\begin{aligned} \mathbf{E} \sup_{0 \leqslant \tau \leqslant T} k^m I_k^N(\tau) &\leqslant \mathbf{E} k^m I_k^N(0) + k^m \mathbf{E} \sup_{0 \leqslant \tau \leqslant T} \left| \int_0^\tau \left( N I_k^{N-1}(s) F_k(v) + \frac{1}{2} N(N-1) I_k^{N-2}(s) \sum_{j=1}^\infty \sigma_{kj}^2 \right) \mathrm{d}s \right| \\ &\quad + k^m \mathbf{E} \sup_{0 \leqslant \tau \leqslant T} \left| \int_0^\tau N I_k^{N-1}(s) \sigma_k(v)\, \mathrm{d}\beta_s \right| \leqslant C(m, N, T). \end{aligned}$$

Doob's inequality, Lemma 3.2 and Theorem 1.2 have been used here. This relation yields the desired estimate. Indeed, by the Hölder inequality we get:

$$\begin{aligned} \mathbf{E}\Big( \sup_{0 \leqslant \tau \leqslant T} \left|I(\tau)\right|_m^{2N} \Big) &= 2^N \mathbf{E} \sup_{0 \leqslant \tau \leqslant T} \left( \sum_{j=1}^\infty \frac{1}{j^2} j^{2m+3} I_j(\tau) \right)^N \\ &\leqslant 2^N \mathbf{E} \sup_{0 \leqslant \tau \leqslant T} \left\{ \left( \sum_{j=1}^\infty j^{N(2m+3)} I_j^N(\tau) \right)^{N \frac{1}{N}} \left( \sum_{j=1}^\infty j^{-\frac{2N}{N-1}} \right)^{N \frac{N-1}{N}} \right\} \\ &\leqslant C_N \mathbf{E} \sup_{0 \leqslant \tau \leqslant T} \left( \sum_{j=1}^\infty j^{N(2m+3)} I_j^N(\tau) \right) \leqslant C_1(m, N, T). \qquad \square \end{aligned}$$

In the further analysis we systematically use the fact that the functionals $F_k(I, \varphi)$ depend weakly on the tails of vectors $\varphi = (\varphi_1, \varphi_2, \ldots)$. Now we state the corresponding auxiliary results.

Let $f \in \mathrm{Lip}_{\mathrm{lock}}(h^{n_1})$ and $v \in h^n$, $n > n_1$. Denoting by $\Pi_M$, $M \geqslant 1$, the projection,

$$\Pi_M : h^0 \to h^0, \qquad \sum v_j f_j \mapsto \sum_{|j| \leqslant M} v_j f_j,$$

we have $|v - \Pi_M v|_{n_1} \leqslant M^{-(n-n_1)} |u|_n$. Accordingly,

$$\big|f(v) - f\big(\Pi_M(v)\big)\big| \leqslant P\big(|v|_n\big) M^{-(n-n_1)}. \tag{4.2}$$

Similar inequalities hold for functions on $h_I^n$, and (2.4) with $r = 0$ implies that

$$\big|W_k(I) - W_k(\Pi_M I)\big| \leqslant P_k\big(|I|_n\big) M^{-n}. \tag{4.3}$$

The torus $\mathbb{T}^M$ acts on the space $\Pi_M h^0$ by linear transformations $\Phi_{\theta_M}$, $\theta_M \in \mathbb{T}^M$, where $\Phi_{\theta_M}$ sends a point $v_M = (I_M, \varphi_M)$ to $(I_M, \varphi_M + \theta_M)$. Similar, the torus $\mathbb{T}^\infty$ acts on $h^0$ by linear transformations $\Phi_\theta : (I, \varphi) \mapsto (I, \varphi + \theta)$. The transformation $\Phi_\theta$ continuously depends on $\theta \in \mathbb{T}^\infty$, in the strong operator topology.

For a function $f \in \mathrm{Lip}_{\mathrm{lock}}(h^{n_1})$ and any $N$ we define the average of $f$ in the first $N$ angles as the function,

$$\langle f \rangle_N(v) = \int_{\mathbb{T}^N} f(\Phi_{\theta_N} \oplus \mathrm{id})(v)\, \mathrm{d}\theta_N$$

(here id stands for the identity transformation in the space $h^0 \ominus \Pi_N h^0$), and define the average in all angles as

$$\langle f\rangle(v) = \int\limits_{\mathbb{T}^\infty} f(\Phi_\theta v)\,\mathrm{d}\theta,$$

where $\mathrm{d}\theta$ is the Haar measure on $\mathbb{T}^\infty$. The estimate (4.2) readily implies that

$$\bigl|\langle f\rangle_N(v) - \langle f\rangle(v)\bigr| \leqslant P(R)N^{-(n-n_1)} \quad \text{if } |v|_n \leqslant R. \tag{4.4}$$

Let $v = (I, \varphi)$. Then $\langle f\rangle_N$ is a function, independent of $\varphi_1, \dots, \varphi_N$, and $\langle f\rangle$ is independent of $\varphi$. I.e., $\langle f\rangle$ can be written as a function $\langle f\rangle(I)$.

**Lemma 4.2.** *Let $f \in \mathrm{Lip}_{\mathrm{lock}}(h^{n_1})$. Then*

(i) *The functions $\langle f\rangle_N(v)$ and $\langle f\rangle(v)$ satisfy* (2.3) *with the same polynomial as $f$ and take the same value at the origin.*
(ii) *They are smooth* (*analytic*) *if $f$ is. Moreover, if $f$ is smooth, then $\langle f\rangle(I)$ is a smooth functions of the vector $(I_1, \dots, I_M)$ for any $M$. If $f(v)$ is analytic in the space $h^{n_1}$, then $\langle f\rangle(I)$ is analytic in the space $h_I^{n_1}$.*

**Proof.** (i) Is obvious.

(ii) The first assertion is obvious. To prove the last two consider the function $g(r_1, r_2, \dots) = \langle f\rangle(\mathbf{v}_1, \mathbf{v}_2, \dots)$, $\mathbf{v}_j = \binom{r_j}{0}$. Then $g(r) = \langle f\rangle(I)$, where $I_l = \frac{1}{2}r_l^2$ for each $l$. The function $g$ is smooth and even in each $r_j$, $j \geqslant 1$. Any function of finitely many arguments with this property is known to be a smooth function of the squared arguments, so the second assertion holds.

Now let $f(v)$ be analytic. Denote by $\mathfrak{h}^{n_1}$ the space of all sequences $r = (r_1, r_2, \dots)$ such that the corresponding vector $v$ belongs to $h^{n_1}$, and provide it with the natural norm. If $f(v)$ is analytic, then $\langle f\rangle(v)$ also is analytic and $g(r)$ extends analytically to an even function in a complex neighbourhood $\mathcal{O}$ of $\mathfrak{h}^{n_1}$ in $\mathfrak{h}^{n_1} \otimes \mathbb{C}$. This neighbourhood may be chosen to be invariant with respect to all involutions:

$$(r_1, r_2, \dots, r_j, \dots) \mapsto (r_1, r_2, \dots, -r_j, \dots), \quad j = 1, 2, \dots.$$

The image $\mathcal{O}_I$ of $\mathcal{O}$ under the map,

$$(r_1, r_2, \dots) \mapsto \left(\frac{1}{2}r_1^2, \frac{1}{2}r_2^2, \dots\right),$$

is a neighbourhood of $h_I^{n_1}$ in the complex space $h_I^{n_1} \otimes \mathbb{C}$. The function,

$$g(\pm\sqrt{2I_1}, \pm\sqrt{2I_2}, \dots) =: g(\sqrt{I}),$$

is a well-defined locally bounded function on $\mathcal{O}_I$.[3] For any $N$ its restriction to $\mathcal{O}_I^N = \mathcal{O}_I \cap \Pi_N(h_I^{n_1} \otimes C)$ is a single-valued algebraic function on a domain in $\mathbb{C}^N$; so $g(\sqrt{I})$ is analytic on $\mathcal{O}_I^N$ for each $N$. Hence, $g(\sqrt{I})$ is analytic on $\mathcal{O}_I$ (see Lemma A.4 in [11]). Since $g(\sqrt{I}) = \langle f\rangle(I)$, then the result follows. □

Let $(I^\nu(\tau), \varphi^\nu(\tau))$ be a solution of (3.7). In the lemma below we show that the processes $I_k^\nu(\tau)$, $k \geqslant 1$, do not asymptotically approach zero as $\nu \to 0$ (concerning the notation $\varkappa(\delta^{-1}; M, T)$, used there, see *Notations*):

**Lemma 4.3.** *For any $M \in \mathbb{N}$ and $T > 0$ we have,*

$$\mathbf{P}\Bigl\{\min_{k \leqslant M} I_k^\nu(\tau) < \delta\Bigr\} \leqslant \varkappa\bigl(\delta^{-1}; M, T\bigr), \tag{4.5}$$

*uniformly in $\nu > 0$ and $0 \leqslant \tau \leqslant T$.*

[3] I.e., it is bounded uniformly on bounded subsets of $\mathcal{O}_I$.

Here the difficulty is that the scalar process $I_k(\tau) = \frac{1}{2}|\mathbf{v}_k(\tau)|^2$ satisfies Eq. (3.7), where the diffusion $\sigma_k$ degenerates when $I_k$ vanishes. The equation for the vector-process $\mathbf{v}_k(\tau)$ (see (7.1) below) has a non-degenerate diffusion, but its drift has a component of order $\nu^{-1}$. To prove the lemma's assertion we construct a new process $\hat{\mathbf{v}}_k(\tau)$ such that $|\mathbf{v}_k(\tau)| = |\hat{\mathbf{v}}_k(\tau)|$ and $\hat{\mathbf{v}}_k$ satisfies an Ito equation with a non-degenerate diffusion and coefficients, bounded uniformly in $\nu$. Then $I_k = \frac{1}{2}|\hat{\mathbf{v}}_k(\tau)|^2$ meets estimate (4.5) by a Krylov's theorem. The problem to perform this scheme is that the process $\hat{\mathbf{v}}_k$ is constructed as a solution of an additional diffusion equation which is ill defined when $v_k$ vanishes. We cannot show that the event,

$$\left\{v_k(\tau) = 0 \text{ for some } 0 \leqslant \tau \leqslant T\right\},$$

has zero probability and resolve this new difficulty by means of some additional (rather involved) construction.

For a complete proof see Section 7.

## 5. Averaging along Kronecker flows

The flow,

$$S^t : \mathbb{T}^\infty \to \mathbb{T}^\infty, \qquad \varphi \mapsto \varphi + tW, \quad t \in \mathbb{R},$$

where $W \in \mathbb{R}^\infty$, is called a *Kronecker flow*. In this section we study averages of functions $f(v) = f(I, \varphi)$ along such flows. That is, we study the quantities:

$$\frac{1}{T}\int_0^T f\big(I, \varphi + W^m t\big)\,\mathrm{d}t, \quad T > 0.$$

**Lemma 5.1.** *Let $f \in \mathrm{Lip}_{\mathrm{lock}}(h^{n_1})$, $v = (I, \varphi) \in h^n$, $n > n_1 \geqslant 0$, and $f$ is analytic in the space $h^{n_1}$. Then for each $R' > 0$, $m \in \mathbb{N}$ and $\delta > 0$ there is a Borel set $\Omega^m_{R'}(\delta) \subset \{x \in \mathbb{R}^m\colon |x| \leqslant R'\}$ such that $|\Omega^m_{R'}(\delta)| < \delta$, and for any $W^m \notin \Omega^m_{R'}(\delta)$, $|W^m| \leqslant R'$ the estimate*:

$$\left|\frac{1}{T}\int_0^T f\big(I, \varphi + W^m t\big)\,\mathrm{d}t - \langle f\rangle(v)\right| \leqslant \frac{1}{T\delta}c_0\big(m, R', |v|_n, f\big) + m^{-(n-n_1)}P\big(|v|_n\big),$$

*holds uniformly in $\varphi \in \mathbb{T}^\infty$. Here $P$ is the continuous function from* (2.3) *and $W^m$ is identified with the vector $(W^m, 0, \ldots) \in \mathbb{R}^\infty$.*

**Proof.** Let us first assume that $f(v) = f(\Pi_m v)$ (i.e., $v$ depends only on finitely-many variables). Then $f = f(I^m, \varphi^m)$ is analytic in $\varphi^m$ and the radius of analyticity is independent of $I$, satisfying $|I|_{h^0_I} \leqslant R'$. Now the estimate with $P := 0$ is a classical result (e.g., see in [16]). In general case we write $f$ as $f \circ \Pi_m + (f - f \circ \Pi_m)$ and use (4.4). $\square$

We will apply this lemma with $W^m = W^m(I)$, where $I = I(\tau)$ is the $I$-component of a solution of (3.6). To do this we have to estimate probabilities of the events $\{W^m(I(\tau)) \in \Omega^m_{R'}(\delta)\}$. To state the corresponding result we introduce more notations. For any events $Q$ and $\mathcal{O}$ we denote:

$$\mathbf{P}_Q(\mathcal{O}) = \mathbf{P}(\complement Q \cap \mathcal{O}),$$

and

$$\mathbf{E}_Q(f) = \mathbf{E}\big((1 - \chi_Q) f\big).$$

Abusing language, we call $\mathbf{P}_Q$ a probability. We fix any,

$$p \geqslant 1,$$

denote

$$\mathcal{B}_R = \big\{I\colon |I|_{h^p_I} \leqslant R\big\},$$

and for $R \geqslant 1$ consider the event,

$$\Omega_R = \Big\{ \sup_{0\leqslant\tau\leqslant T} \big|v^\nu(\tau)\big|_p \geqslant R \Big\},$$

where $v^\nu(\tau)$ is a solution. Noting that $|W^m(I)| \leqslant R' = R'(R,m)$ outside the event $\Omega_R$, we denote:

$$\Omega(\delta) = \Omega^m_{R'}(\delta), \quad R' = R'(R),\ 0 < \delta < 1.$$

Finally, for $M \geqslant m$ and $0 < \gamma < 1$ we define:

$$Q_\gamma = \Big\{ I \in h^0_{I+}\colon \min_{1\leqslant j\leqslant M} I_j < \gamma \Big\}.$$

**Lemma 5.2.** *There exists $M = M(R,m) \geqslant m$ such that*

$$\int_0^T \mathbf{P}_{\Omega_R}\big(\big\{W^m\big(I(s)\big) \in \Omega(\delta)\big\} \setminus \big\{I(s) \in Q_\gamma\big\}\big)\,\mathrm{d}s \leqslant \varkappa\big(\delta^{-1}; R, m, \gamma, T\big), \tag{5.1}$$

*uniformly in $\nu > 0$.*[4]

**Proof.** Consider the function $D(I) = \det(\partial W^m_j/\partial I_r\colon 1 \leqslant j, r \leqslant m)$. It is analytic in $h^0_I$ (see Amplification to Theorem 2.1), and $D \not\equiv 0$ since $D(0) = C^m$, $C \neq 0$ (see [14, Lemma 3.3], and [11]). For a finite non-decreasing sequence of natural numbers $\alpha = (\alpha^1 \leqslant \cdots \leqslant \alpha^N)$ we denote:

$$|\alpha| = \alpha^N, \quad [\alpha] = N,$$

and define the derivative $\partial^\alpha D(I)/\partial I^\alpha$ in the natural way.

**Step 1.** Study of the sets $\{I \in \mathcal{B}_R\colon |D(I)| < \varepsilon\}$, $0 < \varepsilon \ll 1$.

By the analyticity any point $I' \in \mathcal{B}_R$ has a neighbourhood $\mathcal{O} \subset h^0_I$ such that

$$\left| \frac{\partial^\alpha D(I)}{\partial I^\alpha} \right| \geqslant c \quad \forall I \in \mathcal{O},$$

where the sequence $\alpha = (\alpha^1 \leqslant \cdots \leqslant \alpha^N)$ and $c > 0$ depend only on the neighbourhood. Since $\mathcal{B}_R$ is a compact subset of $h^0_I$, we can cover it by a finite system of neighbourhoods $\mathcal{O}_j$, $j = 1, \dots, L$, as above, where $L = L(R,m)$. Then

$$\left\{ I \in \mathcal{B}_R\colon \left| \frac{\partial^{\alpha_j} D(I)}{\partial I^\alpha_j} \right| \ll 1,\ j = 1, \dots, L \right\} = \emptyset. \tag{5.2}$$

Let us denote,

$$M = \max_{1\leqslant j\leqslant L} |\alpha_j|, \qquad N = \max_{1\leqslant j\leqslant L} [\alpha_j],$$

and consider the sequence

$$\varepsilon = \varepsilon_0 < \varepsilon_1 < \cdots < \varepsilon_N < 1, \qquad \varepsilon_j = \varepsilon^{2^{-j} - 2^{-N} + 2^{-j-N}},$$

where $0 < \varepsilon < 1$. Note that

$$\varepsilon_j\, \varepsilon^{-2}_{j+1} = \varepsilon^{(2^{-N})} \quad \text{for } 0 \leqslant j < N.$$

For $m \leqslant [\alpha_j]$ we set:

$$\mathfrak{A}^m_{\alpha_j} = \left\{ I \in \mathcal{B}_R\colon \left| \frac{\partial}{\partial \alpha^1_j} \cdots \frac{\partial}{\partial \alpha^m_j} D(I) \right| < \varepsilon_m \right\}.$$

In particular, $\mathfrak{A}^0_{\alpha_j} = \mathfrak{A}^0 = \{I \in \mathcal{B}_R\colon |D(I)| \leqslant \varepsilon\}$ for each $j$.

[4] We recall that $\varkappa(t; R, m, \gamma, T)$ stands for a function of $t$ which goes to zero when $t \to \infty$, and depends on the parameters $R, m, \gamma$ and $T$.

For $0 < \varepsilon \ll 1$ relation (5.2) implies that

$$\mathfrak{A}^0 = \bigcup_{j=1}^{L} \left( \left(\mathfrak{A}^0 \setminus \mathfrak{A}^1_{\alpha_j}\right) \cup \left(\mathfrak{A}^1_{\alpha_j} \setminus \mathfrak{A}^2_{\alpha_j}\right) \cup \cdots \cup \left(\mathfrak{A}^{[\alpha_j]-1}_{\alpha_j} \setminus \mathfrak{A}^{[\alpha_j]}_{\alpha_j}\right)\right).$$

**Step 2.** An estimate for the integral $\int_0^T \mathbf{P}_{\Omega_R}\{|D(I(s))| < \varepsilon\}\,\mathrm{d}s$.

Due to the last displayed formula, the integral to be estimated is bounded by a finite sum of the terms,

$$\int_0^T \mathbf{P}_{\Omega_R}\left\{I(s) \in \mathfrak{A}^r_{\alpha_j} \setminus \mathfrak{A}^{r+1}_{\alpha_j}\right\}\mathrm{d}s, \quad r < [\alpha_j]. \tag{5.3}$$

To estimate (5.3), we abbreviate $\frac{\partial}{\partial\alpha_j^1} \cdots \frac{\partial}{\partial\alpha_j^r} D(I) = f(I)$. Then

$$\mathfrak{A}^r_{\alpha_j} \setminus \mathfrak{A}^{r+1}_{\alpha_j} = \left\{ I \in \mathcal{B}_R \colon \left|f(I)\right| < \varepsilon_r \text{ and } \left|\frac{\partial}{\partial(\alpha_j^{r+1})} f(I)\right| \geqslant \varepsilon_{r+1}\right\}. \tag{5.4}$$

Consider the Ito process $z(\tau) = f(I(\tau))$. We define the Markov moment $\tau' = \min\{\tau \geqslant 0\colon |I(\tau)|_{h^p_I} \geqslant R^2\} \wedge T$, and re-define $z(\tau)$ for $\tau \geqslant \tau'$ as a continuous process, satisfying:

$$\mathrm{d}z(\tau) = \mathrm{d}\beta^1_\tau \quad \text{for } \tau \geqslant \tau'.$$

Since $\tau' > T$ outside $\Omega_R$, then outside $\Omega_R$ we have $z(\tau) = f(I(\tau))$ for $0 \leqslant \tau \leqslant T$. For $z(\tau)$ we have:

$$\mathrm{d}z(\tau) = c(\tau)\,\mathrm{d}\tau + \sum b_j(\tau)\,\mathrm{d}\beta^j_\tau,$$

where $|c| \leqslant C(R,m)$, $b_j = \delta_{j,1}$ for $\tau \geqslant \tau'$ and $b_j = \sum \frac{\partial f}{\partial I_k}\sigma_{kj}$ for $\tau \leqslant \tau'$. Denoting $a = \sum b_j^2$, we have $a = \sum(\sigma\sigma^t)_{jk}\nabla_j f \nabla_k f$. So $|a(\tau)| \leqslant C(R,m)$. From other hand, (3.4) in Lemma 3.1 implies that

$$\left|a(\tau)\right| \geqslant C(R,m,\gamma)\sum_{j=1}^{M}(\nabla_j f)^2 \quad \text{if } I(\tau) \notin Q_\gamma. \tag{5.5}$$

Applying Theorem 2.3.3 from [13] to the process $z(\tau)$, we get:

$$\mathbf{E}\int_0^T \chi_{\{|z(\tau)| \leqslant \varepsilon_r\}}\left|a(\tau)\right|\mathrm{d}\tau \leqslant C(R,m,T)\varepsilon_r.$$

By (5.4) and (5.5) the integrand is $\geqslant \varepsilon^2_{r+1}C(R,m,\gamma)$ if $I(t) \in (\mathfrak{A}^r_{\alpha_j} \setminus \mathfrak{A}^{r+1}_{\alpha_j}) \setminus Q_\gamma$. Hence,

$$\begin{aligned}\int_0^T \mathbf{P}_{\Omega_R}\left\{I(s) \in \left(\mathfrak{A}^r_{\alpha_j} \setminus \mathfrak{A}^{r+1}_{\alpha_j}\right) \setminus Q_\gamma\right\}\mathrm{d}s &\leqslant \varepsilon_r\varepsilon^{-2}_{r+1}C(R,m,\gamma,T) \\ &= \varepsilon^{(2^{-N})}C(R,m,\gamma,T).\end{aligned}$$

We have seen that

$$\int_0^T \mathbf{P}_{\Omega_R}\left(\left\{\left|D\left(I(s)\right)\right| < \varepsilon\right\} \setminus \left\{I(s) \in Q_\gamma\right\}\right)\mathrm{d}s \leqslant \varepsilon^{(2^{-N})}C_1(R,m,\gamma,T). \tag{5.6}$$

**Step 3.** Proof of (5.1).

We have an inclusion of events:

$$\begin{aligned}
&\big\{W^m(s)\in\Omega(\delta)\big\}\setminus\big\{I(s)\in Q_\gamma\big\}\\
&\quad\subset\big[\big(\big\{W^m(s)\in\Omega(\delta)\big\}\setminus\big(\big\{I(s)\in Q_\gamma\big\}\cup\big\{D\big(I(s)\big)<\varepsilon\big\}\big)\big)\\
&\qquad\cup\big(\big\{D\big(I(s)\big)<\varepsilon\big\}\setminus\big\{I(s)\in Q_\gamma\big\}\big)\big].
\end{aligned}$$

Probability of the second event in the right-hand side is already estimated. To estimate probability of the first event we apply the Krylov estimate to the process $W^m(s)$. Re-defining it after the moment $\tau'$ (see Step 2) and arguing as when deriving (5.6), we get that

$$\begin{aligned}
&\int_0^T \mathbf{P}_{\Omega_R}\big(\big\{W^m(s)\in\Omega(\delta)\big\}\setminus\big(\big\{I(s)\in Q_\gamma\big\}\cup\big\{D\big(I(s)\big)<\varepsilon\big\}\big)\big)\,\mathrm{d}s\\
&\quad\leqslant\big|\Omega(\delta)\big|^{1/m}C(R,m,\gamma,\varepsilon,T).
\end{aligned}\tag{5.7}$$

Finally, choosing first $\varepsilon$ so small that the right-hand side of (5.6) is $\leqslant\tilde\varepsilon$ and next choosing $\delta$ so small that the right-hand side of (5.7) is $\leqslant\tilde\varepsilon$, we see that the left-hand side of (5.1) is $\leqslant 2\tilde\varepsilon$ for any $\tilde\varepsilon>0$, if $\delta$ is sufficiently small. □

## 6. The limiting dynamics

Let us fix any $T>0$, an integer $p\geqslant 3$ and abbreviate:

$$h^p=h,\quad h_I^p=h_I,\quad h_{I+}^p=h_{I+},\quad |I|_{h_I^p}=|I|,\quad |v|_p=|v|.$$

Due to Lemma 4.1 and the equation, satisfied by $I^\nu(\tau)$, the laws $\mathcal{L}\{I^\nu(\cdot)\}$ form a tight family of Borel measures on the space $C([0,T];h_{I+})$. Let us denote by $\mathcal{Q}^0$ any its weak limiting point:

$$\mathcal{Q}^0=\lim_{\nu_j\to 0}\mathcal{L}\big\{I^{\nu_j}(\cdot)\big\}.\tag{6.1}$$

Our aim is to show that $\mathcal{Q}^0$ is a solution to the martingale problem in the space $h_I$ with the drift operator $\langle F\rangle(I)=(\langle F_1\rangle(I),\langle F_2\rangle(I),\dots)$ and the covariance $\langle A\rangle(I)=\{\langle A_{kl}\rangle(I)\}$, where

$$\langle A_{kl}\rangle(I)=\big\langle\big(\sigma(v)\sigma^t(v)\big)_{kl}\big\rangle=\Big\langle\sum_j b_j^2\big(\mathrm{d}\Psi_k(u)e_j,\mathbf{v}_k\big)\big(\mathrm{d}\Psi_l(u)e_j,\mathbf{v}_l\big)\Big\rangle.$$

By Lemmas 3.2 and 4.2 the averages $\langle F_j\rangle$ and $\langle A_{kl}\rangle$ are analytic functions on $h_I$. The covariance $\langle A\rangle$ is non-degenerate outside the boundary of the domain $h_{I+}^p$ in the following sense: let $\xi\in\mathbb{R}^M\subset\mathbb{R}^\infty$ and $I\in h_{I+}$, $|I|\leqslant R$. Then

$$\sum_{k,l\leqslant M}\langle A_{kl}\rangle(I)\xi_k\xi_l\geqslant C|\xi|_{l_2}^2\quad\text{if } |I_j|\geqslant\gamma>0 \text{ for } j\leqslant M,\tag{6.2}$$

where $C>0$ depends on $M$, $R$ and $\gamma$. Indeed, the estimate follows from (3.4) with $v=(I,\varphi)$ by averaging in $\varphi$.

Our study of the limit $\mathcal{Q}^0$ uses the scheme, suggested by R. Khasminskii in [9] and is heavily based on the estimates for solutions $v^\nu(\tau)$, obtained above.

First we show that for any $k$ the difference,

$$I_k(\tau)-\int_0^\tau\langle F_k\rangle\big(I(s)\big)\,\mathrm{d}s,\tag{6.3}$$

is a martingale with respect to $\mathcal{Q}^0$ and the natural filtration of $\sigma$-algebras. A crucial step of the proof is to establish that

$$\mathfrak{A}^\nu:=\mathbf{E}\max_{0\leqslant\tau\leqslant T}\left|\int_0^\tau\big(F_k\big(I^\nu(s),\varphi^\nu(s)\big)-\langle F_k\rangle\big(I^\nu(s)\big)\big)\,\mathrm{d}s\right|\to 0,\tag{6.4}$$

as $\nu\to 0$. Proof of (6.4) occupies most of this section.

Let us fix an integer,

$$m \geqslant 1,$$

denote the first $m$ components of vectors $I^\nu$ and $\varphi^\nu$ by $I^{\nu,m}$ and $\varphi^{\nu,m}$, and rewrite the first $2m$ equations of the system (3.7) as follows:

$$\begin{aligned} dI^{\nu,m} &= F^m\big(I^\nu, \varphi^\nu\big)\, d\tau + \sigma^m\big(I^\nu, \varphi^\nu\big)\, d\beta_\tau, \\ d\varphi^{\nu,m} &= \left(\frac{1}{\nu} W^m\big(I^\nu\big) + G^m\big(I^\nu, \varphi^\nu\big)\right) d\tau + g^m\big(I^\nu, \varphi^\nu\big)\, d\beta_\tau. \end{aligned} \tag{6.5}$$

Here and afterwards we identify the vectors $(I_1^\nu, \dots, I_m^\nu, 0, 0, \dots)$ with $I^{\nu,m}$, and the vectors $(\varphi_1^\nu, \dots, \varphi_m^\nu, 0, 0, \dots)$ with $\varphi^{\nu,m}$.

Denote $\langle F_k\rangle_m(I^m) = \langle F_k\rangle_m(I, \varphi)_{I=(I_m,0),\varphi=0}$. By Lemma 3.2 there is a constant $C_k(R)$ such that for any $v = (I, \varphi)$, $|v| \leqslant R$, we have:

$$\big|F_k(I, \varphi) - F_k\big(I^m, \varphi^m\big)\big| \leqslant C_k(R) m^{-1}, \tag{6.6}$$

$$\big|\langle F_k\rangle_m\big(I^m\big) - \langle F_k\rangle(I)\big| \leqslant C_k(R) m^{-1}. \tag{6.7}$$

Define the event $\Omega_R$ as in Section 5. Due to Lemma 4.1,

$$\mathbf{P}(\Omega_R) \leqslant \varkappa_\infty(R)$$

(here and in similar situations below the function $\varkappa$ is $\nu$-independent). Since by Lemma 3.2 the function $F_k$ has a polynomial growth in $v$, then this estimate implies that

$$\left| \mathbf{E} \max_{0\leqslant\tau\leqslant T} \int_0^\tau F_k\big(v^\nu(s)\big)\, ds - \mathbf{E}_{\Omega_R} \max_{0\leqslant\tau\leqslant T} \int_0^\tau F_k\big(v^\nu(s)\big)\, ds \right| \leqslant \varkappa_\infty(R).$$

The functions $F_k(I^{\nu,m}, \varphi^{\nu,m})$, $\langle F_k\rangle_m(I^{\nu,m})$ and $\langle F_k\rangle(I^{\nu,m})$ satisfy similar relations. So we have:

$$\begin{aligned} \mathfrak{A}^\nu \leqslant{}& \varkappa_\infty(R) + \mathbf{E}_{\Omega_R} \max_{0\leqslant\tau\leqslant T} \left| \int_0^\tau \big\{F_k\big(I^\nu(s), \varphi^\nu(s)\big)\, ds - F_k\big(I^{\nu,m}(s), \varphi^{\nu,m}(s)\big)\big\}\, ds \right| \\ &+ \mathbf{E}_{\Omega_R} \max_{0\leqslant\tau\leqslant T} \left| \int_0^\tau \big\{F_k\big(I^{\nu,m}(s), \varphi^{\nu,m}(s)\big) - \langle F_k\rangle_m\big(I^{\nu,m}(s)\big)\big\}\, ds \right| \\ &+ \mathbf{E}_{\Omega_R} \max_{0\leqslant\tau\leqslant T} \left| \int_0^\tau \big\{\langle F_k\rangle_m\big(I^{\nu,m}(s)\big) - \langle F_k\rangle\big(I^\nu(s)\big)\big\}\, ds \right| \\ \leqslant{}& \varkappa_\infty(R) + C_k(R) m^{-1} \\ &+ \mathbf{E}_{\Omega_R} \max_{0\leqslant\tau\leqslant T} \left| \int_0^\tau \big\{F_k\big(I^{\nu,m}(s), \varphi^{\nu,m}(s)\big) - \langle F_k\rangle_m\big(I^{\nu,m}(s)\big)\big\}\, ds \right|. \end{aligned}$$

The last inequality here follows from (6.6)–(6.7). It remains to estimate the quantity:

$$\max_{0\leqslant\tau\leqslant T} \left| \int_0^\tau \big\{F_k\big(I^{\nu,m}(s), \varphi^{\nu,m}(s)\big) - \langle F_k\rangle_m\big(I^{\nu,m}(s)\big)\big\}\, ds \right| (1 - \chi_{\Omega_R}).$$

To do this we consider a partition of the interval $[0, T]$ to subintervals of length $\nu L$, $L > 1$ by the points:

$$\tau_j = \nu t_0 + \nu j L, \quad 0 \leqslant j \leqslant K + 1,$$

where $\tau_{K+1}$ is the last point $\tau_j$ in $[0, T]$. The constant $L$ such that

$$L \geqslant 2, \quad L \leqslant \frac{1}{2}\nu^{-1} \tag{6.8}$$

and the (deterministic) initial point $t_0 \in [0, L)$ will be chosen later. Note that

$$\frac{1}{2}T \leqslant K \cdot \nu L \leqslant T.$$

Denote

$$\eta_l = \int_{\tau_l}^{\tau_{l+1}} \big(F_k\big(I^{\nu,m}(s), \varphi^{\nu,m}(s)\big)\,\mathrm{d}s - \langle F_k\rangle_m\big(I^{\nu,m}(s)\big)\big)\,\mathrm{d}s, \quad 0 \leqslant l \leqslant K.$$

Since outside the event $\Omega_R$ we have:

$$\left|\int_{\tau'}^{\tau''} \big(F_k\big(I^{\nu,m}(s), \varphi^{\nu,m}(s)\big) - \langle F_k\rangle_m\big(I^{\nu,m}(s)\big)\big)\,\mathrm{d}s\right| \leqslant \nu L C(R)$$

for any $\tau' < \tau''$ such that $\tau'' - \tau' \leqslant \nu L$, then

$$\begin{aligned} &\mathbf{E}_{\Omega_R} \max_{0\leqslant\tau\leqslant T}\left|\int_0^{\tau} \big(F_k\big(I^{\nu,m}(s), \varphi^{\nu,m}(s)\big) - \langle F_k\rangle_m\big(I^{\nu,m}(s)\big)\big)\,\mathrm{d}s\right| \\ &\quad\leqslant \mathbf{E}_{\Omega_R}\sum_{l=0}^{K} |\eta_l| + \nu L C(R). \end{aligned} \tag{6.9}$$

To calculate the contribution from the integral over an $l$th subinterval, we pass there to the slow time $t = \nu^{-1}\tau$. Now the system (6.5) reads as

$$\begin{aligned} \mathrm{d}I^{\nu,m}(t) &= \nu F^m\big(I^\nu, \varphi^\nu\big)\,\mathrm{d}t + \sqrt{\nu}\sigma^m\big(I^\nu, \varphi^\nu\big)\,\mathrm{d}\beta_t, \\ \mathrm{d}\varphi^{\nu,m}(t) &= \big(W^m\big(I^\nu\big) + \nu G^m\big(I^\nu, \varphi^\nu\big)\big)\,\mathrm{d}t + \sqrt{\nu}g^m\big(I^\nu, \varphi^\nu\big)\,\mathrm{d}\beta_t. \end{aligned} \tag{6.10}$$

Denoting $t_j = \tau_j/\nu = t_0 + jL$ we have:

$$\begin{aligned} |\eta_l| \leqslant{}& \nu\left|\int_{t_l}^{t_{l+1}} \big\{F_k\big(I^{\nu,m}(x), \varphi^{\nu,m}(x)\big) - F_k\big(I^{\nu,m}(t_l), \varphi^{\nu,m}(t_l) + W^m\big(I^\nu(t_l)\big)(x - t_l)\big)\big\}\,\mathrm{d}x\right| \\ &+ \nu\left|\int_{t_l}^{t_{l+1}} \big\{F_k\big(I^{\nu,m}(t_l), \varphi^{\nu,m}(t_l) + W^m\big(I^\nu(t_l)\big)(x - t_l)\big) - \langle F_k\rangle_m\big(I^{\nu,m}(t_l)\big)\big\}\,\mathrm{d}x\right| \\ &+ \nu\left|\int_{t_l}^{t_{l+1}} \big\{\langle F_k\rangle_m\big(I^{\nu,m}(t_l)\big) - \langle F_k\rangle_m\big(I^{\nu,m}(x)\big)\big\}\,\mathrm{d}x\right| = \Upsilon_l^1 + \Upsilon_l^2 + \Upsilon_l^3. \end{aligned}$$

To estimate the integrals $\Upsilon_l^1 - \Upsilon_l^3$ we first optimise the choice of $t_0$. Defining the event $\Omega(\delta)$, the number $M(R, m)$ and the set $Q_\gamma$ as in Section 5, we have the:

**Lemma 6.1.** *The non-random number $t_0 \in [0, \nu L)$ (depending on $\nu$ and $\delta$) can be chosen in such a way that*

$$\frac{1}{K}\sum_{l=0}^{K} \mathbf{P}\mathcal{E}_l \leqslant \varkappa_\infty(R) + \varkappa\big(\gamma^{-1}; R, m\big) + \varkappa\big(\delta^{-1}; \gamma, R, m\big), \tag{6.11}$$

*for all* $0 < \delta,\ \gamma < 1$, *where*

$$\mathcal{E}_l = \Omega_R \cup \{I(\tau_l) \in Q_\gamma\} \cup \{W^m(\tau_l) \in \Omega(\delta)\}.$$

**Proof.** Due to Lemmas 5.2 and 4.3,

$$\int_0^T \mathbf{P}\big(\Omega_R \cup \{I(\tau) \in Q_\gamma\} \cup \{W^m(\tau) \in \Omega(\delta)\}\big)\,\mathrm{d}\tau$$
$$\leqslant \varkappa_\infty(R) + \varkappa\big(\gamma^{-1}; R, m\big) + \varkappa\big(\delta^{-1}; R, m, \gamma\big).$$

Writing the left-hand side as $\int_0^{\nu L} \sum_{l=0}^K \mathbf{P}(\mathcal{E}_l)\,\mathrm{d}t_0$, where $\mathcal{E}_l$ is defined in terms of $\tau_l = t_0 + \nu j L$, and applying the meanvalue theorem we get the assertion. □

Applying the Doob inequality and Lemmas 3.2, 4.1 to (3.6) we get that

$$\mathbf{P}_{\Omega_R}\Big(\sup_{t_l \leqslant t \leqslant t_{l+1}} \big|I^\nu(t) - I^\nu(t_l)\big| \geqslant P(R)\nu L + \Delta\Big)$$
$$\leqslant \mathbf{P}\left(\sup_{t_l \leqslant t \leqslant t_{l+1}} \nu \left|\int_{t_l}^t \sigma\big(v(s)\big)\,\mathrm{d}\beta_s\right|^2 \geqslant \Delta^2\right) \leqslant C_N(\nu L)^N \Delta^{-2N},$$

for all $N$ and $\Delta$. Choosing in this inequality $\Delta = (\nu L)^{1/3}$, using (6.8) and denoting,

$$Q_l = \Big\{\sup_{t_l \leqslant t \leqslant t_{l+1}} \big|I^\nu(t) - I^\nu(t_l)\big| \geqslant P_1(R)(\nu L)^{1/3}\Big\},$$

where $P_1$ is a suitable polynomial, we have:

$$\mathbf{P}_{\Omega_R}(Q_l) \leqslant \varkappa_\infty\big((\nu L)^{-1}; m\big). \tag{6.12}$$

Let us set

$$\mathcal{F}_l = \mathcal{E}_l \cup Q_l, \quad l = 0, 1, \ldots, K.$$

Then (6.11) implies the estimate:

$$\frac{1}{K}\sum_{l=0}^K \mathbf{P}\mathcal{F}_l \leqslant \varkappa_\infty(R) + \varkappa\big(\gamma^{-1}; R, m\big) + \varkappa\big(\delta^{-1}; \gamma, R, m\big) + \varkappa_\infty\big((\nu L)^{-1}; m\big) =: \kappa.$$

Since $F_k(I, \varphi)$ has a polynomial growth in $I$, then

$$\sum_{l=0}^K \big|(\mathbf{E} - \mathbf{E}_{\mathcal{F}_l})\Upsilon_l^j\big| \leqslant P(R)\frac{1}{K}\sum_{l=0}^K \mathbf{P}\mathcal{F}_l \leqslant \kappa \quad (j = 1, 2, 3), \tag{6.13}$$

where we denoted by $\kappa$ another function of the same form as above. So it remains to estimate the expectations $\mathbf{E}_{\mathcal{F}_l}\Upsilon_l^j$ and their sums in $l$.

First we study increments of the process $\varphi^{\nu,m}(t)$. Let us denote:

$$\varphi^{\nu,m}(t) - \varphi^{\nu,m}(t_l) - W^m\big(I^\nu(t_l)\big)(t - t_l) =: \Phi_l^\nu(t), \quad t_l \leqslant t \leqslant t_{l+1}.$$

Then

$$\Phi_l^\nu(t) = \int_{t_l}^t \big(W^m\big(I^\nu(x)\big) - W^m\big(I^\nu(t_l)\big)\big)\,\mathrm{d}x + \nu\int_{t_l}^t G^m\,\mathrm{d}x + \sqrt{\nu}\int_{t_l}^t g^m\,\mathrm{d}\beta_x =: J_1 + J_2 + J_3.$$

Outside the event $\mathcal{F}_l$ the term $J_1$ estimates as follows:

$$|J_1| \leqslant P(R, m)(\nu L)^{1/3} L.$$

To estimate $J_2$ and $J_3$ we assume that

$$P(R)(\nu L)^{1/3} \leqslant \frac{1}{2}\gamma. \tag{6.14}$$

Then outside $\mathcal{F}_l$ we have,

$$\left|I_k^\nu(t)\right| \geqslant \frac{1}{2}\gamma \quad \forall t \in [t_l, t_{l+1}],\ k \leqslant m,$$

so by Lemma 3.2 and (6.14) there we have:

$$|J_2| \leqslant \nu L C(R)\gamma^{-1} \leqslant C'(R)(\nu L)^{2/3}.$$

To bound $J_3$ we introduce the stopping time,

$$t' = \min\Big\{t \geqslant t_l\colon \min_{k\leqslant M} I_k^\nu(t) \leqslant \gamma \text{ or } \left|I^\nu(t)\right| \geqslant R\Big\} \wedge t_{l+1}.$$

Then

$$(1-\chi_{\mathcal{F}_l})\left|J_3(t)\right| \leqslant \sqrt{\nu}\left|\int_{t_l}^{t'\wedge t} g^m(s)\,\mathrm{d}\beta_s\right| =: J_3'(t).$$

We have $\nu\mathbf{E}\int_{t_l}^{t'} |g^m|^2\,\mathrm{d}s \leqslant \nu L\gamma^{-1}C(R,m)$. So the Doob inequality implies:

$$\begin{aligned}\mathbf{P}_{\mathcal{F}_l}\Big\{\sup_{t_l\leqslant t\leqslant t_{l+1}} |J_3| \geqslant (\nu L)^{1/3}\Big\} &\leqslant \mathbf{P}\Big\{\sup_{t_l\leqslant t\leqslant t_{l+1}} \left|J_3'\right| \geqslant (\nu L)^{1/3}\Big\}\\ &\leqslant (\nu L)^{1/3}\gamma^{-1}C(R,m).\end{aligned}$$

We have seen that

$$\mathbf{P}_{\mathcal{F}_l}\left\{\Phi_l^\nu \geqslant P'(R,m)\nu^{1/3}L^{4/3}\right\} \leqslant (\nu L)^{1/3}\gamma^{-1}C(R,m). \tag{6.15}$$

Now we may estimate the terms $\Upsilon_l^j$.

**Terms $\boldsymbol{\Upsilon_j^1}$.** Since $F_k \in \mathrm{Lip}_{\mathrm{lock}}(h)$, then by (6.15) 'probability' $\mathbf{P}_{\mathcal{F}_l}$ that the integrand in $\Upsilon_l^1$ is $\geqslant C(R,m)\nu^{1/3}L^{4/3}$ is bounded by $(\nu L)^{1/3}\gamma^{-1}C(R,m)$. Since outside $\mathcal{F}_l$ the integrand is $\leqslant C(R,m)$, then

$$\sum_l \mathbf{E}_{\mathcal{F}_l}\Upsilon_l^1 \leqslant \nu^{1/3}C(R,m,L,\gamma).$$

**Terms $\boldsymbol{\Upsilon_j^2}$.** By Lemma 5.1, outside $\mathcal{F}_l$,

$$\Upsilon_l^2 \leqslant \nu\delta^{-1}C(R,m) + L\nu m^{-1}C(R).$$

So

$$\sum_l \mathbf{E}_{\mathcal{F}_l}\Upsilon_l^2 \leqslant (\delta L)^{-1}C(R,m) + m^{-1}C(R).$$

**Terms $\boldsymbol{\Upsilon_j^3}$.** By Lemma 4.2, outside $\mathcal{F}_l$ we have $\Upsilon_l^3 \leqslant P(R)(\nu L)^{1/3}(\nu L)$. So

$$\sum_l \mathbf{E}_{\mathcal{F}_l}\Upsilon_l^3 \leqslant P(R)(\nu L)^{1/3}.$$

Now (6.13) and the obtained estimates on the terms $\Upsilon_l^j$ imply that

$$\sum_l \mathbf{E}|\eta_l| \leqslant \kappa + \nu^{1/3}C(R,m,L,\gamma) + (\delta L)^{-1}C(R,m) + m^{-1}C(R).$$

Using (6.9) we arrive at the final estimate:

$$\mathfrak{A}^\nu \leqslant \varkappa_\infty(R) + C(R)m^{-1} + \nu LC(R) + \langle \text{same terms as in the right-hand side above} \rangle. \tag{6.16}$$

It is easy to see that for any $\varepsilon > 0$ we can choose our parameters in the following order,

$$R \to m \to \gamma \to \delta \to L \to \nu,$$

so that (6.8), (6.14) hold and the right-hand side of (6.16) is $< \varepsilon$.

Thus, we have proved the:

**Proposition 6.2.** *The limit relation* (6.4) *holds true.*

In the same way one can show that

$$\mathbf{E} \max_{0\leqslant t\leqslant T} \left| \int_0^t \{ F_k\big(I^\nu(s), \varphi^\nu(s)\big) - \langle F_k\rangle\big(I^\nu(s)\big) \} \, \mathrm{d}s \right|^4 \to 0 \quad \text{as } \nu \to 0. \tag{6.17}$$

From Proposition 6.2 taking into account the a priori estimates we finally derive:

**Proposition 6.3.** *The process* (6.3) *is a square integrable martingale with respect to the limit measure $\mathcal{Q}^0$ and the natural filtration of $\sigma$-algebras in $C([0,\infty); h_{I+})$.*

**Proof.** Let us consider the processes:

$$N_k^{\nu_j}(\tau) = I_k^{\nu_j}(\tau) - \int_0^\tau \langle F_k\rangle\big(I^{\nu_j}(s)\big)\, \mathrm{d}s, \quad \tau \in [0,T],\ j = 1, 2, \ldots.$$

Due to (3.7) and (6.4) we can write $N_k^{\nu_j}$ as

$$N_k^{\nu_j}(\tau) = M_k^{\nu_j}(\tau) + \Xi_k^{\nu_j}(\tau).$$

Here $M_k^{\nu_j} = I_k^{\nu_j} - \int F_k(I^{\nu_j}, \varphi^{\nu_j})$ is a martingale, and $\Xi_k^{\nu_j}$ is a process such that

$$\mathbf{E} \sup_{0\leqslant\tau\leqslant T} \big|\Xi_k^{\nu_j}(\tau)\big| \to 0 \quad \text{as } \nu_j \to 0.$$

This convergence implies that

$$\lim_{\nu_j\to 0} \mathcal{L}\big(N_k^{\nu_j}(\cdot)\big) = \lim_{\nu_j\to 0} \mathcal{L}\big(M_k^{\nu_j}(\cdot)\big) \tag{6.18}$$

in the sense that if one limit exists, then another one exists as well and the two are equal.

Due to (6.1) and the Skorokhod theorem, we can find random processes $J^{\nu_j}(\tau)$ and $J(\tau)$, $0 \leqslant t \leqslant T$, such that $\mathcal{L}J^{\nu_j}(\cdot) = \mathcal{L}I^{\nu_j}(\cdot)$, $\mathcal{L}J(\cdot) = \mathcal{Q}^0$, and

$$J^{\nu_j} \to J \text{ in } C\big([0,T], h_I\big) \quad \text{as } \nu_j \to 0, \tag{6.19}$$

almost surely. By Lemma 4.1,

$$\mathbf{P}\Big\{ \sup_{0\leqslant\tau\leqslant T} \big|I^{\nu_j}(\tau)\big| \geqslant R \Big\} \leqslant CR^{-1},$$

uniformly in $\nu_j$. Since $\langle F_k\rangle \in \mathrm{Lip}_{\mathrm{lock}}(h_I)$ by Lemmas 3.2 and 4.2, then (6.19) implies that the left limit in (6.18) exists and equals (6.3). By Lemmas 3.2 and 4.1 the family of martingales $M_k^{\nu_j}(\tau)$ is uniformly integrable. Since they converge in distribution to the process (6.3), then the latter is a martingale as well. $\square$

Denote $Z_k(t) \equiv I_k(t) - \int_0^t \langle F_k\rangle(I(s))\,\mathrm{d}s$. Using the same arguments as above and (6.17) we can show that $Z_k(t)Z_j(t) - \int_0^t \langle A_{kj}\rangle(I(s))\,\mathrm{d}s$ is a $\mathcal{Q}^0$-martingale in $C([0,T); h_{I+})$. Combining the above statement we arrive at the following theorem, where $T > 0$ and $p \geqslant 3$ are any fixed numbers.

**Theorem 6.4.** *Let the process $u^\nu(t)$, $0 < \nu \leqslant 1$, be a solution of Eq.* (1.2) *which either is stationary in time, or satisfies the $\nu$-independent initial condition* (1.3)*, where $u_0$ is non-random and smooth. Let $\Psi(u^\nu(\tau)) = v^\nu(\tau) = (I^\nu(\tau), \varphi^\nu(\tau))$. Then any limiting point $\mathcal{Q}^0$ of the family $\mathcal{L}\{I^\nu(\cdot)\}$ as $\nu \to 0$ is a measure in $C(0,T;h^p_{I+})$ which satisfies the estimates*:

$$\int \sup_{0\leqslant\tau\leqslant T} \left|I(\tau)\right|^N_{h^m_I} \mathcal{Q}_0\big(\mathrm{d}I(\cdot)\big) \leqslant C(N,m,T) < \infty \quad \forall N, m \in \mathbb{N},$$

*and solves the martingale problem in $C(0,T;h^p_I)$ with the drift $\langle F\rangle(I)$ and covariance $\langle A\rangle(I)$.*

Let $\sigma^0(I)$ be a symmetric square root of $\langle A\rangle(I)$ so that $(\sigma^0(I))\sigma^0(I)^t = \langle A\rangle(I)$. We recall that $\langle A\rangle(I)$ is a positive compact operator for each $I \in h$.

**Corollary 6.5.** *Any limiting measure $\mathcal{Q}^0$ as in Theorem* 6.4 *is the distribution of a solution $I(\tau)$ of the following stochastic differential equation*:

$$\mathrm{d}I = \big\langle F(I)\big\rangle \mathrm{d}\tau + \sigma^0(I)\,\mathrm{d}W_\tau, \tag{6.20}$$

*where $W_t$ is a cylindrical Brownian motion on $h^0_I$.*

**Proof.** Denote by $\mathfrak{h}$ the Hilbert space of sequences $\{x_1, x_2 \ldots,\}$ with the norm $|x|^2_{\mathfrak{h}} = \sum_{j=1}^\infty j^{2(2p+4)} x_j^2$. It is easy to check that $\mathfrak{h}$ is continuously embedded in $h^p_I$, thus all the coefficients $\langle F(I)\rangle$, $\sigma^0(I)$ and $\langle A\rangle(I)$ are well defined for any $I \in \mathfrak{h}$.

By Theorem 6.4 and Lemma 4.1 the measure $\mathcal{Q}^0$ is concentrated on $C(0,T;h^{2p+4}_I)$. Since this space is continuously embedded in $C(0,T;\mathfrak{h})$, then $\mathcal{Q}^0$ is also concentrated on $C(0,T;\mathfrak{h})$. Therefore, $\mathcal{Q}^0$ is a solution of the above limit martingale problem in the Hilbert space $\mathfrak{h}$. It remains to use Theorem IV.3.5 in [20] (also see [4]). □

The limiting measure $\mathcal{Q}^0$ and the process $I(\tau)$ inherit the uniform in $\nu$ estimates on the processes $I^\nu(\tau)$, obtained in Sections 1–4. For example,

$$\mathbf{P}\big\{I_k(\tau) < \delta\big\} \leqslant \varkappa\big(\delta^{-1};k\big) \text{ uniformly in } \tau \in [0,T], \quad \text{for any } k \geqslant 1. \tag{6.21}$$

In particular, $I(\tau) \in h^p_{I+} \setminus \partial h^p_{I+}$ a.e., for any $\tau \geqslant 0$.

**Remark.** Eq. (6.20) is the Whitham equation for the damped-driven KdV equation (0.7). Our results show that it has a weak solution in the space $h^p_{I+}$ for a given $I(0)$ which is a deterministic vector in the space $h^\infty_{I+} = \bigcap h^p_{I+}$. In fact, the same arguments apply when $I(0)$ is a random variable in $h^p_{I+}$ such that $\mathbf{E}\|I(0)\|^N_{h^p_I} < \infty$, where $N$ and $p$ are large enough.

Now we assume that $u^\nu(t)$ is a stationary solution of (1.2). Then the limiting process $I(\tau)$ as in (6.20) is stationary in $\tau$. We denote $q^0 = \mathcal{L}(I(0))$ (this is a measure on the space $h^p_{I+}$).

**Theorem 6.6.** *Let a process $u^\nu(t)$ be a stationary solution of Eqs.* (1.2)*. Then,*

(1) *for any $0 \leqslant \tau \leqslant T$ the law of $\varphi^\nu(\tau)$ converges weakly as $\nu \to 0$ to the Haar measure* $\mathrm{d}\varphi$ *on $\mathbb{T}^\infty$.*
(2) *The law of the pair $(I^\nu(\tau), \varphi^\nu(\tau))$ converges, along a subsequence $\{\nu_j\}$, corresponding to the measure $q^0$, to the product measure $q^0 \times \mathrm{d}\varphi$.*
(3) *For any $m$ the measure $q^{0m} = \mathcal{L}(I^m(0))$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^m_+$.*

More precisely, the second assertion of the theorem means the following: Due to (6.21) the limiting measure is supported by the Borel set $h^p \cap \{v\colon \mathbf{v}_j \neq 0\ \forall j\}$, which is measurably isomorphic to $(h^p_{I+} \setminus \partial h^p_{I+}) \times \mathbb{T}^\infty$. Under this isomorphism the limiting measure reeds as $q^0 \times \mathrm{d}\varphi$.

**Proof.** (1) Let us fix any $m$ and take a bounded Lipschitz function $f$, defined on the torus $\mathbb{T}^m \subset \mathbb{T}^\infty$. Then

$$\mathbf{E} f\big(\varphi^\nu(\tau)\big) = \frac{1}{T}\int_0^T \mathbf{E} f\big(\varphi^{\nu,m}(s)\big)\,\mathrm{d}s = \frac{1}{T}\mathbf{E}\int_0^T f\big(\varphi^{\nu,m}(s)\big)\,\mathrm{d}s,$$

where $\varphi^{\nu,m}$ satisfies (6.5). Arguing as when estimating the expectation in the left-hand side of (6.9) in the proof of Theorem 6.4, we get that

$$\mathbf{E}\int_0^T \big(f\big(\varphi^{\nu,m}\big) - \langle f\rangle\,\mathrm{d}s\big) \to 0 \quad \text{as } \nu \to 0.$$

Therefore $\mathbf{E} f(\varphi^\nu(\tau)) \to \langle f\rangle$, and the first assertion of the theorem follows.

(2) Consider an arbitrary bounded Lipschitz test function of the form $\Phi(I,\varphi) = f(I^m)g(\varphi^m)$, $m \geqslant 1$. We have:

$$\mathbf{E}\Phi\big(I^\nu(\tau), \varphi^\nu(\tau)\big) = \frac{\nu}{T}\mathbf{E}\int_0^{\nu^{-1}T} f\big(I^{\nu,m}(t)\big) g\big(\varphi^{\nu,m}(t)\big)\,\mathrm{d}t.$$

Consider a uniform partition of the interval $(0, \nu^{-1}T)$ into sufficiently long subintervals. As was shown in the proof of Theorem 6.4, with high probability on any subinterval of the partition the function $I^{\nu,m}(t)$ does not deviate much from a random constant (see (6.12)), while the normalised integral of $g(\varphi^{\nu,m}(t))$ approaches the integral of $g$ against the Haar measure (see the proof of the first assertion). Therefore when $\nu \to 0$, the right-hand side above can be written as

$$\left(\frac{\nu}{T}\mathbf{E}\int_0^{\nu^{-1}T} f\big(I^{\nu,m}(s)\big)\,\mathrm{d}s\right)\int_{\mathbb{T}^\infty} g\big(\varphi^m\big)\,\mathrm{d}\varphi + o(1) = \int_{h_I^p} f\big(I^m\big)\,\mathrm{d}q^0 \int_{\mathbb{T}^\infty} g\big(\varphi^m\big)\,\mathrm{d}\varphi + \mathrm{o}(1).$$

This completes the proof of (2).

(3) The vector $I^m(\tau)$ satisfies the Ito equation, given by the first $m$ components of (6.20). The corresponding diffusion is non-degenerate by (6.2). Therefore by the Krylov theorem (see [13]) for any Borel set $U \subset [\delta, \delta^{-1}]^m$, $\delta > 0$, we have that

$$q^{0m}(U) = \mathbf{P}\big\{I^{\nu m}(t) \in U\big\} \leqslant C_\delta |U|^{1/m}. \tag{6.22}$$

Let us take any zero-set $Z \subset \mathbb{R}^m_+$ and write it as

$$Z = Z_1 \cup \cdots \cup Z_m \cup \hat{Z}, \quad \text{where } Z_j \subset \{I_j = 0\} \text{ and } \hat{Z} \subset \mathbb{R}^m_{>0}.$$

Then $q^{0m}(Z_j) = 0$ for each $j$ due to (6.21). Writing $\hat{Z} = \bigcup_{\delta>0} Z_\delta$, where $Z_\delta = Z \cap [\delta, \delta^{-1}]^m$, we use (6.22) to get that $q^{0m}(\hat{Z}) = \lim q^{0m}(Z_\delta) = 0$. So $q^{0m}(Z) = 0$ and the theorem's proof is completed. □

**Remark.** For any $j \geqslant 1$ the measure $q_j^0 = \mathcal{L}(I_j(0))$ satisfies an analogy of estimate (6.22) with $m = 1$. Therefore $q_j^0 = f_j(s)\,\mathrm{d}s$, $s \geqslant 0$, where the function $f_j$ is bounded on segments $[\delta, \delta^{-1}]$.

## 7. Proof of Lemma 4.3

**Step 1.** Processes $\tilde{\mathbf{v}}_k^\gamma(\tau)$.

For $\eta_1, \eta_2 \in \mathbb{R}^2 \setminus \{0\}$ we denote by $U(\eta_1, \eta_2)$ the element of $SO(2)$ such that $U(\eta_1, \eta_2)\frac{\eta_2}{|\eta_2|} = \frac{\eta_1}{|\eta_1|}$. Note that $U(\eta_2, \eta_1) = U(\eta_1, \eta_2)^{-1} = U(\eta_1, \eta_2)^*$.

In the fast time $\tau$ Eq. (3.1) reads,

$$\mathrm{d}\mathbf{v}_k = \left(\frac{1}{\nu}\,\mathrm{d}\Psi_k(u)V(u) + A_k(v)\right)\mathrm{d}\tau + \sum_j B_{kj}(v)\,\mathrm{d}\beta_\tau^j, \tag{7.1}$$

where we denoted:

$$A_k(v) = \mathrm{d}\Psi_k(u)u_{xx} + \frac{1}{2}\sum_{j\in\mathbb{Z}_0} b_j^2\,\mathrm{d}^2\Psi_k(u)[e_j, e_j], \qquad B_{kj}(v) = \mathrm{d}\Psi_k(u)b_j e_j.$$

Let $v(\tau) = \{\mathbf{v}_k(\tau),\ k \geqslant 1\}$ be a solution of the system $(7.1)_{k\in\mathbb{N}}$.

We introduce the functions,

$$\tilde{A}_k(\tilde{\mathbf{v}}_k, v) = U(\tilde{\mathbf{v}}_k, \mathbf{v}_k)A_k(v), \qquad \tilde{B}_{kj}(\tilde{\mathbf{v}}_k, v) = U(\tilde{\mathbf{v}}_k, \mathbf{v}_k)B_{kj}(v),$$

smooth in $(\tilde{\mathbf{v}}_k, \mathbf{v}_k)$ from $(\mathbb{R}^2 \setminus \{0\}) \times (\mathbb{R}^2 \setminus \{0\})$, and consider the additional stochastic equation for $\tilde{\mathbf{v}}_k(\tau) \in \mathbb{R}^2$:

$$\mathrm{d}\tilde{\mathbf{v}}_k = \tilde{A}_k(\tilde{\mathbf{v}}_k, v)\,\mathrm{d}\tau + \sum_j \tilde{B}_{kj}(\tilde{\mathbf{v}}_k, v)\,\mathrm{d}\beta_\tau^j. \tag{7.2}$$

Its coefficients are well defined for all non-zero $\mathbf{v}_k$ and $\tilde{\mathbf{v}}_k$.

If $v(\tau) = \{\mathbf{v}_k(\tau),\ k \geqslant 1\}$ is as above, then Eq. (7.2) with a prescribed initial data has a unique solution, defined while

$$|\mathbf{v}_k|, |\tilde{\mathbf{v}}_k| \geqslant c, \qquad |v|_{h^1} \leqslant C,$$

where $c, C$ are any fixed positive constants. This solution may be obtained as the last component of a solution $(v, \tilde{\mathbf{v}}_k)$ of the coupled system $(7.1)_{k\in\mathbb{N}}$, (7.2). This system has a unique solution since $(7.1)_{k\in\mathbb{N}}$ is equivalent to (0.7) (so it has a unique solution), while (7.2) is a Lipschitz equation on the domain, defined by the conditions above.

For a $\gamma \in (0, \frac{1}{2})$ we introduce the stopping times $\tau_i^+$, $i \geqslant 0$, and $\tau_i^-$, $i \geqslant 1$, where $\tau_0^+ = 0$ and for $i \geqslant 1$:

$$\tau_i^- = \inf\left\{\tau \geqslant \tau_{i-1}^+\colon\ \left|\mathbf{v}_k(\tau)\right| \leqslant \gamma \text{ or } \left|v(\tau)\right|_{h^1} \geqslant \frac{1}{\gamma}\right\},$$
$$\tau_i^+ = \inf\left\{\tau \geqslant \tau_i^-\colon\ \left|\mathbf{v}_k(\tau)\right| \geqslant 2\gamma \text{ and } \left|v(\tau)\right|_{h^1} \leqslant \frac{1}{2\gamma}\right\}.$$

Note that $\tau_0^+ \leqslant \tau_1^-$, $\tau_i^- < \tau_i^+ < \tau_{i+1}^-$ if $i > 0$, and $\tau_j^\pm \to \infty$ as $j \to \infty$.

Next we construct a continuous process $\tilde{\mathbf{v}}_k^\gamma(\tau)$, $\tau \geqslant 0$. We set $\tilde{\mathbf{v}}_k^\gamma(\tau_0^+) = \mathbf{v}_k(\tau_0^+)$. For $i = 1$ we extend $\tilde{\mathbf{v}}_k^\gamma(\tau)$ to the segment $\Delta_{i-1} := [\tau_{i-1}^+, \tau_i^-]$ as a solution of Eq. (7.2), and on the segment $\Lambda_i = [\tau_i^-, \tau_i^+]$ we define it as[5]

$$\tilde{\mathbf{v}}_k^\gamma(\tau) = U\big(\tilde{\mathbf{v}}_k(\tau_i^-), \mathbf{v}_k(\tau_i^-)\big)\mathbf{v}_k(\tau), \quad \text{for } \tau \in \Lambda_i. \tag{7.3}$$

**Lemma 7.1.** *If $|\tilde{\mathbf{v}}_k^\gamma(\tau_{i-1}^+)| = |\mathbf{v}_k(\tau_{i-1}^+)|$ and $\tilde{\mathbf{v}}_k^\gamma$ satisfies* (7.2) *on $\Delta_{i-1}$, then $|\tilde{\mathbf{v}}_k^\gamma| = |\mathbf{v}_k|$ everywhere on that segment.*

**Proof.** Application of Ito's formula to the expression $\tilde{I}_k^\gamma = \frac{1}{2}|\tilde{\mathbf{v}}_k^\gamma|^2$ on the segment $\Delta_{i-1}$ yields:

$$\mathrm{d}\tilde{I}_k^\gamma = \big(\tilde{\mathbf{v}}_k^\gamma, \tilde{A}_k(\tilde{\mathbf{v}}_k^\gamma, v)\big)\,\mathrm{d}\tau + \sum_l\left(\frac{1}{2}\left|\tilde{B}_{kl}(\tilde{\mathbf{v}}_k^\gamma, v)\right|^2\mathrm{d}\tau + \big(\tilde{\mathbf{v}}_k^\gamma, \tilde{B}_{kl}(\tilde{\mathbf{v}}_k^\gamma, v)\big)\,\mathrm{d}\beta_\tau^l\right).$$

Similarly, $I_k = \frac{1}{2}|\mathbf{v}_k|^2$ satisfies:

$$\mathrm{d}I_k = \big(\mathbf{v}_k, A_k(v)\big)\,\mathrm{d}\tau + \sum_l\left(\frac{1}{2}\left|B_{kl}(v)\right|^2\mathrm{d}\tau + \big(\mathbf{v}_k, B_{kl}(v)\big)\,\mathrm{d}\beta_\tau^l\right).$$

By construction, the drift and diffusion coefficients of these two equations satisfy the relations:

$$\big(\tilde{\mathbf{v}}_k^\gamma, \tilde{A}_k(\tilde{\mathbf{v}}_k^\gamma, v)\big) + \frac{1}{2}\sum_l\left|\tilde{B}_{kl}(\tilde{\mathbf{v}}_k^\gamma, v)\right|^2 = \frac{|\tilde{\mathbf{v}}_k^\gamma|}{|\mathbf{v}_k|}\big(\mathbf{v}_k, A_k(v)\big) + \frac{1}{2}\sum_l\left|B_{kl}(v)\right|^2,$$
$$\big(\tilde{\mathbf{v}}_k^\gamma, \tilde{B}_{kl}(\tilde{\mathbf{v}}_k^\gamma, v)\big) = \frac{|\tilde{\mathbf{v}}_k^\gamma|}{|\mathbf{v}_k|}\big(\mathbf{v}_k, B_{kl}(v)\big).$$

[5] If $\mathbf{v}_k(0) = 0$, then $\tau_0^+ = \tau_1^- = 0$ and the formula (7.3) is not defined. But it happens with zero probability, and in this case we simply set $\tilde{\mathbf{v}}_k^\gamma \equiv 0$.

For the squared difference $(I_k - \tilde I_k^\gamma)^2$ we have:

$$\mathrm{d}\big(I_k - \tilde I_k^\gamma\big)^2 = \left(2\big(I_k - \tilde I_k^\gamma\big)\frac{|\mathbf{v}_k| - |\tilde{\mathbf{v}}_k^\gamma|}{|\mathbf{v}_k|}\big(\mathbf{v}_k, A_k(v)\big) + \frac{(|\mathbf{v}_k| - |\tilde{\mathbf{v}}_k^\gamma|)^2}{|\mathbf{v}_k|^2}\sum_l \big(\mathbf{v}_k, B_{kl}(v)\big)^2\right)\mathrm{d}\tau + \mathrm{d}\mathcal{M}_\tau, \tag{7.4}$$

where $\mathcal{M}_\tau$ is a square integrable stochastic integral whose structure is of no interest. Denote $J^\gamma(\tau) = (I_k - \tilde I_k^\gamma)^2((\tau \vee \tau_i^+) \wedge \tau_{i+1}^-)$. Since

$$|\mathbf{v}_k| - |\tilde{\mathbf{v}}_k^\gamma| = 2\,\frac{I_k - \tilde I_k^\gamma}{|\mathbf{v}_k| + |\tilde{\mathbf{v}}_k^\gamma|},$$

then it follows from (7.4) that $\mathbf{E}J^\gamma(\tau) \leqslant \mathbf{E}J^\gamma(0) + C(\gamma)\int_0^\tau \mathbf{E}J^\gamma(s)\,\mathrm{d}s$. As $J^\gamma(\tau_{i-1}^+) = 0$, then $J^\gamma(\tau) \equiv 0$ by the Gronwall lemma. That is, $|\tilde{\mathbf{v}}_k^\gamma| = |\mathbf{v}_k|$ on $\Delta_{i-1}$. □

Applying this lemma with $i = 1$ we see that (7.3) with $i = 1$ is well defined, and $|\tilde{\mathbf{v}}_k^\gamma| = |\mathbf{v}_k|$ on $\Delta_0 \cup \Lambda_1$. Repeating the construction above for $i = 2, 3, \ldots$ we get a continuous process $\tilde{\mathbf{v}}_k^\gamma(\tau)$, $\tau \geqslant 0$, satisfying (7.2) on the segments $\Delta_i$, $i \geqslant 0$, satisfying (7.3) on the segments $\Lambda_i$, $i \geqslant 1$, and such that

$$\big|\tilde{\mathbf{v}}_k^\gamma(\tau)\big| \equiv \big|\mathbf{v}_k(\tau)\big|.$$

Let us abbreviate $U_i = U(\tilde{\mathbf{v}}_k(\tau_i^-), \mathbf{v}_k(\tau_i^-))$. Then on the intervals $\Lambda_i$ the process $\tilde{\mathbf{v}}_k^\gamma(\tau)$ satisfies the equation:

$$\mathrm{d}\tilde{\mathbf{v}}_k^\gamma(\tau) = U_i\left(\frac{1}{\nu}\mathrm{d}\Psi_k(u)V(v) + A_k\right)\mathrm{d}\tau + U_i B_{kj}(v)\,\mathrm{d}\beta_\tau^j.$$

Finally, using the notation,

$$\hat A_k(\tilde{\mathbf{v}}_k, v, t) = \begin{cases} \tilde A_k(\tilde{\mathbf{v}}_k, v), & \tau \in \bigcup_i \Delta_i, \\ U_i\big(\frac{1}{\nu}\mathrm{d}\Psi_k(u)V(v) + A_k\big), & \tau \in \bigcup_i (\tau_i^-, \tau_i^+), \end{cases}$$

and

$$\hat B_{kj}(\tilde{\mathbf{v}}_k, v, t) = \begin{cases} \tilde B_{kj}(\tilde{\mathbf{v}}_k, v), & \tau \in \bigcup_i \Delta_i, \\ U_i B_{kj}(v), & \tau \in \bigcup_i (\tau_i^-, \tau_i^+), \end{cases}$$

we represent $\tilde{\mathbf{v}}_k^\gamma(\tau)$ as the Ito process:

$$\tilde{\mathbf{v}}_k^\gamma(\tau) = v_k(0) + \int_0^\tau \hat A_k\big(\tilde{\mathbf{v}}_k^\gamma, v, s\big)\,\mathrm{d}s + \int_0^\tau \hat B_{kj}\big(\tilde{\mathbf{v}}_k^\gamma, v, s\big)\,\mathrm{d}\beta_s^j. \tag{7.5}$$

Letting formally $\frac{|\tilde{\mathbf{v}}_k^\gamma|}{|\mathbf{v}_k|} = 1$ for $|\mathbf{v}_k| = 0$, we make the function $\frac{|\tilde{\mathbf{v}}_k^\gamma|}{|\mathbf{v}_k|} \equiv 1$ along all trajectories.

By the definition of $\hat A_k$ and $\hat B_{kj}$ and by Theorem 2.3 the following bounds hold true with a suitable integer $K$:

$$\begin{aligned} |\hat A_k| &\leqslant C\big(|v|_1^K + 1\big), && \tau \in \bigcup_i \Delta_i, \\ |\hat A_k| &\leqslant C\nu^{-1}\big(|v|_1^K + 1\big), && \tau \in \bigcup_i \big(\tau_i^-, \tau_i^+\big), \\ |\hat B_k|_{h^1} &\leqslant C\big(|v|_1^K + 1\big), && \tau \in [0, \infty) \end{aligned}$$

(cf. Lemma 3.2). Let us fix any $\nu > 0$. The family of processes $\{\mathbf{v}_k^\gamma(\cdot), 0 < \gamma < 1/2\}$ is tight in $C(0, T; \mathbb{R}^2)$. This readily follows from (7.5), Lemma 4.1 and the estimates above.

Since $B_{kj}(v) = \beta_j\,\mathrm{d}\Psi_k(u)e_j$, where $\Psi$ defines diffeomorphisms $H^0 \to h^0$ and $H^1 \to h^1$, then the diffusion $\sum \hat B_{kj}\,\mathrm{d}\beta^j$ in $\mathbb{R}^2$ is non-degenerate and the corresponding diffusion matrix admits lower and upper bounds, uniform if $|v|_1 \leqslant R$ for any fixed $R > 0$.

**Step 2.** Cut-off at a level $|v|_1 = R$.

Let us introduce Markov time $\bar\tau_R = \inf\{\tau \geqslant 0\colon |v(\tau)|_1 \geqslant R\}$. We define the processes $\mathbf{v}_k^R$ equal to $\mathbf{v}_k$ for $\tau \in [0, \bar\tau_R]$ and satisfying the equation:

$$\mathrm{d}\mathbf{v}_k^R(\tau) = \mathrm{d}W_\tau, \quad \tau > \bar\tau_R,$$

where $W_\tau = \binom{\beta_\tau^1}{\beta_\tau^{-1}}$. Also, we define $\tilde{\mathbf{v}}_k^{\gamma,R}$ to be equal to $\tilde{\mathbf{v}}_k^\gamma$ for $\tau \in [0, \bar\tau_R]$ and for $\tau > \bar\tau_R$ satisfying the equation:

$$\mathrm{d}\tilde{\mathbf{v}}_k^{\gamma,R}(\tau) = U\big(\mathbf{v}_k(\bar\tau_R), \tilde{\mathbf{v}}_k^\gamma(\bar\tau_R)\big)\,\mathrm{d}W_\tau, \quad \tau > \bar\tau_R.$$

These processes have positive definite diffusion matrices uniformly in $\gamma$ and $\nu$, and

$$\big|\tilde{\mathbf{v}}_k^{\gamma,R}\big| \equiv \big|\mathbf{v}_k^R\big|.$$

By Lemma 4.1 we have:

$$\begin{aligned}&\mathbf{P}\big\{\tilde{\mathbf{v}}_k^\gamma(\tau) \neq \tilde{\mathbf{v}}_k^{\gamma,R}(\tau) \text{ for some } 0 \leqslant \tau \leqslant T\big\} \to 0,\\ &\mathbf{P}\big\{\big|\mathbf{v}_k(\tau)\big| \neq \big|\mathbf{v}_k^R(\tau)\big| \text{ for some } 0 \leqslant \tau \leqslant T\big\} \to 0,\end{aligned} \tag{7.6}$$

as $R \to \infty$, uniformly in $\gamma$ and $\nu$. Therefore, it suffices to prove the lemma for $\mathbf{v}_k$ replaced by $\mathbf{v}_k^R$ with arbitrary $R$.[6]

**Step 3.** Limit $\gamma \to 0$.

Denote a limiting (as $\gamma \to 0$) law of $\tilde{\mathbf{v}}_k^{\gamma,R}$ in $C(0,T;\mathbb{R}^2)$ by $\tilde{\mathcal{L}}^0$, and let $\hat{\mathbf{v}}_k(\tau)$ be any process such that its law equals $\tilde{\mathcal{L}}^0$. By construction, the relation holds $\tilde{\mathcal{L}}^0\{|\hat{\mathbf{v}}_k(\tau)| \in Q\} = \mathcal{L}\{|\mathbf{v}_k^R(\tau)| \in Q\}$ for any Borel set $Q \subset \mathbb{R}$. So it suffices to prove the lemma's assertion with $\mathbf{v}_k$ replaced by $\hat{\mathbf{v}}_k$.

The process $\tilde{\mathbf{v}}_k^{\gamma,R}$ satisfies the relation:

$$\tilde{\mathbf{v}}_k^{\gamma,R}(\tau) = \mathbf{v}_k^R(0) + \int_0^\tau \hat{A}_{k,R}\big(\tilde{\mathbf{v}}_k^{\gamma,R}, v, s\big)\,\mathrm{d}s + \int_0^\tau \hat{B}_{kj,R}\big(\tilde{\mathbf{v}}_k^{\gamma,R}, v, s\big)\,\mathrm{d}\beta_s^j, \tag{7.7}$$

with

$$\hat{A}_{k,R} = \begin{cases}\hat{A}_k, & s \leqslant \tau_R,\\ 0, & s > \tau_R,\end{cases}$$

and

$$\hat{B}_{kj,R} = \begin{cases}\hat{B}_{kj}, & s \leqslant \tau_R,\\ U\big(\binom{1}{0}\delta_{j,1} + \binom{0}{1}\delta_{-j,1}\big), & s > \tau_R,\end{cases}$$

where $U = U(\mathbf{v}_k(\tau_R), \tilde{\mathbf{v}}_k^\gamma(\tau_R))$.

Denote in (7.7) the drift and martingale parts by $\mathcal{A}^\gamma(\tau)$ and $\mathcal{M}^\gamma(\tau)$, respectively. Then

$$\mathcal{A}^\gamma(\tau) = \int_0^\tau \hat{A}_{k,R}\big(\tilde{\mathbf{v}}_k^{\gamma,R}, v, s\big)\,\mathrm{d}s, \qquad \mathcal{M}^\gamma(\tau) = \int_0^\tau \hat{B}_{kj,R}\big(\tilde{\mathbf{v}}_k^{\gamma,R}, v, s\big)\,\mathrm{d}\beta_s^j.$$

Distributions of the pairs $(\mathcal{A}^\gamma(\cdot), \mathcal{M}^\gamma(\cdot))$ form a tight family of Borel measures in $C(0,T;\mathbb{R}^4)$. Consider a limiting measure and represent it as the distribution of a process $(\mathcal{A}^0(\tau), \mathcal{M}^0(\tau))$. Then $\mathcal{L}\{\mathcal{A}^0(\cdot) + \mathcal{M}^0(\cdot)\} = \tilde{\mathcal{L}}^0$, so we can take for a process $\hat{\mathbf{v}}_k$ above the process $\hat{\mathbf{v}}_k(\tau) = \mathcal{A}^0(\tau) + \mathcal{M}^0(\tau)$. Let $\tau_1$ and $\tau_2$ be arbitrary distinct point of $[0,T]$ and $C_0$—any positive number. The set $\{\varphi \in C(0,T;\mathbb{R}^2)\colon |\varphi(\tau_1) - \varphi(\tau_2)| \leqslant C_0|\tau_1 - \tau_2|\}$ is closed, thus

[6] Indeed, for any $\varepsilon > 0$ choosing first $R$ so big that the probability in (7.6) is $< \varepsilon/2$ and choosing next $\delta = \delta(\varepsilon)$ so small that the left-hand side of (4.5), evaluated for $\mathbf{v}_k$ replaced by $\mathbf{v}_k^R$, also is $< \varepsilon/2$, we see that the left-hand side of (4.5) is $< \varepsilon$, if $\delta$ is sufficiently small.

$$\limsup_{\gamma\to 0}\mathbf{P}\bigl\{\bigl|\mathcal{A}^\gamma(\tau_1)-\mathcal{A}^\gamma(\tau_2)\bigr|\leqslant C_0|\tau_1-\tau_2|\bigr\}$$
$$\leqslant \mathbf{P}\bigl\{\bigl|\mathcal{A}^0(\tau_1)-\mathcal{A}^0(\tau_2)\bigr|\leqslant C_0|\tau_1-\tau_2|\bigr\}. \tag{7.8}$$

Let us choose $C_0 = 2\sup\{|A_k(v)|:\ |v|_1\leqslant R\}$. Then

$$\bigl|A^\gamma(\tau_1)-A^\gamma(\tau_2)\bigr|\leqslant \frac{1}{2}C_0|\tau_1-\tau_2|+\nu^{-1}C(R)\Bigl|\Bigl(\bigcup\Lambda_j\Bigr)\cap[0,T]\Bigr|.$$

Since

$$\mathbf{E}\Bigl|\Bigl(\bigcup\Lambda_j\Bigr)\cap[0,T]\Bigr|\leqslant \mathbf{P}\Bigl\{\sup_{[0,T]}\bigl|v(\tau)\bigr|_{h^1}\geqslant\gamma^{-1}\Bigr\}+\mathbf{E}\int_0^T\chi_{|v_k(\tau)|\leqslant 2\gamma}\,\mathrm{d}\tau,$$

then it follows from (4.1) and Theorem 2.2.4 in [13] that $\mathbf{E}|(\bigcup\Lambda_j)\cap[0,T]|\to 0$ as $\gamma\to 0$. Therefore the limit in the left-hand side of (7.8) equals 1, and we conclude that $\mathbf{P}\{|\mathcal{A}^0(\tau_1)-\mathcal{A}^0(\tau_2)|\leqslant C_0|\tau_1-\tau_2|\}=1$. That is, $\mathcal{A}^0(\tau)$ is $C_0$-Lipschitz continuous and $\mathcal{A}^0(\tau)=\int_0^\tau B^0(s)\,\mathrm{d}s$, where $|B_0|\leqslant C_0$.

We now turn to the martingale part. Since

$$[0,T]\ni\tau\to\mathcal{M}^\gamma(\tau)\in\mathbb{R}^2,\quad 0<\gamma\leqslant 1,$$

is a family of continuous square integrable martingales with respect to the natural filtration and uniformly bounded second moments, then the limiting process $\mathcal{M}^0(\tau)$ is a continuous square integrable martingale as well. Denote $\langle\langle\mathcal{M}^\gamma\rangle\rangle_\tau$ the bracket (quadratic characteristics) of $\mathcal{M}^\gamma$. According to Corollary VI.6.7 in [8], $\langle\langle\mathcal{M}^0\rangle\rangle_\tau=\lim_{\gamma\to 0}\langle\langle\mathcal{M}^\gamma\rangle\rangle_\tau$. Since for $v\in\{v:\ |v|_{h^1}\leqslant R\}$ it holds,

$$c_1(\tau_1-\tau_2)|\xi|^2\leqslant\bigl(\bigl(\langle\!\langle\mathcal{M}^\gamma\rangle\!\rangle_{\tau_1}-\langle\!\langle\mathcal{M}^\gamma\rangle\!\rangle_{\tau_2}\bigr)\xi,\xi\bigr)\leqslant c_1^{-1}(\tau_1-\tau_2)|\xi|^2\quad\forall\xi\in\mathbb{R}^2,$$

with some $c_1>0$, then the bracket $\langle\langle\mathcal{M}^0\rangle\rangle$ satisfies the same estimate. In particular, $\mathrm{d}\langle\langle\mathcal{M}^0\rangle\rangle_\tau=a(\tau)\,\mathrm{d}\tau$ for some progressively measurable symmetric $2\times 2$-matrix $a(\tau)$ such that $c_1\,\mathrm{Id}\leqslant a(\tau)\leqslant c_1^{-1}\,\mathrm{Id}$, a.s. Then $W_\tau=\int_0^\tau a^{-1/2}(s)\,\mathrm{d}\mathcal{M}^0(s)$ is a Wiener process in $\mathbb{R}^2$ and $\mathcal{M}^0(\tau)=\int_0^\tau a^{1/2}(s)\,\mathrm{d}W_s$.

We have seen that for any $\nu>0$ and $R\geqslant 1$ each weak limit of the family $\tilde{\mathbf{v}}_k^{\gamma,R}(\tau)$ is an Ito process of the form:

$$\hat{\mathbf{v}}_k(\tau)=\hat{\mathbf{v}}_k(0)+\int_0^\tau B^0(s)\,\mathrm{d}s+\int_0^\tau a^{1/2}(s)\,\mathrm{d}W_s,$$

where $|B^0(\tau)|\leqslant C_0$ and $c_1^{1/2}\,\mathrm{Id}\leqslant a^{1/2}(\tau)\leqslant c_1^{-1/2}\,\mathrm{Id}$ a.s., uniformly in $t$ and $\nu$. Since all the coefficients of this equation are uniformly bounded and the diffusion matrix is positive definite, the desired statement follows from Theorem 2.2.4 in [13].

## Appendix A

Here we prove the a priori estimates, claimed in Section 1.

Let $F:H^m\to\mathbb{R}$ be a smooth functional (for some $m\geqslant 0$). Applying formally Ito's formula to $F(u(t))$, where $u(t)$ is a solution, and taking the expectation we get:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{E}F\bigl(u(t)\bigr)=\mathbf{E}\bigl\langle\nabla F(u),\nu u_{xx}-V(u)\bigr\rangle+\frac{1}{2}\nu\sum_s b_s^2\mathbf{E}\,\mathrm{d}^2F(u)[e_s,e_s].$$

In particular, if $F(u)$ is an integral of motion for the KdV equation, then $\langle\nabla F(u),V(u)\rangle=0$ and we have:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{E}F\bigl(u(t)\bigr)=\nu\mathbf{E}\bigl\langle\nabla F(u),u_{xx}\bigr\rangle+\frac{1}{2}\nu\sum_s b_s^2\mathbf{E}\,\mathrm{d}^2F(u)[e_s,e_s]. \tag{A.1}$$

Since $\|u\|_0^2$ is an integral of motion, then $F(u)=\exp(\sigma\|u\|_0^2)$, $0<\sigma\leqslant\frac{1}{2}$, also is an integral. We have:

$$\nabla F(u)=2\sigma e^{\sigma\|u\|_0^2}u,\quad \mathrm{d}^2F(u)[e,e]=2\sigma e^{\sigma\|u\|_0^2}\|e\|_0^2+4\sigma^2e^{\sigma\|u\|_0^2}\langle u,e\rangle^2.$$

So (A.1) implies that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{E}e^{\sigma\|u\|_0^2} = -\nu\sigma\mathbf{E}\Big(e^{\sigma\|u\|_0^2}\Big(2\|u\|_1^2 - B_0 - 2\sigma\sum b_s^2u_s^2\Big)\Big),$$

where for $r \geqslant 0$ we set:

$$B_r = \sum j^{2r}b_s^2.$$

Denoting $\hat{B} = \max b_s^2$ and choosing $\sigma \leqslant (2\hat{B})^{-1}$ we get that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{E}e^{\sigma\|u\|_0^2} \leqslant -\nu\sigma\mathbf{E}\big(e^{\sigma\|u\|_0^2}\big(\|u\|_1^2 - B_0\big)\big) \leqslant -\nu\sigma B_0\big(\mathbf{E}e^{\sigma\|u\|_0^2} - 2e^{2\sigma B_0}\big).$$

So the estimate (1.4) holds for all $t \geqslant 0$. In particular, for each $N > 0$ we have:

$$\mathbf{E}\big\|u(t)\big\|_0^N \leqslant M_N = C\sigma^{-N/2}\cdot\big\langle\text{the r.h.s. of (1.4)}\big\rangle. \tag{A.2}$$

The KdV equation has infinitely many integrals of motion $J_m(u)$, $m \geqslant 0$, which can be written as

$$J_m(u) = \|u\|_m^2 + \sum_{r=3}^{m}\sum_{\mathbf{m}}\int C_{r,\mathbf{m}}u^{(m_1)}\dots u^{(m_r)}\,\mathrm{d}x. \tag{A.3}$$

Here the inner sum is taken over all integer $r$-vectors $\mathbf{m} = (m_1,\dots,m_r)$ such that $0 \leqslant m_j \leqslant m-1$ $\forall j$ and $m_1+\dots+m_r = 4+2m-2r$ (in particular, $J_0 = \|u\|_0^2$). E.g., see [11, p. 209].

Let us consider an integral as in (A.3),

$$I = \int u^{(m_1)}\dots u^{(m_f)}\,\mathrm{d}x, \quad m_1+\dots+m_f = M,$$

where $f \geqslant 2$, $M \geqslant 1$ and $0 \leqslant m_j \leqslant \mu-1$, $\Theta := \mu^{-1}(M+f/2-1) < 2$ for some $\mu \geqslant 2$. Then, by Hölder's inequality,

$$|I| \leqslant \big|u^{(m_1)}\big|_{L_{p_1}}\dots\big|u^{(m_f)}\big|_{L_{p_f}}, \quad p_j = \frac{M}{m_j} \leqslant \infty.$$

Applying next the Gagliardo–Nirenberg inequality we find that

$$|I| \leqslant C\|u\|_\mu^\Theta\|u\|_0^{f-\Theta}. \tag{A.4}$$

Finally, evoking the Young inequality we get that

$$|I| \leqslant \delta\|u\|_\mu^2 + C_\delta\|u\|_0^{2\frac{f-\Theta}{2-\Theta}} \quad \forall\delta > 0. \tag{A.5}$$

We have:

$$I_1 := \big\langle\nabla J_m(u), u_{xx}\big\rangle = -2\|u\|_{m+1}^2 + \sum_{r=3}^{m+2}\sum_{\mathbf{m}}C'_{\mathbf{m}}u^{(m_1)}\dots u^{(m_r)}\,\mathrm{d}x,$$

where $m_1+\dots+m_r = 6+2m-2r$. Due to (A.5) with $\delta = 1/2$, $f = r$ and $\mu = m+1$,

$$I_1 \leqslant -\frac{3}{2}\|u\|_{m+1}^2 + C\|u\|_0^{2\frac{r-\Theta}{2-\Theta}} \leqslant -\frac{3}{2}\|u\|_{m+1}^2 + C\big(1+\|u\|_0^{4(m+1)}\big).$$

Next,

$$\mathrm{d}^2J_m(u)[v,v] = 2\|v\|_m^2 + \sum_r\sum_{\mathbf{m}}\int C''_{\mathbf{m}}v^{(m_1)}v^{(m_2)}u^{(m_3)}\dots u^{(m_r)}\,\mathrm{d}x.$$

Hence,

$$I_2 := \mathrm{d}^2J_m(u)[e_j,e_j] \leqslant 2j^{2m} + |e_j|_{C^{m_1}}|e_j|_{C^{m_2}}\sum_{r,\mathbf{m}}\int\hat{C}_{\mathbf{m}}\big|u^{(m_1)}\big|\dots\big|u^{(m_{\hat{r}})}\big|\,\mathrm{d}x,$$

where $\hat r = r-2$ and $m_1+\cdots+m_{\hat r} = 4+2m-2r-m_1-m_2 =: \hat M$, $\hat M \geqslant 0$. Note that $|e_j|_{C^n} = j^n$ for each $j$ and $n$. Assume first that $r \geqslant 4$ and $\hat M > 0$. Then (A.4) implies that

$$I_2 \leqslant 2j^{2m} + C\|u\|_{m+1}^{\Theta} j^{m_1+m_2}\|u\|_0^{r-2-\Theta},$$

with $\Theta = 2 - \frac{(3/2)r+m_1+m_2}{m+1}$.

By the Young inequality,

$$\begin{aligned} I_2 &\leqslant 2j^{2m} + \delta\|u\|_{m+1}^2 + C_\delta\big(j^{m_1+m_2}\|u\|_0^{r-2-\Theta}\big)^{\frac{2}{2-\Theta}} \\ &\leqslant 2j^{2m} + \delta\|u\|_{m+1}^2 + C_\delta j^{2(m+1)}\big(1+\|u\|_0^{\frac{4}{3}(m+1)}\big). \end{aligned}$$

It is easy to see that this estimate also holds for $r=4$ and for $\hat M = 0$.

Using in (A.1) with $F = J_m$ the obtained bounds for $I_1$ and $I_2$ we get that

$$\begin{aligned} \frac{\mathrm d}{\mathrm dt}\mathbf E J_m(u) \leqslant{}& -\frac{3}{2}\nu\mathbf E\|u\|_{m+1}^2 + C_1\nu\mathbf E\big(1+\|u\|_0^{4(m+1)}\big) + \nu\sum|s|^{2m}b_s^2 \\ &+ \frac12\delta\nu\mathbf E\|u\|_{m+1}^2\sum b_s^2 + \frac12 C_\delta\nu\sum b_s^2 s^{2(m+1)}\big(1+\mathbf E\|u\|_0^{\frac43(m+1)}\big). \end{aligned}$$

Choosing $\delta = B_0^{-1}$ and using (A.2) we arrive at the estimate:

$$\frac{\mathrm d}{\mathrm dt}\mathbf E J_m(u) \leqslant -\nu\mathbf E\|u\|_{m+1}^2 + \nu C_m,$$

where $C_m$ depends on $B_{m+1}$ and $M_{4(m+1)}$.

Applying (A.5) with $\mu = m$ to (A.3) we see that

$$\frac12\|u\|_m^2 - C\big(1+\|u\|_0^{4m}\big) \leqslant J_m(u) \leqslant 2\|u\|_m^2 + C\big(1+\|u\|_0^{4m}\big). \tag{A.6}$$

Therefore

$$\frac{\mathrm d}{\mathrm dt}\mathbf E J_m(u) \leqslant -\frac{\nu}{2}\big(\mathbf E J_m(u) - C_m'\big),$$

where $C_m'$ depends on the same quantities as $C_m$. We get that

$$\mathbf E J_m\big(u(t)\big) \leqslant \max\big(\mathbf E J_m\big(u(0)\big), C_m'\big)$$

for each $t \geqslant 0$. Using (A.6) we obtain (1.5).

Let us take any integers $m \geqslant 0$, $k \geqslant 1$. By the interpolation inequality $\|u\|_m^k \leqslant \|u\|_{mk}\|u\|_0^{k-1}$. Therefore

$$\mathbf E\|u\|_m^k \leqslant \big(\mathbf E\|u\|_{mk}^2\big)^{1/2}\big(\mathbf E\|u\|_0^{2(k-1)}\big)^{1/2}.$$

Using this inequality jointly with (A.2) and (1.5) we get the estimate (1.6).

## References

[1] V. Arnold, V.V. Kozlov, A.I. Neistadt, Mathematical Aspects of Classical and Celestial Mechanics, Springer, Berlin, 1989.
[2] P. Djakov, B. Mityagin, Instability zones of 1D periodic Schrödinger and Dirac operators, Russ. Math. Surv. 61 (4) (2006).
[3] B.A. Dubrovin, S.P. Novikov, Hydrodynamics of weakly deformed soliton lattices, differential geometry and Hamiltonian theory, Russ. Math. Surv. 44 (1989) 35–124.
[4] G. Da Prato, J. Zabczyk, Stochastic Equations in Infinite Dimensions, Cambridge University Press, Cambridge, 1992.
[5] H. Flaschka, M.G. Forest, D.W. McLaughlin, Multiphase averaging and the inverse spectral solutions of the Korteweg–de Vries equation, Comm. Pure Appl. Math. 33 (1980) 739–784.
[6] M. Freidlin, A. Wentzell, Random Perturbations of Dynamical Systems, second ed., Springer-Verlag, New York, 1998.
[7] M.I. Freidlin, A.D. Wentzell, Averaging principle for stochastic perturbations of multifrequency systems, Stochastics and Dynamics 3 (2003) 393–408.
[8] J. Jacod, A.N. Shiryaev, Limit Theorems for Stochastic Processes, Springer-Verlag, Berlin, 1987.
[9] R.Z. Khasminskii, On the averaging principle for Ito stochastic differential equations, Kybernetika 4 (1968) 260–279 (in Russian).
[10] E. Korotyaev, Estimates for the Hill operator, II, J. Diff. Equations 223 (2) (2006) 229–260.

[11] T. Kappeler, J. Pöschel, KAM & KdV, Springer, 2003.
[12] I.M. Krichever, The averaging method for two-dimensional "integrable" equations, Funct. Anal. Appl. 22 (1988) 200–213.
[13] N.V. Krylov, Controlled Diffusion Processes, Springer, 1980.
[14] S.B. Kuksin, Analysis of Hamiltonian PDEs, Oxford University Press, Oxford, 2000.
[15] S.B. Kuksin, Randomly forced nonlinear PDEs and statistical hydrodynamics in 2 space dimensions, European Mathematical Society Publishing House, 2006, also see mp_arc 06-178.
[16] J. Moser, C.L. Siegel, Lectures on Celestial Mechanics, Springer, Berlin, 1971.
[17] H. McKean, E. Trubowitz, Hill's operator and hyperelliptic function theory in the presence of infinitely many branching points, Comm. Pure Appl. Math. 29 (1976) 143–226.
[18] A.Yu. Veretennikov, On the averaging principle for systems of stochastic differential equations, Mat. USSR Sb. 69 (1991) 271–284.
[19] G.B. Whitham, Linear and Nonlinear Waves, John Wiley & Sons, New York, 1974.
[20] M. Yor, Existence et unicité de diffusion a valeurs dans un espace de Hilbert, Ann. Inst. Henri Poincaré 10 (1974) 55–88.