Probab. Theory Relat. Fields 129, 157–181 (2004)
Digital Object Identifier (DOI) 10.1007/s00440-003-0326-7

Victor Bakhtin · Yuri Kifer

Diffusion approximation for slow motion in fully coupled averaging

Received: 10 May 2003 / Revised version: 24 October 2003 /
Published online: 3 March 2004 –

Abstract. In systems which combine fast and slow motions it is usually impossible to study directly corresponding two scale equations and the averaging principle suggests to approximate the slow motion by averaging in fast variables. We consider the averaging setup when both fast and slow motions are diffusion processes depending on each other (fully coupled) and show that there exists a diffusion process which approximates the slow motion in the L^2 sense much better than the averaged motion prescribed by the averaging principle.

1. Introduction

Consider a system of stochastic differential equations

$$\begin{cases} dx^\varepsilon(t) = \varepsilon v(x^\varepsilon(t), y^\varepsilon(t), \varepsilon)\,dt + \varepsilon \sum_{i=1}^m u_i(x^\varepsilon(t), y^\varepsilon(t), \varepsilon) \circ dw^i(t), \\ dy^\varepsilon(t) = b(x^\varepsilon(t), y^\varepsilon(t), \varepsilon)\,dt + \sum_{i=1}^m a_i(x^\varepsilon(t), y^\varepsilon(t), \varepsilon) \circ dw^i(t) \end{cases} \tag{1.1}$$

in the Stratonovich form defined on the Cartesian product of two Riemannian manifolds $X \times Y$ where $m \geq \dim Y$, v, u_i, b, a_i are smooth vector fields on $X \times Y$, and $w(t) = (w^1(t), \ldots, w^m(t))$ is a standard Brownian motion (see [14]). Such two scale equations emerge naturally when we study, first, an idealized system described by a family of stochastic differential equations

$$dy_x(t) = b(x, y_x(t))\,dt + \sum_{i=1}^m a_i(x, y_x(t)) \circ dw^i(t) \tag{1.2}$$

V. Bakhtin: Department of Physics, Belarusian State University, Fr. Scoriny 4, Minsk 220050, Belarus. e-mail: `bakhtin@bsu.by`

Y. Kifer: Institute of Mathematics, The Hebrew University, Jerusalem 91904, Israel. e-mail: `kifer@math.huji.ac.il`

The authors are partially supported by INTAS, project No. 99-00559 and by US-Israel BSF, respectively. Part of the work was done during the visit of the 1st author to the Hebrew University.

Mathematics Subject Classification (2000): Primary 34C29; Secondary 60F15, 58J65

Key words or phrases: Averaging – Diffusion – Limit theorems – Stochastic differential equations

on Y depending on a parameter x preserved by the system where $b(x, y) = b(x, y, 0)$ and $a_i(x, y) = a_i(x, y, 0)$. Then the real system viewed as a perturbation of the idealized one should be described by perturbed equations in the form (1.1) which exhibit mutually dependent slow $x^\varepsilon(t)$ and fast $y^\varepsilon(t)$ motions. Such problems (though without a stochastic term) were first encountered already in 18th century in celestial mechanics in the study of perturbations of planetary motion. It was observed later and justified only by heuristic arguments that by averaging in fast variables one obtains a much simpler averaged equation which often gives a good approximation of the slow motion on long time intervals.

In the fully coupled situation as above, i. e. when both slow and fast motions depend on each other the justification of this averaging principle is not easy. Assuming compactness of Y and nondegeneracy of the diffusion terms in (1.2) each process y_x has a unique invariant measure μ_x on Y with a smooth density with respect to the Riemannian volume which depends smoothly on x, as well. Then

$$\bar{B}(x) = \int_Y \left(v(x, y, 0) + \frac{1}{2} \sum_{i=1}^{m} \frac{du_i(x, y, 0)}{dy} a_i(x, y, 0) \right) d\mu_x(y) \tag{1.3}$$

is a smooth vector field and we can consider the averaged ordinary differential equation for the process $z^\varepsilon(t) = x^\varepsilon(t/\varepsilon)$ having the form

$$\frac{d\bar{z}(t)}{dt} = \bar{B}(\bar{z}(t)). \tag{1.4}$$

It follows, essentially, from [1] and [3] that

$$\lim_{\varepsilon \to 0} \mathbf{E} d_X^2(z^\varepsilon(t), \bar{z}(t)) = 0 \tag{1.5}$$

provided $z^\varepsilon(0) = \bar{z}(0)$, where d_X is the distance on X, and, in general, this expectation is of order ε.

Relying on some physical intuition Hasselmann [12] suggested to approximate $z^\varepsilon(t) = x^\varepsilon(t/\varepsilon)$ (in the case when $u_i \equiv 0$) by a diffusion process $r^\varepsilon(t)$ on X solving a stochastic differential equation in the Itô form

$$dr^\varepsilon(t) = \bar{B}(r^\varepsilon(t))\, dt + \varepsilon\eta(r^\varepsilon(t))\, dt + \sqrt{\varepsilon}\sigma(r^\varepsilon(t))\, d\tilde{w}(t), \tag{1.6}$$

where $\eta = (\eta^1, \dots, \eta^n)$, $\sigma = \left(\sigma_j^i\right)$, $i, j = 1, \dots, n$, $\tilde{w}$ is an n-dimensional Brownian motion and $n = \dim X$. Hasselmann wrote (1.6) without the drift η but since we consider this equation on a manifold X and write it in the Itô form in local coordinates this term comes, in general, into the picture, as well (see the corresponding discussion in Section 2). We will justify this approximation showing that for each $\varepsilon > 0$ and an initial condition $x = z^\varepsilon(0) = r^\varepsilon(0)$ a Brownian motion driving this equation can be chosen (on may be a richer probability space) in such a way that

$$\mathbf{E} \sup_{0 \le t \le T} d_X^2(z^\varepsilon(t), r^\varepsilon(t)) \le C\varepsilon^{1+\delta} \tag{1.7}$$

for some $\delta > 0$. This result can be considered as a strong diffusion approximation of the slow motion in averaging which has a global manifold invariant form. This result implies also a weak Gaussian approximation of the normalized error in averaging, i.e. that $\varepsilon^{-1/2}(z^\varepsilon(t) - \bar{z}(t))$ converges weakly to a Gaussian process $g(t)$ (see Corollary 2.2). In the much simpler case of a fast motion independent of the slow one this result was obtained long ago in [15]. Observe that this Gaussian approximation makes sense only in $\mathbb{R}^n$ or in local coordinates and, as any weak convergence result, it cannot truly justify Hasselmann's nonlinear diffusion approximation of $z^\varepsilon(t)$ by $r^\varepsilon(t)$. One may argue that relying on the above weak Gaussian approximation and the Skorokhod representation theorem (see, for instance, [4]) it follows from comparison between r^ε and $\bar{z} + \sqrt{\varepsilon} g$ that after a redefinition on an appropriate probability space $\varepsilon^{-1/2} \sup_{0 \le t \le T} d_X(z^\varepsilon(t), r^\varepsilon(t))$ tends to zero in probability. However, this approach cannot give, in principle, any estimates of the speed of this convergence and it requires a huge (product) probability space which hinders any resemblance of the redefined processes with the original diffusions x^ε and y^ε while our approach provides an explicit redefinition which does not change essentially the structure of these diffusions. Moreover, in our fully coupled situation it is not easy to derive the weak Gaussian approximation, as well. It is clear that δ in (1.7) cannot, in general, exceed 1 and we show that (1.7) holds true with any $\delta < (18 + 8n)^{-1}$ and C depending on δ. Still, (1.7) may hold true with larger δ and it is an interesting open problem to find the optimal bound there.

A similar result was proved in [17] for the case when the fast motion is a sufficiently fast mixing stationary process ξ_t which does not depend on the slow motion so that (1.1) is replaced by one equation $dx^\varepsilon(t)/dt = \varepsilon v(x^\varepsilon(t), \xi_t)$. Though ξ_t here does not have to be necessarily a diffusion this case has been treated by more traditional methods than the perturbations machinery employed in the present paper. Using the technique from [1]–[2] it is possible to extend our results to the setup of fully coupled averaging in difference equations with fast motions being either Markov chains satisfying Doeblin type conditions or Axiom A diffeomorphisms considered in a neighborhood of an attractor.

2. Main results

We will consider the stochastic differential equations (1.1) on a product $X \times Y$ of two Riemannian manifolds, where Y is compact and connected. The small parameter ε in (1.1) will run over an interval $I = [-\varepsilon_0, \varepsilon_0]$. We assume that the vector fields v, u_i, b, a_i are of smoothness class $C^{N+1}(X \times Y \times I)$ with $N \ge 6$ and, furthermore, for any x, y the family of vectors $\{a_i(x, y, 0)\}_{i=1}^m$ spans the tangent space $T_y Y$. In addition, to simplify proofs, we assume that the vector fields v, u_i vanish outside some compact subset $X_0 \subset X$. At the beginning of Section 4 we shall explain how this assumption can be relaxed and replaced by the condition that the Euclid norms $|v|$, $|u_i|$ grow at most lineary with respect to the distance on X. The main result of this paper is the following

Theorem 2.1. *Without changing its distribution, for each $\varepsilon > 0$ and any initial condition $x = x^\varepsilon(0)$ we can redefine the Brownian motion $w(t)$ from* (1.1) *on a* *richer probability space where there exists a family of diffusion processes* $r^\varepsilon(t)$ *on* X *solving* (1.6) *such that* $z^\varepsilon(t) = x^\varepsilon(t/\varepsilon)$ *and* $r^\varepsilon(t)$ *(with* $r^\varepsilon(0) = x$*) satisfy the inequality* (1.7) *with any* $\delta < (18 + 8n)^{-1}$ *while* $C = C(\delta, T)$ *in* (1.7) *depends on* δ *and* T*. The new probability space* $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ *can be taken in the product form* $([0, 1], \mathcal{B}, Leb) \times (\Omega, \mathcal{F}, P)$*, where* $\mathcal{B}$ *is the Borel* σ*-algebra and* Leb *is the Lebesgue measure on* $[0, 1]$*, redefining* $w^i(t)$ *on* $\tilde{\Omega}$ *by* $w^i(t, (u, \omega)) = w^i(t, \omega)$*,* $u \in [0, 1]$*,* $\omega \in \Omega$*. The diffusion matrix* $\sigma = \left(\sigma^i_j\right)$ *in* (1.6) *satisfies* $\sigma(x)\sigma^*(x) = a(x)$ *where* $a(x)$ *is a* C^N *symmetric nonnegatively definite matrix function defined in Proposition 3.13 of the next section.*

The matrix function $a(x)$ is C^N but, in general, it is only nonnegatively definite so we can only be sure that $\sigma(x)$ is Lipschitz (see, for instance, [13], Section 1.3) which does not enable us to write (1.6) in the Stratonovich form which is used usually when dealing with diffusions on manifolds. Still, it is easy to see from the formula for $a(x) = (a^{ij}(x))$ that if it is written in local coordinates $x = (x^1, \ldots, x^n)$ and $\tilde{a}(\tilde{x}) = \left(\tilde{a}^{ij}(\tilde{x})\right)$ is its expression in another set of local coordinates $\tilde{x} = (\tilde{x}^1, \ldots, \tilde{x}^n)$ at the same point then

$$\tilde{a}^{kl}(\tilde{x}) = \sum_{i,j=1}^{n} a^{ij}(x) \frac{\partial \tilde{x}^k}{\partial x^i} \frac{\partial \tilde{x}^l}{\partial x^j}, \tag{2.1}$$

i. e. $a(x)$ is a $(2, 0)$ tensor field. Next, we observe that there exists a (not unique) second order elliptic differential operator L on the manifold X with a prescribed symbol, i. e. coefficients in second derivatives $a(x) = a^{ij}(x)$ provided (2.1) holds true. Define, for instance, L in local coordinats $(x^1, \ldots, x^n)$ by the formula

$$L = \frac{1}{2} \frac{1}{\sqrt{g(x)}} \sum_{i,j=1}^{n} \frac{\partial}{\partial x^i} \left(a^{ij}(x) \sqrt{g(x)} \frac{\partial}{\partial x^j} \right), \tag{2.2}$$

where $\sqrt{g(x)}$ is the density of the Riemannian volume, i. e. $dV(x) = \sqrt{g(x)}\, dx^1 \cdots dx^n$. This is a (weakly) elliptic operator which is self-adjoint with respect to the Riemannian volume and its coefficients obey (as it is easy to check) the correct change of coordinates transformation rules, i. e. L is indeed a differential operator on the manifold. Since

$$\bar{B} = \sum_{i=1}^{n} \bar{B}^i(x) \frac{\partial}{\partial x^i}$$

is a vector field then $L^\varepsilon = \varepsilon L + \bar{B}$ is again an elliptic 2-nd order differential operator on the manifold X. Next, we can proceed as in Section 1.3 of [13] in order to construct a diffusion r^ε which solves a stochastic differential equation of the form (1.6) and has the generator L^ε implying that r^ε is now well defined on the manifold X. Namely, relying on the Whitney embedding theorem embed smoothly X as a closed submanifold into a Euclidean space $\mathbb{R}^k$ of sufficiently high dimension k. As in [13] extend the operator L^ε into a 2-nd order elliptic operator $\tilde{L}^\varepsilon$ with C^N coefficients on the whole $\mathbb{R}^k$. This operator serves as a generator of a diffusion on $\mathbb{R}^k$

solving a stochastic differential equation similar to (1.6) with Lipschitz coefficients and its restriction to X yields the required diffusion $r^\varepsilon(t)$ which solves, in fact, the martingale problem for $f(r^\varepsilon(t)) - f(r^\varepsilon(0)) - \int_0^t L^\varepsilon f(r^\varepsilon(s))\,ds$.

The proof of Theorem 2.1 consists of two main parts. The first one is based on the perturbations machinery and it studies the asymptotic bahaviour as $\varepsilon \to 0$ of characteristic functions of $z^\varepsilon(t)$ (considered, say, in local coordinates). The second part presented in Section 4 below is based on the technique developped in [5], [8], [19] which yields random variables close in probability or in average as soon, as corresponding (conditional) characteristic functions are sufficiently close in a certain sense. This will enable us to construct a Gaussian process on a richer probability space which is sufficiently close in the L^2-sense to both z^ε and r^ε after the latter processes are properly redefined there. It is well known that this type of strong limit theorems cannot be derived without appropriate redefinitions of processes in question since this is impossible even in the central limit theorem setup (see [4], [21], [20] and the discussion in the next paragraph).

Observe, that the Brownian motion $\tilde{w}(t)$ in (1.6) should be chosen separately for each ε and it is impossible to have (1.7) with w in (1.1) and $\tilde{w}$ in (1.6) the same for all ε. Indeed, consider the simple case when b and a_i in (1.1) do not depend on the slow variable x and on ε so that $y(t)$ is a nondegenerate diffusion on a compact manifold Y. Let also $u_i \equiv 0$ and $v(x, y, \varepsilon)$ depends only on the variable y. Then

$$\varepsilon^{-1/2}\big(z^\varepsilon(t) - z^\varepsilon(0)\big) = \varepsilon^{1/2} \int_0^{t/\varepsilon} v(y(s))\,ds. \tag{2.3}$$

Assume that $\bar{B} \equiv 0$. By Strassen's type invariance principle for the law of the iterated logarithm (see, for instance, [21]) it follows that with probability one the right hand side of (2.3) diverges. If the right hand side of (2.3) would converge in the L^2-sense as $\varepsilon \to 0$ to a random variable then the latter should be measurable with respect to the tail σ-algebra of the diffusion y which is strongly mixing, and so this σ-algebra is trivial. So the limit could only be a constant and, in fact, zero since $\bar{B} \equiv 0$. But it is easy to see that, in general, the variance of the right hand side of (2.3) does not tend to zero as $\varepsilon \to 0$. Thus, the right hand side of (2.3) does not converge in the L^2-sense anywhere, in particular, it does not converge to a diffusion.

Hasselmann suggested that the diffusion (1.6) could describe the long time behaviour of the slow motion where large deviations effects should be taken into account, for instance, in the study of rare transitions of the slow motion between attractors of the averaged one. It turns out that this does not hold true and it is not difficult to see that the rate functionals describing large deviations for the slow motion x^ε and for the diffusion r^ε are usually different. For instance, if again $u_i \equiv 0$ then velocity vectors of paths of the process $z^\varepsilon(t)$ must belong to vector fields $v(\cdot, y, \varepsilon)$ while the diffusion r^ε can move rather arbitrarily with some positive probability. In order to clarify this point consider the following simple one dimensional example. Let $u_i \equiv 0$ and $v(x, y) = c(x) + \delta \sin\varphi$ where $c(x) = 2x - 4x^3$ and the fast variable $y = e^{i\varphi}$ belongs to the unit circle S^1. As the fast motion $y(t)$ we take the Brownian motion on S^1. Then $\bar{B}(x) = c(x)$ and the averaged motion has two attracting fixed points $x_- = -2^{-1/2}$, $x_+ = 2^{-1/2}$ and one repelling fixed point $x_0 = 0$. It is easy to see that if $\delta > 0$ is small enough then the slow motion $z^\varepsilon(t)$ cannot pass at all between small neighborhoods of x_- and x_+ though (as a direct computation shows) the corresponding diffusion approximation $r^\varepsilon(t)$ is nondegenerate and with probability one it exhibits rare transitions between these neighborhoods described by the corresponding rate functional (see [10]).

Suppose that $X = \mathbb{R}^n$ and consider a Gaussian process $g(t)$ on $\mathbb{R}^n$ solving the following linear stochastic differential equation

$$dg(t) = \nabla \bar{B}(\bar{z}(t))g(t)\,dt + \sigma(\bar{z}(t))\,d\tilde{w}(t) \tag{2.4}$$

where $\bar{z}(0) = x$ and $\nabla \bar{B}$ is the matrix whose $i, j-$th element is $\partial \bar{B}^i(x)/\partial x^j$. Standard estimates at the end of Section 4 below yield that

$$\mathbf{E} \sup_{0 \le t \le T} \left\| r^\varepsilon(t) - \bar{z}(t) - \sqrt{\varepsilon} g(t) \right\|^2 \le C\varepsilon^2 \tag{2.5}$$

for some $C > 0$ provided $r^\varepsilon(0) = x$ and $g(0) = 0$, and so we arrive at the following result.

Corollary 2.2. *Suppose $X = \mathbb{R}^n$ in Theorem 2.1. Then we have also that*

$$\mathbf{E} \sup_{0 \le t \le T} \left\| z^\varepsilon(t) - \bar{z}(t) - \sqrt{\varepsilon} g(t) \right\|^2 \le C\varepsilon^{1+\delta} \tag{2.6}$$

for some $C > 0$ where δ is the same as in Theorem 2.1 and $z^\varepsilon(0) = \bar{z}(0) = x$. In particular, the process $\varepsilon^{-1/2}(z^\varepsilon(t) - \bar{z}(t))$, $t \in [0, T]$, converges in the weak sense as $\varepsilon \to 0$ to the Gaussian process $g(t)$, $t \in [0, T]$.

3. Semigroup perturbations machinery

The arguments of this section are quite technical by their nature, and so for readers' sake we start with a short overview of our goals here. The main result of this section is Corollary 3.9 providing Gaussian type asymptotics for the characteristic functions of certain functionals of the processes x^ε and y^ε which will be used in the next section in order to construct the required diffusion approximation of $z^\varepsilon(t) = x^\varepsilon(t/\varepsilon)$. The parameters emerging in Corollary 3.9 are further specified in Propositions 3.13 and 3.14. In order to pass from the global setup to local coordinats along the averaged motion we will need also estimates of probabilities of large deviations of the slow motion from the averaged one obtained in Proposition 3.12. Both Corollary 3.9 and Proposition 3.12 follow from the crucial Proposition 3.8 which provides a formula for certain exponential functionals of the processes x^ε and y^ε. The whole proof is based on Propositions 3.1 and 3.3 which are rather standard and whose detailed proof can be found in [2]. These statements deal with the unperturbed case $\varepsilon = 0$. The first one is an exponential ergodicity type result which follows from our nondegeneracy assumption on the diffusion coefficients of the fast motion y^ε. The second one is a result of the type of the Perron-Frobenius theorem for positive operators and it employs also a version of the implicit function theorem to ensure a nice dependence of the corresponding quantities on parameters. In Propositions 3.4–3.6 we continue to elaborate on the quantities obtained in Proposition 3.3 for the unperturbed case after which the perturbation technique of Proposition 3.7 leads to the desired result of Proposition 3.8.

Let us fix some $T > 0$ and a solution $\{\bar{z}(t) \mid 0 \le t \le T\}$ to the equation (1.4). For this averaged path there exists a compact neighborhood X_1 such that we can choose a global coordinate system in it (i. e., X_1 can be covered by one chart). In order to simplify proofs we assume that the compact set X_0 containing supports of the vector fields v, u_i, in turn, is contained in the interior of X_1 and so a trajectory $x^\varepsilon(t)$ cannot leave X_1. At the beginning of the next section this assumption will be dropped.

Let $W^m = \{w(t)\}$ be the set of trajectories of the m-dimensional Wiener process and $\mathbf{P}$ be the Wiener measure on W^m. Denote by $(x^\varepsilon(t), y^\varepsilon(t)) = (x^\varepsilon(t, w), y^\varepsilon(t, w))$ the solution to the system (1.1) that has an initial condition (x, y). It is known from the theory of stochastic differential equations that if $f \in C^i(X \times Y)$, where $i = 1, \ldots, N$, then for every $p \ge 1$ the process $f(x^\varepsilon(t), y^\varepsilon(t))$ forms a continuous curve in the space $L^p(W^m, \mathbf{P})$ which depends i times continuously differentiably on the initial data x, y and on the parameter ε (see, for instance, [14] and [18]). Let us define the semigroup of operators of conditional expectation in the space $C(X_1 \times Y)$ by the formula

$$A^t_\varepsilon f(x, y) = \mathbf{E}_{x,y} f(x^\varepsilon(t), y^\varepsilon(t)) = \int_{W^m} f(x^\varepsilon(t, w), y^\varepsilon(t, w))\, d\mathbf{P}(w). \quad (3.1)$$

Proposition 3.1 (see [2]). *For the semigroup A^t_0 given by* (3.1) *with $\varepsilon = 0$ there exists a projection $\bar{A}: C(X_1 \times Y) \to C(X_1)$ and numbers $C_0 > 0$, $\Lambda_0 \in (0, 1)$ such that for any $i = 0, 1, \ldots, N$ and a function $f \in C^i(X_1 \times Y)$ the following estimate $\|(A^t_0 - \bar{A})f\|_i \le C_0 \Lambda^t_0 \|f\|_i$ is true.*

This proposition implies $\bar{A} = \lim_{t\to\infty} A^t_0$ and therefore $\bar{A} A^t_0 = \bar{A}$. This means that if $\varepsilon = 0$ the following equalities are true

$$\bar{A}\big(f(x, y)\big) = \bar{A}\big(A^t_0 f(x, y)\big) = \bar{A}\big(\mathbf{E}_{x,y} f(x, y^0(t))\big). \quad (3.2)$$

Note that when $\varepsilon = 0$ the system (1.1) is a family of nondegenerate diffusion processes on Y depending on the parameter x. It is well known that each of these processes has a unique invariant probability measure μ_x on Y and the value of the function $\bar{A} f$ at a point $x \in X_1$ is equal to the integral of f with respect to μ_x.

For the system (1.1) let us define the vector fields

$$B(x, y) = v(x, y, 0) + \frac{1}{2} \sum_{i=1}^{m} \frac{du_i(x, y, 0)}{dy} a_i(x, y, 0), \quad (3.3)$$

$$\bar{B}(x) = \bar{A}\big(B(x, y)\big), \qquad \tilde{B}(x, y) = B(x, y) - \bar{B}(x). \quad (3.4)$$

It is clear that the vector field $B(x, y)$ is of the class $C^N(X_1 \times Y)$ and it vanishes outside the compact subset $X_0 \subset \operatorname{Int} X_1$. Moreover, Proposition 3.1 implies that the vector fields $\bar{B}$ and $\tilde{B}$ have the same properties.

For any functions $F, \varphi \in C^{i+2}(X_1),\ i = 1, \dots, N$, let us consider the differentials

$$D_\varepsilon F(x, y) = \frac{1}{\varepsilon}\frac{dF(x)}{dx} \circ dx = \frac{dF(x)}{dv}\,dt + \sum_{i=1}^{m} \frac{dF(x)}{du_i} \circ dw^i(t), \tag{3.5}$$

$$\tilde{D}_\varepsilon \varphi(x, y) = D_\varepsilon \varphi(x, y) - \frac{d\varphi(x)}{d\bar{B}}\,dt, \tag{3.6}$$

where $v = v(x, y, \varepsilon),\ u_i = u_i(x, y, \varepsilon)$. Let $x = (x^1, \dots, x^n)$ be the local coordinates on X and in these coordinates $v = (v^1, \dots, v^n)$ and $u_i = (u_i^1, \dots, u_i^n)$. Then we have

$$\begin{aligned} D_\varepsilon F(x, y) = {} & \frac{dF(x)}{dx}\left(v\,dt + \sum_{i=1}^{m}\left(u_i\,dw^i(t) + \frac{\varepsilon}{2}\frac{du_i}{dx}u_i\,dt + \frac{1}{2}\frac{du_i}{dy}a_i\,dt\right)\right) \\ & + \frac{\varepsilon}{2}\sum_{i,j,k}\frac{\partial^2 F(x)}{\partial x^j \partial x^k}u_i^j u_i^k\,dt. \end{aligned} \tag{3.7}$$

Let as above $(x^\varepsilon(t), y^\varepsilon(t))$ be the solution to the system (1.1) with the initial condition (x, y). By means of the differentials (3.5)–(3.6) we define the families of stochastic processes

$$F_t^\varepsilon = F_t^\varepsilon(x, y) = \int_0^t D_\varepsilon F(x^\varepsilon(\tau), y^\varepsilon(\tau)), \tag{3.8}$$

$$\tilde{\varphi}_t^\varepsilon = \tilde{\varphi}_t^\varepsilon(x, y) = \int_0^t \tilde{D}_\varepsilon \varphi(x^\varepsilon(\tau), y^\varepsilon(\tau)). \tag{3.9}$$

In particular, by the Itô formula we have $\varepsilon F_t^\varepsilon(x, y) = F(x^\varepsilon(t)) - F(x)$. And if $\varepsilon = 0$ then (3.7), (3.3), and (3.4) imply

$$F_t^0(x, y) = \int_0^t \frac{dF(x)}{dx}\left(B(x, y^0(\tau))\,d\tau + \sum_{i=1}^{m} u_i(x, y^0(\tau), 0)\,dw^i(\tau)\right), \tag{3.10}$$

$$\tilde{\varphi}_t^0(x, y) = \int_0^t \frac{d\varphi(x)}{dx}\left(\tilde{B}(x, y^0(\tau))\,d\tau + \sum_{i=1}^{m} u_i(x, y^0(\tau), 0)\,dw^i(\tau)\right). \tag{3.11}$$

The processes $\exp(F_t^\varepsilon)$ and $\exp(\tilde{\varphi}_t^\varepsilon)$ satisfy the linear stochastic differential equations with bounded coefficients

$$d\exp(F_t^\varepsilon) = \exp(F_t^\varepsilon) \circ D_\varepsilon F(x^\varepsilon(t), y^\varepsilon(t)),$$
$$d\exp(\tilde{\varphi}_t^\varepsilon) = \exp(\tilde{\varphi}_t^\varepsilon) \circ \tilde{D}_\varepsilon \varphi(x^\varepsilon(t), y^\varepsilon(t)).$$

From the theory of stochastic differential equations it is well known that if $F, \varphi \in C^{i+2}(X_1)$, where $i = 1, \dots, N$, then for every $p \geq 1$ the processes F_t^ε, $\tilde{\varphi}_t^\varepsilon$, $\exp(F_t^\varepsilon)$ and $\exp(\tilde{\varphi}_t^\varepsilon)$ form continuous curves in the space $L^p(W^m, \mathbf{P})$ and these curves depend i times continuously differentiably on the initial data $x,\ y$ and the parameter ε. For any functions $F, \varphi \in C^{i+2}(X_1),\ i = 1, \ldots, N$, let us define the family of semigroups $A^t_\varepsilon[F, \varphi]$ on the space $C^i(X_1 \times Y)$ by the following formula

$$A^t_\varepsilon[F, \varphi] f(x, y) = \mathbf{E}_{x,y}\big(\exp\big(F^\varepsilon_t(x, y) + \tilde{\varphi}^\varepsilon_t(x, y)\big) f(x^\varepsilon(t), y^\varepsilon(t))\big). \tag{3.12}$$

From (3.12) and the equality $\varepsilon F^\varepsilon_t(x, y) = F(x^\varepsilon(t)) - F(x)$ it follows that when $\varepsilon \neq 0$,

$$A^t_\varepsilon[F, \varphi] f = e^{-F/\varepsilon} A^t_\varepsilon[0, \varphi]\left(e^{F/\varepsilon} f\right). \tag{3.13}$$

Proposition 3.2. *If $F, \varphi \in C^{i+2}(X_1)$, where $1 \le i \le N$, then the linear operator $f(x, y) \mapsto g(x, y, \varepsilon) = A^t_\varepsilon[F, \varphi] f(x, y)$ maps continuously the space $C^i(X_1 \times Y)$ into the space $C^i(X_1 \times Y \times [-\varepsilon_0, \varepsilon_0])$ and this operator depends analytically on $F,\ \varphi$. If the parameters $t,\ F,\ \varphi$ range over bounded domains then this family of operators is uniformly bounded in norm.*

Proof. The statement follows from formula (3.12) and the smooth dependence of the processes $\exp(F^\varepsilon_t(x, y))$, $\exp(\tilde{\varphi}^\varepsilon_t(x, y))$ and $f(x^\varepsilon(t), y^\varepsilon(t))$ on $x,\ y,\ \varepsilon$. □

Proposition 3.3. *For all sufficiently small $F, \varphi \in C^{i+1}(X_1),\ i = 1, \ldots, N$ there exist uniquely defined functions $\lambda[F, \varphi] \in C^i(X_1),\ h[F, \varphi] \in C^i(X_1 \times Y)$ and a $C^i(X_1)$-linear functional $\nu[F, \varphi]\colon C^i(X_1 \times Y) \to C^i(X_1)$ that depend analytically on $F,\ \varphi$ and satisfy the conditions*

$$\lambda[0,0] \equiv 0,\quad h[0,0] \equiv 1,\quad \nu[0,0] = \bar{A},\quad \bar{A}h[F, \varphi] \equiv 1,\quad \nu[F, \varphi](h[F, \varphi]) \equiv 1;$$
$$A^t_0[F, \varphi] h[F, \varphi] = e^{t\lambda[F,\varphi]} h[F, \varphi], \quad \nu[F, \varphi] \circ A^t_0[F, \varphi] = e^{t\lambda[F,\varphi]} \nu[F, \varphi];$$
$$\lim_{t\to\infty} e^{-t\lambda[F,\varphi]} A^t_0[F, \varphi] = \bar{A}[F, \varphi], \quad \textit{where}\ \ \bar{A}[F, \varphi] f = \nu[F, \varphi](f) h[F, \varphi].$$

The proof of this proposition can be obtained by a simple application of the implicit function theorem. It is given in Proposition 2.1 of [2] for the family of operators $A^t_0[F, 0]$ (that is in the case $\varphi \equiv 0$). In the situation when $\varphi \neq 0$ the proof is absolutely the same.

Proposition 3.4. *Under the conditions and notation of Proposition 3.3 the functions $\lambda[F, \varphi]$ and $h[F, \varphi]$ have the form $\lambda[F, \varphi](x) = \lambda(x, p, q)$ and $h[F, \varphi](x, y) = h(x, y, p, q)$ respectively, where $p = dF(x)/dx$ and $q = d\varphi(x)/dx$. The functions $\lambda(x, p, q)$ and $h(x, y, p, q)$ depend analytically on the covectors $p, q \in T^*_x X_1$ and they are N times continuously differentiable with respect to $x \in X_1,\ y \in Y$. In addition, if $x \in X_1 \setminus X_0$ then $\lambda(x, p, q) = 0$ and $h(x, y, p, q) = 1$.*

Proof. We have by definition that $A^t_0[F, \varphi] f(x, y) = \mathbf{E}_{x,y}\big(\exp(F^0_t + \tilde{\varphi}^0_t)\, f(x, y^0(t))\big)$. If we substitute in this formula the expressions (3.10) and (3.11) for the processes F^0_t and $\tilde{\varphi}^0_t$ then it becomes clear that the operator $A^t_0[F, \varphi]$ acts independently on every fiber $x = \text{const}$ and its restriction onto this fiber depends analytically on the covectors $p = dF(x)/dx$ and $q = d\varphi(x)/dx$. For any fixed x the number $e^{t\lambda[F,\varphi](x)}$ is the maximal eigenvalue for the restriction of the operator $A^t_0[F, \varphi]$ onto the fiber $x = \text{const}$ and the function $h[F, \varphi](x, y)$ is the corresponding eigenvector normalized by the condition $\bar{A}h[F, \varphi] = 1$. Therefore the functions $\lambda[F, \varphi](x)$ and $h[F, \varphi](x, y)$ also depend analytically on $p = dF(x)/dx$ and $q = d\varphi(x)/dx$ and thus they have the form $\lambda(x, p, q)$ and $h(x, y, p, q)$ respectively. Further, if we choose the functions F and φ in such a way that in some local coordinates x $p = dF(x)/dx = \text{const}$ and $q = d\varphi(x)/dx = \text{const}$ then Proposition 3.3 implies that the functions $\lambda(x, p, q)$ and $h(x, y, p, q)$ are at least N times continuously differentiable with respect to x, y. Finally, the vector fields $B(x, y)$, $\tilde{B}(x, y)$, $u_i(x, y, \varepsilon)$ vanish outside X_0. In view of (3.10) and (3.11) we have $F_t^0 = 0$ and $\tilde{\varphi}_t^0 = 0$ if $x \in X_1 \setminus X_0$. Thus the operator $A_0^t[F, \varphi]$ coincides with A_0^t in $X_1 \setminus X_0$ and as a result $\lambda(x, p, q) = 0$ and $h(x, y, p, q) = 1$ there. □

Proposition 3.5. *Under the conditions and notation of Proposition 3.3 we have*

$$\left.\frac{d\lambda[\xi F, 0]}{d\xi}\right|_{\xi=0} = \frac{dF}{d\bar{B}}, \qquad \left.\frac{d\lambda[0, \xi\varphi]}{d\xi}\right|_{\xi=0} = 0. \tag{3.14}$$

Proof. From (3.2) and (3.4) it follows that $\bar{A}\big(\mathbf{E}_{x,y}B(x, y^0(\tau))\big) = \bar{A}\big(B(x, y)\big) = \bar{B}(x)$ and $\bar{A}\big(\mathbf{E}_{x,y}\tilde{B}(x, y^0(\tau))\big) = \bar{A}\big(\tilde{B}(x, y)\big) = 0$. In view of these equalities along with (3.10) and (3.11) we have

$$\bar{A}\left(\mathbf{E}_{x,y}F_t^0(x, y)\right) = \int_0^t \frac{dF(x)}{dx}\bar{A}\Big(\mathbf{E}_{x,y}B(x, y^0(\tau))\Big)d\tau = t\frac{dF(x)}{dx}\bar{B}(x), \tag{3.15}$$

$$\bar{A}\left(\mathbf{E}_{x,y}\tilde{\varphi}_t^0(x, y)\right) = \int_0^t \frac{d\varphi(x)}{dx}\bar{A}\Big(\mathbf{E}_{x,y}\tilde{B}(x, y^0(\tau))\Big)d\tau = 0. \tag{3.16}$$

The equalities (3.12), (3.15), and (3.16) imply

$$\bar{A}\left(\left.\frac{d}{d\xi}\right|_{\xi=0} A_0^t[\xi F, 0]1\right) = \bar{A}\big(\mathbf{E}_{x,y}F_t^0(x, y)\big) = t\frac{dF(x)}{d\bar{B}(x)},$$
$$\bar{A}\left(\left.\frac{d}{d\xi}\right|_{\xi=0} A_0^t[0, \xi\varphi]1\right) = \bar{A}\big(\mathbf{E}_{x,y}\tilde{\varphi}_t^0(x, y)\big) = 0.$$

Moreover, it follows from Proposition 3.3 that $\bar{A}h[\xi F, \zeta\varphi] \equiv 1$. So $\bar{A}(dh[\xi F, 0]/d\xi) \equiv 0$ and $\bar{A}(dh[0, \zeta\varphi]/d\zeta) \equiv 0$. Now to finish the proof it is enough to differentiate the identity $\bar{A}\big(A_0^t[\xi F, \zeta\varphi]h[\xi F, \zeta\varphi]\big) \equiv e^{t\lambda[\xi F, \zeta\varphi]}$ with respect to ξ and ζ (when $\xi = \zeta = 0$). □

We shall fix once and for all a family of functions $\varphi = (\varphi_1(x), \ldots, \varphi_n(x)) \in (C^{N+2}(X_1))^n$. Let us introduce the n-dimensional parameter $\xi = (\xi_1, \ldots, \xi_n)$ and the notation $\xi\varphi(x) = \xi_1\varphi_1(x) + \cdots + \xi_n\varphi_n(x)$. Consider in the domain X_1 the first order partial differential equation depending on the parameter $\xi \in \mathbb{R}^n$:

$$\dot{F}_t = \lambda[F_t, \xi\varphi], \qquad F_0 \equiv 0. \tag{3.17}$$

We shall denote by $F_t(\xi) = F_t(\xi, x)$ a solution to this equation. By construction, the function $\lambda[F, \varphi]$ vanishes in $X_1 \setminus X_0$ and in addition $\lambda[0, 0] \equiv 0$. Thus for $\xi = 0$ we have $F_t(0, x) \equiv 0$. Under these conditions and in view of Proposition 3.3

it follows from general theory of the first order partial differential equations that for any $T > 0$ there exists a (small) neighbourhood U of the origin in $\mathbb{R}^n$ such that equation (3.17) has a solution $F_t(\xi, x)$ defined for every $t \in [0, T]$, $\xi \in U$, $x \in X_1$ and this solution is at least $N - 1$ times continuously differentiable with respect to t, ξ, x. Let us fix a time segment $[0, T]$, the corresponding neighborhood $U \subset \mathbb{R}^n$ and the solution $F_t(\xi) = F_t(\xi, x)$.

By Proposition 3.3 to any function $F_t(\xi) \in C^{N-1}(X_1)$ there corresponds the function $h[F_t(\xi), \xi\varphi] \in C^{N-2}(X_1 \times Y)$. For the sake of brevity we introduce the notation $h_t(\xi) = h[F_t(\xi), \xi\varphi]$. Proposition 3.4 implies that the function $h_t(\xi) = h(x, y, dF_t(\xi, x)/dx, d\xi\varphi(x)/dx)$ is at least $N - 2$ times continuously differentiable with respect to the totality of variables $t \in [0, T]$, $\xi \in U$, $x \in X_1$, $y \in Y$.

Let $F_t^k(\xi)$ and $h_t^k(\xi)$ be the Taylor polynomials for the functions $F_t(\xi)$ and $h_t(\xi)$ of power k on ξ:

$$F_t^k(\xi) = \sum_{i=0}^{k} \frac{1}{i!} F_{it}(\xi^i), \qquad F_{it} = \left.\frac{d^i F_t(\xi)}{d\xi^i}\right|_{\xi=0}, \tag{3.18}$$

$$h_t^k(\xi) = \sum_{i=0}^{k} \frac{1}{i!} h_{it}(\xi^i), \qquad h_{it} = \left.\frac{d^i h_t(\xi)}{d\xi^i}\right|_{\xi=0}. \tag{3.19}$$

Proposition 3.6. *In the notation* (3.18), (3.19) *we have* $F_{0t} \equiv 0$, $F_{1t} \equiv 0$, $h_{0t} \equiv 1$ *and the quadratic form* F_{2t}*, being considered as a function of* t *and* x*, satisfies the linear partial differential equation*

$$\dot F_{2t} = \frac{dF_{2t}}{d\bar B} + \left.\frac{d^2\lambda[0, \xi\varphi]}{d\xi^2}\right|_{\xi=0}, \qquad F_{20} = 0. \tag{3.20}$$

Proof. We know already that $F_{0t} = F_t(0) \equiv 0$. By Proposition 3.3, $h[0, 0] \equiv 1$. Therefore, $h_{0t} = h[F_t(0), 0] \equiv 1$. Linearizing equation (3.17) by means of (3.14), we obtain an equation for the family of linear functionals F_{1t}:

$$\dot F_{1t} = \left.\frac{d\lambda[F_{1t}(\xi), 0]}{d\xi}\right|_{\xi=0} + \left.\frac{d\lambda[0, \xi\varphi]}{d\xi}\right|_{\xi=0} = \frac{dF_{1t}}{d\bar B}, \qquad F_{10} \equiv 0.$$

Evidently, it has zero solution. Then, extracting the quadratic part in (3.17), we get exactly (3.20). □

In the notation of (3.18), (3.19) consider the expression

$$r_{t\tau}^k(\xi, \varepsilon) = \exp\bigl\{\varepsilon^{-1}\bigl(F_t^k(\xi) - F_{t+\varepsilon\tau}^k(\xi)\bigr)\bigr\} A_\varepsilon^\tau\bigl[F_t^k(\xi), \xi\varphi\bigr] h_t^k(\xi) - h_{t+\varepsilon\tau}^k(\xi). \tag{3.21}$$

Since $F_t^k(\xi)$ and $h_t^k(\xi)$ are polynomials in $\xi \in \mathbb{R}^n$, the last formula is well defined for all $\xi \in \mathbb{C}^n$.

Proposition 3.7. *Let $k \le N-4$. Then given $T > 0$ there exists a neighborhood V_T of the origin in $\mathbb{C}^n$ such that the expression $r^k_{t\tau}(\xi,\varepsilon)$ from* (3.21) *can be represented in the form $r^k_{t\tau}(\xi,\varepsilon) = \varepsilon\alpha_{t\tau}(\xi,\varepsilon) + \beta_{t\tau}(\xi)$, where the families $\alpha_{t\tau}(\xi,\varepsilon) \in C^{N-4-k}(X_1 \times Y)$ and $\beta_{t\tau}(\xi) \in C^{N-3-k}(X_1 \times Y)$ are analytic in $\xi \in V_T$ and for all $t,\tau \in [0,T]$, $\xi \in V_T$, $\varepsilon \in [-\varepsilon_0,\varepsilon_0]$ satisfy estimates $\|\alpha_{t\tau}(\xi,\varepsilon)\|_{N-4-k} \le C|\xi|$, $\|\beta_{t\tau}(\xi)\|_{N-3-k} \le C|\xi|^{k+1}$, the constant C being independent of $t,\ \tau,\ \xi,\ \varepsilon$.*

Proof. First of all note that the family $F^k_t(\xi)$ is a polynomial in ξ with coefficients belonging to the space $C^{N-1-k}\{(t,x) \in [0,T] \times X_1\}$ and the family $h^k_t(\xi)$ is a polynomial in ξ with coefficients belonging to the space $C^{N-2-k}\{(t,x,y) \in [0,T] \times X_1 \times Y\}$. Hence Proposition 3.2 implies that the family $r^k_{t\tau}(\xi,\varepsilon)$ is analytic in ξ and at least $N-3-k$ times continuously differentiable with respect to $t,\ x,\ y,\ \varepsilon$ and all its partial derivatives up to order mentioned above are uniformly bounded if $t,\tau \in [0,T]$. Put $\beta_{t\tau}(\xi) = r^k_{t\tau}(\xi,0)$ and

$$\varepsilon\alpha_{t\tau}(\xi,\varepsilon) = r^k_{t\tau}(\xi,\varepsilon) - r^k_{t\tau}(\xi,0) = \int_0^1 \frac{dr^k_{t\tau}(\xi,\theta\varepsilon)}{d\theta}\,d\theta.$$

Recall that $e^{-\tau\lambda[F_t(\xi),\xi\varphi]}A^\tau_0[F_t(\xi),\xi\varphi]h_t(\xi) \equiv h_t(\xi)$ (by Proposition 3.3) and $\dot F_t(\xi) = \lambda[F_t(\xi),\xi\varphi]$ (by (3.17)). So (3.21) implies that $r^k_{t\tau}(\xi,0) = O\left(|\xi|^{k+1}\right)$ and $r^k_{t\tau}(0,\varepsilon) \equiv 0$. From here we get easily the required estimates for $\beta_{t\tau}(\xi)$ and $\alpha_{t\tau}(\xi,\varepsilon)$. □

Proposition 3.8. *Suppose $\varphi = (\varphi_1,\dots,\varphi_n) \in (C^{N+2}(X_1))^n$ and $\tilde\varphi^\varepsilon_t = \tilde\varphi^\varepsilon_t(x,y)$ is a family of n-dimensional stochastic processes defined by* (3.9)*. Then given $T > 0$ and $k \le N-4$ there exist a small neighborhood V_T of zero in $\mathbb{C}^n$, a number $\varepsilon_T > 0$ and a family of stochastic processes $G^\varepsilon_t(\xi) = G^\varepsilon_t(\xi,x,y)$ such that for all $\varepsilon \in (0,\varepsilon_T)$, $t \in [0,T/\varepsilon]$, $\xi \in V_T$, $x \in X_1$, $y \in Y$ the following equality holds:*

$$\mathbf{E}_{x,y}\exp\bigl(\xi\tilde\varphi^\varepsilon_t + G^\varepsilon_t(\xi)\bigr) = \exp\bigl(F^k_{\varepsilon t}(\xi)/\varepsilon\bigr), \tag{3.22}$$

in which $F^k_t(\xi)$ is the Taylor polynomial (3.18) *for the solution of equation* (3.17)*. Here $G^\varepsilon_t(\xi)$ depends analytically on ξ, is real-valued for real ξ and satisfies the estimate $|G^\varepsilon_t(\xi,x,y)| \le C|\xi| + Ct|\xi|^{k+1}$, the constant C being independent of $\varepsilon,\ t,\ \xi,\ x,\ y$.*

Proof. It follows from (3.13) and (3.21) that

$$r^k_{t\tau}(\xi,\varepsilon) = e^{-F^k_{t+\varepsilon\tau}(\xi)/\varepsilon}A^\tau_\varepsilon[0,\xi\varphi]\left(e^{F^k_t(\xi)/\varepsilon}h^k_t(\xi)\right) - h^k_{t+\varepsilon\tau}(\xi). \tag{3.23}$$

Define families of nonrandom functions $g^\varepsilon_i(\xi) = g^\varepsilon_i(\xi,x,y)$ by the formulas

$$g^\varepsilon_0(\xi) = \ln h^k_0(\xi),$$

$$g^\varepsilon_i(\xi) = \ln\frac{h^k_{\varepsilon i}(\xi)}{h^k_{\varepsilon i}(\xi) + r^k_{\varepsilon i-\varepsilon,1}(\xi,\varepsilon)}, \qquad i = 1,\ 2,\ \dots,\ [t], \tag{3.24}$$

$$g^\varepsilon_t(\xi) = -\ln\left\{e^{-F^k_{\varepsilon t}(\xi)/\varepsilon}A^{\{t\}}_\varepsilon[0,\xi\varphi]\left(e^{F^k_{\varepsilon[t]}(\xi)/\varepsilon}h^k_{\varepsilon[t]}(\xi)\right)\right\} \tag{3.25}$$

(where $[\,\cdot\,]$ and $\{\,\cdot\,\}$ denote the integer and the fractional part, respectively, and the principal value of the logarithm function is taken everywhere). By (3.23) and (3.24) we have

$$e^{g_i^\varepsilon(\xi)} A_\varepsilon^1[0, \xi\varphi] \left(e^{F_{\varepsilon i-\varepsilon}^k(\xi)/\varepsilon} h_{\varepsilon i-\varepsilon}^k(\xi) \right) = e^{F_{\varepsilon i}^k(\xi)/\varepsilon} h_{\varepsilon i}^k(\xi). \tag{3.26}$$

As before, let $(x^\varepsilon(t), y^\varepsilon(t))$ be a solution to system (1.1)) with initial data (x, y). Put

$$G_t^\varepsilon(\xi, x, y) = g_t^\varepsilon(\xi, x, y) + \sum_{i=0}^{[t]} g_i^\varepsilon(\xi, x^\varepsilon(t-i), y^\varepsilon(t-i)). \tag{3.27}$$

Proposition 3.7 implies the desired estimate for the family $G_t^\varepsilon(\xi, x, y)$. From the definition (3.9) of the stochastic process $\tilde{\varphi}_t^\varepsilon(x, y)$ it follows that

$$\tilde{\varphi}_t^\varepsilon(x, y) = \tilde{\varphi}_{\{t\}}^\varepsilon(x, y) + \sum_{i=1}^{[t]} \tilde{\varphi}_1^\varepsilon(x^\varepsilon(t-i), y^\varepsilon(t-i)). \tag{3.28}$$

Therefore, by (3.27) and (3.28),

$$\begin{aligned}
&\mathbf{E}_{x,y} \exp\big(\xi\tilde{\varphi}_t^\varepsilon(x, y) + G_t^\varepsilon(\xi, x, y)\big) \\
&\quad = e^{g_t^\varepsilon(\xi,x,y)} \mathbf{E}_{x,y} \Big(e^{\xi\tilde{\varphi}_{\{t\}}^\varepsilon(x,y)} e^{g_{[t]}^\varepsilon(\xi, x^\varepsilon(\{t\}), y^\varepsilon(\{t\}))} \\
&\qquad \times \mathbf{E}_{x^\varepsilon(\{t\}), y^\varepsilon(\{t\})} \Big(e^{\xi\tilde{\varphi}_1^\varepsilon(x^\varepsilon(\{t\}), y^\varepsilon(\{t\}))} \dots e^{g_1^\varepsilon(\xi, x^\varepsilon(t-1), y^\varepsilon(t-1))} \\
&\qquad \times \mathbf{E}_{x^\varepsilon(t-1), y^\varepsilon(t-1)} \Big(e^{\xi\tilde{\varphi}_1^\varepsilon(x^\varepsilon(t-1), y^\varepsilon(t-1))} e^{g_0^\varepsilon(\xi, x^\varepsilon(t), y^\varepsilon(t))} \Big)\Big) \dots \Big) \\
&\quad = e^{g_t^\varepsilon(\xi)} A_\varepsilon^{\{t\}}[0, \xi\varphi] \Big(e^{g_{[t]}^\varepsilon(\xi)} A_\varepsilon^1[0, \xi\varphi] \ \dots \ \Big(e^{g_1^\varepsilon(\xi)} A_\varepsilon^1[0, \xi\varphi] h_0^k(\xi) \Big) \dots \Big).
\end{aligned}$$

To calculate the last expression use successively (from right to left) equality (3.26) for $i = 1, 2, \dots, [t]$ and equality (3.25) at the end. As a result we obtain exactly (3.22). □

Corollary 3.9. *In the setting of Proposition 3.8 for all $\varepsilon \in (0, \varepsilon_T)$, $t \in [0, T]$, $(x, y) \in (X_1 \times Y)$ and $\zeta \in \mathbb{R}^n$ we have*

$$\mathbf{E}_{x,y} \exp\big(i\sqrt{\varepsilon}\zeta\tilde{\varphi}_{t/\varepsilon}^\varepsilon(x, y)\big) - \exp\big(-\tfrac{1}{2}\langle F_{2,t}(x)\zeta, \zeta\rangle\big) = O\left(\sqrt{\varepsilon}|\zeta| + t\sqrt{\varepsilon}|\zeta|^3\right),$$

where the matrix $F_{2,t}(x)$ is defined by the formulas

$$F_{2,t}(x) = \int_0^t a(\bar{z}(s))\, ds, \qquad a(x) = \left.\frac{d^2\lambda[0, \xi\varphi]}{d\zeta^2}\right|_{\xi=0},$$

provided $\bar{z}(0) = x$. Here $\langle \xi, \zeta \rangle = \xi_1\zeta_1 + \dots + \xi_n\zeta_n$, and $O(\zeta)$ is a function satisfying the estimate $|O(\zeta)| \le C_T|\zeta|$ whenever $|\zeta| < \delta_T$, the constants C_T, δ_T being positive and independent of ζ.

Proof. The first formula above follows from Proposition 3.8 if we take $k = 2$, $\xi = i\sqrt{\varepsilon}\zeta$ and replace t by t/ε in it. The second formula merely defines the solution to (3.20). $\square$

Proposition 3.10. *Given $T > 0$ and a function $\varphi \in C^{N+2}(X_1)$ there exist $\varepsilon_T > 0$ so small and $C = C(T)$ so large that for all $\varepsilon \in (0, \varepsilon_T)$, $t \in [0, T/\varepsilon]$, $(x, y) \in X_1 \times Y$ and positive $\varkappa$ we have the estimates*

$$\mathbf{P}\big\{\tilde{\varphi}_t^\varepsilon(x, y) \geq \varkappa\big\} \leq C \exp\Big(-\varkappa/\sqrt{C + Ct}\Big), \tag{3.29}$$

$$\mathbf{P}\Bigg\{\sup_{0\leq s\leq t} \big|\tilde{\varphi}_s^\varepsilon(x, y)\big| \geq \varkappa\Bigg\} \leq C(1 + t) \exp\Big(-\varkappa/\sqrt{C + Ct}\Big). \tag{3.30}$$

Proof. Apply Proposition 3.8 for $n = k = 1$. In this case $F^1_{\varepsilon t}(\xi) \equiv 0$. Hence (3.22) yields for all small enough $\xi > 0$ inequality

$$\begin{aligned}\mathbf{P}\big\{\tilde{\varphi}_t^\varepsilon(x, y) \geq \varkappa\big\} &\leq \mathbf{E}_{x,y} \exp\big(\xi(\tilde{\varphi}_t^\varepsilon - \varkappa)\big)\\ &\leq \mathbf{E}_{x,y} \exp\big(\xi\tilde{\varphi}_t^\varepsilon + G_t^\varepsilon(\xi) + C\xi + Ct\xi^2 - \xi\varkappa\big)\\ &= \exp\Big(C\xi + Ct\xi^2 - \xi\varkappa\Big).\end{aligned}$$

Taking $\xi = (C + Ct)^{-1/2}$ in it, we get

$$\mathbf{P}\big\{\tilde{\varphi}_t^\varepsilon(x, y) \geq \varkappa\big\} \leq \exp\Big(\sqrt{C} + 1 - \varkappa/\sqrt{C + Ct}\Big).$$

Thus (3.29) is proved. From (3.6)–(3.9) it follows that for any $t > s$,

$$\tilde{\varphi}_t^\varepsilon(x, y) - \tilde{\varphi}_s^\varepsilon(x, y) = \int_s^t R_\tau^\varepsilon(x, y)\, d\tau + \int_s^t Q_\tau^\varepsilon(x, y)\, dw(\tau),$$

where $R_t^\varepsilon(x, y)$ and $Q_t^\varepsilon(x, y)$ are uniformly bounded stochastic processes. By standard estimates for stochastic integrals (see, for instance, [14] or [18]) we see that

$$\mathbf{P}\Bigg\{\sup_{s\leq\tau\leq s+1} \big|\tilde{\varphi}_\tau^\varepsilon(x, y) - \tilde{\varphi}_s^\varepsilon(x, y)\big| \geq \varkappa\Bigg\} \leq Ce^{-\varkappa^2/C}$$

for some C independent of ε, s, x, y, $\varkappa$. Combining this with (3.29) for all integer $s \in [0, t]$, we obtain (3.30) with some larger C. $\square$

Proposition 3.11. *For any function $\varphi \in C^{N+2}(X_1)$ and numbers $T > 0$ and $p > 0$ there exist $\varepsilon_T > 0$ so small and $C = C(p, T)$ so large that*

$$\mathbf{E}_{x,y}\big|\tilde{\varphi}_t^\varepsilon(x, y)\big|^{2p} \leq C(1 + t)^p, \tag{3.31}$$

$$\mathbf{E}_{x,y} \sup_{0\leq s\leq t} \big|\tilde{\varphi}_s^\varepsilon(x, y)\big|^{2p} \leq C(1 + t)^p \ln^{2p}(2 + t) \tag{3.32}$$

for all $\varepsilon \in (0, \varepsilon_T)$, $t \in [0, T/\varepsilon]$, $(x, y) \in X_1 \times Y$.

Proof. Denote by $\Phi(\varkappa)$ the distribution function for the random variable $\tilde{\varphi}_t^{\varepsilon}(x, y)$ and by $\Psi(\varkappa)$ the distribution function for $-\tilde{\varphi}_t^{\varepsilon}(x, y)$. Obviously,

$$\mathbf{E}_{x,y}\big|\tilde{\varphi}_t^{\varepsilon}(x, y)\big|^{2p} = \int_0^{\infty} \varkappa^{2p}\, d\Phi(\varkappa) + \int_0^{\infty} \varkappa^{2p}\, d\Psi(\varkappa). \tag{3.33}$$

Consider the first of two integrals in the right-hand side of (3.33). Take $\varkappa_i = i\sqrt{C + Ct}$. Then by (3.29),

$$\begin{aligned}\int_0^{\infty} \varkappa^{2p}\, d\Phi(\varkappa) &= \sum_{i=0}^{\infty} \int_{\varkappa_i}^{\varkappa_{i+1}} \varkappa^{2p}\, d\Phi(\varkappa) \le \sum_{i=0}^{\infty} \varkappa_{i+1}^{2p} \mathbf{P}\big\{\tilde{\varphi}_t^{\varepsilon}(x, y) \ge \varkappa_i\big\} \\ &\le (C + Ct)^p \sum_{i=0}^{\infty} (i+1)^{2p} C e^{-i}.\end{aligned}$$

The second integral in the right-hand side of (3.33) can be evaluated in the same manner. So (3.31) is proved. Now take $\varkappa_i = i\sqrt{C + Ct}\ln(2 + t)$. Then by (3.30),

$$\begin{aligned}\mathbf{E}_{x,y} \sup_{0\le s\le t} \big|\tilde{\varphi}_s^{\varepsilon}(x, y)\big|^{2p} &\le \varkappa_1^{2p} + \sum_{i=1}^{\infty} \varkappa_{i+1}^{2p} \mathbf{P}_{x,y}\left\{\sup_{0\le s\le t} \big|\tilde{\varphi}_s^{\varepsilon}(x, y)\big| \ge \varkappa_i\right\} \\ &\le (C + Ct)^p \ln^{2p}(2 + t) + (C + Ct)^p \ln^{2p}(2 + t) \\ &\quad \times \sum_{i=1}^{\infty} (i+1)^{2p} C e^{-(i-1)\ln(2+t)}\end{aligned}$$

and (3.32) is proved too. □

Proposition 3.12. *Given $T > 0$ and $p > 0$ there exist $\varepsilon_T > 0$ so small and $C = C(p, T)$ so large that for all $\varepsilon \in (0, \varepsilon_T)$, $(x, y) \in X_1 \times Y$ and $\varkappa > 0$,*

$$\mathbf{P}_{x,y}\left\{\sup_{0\le t\le T} d_X(z^{\varepsilon}(t), \bar{z}(t)) \ge \varkappa\right\} \le C\varepsilon^{-1} \exp\left(-\frac{\varkappa}{C\sqrt{\varepsilon}}\right), \tag{3.34}$$

$$\mathbf{E}_{x,y}\left(\sup_{0\le t\le T} d_X(z^{\varepsilon}(t), \bar{z}(t))\right)^{2p} \le C\varepsilon^p \ln^{2p}\big(\varepsilon^{-1}\big), \tag{3.35}$$

where d_X is the distance function on X.

Proof. Choose a set of coordinate functions $\varphi = (\varphi_1, \dots, \varphi_n) \in \big(C^{N+2}(X_1)\big)^n$ on X. From (3.6) and (3.9) it follows that

$$\begin{aligned}\tilde{\varphi}_t^{\varepsilon}(x, y) &= \int_0^t D_{\varepsilon}\varphi(x^{\varepsilon}(s), y^{\varepsilon}(s)) - \frac{d\varphi}{d\bar{B}}(\bar{z}(\varepsilon s))\, ds \\ &\quad + \int_0^t \left(\frac{d\varphi}{d\bar{B}}(\bar{z}(\varepsilon s)) - \frac{d\varphi}{d\bar{B}}(x^{\varepsilon}(s))\right) ds \\ &= \varepsilon^{-1}\big(\varphi(x^{\varepsilon}(t)) - \varphi(\bar{z}(\varepsilon t))\big) \\ &\quad + \int_0^t \left(\frac{d\varphi}{d\bar{B}}(\bar{z}(\varepsilon s)) - \frac{d\varphi}{d\bar{B}}(x^{\varepsilon}(s))\right) ds.\end{aligned} \tag{3.36}$$

Hence,

$$d_X\big(x^\varepsilon(t), \bar{z}(\varepsilon t)\big) \le C\varepsilon\big|\tilde{\varphi}_t^\varepsilon(x, y)\big| + C\varepsilon \int_0^t d_X\big(x^\varepsilon(s), \bar{z}(\varepsilon s)\big)ds$$

for some $C > 0$. It follows by Gronwall's inequality that

$$\sup_{0\le t\le T} d_X\big(z^\varepsilon(t), \bar{z}(t)\big) \le e^{CT} C\varepsilon \sup_{0\le t\le T/\varepsilon} \big|\tilde{\varphi}_t^\varepsilon(x, y)\big|.$$

Combining this with (3.30) and (3.32), we obtain (3.34) and (3.35) respectively. □

Proposition 3.13. *The matrix*

$$a(x) = \left.\frac{d^2\lambda[0, \xi\varphi]}{d\xi^2}\right|_{\xi=0}$$

has the C^N dependence on x and satisfies the identity

$$\langle a(x)\xi, \xi\rangle = \lim_{t\to\infty} \frac{1}{t}\mathbf{E}_{x,y}\left(\xi\tilde{\varphi}_t^0(x, y)\right)^2. \tag{3.37}$$

So it is symmetric and nonnegatively definite.

Proof. Take $k = 2$ in Proposition 3.8. Then

$$\mathbf{E}_{x,y}\exp\big(\xi\tilde{\varphi}_t^\varepsilon + G_t^\varepsilon(\xi)\big) = \exp\big(\langle F_{2,\varepsilon t}(x)\xi, \xi\rangle/2\varepsilon\big), \tag{3.38}$$

where $F_{2,t}(x)$ is a matrix satistying differential equation (3.20) and $G_t^\varepsilon(\xi)$ is a family of stochastic processes which is analytic in $\xi = (\xi_1, \ldots, \xi_n) \in V_T \subset \mathbb{C}^n$ and satisfies the inequality $\big|G_t^\varepsilon(\xi)\big| \le C|\xi| + Ct|\xi|^3$ for all $\xi \in V_T$. Consider the Taylor expansion of $G_t^\varepsilon(\xi)$ in ξ:

$$G_t^\varepsilon(\xi) = \sum_{i=1}^n g_i^\varepsilon\xi_i + \frac{1}{2}\sum_{i,j=1}^n g_{ij}^\varepsilon\xi_i\xi_j + \cdots.$$

Let t be so large that if $|\xi| \le t^{-1/2}$ then $\xi \in V_T$. In this case by the Cauchy inequality from complex variable analysis we have

$$|g_i^\varepsilon| \le \sqrt{t} \sup_{|\xi|\le t^{-1/2}} \big|G_t^\varepsilon(\xi)\big| \le 2C, \tag{3.39}$$

$$|g_{ij}^\varepsilon| \le t \sup_{|\xi|\le t^{-1/2}} \big|G_t^\varepsilon(\xi)\big| \le C\sqrt{t}. \tag{3.40}$$

Let us differentiate (3.38) twice in ξ at the origin:

$$\mathbf{E}_{x,y}\left\{\widetilde{(\varphi_i)}_t^\varepsilon\widetilde{(\varphi_j)}_t^\varepsilon + \widetilde{(\varphi_i)}_t^\varepsilon g_j^\varepsilon + \widetilde{(\varphi_j)}_t^\varepsilon g_i^\varepsilon + g_{ij}^\varepsilon\right\} = \varepsilon^{-1}F_{2,\varepsilon t}(x). \tag{3.41}$$

From (3.31) it follows that $\mathbf{E}_{x,y}\left|\widetilde{(\varphi_i)}_t^{\varepsilon}\right| \le C\sqrt{1+t}$. Now, if we take the limit in (3.41) as $\varepsilon \to 0$ and use (3.39), (3.40) and (3.20), we get

$$\mathbf{E}_{x,y}\left\{\widetilde{(\varphi_i)}_t^0 \widetilde{(\varphi_j)}_t^0\right\} + O\left(\sqrt{1+t}\right) = t\, \frac{\partial^2\lambda[0,\xi\varphi]}{\partial\xi_i\partial\xi_j}\bigg|_{\xi=0}.$$

Thus (3.37) is proved. The C^N-smoothness of $a(x)$ follows from Proposition 3.4. □

Proposition 3.14. *Consider the vector fields $u_i = u_i(x, y, 0)$, $a_i = a_i(x, y, 0)$ from the system* (1.1) *and $\tilde{V}(x, y) = \int_0^\infty A_0^t \tilde{B}(x, y)\,dt$, where $\tilde{B}$ is defined by* (3.4). *Let $\varphi = (\varphi_1, \dots, \varphi_n) \in (C^{N+2}(X_1))^n$ and $\xi = (\xi_1, \dots, \xi_n) \in \mathbb{R}^n$. Then the function $\lambda[0, \xi\varphi]$ from Proposition 3.3 satisfies the equality*

$$\begin{aligned}\frac{\partial^2\lambda[0,\xi\varphi]}{\partial\xi_j\partial\xi_k}\bigg|_{\xi=0} &= \bar{A}\Bigg(\sum_{i=1}^m \frac{d\varphi_j}{du_i}\frac{d\varphi_k}{du_i} + \frac{d\varphi_j}{d\tilde{B}}\frac{d\varphi_k}{d\tilde{V}} + \frac{d\varphi_k}{d\tilde{B}}\frac{d\varphi_j}{d\tilde{V}}\\ &\quad + \sum_{i=1}^m \frac{d\varphi_j}{du_i}\cdot\frac{d\varphi_k}{dx}\left(\frac{d\tilde{V}}{dy}a_i\right) + \sum_{i=1}^m \frac{d\varphi_k}{du_i}\cdot\frac{d\varphi_j}{dx}\left(\frac{d\tilde{V}}{dy}a_i\right)\Bigg).\end{aligned} \tag{3.42}$$

Proof. By Proposition 3.3, $\lambda[0,0] \equiv 0$, $h[0,0] \equiv 1$, and $\bar{A}h[0,\xi\varphi] \equiv 1$. Hence,

$$\bar{A}\frac{\partial h[0,\xi\varphi]}{\partial\xi_j} = 0, \qquad \bar{A}\frac{\partial^2 h[0,\xi\varphi]}{\partial\xi_j\partial\xi_k} = 0. \tag{3.43}$$

Recall that by (3.11) and (3.12) for any function $\psi \in C^{N+2}(X_1)$

$$A_0^t[0,\psi]f(x,y) = \mathbf{E}_{x,y}\left(\exp\left\{\tilde{\psi}_t^0(x,y)\right\} f(x, y^0(t))\right), \tag{3.44}$$

$$\tilde{\psi}_t^0(x,y) = \int_0^t \frac{d\psi(x)}{dx}\left(\tilde{B}(x, y^0(s))\,ds + \sum_{i=1}^m u_i(x, y^0(s), 0)\,dw^i(s)\right), \tag{3.45}$$

where $(x, y^0(t))$ is a solution to the system (1.1) for $\varepsilon = 0$ with initial condition (x, y). Differentiating the identity $e^{t\lambda[0,\xi\varphi]}h[0,\xi\varphi] \equiv A_0^t[0,\xi\varphi]h[0,\xi\varphi]$ (from Proposition 3.3) in ξ_j when $\xi = 0$ and taking into account (3.15), (3.44), (3.45) we obtain

$$\begin{aligned}\frac{\partial h[0,\xi\varphi]}{\partial\xi_j}\bigg|_{\xi=0} &= \mathbf{E}_{x,y}\left(\widetilde{(\varphi_j)}_t^0(x,y)\right) + A_0^t\,\frac{\partial h[0,\xi\varphi]}{\partial\xi_j}\bigg|_{\xi=0}\\ &= \int_0^t \frac{d\varphi_j(x)}{dx}\mathbf{E}_{x,y}\tilde{B}(x, y^0(s))\,ds + A_0^t\,\frac{\partial h[0,\xi\varphi]}{\partial\xi_j}\bigg|_{\xi=0}\\ &= \int_0^t \frac{d\varphi_j(x)}{dx}A_0^s\tilde{B}(x,y)\,ds + A_0^t\,\frac{\partial h[0,\xi\varphi]}{\partial\xi_j}\bigg|_{\xi=0}.\end{aligned}$$

Passing to the limit here as $t \to \infty$ and combining the equalities $\bar{A} = \lim_{t\to\infty} A_0^t$ and (3.43) we derive

$$\frac{\partial h[0, \xi\varphi]}{\partial \xi_j}\bigg|_{\xi=0} = \int_0^\infty \frac{d\varphi_j(x)}{dx} A_0^s \tilde{B}(x, y)\, ds = \frac{d\varphi_j(x)}{dx} \tilde{V}(x, y). \tag{3.46}$$

Consider the identity $e^{t\lambda[0,\xi\varphi]} \equiv \bar{A} A_0^t[0, \xi\varphi] h[0, \xi\varphi]$. Differentiating it in ξ_j we get

$$te^{t\lambda[0,\xi\varphi]} \frac{\partial \lambda[0, \xi\varphi]}{\partial \xi_j} = \bar{A}\left(\frac{\partial A_0^t[0, \xi\varphi]}{\partial \xi_j} h[0, \xi\varphi] + A_0^t[0, \xi\varphi] \frac{\partial h[0, \xi\varphi]}{\partial \xi_j} \right).$$

Next, differentiating this expression in ξ_k at the point $\xi = 0$ we have

$$t \frac{\partial^2 \lambda[0, \xi\varphi]}{\partial \xi_j \partial \xi_k}\bigg|_{\xi=0} = \bar{A}\left(\frac{\partial^2 A_0^t[0, \xi\varphi]}{\partial \xi_j \partial \xi_k}(1) + \frac{\partial A_0^t[0, \xi\varphi]}{\partial \xi_j} \frac{\partial h[0, \xi\varphi]}{\partial \xi_k} + \right.$$
$$\left. + \frac{\partial A_0^t[0, \xi\varphi]}{\partial \xi_k} \frac{\partial h[0, \xi\varphi]}{\partial \xi_j} + A_0^t \frac{\partial^2 h[0, \xi\varphi]}{\partial \xi_j \partial \xi_k} \right)\bigg|_{\xi=0}.$$

Using (3.43), (3.44), and (3.46) we conclude from the last equality that

$$t\, \frac{\partial^2 \lambda[0, \xi\varphi]}{\partial \xi_j \partial \xi_k}\bigg|_{\xi=0} = \bar{A}\mathbf{E}_{x,y}\left(\widetilde{(\varphi_j)}_t^0 \cdot \widetilde{(\varphi_k)}_t^0 \right) + \bar{A}\mathbf{E}_{x,y}\left(\widetilde{(\varphi_j)}_t^0 \cdot \frac{d\varphi_k(x)}{dx} \tilde{V}(x, y^0(t)) \right.$$
$$\left. + \widetilde{(\varphi_k)}_t^0 \cdot \frac{d\varphi_j(x)}{dx} \tilde{V}(x, y^0(t)) \right). \tag{3.47}$$

Differentiation of (3.47) in t for $t = 0$ by means of the Itô formula and (3.45) gives us (3.42). □

Now let us make some supplementary remarks, which can be proved easily by the methods developed above.

Remark 3.15. In the special case when the fast motion y^ε does not depend on the slow one x^ε, i.e. when the coefficients of the second equation in (1.1) do not depend on the slow variable x, the formula (3.37) for $a(x)$ reduces to the formula (3.2) obtained in [15].

Remark 3.16. If we replace $\tilde{\varphi}_t^\varepsilon = \int_0^t \tilde{D}_\varepsilon \varphi$ by $\varphi_t^\varepsilon = \int_0^t D_\varepsilon \varphi$ in the definition of the operator $A_\varepsilon^t[F, \varphi]$, then we get immediately the central limit theorem and the Cramer asymptotics for $\varphi(x^\varepsilon(t))$, where $t \le T/\varepsilon$ (see [1]–[2]). In this case we have in (3.14) the equality $\frac{d\lambda[0,\xi\varphi]}{d\xi}|_{\xi=0} = \frac{d\varphi}{d\bar{B}}$ and respectively $\dot{F}_{1t} = \frac{d\varphi}{d\bar{B}} + \frac{dF_{1t}}{d\bar{B}}$.

Remark 3.17. The function $\lambda[F, \varphi]$ is convex with respect to the pair (F, φ).

4. Diffusion approximation

Since we are interested in uniform in time estimates we always consider in this section continuous in time modifications of the processes and stochastic integrals in question. We will estimate the distance between the processes z^ε and r^ε using first the standard moment inequalities for stochastic integrals and then employ Corollary 3.9 which shows that as $\varepsilon \to 0$ the characteristic functions of increments $x^\varepsilon(t) - x^\varepsilon(s) + \int_s^t \bar{B}(x^\varepsilon(u))du$ become close to the Gaussian ones. This together with the technique from [5], [8], and [19] extended for the averaging setup in [17] will enable us to construct on a richer probability space certain Gaussian process which is sufficiently close in the L^2-sense both to z^ε and to the diffusion r^ε solving (1.6).

Assuming that the vector fields v, u_i from system (1.1) vanish outside some compact neighborhood X_0 of an averaged path $\{\bar{z}(t) \mid 0 \le t \le T\}$, we have established in Proposition 3.12 that $\mathbf{P}\left(\sup_{0\le t\le T} d_X(z^\varepsilon(t), \bar{z}(t)) \ge \delta\right)$ is exponentially small for any $\delta > 0$. So the behaviour of v and u_i outside a δ-vicinity of the averaged path affects the process $\{z^\varepsilon(t) \mid 0 \le t \le T\}$ only with exponentially small probability. We shall see soon that the same is true for the process $r^\varepsilon(t)$ from (1.6). Hence the desired estimate (1.7) does not depend on the size of X_0, provided this size is finite, and it is enough to prove (1.7) for *any* X_0. Moreover, if we allow the vector fields v, u_i not to have compact supports but to grow at most lineary in the distance from some point in X, this still has no affect on (1.7). Though this can be proved by means of usual large deviations estimates, we will not go into details and proceed in the setting of compact supports contained in small X_0. Then we may work in local coordinates and it is enough to prove Theorem 2.1 for the case when $X = \mathbb{R}^n$ which we assume in this section.

Taking $\varphi(x) = x = (x^1, \dots, x^n)$ and denoting $\tilde{\varphi}_t^\varepsilon(x, y)$ in this case by $\tilde{x}_t^\varepsilon(x, y)$ we obtain from (3.36) that

$$x^\varepsilon(t) - x^\varepsilon(0) = \varepsilon\tilde{x}_t^\varepsilon(x, y) + \varepsilon \int_0^t \bar{B}(x^\varepsilon(s))\, ds. \tag{4.1}$$

The matrix $a(x)$ from Proposition 3.13 is bounded in C^N-norm and it is symmetric and nonnegatively definite. Hence, by Theorem 2.1 from § 3.2 in [9] there exists a bounded Lipschitz continuous symmetric square root $\sigma(x)$ of $a(x)$, i.e. $a(x) = (\sigma(x))^2 = \sigma(x)\sigma^*(x)$. In what follows we consider (1.6) with such σ and

$$\eta(x) = \frac{1}{2}\frac{1}{\sqrt{g(x)}} \sum_{i,j=1}^n \frac{\partial}{\partial x^i}\left(a^{ij}(x)\sqrt{g(x)}\right)\frac{\partial}{\partial x^j}. \tag{4.2}$$

Then the diffusion $r^\varepsilon(t)$ defined by (1.6) has the generator (2.2) written in a manifold invariant form. Taking initial conditions $z^\varepsilon(0) = r^\varepsilon(0) = x$ we obtain from (4.1), (1.6) that

$$\begin{aligned} z^\varepsilon(t) - r^\varepsilon(t) = &\int_0^t \left(\bar{B}(z^\varepsilon(s)) - \bar{B}(r^\varepsilon(s))\right) ds + \varepsilon\tilde{x}_{t/\varepsilon}^\varepsilon(x, y) \\ &-\sqrt{\varepsilon}\int_0^t \sigma(\bar{z}(s))\, d\tilde{w}(s) - R^\varepsilon(t) - \tilde{R}^\varepsilon(t), \end{aligned} \tag{4.3}$$

where $R^\varepsilon(t) = \sqrt{\varepsilon}\int_0^t (\sigma(r^\varepsilon(s)) - \sigma(\bar{z}(s)))\, d\tilde{w}(s)$ and $\tilde{R}^\varepsilon(t) = \varepsilon\int_0^t \eta(r^\varepsilon(s))\, ds$. Recall, that $\bar{B} \in C^N(X_1)$ (see remark following (3.4)) and so it is Lipschitz continuous. Hence (4.3) and the Gronwall inequality yield

$$\sup_{0\le t\le T}\left|z^\varepsilon(t) - r^\varepsilon(t)\right| \le e^{CT}\left(\sup_{0\le t\le T}\left|\varepsilon\tilde{x}^\varepsilon_{t/\varepsilon}(x,y) - \sqrt{\varepsilon}\int_0^t \sigma(\bar{z}(s))\, d\tilde{w}(s)\right| \right.$$
$$\left. + \sup_{0\le t\le T}\left|R^\varepsilon(t)\right| + \sup_{0\le t\le T}\left|\tilde{R}^\varepsilon(t)\right|\right) \tag{4.4}$$

for some $C > 0$ independent of ε.

By Lipschitz continuity of σ and standard martingale estimates of stochastic integrals,

$$\mathbf{E}\sup_{0\le t\le T}\left|R^\varepsilon(t)\right|^2 \le C\varepsilon\int_0^T \mathbf{E}\left|r^\varepsilon(t) - \bar{z}(t)\right|^2 dt \tag{4.5}$$

for some $C > 0$. Next, by Lipschitz continuity of $\bar{B}$ and boundedness of η we obtain from (1.4) and (1.6) that

$$\left|r^\varepsilon(t) - \bar{z}(t)\right| \le C\int_0^t \left|r^\varepsilon(s) - \bar{z}(s)\right| ds + Ct\varepsilon + \sqrt{\varepsilon}\left|\int_0^t \sigma(r^\varepsilon(s))\, d\tilde{w}(s)\right| \tag{4.6}$$

for some $C > 0$, and so by Gronwall's inequality

$$\left|r^\varepsilon(t) - \bar{z}(t)\right| \le e^{Ct}\left(Ct\varepsilon + \sqrt{\varepsilon}\sup_{0\le s\le t}\left|\int_0^s \sigma(r^\varepsilon(u))\, d\tilde{w}(u)\right|\right). \tag{4.7}$$

Taking squares of both parts of (4.7), applying the expectation, using moment inequalities for stochastic integrals and taking into account (4.5) we arrive at

$$\mathbf{E}\sup_{0\le t\le T}\left|R^\varepsilon(t)\right|^2 \le 4C^2T^2e^{2CT}\varepsilon^2 \tag{4.8}$$

for some $C > 0$ independent of ε. Clearly, we have also

$$\mathbf{E}\sup_{0\le t\le T}\left|\tilde{R}^\varepsilon(t)\right|^2 \le \tilde{C}^2T^2\varepsilon^2 \tag{4.9}$$

for some $\tilde{C} > 0$ independent of ε.

It remains to deal with the first term in the right hand side of (4.4) where we will have to pick up an appropriate Brownian motion $\tilde{w}$ in order to obtain the desired estimates. Denote by $\mathcal{F}_t$ the $\sigma-$algebra generated by $\{w(s) \mid 0 \le s \le t\}$. By the Markov property for any $t > s \ge 0$ and $\zeta = (\zeta_1, \ldots, \zeta_n) \in \mathbb{R}^n$,

$$\mathbf{E}_{x,y}\left(\exp\left\{i\sqrt{\varepsilon}\left\langle\zeta,\, \tilde{x}^\varepsilon_t(x,y) - \tilde{x}^\varepsilon_s(x,y)\right\rangle\right\}\middle|\mathcal{F}_s\right)$$
$$= \mathbf{E}_{x^\varepsilon(s),y^\varepsilon(s)}\exp\left\{i\sqrt{\varepsilon}\left\langle\zeta,\, \tilde{x}^\varepsilon_{t-s}(x^\varepsilon(s), y^\varepsilon(s))\right\rangle\right\} \tag{4.10}$$

where, again, $\langle \zeta, x\rangle = \zeta_1 x^1 + \cdots + \zeta_n x^n$. The matrix $a(x)$ from Proposition 3.13 is Lipschitz continuous and nonnegatively definite, and so is the matrix $F_{2,t}(x)$ from Corollary 3.9. Then, using the inequality $|e^{-a} - e^{-b}| \le |b-a|$, where $a, b > 0$, we obtain that for any $0 \le s \le t \le T$,

$$\begin{aligned}\mathbf{E}_{x,y}\Big|&\exp\big\{-\tfrac{1}{2}\big\langle F_{2,t-s}(z^\varepsilon(s))\zeta, \zeta\big\rangle\big\} - \exp\big\{-\tfrac{1}{2}\big\langle F_{2,t-s}(\bar z(s))\zeta, \zeta\big\rangle\big\}\Big|\\ &\le C(t-s)|\zeta|^2 \mathbf{E}_{x,y}\big|z^\varepsilon(s) - \bar z(s)\big|. \end{aligned}\tag{4.11}$$

Now (4.10), (4.11) together with Corollary 3.9 and (3.35) yield

$$\begin{aligned}\mathbf{E}_{x,y}\Big|&\mathbf{E}_{x,y}\Big(\exp\big\{i\sqrt{\varepsilon}\big\langle\zeta,\ \tilde x^\varepsilon_{t/\varepsilon}(x,y) - \tilde x^\varepsilon_{s/\varepsilon}(x,y)\big\rangle\big\}\Big|\mathcal{F}_{s/\varepsilon}\Big)\\ &\quad - \exp\big\{-\tfrac{1}{2}\big\langle F_{2,t-s}(\bar z(s))\zeta, \zeta\big\rangle\big\}\Big|\\ &\le C\Big(\sqrt{\varepsilon}|\zeta| + \sqrt{\varepsilon}(t-s)|\zeta|^3 + (t-s)|\zeta|^2\sqrt{\varepsilon}\ln\varepsilon^{-1}\Big)\end{aligned}\tag{4.12}$$

for some $C = C(T)$.

Take some small $\tau > 0$ and set

$$U_k = U_k(\tau,\varepsilon) = \sqrt{\varepsilon}\big(\tilde x^\varepsilon_{k\tau/\varepsilon}(x,y) - \tilde x^\varepsilon_{(k-1)\tau/\varepsilon}(x,y)\big),\tag{4.13}$$

$$H_k(\tau) = \int_{(k-1)\tau}^{k\tau} \sigma(\bar z(t))\,dW(t),\qquad k = 1,\ 2,\ \ldots,\ [T/\tau],\tag{4.14}$$

where $W(t)$ is a standard continuous n-dimensional Brownian motion. Clearly, $H_k(\tau)$ has the characteristic function

$$\exp\left\{-\frac{1}{2}\int_{k\tau-\tau}^{k\tau}\big\langle a(\bar z(t))\zeta, \zeta\big\rangle\,dt\right\} = \exp\big(-\tfrac{1}{2}\big\langle F_{2,\tau}(\bar z(k\tau - \tau))\zeta, \zeta\big\rangle\big)\tag{4.15}$$

and so it satisfies

$$\mathbf{P}\left\{H_k(\tau) \ge \tfrac{1}{4}L\right\} \le \exp(-C_T L^2/\tau)\tag{4.16}$$

for any $L > 0$, where $C_T > 0$ depends only on T. This together with (4.12), (4.15) and Theorem 4.6 from [7] (or with its older version Theorem 1 from [5]) yield that we can redefine the sequence $U_k,\ k = 1, 2, \ldots, [T/\tau]$ on a richer probability space where there exists a sequence $V_k = V_k(\tau),\ k = 1, 2, \ldots, [T/\tau]$ of independent random vectors such that V_k has the same distribution as $H_k(\tau)$ and

$$\mathbf{P}\big\{|U_k - V_k| \ge \beta\big\} \le \beta,\qquad k = 1,\ 2,\ \ldots,\ [T/\tau],\tag{4.17}$$

where

$$\begin{aligned}\beta &= 16nL^{-1}\log L + 4\sqrt{\mu}L^n + \exp(-C_T L^2/\tau),\\ \mu &= C\sqrt{\varepsilon}\big(L + L^3\tau + L^2\tau\ln\varepsilon^{-1}\big).\end{aligned}\tag{4.18}$$

It follows easily that for any $R > 0$ and $p > 1$ (see (4.23) in [17]),

$$\mathbf{E}|U_k - V_k|^2 \le \varkappa = \beta^2 + 4R^2\beta + R^{-2(p-1)}\big(\mathbf{E}|U_k|^{2p} + \mathbf{E}|V_k|^{2p}\big).\tag{4.19}$$

Since the sequences V_k and $H_k(\tau)$ have the same distributions the Kolmogorov existence theorem yields that for each ε and k we can redefine the Brownian motion in (1.1) without changing its distribution on a richer probability space where there exists another n-dimensional Brownian motion $W(t)$ such that $U_k = U_k(\tau, \varepsilon)$ defined by (4.13) and $H_k(\tau)$ defined by (4.14) via these new processes satisfy

$$\mathbf{E}|U_k - H_k(\tau)|^2 \le \varkappa. \tag{4.20}$$

In fact, according to [8] and [19] (see also the discussion concerning this question on p. 551 in [20]) if the original probability space is already rich enough then there is no need in redefining both the sequence $U_k,\ k = 1, 2, \ldots, [T/\tau]$ and the Brownian motion from (1.1) and, in any case, it is possible to enrich the original probability space in the specific way multiplying it by the interval [0, 1] and redefining the above processes explicitly via projections as described in Theorem 2.1.

Now we have

$$\begin{aligned} I &= \sup_{0\le t\le T}\left|\sqrt{\varepsilon}\tilde{x}^{\varepsilon}_{t/\varepsilon}(x,y) - \int_0^t \sigma(\bar{z}(s))\,dW(s)\right| \\ &\le \sum_{k=1}^{[T/\tau]} |U_k - H_k(\tau)| + \max_{0\le k\le[T/\tau]} J_k^{(1)}(\tau) + \max_{0\le k\le[T/\tau]} J_k^{(2)}(\tau,\varepsilon), \end{aligned} \tag{4.21}$$

where

$$J_k^{(1)}(\tau) = \sup_{0\le\theta\le\tau}\left|\int_{k\tau}^{k\tau+\theta} \sigma(\bar{z}(t))\,dW(t)\right|,$$

$$J_k^{(2)}(\tau,\varepsilon) = \sup_{k\tau\le t\le k\tau+\tau} \sqrt{\varepsilon}\left|\tilde{x}^{\varepsilon}_{t/\varepsilon}(x,y) - \tilde{x}^{\varepsilon}_{k\tau/\varepsilon}(x,y)\right|.$$

By (4.20),

$$\mathbf{E}\left(\sum_{k=1}^{[T/\tau]}\left|U_k - H_k(\tau)\right|\right)^2 \le [T/\tau]^2\varkappa. \tag{4.22}$$

Employing standard martingale moment estimates for stochastic integrals (see, for instance, [14]) we obtain that for any $p \ge 1$,

$$\mathbf{E}\left|J_k^{(1)}(\tau)\right|^{2p} \le C\tau^p \tag{4.23}$$

for some $C > 0$. It follows from (3.32) that for any $p \ge 1$,

$$\mathbf{E}\left|J_k^{(2)}(\tau,\varepsilon)\right|^{2p} \le C(\varepsilon+\tau)^p \ln^{2p}(2+\tau/\varepsilon) \tag{4.24}$$

for some $C > 0$.

It is easy to derive from (4.23) and (4.24) similarly to (4.33) and (4.34) in [17] that for any $Q > 0$ and $p \ge 1$,

$$\mathbf{E}\max_{0\le k\le[T/\tau]}\left|J_k^{(1)}(\tau)\right|^2 \le Q + C_p Q^{-(p-1)}[T/\tau]\tau^p \tag{4.25}$$

and

$$\mathbf{E}\max_{0\le k\le[T/\tau]}\left|J_k^{(2)}(\tau,\varepsilon)\right|^2\le Q+C_pQ^{-(p-1)}[T/\tau](\varepsilon+\tau)^p\ln^{2p}(2+\tau/\varepsilon) \tag{4.26}$$

for some $C_p>0$ independent of $\varepsilon,\ Q,\ \tau$. It follows from (4.21), (4.22), (4.25), (4.26) that

$$\mathbf{E}I^2\le C(p,T)\rho_\varepsilon, \tag{4.27}$$

where

$$\rho_\varepsilon=\tau^{-2}\varkappa+Q+Q^{-(p-1)}\tau^{-1}(\varepsilon+\tau)^p\ln^{2p}(2+\tau/\varepsilon). \tag{4.28}$$

Employing (4.13) and (3.31) we obtain that

$$\mathbf{E}|U_k|^{2p}\le C(\varepsilon+\tau)^p$$

and by martingale moment inequalities we see also that

$$\mathbf{E}|V_k|^{2p}\le C\tau^p$$

for some $C=C(p)>0$. Hence, by (4.19) and (4.28) we get the estimate

$$\begin{aligned}\rho_\varepsilon\le C\Big(&\tau^{-2}\big(\beta^2+4R^2\beta+R^{-2(p-1)}(\varepsilon+\tau)^p\big)\\&+Q+Q^{-(p-1)}\tau^{-1}(\varepsilon+\tau)^p\ln^{2p}(2+\tau/\varepsilon)\Big).\end{aligned} \tag{4.29}$$

Now we have to minimize ρ_ε by appropriate choice of the parameters τ, R, Q, p here and L in (4.18). To do this, let us take

$$\delta_0=(18+8n)^{-1},\ \tau=\varepsilon^{\delta_0},\ Q=\tau^{1-\gamma},\ R=\tau^{1/2-\gamma},\ p=\gamma^{-1}+1,\ L=\tau^{-2},$$

where $\gamma>0$ is arbitrarily small. It is easy to check in this case that $\beta<\varepsilon^{2\delta_0-\gamma}$ and $\rho_\varepsilon<\varepsilon^{\delta_0-2\gamma}$ as $\varepsilon\to 0$. So Theorem 2.1 is proved. □

Remark 4.1. The right hand side of (4.29) cannot be made less than ε^{δ_0} by order. Indeed, omitting the constant C, we may write it in the form

$$\begin{aligned}\rho=\frac{\beta^2}{\tau^2}+\frac{4R^2\beta}{\tau^2}+\Big(\frac{\varepsilon+\tau}{R^2}\Big)^{p-2}\Big(\frac{\varepsilon}{\tau}+1\Big)^2\frac{1}{R^2}\\+Q+\Big(\frac{\tau}{Q}\Big)^{p-1}\Big(\frac{\varepsilon}{\tau}+1\Big)^p\ln^{2p}(2+\tau/\varepsilon).\end{aligned}$$

Observing the last term here, we conclude that if ρ is small then $\tau<Q$. Observing the third term, we see that R^2 must be greater than τ. Hence the second term is greater than $\tau^{-1}\beta$. From (4.18) it follows that $\beta>L^{-1}\log L+\sqrt{\mu}L^n>L^{-1}+\varepsilon^{1/4}\sqrt{L^3\tau}L^n$. This inequality implies $\beta>\varepsilon^{\frac{1}{4n+10}}\tau^{\frac{1}{2n+5}}$ for any positive L. Thus, we have $\rho>Q+\tau^{-1}\beta>\tau+\varepsilon^{\frac{1}{4n+10}}\tau^{\frac{-2n-4}{2n+5}}$ and so $\rho>\varepsilon^{\frac{1}{8n+18}}$ for any positive τ.

Finally, in order to establish Corollary 2.2 we have to obtain (2.5). Taking in (1.6) three terms of the Taylor expansion of $\bar{B}$ and subtracting (1.4), (2.4) we obtain that

$$d\left(r^{\varepsilon}(t)-\bar{z}(t)-\sqrt{\varepsilon}g(t)\right)=\nabla\bar{B}(\bar{z}(t))\left(r^{\varepsilon}(t)-\bar{z}(t)-\sqrt{\varepsilon}g(t)\right)dt$$
$$+O\left(|r^{\varepsilon}(t)-\bar{z}(t)|^{2}\right)dt+\varepsilon\eta(r^{\varepsilon}(t))\,dt+\sqrt{\varepsilon}\big(\sigma(r^{\varepsilon}(t))-\sigma(\bar{z}(t))\big)\,d\tilde{w}(t).$$

Then by the Gronwall inequality,

$$\sup_{0\le t\le T}\left|r^{\varepsilon}(t)-\bar{z}(t)-\sqrt{\varepsilon}g(t)\right|\le Ce^{CT}\int_0^T\left|r^{\varepsilon}(t)-\bar{z}(t)\right|^2dt$$
$$+Ce^{CT}\sup_{0\le t\le T}\left|\varepsilon\int_0^t\eta(r^{\varepsilon}(s))\,ds+\sqrt{\varepsilon}\int_0^t\big(\sigma(r^{\varepsilon}(s))-\sigma(\bar{z}(s))\big)\,d\tilde{w}(s)\right|.$$

Taking the square and employing (4.7), (4.8) and (4.9) we arrive at (2.5). □

References

[1] Bakhtin, V.I.: Cramér asymptotics in a system with slow and fast Markovian motions. Theory Probab. Appl. **44**, 1–17 (1999)
[2] Bakhtin, V.I.: Asymptotics of superregular perturbations of fiber ergodic semigroups. Stochastics and Stochastics Reports **75**, 295–318 (2003)
[3] Bakhtin, V.I.: Cramér asymptotics in a system with fast and slow motions. Stochastics and Stochastics Reports **75**, 319–341 (2003)
[4] Billingsley, P.: Convergence of Probability Measures. 2nd (ed.), J. Wiley, New York, 1999
[5] Berkes, I., Philipp, W.: Approximation theorems for independent and weakly dependent random vectors. Ann. Probab. **7**, 29–54 (1979)
[6] Doukhan, P.: Mixing. Springer-Verlag, New York, 1994
[7] Dehling, H., Philipp, W.: Empirical process techniques for dependent data. In: Empirical Process Techniques for Dependent Data H.G. Dehling, T. Mikosch, M. Sorensen, (eds.), Birkhäuser, 2002
[8] Dudley, R.M., Philipp, W.: Invariance principles for sums of Banach space valued random elements and empirical processes. Z. Wahrscheinlichkeitstheor. Verw. Geb. **62**, 509–552 (1983)
[9] Freidlin, M.I.: Functional Integration and Partial Differential Equations. Princeton Univ. Press, Princeton, 1985
[10] Freidlin, M.I., Wentzell, A.D.: Random Perturbations of Dynamical Systems. 2nd ed., Springer-Verlag, New York, 1998
[11] Gray, A.: Tubes, Addison-Wesley, Redwood City, Ca., 1990
[12] Hasselmann, K.: Stochastic climate models. Part I. Theory. Tellus **28**, 473–485 (1976)
[13] Hsu, E.P.: Stochastic analysis on manifolds. American Math. Soc., Providence, RI, 2002
[14] Ikeda, N., Watanabe, S.: Stochastic Differential Equations and Diffusion Processes. North-Holland, Amsterdam (Kodansha Ltd. Tokyo), 1981
[15] Khasminskii, R.Z.: On stochastic processes defined by differential equations with a small parameter. Th. Probab. Appl. **11**, 211–228 (1966)
[16] Kifer, Y.: Averaging and climate models. Stochastic Climate Models, Progress in Probability **49** Birkhäuser, 2001

[17] Kifer, Y.: L^2 diffusion approximation for slow motion in averaging. Stochastics and Dynamics **3**, 213–246 (2003)
[18] Kunita, H.: Stochastic Flows and Stochastic Differential Equations. Cambridge Univ. Press, Cambridge, 1990
[19] Monrad, D., Philipp, W.: Nearby variables with nearby conditional laws and a strong approximation theorem for Hilbert space valued martingales. Probab. Th. Rel. Fields **88**, 381–404 (1991)
[20] Philipp, W.: The profound impact of Paul Erdoes on probability limit theory - a personal account. In: "Paul Erdös and his Mathematics", **I**, pp. 549–566, Budapest, 2002
[21] Philipp, W., Stout, W.: Almost Sure Invariance Principles for Partial Sums of Weakly Dependent Random Variables. Mem. Amer. Math. Soc., **161** Providence, R.I., 1975