

EXACT SIMULATION FOR MULTIVARIATE ITÔ DIFFUSIONS

JOSE BLANCHET,^{**} AND
FAN ZHANG,^{***} *Stanford University*

Abstract

We provide the first generic exact simulation algorithm for multivariate diffusions. Current exact sampling algorithms for diffusions require the existence of a transformation which can be used to reduce the sampling problem to the case of a constant diffusion matrix and a drift which is the gradient of some function. Such a transformation, called the Lamperti transformation, can be applied in general only in one dimension. So, completely different ideas are required for the exact sampling of generic multivariate diffusions. The development of these ideas is the main contribution of this paper. Our strategy combines techniques borrowed from the theory of rough paths, on the one hand, and multilevel Monte Carlo on the other.

Keywords: Exact simulation; stochastic differential equation; Brownian motion; Monte Carlo method

2010 Mathematics Subject Classification: Primary 34K50; 65C05; 82B80
Secondary 97K60

1. Introduction

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and an Itô stochastic differential equation (SDE)

$$dX(t) = \mu(X(t))dt + \sigma(X(t))dW(t), \quad X(0) = x_0, \quad (1)$$

where $W(\cdot)$ is a d' -dimensional Brownian motion under \mathbb{P} , and $\mu(\cdot) = (\mu_i(\cdot))_d : \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $\sigma(\cdot) = (\sigma_{ij}(\cdot))_{d \times d'} : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d'}$ satisfy suitable regularity conditions. For instance, in order for (1) to have a strong solution, it is sufficient to assume that both $\mu(\cdot)$ and $\sigma(\cdot)$ are uniformly Lipschitz.

Under additional regularity conditions on $\mu(\cdot)$ and $\sigma(\cdot)$, this paper provides the first Monte Carlo simulation algorithm which allows sampling of any discrete skeleton $X(t_1), \dots, X(t_m)$ exactly, without any bias.

The precise regularity conditions that we impose on $\mu(\cdot)$ and $\sigma(\cdot)$ are stated in Section 3. In particular, it is sufficient for the validity of our Monte Carlo method to assume $\mu(\cdot)$ and $\sigma(\cdot)$ to be three times continuously differentiable, both with Lipschitz continuous derivatives of order three. In addition, we must assume that $\sigma(\cdot)$ is uniformly elliptic.

Exact simulation of SDEs has generated a substantial amount of interest in the applied probability and Monte Carlo simulation communities. The landmark paper [5] introduced what has become the standard procedure for the design of generic exact simulation algorithms for diffusions. The authors of [5] propose a clever acceptance-rejection sampler which uses Brownian

Received 12 December 2017; revision received 1 May 2020.

* Postal address: Huang Engineering Center, 475 Via Ortega, Stanford, CA 94305, United States.

** Email address: jose.blanchet@stanford.edu

*** Email address: fzh@stanford.edu

motion as a proposal distribution. The authors of [7] apply a localization technique which eliminates certain boundedness assumptions originally present in [5]; see also [4] for the use of retrospective simulation ideas to dispense with boundedness assumptions.

The fundamental assumption underlying the work of [5] and its extensions is that the underlying (target) process has a constant diffusion coefficient; i.e., $\sigma(x) = \sigma$ for every x . Beskos and Roberts [5] note that in the case $d = 1$, owing to Lamperti's transformation, the constant diffusion coefficient assumption comes basically at no cost in generality.

Unfortunately, however, Lamperti's transformation is only generally applicable in one dimension. In fact, [1] characterizes the multidimensional diffusions for which Lamperti's transformation can be successfully applied, and these models are very restrictive.

Moreover, even if Lamperti's transformation is applicable in a multidimensional setting, another implicit assumption in the application of the Beskos and Roberts acceptance-rejection procedure is that the drift coefficient $\mu(\cdot)$ is the gradient of some function (i.e. $\mu(x) = \nabla v(x)$ for some $v(\cdot)$). This assumption, once again, comes at virtually no cost in generality in the one-dimensional setting, but it may be very restrictive in the multidimensional case.

Because of these limitations, a generic algorithm for exact simulation of multidimensional diffusions, even under the regularity conditions that we impose here, requires a completely different set of ideas.

The contribution in this paper is therefore not only the production of such a generic exact simulation algorithm, but also the development of the ideas that are behind its construction. In Section 2 we introduce an algorithm which assumes a constant diffusion coefficient, but removes the assumption of the drift coefficient being the gradient of some function. In Section 3, we eventually remove the requirement of a constant diffusion matrix and propose an algorithm applicable to general diffusions. The algorithms in Sections 2 and 3 are different in nature. However, they share some common elements, such as the use of so-called Tolerance-Enforced Simulation (TES) techniques based on rough path estimates. Even though the algorithm in Section 3 is more general, we believe that there is significant value in developing the algorithm in Section 2, for two reasons. The first one is pedagogical: the algorithm in Section 2 is easier to understand, while building on a key idea, which involves localizing essential quantities within specific compact domains with probability one. The second reason is that the algorithm in Section 2, being simpler, may be subject to potential improvement methodologies to be pursued in future research.

Potential improvements are particularly interesting directions, especially given that, unfortunately, the algorithms that we present have infinite expected termination time. We recognize that this issue should be resolved for the algorithms to be widely used in practice, and we discuss the elements which lead to infinite expected running time in Section 4. There are basically two types of elements that affect the running time of the algorithm in Section 3; one of them has to do with the types of issues that arise when trying to fully remove the bias in TES and related approximations, and the other has to do with the use of a density approximation coupled with Bernoulli factories. In contrast, the algorithm in Section 2 is only affected by removing the bias in TES-type approximations. We must stress, however, that the present paper shows for the first time that it is possible to perform exact sampling of multidimensional diffusions in substantial generality, and, in doing so, it provides a conceptual framework different from the prevailing use of Lamperti transformation, which is the only available generic approach for producing exact sampling of diffusions.

Now, despite the algorithm's practical limitations, it is vital to recognize the advantages that unbiased samplers have over biased samplers in the context of a massive parallel computing

environment, because it is straightforward to implement a parallel procedure to reduce the estimation error of an unbiased sampler.

Recently, there have been several unbiased estimation procedures which have been proposed for expectations of the form $\alpha = \mathbb{E}(f(X(t)))$, assuming $\text{Var}(f(X(t))) < \infty$. For example, the work of [21] shows that if $f(\cdot)$ is twice continuously differentiable (with Lipschitz derivatives) and if there exists a discretization scheme which can be implemented with a strong convergence error of order 1, then it is possible to construct an unbiased estimator for α with finite variance and finite expected termination time. The work of [12] shows that such a discretization scheme can be developed if $\mu(\cdot)$ and $\sigma(\cdot)$ are sufficiently smooth under certain boundedness assumptions. The paper [13] also develops an unbiased estimator for α using a regime-switching technique. Our work here is somewhat related to this line of research, but an important difference is that we *actually generate* $X(t)$ exactly, while *all* of the existing algorithms which apply in multidimensional diffusion settings generate Z such that $\mathbb{E}(Z) = \alpha$. So, for example, if $f(\cdot)$ is positive, one cannot guarantee that Z is positive using the type of samplers suggested in [21]. However, by sampling $X(t)$ directly, one maintains the positivity of the estimator.

Another instance in which direct exact samplers are useful arises in the context of stochastic optimization. For instance, consider the case in which one is interested in optimizing a convex function of the form $g(\theta) = \mathbb{E}(h(X(t), \theta))$, where $h(x, \cdot)$ is differentiable. In this case, one can naturally construct an estimator $Z(\theta)$ such that $g(\theta) = \mathbb{E}(Z(\theta))$ using the results in [21] and optimize the mapping $\theta \rightarrow n^{-1} \sum_{i=1}^n Z_i(\theta)$, which unfortunately will typically not be convex. So, having access to a direct procedure to sample $X(t)$ in this setting is particularly convenient, as convexity is preserved.

The rest of the paper is organized as follows. In Section 2, we consider the case of multidimensional diffusions with a constant diffusion coefficient and a Lipschitz continuous (suitably smooth) drift. The general case is discussed in Section 3. Our development uses localization ideas which are introduced in Section 2, as well as some basic estimates of the transition density of the underlying diffusion (e.g. Lipschitz continuity), which are developed in Appendix A. As mentioned before, we discuss the bottlenecks in the expected running time of the algorithm in Section 4.

2. Exact simulation of SDEs with identity diffusion coefficient

In the case that Lamperti's transformation is applicable, the SDE of interest is reducible to another SDE whose diffusion matrix is the identity. As a result, it suffices to consider simulating the following SDE:

$$dX(t) = \mu(X(t))dt + dW(t), \quad X(0) = x_0, \quad (2)$$

where $W = \{W(t) = (W_1(t), \dots, W_d(t)) : 0 \leq t < \infty\}$ is a d -dimensional Brownian motion. In this section we concentrate on the identity diffusion case (2), but the development can be immediately extended to the case of a constant diffusion matrix. However, throughout this section we must impose the following assumptions.

Assumption 1. *The SDE (2) has a strong solution.*

Assumption 2. *The drift coefficient $\mu(\cdot)$ is three times continuously differentiable.*

Assumption 2 is the requirement of TES, the theoretical foundation of our algorithm, which we shall introduce later.

Let us introduce some notation first. For any set G and $x \in \mathbb{R}^d$, we use $d(x, G) = \inf \{ \|x - y\|_2 : y \in G \}$ to denote the distance between x and G ; $\overset{\circ}{G}$ denotes the interior of G ; ∂G denotes the boundary of G ; G^c denotes the complement of G .

Consider a probability space $(\Omega, \mathcal{F}, \tilde{\mathbb{P}})$ endowed with a filtration $\{\mathcal{F}_t : 0 \leq t \leq 1\}$, and supporting a d -dimensional Brownian motion

$$X(t) = (X_1(t), \dots, X_d(t)), \quad 0 \leq t \leq 1.$$

Let $\{L(t) : 0 \leq t \leq 1\}$ be a $\tilde{\mathbb{P}}$ -local martingale defined as

$$L(t) = \exp \left(\int_0^t \mu^T(X(s)) dX(s) - \frac{1}{2} \int_0^t \|\mu(X(s))\|_2^2 ds \right), \tag{3}$$

where $\mu^T(\cdot)$ denotes the transpose of the column vector $\mu(\cdot)$. Under Assumption 1, $L(\cdot)$ is a $\tilde{\mathbb{P}}$ -martingale; see Corollary 3.5.16 of [15].

In this case we can define a probability measure \mathbb{P} through

$$\mathbb{P}(A) = \mathbb{E}^{\tilde{\mathbb{P}}} [I(A)L(1)] \quad \forall A \in \mathcal{F},$$

where $I(A)$ denotes the indicator function of the set A and $\mathbb{E}^{\tilde{\mathbb{P}}}(\cdot)$ is the expectation operator under $\tilde{\mathbb{P}}$.

Let

$$W(t) = (W_1(t), \dots, W_d(t)), \quad 0 \leq t \leq 1,$$

be a d -dimensional process defined by

$$W(t) = X(t) - \int_0^t \mu(X(s)) ds, \quad 0 \leq t \leq 1. \tag{4}$$

The following theorem provides the distribution of $W(\cdot)$.

Theorem 1 (Girsanov’s theorem). *If Assumption 1 is satisfied, then the process $W(\cdot)$ is a d -dimensional Brownian motion on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$.*

Proof. See, for instance, Theorem 3.5.1 of [15]. □

It is readily apparent from (4) that $X(\cdot)$ is a weak solution to the SDE (2) on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The exact simulation problem becomes sampling $X(1)$ under the measure \mathbb{P} . Since $X(1)$ follows a normal distribution under the measure $\tilde{\mathbb{P}}$, we can attempt to use acceptance-rejection to sample $X(1)$. A direct application of acceptance-rejection may proceed by using the $\tilde{\mathbb{P}}$ distribution of $X(1)$ (which is simply a normal distribution) as the proposal, which, if acceptance-rejection is applicable, would then be accepted with probability proportional to $L(1)$. However, there are two obstacles when trying to apply such a direct acceptance-rejection approach. First, the presence of the general stochastic integral appearing in the definition of $L(1)$ makes the likelihood ratio difficult to compute directly. Second, a direct application of acceptance-rejection requires the likelihood ratio, $L(1)$, to be bounded, which is unfortunately violated.

In order to deal with the first obstacle, we note that it is really not necessary to accurately evaluate the likelihood ratio. In the standard procedure of acceptance-rejection, the likelihood ratio is only used for comparison with an independent uniform random variable. Thus, to address the first obstacle, we can approximate the likelihood ratio with a deterministic error

bound, and keep refining until we can decide to either accept or reject the proposal. It turns out, as we shall see in Corollary 1, that the same approximation technique can actually be used to localize $L(1)$ and also resolve the second obstacle. Then, we will sample the distribution of $X(1)$ conditional on the localization of $L(1)$ using acceptance-rejection, where the rejection scheme is suggested by Lemma 1.

The theoretical foundation for such an approximation and refinement strategy is given by Tolerance-Enforced Simulation (TES), which is presented in Theorem 2.

Theorem 2 (Tolerance-Enforced Simulation). *Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and the following SDE:*

$$dY(t) = \alpha(Y(t))dt + v(Y(t))dW(t), \quad Y(t) = y_0, \tag{5}$$

where $\alpha(\cdot) = (\alpha_i(\cdot))_d : \mathbb{R}^d \rightarrow \mathbb{R}^d$, $v(\cdot) = (v_{ij}(\cdot))_{d \times d'} : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d'}$, and $W(\cdot)$ is a d' -dimensional Brownian motion. Suppose that $\alpha(\cdot)$ is continuously differentiable and that $v(\cdot)$ is three times continuously differentiable. Then, given any deterministic $\varepsilon > 0$, there is an explicit Monte Carlo procedure that allows us to simulate a piecewise constant process $Y_\varepsilon(\cdot)$ such that

$$\sup_{t \in [0,1]} \|Y_\varepsilon(t) - Y(t)\|_2 \leq \varepsilon$$

with probability one. Furthermore, for any $m > 1$ and $0 < \varepsilon_m < \dots < \varepsilon_1 < 1$, we can simulate Y_{ε_m} conditional on $Y_{\varepsilon_1}, \dots, Y_{\varepsilon_{m-1}}$.

Proof. See Theorem 2.1, Theorem 2.2, and Section 2.1 of [6], where a detailed TES procedure is also provided. □

Remark 1. TES is based on the Lévy–Ciesielski construction of the driving Brownian motion $W(\cdot)$ up to a random level. Consequently, $W(1)$ is obtained for free when we run TES in which a skeleton of the driving Brownian motion $W(\cdot)$ is simulated. In particular, for any $m > 1$ and $0 < \varepsilon_m < \dots < \varepsilon_1 < 1$, we can simulate Y_{ε_m} conditional on $Y_{\varepsilon_1}, \dots, Y_{\varepsilon_{m-1}}$ and $W(1)$.

As a straightforward consequence of Theorem 2, we develop a localization procedure for SDE in Corollary 1. Before moving forward to state the result, we define some notation that will be used therein.

Definition 1. A family of (Borel measurable) sets $\mathcal{G} = \{G_i \subset \mathbb{R}^d : i \in \mathbb{N}\}$ is said to be a countable continuous partition for a d -dimensional random vector Y if and only if the following hold:

1. The sets in \mathcal{G} are mutually disjoint; that is, $G_i \cap G_j = \emptyset$ for $i \neq j$.
2. Y is concentrated on \mathcal{G} ; that is, $\mathbb{P}(Y \in \cup_{i \in \mathbb{N}} G_i) = 1$.
3. $\mathbb{P}(Y \in \partial G_i) = 0$ for all $i \in \mathbb{N}$.

In addition, a function $\Xi_{\mathcal{G}}(x) : \text{supp}(Y) \rightarrow \mathbb{N}$ is defined such that $\Xi_{\mathcal{G}}(x) = i$ if and only if $x \in G_i$.

Corollary 1. *In the setting of Theorem 2, let $\mathcal{G} = \{G_i : i \in \mathbb{N}\}$ be a countable continuous partition for $Y(1)$; then there is an algorithm for simulating $\Xi_{\mathcal{G}}(Y(1))$ that terminates in finite time with probability one. In particular, for any set G such that $\mathbb{P}(Y(1) \in \partial G) = 0$, there is an algorithm for simulating $I(Y(1) \in G)$.*

Algorithm 1: Localization of SDE over countable continuous partition \mathcal{G} .

- 1: **Initialize** $\varepsilon \leftarrow 1/2$.
 - 2: Apply TES to simulate $Y_\varepsilon(1)$, $i \leftarrow \Xi_{\mathcal{G}}(Y_\varepsilon(1))$.
 - 3: **while** $d(Y_\varepsilon(1), G_i^c) \leq \varepsilon$ **do**
 - 4: Apply TES to simulate $Y_{\varepsilon/2}(1)$ conditional on $Y_{1/2}(1), \dots, Y_\varepsilon(1)$.
 - 5: $i \leftarrow \Xi_{\mathcal{G}}(Y_{\varepsilon/2}(1))$.
 - 6: $\varepsilon \leftarrow \varepsilon/2$.
 - 7: **Output** i .
-

Proof. Notice that $\mathbb{P}(Y(1) \in \partial G_i) = 0$, so $Y(1) \in \bigcup_{i \in \mathbb{N}} \overset{\circ}{G}_i$ holds almost surely. Recall from Theorem 2 that $\|Y_\varepsilon(1) - Y(1)\|_2 \leq \varepsilon$ almost surely, which suggests that

$$\mathbb{P}\left(\{\omega \in \Omega : Y(1) \in \overset{\circ}{G}_i\}\right) = \mathbb{P}\left(\bigcup_{\varepsilon > 0} \{\omega \in \Omega : d(Y_\varepsilon(1), G_i^c) > \varepsilon\}\right).$$

Thus, we pick $\varepsilon \in (0, 1)$ and apply TES to simulate the approximation process $Y_\varepsilon(1)$. If

$$d\left(Y_\varepsilon(1), G_{\Xi_{\mathcal{G}}(Y_\varepsilon(1))}^c\right) > \varepsilon,$$

then $\Xi_{\mathcal{G}}(Y(1)) = \Xi_{\mathcal{G}}(Y_\varepsilon(1))$, which terminates the algorithm. Otherwise we keep refining the approximation of TES, by setting $\varepsilon \leftarrow \varepsilon/2$, until $d(Y_\varepsilon(1), G_{\Xi_{\mathcal{G}}(Y_\varepsilon(1))}^c) > \varepsilon$. The algorithm will ultimately terminate since

$$\mathbb{P}\left(\bigcup_{i \in \mathbb{N}} \bigcup_{\varepsilon > 0} \{\omega \in \Omega : d(Y_\varepsilon(1), G_i^c) > \varepsilon\}\right) = 1.$$

The procedure for simulation of $I(Y(1) \in G)$ is just a particular case, obtained by setting $\mathcal{G} = \{G, G^c\}$. The details of the algorithm are given in Algorithm 1. □

The algorithm for simulating $X(1)$ is performed in a two-stage fashion. At the first stage, the likelihood ratio $L(1)$ is localized with the help of Corollary 1. (The efficiency of the algorithm may be slightly improved if we localize $X(1)$ and $L(1)$ simultaneously at the first step, then apply acceptance-rejection based on localization. However, this does not solve the problem of the infinite expected running time.) Then, at the second stage, $X(1)$ is sampled conditional on the result of localization.

We now illustrate how to localize $L(1)$ in detail. In order to write the dynamics of $Y(1)$ in standard form as in (5), we consider the SDE of $(L(\cdot), X(\cdot))$ under the measure \mathbb{P} as follows:

$$\begin{cases} dL(t) = L(t)\|\mu(X(t))\|_2^2 dt + L(t)\mu^T(X(t))dW(t), \\ dX(t) = \mu(X(t))dt + dW(t). \end{cases} \tag{6}$$

Let $\mathcal{G} = \{G_i = [i, i + 1) \times \mathbb{R}^d : i \in \mathbb{N}\}$ in the rest of this section. As (3) guarantees that $L(1)$ is nonnegative, it follows immediately that \mathcal{G} is a countable continuous partition for $L(1)$. Therefore, Algorithm 1 is directly applicable to sample $\Xi_{\mathcal{G}}((L(1), X(1)))$ using the SDE (6).

Without loss of generality, we assume $\Xi_{\mathcal{G}}((L(1), X(1))) = i$ in the rest of this section. It remains to sample $X(1)$ conditional on $\Xi_{\mathcal{G}}((L(1), X(1))) = i$ under the probability measure \mathbb{P} .

The following lemma provides an alternative expression of the conditional distribution of $X(1)$, which facilitates the simulation of $X(1)$ conditional on localization.

Lemma 1. *Let $U \sim \text{Unif}(0, 1)$ independent of everything else under the probability measure $\tilde{\mathbb{P}}$; then we have*

$$\mathbb{P}(X(1) \in dx \mid \Xi_{\mathcal{G}}((L(1), X(1))) = i) = \tilde{\mathbb{P}}(X(1) \in dx \mid \max(i, (i + 1)U) < L(1) < i + 1).$$

Proof. From the definition of conditional probability,

$$\mathbb{P}(X(1) \in dx \mid \Xi_{\mathcal{G}}((L(1), X(1))) = i) = \frac{\mathbb{P}(X(1) \in dx; \Xi_{\mathcal{G}}((L(1), X(1))) = i)}{\mathbb{P}(\Xi_{\mathcal{G}}((L(1), X(1))) = i)}.$$

Recall that $d\tilde{\mathbb{P}} = L(1)d\mathbb{P}$, so we have

$$\mathbb{P}(X(1) \in dx \mid \Xi_{\mathcal{G}}((L(1), X(1))) = i) = \frac{\mathbb{E}^{\tilde{\mathbb{P}}}[L(1)I(X(1) \in dx; \Xi_{\mathcal{G}}((L(1), X(1))) = i)]}{\mathbb{P}(\Xi_{\mathcal{G}}((L(1), X(1))) = i)}.$$

Since on $\Xi_{\mathcal{G}}((L(1), X(1))) = i$,

$$i \leq L(1) \leq i + 1,$$

we can rewrite the expectation into a probability by introducing $U \sim \text{Unif}(0, 1)$, as follows:

$$\begin{aligned} &\mathbb{E}^{\tilde{\mathbb{P}}}[L(1)I(X(1) \in dx; \Xi_{\mathcal{G}}((L(1), X(1))) = i)] \\ &= (i + 1)\tilde{\mathbb{P}}(X(1) \in dx; \Xi_{\mathcal{G}}((L(1), X(1))) = i; (i + 1)U < L(1)) \\ &= (i + 1)\tilde{\mathbb{P}}(X(1) \in dx; \max(i, (i + 1)U) < L(1) < i + 1). \end{aligned}$$

By substitution, it follows easily that

$$\begin{aligned} \mathbb{P}(X(1) \in dx \mid \Xi_{\mathcal{G}}((L(1), X(1))) = i) &= \frac{(i + 1)\tilde{\mathbb{P}}(\max(i, (i + 1)U) < L(1) < i + 1)}{\mathbb{P}(\Xi_{\mathcal{G}}((L(1), X(1))) = i)} \\ &\quad \times \tilde{\mathbb{P}}(X(1) \in dx \mid \max(i, (i + 1)U) < L(1) < i + 1). \end{aligned}$$

It remains to prove that

$$(i + 1)\tilde{\mathbb{P}}(\max(i, (i + 1)U) < L(1) < i + 1) = \mathbb{P}(\Xi_{\mathcal{G}}((L(1), X(1))) = i).$$

By a similar argument we can deduce that

$$\begin{aligned} &\mathbb{P}(\Xi_{\mathcal{G}}((L(1), X(1))) = i) \\ &= \mathbb{E}^{\tilde{\mathbb{P}}}[L(1)I(\Xi_{\mathcal{G}}((L(1), X(1))) = i)] \\ &= (i + 1)\tilde{\mathbb{P}}(\Xi_{\mathcal{G}}((L(1), X(1))) = i; (i + 1)U < L(1)) \\ &= (i + 1)\tilde{\mathbb{P}}(\max(i, (i + 1)U) < L(1) < i + 1), \end{aligned}$$

which ends the proof. □

Algorithm 2: Exact simulation for SDE with constant diffusion coefficient.

- 1: Apply Algorithm 1 to simulate random variable $\Xi_{\mathcal{G}}(L(1), W(1))$ associated with SDE (6); record the result as $i \leftarrow \Xi_{\mathcal{G}}(L(1), W(1))$.
- 2: **repeat**
- 3: Draw a sample u from $\text{Unif}(0, i + 1)$.
- 4: Apply Algorithm 1 to sample $I(\max(i, u) < L(1) < i + 1)$ using SDE (7). The end of $\tilde{\mathbb{P}}$ -Brownian path $x \leftarrow X(1)$ is also sampled as a by-product of TES.
- 5: **until** $I(\max(i, u) < L(1) < i + 1) = 1$.
- 6: **Output** x as a sample of $X(1)$.

As a direct implication of Lemma 1, in order to obtain an example sample for $X(1)$ under \mathbb{P} , given $\Xi_{\mathcal{G}}(L(1), X(1)) = i$, we can simply simulate $X(1)$ conditional on $\max(i, (i + 1)U) < L(1) < i + 1$ under the probability measure $\tilde{\mathbb{P}}$. In order to do this sampling under $\tilde{\mathbb{P}}$, we can sample U first and denote the value by u . Then, observing that $X(\cdot)$ is the driving Brownian motion under the probability measure \mathbb{P} , Algorithm 1 is applied to the SDE

$$dL(t) = L(t)\mu^T(X(t))dX(t) \tag{7}$$

to simulate the indicator function $I(\max(i, u) < L(1) < i + 1)$. In addition, according to Remark 1, when TES is employed in Algorithm 1, a sample of $X(1)$ is also produced simultaneously. Thereafter, the value of $X(1)$ is accepted if and only if $I(\max(i, u) < L(1) < i + 1) = 1$; otherwise we repeat the procedure in this paragraph, but we fix the parameter i , because $\Xi_{\mathcal{G}}(L(1), X(1)) = i$ has already been sampled under the correct distribution \mathbb{P} . The output of the algorithm, once the value $X(1)$ is finally accepted, follows the distribution of $X(1)$ under \mathbb{P} without any bias.

We summarize the discussion in this section in the following theorem.

Theorem 3. *If Assumptions 1 and 2 are satisfied, then there is an exact simulation algorithm for $X(1)$ that terminates with probability one; see Algorithm 2.*

3. Exact simulation for general SDEs

In this section, we will develop an exact simulation algorithm for the SDE (1). We shall fix $X(0) = x_0$, and the dependence of x_0 on some objects (such as the transition density of $X(1)$) will be omitted.

We are still going to construct an exact simulation algorithm based on acceptance-rejection in this section. However, for SDEs with non-constant diffusion matrix, applying Girsanov’s theorem no longer provides a Brownian-type proposal distribution for acceptance-rejection, so we will construct an acceptance-rejection algorithm based on the density of $X(1)$.

Throughout the rest of this section, we shall assume the following assumptions and conditions.

Assumption 3. *The drift coefficient $\mu(\cdot)$ is continuously differentiable, and the diffusion coefficient $\sigma(\cdot)$ is three times continuously differentiable. Moreover, a strong solution to the SDE (1) exists.*

Condition 1. The probability distribution of $X(1)$ is absolutely continuous with respect to the Lebesgue measure. In other words, $X(1)$ has a density function, denoted by $p(\cdot)$, with respect to the Lebesgue measure.

Condition 2. For any relatively compact set S , the density $p(\cdot)$ is Lipschitz continuous with Lipschitz constant C_S ; i.e.

$$|p(x) - p(y)| \leq C_S|x - y| \quad \forall x, y \in S.$$

Condition 3. For any relatively compact set S , there exist $\delta_S > 0$ such that

$$p(x) \geq \delta_S \quad \forall x \in S.$$

As we have seen in the previous section, Assumption 3 is the necessary condition for the applicability of the TES result introduced in Theorem 2, which enables us to strongly approximate $X(1)$. Condition 1 will eventually be used to apply the acceptance-rejection technique using an absolutely continuous (with respect to the Lebesgue measure) proposal distribution. Conditions 2 and 3, as we shall see, will allow us to control the bound of the likelihood ratio when applying acceptance-rejection.

It is important to ensure that the constants C_S and δ_S are explicitly computable in terms of $\mu(\cdot)$ and $\sigma(\cdot)$ only, but we should also emphasize that we are not assuming that the density $p(\cdot)$ is known.

There are many ways in which the computability of C_S and δ_S can be enforced. For instance, in Appendix A we discuss a set of assumptions involving classical estimators of the fundamental solutions of parabolic equations, which we review in order to compute C_S and δ_S explicitly.

The standard use of the acceptance-rejection algorithm requires knowing the density $p(x)$, which seems hopeless for the general SDE problem that we study. An alternative approach is constructing a nonnegative, bounded, and unbiased estimator of $p(x)$. While the density $p(x)$ is unknown, an unbiased estimator of $p(x)$, denoted by $\Lambda_N(x)$ in Section 3.1, can be constructed by means of a local approximation of the density. However, the unbiased estimator $\Lambda_N(x)$ may be negative, so it cannot be directly used in acceptance-rejection. To remedy this problem, in Lemma 3 we construct a nonnegative and unbiased estimator $\Lambda_N^+(x)$ of $p(x)$ using a random walk and a suitable Bernoulli factory. However, the estimator $\Lambda_N^+(x)$ is unbounded, so we propose to sample enough information about the SDEs (the ancillary variable N') so that the estimator $\Lambda_N^+(x)$ conditional on the sampled information is locally bounded. Consequently, conditional on the localization of $X(1)$ and the additional information $N = N'$, the estimator $\Lambda_N^+(x)$ is bounded and nonnegative.

We now state the outline of our exact simulation algorithm. First of all, we apply a localization technique on the countable continuous partition \mathcal{G}_{loc} defined as

$$\mathcal{G}_{\text{loc}} = \{[i_1, i_1 + 1) \times \cdots \times [i_d, i_d + 1) : (i_1, \dots, i_d) \in \mathbb{Z}^d\}.$$

Since \mathcal{G}_{loc} has countably many components, we can enumerate \mathcal{G}_{loc} and rewrite it in the form $\mathcal{G}_{\text{loc}} = \{G_i : i \in \mathbb{N}\}$, where each G_i is a unit hypercube. Obviously, Algorithm 1 is applicable to $X(1)$ with respect to the countable continuous partition \mathcal{G}_{loc} .

Then, we introduce an ancillary random variable N' coupled with $X(1)$, and simulate $(N'|X(1) \in G_i)$. As we shall see, the random variable N' will play an important role after we introduce a suitable family of random variables whose expectations converge to the density

Algorithm 3: Exact simulation for multivariate SDE.

- 1: Simulate $\Xi_{\mathcal{G}_{\text{loc}}}(X(1))$ applying Algorithm 1. Set $i \leftarrow ts\Xi_{\mathcal{G}_{\text{loc}}}(X(1))$.
 - 2: Simulate $(N'|X(1) \in G_i)$; denote the result by n' .
 - 3: Simulate $(X(1)|N' = n', X(1) \in G_i)$; denote the result by x .
 - 4: **Output** x .
-

of $X(1)$ at a given point. In the end, we will be able to sample $X(1)$ conditional on N' and $X(1) \in G_i$, using the estimator $\Lambda_N^+(x)$, which is bounded and nonnegative for $x \in G_i$ and $N = N'$.

The following theorem provides the main contribution of this paper.

Theorem 4. *If Assumption 3 and Conditions 1–3 are satisfied, then there is an algorithm for exactly simulating $X(1)$ which terminates in finite time with probability one; see Algorithm 3.*

The rest of this section is organized as follows. Section 3.1 applies a technique borrowed from multilevel Monte Carlo to construct the unbiased density estimator and the ancillary random variable N' . Section 3.2 explains how to sample N' using acceptance-rejection and a suitable Bernoulli factory [19, 17, 14] conditional on localization. Section 3.3 demonstrates how to sample $X(1)$ conditional on N' , once again using a suitable localization.

3.1. A multilevel representation of the density

In this section, we borrow an idea from multilevel Monte Carlo [11] to construct an unbiased estimator for $p(\cdot)$, and we also introduce the ancillary random variable N' .

In order to illustrate our idea, first we need to introduce some notation. For any x in G_i , we define $\{B_{r_n}(x) : n \geq 1\}$ as a sequence of open balls centered at x , whose radii $\{r_n : n \geq 1\}$ form a decreasing sequence with $r_n \rightarrow 0$ as $n \rightarrow \infty$.

Let $V(r)$ denote the volume of a d -dimensional ball with radius r (i.e. the volume of $B_r(0)$). We define $\overline{p}_n(x)$ to be the average density over the ball $B_{r_n}(x)$, i.e.

$$\overline{p}_n(x) = [V(r_n)]^{-1} \int_{B_{r_n}(x)} p(x) dx.$$

Let $\hat{p}_n(x)$ denote a nonnegative unbiased estimator for $\overline{p}_n(x)$, i.e.

$$\hat{p}_n(x) = [V(r_n)]^{-1} \times I(X(1) \in B_{r_n}(x))$$

for $n \geq 1$, where $\hat{p}_n(x)$ is defined using the same realization $X(1)$ for all n and x . We define $\hat{p}_0(x) := 0$ and $\overline{p}_0 := 0$ for notational simplicity. It follows immediately that $\mathbb{E}[\hat{p}_n(x)] = \overline{p}_n(x)$ for $n \geq 0$.

The density $p(x)$ is first decomposed into an infinite telescoping sum,

$$p(x) = \sum_{n=0}^{\infty} (\overline{p}_{n+1}(x) - \overline{p}_n(x)).$$

Then we introduce a randomization technique inspired by randomized multilevel Monte Carlo (see [18, 22]). The density $p(x)$ can be decomposed as the expectation of an infinite sum of estimators, which is truncated to a finite but random level so that the expectation is invariant. The idea is to introduce an integer-valued random variable N , which is independent of

everything else. Then $p(x)$ can be expressed as

$$\begin{aligned}
 p(x) &= \sum_{n=0}^{\infty} (\overline{p}_{n+1}(x) - \overline{p}_n(x)) \\
 &= \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \frac{(\overline{p}_{n+1}(x) - \overline{p}_n(x))}{\mathbb{P}(N \geq n)} \mathbb{P}(N = k) I(n \leq k) \\
 &= \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} \frac{(\overline{p}_{n+1}(x) - \overline{p}_n(x))}{\mathbb{P}(N \geq n)} \mathbb{P}(N = k) I(n \leq k) \\
 &= \sum_{k=0}^{\infty} \sum_{n=0}^k \frac{(\overline{p}_{n+1}(x) - \overline{p}_n(x))}{\mathbb{P}(N \geq n)} \mathbb{P}(N = k) \\
 &= \mathbb{E} \left[\sum_{n=0}^N \frac{(\overline{p}_{n+1}(x) - \overline{p}_n(x))}{\mathbb{P}(N \geq n)} \right],
 \end{aligned}$$

where the third equality follows from Fubini’s theorem, which can be justified if

$$\begin{aligned}
 \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \frac{|\overline{p}_{n+1}(x) - \overline{p}_n(x)|}{\mathbb{P}(N \geq n)} \mathbb{P}(N = k) I(n \leq k) &= \sum_{n=0}^{\infty} |\overline{p}_{n+1}(x) - \overline{p}_n(x)| \\
 &\leq 2C_{G_i, r_1} \sum_{n=0}^{\infty} r_n \\
 &< \infty.
 \end{aligned}$$

We will show $\sum_{n=0}^{\infty} r_n < \infty$ in the sequel. Moreover, by the tower property we have

$$\begin{aligned}
 \mathbb{E} \left[\sum_{n=0}^N \frac{(\overline{p}_{n+1}(x) - \overline{p}_n(x))}{\mathbb{P}(N \geq n)} \right] &= \mathbb{E} \left[\sum_{n=0}^N \frac{\mathbb{E}[\hat{p}_{n+1}(x) - \hat{p}_n(x) | N]}{\mathbb{P}(N \geq n)} \right] \\
 &= \mathbb{E} \left[\mathbb{E} \left[\sum_{n=0}^N \frac{\hat{p}_{n+1}(x) - \hat{p}_n(x)}{\mathbb{P}(N \geq n)} \middle| N \right] \right] \\
 &= \mathbb{E} \left[\sum_{n=0}^N \frac{\hat{p}_{n+1}(x) - \hat{p}_n(x)}{\mathbb{P}(N \geq n)} \right].
 \end{aligned}$$

Therefore, if we define

$$\Lambda_n(x) = \sum_{k=0}^n \frac{\hat{p}_{k+1}(x) - \hat{p}_k(x)}{\mathbb{P}(N \geq k)} \quad \text{for } n \geq 0,$$

it follows easily that

$$p(x) = \mathbb{E} [\Lambda_N(x)]. \tag{8}$$

We now are interested in obtaining bounds for $\Lambda_n(x)$ and its expectation for $x \in G_i$. To this end we first define G_{i,r_1} as an r_1 -neighborhood of the set G_i , which consists of all points at a distance less than r_1 from G_i ; i.e.

$$G_{i,r_1} = \bigcup_{x \in G_i} B_{r_1}(x).$$

It is not hard to observe that G_{i,r_1} is a relatively compact set, to which Conditions 1–3 are applicable. In the following lemma, we will demonstrate that under these conditions, one can judiciously pick the distribution of N and the radii $\{r_n : n \geq 1\}$ in order to establish explicit bounds for $\Lambda_n(x)$ and $\mathbb{E}[\Lambda_n(x)]$, respectively.

Lemma 2. *Suppose that $x \in G_i$ and Conditions 1–3 are satisfied. If we pick*

$$r_n = (3\delta_{G_{i,r_1}})/(2\pi^2 n^3 C_{G_{i,r_1}}) \quad \text{and} \quad \mathbb{P}(N = n) = 1/[n(n + 1)]$$

for $n \geq 1$, then we have

$$\delta_{G_{i,r_1}}/2 \leq \mathbb{E}[\Lambda_n(x)] \leq [V(r_1)]^{-1} + \delta_{G_{i,r_1}}/2 \tag{9}$$

and

$$|\Lambda_n(x)| \leq [V(r_1)]^{-1} + \sum_{k=1}^n \left(k[V(r_{k+1})]^{-1} + k[V(r_k)]^{-1} \right) =: m_n. \tag{10}$$

Proof. Let us construct the lower bound of $\mathbb{E}[\Lambda_n(x)]$ first. By the triangle inequality,

$$\mathbb{E}[\Lambda_n(x)] \geq \mathbb{E}[\Lambda_0(x)] - \sum_{k=1}^n \mathbb{E} |\Lambda_k(x) - \Lambda_{k-1}(x)|.$$

On the one hand, from the definition of $\Lambda_0(x)$ and Condition 3, we can conclude that

$$\mathbb{E}[\Lambda_0(x)] = \mathbb{E}[\hat{p}_1(x)] = \overline{p}_1(x) \geq \delta_{G_{i,r_1}}.$$

On the other hand, using the triangle inequality,

$$\begin{aligned} \mathbb{E} |\Lambda_k(x) - \Lambda_{k-1}(x)| &= \mathbb{E} \left| \frac{\hat{p}_{k+1}(x) - \hat{p}_k(x)}{\mathbb{P}(N \geq k)} \right| \\ &\leq \left| \frac{\mathbb{E}\hat{p}_{k+1}(x) - \mathbb{E}\hat{p}_k(x)}{\mathbb{P}(N \geq k)} \right| \\ &= (\mathbb{P}(N \geq k))^{-1} |\overline{p}_{k+1}(x) - \overline{p}_k(x)|. \end{aligned}$$

Then, from Condition 2 we have

$$|p(x) - p(y)| \leq C_{G_{i,r_1}} |x - y| \quad \text{for } x, y \in C_{G_{i,r_1}}. \tag{11}$$

Recall that $\overline{p}_k(x)$ is the average density over the ball $B_{r_k}(x)$ and $B_{r_k}(x) \subseteq B_{r_1}(x) \subseteq G_{i,r_1}$ for $x \in G_i$. It then follows from (11) that

$$|\overline{p}_{k+1}(x) - \overline{p}_k(x)| \leq C_{G_{i,r_1}} \text{diam}(B_{r_1}(x)) = 2C_{G_{i,r_1}} r_1.$$

Consequently,

$$\sum_{k=1}^n \mathbb{E} |\Lambda_k(x) - \Lambda_{k-1}(x)| \leq \sum_{k=1}^n 2C_{G_i, r_1} (\mathbb{P}(N \geq k))^{-1} r_k \leq \delta_{G_i, r_1} / 2.$$

Combining the above inequality with $\mathbb{E}[\Lambda_0(x)] \geq \delta_{G_i, r_1}$ yields

$$\mathbb{E}[\Lambda_n(x)] \geq \delta_{G_i, r_1} / 2.$$

Similarly, observing that

$$\mathbb{E}[\Lambda_0(x)] = \mathbb{E}[\hat{p}_1(x)] = [V(r_1)]^{-1} \times \mathbb{P}(X(1) \in B_{r_1}(x)) \leq [V(r_1)]^{-1},$$

for the upper bound we have

$$\mathbb{E}[\Lambda_n(x)] \leq \mathbb{E}[\Lambda_0(x)] + \sum_{k=1}^n \mathbb{E} |\Lambda_k(x) - \Lambda_{k-1}(x)| \leq [V(r_1)]^{-1} + \delta_{G_i, r_1} / 2.$$

We can also derive an upper bound of $|\Lambda_n(x)|$:

$$\begin{aligned} |\Lambda_n(x)| &\leq \sum_{k=0}^n (\mathbb{P}(N \geq k))^{-1} |\hat{p}_{k+1}(x) - \hat{p}_k(x)| \\ &\leq [V(r_1)]^{-1} + \sum_{k=1}^n \left(k[V(r_{k+1})]^{-1} + k[V(r_k)]^{-1} \right) =: m_n < \infty, \end{aligned}$$

which ends the proof. □

In the rest of this paper, we will adopt the value of r_n and the distribution of N in Lemma 2.

Even though we have constructed an unbiased estimator $\Lambda_N(x)$ for $p(x)$, acceptance-rejection is not applicable because $\Lambda_N(x)$ may be negative and unbounded. In order to apply acceptance-rejection, we need a nonnegative unbiased estimator for $p(x)$, which will be constructed in Lemma 3. The idea of the construction is borrowed from [8]. Let $\{\Lambda_{n,k}(x) : k \geq 1\}$ be independent and identically distributed copies of $\Lambda_n(x)$. We then define

$$S_{n,k}(x) := \Lambda_{n,1}(x) + \dots + \Lambda_{n,k}(x)$$

and

$$\tau_n(x) := \inf\{k \geq 1 : S_{n,k}(x) \geq 0\}.$$

By Wald's first equation,

$$\mathbb{E}[\Lambda_n(x)] = \mathbb{E} [S_{n, \tau_n(x)}(x)] / \mathbb{E}[\tau_n(x)]. \tag{12}$$

Note that $S_{n, \tau_n(x)}(x) \geq 0$, but now we have an additional contribution $1/\mathbb{E}[\tau_n(x)]$, which can be interpreted as a probability. In order to sample a Bernoulli random variable with such a probability, we will need the following result, which we refer to as the Bernoulli factory.

Theorem 5 (Bernoulli factory [19, 14]). *Assume that $\epsilon \in (0, 1/2]$ and $\alpha > 0$ are two known constants and that we have an oracle that outputs independent and identically distributed Bernoullis with parameter $p \in (0, (1 - \epsilon) / \alpha]$. Then there is an algorithm which takes the*

output of the oracle and produces a Bernoulli random variable with parameter $f(p) = \min(\alpha p, 1 - \epsilon) = \alpha p$. Moreover, if $\bar{N}(\alpha, \epsilon)$ is the number of Bernoulli(p) random variables required to output Bernoulli($f(p)$), then $.004 \cdot \alpha/\epsilon \leq \mathbb{E}(\bar{N}(\alpha, \epsilon)) \leq 10 \cdot \alpha/\epsilon$.

We can now explain how to construct $\Lambda_n^+(x) \geq 0$ such that $\mathbb{E}[\Lambda_n^+(x)] = \mathbb{E}[\Lambda_n(x)]$.

Lemma 3. *There exists a family of random variables $\{\Lambda_n^+(x) : n \in \mathbb{N}, x \in G_i\}$ such that the following properties hold:*

1. $0 \leq \Lambda_n^+(x) \leq m_n$.
2. $\mathbb{E}[\Lambda_n^+(x)] = \mathbb{E}[\Lambda_n(x)]$.
3. Given n and x , there is an algorithm for simulating $\Lambda_n^+(x)$.

Proof. Let $\bar{\Gamma}_n(x)$ be a Bernoulli random variable with parameter $(\mathbb{E}[\tau_n(x)])^{-1}$, which is independent of everything else. It follows that

$$\mathbb{E}[\Lambda_n(x)] = \mathbb{E}[\bar{\Gamma}_n(x)S_{n,\tau_n(x)}(x)].$$

We write $\Lambda_n^+(x) := \bar{\Gamma}_n(x)S_{n,\tau_n(x)}(x)$. Property 1 follows from the facts that $0 \leq S_{n,\tau_n(x)}(x) \leq \Lambda_{n,\tau_n(x)}(x) \leq m_n$ and that $0 \leq \bar{\Gamma}_n(x) \leq 1$. Property 2 is justified directly by (12). To show that $\Lambda_n^+(x)$ can be simulated, we just need to provide an algorithm for simulating $\bar{\Gamma}_n(x)$.

Recall that $\mathbb{E}[\Lambda_n(x)] \geq \delta_{G_i,r_1}/2$, so we have

$$\mathbb{E}[\tau_n(x)] = \frac{\mathbb{E}[S_{n,\tau_n(x)}(x)]}{\mathbb{E}[\Lambda_n(x)]} \leq 2\delta_{G_i,r_1}^{-1} m_n.$$

Consider Wald’s second equation

$$\mathbb{E}\left[(S_{n,\tau_n(x)}(x) - \mathbb{E}(\Lambda_n(x))\tau_n(x))^2\right] = \text{Var}(\Lambda_n(x))\mathbb{E}[\tau_n(x)],$$

which implies

$$\begin{aligned} \mathbb{E}\left[(\tau_n(x))^2\right] &\leq \frac{2\mathbb{E}[\tau_n(x)S_{n,\tau_n(x)}(x)] \mathbb{E}[\Lambda_n(x)] + \text{Var}(\Lambda_n(x))\mathbb{E}[\tau_n(x)]}{\mathbb{E}[\Lambda_n(x)]^2} \\ &\leq \frac{(2m_n^2 + m_n^2)\mathbb{E}(\tau_n(x))}{(\delta_{G_i,r_1}/2)^2} \\ &\leq \frac{3m_n^3}{(\delta_{G_i,r_1}/2)^3} =: m_{\tau,n}. \end{aligned} \tag{13}$$

Now we shall proceed to simulate the random variable $\bar{\Gamma}_n(x)$. Consider a random variable $T_n(x)$ with distribution

$$\mathbb{P}(T_n(x) = k) = \mathbb{P}(\tau_n(x) \geq k) / \mathbb{E}[\tau_n(x)] \quad \text{for } k \geq 1.$$

Since $I(T_n(x) = 1)$ is the desired Bernoulli random variable with parameter $(\mathbb{E}[\tau_n(x)])^{-1}$, it then suffices to simulate $T_n(x)$. Towards this end, we apply acceptance-rejection again using another random variable T' as proposal, whose distribution is

$$\mathbb{P}(T' = k) = \frac{6}{\pi^2 k^2} \quad \text{for } k \geq 1.$$

Algorithm 4: Simulation of $\Lambda_n^+(x)$.

- 1: **repeat**
- 2: Sample random variable T' ; set $k \leftarrow T'$.
- 3: **repeat**
- 4: Sample an independent copy of $I(\tau_n(x) \geq k)$ as an input of Bernoulli factory associated with function

$$f(p) = \min \left(\frac{k^2}{2m_{\tau,n}} p, \frac{1}{2} \right)$$

- 5: **until** Bernoulli factory produces an output γ .
 - 6: **until** $\gamma = 1$.
 - 7: **if** $k=1$ **then**
 - 8: Sample $S_{n,\tau_n(x)}(x)$ and **output** the result.
 - 9: **else**
 - 10: **Output** 0.
-

Since $\tau_n(x)$ is nonnegative, Markov’s inequality asserts that

$$\mathbb{P}(\tau_n(x) \geq k) = \mathbb{P}\left((\tau_n(x))^2 \geq k^2\right) \leq k^{-2} \mathbb{E}[\tau_n(x)^2] \leq k^{-2} m_{\tau,n}, \tag{14}$$

where the last inequality follows from (13). Also, from the definition of $\tau_n(x)$ we know that $\mathbb{E}[\tau_n(x)] \geq 1$. Consequently, the likelihood ratio between $T_n(x)$ and T' is given by

$$\frac{\mathbb{P}(T_n(x) = k)}{\mathbb{P}(T' = k)} = \frac{\pi^2 k^2 \mathbb{P}(\tau_n(x) \geq k)}{6 \mathbb{E}[\tau_n(x)]} \leq \frac{\pi^2}{6} m_{\tau,n}.$$

From the above inequality we see that the likelihood ratio is bounded, so the acceptance-rejection procedure is applicable.

Conditional on the proposal $T' = k$, we then introduce a new Bernoulli random variable $\tilde{\Gamma}_{n,k}(x)$ to decide whether or not the proposal is accepted as $T_n(x)$. The distribution of $\tilde{\Gamma}_{n,k}(x)$ is defined as

$$\mathbb{P}(\tilde{\Gamma}_{n,k}(x) = 1) = 1 - \mathbb{P}(\tilde{\Gamma}_{n,k}(x) = 0) = \frac{k^2}{2m_{\tau,n}} \mathbb{P}(\tau_n(x) \geq k).$$

Hence it follows from (14) that

$$\mathbb{P}(\tilde{\Gamma}_{n,k}(x) = 1) \leq 1/2. \tag{15}$$

Observe that the indicator $I(\tau_n(x) \geq k)$ is simulable, but its distribution is not explicitly accessible, so it is natural to sample $\tilde{\Gamma}_{n,k}(x)$ via the Bernoulli factory introduced in Theorem 5. By (15), the function $f(\cdot)$ involved in the Bernoulli factory is a linear function, as follows:

$$f(p) = \min \left(\frac{k^2}{2m_{\tau,n}} p, \frac{1}{2} \right) = \frac{k^2}{2m_{\tau,n}} p.$$

We summarize the procedure for simulating $\Lambda_n^+(x)$ in Algorithm 4. □

We now introduce an ancillary random variable N' coupled with $X(1)$ in the following way:

$$\mathbb{P}(N' = n|X(1) \in dx) \propto \mathbb{P}(N = n) \times \mathbb{E}[\Lambda_n^+(x)]. \tag{16}$$

Assuming $(N'|X(1) \in G_i)$ can be simulated, $(X(1)|N', X(1) \in G_i)$ can easily be simulated by acceptance-rejection as well, thanks to the convenient density representation given by (18). The algorithm for sampling $(N'|X(1) \in G_i)$ will be explained in the next section.

3.2. Conditional sampling of N'

In this section we will focus on the procedure for simulating N' conditional on $X(1) \in G_i$. First we derive from (16) the probability mass function of $(N'|X(1) \in G_i)$, namely

$$\mathbb{P}(N' = n|X(1) \in G_i) = \frac{\mathbb{P}(N = n)}{\mathbb{P}(X(1) \in G_i)} \times \int_{G_1} \mathbb{E}[\Lambda_n^+(x)]dx.$$

From the upper bound of $\mathbb{E}[\Lambda_n]$ given by Lemma 2, we have the following inequality:

$$\mathbb{P}(N' = n|X(1) \in G_i) \leq \frac{\mathbb{P}(N = n)}{\mathbb{P}(X(1) \in G_i)} \times \left([V(r_1)]^{-1} + \delta_{G_{i,r_1}}/2 \right).$$

Simulation of $(N'|X(1) \in G_i)$ can be achieved by acceptance-rejection. Consider a Bernoulli random variable $\Gamma_n(x)$ defined as

$$\mathbb{P}(\Gamma_n(x) = 1) = 1 - \mathbb{P}(\Gamma_n(x) = 0) = \frac{1}{2} \left([V(r_1)]^{-1} + \delta_{G_{i,r_1}}/2 \right)^{-1} \mathbb{E}[\Lambda_n^+(x)] \leq \frac{1}{2}. \tag{17}$$

Then the outline of the acceptance-rejection algorithm for simulating N' would be as follows:

- Step 1.** Sample n from random variable N .
- Step 2.** Sample x from uniform distribution $U_{G_i} \sim \text{Unif}(G_i)$.
- Step 3.** Simulate $\Gamma_n(x)$. If $\Gamma_n(x) = 0$, go to Step 1. Otherwise accept n as a sample of N' .

The only difficult step in the above procedure is Step 3, namely, simulating $\Gamma_n(x)$, which we will discuss now.

Lemma 2 implies that $0 \leq \Lambda_n^+(x) \leq m_n$. Let $U \sim \text{Unif}(0, m_n)$, which is independent of everything else; then $I(U \leq \Lambda_n^+(x))$ is a Bernoulli random variable with parameter $(m_n)^{-1} \mathbb{E}[\Lambda_n^+(x)]$. By (17), the Bernoulli factory given in Theorem 5 can be applied to simulate $\Gamma_n(x)$, using $I(U \leq \Lambda_n^+(x))$ as input and using the function

$$f(p) = \min \left(\frac{m_n}{2} \left([V(r_1)]^{-1} + \delta_{G_{i,r_1}}/2 \right)^{-1} p, \frac{1}{2} \right) = \frac{m_n}{2} \left([V(r_1)]^{-1} + \delta_{G_{i,r_1}}/2 \right)^{-1} p.$$

To conclude, we synthesize the complete steps for simulating N' in the following algorithm.

3.3. Sampling of $(X(1)|N', X(1) \in G_i)$

In this section, we will focus on sampling $(X(1)|N', X(1) \in G_i)$.

Without loss of generality, let us assume $N' = n$ throughout the the rest of this section. According to (16) and Lemma 3, the conditional distribution of $X(1)$ can be written as

$$\mathbb{P}(X(1) \in dx|N' = n, X(1) \in G_i) = C_{n,G_i} \mathbb{E}[\Lambda_n^+(x)], \tag{18}$$

Algorithm 5: Simulation of $(N' | x \in G_i)$.

1: **repeat**
 2: Sample n from random variable N .
 3: Sample x from uniform distribution $U_{G_i} \sim \text{Unif}(G_i)$.
 4: **repeat**
 5: Sample u from $U \sim \text{Unif}(0, m_n)$.
 6: Sample λ from distribution of $\Lambda_n^+(x)$ using Algorithm 4.
 7: Use $I(u < \lambda)$ as an input of Bernoulli factory associated with function

$$f(p) = \min \left(\frac{m_n}{2} \left([V(r_1)]^{-1} + \delta_{G_i, r_1} / 2 \right)^{-1} p, \frac{1}{2} \right).$$

8: **until** Bernoulli factory produces an output γ .
 9: **until** $\gamma = 1$.
 10: **Output** n .

Algorithm 6: Simulation of $X(1)$ conditional on $N' = n, X(1) \in G_i$.

1: **repeat**
 2: Sample x from uniform distribution $U_{G_i} \sim \text{Unif}(G_i)$.
 3: Sample u from $U \sim \text{Unif}(0, m_n)$.
 4: Sample λ from distribution of $\Lambda_n^+(x)$ using Algorithm 4.
 5: **until** $u \leq \lambda$.
 6: **Output** x .

where C_{n, G_i} is a constant independent of x . Once again, as we shall see, $(X(1) | N' = n, X(1) \in G_i)$ can be simulated by acceptance-rejection. We use the uniform distribution U_{G_i} as the proposal distribution, and accept the proposal with probability $m_n^{-1} \times \mathbb{E}[\Lambda_n^+(x)]$, so we can accept if and only if $I(U \leq \Lambda_n^+(x)) = 1$, where $U \sim \text{Unif}(0, m_n)$ is independent of everything else. The output of the acceptance-rejection follows the desired distribution. The explicit procedure for simulating $(X(1) | N', X(1) \in G_i)$ is given in Algorithm 6.

4. Conclusion

The main contribution of this paper is the construction of the first generic exact simulation algorithm for multidimensional diffusions. The algorithm extensively uses several localization ideas and TES techniques. But it also combines ideas from multilevel Monte Carlo in order to construct a sequence of random variables which ultimately provides an unbiased estimator for the underlying transition density.

Although the overall construction can be implemented with a finite termination time almost surely, the expected running time is infinite. Thus, the contribution of the paper is of theoretical nature, showing that it is possible to perform exact sampling of multivariate diffusions without applying Lamperti's transformation. However, more research is needed to investigate whether the algorithm can be modified to be implemented in finite expected time, perhaps under additional assumptions. Alternatively, perhaps some controlled bias can be introduced while preserving features such as positivity and convexity in the applications discussed at the end of the introduction. To this end, we conclude with a discussion of the elements in the algorithm which are behind the infinite expected termination time.

There are three basic problems that cause the algorithm to have infinite expected termination time. Two of them can be appreciated already from the constant diffusion discussion and involve the use of the localization techniques that we have introduced. The third problem has to do with the application of the Bernoulli factory.

- **Problem 1:** The first problem arises when trying to sample a Bernoulli of the form $I(X(1) \in G)$. Given $\epsilon_n > 0$, sampling $X_{\epsilon_n}(1)$ such that $\|X_{\epsilon_n}(1) - X(1)\| \leq \epsilon_n$ takes an $O(\epsilon_n^{-(2+\delta)})$ computational cost for any $\delta > 0$. So, if G is a unit hypercube in d dimensions, using the density estimates for $X(1)$ we obtain

$$\mathbb{P}(d(X(1), \partial G_i) \leq \epsilon) \geq c_G \epsilon$$

for some $c_G > 0$. Therefore, if N_0 is the computational cost required to sample $I(X(1) \in G)$, we have that for some $\delta_0 > 0$,

$$\mathbb{P}\left(N_0 > \frac{1}{\epsilon^{2+\delta}}\right) \geq \mathbb{P}(d(X(1), \partial G_i) \leq \delta_0 \epsilon) \geq c_G \delta_0 \epsilon.$$

Therefore,

$$\mathbb{P}(N_0 > x) \geq c_G \delta_0 \frac{1}{x^{1/(2+\delta)}},$$

which yields that $\mathbb{E}(N_0) = \infty$.

- **Problem 2:** The second problem arises in the acceptance-rejection step applied in Lemma 1, which requires sampling $X(1)$ under $\tilde{\mathbb{P}}$ conditional on $\max(i, (i + 1)U) < L(1) < i + 1$. Directly sampling from this conditional law might be inefficient if i is large. However, this problem can be mitigated using rare-event simulation techniques, which might be available in the presence of additional structure on the drift, because under $\tilde{\mathbb{P}}(\cdot)$, $X(\cdot)$ follows a Brownian motion.
- **Problem 3:** This arises because, as indicated in Theorem 5, the computational complexity of the Bernoulli factory of a linear function of the form $f(p) = \min(\alpha p, 1 - \epsilon)$ scales with order $O(\alpha/\epsilon)$. In our case, we are able to select $\epsilon = 1/2$, and we invoke the Bernoulli factory in Algorithms 4 and 5. In Algorithm 4, $\alpha = O(k^2)$, given $T' = k$ and $\mathbb{E}(T') = \infty$. In Algorithm 5, $\alpha = O(m_n)$, given $N = n$. Although the bound which is used to define m_n in Lemma 3 is far from optimal, in its current form, $m_n = O(n^{3d+2})$, we have that $\mathbb{E}(N) = \infty$.

Appendix A. Transition density estimates

In the appendix, we will discuss some assumptions which are sufficient for the applicability of Conditions 1, 2, and 3. In addition, we also give explicit procedures for computing the constants which appear in these conditions.

Consider a matrix-valued function $a(\cdot) = (a_{ij}(\cdot))_{d \times d} : \mathbb{R}^d \mapsto \mathbb{R}^{d \times d}$ defined as

$$a_{ij}(x) := \sum_{k=1}^d \sigma_{ik}(x)\sigma_{jk}(x) \quad \text{for } 1 \leq i, j \leq d.$$

Assumption 4. Every component of μ and a is three times continuously differentiable. Moreover, there exists a constant M such that $\|\mu^{(i)}\|_\infty \leq M$ and $\|a^{(i)}\|_\infty \leq M$ for $i = 0, 1, 2, 3$.

Algorithm 7: Computation of local Lipchitz constant C_S .

- 1: **Input:** M in Assumption 4, λ_\downarrow and λ_\uparrow in Assumption 5, dimension d , time T , an arbitrary number $\epsilon \in (0, 1)$.
 - 2: $C_0 \leftarrow (2\sqrt{\pi})^{-d} \lambda_\uparrow^{d/2}$.
 - 3: $C_1 \leftarrow (2\epsilon e)^{-1/2} \lambda_\downarrow^{1/2} \lambda_\uparrow^{-1} C_0$.
 - 4: $C_2 \leftarrow C_0 \left(\frac{4\lambda_\downarrow}{\epsilon \lambda_\uparrow} \right)^2$.
 - 5: $C_3 \leftarrow C_0 \left(\frac{4\lambda_\downarrow}{\epsilon \lambda_\uparrow} \right)^2 + C_0 \frac{M}{4} \left(\frac{2\lambda_\downarrow}{\epsilon e} \right)^{\frac{1}{2}}$.
 - 6: $C_4 \leftarrow dMC_3 + d(d-1)MC_2 + d(d+1)MC_1 + T^{\frac{1}{2}}(0.5d^2 + d)MC_0$.
 - 7: $C_5 \leftarrow \left(\frac{4\pi\lambda_\downarrow}{1-\epsilon} \right)^{-\frac{d}{2}}$.
 - 8: $C_6 \leftarrow C_4 \left(\frac{4\pi\lambda_\downarrow}{1-\epsilon} \right)^{\frac{d}{2}}$.
 - 9: $C_7 \leftarrow C_5 C_6 \exp(C_6 T)$.
 - 10: $C_8 \leftarrow 2C_1 C_7 \left(\frac{4\pi\lambda_\downarrow}{1-\epsilon} \right)^{\frac{d}{2}}$.
 - 11: $C_S \leftarrow \left[\frac{dC_1}{T^{\frac{d+1}{2}}} + \frac{dC_8}{T^{\frac{d-1}{2}}} \right] \exp\left(-\frac{(1-\epsilon) \inf_{\bar{x} \in S} \|\bar{x} - x_0\|_2}{4\lambda_\downarrow T}\right)$.
 - 12: **Output** C_S .
-

Assumption 5. *There exist constants $0 < \lambda_\downarrow < \lambda_\uparrow < \infty$ such that for all $x \in \mathbb{R}^d$ and $\xi = (\xi_i)_d \in \mathbb{R}^d$, we have*

$$\lambda_\downarrow \|\xi\|_2 \leq \sqrt{\xi^T a(x) \xi} \leq \lambda_\uparrow \|\xi\|_2.$$

Under Assumptions 4 and 5, it is proved in [10] that the SDE (1) possesses a transition density denoted by $p(x, t; y, \tau)$, which satisfies

$$\mathbb{P}(X(t) \in dx | X(\tau) = y) = p(x, t; y, \tau) dx$$

for $\tau < t$. Therefore, Condition 1 is proved given Assumptions 4 and 5.

In the following proposition, we will establish Condition 2 via the Kolmogorov forward equation.

Proposition 1. *Suppose Assumptions 4 and 5 are satisfied. Then for any relatively compact set S , the density $p(\cdot) = p(\cdot, T; x_0, 0)$ is Lipschitz continuous with Lipschitz constant C_S ; i.e.*

$$|p(x) - p(y)| \leq C_S \|x - y\|_2 \quad \forall x, y \in S.$$

Furthermore, C_S can be computed by Algorithm 7.

Proof. Our methodology is closely related to the parametrix method introduced in [10]. Following the same scheme, we focus on explicit computation of the constants.

Under Assumptions 4 and 5, $p(\cdot, \cdot; y, \tau)$ is a solution of the Kolmogorov forward equation:

$$\frac{\partial}{\partial t} p(x, t; y, \tau) = - \sum_{i=1}^d \frac{\partial}{\partial x_i} [\mu_i(x) p(x, t; y, \tau)] + \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d \frac{\partial^2}{\partial x_i \partial x_j} [a_{ij}(x) p(x, t; y, \tau)]. \quad (19)$$

We shall rewrite (19) into its non-divergence form as

$$\mathcal{L}_f p := \sum_{i,j=1}^d a_{ij}(x) \frac{\partial^2 p(x, t; y, \tau)}{\partial x_i \partial x_j} + \sum_{i=1}^d b_i(x) \frac{\partial p(x, t; y, \tau)}{\partial x_i} + c(x)p(x, t; y, \tau) - \frac{\partial p(x, t; y, \tau)}{\partial t} = 0,$$

where

$$b_i(x) := \sum_{j=1}^d \frac{\partial a_{ij}(x)}{\partial x_j} - \mu_i(x), \quad c(x) := \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d \frac{\partial^2 a_{ij}(x)}{\partial x_i \partial x_j} - \sum_{i=1}^d \frac{\partial \mu_i(x)}{\partial x_i},$$

and \mathcal{L}_f is a uniform parabolic operator. By Assumption 4, it follows that

$$\|b(x)\|_\infty \leq (d + 1)M, \quad |c(x)| \leq (0.5d^2 + d)M. \tag{20}$$

We denote by $a^{-1}(x) = (a_{ij}^{-1}(x))_{d \times d}$ the inverse matrix of $(a_{ij}(x))_{d \times d}$, and define

$$\theta(x, \xi) := \sum_{i,j=1}^d a_{ij}^{-1}(\xi)(x_i - \xi_i)(x_j - \xi_j).$$

Assumption 5 implies that for all $\xi \in \mathbb{R}^d$,

$$\lambda_\uparrow^{-1} \|\xi\|_2 \leq \sqrt{\xi^T a^{-1}(x) \xi} \leq \lambda_\downarrow^{-1} \|\xi\|_2.$$

Following the idea of the parametrix method, we also define a partial differential equation with constant coefficients, namely,

$$\mathcal{L}_0^y u := \sum_{i,j=1}^n a_{ij}(y) \frac{\partial^2 u(x, t)}{\partial x_i \partial x_j} - \frac{\partial u(x, t)}{\partial t} = 0.$$

The fundamental solution of function $\mathcal{L}_0^y u = 0$ is given by

$$Z(x, t; \xi, \tau) = C_Z(\xi)w(x, t; \xi, \tau),$$

where

$$C_Z(\xi) := (2\sqrt{\pi})^{-d} [\det(a_{ij}(\xi))]^{1/2} \leq (2\sqrt{\pi})^{-d} \lambda_\uparrow^{d/2} =: C_0,$$

$$w(x, t; \xi, \tau) := (t - \tau)^{-d/2} \exp\left(-\frac{\theta(x, \xi)}{4(t - \tau)}\right).$$

According to Theorem 1.3 and Theorem 1.10 in [10], $p(x, t; \xi, \tau)$ can be represented by the parametrix method as

$$p(x, t; \xi, \tau) = Z(x, t; \xi, \tau) + \int_\tau^t \int_{\mathbb{R}^d} Z(x, t; y, s) \Phi(y, s; \xi, \tau) dy ds,$$

where

$$\Phi(x, t; \xi, \tau) := \sum_{k=1}^\infty (\mathcal{L}_f Z)_k(x, t; \xi, \tau),$$

$$(\mathcal{L}_f Z)_1 := \mathcal{L}_f Z,$$

$$(\mathcal{L}_f Z)_{k+1} := \int_\tau^t \int_{\mathbb{R}^d} \mathcal{L}_f Z(x, t; y, s) (\mathcal{L}_f Z)_k(y, s; \xi, \tau) dy ds$$

for $k \geq 1$. Furthermore, the partial derivatives of the fundamental solution admit the following expression:

$$\frac{\partial}{\partial x_i} p(x, t; \xi, \tau) = \frac{\partial}{\partial x_i} Z(x, t; \xi, \tau) + \int_{\tau}^t \int_{\mathbb{R}^d} \frac{\partial}{\partial x_i} Z(x, t; y, s) \Phi(y, s; \xi, \tau) dy ds. \tag{21}$$

Let us pick $\epsilon \in (0, 1)$; then we can derive a bound for Z as

$$|Z(x, t; \xi, \tau)| \leq C_0 \times (t - \tau)^{-d/2} \exp\left(-\frac{(1 - \epsilon)\|x - \xi\|_2^2}{4\lambda_{\downarrow}(t - \tau)}\right).$$

For the bound of $\partial Z(x, t; \xi, \tau)/\partial x_i$, note that

$$\left| \frac{\partial Z(x, t; \xi, \tau)}{\partial x_i} \right| = [4(t - \tau)]^{-1} \left| \frac{\partial \theta(x, \xi)}{\partial x_i} \right| C_Z(\xi) w(x, t; \xi, \tau),$$

and that

$$\left| \frac{\partial \theta(x, \xi)}{\partial x_i} \right| = \left| 2 \sum_{j=1}^d a_{ij}^{-1}(\xi)(x_j - \xi_j) \right| \leq 2\lambda_{\uparrow}^{-1} \|x - \xi\|_2.$$

Combining the definition of $C_Z(\cdot)$ and the above two equations implies that

$$\begin{aligned} \left| \frac{\partial Z(x, t; \xi, \tau)}{\partial x_i} \right| &\leq \frac{\lambda_{\uparrow}^{-1}}{2} C_0 \|x - \xi\| (t - \tau)^{-\frac{d+1}{2}} (\theta(x, \xi))^{-1/2} \left[\frac{\theta(x, \xi)}{t - \tau} \right]^{1/2} \\ &\quad \times \exp\left(-\frac{\epsilon \theta(x, \xi)}{4(t - \tau)}\right) \exp\left(-\frac{(1 - \epsilon)\theta(x, \xi)}{4(t - \tau)}\right). \end{aligned}$$

Applying the inequalities

$$\left[\frac{\theta(x, \xi)}{t - \tau} \right]^{1/2} \exp\left(-\frac{\epsilon \theta(x, \xi)}{4(t - \tau)}\right) \leq \sup_{x \in [0, +\infty)} x^{\frac{1}{2}} e^{-\frac{\epsilon x}{4}} = \left(\frac{2}{\epsilon e}\right)^{1/2}$$

and

$$\|x - \xi\| (\theta(x, \xi))^{-1/2} \leq \lambda_{\downarrow}^{1/2},$$

we obtain

$$\left| \frac{\partial Z(x, t; \xi, \tau)}{\partial x_i} \right| \leq \frac{C_1}{(t - \tau)^{\frac{d+1}{2}}} \exp\left(-\frac{(1 - \epsilon)\|x - \xi\|_2^2}{4\lambda_{\downarrow}(t - \tau)}\right)$$

by setting

$$C_1 := (2\epsilon e)^{-1/2} \lambda_{\downarrow}^{1/2} \lambda_{\uparrow}^{-1} C_0.$$

Similarly, we can derive bounds for $\partial^2 Z(x, t; \xi, \tau)/(\partial x_i \partial x_j)$ and $\partial^2 Z(x, t; \xi, \tau)/(\partial x_i^2)$. For $i \neq j$,

$$\begin{aligned} \left| \frac{\partial^2 Z(x, t; \xi, \tau)}{\partial x_i \partial x_j} \right| &\leq \frac{C_2}{(t - \tau)^{\frac{d+1}{2}} \|x - \xi\|} \exp\left(-\frac{(1 - \epsilon)\|x - \xi\|_2^2}{4\lambda_{\downarrow}(t - \tau)}\right), \\ \left| \frac{\partial^2 Z(x, t; \xi, \tau)}{\partial x_i^2} \right| &\leq \frac{C_3}{(t - \tau)^{\frac{d+1}{2}} \|x - \xi\|} \exp\left(-\frac{(1 - \epsilon)\|x - \xi\|_2^2}{4\lambda_{\downarrow}(t - \tau)}\right), \end{aligned}$$

where

$$C_2 := C_0 \left(\frac{4\lambda_{\downarrow}}{\epsilon \in \lambda_{\uparrow}} \right)^2,$$

$$C_3 := C_0 \left(\frac{4\lambda_{\downarrow}}{\epsilon \in \lambda_{\uparrow}} \right)^2 + C_0 \frac{M}{4} \left(\frac{2\lambda_{\downarrow}}{\epsilon \epsilon} \right)^{\frac{1}{2}}.$$

By definition of $Z(\cdot)$ we can observe that

$$\begin{aligned} \mathcal{L}_f Z(x, t; \xi, \tau) &= \sum_{i,j=1}^d [a_{ij}(x) - a_{ij}(\xi)] \frac{\partial^2 Z(x, t; \xi, \tau)}{\partial x_i \partial x_j} \\ &+ \sum_{i=1}^d b_i(x) \frac{\partial Z(x, t; \xi, \tau)}{\partial x_i} + c(x)Z(x, t; \xi, \tau). \end{aligned}$$

Suppose $0 \leq t - \tau \leq T$ in the sequel. By considering the upper bounds of the partial derivatives of Z , as well as (20) and Assumption 4, we obtain

$$|\mathcal{L}_f Z(x, t; \xi, \tau)| \leq \frac{C_4}{(t - \tau)^{\frac{d+1}{2}}} \exp \left(-\frac{(1 - \epsilon)\|x - \xi\|_2^2}{4\lambda_{\downarrow}(t - \tau)} \right), \tag{22}$$

where

$$C_4 := dMC_3 + d(d - 1)MC_2 + d(d + 1)MC_1 + T^{\frac{1}{2}}(0.5d^2 + d)MC_0.$$

Now, in order to find a bound for $\Phi(x, t; \xi, \tau)$, we need to introduce a technical lemma.

Lemma 4 (Lemma 1.3 of [10]). *If β and γ are two constants in $(-\infty, \frac{d}{2} + 1)$, then*

$$\begin{aligned} &\int_{\tau}^t \int_{\mathbb{R}^d} (t - s)^{-\beta} \exp \left(-\frac{h\|x - y\|_2^2}{4(t - s)} \right) (s - \tau)^{-\gamma} \exp \left(-\frac{h\|y - \xi\|_2^2}{4(s - \tau)} \right) dy ds \\ &= \left(\frac{4\pi}{h} \right)^{\frac{d}{2}} \text{Beta} \left(\frac{d}{2} - \beta + 1, \frac{d}{2} - \gamma + 1 \right) (t - \tau)^{\frac{d}{2} + 1 - \beta - \gamma} \exp \left(-\frac{h\|x - \xi\|_2^2}{4(t - \tau)} \right), \end{aligned}$$

where $\text{Beta}(\cdot)$ is the beta function.

From (22) and Lemma 4, we can derive

$$\begin{aligned} |(\mathcal{L}_f Z)_2(x, t; \xi, \tau)| &\leq \int_{\tau}^t \int_{\mathbb{R}^d} |\mathcal{L}_f Z(x, t; y, s)| |\mathcal{L}_f Z(y, s; \xi, \tau)| dy ds. \\ &\leq \frac{C_5 C_6^2}{1!} (t - \tau)^{1 - \frac{d}{2}} \exp \left(-\frac{(1 - \epsilon)\|x - \xi\|_2^2}{4\lambda_{\downarrow}(t - \tau)} \right), \end{aligned}$$

where

$$C_5 := \left(\frac{4\pi\lambda_{\downarrow}}{1 - \epsilon} \right)^{-\frac{d}{2}}, \quad C_6 := C_4 \left(\frac{4\pi\lambda_{\downarrow}}{1 - \epsilon} \right)^{\frac{d}{2}}.$$

By induction we can show that, for any positive integer m ,

$$|(\mathcal{L}_f Z)_m(x, t; \xi, \tau)| \leq \frac{C_5 C_6^m}{(m-1)!} (t-\tau)^{m-\frac{d}{2}-1} \exp\left(-\frac{(1-\epsilon)\|x-\xi\|_2^2}{4\lambda_\downarrow(t-\tau)}\right).$$

It turns out that

$$\begin{aligned} \Phi(x, t; \xi, \tau) &\leq \sum_{m=1}^{\infty} |(\mathcal{L}Z)_m(x, t; \xi, \tau)| \\ &\leq \frac{C_7}{(t-\tau)^{\frac{d}{2}}} \exp\left(-\frac{(1-\epsilon)\|x-\xi\|_2^2}{4\lambda_\downarrow(t-\tau)}\right), \end{aligned}$$

where

$$C_7 := \sum_{m=1}^{\infty} \frac{C_5 C_6^m}{(m-1)!} T^{m-1} = C_5 C_6 \exp(C_6 T).$$

Recalling (21), we can apply Lemma 4 again and conclude that

$$\begin{aligned} \left| \frac{\partial}{\partial x_i} p(x, t; \xi, \tau) \right| &\leq \left| \frac{\partial}{\partial x_i} Z(x, t; \xi, \tau) \right| + \int_{\tau}^t \int_{\mathbb{R}^d} \left| \frac{\partial}{\partial x_i} Z(x, t; y, s) \Phi(y, s; \xi, \tau) \right| dy ds. \\ &\leq \left[\frac{C_1}{(t-\tau)^{\frac{d+1}{2}}} + \frac{C_8}{(t-\tau)^{\frac{d-1}{2}}} \right] \exp\left(-\frac{(1-\epsilon)\|x-\xi\|_2^2}{4\lambda_\downarrow(t-\tau)}\right), \end{aligned}$$

where

$$C_8 := 2C_1 C_7 \left(\frac{4\pi\lambda_\downarrow}{1-\epsilon} \right)^{\frac{d}{2}}.$$

Therefore, we obtain an upper bound for $|\nabla_x p(x, t; \xi, \tau)|$ by considering

$$|\nabla_x p(x, t; \xi, \tau)| \leq d \times \left| \frac{\partial}{\partial x_i} p(x, t; \xi, \tau) \right|.$$

Therefore, for all $x, y \in S$ we have

$$|p(x) - p(y)| \leq C_S \|x - y\|_2,$$

where

$$C_S = \left[\frac{dC_1}{T^{\frac{d+1}{2}}} + \frac{dC_8}{T^{\frac{d-1}{2}}} \right] \exp\left(-\frac{(1-\epsilon) \inf_{\bar{x} \in S} \|\bar{x} - x_0\|_2}{4\lambda_\downarrow T}\right). \quad \square$$

Next, we will propose a computational procedure for lower bounds of transition density. There is a substantial amount of literature that studies lower bounds for the transition density of diffusions, through analytical approaches or probabilistic approaches. For instance, Aronson [2] develops estimates of lower bounds of fundamental solutions of second-order parabolic partial differential equations in divergence form. Using Malliavin calculus, Kusuoka

and Stroock [16] derived a lower bound for the transition density of uniformly elliptic diffusions. Bally [3] generalized the idea of [16] to locally elliptic diffusions. We follow the approach suggested by Sheu [23] and review it in order to find explicit expressions to obtain a computable lower bound.

In order to keep our paper self-contained, we first introduce some notation to be used later. Let \mathcal{L}_b be the generator of the Komolgorov backward equation:

$$\mathcal{L}_b u(x, t) := \frac{1}{2} \sum_{i,j=1}^n a_{ij}(x) \frac{\partial^2 u(x, t)}{\partial x_i \partial x_j} + \sum_{i=1}^d \mu_i(x) \frac{\partial u(x, t)}{\partial x_i} - \frac{\partial u(x, t)}{\partial t}.$$

The transition density as a function of $(x, t) \mapsto p(y, t; x, 0)$ coincides with the fundamental solution of the Komolgorov backward equation:

$$\begin{aligned} \mathcal{L}_b u(x, t) &= 0, & t > 0, & \quad x \in \mathbb{R}^d, \\ u(0, x) &= u_0(x). \end{aligned}$$

Throughout the rest of this section, we suppose that Assumptions 4 and 5 are in force. Recall that $a^{-1}(x)$ is the inverse matrix of $a(x)$, and define

$$k(x, \psi) = \frac{1}{2} \sum_{i,j=1}^d a_{ij}^{-1}(x) (\mu_i(x) - \psi_i) (\mu_j(x) - \psi_j).$$

For a fixed $y_0 \in \mathbb{R}^d$, we define

$$f^\beta(y; y_0) := \left(\frac{1}{\sqrt{2\pi\beta}} \right)^d \frac{1}{\sqrt{\det a(y_0)}} \exp \left(-\frac{1}{2\beta} \sum_{i,j=1}^d a_{ij}^{-1}(y_0) (y - y_0)_i (y - y_0)_j \right),$$

and

$$p^\beta(y_0, t; x, 0) := \mathbb{E}_x[f^\beta(X(t); y_0)].$$

The continuity of the density implies

$$\lim_{\beta \rightarrow 0} p^\beta(y_0, t; x, 0) = p(y_0, t; x, 0). \tag{23}$$

For simplicity, we also define the logarithmic transform of p and p^β as

$$\begin{aligned} J(t, x) &:= -\log(p(y_0, t; x, 0)), \\ J^\beta(t, x) &:= -\log(p^\beta(y_0, t; x, 0)). \end{aligned}$$

To prepare the analysis, which is based on stochastic control, we introduce the space of control functions, denoted by $\mathcal{F}_{T,x}$. The class $\mathcal{F}_{T,x}$ is defined as a family of measurable functions $\psi : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that the SDE

$$d\eta(t) = \psi(t, \eta(t))dt + \sigma(\eta(t))dW(t), \quad \eta(0) = x$$

has a weak solution $\eta(\cdot)$ that satisfies

$$\mathbb{E} \left(\int_0^T \|\psi(t, \eta(t))\|_2^2 dt \right) < \infty.$$

Algorithm 8: Computation of the lower bound δ_S .

- 1: $D_S \leftarrow \sup_{x \in S} \|x - x_0\|_2$.
 - 2: Evaluate $\Psi_i(D_S)$; $i = 1, 2, 3, 4$ by (31), (32), (33), and (34).
 - 3: Evaluate $J_{\uparrow}(D_S; T)$ by (35).
 - 4: $\delta_S \leftarrow \exp(-J_{\uparrow}(D_S; T))$.
 - 5: **Output** δ_S .
-

Now we state a lemma that is crucial for proving the main result of this section.

Lemma 5. Recall the definition of $\mathcal{F}_{T,x}$ and $\eta(\cdot)$ from the previous paragraph; then we have

$$J^\beta(T, x) = \inf_{\psi \in \mathcal{F}_{T,x}} \mathbb{E} \left(\int_0^T k(\eta(t), \psi(t)) dt + J^\beta(0, \eta(T)) \right).$$

Together with (23), we see that

$$J(T, x) = \lim_{\beta \rightarrow 0} \inf_{\psi \in \mathcal{F}_{T-\beta,x}} \mathbb{E} \left(\int_0^{T-\beta} k(\eta(t), \psi(t)) dt + J^\beta(0, \eta(T-\beta)) \right). \tag{24}$$

Proof. See [9]. □

Theorem 6. Suppose Assumptions 4 and 5 are satisfied. Then, for any relatively compact set S , the density $p(\cdot) = p(\cdot, T; x_0, 0)$ has a uniform lower bound $\delta_S > 0$ in S ; i.e.

$$p(x) \geq \delta_S \quad \forall x \in S.$$

Furthermore, δ_S can be computed by Algorithm 8.

Proof. Finding a lower bound of the density $p(y_0, T; x_0, 0)$ is equivalent to finding an upper bound for $J(T, x_0)$. Towards this end, it suffices to find an upper bound for $J^\beta(T, x_0)$ that is uniform in β . We shall define $\phi(\cdot)$ as a linear function such that $\phi(0) = x_0, \phi(T) = y_0$. Write

$$\psi(t, x) = \frac{y_0 - x_0}{T} - \frac{x - \phi(t)}{T - t}, \quad 0 \leq t \leq T - \beta.$$

It is not hard to see that $\psi \in \mathcal{F}_{T-\beta, x_0}$. Therefore,

$$J^\beta(T, x_0) \leq \mathbb{E} \left(\int_0^{T-\beta} k(\eta(t), \psi(t)) dt + J^\beta(0, \eta(T-\beta)) \right), \tag{25}$$

according to Lemma 5. Notice that

$$(\eta(t) - \phi(t))_i = (T - t) \sum_{l=1}^{d'} \int_0^t \frac{1}{T - s} \sigma_{il}(\eta(s)) dW_l(s), \quad \text{for } i = 1, \dots, d. \tag{26}$$

It follows that

$$\mathbb{E}((\eta(t) - \phi(t))_i (\eta(t) - \phi(t))_j) = (T - t)^2 \mathbb{E} \left(\int_0^t \frac{1}{(T - s)^2} a_{ij}(\eta(s)) ds \right)$$

and

$$\mathbb{E}(\|\eta(t) - \phi(t)\|_2^2) = (T - t)^2 \mathbb{E} \left(\int_0^t \frac{1}{(T - s)^2} \sum_{i=1}^d a_{ii}(\eta(s)) ds \right) \leq d\lambda_{\uparrow}(T - t). \tag{27}$$

We now apply a Taylor expansion of $k(\eta(t), \psi(t))$ around $(\phi(t), \dot{\phi}(t))$, where $\dot{\phi}(\cdot)$ denotes the derivative of $\phi(\cdot)$. For notational simplicity, we define

$$\begin{aligned} \Delta_1(t) &= \eta(t) - \phi(t), \\ \Delta_2(t) &= \psi(t) - \dot{\phi}(t) = -\frac{1}{T - t} \Delta_1(t). \end{aligned}$$

We also define

$$D_{x_i} k(\lambda) = \frac{\partial}{\partial x_i} k(\phi(t) + \lambda \Delta_1(t), \dot{\phi}(t) + \lambda \Delta_2(t)),$$

and similarly for $D_{\psi_i} k, D_{x_i, x_j} k, D_{x_i, \psi_j} k$, and $D_{\psi_i, \psi_j} k$. The Taylor expansion with remainders of third order is given as follows:

$$k(\eta(t), \psi(t)) = k_0(t) + k_1(t) + k_{2,1}(t) + k_{2,2}(t) + k_{2,3}(t) + k_{3,1}(t) + k_{3,2}(t) + k_{3,3}(t),$$

where

$$\begin{aligned} k_0(t) &:= k(\phi(t), \dot{\phi}(t)), \\ k_1(t) &:= \sum_{i=1}^d (D_{x_i} k(0) \Delta_{1,i}(t) + D_{\psi_i} k(0) \Delta_{2,i}(t)), \\ k_{2,1}(t) &:= \frac{1}{2} \sum_{i,j=1}^d D_{x_i, x_j} k(0) \Delta_{1,i}(t) \Delta_{1,j}(t), \\ k_{2,2}(t) &:= \sum_{i,j=1}^d D_{x_i, \psi_j} k(0) \Delta_{1,i}(t) \Delta_{2,j}(t), \\ k_{2,3}(t) &:= \frac{1}{2} \sum_{i,j=1}^d D_{\psi_i, \psi_j} k(0) \Delta_{2,i}(t) \Delta_{2,j}(t), \\ k_{3,1}(t) &:= \sum_{i,j=1}^d \int_0^1 \int_0^1 (D_{x_i, x_j} k(\lambda\mu) - D_{x_i, x_j} k(0)) \Delta_{1,i}(t) \Delta_{1,j}(t) \lambda d\mu d\lambda, \\ k_{3,2}(t) &:= \sum_{i,j=1}^d 2 \int_0^1 \int_0^1 (D_{x_i, \psi_j} k(\lambda\mu) - D_{x_i, \psi_j} k(0)) \Delta_{1,i}(t) \Delta_{2,j}(t) \lambda d\mu d\lambda, \\ k_{3,3}(t) &:= \sum_{i,j=1}^d \int_0^1 \int_0^1 (D_{\psi_i, \psi_j} k(\lambda\mu) - D_{\psi_i, \psi_j} k(0)) \Delta_{2,i}(t) \Delta_{2,j}(t) \lambda d\mu d\lambda. \end{aligned}$$

Now we integrate all the above terms from 0 to $T - \beta$ with respect to the variable t , then take expectations, and analyze the upper bounds of the result term by term.

- **Zeroth-order term:** Notice that k is in quadratic form, with matrix $(a_{ij}^{-1}(x))$, so

$$\mathbb{E}\left(\int_0^{T-\beta} k_0(t)dt\right) = \mathbb{E}\left(\int_0^{T-\beta} k(\phi(t), \dot{\phi}(t))dt\right) \leq \lambda_{\downarrow}^{-1}T \left(M + \frac{|y_0 - x_0|}{T}\right)^2.$$

- **First-order term:** We treat the first-order term $k_1(t)$ first. Since $\Delta_{2,i}(t)$ is a martingale by (26), the first-order term satisfies

$$\mathbb{E}\left(\int_0^{T-\beta} k_1(t)dt\right) = \mathbb{E}\left(\int_0^{T-\beta} (D_{x_i}k(0)\Delta_{1,i}(t) + D_{\psi_i}k(0)\Delta_{2,i}(t)) dt\right) = 0.$$

- **Second-order terms:** We then treat the second-order terms. As $D_{\psi_i, \psi_j}k(0) = a_{ij}^{-1}(\phi(t))$,

$$\begin{aligned} \mathbb{E}\left(\int_0^{T-\beta} k_{2,3}(t)dt\right) &= \mathbb{E}\left(\int_0^{T-\beta} \frac{1}{2} \sum_{i,j=1}^d D_{\psi_i, \psi_j}k(0)\Delta_{2,i}(t)\Delta_{2,j}(t)dt\right) \\ &= \frac{1}{2} \int_0^{T-\beta} \mathbb{E}\left(\int_0^t \frac{1}{(T-s)^2} \sum_{i,j=1}^d a_{ij}^{-1}(\phi(t))a_{ij}(\eta(s))ds\right) dt. \end{aligned}$$

Writing

$$a_{ij}(\eta(s)) = (a_{ij}(\eta(s)) - a_{ij}(\phi(s))) + (a_{ij}(\phi(s)) - a_{ij}(\phi(t))) + a_{ij}(\phi(t)),$$

and noticing that $(a_{ij}(t))$ is symmetric, we see that

$$\frac{1}{2} \int_0^{T-\beta} \mathbb{E}\left(\int_0^t \frac{1}{(T-s)^2} \sum_{i,j=1}^d a_{ij}^{-1}(\phi(t))a_{ij}(\phi(t))ds\right) dt = \frac{d}{2}(\log(T) - \log(\beta)). \tag{28}$$

Assumption 4 implies the Lipschitz continuity of $a(\cdot)$, which gives

$$\begin{aligned} &\frac{1}{2} \int_0^{T-\beta} \mathbb{E}\left(\int_0^t \frac{1}{(T-s)^2} \sum_{i,j=1}^d g_{ij}(\phi(t))(a_{ij}(\eta(s)) - a_{ij}(\phi(s)))ds\right) dt \\ &\leq \frac{1}{2} \int_0^{T-\beta} \mathbb{E}\left(\int_0^t \frac{M}{(T-s)^2} \sum_{i,j=1}^d |g_{ij}(\phi(t))| \times \|\eta(s) - \phi(s)\|_2 ds\right) dt \\ &= \frac{1}{2} \int_0^{T-\beta} \int_0^t \frac{M}{(T-s)^2} \sum_{i,j=1}^d |a_{ij}^{-1}(\phi(t))| \times \mathbb{E}(\|\eta(s) - \phi(s)\|_2) ds dt. \end{aligned}$$

By (27) and Jensen’s inequality,

$$\mathbb{E}[\|\eta(s) - \phi(s)\|_2] \leq (d\lambda_{\uparrow}(T-t))^{1/2}.$$

Observe that $\sum_{i,j=1}^d |a_{ij}^{-1}(\phi(t))| \leq d\lambda_{\downarrow}^{-1}$, so we have

$$\begin{aligned} &\frac{1}{2} \int_0^{T-\beta} \mathbb{E}\left(\int_0^t \frac{1}{(T-s)^2} \sum_{i,j=1}^d a_{ij}^{-1}(\phi(t))(a_{ij}(\eta(s)) - a_{ij}(\phi(s)))ds\right) dt \\ &\leq M(d\lambda_{\uparrow}T)^{1/2}d\lambda_{\downarrow}^{-1}. \end{aligned} \tag{29}$$

By the Lipschitz continuity of $(a_{ij}(\cdot))$, it follows that

$$|a_{ij}(\phi(s)) - a_{ij}(\phi(t))| \leq MT^{-1} \|x_0 - y_0\|_2 |s - t|.$$

Therefore,

$$\begin{aligned} & \frac{1}{2} \int_0^{T-\beta} \mathbb{E} \left(\int_0^t \frac{1}{(T-s)^2} \sum_{i,j=1}^d a_{ij}^{-1}(\phi(t))(a_{ij}(\phi(s)) - a_{ij}(\phi(t))) ds \right) dt \\ & \leq \frac{1}{2} d \lambda_{\downarrow}^{-1} M T^{-1} \|x_0 - y_0\|_2 \int_0^{T-\beta} \left(\int_0^t \frac{t-s}{(T-s)^2} ds \right) dt \\ & \leq \frac{1}{2} d \lambda_{\downarrow}^{-1} M \|x_0 - y_0\|_2. \end{aligned} \tag{30}$$

Combining (28), (29), and (30) yields

$$\begin{aligned} \mathbb{E} \left(\int_0^{T-\beta} k_{2,3}(t) dt \right) &= \mathbb{E} \left(\int_0^{T-\beta} \frac{1}{2} \sum_{i,j=1}^d D_{\psi_i, \psi_j} k(0) \Delta_{2,i}(t) \Delta_{2,j}(t) dt \right) \\ &\leq \frac{d}{2} (\log(T) - \log(\beta)) + M(d\lambda_{\uparrow} T)^{1/2} d\lambda_{\downarrow}^{-1} + \frac{d}{2} \lambda_{\downarrow}^{-1} M \|x_0 - y_0\|_2. \end{aligned}$$

By the chain rule and Assumption 4, we obtain

$$\begin{aligned} |D_{x_i} a_{ij}^{-1}(x)| &\leq d^2 \lambda_{\downarrow}^{-2} M, \\ |D_{x_i \psi_j} k(0)| &\leq \Psi_1(\|x_0 - y_0\|_2/T), \\ |D_{x_i x_j} k(0)| &\leq \Psi_2(\|x_0 - y_0\|_2/T), \end{aligned}$$

where $\Psi_i(\cdot) : \mathbb{R} \rightarrow \mathbb{R}, i = 1, 2$, are defined as

$$\Psi_1(x) := d^2 \lambda_{\downarrow}^{-2} M(M + x) + d \lambda_{\downarrow}^{-1} M, \tag{31}$$

$$\begin{aligned} \Psi_2(x) &:= (M + x)^2 d^2 \left(\frac{1}{2} \lambda_{\downarrow}^{-2} M d + \lambda_{\downarrow}^{-3} M^2 d^2 \right) + 2 \lambda_{\downarrow}^{-1} M^2 d^2 + \lambda_{\downarrow}^{-1} M d x \\ &\quad + 2 \lambda_{\downarrow}^{-2} M^3 d^3 + 2 \lambda_{\downarrow}^{-2} M^2 d^2 x. \end{aligned} \tag{32}$$

Taking (27) into consideration, we obtain

$$\begin{aligned} \mathbb{E} \left(\int_0^{T-\beta} k_{2,2}(t) dt \right) &= \mathbb{E} \left(\int_0^{T-\beta} \frac{1}{2} \sum_{i,j=1}^d D_{x_i, \psi_j} k(0) \Delta_{1,i}(t) \Delta_{2,j}(t) dt \right) \\ &\leq \frac{1}{2} d \lambda_{\uparrow} T \Psi_1(\|x_0 - y_0\|_2/T), \\ \mathbb{E} \left(\int_0^{T-\beta} k_{2,1}(t) dt \right) &= \mathbb{E} \left(\int_0^{T-\beta} \frac{1}{2} \sum_{i,j=1}^d D_{x_i, x_j} k(0) \Delta_{1,i}(t) \Delta_{1,j}(t) dt \right) \\ &\leq \frac{1}{4} d \lambda_{\uparrow} T^2 \Psi_2(\|x_0 - y_0\|_2/T), \end{aligned}$$

- **Third-order terms:** We proceed to analyze the third-order remainder terms. Let us consider $k_{3,3}(t)$ first. Notice that

$$|D_{\psi_i, \psi_j} k(\lambda, \mu) - D_{\psi_i, \psi_j} k(0)| \leq \lambda_{\downarrow}^{-2} d^2 M \lambda \mu \|\Delta_1(t)\|_2;$$

thus,

$$\begin{aligned} & \left| \mathbb{E} \left(\int_0^{T-\beta} k_{3,3}(t) dt \right) \right| \\ &= \left| \mathbb{E} \left(\sum_{i,j=1}^d \int_0^1 \int_0^1 (D_{\psi_i, \psi_j} k(\lambda, \mu) - D_{\psi_i, \psi_j} k(0)) \Delta_{2,i}(t) \Delta_{2,j}(t) \lambda d\mu d\lambda \right) \right| \\ &\leq \frac{1}{6} M d^3 \lambda_{\downarrow}^{-2} \mathbb{E}(\|\Delta^1(t)\|_2 \|\Delta_2(t)\|_2^2). \end{aligned}$$

Then, by the Burkholder–Davis–Gundy inequality,

$$\begin{aligned} \mathbb{E}(\|\Delta_1(t)\|_2 \|\Delta_2(t)\|_2^2) &\leq (T-t) C_{\text{BDG}}(3) d^{\frac{1}{2}} \sum_{i=1}^d \mathbb{E} \left(\left(\int_0^t \frac{1}{(T-s)^2} a_{ii}(\eta(s)) ds \right)^{\frac{3}{2}} \right) \\ &\leq C_{\text{BDG}}(3) d^{\frac{3}{2}} \lambda_{\uparrow}^{\frac{3}{2}} (T-t)^{-\frac{1}{2}}, \end{aligned}$$

where $C_{\text{BDG}}(3)$ is the explicit constant in the Burkholder–Davis–Gundy inequality. We can pick

$$C_{\text{BDG}}(p) = \left(\frac{p(p-1)}{2} \left(\frac{p}{p-1} \right)^p \right)^{p/2}$$

(see Proposition 4.4.3 of [20]). To summarize, we obtain

$$\begin{aligned} & \left| \mathbb{E} \left(\int_0^{T-\beta} k_{3,3}(t) dt \right) \right| \\ &= \left| \int_0^{T-\beta} \mathbb{E} \left(\sum_{i,j=1}^d \int_0^1 \int_0^1 (D_{\psi_i, \psi_j} k(\lambda, \mu) - D_{\psi_i, \psi_j} k(0)) \Delta_{2,i}(t) \Delta_{2,j}(t) \lambda d\mu d\lambda \right) dt \right| \\ &\leq \frac{1}{3} C_{\text{BDG}}(3) M d^{\frac{9}{2}} \lambda_{\downarrow}^{-2} \lambda_{\uparrow}^{\frac{3}{2}} T^{\frac{1}{2}}. \end{aligned}$$

Next, we consider the other two remainders $k_{3,2}(t)$, $k_{3,1}(t)$. Observe that

$$|D_{x_i \psi_j} k(\lambda, \mu)| \leq \Psi_1(\|x_0 - y_0\|_2 / T + \lambda \mu \|\Delta_2(t)\|_2)$$

and

$$|D_{x_i x_j} k(\lambda, \mu)| \leq \Psi_2(\|x_0 - y_0\|_2 / T + \lambda \mu \|\Delta_2(t)\|_2).$$

Thus, by a similar argument, we can also derive

$$\left| \mathbb{E} \left(\int_0^{T-\beta} k_{3,2}(t) dt \right) \right| \leq \Psi_3(\|x_0 - y_0\|_2 / T)$$

and

$$\left| \mathbb{E} \left(\int_0^{T-\beta} k_{3,1}(t) dt \right) \right| \leq \Psi_4(\|x_0 - y_0\|_2/T),$$

where $\Psi_3(\cdot)$ and $\Psi_4(\cdot)$ are defined as

$$\begin{aligned} \Psi_3(x) &:= d^4 T \lambda_{\uparrow} \lambda_{\downarrow}^{-2} M x + d^4 T \lambda_{\uparrow} \lambda_{\downarrow}^{-2} M^2 + d^3 T \lambda_{\uparrow} \lambda_{\downarrow}^{-1} M + \frac{1}{3} C_{\text{BDG}}(3) d^{\frac{7}{2}} T^{\frac{1}{2}} \lambda_{\uparrow}^{\frac{3}{2}} \lambda_{\downarrow}^{-2} M, \\ \Psi_4(x) &:= \left(\frac{1}{4} d^5 T^2 \lambda_{\uparrow} \lambda_{\downarrow}^{-2} M + \frac{1}{2} d^6 T^2 \lambda_{\uparrow} \lambda_{\downarrow}^{-3} M^2 \right) x^2 \end{aligned} \tag{33}$$

$$\begin{aligned} &+ \frac{1}{9} C_{\text{BDG}}(3) (d \lambda_{\uparrow} T)^{\frac{3}{2}} (\lambda_{\downarrow}^{-2} M d^3 + 2 \lambda_{\downarrow}^{-3} M^3 d^4) x + d^5 T^2 \lambda_{\uparrow} \lambda_{\downarrow}^{-2} M^2 x + d^6 T^2 \lambda_{\uparrow} \lambda_{\downarrow}^{-3} M^3 x \\ &+ d^3 T^2 \lambda_{\uparrow} \lambda_{\downarrow}^{-1} M x + \frac{1}{24} C_{\text{BDG}}(4) d^5 T \lambda_{\uparrow} \lambda_{\downarrow}^{-2} M + \frac{1}{12} C_{\text{BDG}}(4) d^6 T \lambda_{\uparrow} \lambda_{\downarrow}^{-3} M^2 \\ &+ \frac{2}{9} C_{\text{BDG}}(3) d^{\frac{9}{2}} T^{\frac{3}{2}} \lambda_{\uparrow}^{\frac{3}{2}} \lambda_{\downarrow}^{-2} M^2 + \frac{2}{9} C_{\text{BDG}}(3) d^{\frac{11}{2}} T^{\frac{3}{2}} \lambda_{\uparrow}^{\frac{3}{2}} \lambda_{\downarrow}^{-3} M^3 + \frac{1}{9} C_{\text{BDG}}(3) d^{\frac{5}{2}} T^{\frac{3}{2}} \lambda_{\uparrow}^{\frac{3}{2}} \lambda_{\downarrow}^{-1} M \\ &+ \frac{1}{2} d^6 T^2 \lambda_{\uparrow} \lambda_{\downarrow}^{-3} M^4 + \frac{1}{2} d^4 T^2 \lambda_{\uparrow} \lambda_{\downarrow}^{-1} M^2 + \frac{3}{4} d^5 T^2 \lambda_{\uparrow} \lambda_{\downarrow}^{-2} M^3. \end{aligned} \tag{34}$$

Finally, let us consider $\mathbb{E}(J^\beta(0, \eta(T - \beta)))$. Since

$$\begin{aligned} &\mathbb{E}((\eta(T - \beta) - y_0)_i (\eta(T - \beta) - y_0)_j) \\ &= \mathbb{E}((\eta(T - \beta) - \phi(T - \beta))_i (\eta(T - \beta) - \phi(T - \beta))_j \\ &\quad + (\phi(T - \beta) - \phi(T))_i (\phi(T - \beta) - \phi(T))_j) \\ &\leq d \lambda_{\uparrow} \beta + \beta^2 |x_0 - y_0|^2 / T^2, \end{aligned}$$

it follows that

$$\begin{aligned} &\mathbb{E}(J^\beta(0, \eta(T - \beta))) \\ &= \frac{d}{2} \log(2\pi\beta) + \frac{1}{2} \log \det(a(y_0)) + \frac{1}{2\beta} \sum_{i,j=1}^d g_{ij}(y_0) \mathbb{E}((\eta(T - \beta) - y_0)_i (\eta(T - \beta) - y_0)_j) \\ &\leq \frac{d}{2} \log(2\pi\beta) + \frac{d}{2} \log \lambda_{\uparrow} + \frac{1}{2} d^2 \lambda_{\uparrow} \lambda_{\downarrow}^{-1} + \frac{d}{2} \frac{\|x_0 - y_0\|_2^2}{T}. \end{aligned}$$

To conclude, let us summarize all the intermediate results and substitute them into (25). We have

$$J(T, x_0) = \lim_{\beta \rightarrow 0} J^\beta(T - \beta, x_0) \leq J_{\uparrow}(\|x_0 - y_0\|_2; T),$$

where $J_{\uparrow}(\cdot; T)$ is defined as

$$\begin{aligned} J_{\uparrow}(x; T) &:= \lambda_{\downarrow}^{-1} T \left(M + \frac{\|y_0 - x_0\|_2}{T} \right)^2 + \frac{d}{2} (\log(2\pi T)) + M (d \lambda_{\uparrow} T)^{1/2} d \lambda_{\downarrow}^{-1} \\ &\quad + \frac{d}{2} \lambda_{\downarrow}^{-1} M x + \frac{1}{2} d \lambda_{\uparrow} T \Psi_1(x/T) + \frac{1}{4} d \lambda_{\uparrow} T^2 \Psi_2(x/T) \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{3} C_{\text{BDG}}(3) M d^{\frac{9}{2}} \lambda_{\downarrow}^{-2} \lambda_{\uparrow}^{\frac{3}{2}} T^{\frac{1}{2}} + \Psi_3(x/T) + \Psi_4(x/T) + \frac{d}{2} \log \lambda_{\uparrow} \\
& + \frac{1}{2} d^2 \lambda_{\uparrow} \lambda_{\downarrow}^{-1} + \frac{d x^2}{2 T}, \quad \forall \beta \in (0, T).
\end{aligned} \tag{35}$$

Therefore, if we pick $D_S = \sup_{x \in S} \|x - x_0\|_2$, it follows that

$$p(x, T; x_0, y) \geq \exp(-J_{\uparrow}(D_S; T)), \quad \forall x \in S,$$

which ends the proof. \square

Acknowledgements

We gratefully acknowledge support from the NSF grants 1915967, 1820942, and 1838676, as well as from AFOSR.

References

- [1] AIT-SAHALIA, Y. (2008). Closed-form likelihood expansions for multivariate diffusions. *Ann. Statist.* **36**, 906–937.
- [2] ARONSON, D. (1967). Bounds for the fundamental solution of a parabolic equation. *Bull. Amer. Math. Soc.* **73**, 890–896.
- [3] BALLY, V. (2006). Lower bounds for the density of locally elliptic Itô processes. *Ann. Probab.* **34**, 2406–2440.
- [4] BESKOS, A., PAPASPILIOPOULOS, O. AND ROBERTS, G. O. (2006). Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli* **12**, 1077–1098.
- [5] BESKOS, A. AND ROBERTS, G. O. (2005). Exact simulation of diffusions. *Ann. Appl. Probab.* **15**, 2422–2444.
- [6] BLANCHET, J., CHEN, X. AND DONG, J. (2017). ϵ -strong simulation for multidimensional stochastic differential equations via rough path analysis. *Ann. Appl. Probab.* **27**, 275–336.
- [7] CHEN, N. AND HUANG, Z. (2013). Localization and exact simulation of Brownian motion-driven stochastic differential equations. *Math. Operat. Res.* **38**, 591–616.
- [8] FEARNHEAD, P., PAPASPILIOPOULOS, O., ROBERTS, G. O. AND STUART, A. (2010). Random-weight particle filtering of continuous time processes. *J. R. Statist. Soc. B [Statist. Methodology]* **72**, 497–512.
- [9] FLEMING, W. H. AND RISHEL, R. W. (2012). *Deterministic and Stochastic Optimal Control*. Springer Science & Business Media, New York.
- [10] FRIEDMAN, A. (2013). *Partial Differential Equations of Parabolic Type*. Courier Corporation, Chelmsford, MA.
- [11] GILES, M. B. (2008). Multilevel Monte Carlo path simulation. *Operat. Res.* **56**, 607–617.
- [12] GILES, M. B. AND SZPRUCH, L. (2014). Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation. *Ann. Appl. Probab.* **24**, 1585–1620.
- [13] HENRY-LABORDÈRE, P., TAN, X., TOUZI, N. (2017). Unbiased simulation of stochastic differential equations. *Ann. Appl. Probab.* **27**, 3305–3341.
- [14] HUBER, M. (2016). Nearly optimal Bernoulli factories for linear functions. *Combinatorics Prob. Comput.* **25**, 577–591.
- [15] KARATZAS, I. AND SHREVE, S. (2012). *Brownian Motion and Stochastic Calculus*. Springer Science & Business Media, New York.
- [16] KUSUOKA, S. AND STROOCK, D. (1987). Applications of the Malliavin calculus, part 3. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **34**, 397–442.
- [17] ŁATUSZYŃKI, K., KOSMIDIS, I., PAPASPILIOPOULOS, O. AND ROBERTS, G. O. (2011). Simulating events of unknown probabilities via reverse time martingales. *Random Structures Algorithms* **38**, 441–452.
- [18] MCLEISH, D. (2011). A general method for debiasing a Monte Carlo estimator. *Monte Carlo Meth. Appl.* **17**, 301–315.
- [19] NACU, Ş. AND PERES, Y. (2005). Fast simulation of new coins from old. *Ann. Appl. Probab.* **15**, 93–115.
- [20] REVUZ, D. AND YOR, M. (2013). *Continuous Martingales and Brownian Motion*. Springer Science & Business Media, New York.

- [21] RHEE, C.-H. AND GLYNN, P. W. (2012). A new approach to unbiased estimation for SDE's. In *WSC '12: Proceedings of the Winter Simulation Conference*, Berlin, pp. 1–7.
- [22] RHEE, C.-H. AND GLYNN, P. W. (2015). Unbiased estimation with square root convergence for SDE models. *Operat. Res.* **63**, 1026–1043.
- [23] SHEU, S.-J. (1991). Some estimates of the transition density of a nondegenerate diffusion Markov process. *Ann. Prob.* **19**, 538–561.