Теорія ймовірностей
та математична статистика
Вип. 2(101)/2019, с. 15–39

Teoriya Ĭmovirnosteĭ
ta Matematychna Statystyka
No. 2(101)/2019, pp. 15–39

UDC 519.21

IMPROVED LOCAL APPROXIMATION FOR MULTIDIMENSIONAL DIFFUSIONS: THE G-RATES

S. BODNARCHUK, D. IVANENKO, A. KOHATSU-HIGA, A. KULIK

Abstract. In this article, we consider the problem of improving the local approximations for multidimensional diffusions. In particular, our proposed explicit approximation improves the Milshtein approximation. We also provide a semi-explicit convergence rate estimate (we call it G-rate) for the proposed local approximation. The main error term in the difference of densities is bounded by a polynomial multiplied by a Gaussian density and the remainder is exponentially small as time goes to zero.

Key words and phrases. Expansions, Stochastic Differential Equations, Total Variation Distance.

2010 *Mathematics Subject Classification.* Primary:60H35 .

1. Introduction

The main aim of this paper is to provide small time approximations for a multidimensional diffusion X, which is *explicit* and is an *improvement* of the Euler-Maruyama approximation for X. We will control the accuracy of approximation by means of the total variation distance.

To explain our setting, let X be the solution to the uniformly elliptic d-dimensional SDE

$$\mathrm{d}X^x(t) = a(X^x(t))\,\mathrm{d}t + \sigma(X^x(t))\,\mathrm{d}W(t), \quad X^x(0) = x. \tag{1.1}$$

Here, W is an m-dimensional Brownian motion, $a : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ are given functions. For the above equation, we define the Euler-Maruyama approximation to X as

$$\overline{X}^x(t) = x + a(x)t + \sigma(x)W(t). \tag{1.2}$$

The standard *weak* and *strong* approximation rates for (1.2) in small time are known to equal respectively 2 and 1 for sufficiently smooth coefficients a, σ; for a detailed discussion and references on this subject, see [7]. That is, for any sufficiently smooth test function $f : \mathbb{R}^d \to \mathbb{R}$,

$$\mathbf{E}f(X^x(t)) - \mathbf{E}f(\overline{X}^x(t)) = O(t^2), \quad t \to 0+, \tag{1.3}$$

and for any $p \geq 1$,

$$\Big(\mathbf{E}|X^x(t) - \overline{X}^x(t)|^p\Big)^{1/p} = O(t), \quad t \to 0+. \tag{1.4}$$

Many efforts have been devoted to the construction of *improved* approximation schemes which would exhibit better rates (for example, see [9] and [10]). In this paper we will refer to only one of them, the Milshtein scheme, which for the scalar case (i.e. $d = 1$) is given by

$$\overline{X}^{x,2}(t) = x + a(x)t + \sigma(x)W(t) + \frac{1}{2}\sigma'(x)\sigma(x)(W^2(t) - t). \tag{1.5}$$

This scheme has the strong rate 2 (i.e. $O(t^2)$ instead of $O(t)$ in (1.4)) and is given by the explicit formula (1.5). However, the Milshtein scheme is not explicit when extended to the multidimensional case because it requires the computation of double iterated stochastic integrals with respect to independent Brownian motions.

In order to describe a restrictive multidimensional situation where the Milshtein approximation is still explicit, we introduce first

$$\rho_{iuv}(x) = \sum_{l=1}^{d} \partial_{x_l} \sigma_{iu}(x)\sigma_{lv}(x), \quad i = 1, \dots, d, \quad u, v = 1, \dots, m$$

(∂_{x_l} denotes the partial derivative w.r.t. x_l) and define the $\mathbb{R}^d$-valued process $\overline{X}_t^{x,2}$ as

$$\overline{X}^{x,2}(t) = x + a(x)t + \sigma(x)W(t) + \frac{1}{2}\sum_{i=1}^{d}\sum_{u,v=1}^{m} \rho_{iuv}(x)\Big(W_u(t)W_v(t) - 1_{u=v}t\Big)\mathbf{e}_i, \quad (1.6)$$

where $\mathbf{e}_i, i = 1, \dots, d$ are the coordinate vectors in $\mathbb{R}^d$.

It is known that, under the *commutativity condition*

$$\rho_{iuv}(x) = \rho_{ivu}(x), \quad i = 1, \dots, d, \quad u, v = 1, \dots, m, \quad (1.7)$$

the Milshtein scheme (1.6) has the strong rate 2. This however is not true in general: if (1.7) fails, then (1.6) has the strong rate 1, which is the same rate as the Euler-Maruyama scheme (1.2).

The G-rates studied in the paper (see Definition 2.2 below) provide error bounds for the approximation of the law of $X^x(t)$ in a sense which is stronger than in the Total Variation (TV) distance. Therefore, in particular, our results provide error bounds for the approximation of $\mathbf{E}f(X^x(t))$ with measurable functions f of exponential growth at infinity, which is a stronger statement than those provided by weak rates (1.3) and is linked to strong rates (1.4) (see e.g. [2]). A good illustration of our main result is the following:

Theorem 1.1. *Let $d = m$, $a \in C_b^1(\mathbb{R}^d, \mathbb{R}^d)$, $\sigma \in C_b^2(\mathbb{R}^d, \mathbb{R}^{d\times d})$ and $b = \sigma\sigma^*$ be uniformly elliptic. Then the TV-distance between the laws of $X^x(t)$ and $\overline{X}^{x,2}(t)$ converges to zero as $O(t)$ when $t \to 0+$.*

Let us stress that this theorem does not require the commutativity assumption (1.7).

It is not difficult to see (cf. Remark 3.1 below) that the G-rate for the Euler-Maruyama approximation (1.2) is 1/2, hence Theorem 1.1 states that the scheme (1.6) improves the G-rate by 1/2. Summarizing the above discussion, we conclude that (1.6) is an *explicit* scheme, which *improves* the Euler-Maruyama scheme (1.2) when using the G-rate as a measure of the error.

The scheme with such properties is not unique, and one of our main goals is to describe an algorithm that derives many schemes of this type. The first step in order to obtain such an algorithm is the improved approximation for the transition density of the process X; see Theorem 3.1 below. Such an approximation is of a separate interest and has clear applications. For example, in statistical procedures for diffusion processes (see e.g. [1]) or Monte Carlo simulations in the multi-dimensional case.

We remark that the expansions provided here are in essence different from the ones in [3] where geometrical concepts together with implicit solutions of the Kolmogorov equations are used in order to design theoretical expansions of densities. These ideas have been extensively used in statistical applications such as [1]. Note that in Malliavin calculus a theory of expansions is also available as in [14] and [15] with statistical applications provided in [16]. This theory seems difficult to apply in the present situation given the limited regularity that we impose in our current setting.

The second main step consists in studying polynomial transformations of the Euler-Maruyama scheme whose image measure may be asymptotically equal to the expansion obtained in the first step in Theorem 3.1. This will require the use of Ramer's theorem in order to carry out the change of variables on Gaussian measures and the final result is Theorem 3.2. Finally, after obtaining the two expansions, we just need to match them in order to obtain the improved approximation.

The structure of the rest of the paper is as follows. In Section 2 we introduce notation and give necessary preliminaries. In Section 3, we state the main results of the paper, Theorems 3.1 and 3.2. Section 4 and Section 5 contain the proofs of Theorem 3.1 and Theorem 3.2, respectively. Theorem 1.1 is proved at the end of Section 5. In the Appendix, we provide a series of properties of Hermite polynomials and some additional comments on the interpretation of the divergence operator within this approximation problem.

This paper can be considered as a first step towards a larger subject which aims at obtaining general explicit G-expansions (like the one proposed in Definition 2.2) of higher order for the multi-dimensional densities of X and therefore high order approximations which will improve the Euler-Maruyama scheme. These issues as well as numerical implementations may be discussed in future research.

2. Notation and preliminaries

Notation: We will always work on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$, on which all random variables and processes that will appear from now on are well defined. Expectations will be denoted by $\mathbf{E}$.

Constants are usually denoted by C, c. As usual, they may change values from one line to another.

Functions with the same arguments may be aggregated as e.g. $\sigma\sigma^*(x) = \sigma(x)\sigma^*(x)$. The class of ϱ-Hölder continuous functions is denoted by C^ϱ, the class of differentiable functions with ϱ-Hölder continuous derivatives is denoted by $C^{1+\varrho}$. The derivative of function $f(x), x \in \mathbb{R}^d$ is denoted by $\nabla f(x)$. In the case of several space variables, we denote by $\nabla_1, \nabla_2, \dots,$ the derivatives with respect to the first, second, ... variables. The derivative in time is denoted separately as ∂_t. For example,

$$\nabla_2 f_t(x,y) = \left(\frac{\partial}{\partial y_1} f_t(x,y) \dots, \frac{\partial}{\partial y_d} f_t(x,y)\right), \quad \partial_t f_t(x,y) = \frac{\partial}{\partial t} f_t(x,y).$$

We denote by $\mathbb{R}^{d\times d}_{\text{sym}}$ the set of strictly positive definite matrices of order $d \times d$. For arbitrary vectors $\mathbf{a}, x, y \in \mathbb{R}^d$ and $\mathbf{b} \in \mathbb{R}^{d\times d}_{\text{sym}}$, we denote by

$$\Phi_t(\mathbf{a}, \mathbf{b}; x, y) = (2\pi t)^{-d/2}\Big(\det \mathbf{b}\Big)^{-1/2} \exp\left(-\frac{1}{2t}\Big(\mathbf{b}^{-1}(y - x - \mathbf{a}t), y - x - \mathbf{a}t\Big)\right), \quad (2.1)$$

the *Hermite function of order zero* with coefficients $\mathbf{a} \in \mathbb{R}^d, \mathbf{b} \in \mathbb{R}^{d\times d}_{\text{sym}}$. This is clearly the same as the multivariate density of a Brownian motion with drift $\mathbf{a} \in \mathbb{R}^d$ and covariance matrix $\mathbf{b} \in \mathbb{R}^{d\times d}_{\text{sym}}$. The *Hermite functions of higher orders* are defined as follows: for any vector of indexes $(i_1, \dots, i_n) \in \{1, \dots, d\}^n$,

$$\Phi_t^{(i_1,\dots,i_n)}(\mathbf{a}, \mathbf{b}; x, y) = \partial_{x_{i_1}} \dots \partial_{x_{i_n}} \Phi_t(\mathbf{a}, \mathbf{b}; x, y).$$

The corresponding *Hermite polynomials* are defined by

$$\Phi_t^{(i_1,\dots,i_n)}(\mathbf{a}, \mathbf{b}; x, y) = H_t^{(i_1,\dots,i_n)}(\mathbf{a}, \mathbf{b}; x, y)\Phi_t(\mathbf{a}, \mathbf{b}; x, y), \quad (2.2)$$

the *Hermite polynomial of order zero* identically equals 1. More information about Hermite polynomials appears in Section A in the Appendix.

When the $d \times d$-matrix $b(x) = \sigma\sigma(x)^*$ is positively defined, the distribution density of the Gaussian approximation (1.2) equals

$$p_t^{EM}(x,y) = \Phi_t(a(x), b(x); x, y). \quad (2.3)$$

With this observation in mind, we define the basic elements to be used in the further approximations for the transition density $p_t(x,y)$ of the process X. For a parameter $\xi\in\mathbb{R}^d$ define

$$\Phi_t(\xi;x,y):=\Phi_t(a(\xi),b(\xi);x,y),\quad \Phi_t^{(i_1,\dots,i_n)}(\xi;x,y):=\Phi_t^{(i_1,\dots,i_n)}(a(\xi),b(\xi);x,y),$$
$$H_t^{(i_1,\dots,i_n)}(\xi;x,y):=H_t^{(i_1,\dots,i_n)}(a(\xi),b(\xi);x,y).$$

In the case $\xi=x$ the above function corresponds to the density of the Euler–Maruyama scheme and for this reason, we use the simplified notation:

$$\Phi_t(x,y):=\Phi(x;x,y),\qquad \Phi_t^{(i_1,\dots,i_n)}(x,y):=\Phi_t^{(i_1,\dots,i_n)}(x;x,y),$$
$$H_t^{(i_1,\dots,i_n)}(x,y)=H_t^{(i_1,\dots,i_n)}(x;x,y).$$

Therefore, in particular,

$$p_t^{EM}(x,y)=\Phi_t(x,y). \tag{2.4}$$

Finally, we introduce two closely related notions about the rate at which time dependent Gaussian type kernels and time dependent functions behave as $t\downarrow 0$.

Definition 2.1. For a kernel $\Gamma_t(x,y)$, its t-order (order in time) is the maximal $p\in\mathbb{R}$ such that the following bound holds: for some[1] $c>0, C<\infty$

$$|\Gamma_t(x,y)|\le Ct^{p-d/2}\exp\left(-c\frac{|y-x|^2}{t}\right),\quad t\in(0,t_0],\quad x,y\in\mathbb{R}^d. \tag{2.5}$$

For a function $F_t(x,y)$ its t-order is the maximal $p\in\mathbb{R}$ such that, for arbitrary $\varepsilon>0$, there exists $C=C(\varepsilon)<\infty$ such that

$$|F_t(x,y)|\le Ct^{p}\exp\left(\varepsilon\frac{|y-x|^2}{t}\right),\quad t\in(0,t_0],\quad x,y\in\mathbb{R}^d. \tag{2.6}$$

The slight abuse of terminology (the same name 't-order' for two different notions) will not cause misunderstanding since we will systematically use different notation for kernels (capital Greek letters) and for functions (capital Roman letters). On the other hand, this terminology is natural and convenient, e.g if $F_t(x,y)$, $\Gamma_t(x,y)$ have t-orders p_1,p_2 then their product $\Xi_t(x,y)=F_t(x,y)\Gamma_t(x,y)$ has the t-order p_1+p_2. The same observation applies to other combinations of products between functions and kernels.

In some cases, we will need to estimate the kernels/functions which depend on additional parameters, such as $\xi,\mathbf{a}\in\mathbb{R}^d,\mathbf{b}\in\mathbb{R}^{d\times d}_{\mathrm{sym}}$ in the notation above. We will say that the kernel (resp. the function) has uniform t-order if (2.5) (resp. (2.6)) holds true for all values of the parameters with the same constants c,C (resp. family of constants $C(\varepsilon),\varepsilon>0$). In particular, the following properties hold true for any bounded set $A\subset\mathbb{R}^d$ and any compact subset B of the set of symmetric positively definite $d\times d$ matrices:

- $\Phi_t(\mathbf{a},\mathbf{b};x,y)$ and $\Phi_t^{(i_1,\dots,i_n)}(\mathbf{a},\mathbf{b};x,y)$ have respectively t-orders 0 and $-n/2$ uniformly for $\mathbf{a}\in A,\mathbf{b}\in B$;
- $H_t^{(i_1,\dots,i_n)}(\mathbf{a},\mathbf{b};x,y)$ has t-order $-n/2$ uniformly for $\mathbf{a}\in A,\mathbf{b}\in B$.

These statements can be easily obtained from the definition of the Hermite functions and polynomials together with the inequality $|x|^ke^{-\varepsilon|x|^2}\le C$ valid for all $x\in\mathbb{R}^d$ with some $C=C(k,\varepsilon)<\infty$. The idea of applying this inequality as well as the above t-order results are important in order to obtain many results in this work and will be used frequently without further reference.

[1] Constants will depend on the fixed value $t_0\in(0,1)$ (see Proposition 5.1 and the argument in (5.13)).

Clearly, knowing that kernel $\Gamma_t(x,y)$ has t-order p yields the integral bound

$$\int_{\mathbb{R}^d} |\Gamma_t(x,y)|\,\mathrm{d}y \le Ct^p.$$

However, the latter bound is much weaker than the Gaussian type estimate from the definition of the t-order. Below we will use these two types of estimates simultaneously in quite a different manner: one (the 'main', 'well structured') part of an approximation error will have a density of a given t-order, while the other (the 'nuisance', 'microstructural') part will decay much faster, but only in the TV sense. This motivates the following definition.

Definition 2.2. A family $\widetilde{X}^x(t), x \in \mathbb{R}^d, t > 0$ is said to approximate the law of the solution to (1.1) with G-rate p if there exists a decomposition[2]

$$\mathbf{P}(X^x(t) \in \mathrm{d}y) - \mathbf{P}(\widetilde{X}^x(t) \in \mathrm{d}y) = t^p\Gamma_t(x,y)\,\mathrm{d}y + \nu_{x,t}(\mathrm{d}y) \tag{2.7}$$

where $t^p\Gamma_t(x,y)$ has the t-order p and the signed measure $\nu_{x,t}$ is exponentially negligible as $t \to 0+$ in the TV sense. That is, for some $r > 0, c > 0, C < \infty$

$$|\nu_{x,t}|(\mathbb{R}^d) \le Ce^{-ct^{-r}}. \tag{2.8}$$

It is a consequence of this definition that the TV distance between the laws of $X^x(t)$ and $\widetilde{X}^x(t)$ converges to 0 as $O(t^p)$.

We use the following notation for space and space-time convolutions of functions $\Gamma_t(x,y), \Lambda_t(x,y)$

$$(\Gamma * \Lambda)_t(x,y) = \int_{\mathbb{R}^d} \Gamma_t(x,z)\Lambda_t(z,y)\mathrm{d}z,$$

$$(\Gamma \circledast \Lambda)_t(x,y) = \int_0^t \int_{\mathbb{R}^d} \Gamma_{t-s}(x,z)\Lambda_s(z,y)\mathrm{d}z\mathrm{d}s. \tag{2.9}$$

We also define the iterated convolutions as $\Gamma^{\circledast k}(x,y) = \Gamma \circledast \Gamma^{\circledast(k-1)}(x,y)$ with $\Gamma^{\circledast 1}(x,y) = \Gamma(x,y)$.

3. Main statements

Our first main result provides an expansion of the error between the densities of X and the Euler-Maruyama scheme.

Theorem 3.1. *Assume the following:*

- $a \in C^\varrho, b \in C^{1+\varrho}$, *for some* $\varrho \in (0,1]$;
- $a, b, \nabla b$ *are bounded;*
- b *is uniformly elliptic.*

Then the density of the law of $X^x(t)$ *exists and is denoted by* $p_t(x,y)$. *Furthermore, the following expansion is satisfied:*

$$p_t(x,y) = \Phi_t(x,y) + t^2 \sum_{i,j,k=1}^d c_{ijk}(x)\Phi_t^{(i,j,k)}(x,y) + \Delta_t(x,y), \tag{3.1}$$

where

$$c_{ijk}(x) = \frac{1}{4}\sum_{l=1}^d b_{kl}(x)\partial_{x_l} b_{ij}(x), \quad i,j,k = 1,\dots,d, \tag{3.2}$$

and the remainder kernel $\Delta_t(x,y)$ *has t-order* $1/2 + \varrho/2$.

[2] G stands for Gaussian.

Theorem 3.1 provides an approximation to the transition density $p_t(x,y)$ by an explicit function

$$p_t^{(1)}(x,y)=\Phi_t(x,y)+t^2\sum_{i,j,k=1}^d c_{ijk}(x)\Phi_t^{(i,j,k)}(x,y)$$

with the accuracy $t^{1/2+\varrho/2}$ (in the t-order sense and thus in the TV-distance). Note that the ratio

$$\frac{p_t^{(1)}(x,y)}{p_t^{EM}(x,y)}=1+t^2\sum_{i,j,k=1}^d c_{ijk}(x)H_t^{(i,j,k)}(x,y)$$

is the sum of a Hermite polynomial of order zero with the mixture of Hermite polynomials of order 3, and thus the approximation $p_t^{(1)}(x,y)$ is not necessarily positive.

Our second main result allows us to decompose, in the same spirit as (3.1), the *image measure* $\mathbf{P}(F_t(x,\overline{X}^x(t))\in \mathrm{d}y)$, where $\overline{X}^x(t)$ is defined by (1.2), and the mapping $F_t(x,\cdot):\mathbb{R}^d\to\mathbb{R}^d$ is a function of the form

$$F_t(x,y)=y+U_t(x,y), \tag{3.3}$$

where $U_t(x,y)$ is the $\mathbb{R}^d$-valued function defined through a given family of functions $\{f_{ijk}; i,j,k=1,...,d\}$ as

$$U_t(x,y)=t^2\sum_{i,j,k=1}^d f_{ijk}(x)H_t^{(j,k)}(x,y)\mathbf{e}_i. \tag{3.4}$$

Associated to the above function U, we define the corresponding scalar-valued function[3] $\delta_t^U(x,y)$ by

$$\delta_t^U(x,y)=t^2\sum_{i,j,k=1}^d f_{ijk}(x)H_t^{(i,j,k)}(x,y). \tag{3.5}$$

Theorem 3.2. *Assume that b is uniformly elliptic. Let $F_t(x,y)=y+U_t(x,y)$ and assume that the functions $f_{ijk}(x)$ in* (3.4) *are bounded. Then*

$$\begin{aligned}\mathbf{P}(F_t(x,\overline{X}^x(t))\in \mathrm{d}y)=\Phi_t(x,y)\mathrm{d}y+\delta_t^U(x,y)\Phi_t(x,y)\,\mathrm{d}y\\ +Q_t(x,y)\Phi_t(x,y)\,\mathrm{d}y+\mu_{x,t}(\mathrm{d}y),\end{aligned} \tag{3.6}$$

where $Q_t(x,y)$ has t-order 1 *and*

$$|\mu_{x,t}|(\mathbb{R}^d)\le Ce^{-ct^{-1/3}} \tag{3.7}$$

for some $c>0, C<\infty$.

We remark that the conditions on Theorem 3.2 are weaker than in Theorem 3.1. Now, combining Theorem 3.1 with Theorem 3.2, we can solve the problem posed in the Introduction. Namely, we are able to construct a family of approximations for $X^x(t)$, which will have a faster G-rate than the Euler-Maruyama scheme $\overline{X}^x(t)$. This improved approximation will be given by the explicit formula

$$\overline{X}^{x,U}(t)=\overline{X}^x(t)+U_t(x,\overline{X}^x(t)), \tag{3.8}$$

where the (quadratic in y) function $U_t(x,y)$ is given by (3.4) with the condition on the coefficients $f_{ijk}(x)$ which we will now introduce. Define

$$f_{ijk}^{\mathrm{sym}}(x)=\frac{1}{6}\Big(f_{ijk}(x)+f_{ikj}(x)+f_{jik}(x)+f_{jki}(x)+f_{kij}(x)+f_{kji}(x)\Big),$$

[3] This notation is purposely similar to Skorohod integrals. For more on this, we refer the reader to Section B in the Appendix.

and

$$c^{\rm sym}_{ijk}(x) = \frac{1}{6}\Big(c_{ijk}(x) + c_{ikj}(x) + c_{jik}(x) + c_{jki}(x) + c_{kij}(x) + c_{kji}(x)\Big) = \\ = \frac{1}{3}\Big(c_{ijk}(x) + c_{ikj}(x) + c_{jki}(x)\Big), \tag{3.9}$$

as the function symmetrization of the coefficients $f_{ijk}(x)$ and $c_{ijk}(x)$. The second identity in (3.9) reflects the fact that the coefficients $c_{ijk}(x)$ are already symmetric w.r.t. i, j, see (3.2). The condition required for (3.8) to define an improved scheme is the following:

$$c^{\rm sym}_{ijk}(x) = f^{\rm sym}_{ijk}(x), \quad i, j, k = 1, \dots, d. \tag{3.10}$$

That is, the symmetrization for the coefficients of the scheme (3.8) should coincide with the one for the coefficients (3.2) which appear in the density expansion (3.1). The following statement is a straightforward corollary of Theorems 3.1 and 3.2.

Corollary 3.1. *Let conditions of Theorem 3.1 hold true and assume that the function* $U_t(x, y)$*, given by* (3.4)*, has bounded coefficients* $f_{ijk}(x)$ *such that* (3.10) *and* (3.2) *holds.*

Then the family $\overline{X}^{x,U}(t), x \in \mathbb{R}^d, t > 0$ *approximates in law the solution to* (1.1) *with the G-rate* $1/2 + \varrho/2$.

Proof. Remark that the Hermite functions and polynomials are symmetric w.r.t. permutations of their indices. Then, in the formulae (3.1) and (3.5), the coefficients $c_{ijk}(x)$ and $f_{ijk}(x)$ can be changed to $c^{\rm sym}_{ijk}(x)$ and $f^{\rm sym}_{ijk}(x)$ respectively. That is, (3.10) yields

$$\delta^U_t(x, y) = t^2 \sum_{i,j,k=1}^{d} c_{ijk}(x) H^{(i,j,k)}_t(x, y). \tag{3.11}$$

Hence, noting (2.2) and combining (3.1) and (3.6), we obtain for $\widetilde{X}^x(t) = \overline{X}^{x,U}(t)$, that the representation (2.7) is satisfied with $\Gamma_t(x, y) = \Delta_t(x, y) - Q_t(x, y)\Phi_t(x, y)$ and $\nu_{x,t}(dy) = -\mu_{x,t}(dy)$. □

The next corollary gives the weak error for small times.

Corollary 3.2. *Assume the conditions of Corollary 3.1. Let* $f : \mathbb{R}^d \to \mathbb{R}$ *be a measurable function which satisfies*

$$|f(x)| \le Ae^{B|x|}.$$

Then there exist $C > 0$ *and* $t_0 > 0$*, dependent only on* A, B *and the coefficients of the SDE* (1.1) *such that*

$$|\mathbf{E}f(X^x(t)) - \mathbf{E}f(\overline{X}^{x,U}(t))| \le Ce^{B|x|}t^{1/2+\varrho/2}, \quad x \in \mathbb{R}^d, \quad t \in (0, t_0).$$

Proof. Fix $x \in \mathbb{R}^d$ and decompose

$$f(y) = f(y)1_{|y-x|\le 1} + f(y)1_{|y-x|>1} =: f_{x,1}(y) + f_{x,2}(y),$$

and estimate separately

$$\Delta^f_{x,t,1} := |\mathbf{E}f_{x,1}(X^x(t)) - \mathbf{E}f_{x,1}(\overline{X}^{x,U}(t))|, \quad \Delta^f_{x,t,2} := |\mathbf{E}f_{x,2}(X^x(t)) - \mathbf{E}f_{x,2}(\overline{X}^{x,U}(t))|.$$

It follows from Theorem 3.2 that the TV-distance between the law of $\overline{X}^{x,U}(t)$ and the law of $X^x(t)$ is bounded by $Ct^{1/2+\varrho/2}$. Since

$$\sup_{y\in\mathbb{R}^d} |f_{x,1}(y)| \le Ae^B e^{B|x|},$$

this gives the bound

$$\Delta^f_{x,t,1} \le Ce^{B|x|}t^{1/2+\varrho/2}.$$

To estimate $\Delta^f_{x,t,2}$, we simply write

$$\Delta^f_{x,t,2} \le |\mathbf{E} f_{x,2}(X^x(t))| + |\mathbf{E} f_{x,2}(\overline{X}^{x,U}(t))| \le \Big(\mathbf{E} f(X^x(t))^2\Big)^{1/2} \mathbf{P}(|X^x(t)-x|>1)^{1/2} + \\ + \Big(\mathbf{E} f(\overline{X}^{x,U}(t))^2\Big)^{1/2} \mathbf{P}(|\overline{X}^{x,U}(t)-x|>1)^{1/2}.$$

Recall that the coefficients a, σ of the SDE (1.1) and the functions f_{ijk} in the formula (3.4) for $U_t(x,y)$ are bounded, and b is elliptic. Then it is easy to show using the exponential martingale inequality that

$$\mathbf{P}(|X^x(t)-x|>1) \le Ce^{-ct^{-1}}, \quad \mathbf{P}(|\overline{X}^{x,U}(t)-x|>1) \le Ce^{-ct^{-1}}.$$

Furthermore, standard estimates for solutions of SDE's and for the Brownian motion yield

$$\mathbf{E}e^{B|X^x(t)|} \le Ce^{B|x|}, \quad t\le 1, \quad \mathbf{E}e^{B|\overline{X}^{x,U}(t)|} \le Ce^{B|x|}, \quad t \le t_0,$$

where t_0 should be taken small enough in order to guarantee that

$$\mathbf{E}e^{B|U(x,\overline{X}^x(t))|} \le C.$$

Combining these estimates, we get

$$\Delta^f_{x,t,2} \le Ce^{B|x|}e^{-ct^{-1}}. \qquad \square$$

Remark 3.1. Condition (3.10) is actually both *necessary and sufficient* for (3.11); for more details see Section B below. That is, if (3.10) fails, the difference $\mathbf{P}(X^x(t) \in dy) - \mathbf{P}(\overline{X}^{x,U}(t) \in dy)$ will contain a non-trivial term

$$\left(\delta^U_t(x,y) - t^2 \sum_{i,j,k=1}^d c_{ijk}(x) H^{(i,j,k)}_t(x,y)\right)\Phi_t(x,y)\,dy$$

of t-order $1/2$. This, in particular, means that the Euler-Maruyama approximation (1.2) has the G-rate $1/2$, unless

$$c^{\mathrm{sym}}_{ijk}(x) = 0, \quad i,j,k = 1,\dots,d.$$

It is clear that for coefficients $\{c_{ijk}\}$, we can choose $\{f_{ijk}\}$ satisfying (3.10) in different ways. Therefore the algorithm to derive many *explicit* approximation schemes with *improved G-rates* is as follows. We first calculate the coefficients $c_{ijk}(x)$ in the improved approximation for the transition density of the process. Then we take any mapping of the form (3.3), (3.4) with coefficients $f_{ijk}(x)$ satisfying (3.10). Then the corresponding approximation (3.8) has the (improved) G-rate 1.

In Section 6 we show that the second order approximation (1.6) is a particular representative of the family of improved approximations (3.8) with the coefficients of the function $U_t(x,y)$ satisfying (3.10). This argument will complete the proof of Theorem 1.1.

4. Proof of Theorem 3.1

4.1. **The parametrix expansion for $p_t(x,y)$. Outline of the proof.** In this section we briefly recall the classic *parametrix method* which provides a series expansion of the transition density of a diffusion process based on a basic function. For a detailed exposition we refer the reader to the classical monograph [6], Chapter 1.

Denote by L the following second order differential operator, defined on the space $C^2_0(\mathbb{R}^d)$ of twice continuously differentiable functions which vanish at ∞ together with their derivatives:

$$L = \sum_{i=1}^d a_i(x)\partial_{x_i} + \frac{1}{2}\sum_{i,j=1}^d b_{ij}(x)\partial^2_{x_ix_j}.$$

Clearly, by Itô's formula L is a restriction to $C_0^2(\mathbb{R}^d)$ of the generator of the semigroup generated by the Markov process X defined by the SDE (1.1). Under the assumptions of Theorem 3.1, the transition density $p_t(x,y)$ associated to $X^x(t)$ exists and it can be represented as the sum of the following absolutely convergent series

$$p_t(x,y) = \Xi_t(x,y) + \sum_{k=1}^{\infty} \big(\Xi \circledast \Upsilon^{\circledast k}\big)_t(x,y), \tag{4.1}$$

where $\circledast$ is the space-time convolution defined in (2.9), with

$$\Xi_t(x,y) = \Phi_t(y;x,y), \tag{4.2}$$

and the auxiliary kernel $\Upsilon_t(x,y)$ is defined by

$$\Upsilon_t(x,y) = -\Big(\frac{\partial}{\partial t} - L_x\Big)\Xi_t(x,y). \tag{4.3}$$

Here and below the notation L_x means that the derivatives in the differential operator L are applied w.r.t. the variable x. In the present case, the expansion is obtained based on the kernel function $\Xi_t(x,y)$ in (4.2). Notably, this differs from the kernel function used in [6], Chapter 1, where the drift term is neglected in the kernel function, but this difference is just technical and the entire argument remains literally the same; see in particular Section 4.2 below where the convergence of the series (4.1) is proved.

The series representation (4.1) is our starting point to obtain the expansion (3.1). Let us outline the methodology.

First, we show that the sum of all the terms with $k \geq 2$ (we call them 'tail' terms) in (4.1) has t-order $1/2 + \varrho/2$ and thus can be considered as a residual term in the representation (3.1). That is, the sum of several residual terms will constitute the 'error' kernel $\Delta_t(x,y)$. This is one of the purposes of Step 1.

After this first 'tail cut-off' step, only two terms from the entire series (4.1) will remain: the term $\Xi_t(x,y)$ which has t-order 0 and the convolution term $(\Xi \circledast \Upsilon)_t(x,y)$. These terms are not yet the ones we announced in (3.1). In particular, the zero-order term in (3.1) is the density of the Euler-Maruyama scheme and equals $\Phi(x;x,y)$, while the function $\Xi_t(x,y)$ has the form $\Phi(y;x,y)$, which does not correspond to a density. Furthermore, the convolution term has to be expanded in an explicit form. In the next step of the proof we will expand the function $\Phi(y;x,y)$ around $\Phi(x;x,y)$; in other words, this step consists of exchanging the value of $\sigma(y)$ in $\Phi(y;x,y)$ into $\sigma(x)$. The last step is to perform the similar expansions for the terms involved into the convolution, and to compute explicitly the principal part of the convolution. These two steps will also produce residual terms that will be included into the 'error' kernel $\Delta_t(x,y)$.

We recall the reader that through this section we assume the conditions on the functions a and b as stated in Theorem 3.1.

4.2. **Step 1: Convergence of** (4.1) **and the 'tail cut-off'.** The kernel $\Phi_t(\xi;x,y)$ is a fundamental solution to the following version of the operator L with the coefficients 'frozen' at the point ξ:

$$L_x^\xi = \sum_{i=1}^d a_i(\xi)\partial_{x_i} + \frac{1}{2}\sum_{i,j=1}^d b_{ij}(\xi)\partial^2_{x_i x_j}.$$

Using these observations and the definition of the Hermite functions of t-order 1 and 2 we can write the function $\Upsilon_t(x,y)$ as

$$\Upsilon_t(x,y) = \Big(L_x - \frac{\partial}{\partial t}\Big)\Phi_t(y;x,y) = (L_x - L_x^y)\Phi_t(y;x,y) =$$

$$
\begin{aligned}
&= \sum_{i=1}^{d}\Big(a_i(x)-a_i(y)\Big)\Phi_t^{(i)}(y;x,y)+\frac{1}{2}\sum_{i,j=1}^{d}\Big(b_{ij}(x)-b_{ij}(y)\Big)\Phi_t^{(i,j)}(y;x,y) =:\\
&=: \Upsilon_t^{\mathrm{drift}}(x,y)+\Upsilon_t^{\mathrm{diffusion}}(x,y). \qquad (4.4)
\end{aligned}
$$

Lemma 4.1. *The kernels $\Upsilon_t^{\mathrm{drift}}(x,y)$ and $\Upsilon_t^{\mathrm{diffusion}}(x,y)$ have t-orders $-1/2+\varrho/2$ and $-1/2$, respectively.*

Proof. For given $R>0$, define

$$A_R=\{\mathbf{a}\in\mathbb{R}^d : |\mathbf{a}|\le R\},$$

$$B_R=\{\mathbf{b}\in\mathbb{R}^{d\times d}_{\mathrm{sym}} : \text{ for all } v\in\mathbb{R}^d,\ R^{-1}|v|^2\le(\mathbf{b}v,v)\le R|v|^2\}.$$

Since the functions a and b are bounded and b is uniformly elliptic, we have that there exists $R>0$ large enough such that

$$a(\xi)\in A_R,\quad b(\xi)\in B_R,\quad \xi\in\mathbb{R}^d.$$

The sets A_R, B_R are compact, thus the kernels $\Phi_t^{(i)}(\mathbf{a},\mathbf{b};x,y),\Phi_t^{(i,j)}(\mathbf{a},\mathbf{b};x,y)$ have uniform t-orders $-1/2,-1$ w.r.t. $\mathbf{a}\in A_R,\mathbf{b}\in B_R$; see Section 2. This yields that $\Phi_t^{(i)}(y;x,y),\Phi_t^{(i,j)}(y;x,y)$ have t-orders $-1/2,-1$.

On the other hand, the functions $|a_i(x)-a_i(y)|,|b_{ij}(x)-b_{ij}(y)|$ are dominated by the functions

$$F_t^1(x,y)=C|x-y|^{\varrho},\quad F_t^2(x,y)=C|x-y|$$

which have the orders $\varrho/2,1/2$. Recall that the product of a function and a kernel has as order the sum of their orders, which yields that $\Upsilon_t^{\mathrm{drift}}(x,y)$ has t-order $-1/2+\varrho/2$ and $\Upsilon_t^{\mathrm{diffusion}}(x,y)$ has t-order $-1+1/2=-1/2$. □

The following statement is standard and general but for the sake of completeness we provide its proof for the reader's convenience.

Proposition 4.1.
1. *Let $\Xi_t(x,y),\Theta_t(x,y)$ be kernels of t-orders p and r respectively with $p,\ r>-1$. Then the convolution $(\Xi\circledast\Theta)_t(x,y)$ is well defined, and has t-order $p+r+1$.*
2. *Let $\Theta_t(x,y)$ be a kernel of t-order $r>-1$, then the series $\sum_{k=2}^{\infty}\Theta_t^{\circledast k}(x,y)$ absolutely converges and has t-order $2r+1$.*

Proof. 1. The definition of t-order for kernel functions is equivalent to the following: there exist positive constants C,c such that for $(t,x,y)\in(0,t_0)\times\mathbb{R}^d\times\mathbb{R}^d$

$$|\Gamma_t(x,y)|\le Ct^p\varphi_{ct}(x,y),$$

where

$$\varphi_t(x,y)=(2\pi t)^{-d/2}e^{-\frac{|y-x|^2}{2t}}$$

is the transition density of the Wiener process. Without loss of generality we can assume that the constants c,C in this alternative definition for the kernels $\Xi_t(x,y),\Theta_t(x,y)$ are the same. Then

$$|(\Xi_{t-s}*\Theta_s)(x,y)|\le C^2(t-s)^p s^r(\varphi_{c(t-s)}*\varphi_{cs})(x,y)=C^2(t-s)^p s^r\varphi_{ct}(x,y)$$

and thus

$$|(\Xi\circledast\Theta)_t(x,y)|\le C^2\varphi_{ct}(x,y)\int_0^t(t-s)^p s^r\,ds=C^2t^{p+r+1}\varphi_t(x,y)\mathrm{B}(p+1,r+1)$$

where $\mathrm{B}(\cdot,\cdot)$ denotes the Beta function.

2. By the same reasoning, for any $k \geq 2$

$$\left|\left(\Theta^{\circledast k}\right)_t(x,y)\right| \leq C^k t^{kr+k-1} \varphi_t(x,y) \prod_{j=1}^{k-1} B(j(r+1), r+1) =$$
$$= \varphi_t(x,y) t^{kr+k-1} \frac{\left(C\Gamma(r+1)\right)^k}{\Gamma(k(r+1))},$$

where $\Gamma(\cdot)$ denotes the Gamma function. Using the Stirling formula for the Gamma function, we obtain that for any $C' > 0$

$$\Gamma(k(r+1)) \geq (C')^k$$

for sufficiently large k. Hence, taking C' large enough, we conclude that the convolution series absolutely converges, and

$$\left|\sum_{k=2}^{\infty} \left(\Theta^{\circledast k}\right)_t(x,y)\right| \leq C t^{2r+1} \varphi_t(x,y). \qquad \square$$

Now we can prove that the series (4.1) converges and realize the 'tail cut off' heuristically explained in subsection 4.1. We know from Lemma 4.1 and (4.3) that the kernel $\Upsilon_t(x,y)$ has the t-order $-1/2 = \min(-1/2+\varrho/2, -1/2)$. Then by statement 2 of Proposition 4.1, the series $\sum_{k=2}^{\infty} \Upsilon_t^{\circledast k}(x,y)$ absolutely converges and has t-order 0. Since $\Xi_t(x,y)$ has t-order 0 then using statement 1 of Proposition 4.1, the sum $\sum_{k=2}^{\infty} \left(\Xi \circledast \Upsilon^{\circledast k}\right)_t(x,y)$ has the t-order 1 (note that one has to use Fubini's Theorem). Similarly, we observe that the kernel $(\Xi \circledast \Upsilon^{\text{drift}})_t(x,y)$ has the t-order $-1/2 + \varrho/2 + 1 = 1/2 + \varrho/2$. Summarizing all the above, we get the following:

$$p_t(x,y) = \Xi_t(x,y) + (\Xi \circledast \Upsilon^{\text{diffusion}})_t(x,y) + \Delta_t^1(x,y), \tag{4.5}$$

where the kernel $\Delta_t^1(x,y)$ has t-order $1/2 + \varrho/2$.

4.3. **Step 2: The change of 'freezing point'.** In this step, we will expand the kernel function $\Xi_t(x,y)$ using the density of the Euler-Maruyama scheme.

Lemma 4.2. *The following expansion formula holds true for a kernel $\Delta_t^2(x,y)$ which has t-order $1/2 + \varrho/2$:*

$$\Xi_t(x,y) = \Phi_t(x,y) + \frac{t}{2} \sum_{i,j=1}^{d} (b_{ij}(y) - b_{ij}(x)) \Phi_t^{(i,j)}(x,y) + \Delta_t^2(x,y).$$

Proof. Recall that $\Xi_t(x,y) = \Phi_t(a(y), b(y); x, y)$, $\Phi_t(x,y) = \Phi_t(a(x), b(x); x, y)$, thus

$$\Xi_t(x,y) - \Phi_t(x,y) = \Phi_t(a(y), b(y); x, y) - \Phi_t(a(x), b(y); x, y) +$$
$$+ \Phi_t(a(x), b(y); x, y) - \Phi_t(a(x), b(x); x, y).$$

We have by (A.6) for any $\mathbf{a}, \mathbf{a}' \in \mathbb{R}^d$, $\mathbf{b} \in \mathbb{R}^{d\times d}_{\text{sym}}$

$$\Phi_t(\mathbf{a}', \mathbf{b}; x, y) - \Phi_t(\mathbf{a}, \mathbf{b}; x, y) = t \sum_{i=1}^{d} (\mathbf{a}'_i - \mathbf{a}_i) \Phi_t^{(i)}(\mathbf{a}'', \mathbf{b}; x, y)$$

where $\mathbf{a}''$ is a convex combination of $\mathbf{a}, \mathbf{a}'$. Note that the sets A_R, B_R specified in the proof of Lemma 4.1 are convex. Take $\mathbf{a} = a(x), \mathbf{a}' = a(y), \mathbf{b} = b(y)$, then $\mathbf{a}, \mathbf{a}', \mathbf{a}'' \in A_R$, $\mathbf{b} \in B_R$ and the same argument as in the proof of Lemma 4.1 yields that the kernel

$$\Phi_t(a(y), b(y); x, y) - \Phi_t(a(x), b(y); x, y)$$

has t-order $1 + \varrho/2 + (-1/2) = 1/2 + \varrho/2$.

Next, using (A.7) and the 1st order Taylor formula, we get

$$\Phi_t(\mathbf{a},\mathbf{b}';x,y)-\Phi_t(\mathbf{a},\mathbf{b};x,y)=\frac{t}{2}\sum_{i,j=1}^d(\mathbf{b}'_{ij}-\mathbf{b}_{ij})\Phi_t^{(i,j)}(\mathbf{a},\mathbf{b};x,y)+$$
$$+\frac{t^2}{8}\sum_{i,j,k,l=1}^d(\mathbf{b}'_{ij}-\mathbf{b}_{ij})(\mathbf{b}'_{kl}-\mathbf{b}_{kl})\Phi_t^{(i,j,k,l)}(\mathbf{a},\mathbf{b}'';x,y),$$

where $\mathbf{b}''$ is a convex combination of $\mathbf{b},\mathbf{b}'\in\mathbb{R}^{d\times d}_{\mathrm{sym}}$. The kernel $\Phi_t^{(i,j,k,l)}(\mathbf{a},\mathbf{b};x,y)$ has t-order -2 uniformly in $\mathbf{a}\in A_R,\mathbf{b}\in B_R$. Taking $\mathbf{a}=a(x),\mathbf{b}=b(x),\mathbf{b}'=b(y)$ we get $\mathbf{a}\in A_R,\mathbf{b}''\in B_R$, hence

$$\Phi_t(a(x),b(y);x,y)-\Phi_t(a(x),b(x);x,y)-\frac{t}{2}\sum_{i,j=1}^d(b_{ij}(y)-b_{ij}(x))\Phi_t^{(i,j)}(x,y)$$

has t-order

$$2+1/2+1/2+(-2)=1\geq 1/2+\varrho/2. \qquad \square$$

Lemma 4.3. *The following expansion formula holds true*

$$\frac{t}{2}\sum_{i,j=1}^d(b_{ij}(y)-b_{ij}(x))\Phi_t^{(i,j)}(x,y)=2t^2\sum_{i,j,k=1}^d c_{ijk}(x)\Phi_t^{(i,j,k)}(x,y)+$$
$$+t\sum_{i,j=1}^d\partial_{x_j}b_{ij}(x)\Phi_t^{(i)}(x,y)+\Delta_t^3(x,y), \qquad (4.6)$$

where the coefficients $c_{ijk}(x)$ are defined by (3.2) *and the kernel $\Delta_t^3(x,y)$ has t-order $1/2+\varrho/2$.*

Proof. The proof here is similar and simpler than the one from the proof of Lemma 4.2, thus we omit some details.

First, we apply the 1st order Taylor formula to the differences $b_{ij}(y)-b_{ij}(x)$ on the left hand side of (4.6) to obtain that

$$\frac{t}{2}\sum_{i,j=1}^d(b_{ij}(y)-b_{ij}(x))\Phi_t^{(i,j)}(x,y)-\frac{t}{2}\sum_{i,j,l=1}^d\partial_{x_l}b_{ij}(x)(y-x)_l\Phi_t^{(i,j)}(x,y)$$

has t-order $1/2+\varrho/2$. Note that here we have used the fact that $\nabla b\in C^\varrho$.

Next, we replace in the second term of the above formula $(y-x)_l$ by $(y-x-ta(x))_l$. Note that the corresponding difference has the t-order $2+1-2=1\geq 1/2+\varrho/2$. On the other hand, by (A.5)

$$\frac{t}{2}\sum_{i,j,l=1}^d\partial_{x_l}b_{ij}(x)(y-x-ta(x))_l\Phi_t^{(i,j)}(x,y)=\frac{t^2}{2}\sum_{i,j,l,k=1}^d\partial_{x_l}b_{ij}(x)b_{kl}(x)\Phi_t^{(i,j,k)}(x,y)+$$
$$+\frac{t}{2}\sum_{i,j,l=1}^d\partial_{x_l}b_{ij}(x)1_{j=l}\Phi_t^{(i)}(x,y)+\frac{t}{2}\sum_{i,j,l=1}^d\partial_{x_l}b_{ij}(x)1_{i=l}\Phi_t^{(j)}(x,y)=$$
$$=2t^2\sum_{i,j,k=1}^d c_{ijk}(x)\Phi_t^{(i,j,k)}(x,y)+t\sum_{i,j=1}^d\partial_{x_j}b_{ij}(x)\Phi_t^{(i)}(x,y).$$

Remark that in the last identity we have used symmetry of $b_{ij}(x)$. This proves the required statement. $\square$

Summarizing the above calculation, we conclude that the following expansion hold true

$$\Xi_t(x,y) = \Phi_t(x,y) + 2t^2 \sum_{i,j,k=1}^{d} c_{ijk}(x)\Phi_t^{(i,j,k)}(x,y) + t \sum_{i,j=1}^{d} \partial_{x_j} b_{ij}(x)\Phi_t^{(i)}(x,y) + \Delta_t^4(x,y), \tag{4.7}$$

where the kernel $\Delta_t^4(x,y)$ has t-order $1/2+\varrho/2$.

4.4. **Step 3: Expanding the convolution term. End of the proof.** In this section, we provide an expansion, similar to (4.7), for the convolution term $(\Xi \circledast \Upsilon^{\text{diffusion}})_t(x,y)$. First, we expand each kernel under the convolution.

Let us start with $\Xi_t(x,y)$. By (4.7), the difference $\Xi_t(x,y) - \Phi_t(x,y)$ has t-order $1/2$. Since $\Upsilon_t^{\text{diffusion}}(x,y)$ has t-order $-1/2$, then by Proposition 4.1

$$(\Xi \circledast \Upsilon^{\text{diffusion}})_t(x,y) - (\Phi \circledast \Upsilon^{\text{diffusion}})_t(x,y)$$

has t-order $1/2+1-1/2 = 1 \geq 1/2+\varrho/2$.

Now, we continue with the expansion of $\Upsilon_t^{\text{diffusion}}(x,y)$. Define the kernel

$$\Lambda_t(x,y) = -2t \sum_{i,j,k=1}^{d} c_{ijk}(x)\Phi_t^{(i,j,k)}(x,y) - \sum_{i,j=1}^{d} \partial_{x_j} b_{ij}(x)\Phi_t^{(i)}(x,y).$$

We claim that $\Upsilon_t^{\text{diffusion}}(x,y) - \Lambda_t(x,y)$ has t-order $-1/2+\varrho/2$. The proof of this claim repeats arguments from the previous proofs (in particular, Lemmas 4.1, 4.2 and 4.3), hence we omit the details and just outline the argument.

First, we change $\Phi_t^{(i,j)}(y;x,y)$ in the definition of $\Upsilon_t^{\text{diffusion}}(x,y)$ to $\Phi_t^{(i,j)}(x,y)$. The argument here is similar to Lemma 4.1 and is even simpler, since for both of the variables $\mathbf{a} \in \mathbb{R}^d$ and $\mathbf{b} \in \mathbb{R}^{d\times d}_{\text{sym}}$ we apply the 0-th order Taylor formula (also known as the mean value theorem). The corresponding error will be of t-order

$$1/2 + \min\Big(1+\varrho/2-3/2, 1+1/2-4/2\Big) = 0.$$

After this argument, $\Upsilon_t^{\text{diffusion}}(x,y)$ can be approximated by

$$\frac{1}{2}\sum_{i,j=1}^{d} (b_{ij}(x)-b_{ij}(y))\Phi_t^{(i,j)}(x,y),$$

which is just the left hand side of (4.6) divided by $(-t)$. Applying Lemma 4.3 we complete the proof of the claim. That is, $\Upsilon_t^{\text{diffusion}}(x,y) - \Lambda_t(x,y)$ has t-order $-1/2+\varrho/2$.

Now, we apply Proposition 4.1 once again, in order to obtain that

$$(\Phi \circledast \Upsilon^{\text{diffusion}})_t(x,y) - (\Phi \circledast \Lambda)_t(x,y)$$

has t-order $1/2+\varrho/2$. That is, the convolution term $(\Xi \circledast \Upsilon^{\text{diffusion}})_t(x,y)$ can be replaced by the (simpler) term $(\Phi \circledast \Lambda)_t(x,y)$ with an error kernel of t-order $1/2+\varrho/2$.

Next, we will consider the kernel $(\Phi \circledast \Lambda)_t(x,y)$ using the convolution property (A.9). With that aim in mind, we perform further transformations of the kernel $\Lambda_t(x,y)$. Denote

$$\Lambda_t(\xi;x,y) = -2t \sum_{i,j,k=1}^{d} c_{ijk}(\xi)\Phi_t^{(i,j,k)}(\xi;x,y) - \sum_{i,j=1}^{d} \partial_{\xi_j} b_{ij}(\xi)\Phi_t^{(i)}(\xi;x,y),$$

and observe that by (A.9) with $\mathbf{a} = a(x)$, $\mathbf{b} = b(x)$

$$\int_{\mathbb{R}^d} \Phi_{t-s}(x,z)\Lambda_s(x;z,y)\,dz = -2s \sum_{i,j,k=1}^d c_{ijk}(x) \int_{\mathbb{R}^d} \Phi_{t-s}(x,z)\Phi_s^{(i,j,k)}(x;z,y)\,dz - \\ - \sum_{i,j=1}^d \partial_{x_j} b_{ij}(x) \int_{\mathbb{R}^d} \Phi_{t-s}(x,z)\Phi_t^{(i)}(x;z,y)\,dz = \\ = -2s \sum_{i,j,k=1}^d c_{ijk}(x)\Phi_t^{(i,j,k)}(x,y) - \sum_{i,j=1}^d \partial_{x_j} b_{ij}(x)\Phi_t^{(i)}(x,y).$$

Taking the integral in s we obtain

$$\int_0^t \int_{\mathbb{R}^d} \Phi_{t-s}(x,z)\Lambda_s(x;z,y)\,dzds = -t^2 \sum_{i,j,k=1}^d c_{ijk}(x)\Phi_t^{(i,j,k)}(x,y) - \\ - t \sum_{i,j=1}^d \partial_{x_j} b_{ij}(x)\Phi_t^{(i)}(x,y).$$

On the other hand, the following bound is available: there exists a kernel $\Theta_t(x,y)$ of t-order $-1/2$ such that

$$|\Lambda_t(\xi;x,y) - \Lambda_t(\xi';x,y)| \le |\xi - \xi'|^{\varrho}\Theta_t(x,y).$$

The proof of this bound is analogous to the proofs of Lemmas 4.1 and 4.2 and therefore details are omitted. Using this bound,

$$\left|(\Phi \circledast \Lambda)_t(x,y) - \int_0^t \int_{\mathbb{R}^d} \Phi_{t-s}(x,z)\Lambda_s(x;z,y)\,dzds\right| = \\ = \left|\int_0^t \int_{\mathbb{R}^d} \Phi_{t-s}(x,z)\Big(\Lambda_s(z;z,y) - \Lambda_s(x;z,y)\Big)\,dzds\right| \le \\ \le \int_0^t \int_{\mathbb{R}^d} \Phi_{t-s}(x,z)|z-x|^{\varrho}\Theta_s(z,y)\,dzds = (\Gamma \circledast \Theta)_t(x,y),$$

where we denote by $\Gamma_t(x,y) = |y-x|^{\varrho}\Phi_t(x,y)$. Since the t-orders of the kernels $\Gamma_t(x,y), \Theta_t(x,y)$ are $\varrho/2$, $-1/2$, using Proposition 4.1 we conclude that the kernel

$$(\Phi \circledast \Lambda)_t(x,y) - \int_0^t \int_{\mathbb{R}^d} \Phi_{t-s}(x,z)\Lambda_s(x;z,y)\,dzds$$

has t-order $\varrho/2 - 1/2 + 1 = 1/2 + \varrho/2$. Summarizing the above calculation, we conclude that

$$(\Xi \circledast \Upsilon^{\text{diffusion}})_t(x,y) = -t^2 \sum_{i,j,k=1}^d c_{ijk}(x)\Phi_t^{(i,j,k)}(x,y) - \\ - t \sum_{i,j=1}^d \partial_{x_j} b_{ij}(x)\Phi_t^{(i)}(x,y) + \Delta_t^5(x,y), \tag{4.8}$$

where the kernel $\Delta_t^5(x,y)$ has t-order $1/2 + \varrho/2$.

Now the statement of the theorem is obtained as a straightforward combination of (4.5), (4.7), and (4.8). □

5. Proof of Theorem 3.2

Before giving the proof, we give a brief introduction to the Ramer change of variables formula in a simple finite dimensional setting.

5.1. **Change of variables for a Gaussian measure.** Let $P(\mathrm{d}y)$ be a Gaussian measure on $\mathbb{R}^d$ with mean vector $\mathbf{a} \in \mathbb{R}^d$ and strictly positive definite covariance matrix $\mathbf{b} \in \mathbb{R}^{d\times d}$. Then for a C^1-mapping $F\colon \mathbb{R}^d \to \mathbb{R}^d$ for any bounded measurable functions $g, h\colon \mathbb{R}^d \to \mathbb{R}_+$ with compact support, we have

$$\int_{\mathbb{R}^d} g(F(y))h(y)J^{F,P}(y)\,P(\mathrm{d}y) = \int_{\mathbb{R}^d} g(y)\left(\sum_{z:F(z)=y} h(z)\right)P(\mathrm{d}y), \tag{5.1}$$

where the above sum is well defined if one assumes that ∇F exists (the index set is in fact a finite set P-a.e.) and

$$J^{F,P}(y) = |\det(\nabla F(y))|\exp\Big\{-\Big(\mathbf{b}^{-1}(y-\mathbf{a}), U(y)\Big) - \frac{1}{2}\Big(\mathbf{b}^{-1}U(y), U(y)\Big)\Big\}, \tag{5.2}$$
$$U(y) := F(y) - y.$$

This is actually the classical Jacobi formula[4] combined with an elementary calculation which involves the Gaussian distribution density of $P \sim \mathcal{N}(\mathbf{a}, \mathbf{b})$; we refer to (1.1) and (1.2) in [13] for the case of canonical Gaussian measure, the general case is completely analogous. We will use an alternative formula for (5.2), which dates back to Ramer [11], see [13, Section III] for an excellent overview of the history of the subject.

To introduce Ramer's formula, we first recall the definition of the *Carleman-Fredholm determinant* of a mapping $I + A$ which is denoted by $\det_2(I + A)$ and defined as

$$\det{}_2(I + A) = \det(I + A)\exp(-\operatorname{trace} A).$$

Ramer's formula has a modern expression using stochastic integrals and Girsanov change of measure. To introduce this formula, we define the *divergence operator*, or *the Skorokhod integral*[5] of U denoted by $\delta_P U$ as

$$[\delta_P U](y) = \Big(\mathbf{b}^{-1}(y-\mathbf{a}), U(y)\Big) - \operatorname{trace}\nabla U(y). \tag{5.3}$$

Then (5.2) can be re-written as follows:

$$J^{F,P}(y) = |\det{}_2(I_d + \nabla U(y))|\exp\Big\{-[\delta_P U](y) - \frac{1}{2}\Big(\mathbf{b}^{-1}U(y), U(y)\Big)\Big\}. \tag{5.4}$$

Here I_d denotes the identity matrix of order $d \times d$. Formula (5.4) is commonly used in the Gaussian infinite-dimensional version of (5.1) instead of (5.2) because $\det_2(I + A)$ is well defined for Hilbert-Schmidt mappings A, while $\det(I + A)$ requires A to be from the trace class.

In our finite-dimensional setting this difference is not so crucial, however, using (5.4) instead of (5.2) will simplify the calculations significantly.

Formula (5.1) leads to the following: let a measurable set $A \subset \mathbb{R}^d$ be such that $\nabla F(y)$ is non-degenerate for P-a.e. for $y \in A$. Then the image measure of the restriction $P|_A(\cdot) = P(\cdot \cap A)$ under F has the density w.r.t. P equal

$$\rho^F_A(y) = \sum_{z\in A:F(z)=y} \Big[J^{F,P}(z)\Big]^{-1}. \tag{5.5}$$

Here, the index set $\{z \in A : F(z) = y\}$ has at most an infinite countable number of elements and the above sum is well defined P-a.e. for $y \in F(A)$.

One easily gets (5.5) from (5.1) by monotone approximation of

$$h(y) = \Big[J^{F,P}(y)\Big]^{-1} 1_A(y)$$

---

[4] Also linked to the simpler change of variables formula for multivariate integrals and many Sard's type theorems.

[5] One can prove that $\delta_P U$ corresponds to the adjoint operator to ∇ in the $L_2(\mathbb{R}^d, P)$ sense.

using bounded measurable functions.

5.2. **Outline of the proof for Theorem 3.2.** Rather than giving the proof directly, we first give an outline of the proof to guide the reader.

The goal is to compute an expansion of the image measure

$$P^F_{x,t}(\mathrm{d}y) := \mathbf{P}\Big(F_t(x, \overline{X}^x(t)) \in \mathrm{d}y\Big). \tag{5.6}$$

To achieve this goal, we will apply (5.5) for $F(\cdot) \equiv F_{x,t}(\cdot) := F_t(x,\cdot)$ (i.e. (t,x) is considered as a fixed parameter) and P given by the law of $\overline{X}^x(t)$ which is the Gaussian distribution $\mathcal{N}(x + ta(x), tb(x))$. The set A will be chosen in a way so as to avoid the technical difficulty caused by the equation

$$F_t(x, z) = y, \tag{5.7}$$

which, in general, is not easy to solve for all $y \in \mathbb{R}^d$ [6] .

That is, we will consider a local problem on the set $A = B_{x,t,1}$, where we use the following parametrized set

$$B_{x,t,R} = \{y : |y - x - ta(x)| \le Rt^{1/3}\}, \quad R > 0.$$

The choice of the factor $t^{1/3}$ in the definition of $B_{x,t,R}$ is intrinsic, and will become clear later; see Remark 5.1. In Section 5.3, below we prove the following result which states the existence of unique solutions for (5.7) locally.

Proposition 5.1. *There exists $t_0 > 0$ such that for $t \le t_0$ the following properties hold:*

(i) $\det(\nabla_2 F_t(x, y)) \neq 0$ *for any $y \in B_{x,t,2}$.*
(ii) For any $r > 0$, the image of $B_{x,t,r}$ under $F_t(x,\cdot)$ is contained in $B_{x,t,2r}$.
(iii) For any $y \in B_{x,t,2}$ there exists a unique solution $z \in B_{x,t,4}$ to (5.7) *which we denote as $z = G_t(x, y)$ for $G_t(x,\cdot) : B_{x,t,2} \to B_{x,t,4}$ which satisfies that* $|G_t(x, y) - y| < 2^{-1}t^{1/3}$ *for $y \in B_{x,t,2}$.*

With the above result, we decompose the image measure as follows:

$$\begin{aligned} P^F_{x,t}(\mathrm{d}y) &:= \mathbf{P}\Big(F_t(x, \overline{X}^x(t)) \in \mathrm{d}y,\, \overline{X}^x(t) \in B_{x,t,1}\Big) + \\ &\quad + \mathbf{P}\Big(F_t(x, \overline{X}^x(t)) \in \mathrm{d}y,\, \overline{X}^x(t) \notin B_{x,t,1}\Big) =: P^{F,\mathrm{loc}}_{x,t}(\mathrm{d}y) + P^{F,\mathrm{tail}}_{x,t}(\mathrm{d}y). \end{aligned}$$

As we assume that b is uniformly elliptic, the 'tail' part of this decomposition is exponentially negligible which is part of the statement in (3.7). That is, there exists positive constants C_0, C and c such that :

$$P^{F,\mathrm{tail}}_{x,t}(\mathbb{R}^d) = \mathbf{P}(\overline{X}^x(t) \notin B_{x,t,1}) \le \mathbf{P}(|W(t)| > C_0 t^{1/3}) \le Ce^{-ct^{-1/3}}. \tag{5.8}$$

For the 'local' part, the formula for the density is available following Section 5.1. Namely, using Proposition 5.1 (ii) with $r = 1$ and (iii), we obtain that the support of the measure $P^{F,\mathrm{loc}}_{x,t}$ is a subset of $B_{x,t,2}$, and on this set the unique solution to (5.7) in $B_{x,t,4}$ is $z = G_t(x, y)$. Thus, by (5.5) with $A = B_{x,t,1}$

$$\rho^F_{x,t}(y) := \frac{P^F_{x,t}(\mathrm{d}y)}{P_{x,t}(\mathrm{d}y)} = 1_{B_{x,t,2}}(y)\Big[J^{F_{x,t},P_{x,t}}\Big(G_t(x, y)\Big)\Big]^{-1} 1_{B_{x,t,1}}\Big(G_t(x, y)\Big).$$

Define $C_{x,t} = \{y : G_t(x, y) \in B_{x,t,1}\}$, then by Proposition 5.1 (ii) with $r = 1$, $C_{x,t} \subset B_{x,t,2}$. This gives finally the formula for the density of the 'local part':

$$\rho^F_{x,t}(y) = \Big[J^{F_{x,t},P_{x,t}}\Big(G_t(x, y)\Big)\Big]^{-1} 1_{C_{x,t}}(y). \tag{5.9}$$

In Section 5.4 below, we will prove the following result.

[6] To see this, it is enough to consider the quadratic function $F_t(x,\cdot) : \mathbb{R}^d \to \mathbb{R}^d$ given by (3.3) and (3.4).

Proposition 5.2. *The function*

$$Q_t(x,y) = \left\{ \left[J^{F_{x,t},P_{x,t}}\Big(G_t(x,y)\Big)\right]^{-1} - 1 - \delta_t^U(x,y)\right\} 1_{C_{x,t}}(y) \tag{5.10}$$

has t-order 1.

The rest of the proof for Theorem 3.2 is obtained as follows. We have (recall (3.5))

$$\begin{aligned} P^F_{x,t}(\mathrm{d}y) &= \rho^F_{x,t}(y)\Phi_t(x,y)\,\mathrm{d}y + P^{F,tail}_{x,t}(\mathrm{d}y) = \\ &= \left[J^{F_{x,t},P_{x,t}}\Big(G_t(x,y)\Big)\right]^{-1} 1_{C_{x,t}}(y)\Phi_t(x,y)\,\mathrm{d}y + P^{F,tail}_{x,t}(\mathrm{d}y) = \\ &= \Big(1+\delta_t^U(x,y)+Q_t(x,y)\Big)1_{C_{x,t}}(y)\Phi_t(x,y)\,\mathrm{d}y + P^{F,tail}_{x,t}(\mathrm{d}y) = \\ &= \Big(1+\delta_t^U(x,y)+Q_t(x,y)\Big)\Phi_t(x,y)\,\mathrm{d}y + \mu_{x,t}(\mathrm{d}y), \end{aligned}$$

with

$$\mu_{x,t}(\mathrm{d}y) = P_t^{F,\mathrm{tail}}(x,\mathrm{d}y) - \Big(1+\delta_t^U(x,y)\Big)\Phi_t(x,y)1_{\mathbb{R}^d\setminus C_{x,t}}(y)\,\mathrm{d}y,$$

which is actually the decomposition (3.6).

The first term, $P_t^{F,\mathrm{tail}}(x,\mathrm{d}y)$ in $\mu_{x,t}(\mathrm{d}y)$ has already been proven to be exponentially negligible in (5.8). To handle the second part, we note that by Proposition 5.1 (iii) and the triangular inequality it follows that $C_{x,t} \supset B_{x,t,1/2}$. Since 1 and $\delta_t^U(x,y)$ are functions with t-orders 0 and 1/2 respectively, this yields

$$\begin{aligned} &\int_{\mathbb{R}^d\setminus C_{x,t}} \Big|1+\delta_t^U(x,y)\Big|\Phi_t(x,y)\,\mathrm{d}y \le \\ &\qquad \le Ct^{-d/2}\int_{|y-x-ta(x)|\ge t^{1/3}/2} \exp\left(-c\frac{|y-x|^2}{t}\right)\mathrm{d}y \le Ce^{-ct^{-1/3}}, \end{aligned}$$

which completes the proof of (3.7).

The plan of the rest of Section 5 is the following. In Section 5.3 we study the properties of the inverse function $G_t(x,\cdot)$. In Section 5.4 we prove Proposition 5.2, which completes the proof of Theorem 3.2.

5.3. **The inverse function $G_t(x,\cdot)$: construction and approximation. Proof of Proposition 5.1.** Using (3.4) and the explicit formulas for the derivatives of Hermite polynomials (see formulas after (A.1) and before (A.3)), we have

$$U_t(x,y) = \sum_{i,j,k=1}^d f_{ijk}(x)\Big(\Big(b(x)^{-1}(y-x-a(x)t)\Big)_j\Big(b(x)^{-1}(y-x-a(x)t)\Big)_k - tb(x)^{-1}_{jk}\Big)\mathbf{e}_i, \tag{5.11}$$

$$\begin{aligned} \nabla_2\big(U_t(x,y)\big) = \sum_{i,j,k,l=1}^d f_{ijk}(x)\Big[&\Big(b(x)^{-1}(y-x-a(x)t)\Big)_j b(x)^{-1}_{kl} + \\ &+\Big(b(x)^{-1}(y-x-a(x)t)\Big)_k b(x)^{-1}_{jl}\Big]\mathbf{e}_i\otimes\mathbf{e}_l. \end{aligned} \tag{5.12}$$

Then for any $r>0$ there exists $C=C(r)$ such that for $t_0\le 1$

$$|U_t(x,y)|\le Ct^{2/3}, \quad |\nabla_2 U_t(x,y)|\le Ct^{1/3}, \quad y\in B_{x,t,r}, \quad t\le t_0. \tag{5.13}$$

The first inequality in (5.13) can be rewritten using (3.3) as $|F_t(x,y)-y|\le Ct^{2/3}$ for $y\in B_{x,t,r}$ and therefore by triangular inequality there exists $c>0$ such that

$$F_t(x,y)\in B_{x,t,r+ct^{1/3}}, \quad y\in B_{x,t,r},$$

which proves statement (ii). The second inequality in (5.13) shows that, for t_0 small enough, $\nabla_2 F_t(x,y)$ is a small perturbation of identity for $y \in B_{x,t,2}$, which proves statement (i).

Finally, we prove (iii). Note that the second inequality in (5.13) with t_0 small enough implies that $U_t(x,\cdot)$ is a 1/2-contraction on $B_{x,t,4}$. Thus the standard argument based on the Banach fixed point theorem can be used to solve the equation $y - U_t(x,z) = z$ (which is equivalent to (5.7)) for $y \in B_{x,t,2}$.

In order to describe the method in detail, define iteratively

$$\begin{aligned} G_t^0(x,y) &= y, \\ &\dots \\ G_t^k(x,y) &= y - U_t\Big(x, G_t^{k-1}(x,y)\Big), \\ &\dots \end{aligned} \tag{5.14}$$

Note that if we choose $t<1$ small enough so that $Ct^{1/3} < 2^{-2}$ where C is the constant in (5.13), then

$$|G_t^1(x,y) - G_t^0(x,y)| = |U_t(x,y)| \le Ct^{2/3} < 2^{-2}t^{1/3}, \quad y \in B_{x,t,2}.$$

Then one shows by induction that for any $y \in B_{x,t,4}$ and $k \ge 0$

$$|G_t^{k+1}(x,y) - G_t^k(x,y)| \le 2^{-k}|G_t^1(x,y) - G_t^0(x,y)| \le 2^{-k-2}t^{1/3}. \tag{5.15}$$

Using telescopic sums, one also obtains that $G_t^{k+1}(x,y) \in B_{x,t,4}$ for any $y \in B_{x,t,4}$ and $k \ge 0$. Then the function

$$G_t(x,y) = \lim_{k\to\infty} G_t^k(x,y), \quad y \in B_{x,t,2},$$

is well defined and takes values in $B_{x,t,4}$. Moreover, $z = G_t(x,y)$ is the unique solution to the equation (5.7) in the set $B_{x,t,4}$. This completes the entire proof. □

The construction of the function $G_t(x,\cdot)$ yields the following bound.

Corollary 5.1. *There exists a function $R_t(x,y)$, $(t,x,y) \in (0,t_0)\times\mathbb{R}^d\times\mathbb{R}^d$ of t-order 1 such that it verifies*

$$|G_t(x,y) - y| = R_t(x,y), \quad y \in B_{x,t,2}. \tag{5.16}$$

Proof. The estimates in (5.15) yield for any $y \in B_{x,t,2}$

$$|G_t(x,y) - G_t^0(x,y)| \le \sum_{k=0}^{\infty} |G_t^{k+1}(x,y) - G_t^k(x,y)| \le 2^{-1}|G_t^1(x,y) - G_t^0(x,y)|.$$

Recall that $G_t^0(x,y) = y$, $G_t^1(x,y) = y - U_t(x,y)$. Since $|U_t(x,y)|$ has t-order 1, this completes the proof if one defines $R_t(x,y) = 0$ for $y \notin B_{x,t,2}$. □

Note that in the above proof we have first defined the function $R_t(x,y)$ for $y \in B_{x,t,2}$ and then later extended it to the whole space in order to prove that it is a function of t-order 1. This argument will be used repeatedly without further explanation.

5.4. **Proof of Proposition 5.2.** The proof follows by applying an expansion technique to the function $\Big[J^{F_{x,t},P_{x,t}}\Big(G_t(x,y)\Big)\Big]^{-1}$. That is, using (5.4), (5.11), (5.12) and (A.2), we have for $y \in C_{x,t}$,

$$\begin{aligned} \Big[J^{F_{x,t},P_{x,t}}\Big(G_t(x,y)\Big)\Big]^{-1} &= \Big[\det_2\Big(I_d + \nabla_2 U_t\Big(x, G_t(x,y)\Big)\Big)\Big]^{-1} \times \\ &\times \exp\Big[\delta_t^U\Big(x, G_t(x,y)\Big) + \frac{1}{2t}\Big(b^{-1}(x)U_t\Big(x,G_t(x,y)\Big), U\Big(x,G_t(x,y)\Big)\Big)\Big]. \end{aligned} \tag{5.17}$$

We will expand the functions $[\det_2(I_d+\cdot)]^{-1}, \exp[\cdot]$ in this expression around the point $(0,0)\in\mathbb{R}^{d\times d}\times\mathbb{R}$ using the Taylor expansion formula. For that, we first analyze the terms inside these functions, giving two types of bounds: (i) t-order; (ii) uniform for $y\in C_{x,t}$. The latter one is needed because we would like to apply the Taylor formula in a bounded domain.

Lemma 5.1. *There exists a function $Q_t^1(x,y)$, $(t,x,y)\in(0,t_0)\times\mathbb{R}^d\times\mathbb{R}^d$ of t-order 1/2 such that*

$$\Big|\nabla_2 U_t\Big(x, G_t(x,y)\Big)\Big| = Q_t^1(x,y), \quad y\in C_{x,t}. \tag{5.18}$$

In addition,

$$\Big|\nabla_2 U_t\Big(x, G_t(x,y)\Big)\Big| \le Ct^{1/3}, \quad y\in C_{x,t}. \tag{5.19}$$

Proof. Recall that $\nabla_2 U_t(x,\cdot)$ is a linear function, see (5.12). Then

$$\Big|\nabla_2 U_t\Big(x, G_t(x,y)\Big)\Big| \le |\nabla_2 U_t(x,y)| + C|G_t(x,y)-y|.$$

By Corollary 5.1, the second term is dominated on $C_{x,t}$ by a function with t-order 1. The first term $|\nabla_2 U_t(x,y)|$ itself has t-order 1/2. This proves the first statement.

To prove the second statement we recall that $G_t(x,y)\in B_{x,t,1}$ for $y\in C_{x,y}$. In addition, it follows directly from the formula (5.12) that $|\nabla_2 U_t(x,z)|\le Ct^{1/3}$ for $z\in B_{x,t,1}$. This completes the proof. ☐

Lemma 5.2. *There exists a function $Q_t^2(x,y)$ of t-order 1 such that*

$$\begin{aligned}\delta_t^U\Big(x,G_t(x,y)\Big) + \frac{1}{2t}\Big(b^{-1}(x)U_t\Big(x,G_t(x,y)\Big), U_t\Big(x,G_t(x,y)\Big)\Big) =\\ = \delta_t^U(x,y) + Q_t^2(x,y), \quad y\in C_{x,t}.\end{aligned} \tag{5.20}$$

In addition,

$$\Big|\delta_t^U\Big(x,G_t(x,y)\Big) + \frac{1}{2t}\Big(b^{-1}(x)U_t\Big(x,G_t(x,y)\Big), U_t\Big(x,G_t(x,y)\Big)\Big)\Big| \le C, \quad y\in C_{x,t}. \tag{5.21}$$

Proof. By the explicit expression (3.5), the function $\delta_t^U(x,G_t(x,y))$ is a 3rd order polynomial in the variable y. Since $G_t(x,y)-x-a(x)t = (G_t(x,y)-y)+(y-x-a(x)t)$, this yields

$$\begin{aligned}\Big|\delta_t^U\Big(x,G_t(x,y)\Big) - \delta_t^U(x,y)\Big| &\le C\sum_{r=1}^{3}|y-x-a(x)t|^{3-r}t^{-1}|G_t(x,y)-y|^r =\\ &= C\sum_{r=1}^{3}|y-x-a(x)t|^{3-r}t^{-1}|R_t(x,y)|^r, \quad y\in C_{x,t}.\end{aligned}$$

Note that in the last line we have used Corollary 5.1. The function on the right hand side has t-order

$$\min\Big(0+1, -1/2+2, -1+3\Big) = 1.$$

Similarly, the map

$$y\mapsto \frac{1}{2t}\Big(b^{-1}(x)U_t(x,y), U_t(x,y)\Big)$$

is a 4th order polynomial which has t-order 1, and

$$\Big|\frac{1}{2t}\Big(b^{-1}(x)U_t\Big(x,G_t(x,y)\Big), U\Big(x,G_t(x,y)\Big)\Big) - \frac{1}{2t}\Big(b^{-1}(x)U_t(x,y), U_t(x,y)\Big)\Big|$$

can be bounded on $C_{x,t}$ by a function which has t-order

$$\min\Big(1/2+1, 0+2, -1/2+3, -1+4\Big) = 3/2.$$

Collecting these bounds together, we get (5.20).

To get the second statement, (5.21), we use again that $G_t(x,y) \in B_{x,t,1}$ if $y \in C_{x,t}$. Using the explicit formulae for the Hermite polynomials (see Section A), for $z \in B_{x,t,1}$ one obtains estimates which can be applied to (3.4), (3.5) as follows

$$\begin{aligned}
|H_t^{(i,j)}(x,z)| &\le C\frac{(t^{1/3})^2}{t^2} + C\frac{1}{t} \le Ct^{2/3-2},\\
|H_t^{(i,j,k)}(x,z)| &\le C\frac{(t^{1/3})^3}{t^3} + C\frac{(t^{1/3})}{t^2} \le Ct^{-2},\\
0 \le \frac{1}{2t}\Big(b^{-1}(x)U_t(x,z), U_t(x,z)\Big) &\le Ct^{-1}\Big(t^2 t^{2/3-2}\Big)^2 = Ct^{1/3},\\
|\delta_t^U(x,z)| &\le Ct^2 t^{-2} = C,
\end{aligned} \tag{5.22}$$

which completes the proof of the second statement if we use the above bounds for $z = G_t(x,y)$. □

Remark 5.1. Note that the bound in (5.22) is the essential estimate which requires the factor $t^{1/3}$ in the definition of $C_{x,t}$. In turn, the estimate (5.22) is required to get the uniform bound (5.21), which is needed since we will apply the Taylor expansion in a bounded domain.

Now it is easy to complete the proof of Proposition 5.2. The Taylor formulae for the functions

$$\det{}_2(I_d + A) = \det(I_d + A)\exp(-\operatorname{trace} A), \quad A \in \mathbb{R}^{d\times d}, \quad \exp(w), \qquad w \in \mathbb{R},$$

yield that, for $r_0 > 0$ small enough and arbitrary $R > 0$,

$$\Big|[\det{}_2(I_d + A)]^{-1}\exp(w) - 1 - w\Big| \le C(|A|^2 + w^2), \quad |A| \le r_0, \quad |w| \le R.$$

We apply the above expansion for $A = \nabla_2 U_t(x, G_t(x,y))$ and w given by the left hand side of (5.20). Then, by Lemmas 5.1 and 5.2, for t_0 small enough we have for $t \le t_0$, $y \in C_{x,t}$

$$\Big[J^{F_{x,t},P_{x,t}}\Big(G_t(x,y)\Big)\Big]^{-1} = 1 + \delta_t^U(x,y) + Q_t^2(x,y) + Q_t^3(x,y)$$

with

$$|Q_t^3(x,y)| \le C\Big[\Big(Q_t^1(x,y)\Big)^2 + \Big(\delta_t^U(x,y) + Q_t^2(x,y)\Big)^2\Big].$$

Since $Q_t^1(x,y), \delta_t^U(x,y), Q_t^2(x,y)$ have t-orders $1/2, 1/2, 1$ respectively, function $Q_t^3(x,y)$ has t-order 1. Now we have (5.10) with

$$Q_t(x,y) = \Big[Q_t^2(x,y) + Q_t^3(x,y)\Big]1_{C_{x,t}}(y),$$

which has t-order 1. □

6. Proof of Theorem 1.1

We consider the second order scheme (1.6). The process $\overline{X}^{x,2}(t)$ can be written as

$$\overline{X}_i^{x,2}(t) = \overline{X}_i^x(t) + \frac{1}{2}\sum_{u,v=1}^m \rho_{iuv}(x)t^2\mathcal{H}_t^{(u,v)}(W(t))\mathbf{e}_i.$$

Recall that $m = d$ and write (A.1) in the form

$$\mathcal{H}_t^{(j_1,\dots,j_n)}(w) = \sum_{i_1,\dots,i_n=1}^{d} \sigma^*_{j_1 i_1}\dots\sigma^*_{j_n i_n} H_t^{(i_1,\dots,i_n)}(\mathbf{a},\mathbf{b},x,y) =$$
$$= \sum_{i_1,\dots,i_n=1}^{d} \sigma_{i_1 j_1}\dots\sigma_{i_n j_n} H_t^{(i_1,\dots,i_n)}(\mathbf{a},\mathbf{b},x,y), \quad w = \sigma^{-1}(y - x - t\mathbf{a}).$$

Note that $b(x) = \sigma\sigma^*(x)$ and

$$W(t) = \sigma^{-1}(x)(\overline{X}^x(t) - x - ta(x)).$$

Hence the above formula gives

$$\mathcal{H}_t^{(u,v)}(W(t)) = \sum_{j,k=1}^{d} \sigma_{ju}(x)\sigma_{kv}(x) H_t^{(j,k)}(x, \overline{X}^x(t)).$$

That is, the process $\overline{X}^{x,2}(t)$ can be written in the form (3.8), where the function $U_t(x,y)$ has representation (3.4) with

$$f_{ijk}(x) = \frac{1}{2}\sum_{u,v=1}^{d} \rho_{iuv}(x)\sigma_{ju}\sigma_{kv}(x) = \frac{1}{2}\sum_{u,v,l=1}^{d} \partial_{x_l}\sigma_{iu}\sigma_{lv}\sigma_{ju}\sigma_{kv}(x) =$$
$$= \frac{1}{2}\sum_{u,l=1}^{d} \partial_{x_l}\sigma_{iu}\sigma_{ju} b_{kl}(x).$$

Then

$$f^{\mathrm{sym}}_{ijk}(x) = \frac{1}{12}\sum_{u,l=1}^{d}\Big(\partial_{x_l}\sigma_{iu}\sigma_{ju}b_{kl} + \partial_{x_l}\sigma_{iu}\sigma_{ku}b_{jl} + \partial_{x_l}\sigma_{ju}\sigma_{iu}b_{kl} +$$
$$+ \partial_{x_l}\sigma_{ju}\sigma_{ku}(x)b_{il} + \partial_{x_l}\sigma_{ku}\sigma_{iu}b_{jl} + \partial_{x_l}\sigma_{ku}\sigma_{ju}b_{il}\Big)(x).$$

On the other hand, the coefficients $c_{ijk}(x)$ are symmetric w.r.t. i, j and thus, taking into account that $b = \sigma\sigma^*$, we get

$$c^{\mathrm{sym}}_{ijk}(x) = \frac{1}{3}\Big(c_{ijk} + c_{ikj} + c_{jki}\Big)(x) = \frac{1}{12}\sum_{l=1}^{d}\Big(b_{kl}\partial_{x_l}b_{ij} + b_{il}\partial_{x_l}b_{jk} + b_{jl}\partial_{x_l}b_{ik}\Big)(x) =$$
$$= \frac{1}{12}\sum_{u,l=1}^{d}\Big(b_{kl}\big(\partial_{x_l}\sigma_{iu}\sigma_{ju} + \partial_{x_l}\sigma_{ju}\sigma_{iu}\big) + b_{il}(x)\big(\partial_{x_l}\sigma_{ju}\sigma_{ku} + \partial_{x_l}\sigma_{ku}\sigma_{ju}\big) +$$
$$+ b_{jl}\big(\partial_{x_l}\sigma_{iu}\sigma_{ku} + \partial_{x_l}\sigma_{ku}\sigma_{iu}\big)\Big)(x) =$$
$$= \frac{1}{12}\sum_{u,l=1}^{d}\Big(\partial_{x_l}\sigma_{iu}\sigma_{ju}b_{kl} + \partial_{x_l}\sigma_{ju}\sigma_{iu}b_{kl} + \partial_{x_l}\sigma_{ju}\sigma_{ku}(x)b_{il} + \partial_{x_l}\sigma_{ku}\sigma_{ju}(x)b_{il} +$$
$$+ \partial_{x_l}\sigma_{iu}\sigma_{ku}(x)b_{jl} + \partial_{x_l}\sigma_{ku}\sigma_{iu}(x)b_{jl}\Big)(x) = f^{\mathrm{sym}}_{ijk}(x),$$

and (3.10) is satisfied. Under the conditions of Theorem 3.1 the coefficients $f_{ijk}(x)$ are bounded, thus Corollary 3.1 can be applied, which completes the proof of the claim. □

Appendix A. Hermite functions and polynomials

In this section, we gather a series of identities and properties of general Hermite functions and polynomials as introduced in Section 2. These functions/polynomials are not usually used in such generality in the literature, which typically focuses on standardized

Gaussian measures $\mathcal{N}(0, \mathrm{Id}_{\mathbb{R}^d})$ or $\mathcal{N}(0, t\mathrm{Id}_{\mathbb{R}^d})$; thus their properties can be hardly found explicitly written[7]

To provide a comprehensive discussion, in addition to the Hermite functions introduced in Section 2, we will also consider their standardized analogues

$$\varphi_t(0,w) = (2\pi t)^{-d/2}\exp\left(-\frac{1}{2t}|w|^2\right),$$
$$\varphi_t^{(i_1,\dots,i_n)}(0,w) = (-1)^n\partial_{w_{i_1}}\dots\partial_{w_{i_n}}\varphi_t(0,w), \quad w\in\mathbb{R}^d,$$

and corresponding Hermite polynomials defined by

$$\varphi_t^{(i_1,\dots,i_n)}(0,w) = \mathcal{H}_t^{(i_1,\dots,i_n)}(w)\varphi_t(0,w).$$

Note that the Hermite functions/polynomials introduced in Section 2 reduce to the above in the particular case that $\mathbf{a}=0$ and $\mathbf{b}=I_d$. Furthermore, let $\mathbf{a}\in\mathbb{R}^d, \mathbf{b}\in\mathbb{R}^{d\times d}_{\mathrm{sym}}$ be given. As $\mathbf{b}$ is assumed to be a strictly positive definite matrix then there exists a $d\times d$ matrix σ such that $\mathbf{b}=\sigma\sigma^*$. From now on, σ is considered to be fixed and it is invertible since $\mathbf{b}\in\mathbb{R}^{d\times d}_{\mathrm{sym}}$. It follows simply by the definition that if we let $w=\sigma^{-1}(y-x-t\mathbf{a})$ then

$$\Phi_t(\mathbf{a},\mathbf{b};x,y) = |\det\sigma|^{-1}\varphi_t(0,w).$$

Thus by the chain rule we have the following relation:

$$\begin{aligned} H_t^{(i_1,\dots,i_n)}(\mathbf{a},\mathbf{b};x,y) &= \sum_{j_1,\dots,j_n=1}^{d}\mathcal{H}_t^{(j_1,\dots,j_n)}(w)(\sigma^{-1})_{j_1i_1}\dots(\sigma^{-1})_{j_ni_n} = \\ &= \sum_{j_1,\dots,j_n=1}^{d}((\sigma^*)^{-1})_{i_1j_1}\dots((\sigma^*)^{-1})_{i_nj_n}\mathcal{H}_t^{(j_1,\dots,j_n)}(w). \end{aligned} \tag{A.1}$$

Therefore the above formulas show that any property on the standardized Hermite polynomials $\mathcal{H}$ can be transferred to the general Hermite polynomials H.

For example,

$$\mathcal{H}_t^{(i)}(w) = \frac{1}{t}w_i, \quad H_t^{(i)}(\mathbf{a},\mathbf{b};x,y) = \frac{1}{t}\Big(\mathbf{b}^{-1}(y-x-\mathbf{a}t)\Big)_i,$$

the second identity can be either derived directly or obtained from the first one using (A.1) and the fact that $\mathbf{b}^{-1} = (\sigma^*)^{-1}\sigma^{-1}$. Similarly,

$$\mathcal{H}_t^{(i,j)}(w) = \frac{1}{t^2}w_iw_j - \frac{1}{t}1_{i=j},$$
$$H_t^{(i,j)}(\mathbf{a},\mathbf{b};x,y) = \frac{1}{t^2}\Big(\mathbf{b}^{-1}(y-x-\mathbf{a}t)\Big)_i\Big(\mathbf{b}^{-1}(y-x-\mathbf{a}t)\Big)_j - \frac{1}{t}\mathbf{b}^{-1}_{ij}$$

and

$$\mathcal{H}_t^{(i,j,k)}(w) = \frac{1}{t^3}w_iw_jw_k - \frac{1}{t^2}w_i1_{j=k} - \frac{1}{t^2}w_j1_{i=k} - \frac{1}{t^2}w_k1_{i=j},$$

$$\begin{aligned} H_t^{(i,j,k)}(\mathbf{a},\mathbf{b};x,y) = &\frac{1}{t^3}\Big(\mathbf{b}^{-1}(y-x-\mathbf{a}t)\Big)_i\Big(\mathbf{b}^{-1}(y-x-\mathbf{a}t)\Big)_j\Big(\mathbf{b}^{-1}(y-x-\mathbf{a}t)\Big)_k - \\ &- \frac{1}{t^2}\mathbf{b}^{-1}_{ij}\Big(\mathbf{b}^{-1}(y-x-\mathbf{a}t)\Big)_k - \frac{1}{t^2}\mathbf{b}^{-1}_{ik}\Big(\mathbf{b}^{-1}(y-x-\mathbf{a}t)\Big)_j - \\ &- \frac{1}{t^2}\mathbf{b}^{-1}_{jk}\Big(\mathbf{b}^{-1}(y-x-\mathbf{a}t)\Big)_i. \end{aligned}$$

[7] For an exception using tensorial notation, see Section 5.7 in [4].

Next, we have the identities

$$\begin{aligned} H_t^{(i_1,\dots,i_{m-1},i,i_{m+1},\dots,i_n)}&(\mathbf{a},\mathbf{b};x,y) = \\ &= \frac{1}{t}\Big[\mathbf{b}^{-1}(y-x-t\mathbf{a})\Big]_i H_t^{(i_1,\dots,i_{m-1},i_{m+1},\dots,i_n)}(\mathbf{a},\mathbf{b};x,y) - \\ &\quad - \partial_{y_i} H_t^{(i_1,\dots,i_{m-1},i_{m+1},\dots,i_n)}(\mathbf{a},\mathbf{b};x,y), \end{aligned} \tag{A.2}$$

$$\partial_{y_i} H_t^{(i_1,\dots,i_n)}(\mathbf{a},\mathbf{b};x,y) = \frac{1}{t}\sum_{m=1}^n \mathbf{b}^{-1}_{i_m i} H_t^{(i_1,\dots,i_{m-1},i_{m+1},\dots,i_n)}(\mathbf{a},\mathbf{b};x,y), \tag{A.3}$$

which can be derived either directly or using (A.1) and the well known identities for canonic Hermite polynomials:

$$\mathcal{H}_t^{(i_1,\dots,i_{m-1},i,i_{m+1},\dots,i_n)}(w) = \frac{1}{t} w_i \mathcal{H}_t^{(i_1,\dots,i_{m-1},i_{m+1},\dots,i_n)}(w) - \partial_{w_i}\mathcal{H}_t^{(i_1,\dots,i_{m-1},i_{m+1},\dots,i_n)}(w),$$

$$\partial_{w_i}\mathcal{H}_t^{(i_1,\dots,i_n)}(w) = \frac{1}{t}\sum_{m=1}^n 1_{i_m=i}\mathcal{H}_t^{(i_1,\dots,i_{m-1},i_{m+1},\dots,i_n)}(x,y).$$

Combining (A.2) and (A.3) and re-arranging the notation for the multi-indices we get

$$\begin{aligned} \Big[\mathbf{b}^{-1}&(y-x-t\mathbf{a})\Big]_i H_t^{(i_1,\dots,i_n)}(\mathbf{a},\mathbf{b};x,y) = \\ &= tH_t^{(i,i_1,\dots,i_n)}(x,y) + \sum_{m=1}^n \mathbf{b}^{-1}_{i_m i} H_t^{(i_1,\dots,i_{m-1},i_{m+1},\dots,i_n)}(x,y). \end{aligned} \tag{A.4}$$

Multiplying (A.4) by $\mathbf{b}_{ji}$ and taking the sum over i, we obtain

$$\begin{aligned} (y-x-t\mathbf{a})_j H_t^{(i_1,\dots,i_n)}(\mathbf{a},\mathbf{b};x,y) &= t\sum_{i=1}^d \mathbf{b}_{ji} H_t^{(i,i_1,\dots,i_n)}(\mathbf{a},\mathbf{b};x,y) + \\ &\quad + \sum_{m=1}^n 1_{i_m=j} H_t^{(i_1,\dots,i_{m-1},i_{m+1},\dots,i_n)}(x,y). \end{aligned} \tag{A.5}$$

The Hermite functions $\Phi_t^{(i_1,\dots,i_n)}(\mathbf{a},\mathbf{b};x,y)$ also have nice properties as functions of the parameters $\mathbf{a}\in\mathbb{R}^d, \mathbf{b}\in\mathbb{R}^{d\times d}_{\text{sym}}$:

$$\partial_{\mathbf{a}_j}\Phi_t^{(i_1,\dots,i_n)}(\mathbf{a},\mathbf{b};x,y) = t\Phi_t^{(i_1,\dots,i_n,j)}(\mathbf{a},\mathbf{b};x,y), \tag{A.6}$$

$$\partial_{\mathbf{b}_{jk}}\Phi_t^{(i_1,\dots,i_n)}(\mathbf{a},\mathbf{b};x,y) = \frac{t}{2}\Phi_t^{(i_1,\dots,i_n,j,k)}(\mathbf{a},\mathbf{b};x,y). \tag{A.7}$$

These properties are difficult to prove using the canonic Hermite functions, thus we provide the direct proof based on the Fourier transform. Namely, let

$$\widehat{\Phi}_t^{(i_1,\dots,i_n)}(\mathbf{a},\mathbf{b};x,\lambda) = \int_{\mathbb{R}^d} e^{\mathbf{i}y\cdot\lambda}\Phi_t^{(i_1,\dots,i_n)}(\mathbf{a},\mathbf{b};x,y)\,dy, \quad \lambda\in\mathbb{R}^d,$$

then

$$\widehat{\Phi}_t(\mathbf{a},\mathbf{b};x,\lambda) = e^{t\mathbf{i}\mathbf{a}\cdot\lambda - \frac{t}{2}(\mathbf{b}\lambda,\lambda)}$$

and

$$\begin{aligned} \widehat{\Phi}_t^{(i_1,\dots,i_n)}(\mathbf{a},\mathbf{b};x,\lambda) &= \int_{\mathbb{R}^d} e^{\mathbf{i}y\cdot\lambda}(-\partial_{y_{i_1}})\dots(-\partial_{y_{i_n}})\Phi_t(\mathbf{a},\mathbf{b};x,y)\,dy = \\ &= \int_{\mathbb{R}^d}\partial_{y_{i_1}}\dots\partial_{y_{i_n}} e^{\mathbf{i}y\cdot\lambda}\Phi_t(\mathbf{a},\mathbf{b};x,y)\,dy = \\ &= (\mathbf{i}\lambda_{i_1})\dots(\mathbf{i}\lambda_{i_n})\widehat{\Phi}_t(\mathbf{a},\mathbf{b};x,\lambda) = (\mathbf{i}\lambda_{i_1})\dots(\mathbf{i}\lambda_{i_n})e^{\mathbf{i}\mathbf{a}\cdot\lambda - \frac{t}{2}(\mathbf{b}\lambda,\lambda)}, \quad \lambda\in\mathbb{R}^d. \end{aligned} \tag{A.8}$$

Then, we conclude that

$$\partial_{\mathbf{b}_{jk}} \widehat{\Phi}_t^{(i_1,\dots,i_n)}(\mathbf{a},\mathbf{b};x,\lambda) = -\frac{t}{2}(\mathbf{i}\lambda_{i_1})\dots(\mathbf{i}\lambda_{i_n})\lambda_j\lambda_k e^{\mathbf{i}\mathbf{a}\cdot\lambda-\frac{1}{2}(\mathbf{b}\lambda,\lambda)} = \frac{t}{2}\widehat{\Phi}_t^{(i_1,\dots,i_n,j,k)}(\mathbf{a},\mathbf{b};x,\lambda),$$

which proves (A.7). The proof of (A.6) is similar and omitted.

Finally, we mention the following convolution property of the Hermite functions: for any $t, s > 0$ and $m, n \geq 0$, $i_1, \dots, i_n, j_1, \dots, j_m$

$$\Big(\Phi_t^{(i_1,\dots,i_n)} * \Phi_s^{(j_1,\dots,j_m)}\Big)(\mathbf{a},\mathbf{b};x,y) = \Phi_{t+s}^{(i_1,\dots,i_n,j_1,\dots,j_m)}(\mathbf{a},\mathbf{b};x,y). \tag{A.9}$$

Identity (A.9) easily follows from (A.8) and the fact that the Fourier transform of a convolution is a product.

Appendix B. The divergence operator

In this section we briefly discuss the properties of the divergence operator (5.3) used in the proof of Theorem 3.2; with this perspective in mind, we will only consider the particular case of $P = P_{x,t} \sim \mathcal{N}(ta(x), tb(x))$. It follows from (5.3) and (A.2) that for

$$V_t(x,y) = H_t^{(i_1,\dots,i_n)}(x,y)\mathbf{e}_i,$$

$$\begin{aligned}\Big[\delta_{P_{x,t}} V_t(x,\cdot)\Big](y) &= \frac{1}{t}\Big[b^{-1}(x)(y-x-ta(x))\Big]_i H_t^{(i_1,\dots,i_n)}(x,y) - \partial_{y_i} H_t^{(i_1,\dots,i_n)}(x,y) = \\ &= H_t^{(i,i_1,\dots,i_n)}(x,y).\end{aligned} \tag{B.1}$$

Next, recall that the function $U_t(x,y)$ is defined in (3.4), then by (B.1) (see (3.5))

$$\Big[\delta_{P_{x,t}} U_t(x,\cdot)\Big](y) = t^2 \sum_{i,j,k=1}^{d} f_{ijk}(x) H_t^{(i,j,k)}(x,y) = \delta_t^U(x,y), \tag{B.2}$$

That is, the function $\delta_t^U(x,y)$ in Theorem 3.2 is just the Skorokhod integral of $U_t(x,y)$ w.r.t. $P_{x,t}$ and is the function that matches the second term on the right hand side of (3.1).

Therefore (3.11) can be naturally understood as just an equation of the form

$$\delta_P U = g \tag{B.3}$$

for a given polynomial function g and unknown U. The structure of the set of solutions to (B.3) is well understood; e.g. [12]. Namely, *any* solution has the form $U = U^0 + V$, where U^0 is *some* solution, and V is a solution to the homogeneous equation $\delta_P V = 0$. The latter is equivalently described as an $\mathbb{R}^d$-valued function such that, for any $Q \in C^1(\mathbb{R}^d)$ with compact support,

$$\int_{\mathbb{R}^d} (V_t(x,y), \nabla Q(y)) P_{x,t}(\mathrm{d}y) = 0.$$

Therefore, the set of solutions to (B.3) is rather wide. In Corollary 3.1 we actually list a subclass of all the solutions of such an equation, with particular 3rd order polynomial for g, given *only* by the 2nd order polynomials. Such a restriction is natural, since our main goal is to make the improved approximations (3.8) as simple as is possible.

Acknowledgment

The research of the first author and the second author was supported by Alexander von Humboldt Foundation within the Research Group Linkage Programme between the Institute of Mathematics at the University of Potsdam and the Institute of Mathematics of National Academy of Sciences of Ukraine. The research of the third author was supported by KAKENHI grants 24340022 and 16H03642.

References

1. Y. Ait-Sahahia. *Closed-form likelihood expansions for multivariate diffusions*, Annals of Statistics, **36** (2008), 906–937.
2. A. Alfonsi, B. Jourdain, A. Kohatsu-Higa. *Optimal transport bounds between the time-marginals of a multidimensional diffusion and its Euler scheme*, Electronic Journal of Probability, **20** (2015), 1–31.
3. R. Azencott. *Densité des diffusions en temps petit: développements asymptotiques. I,* Seminar on probability, XVIII, Lecture Notes in Math. 1059, Springer, Berlin, 1984, 402–498.
4. O. E. Barndorff-Nielsen, D. R. Cox. *Asymptotic Techniques for Use in Statistics*, Chapman and Hall, 1989.
5. A. B. Cruzeiro, P. Malliavin, A. Thalmaier. *Geometrization of Monte-Carlo numerical analysis of an elliptic operator: strong approximation*, C. R. Math. Acad. Sci. Paris, **338** (2004), 481–486.
6. A. Friedman. *Partial differential equations of parabolic type*, Dover Publications, Inc., 1964.
7. P. E. Kloeden, E. Platen. *Numerical Solution of Stochastic Differential Equations*, Series: Stochastic Modelling and Applied Probability, Vol. 23. 1992.
8. D. O. Ivanenko, A. Kohatsu-Higa, A. M. Kulik, *LA(M)N property for diffusion models with low regularity coefficients*, *Manuscript.*
9. S. Kusuoka. *Approximation of expectation of diffusion process and mathematical finance*, Taniguchi Conference on Mathematics Nara '98, Adv. Stud. Pure Math., Math. Soc. Japan, Tokyo, **131** (2001), 147–165.
10. K. Oshima, J. Teichmann, D. Velušček. *A new extrapolation method for weak approximation schemes with applications*, Ann. Appl. Probab., **22** (2012), 1008–1045.
11. R. Ramer. On nonlinear transformations of Gaussian measures. J. Funct. Anal., 1974, **15**, 166–187.
12. A. Kulik. *Integral representation for functionals on a space with a smooth measure*, Theory of stochastic processes, **3** (1997), 235–243.
13. A. Üstünel, M. Zakai. *The change of variables formula on Wiener space*, Séminaire de probabilités de Strasbourg, **31** (1997), 24–39.
14. S. Watanabe. *Short time asymptotic problems in Wiener functional integration theory.* Applications to heat kernels and index theorems. Stochastic analysis and related topics, II (Silivri, 1988), Lecture Notes in Math., **1444**, Springer, Berlin, 1990, 1–62.
15. S. Watanabe. *Analysis of Wiener functionals (Malliavin calculus) and its applications to heat kernels*, Ann. Probab., **15** (1987), 1–39.
16. N. Yoshida. *Asymptotic expansions of maximum likelihood estimators for small diffusions via the theory of Malliavin–Watanabe*, Probab. Theory Related Fields, **92** (1992), 275–311.

Department of Mathematical Analysis and Probability Theory, Igor Sikorsky Kyiv Polytechnic Institute, Prosp. Peremogy 37, 03056 Kyiv, Ukraine
E-mail address: sem_bodn@ukr.net

Department of Radio Physics, Electronics and Computer Systems, of Taras Shevchenko National University of Kyiv, Hlushkova Avenue 4g, 03127 Kyiv, Ukraine
E-mail address: divanenko1979@gmail.com

Department of Mathematical Sciences, Ritsumeikan University, 1-1-1 Nojihigashi, Kusatsu, Shiga, 525-8577, Japan.
E-mail address: khts00@fc.ritsumei.ac.jp

Wroclaw University of Science and Technology, Wybrzeźe Wyspiańskiego Str. 27, 50-370 Wroclaw, Poland
E-mail address: kulik.alex.m@gmail.com

ПОКРАЩЕНА ЛОКАЛЬНА АПРОКСИМАЦІЯ ДЛЯ БАГАТОВИМІРНИХ ДИФУЗІЙ: G-ОЦІНКИ

С. БОДНАРЧУК, Д. ІВАНЕНКО, А. КОХАЦУ-ХІГА, О. КУЛИК

Анотація. Розглянуто задачу покращення локальної апроксимації для багатовимірних дифузій. Зокрема, запропонована нами явна схема апроксимації покращує схему Мільштейна. Ми також даємо частково явну оцінку точності апроксимації (ми називаємо її G-оцінкою), в якій головний член обмежений добутком полінома на гауссову щільність, а залишок є експоненційно малим для малого часу.