The Annals of Applied Probability
2014, Vol. 24, No. 5, 1994–2032
DOI: 10.1214/13-AAP969

SIMULATION OF FORWARD-REVERSE STOCHASTIC REPRESENTATIONS FOR CONDITIONAL DIFFUSIONS

BY CHRISTIAN BAYER AND JOHN SCHOENMAKERS[1]

WIAS Berlin

In this paper we derive stochastic representations for the finite dimensional distributions of a multidimensional diffusion on a fixed time interval, conditioned on the terminal state. The conditioning can be with respect to a fixed point or more generally with respect to some subset. The representations rely on a reverse process connected with the given (forward) diffusion as introduced in Milstein, Schoenmakers and Spokoiny [*Bernoulli* **10** (2004) 281–312] in the context of a forward-reverse transition density estimator. The corresponding Monte Carlo estimators have essentially root-N accuracy, and hence they do not suffer from the curse of dimensionality. We provide a detailed convergence analysis and give a numerical example involving the realized variance in a stochastic volatility asset model conditioned on a fixed terminal value of the asset.

1. Introduction. The central result in this paper is the development of a new generic procedure for simulation of conditioned diffusions, also called diffusion bridges or pinned diffusions. More specifically, for some given (unconditional) diffusion process X, we aim to simulate the functional

$$\mathbb{E}\big[g\big(X(s_1),\dots,X(s_R)\big)|X(T)\in A, X(0)=x\big], \tag{1}$$

where $0\le s_1<s_2<\cdots<s_R<T$, A is some set that may consist of only one point, g is an arbitrarily given suitable test function and $x\in\mathbb{R}^d$ is a given state. In recent years, the problem of computing terms such as (1) has attracted a lot of attention in the literature, sparked by several applications. Indeed, many relevant properties of a diffusion process X can be advantageously analyzed by considering the process conditioned on certain appropriate events. One so allows "to study rare events by conditioning on the event happening or to analyze the behaviour of a composite system when only some of its components can be observed," as is eloquently put by Hairer, Stuart and Voss (2009). For instance, in statistical inference based on a continuous time model, discrete time observations can be enriched to continuous time observations by sampling from the diffusion bridges between

Received February 2013; revised July 2013.

[1]Supported in part by the DFG Research Center MATHEON "Mathematics for Key Technologies" in Berlin.

MSC2010 subject classifications. Primary 65C05; secondary 65C30.

Key words and phrases. Forward-reverse representations, pinned or conditional diffusions, Monte Carlo simulation.

the discrete time data; see Lin, Chen and Mykland (2010) and Bladt and Sørensen (2012) for more information. Conditional diffusions have further been successfully used for critical calculations in rare event situations. As an example from computational chemistry, we refer to the review paper of Bolhuis et al. (2002), where diffusion bridges are used for detection of the transition state surface between two stable regions A and B in configuration space. Here, standard Monte Carlo simulation is prohibitively costly, as the event of such a transition is rare, provided that the "walls" in the energy surface between A and B are high. However, by studying the process conditioned on starting in A and ending in B, one can efficiently observe on which paths the configuration typically travels from A to B. Other possible applications appear in the field of stochastic environmental models, for instance, regarding the concentration evolutions of pollution in water; for example see Spivakovskaya, Heemink and Schoenmakers (2007) and references therein for a related problem.

Several approaches for simulation of diffusion bridges have already been studied in the literature. For the theory of diffusion bridges we refer to Lyons and Zheng (1990) and the references therein. Many existing approaches utilize known Radon–Nikodym densities of the law of the diffusion X conditioned on initial and terminal values, with respect to the law of a standard diffusion bridge process (e.g., Wiener bridge) on path-space [as a Radon–Nikodym derivative obtained by Doob's h-transform; see, e.g., Rogers and Williams (2000) or Lyons and Zheng (1990)]. Several other approaches are based on (partial) knowledge of the transition densities of the unconditional diffusion (that is not generically available, of course). For an overview of many different techniques, we refer to Lin, Chen and Mykland (2010).

First, let us mention the work by Beskos, Papaspiliopoulos and Roberts (2006) who construct a general, rejection-based algorithm for solutions of *one-dimensional* SDEs, based on the Radon–Nikodym derivative of the law of the solution with respect to the Wiener measure. The algorithm gives (in finite, but random time) discrete samples of the exact solution of the SDE. A simple adaption of this algorithm gives samples of the exact diffusion process conditioned on $X(T) = y$, by using the law of the corresponding Brownian bridge as reference measure (instead of the Wiener measure). An overview of related importance sampling techniques is given by Papaspiliopoulos and Roberts (2012). On the other hand, by relying on knowledge of the transition densities of X, Lin, Chen and Mykland (2010) use a sequential weighted Monte Carlo framework, including resampling with optimal priority scores.

Another general technique used for simulation of diffusion bridges is the Markov chain Monte Carlo method. Indeed, Stuart, Voss and Wiberg (2004) and Hairer, Stuart and Voss (2009) show how the law of a (multi-dimensional, uniformly elliptic, additive-noise) diffusion X conditioned on $X(T) = y$ can be regarded as the invariant distribution of a stochastic differential equation of Langevin type on path-space, that is, of a Langevin-type stochastic partial differential equation (SPDE). Thus, in principle MCMC methods are applicable as explored by

Stuart, Voss and Wiberg (2004) and Beskos et al. (2008). However, this requires the numerical solution of the SPDE involved. It should be noted that in Hairer, Stuart and Voss (2011) the uniform ellipticity condition is relaxed leading to a fourth order parabolic SPDE rather than a second order one.

Other notable approaches include those of Milstein and Tretyakov (2004), which treat the case of physically relevant functionals of Wiener integrals with respect to Brownian bridges, and Stinis (2011), who uses an MCMC approach based on successive modifications of the drift of the diffusion process.

Another approach is the one of Bladt and Sørensen (2012) developed for one-dimensional diffusions. In order to obtain a sample from the process X conditioned on $X(0) = x$ and $X(T) = y$, Bladt and Sørensen (2012) start a path of the diffusion from $(0, x)$ and another path of the diffusion in *reversed time* at (T, y). If these paths hit at time τ, consider the concatenated path Z. The distribution of the process Z (conditional on $0 \le \tau \le T$) equals the distribution of the bridge conditional on being hit by an independent path of the underlying diffusion with initial distribution $p(0, y, T, \cdot)$. As proved by Bladt and Sørensen (2012), the probability of this event approaches 1 when $T \to \infty$. Finally, in order to improve the accuracy, Z is used as initial value of an MCMC algorithm on path space, converging to a sample from the true diffusion bridge.

A more general approach is given by Delyon and Hu (2006) which relies on the explicit Radon–Nikodym derivative of the diffusion X conditioned on its initial and terminal values and another diffusion Y, which is modeled like the Brownian bridge. In fact, Y has the same dynamics as X, except for an extra term $-\frac{Y(t)-y}{T-t}$ in the drift, which enforces $Y(T) = y$. Under certain regularity conditions—in particular invertibility of the diffusion matrix $\sigma = \sigma(t, x)$—Delyon and Hu (2006) provide a Girsanov-type theorem, which leads to a representation of the form

$$\mathbb{E}\big[g(X)|X(0) = x, X(T) = y\big] = \mathbb{E}\big[g(Y)Z(Y)\big]$$

for functionals g defined on path-space and a factor $Z(Y)$ explicitly given as a functional of the path Y together with quadratic variations of functions of Y. As such this approach allows for direct Monte Carlo simulation of (1). However, we stress that $Z(Y)$ explicitly depends on σ^{-1} which does not exist in many hypoelliptic applications. On the other hand, simulation of the bridge-type process Y is numerically troublesome because of the exploding drift term.

The new method presented in this article is inspired by the forward-reverse estimator for the transition density $p(0, x, T, y)$ constructed by Milstein, Schoenmakers and Spokoiny (2004). Given a grid $0 \le s_0 < s_1 < \cdots < s_K = t^* < t_1 < \cdots < t_L = T$, we prove that

$$\mathbb{E}\big[g\big(X(s_1), \ldots, X(s_K), X(t_1), \ldots, X(t_{L-1})\big)|X(s_0) = x, X(T) = y\big]$$

equals

$$\lim_{\varepsilon \to 0} \frac{\mathbb{E}[g(X(s_1), \ldots, X(s_K), Y(\widehat{t_{L-1}}), \ldots, Y(\widehat{t_1}))K_\varepsilon(Y(\widehat{t_L}) - X(t^*))\mathcal{Y}(\widehat{t_L})]}{\mathbb{E}[K_\varepsilon(Y(\widehat{t_L}) - X(t^*))\mathcal{Y}(\widehat{t_L})]}, \tag{2}$$

which can be implemented by Monte Carlo simulation for any $\varepsilon > 0$. In (2) $s_0 < t^* < T$ is a given grid-point chosen by the user. The process X solves the original SDE with initial value $X(s_0) = x$ on the time-interval $[s_0, t^*]$. On the other hand, Y is an (independent) *reverse process* as defined in Section 3 started at $Y(0) = y$ and simulated until time $\widehat{t}_L := T - t^*$, not on the original grid, but on a "perturbed" grid defined in (14). [Note that Y is different from the time-reversed diffusion in the sense of Haussmann and Pardoux (1986) that explicitly requires the transition density of X.] Indeed, the dynamics of Y are explicitly given below in terms of the dynamics of X—not relying on the transition density—and, usually, share the same regularity properties; see (4) and (5). Next, we weight the trajectories according to the distance between $X(t^*)$ and $Y(T)$ using a kernel K with bandwidth ε. Finally, we have an exponential weighting factor $\mathcal{Y}$, similar to the Radon–Nikodym derivative in Delyon and Hu (2006). The denominator in (2) actually corresponds to the forward-reverse estimator for the transition density $p(s_0, x, T, y)$ of X introduced by Milstein, Schoenmakers and Spokoiny (2004). The details of the Monte Carlo simulation are spelled out in Section 4, but we note that (2) can be computed to an accuracy of ε with a complexity of $\mathcal{O}(\varepsilon^{-2})$ in any dimension less or equal to four[2] (disregarding possible discretization errors due to the construction of samples X, Y and $\mathcal{Y}$).[3] Thus our algorithm essentially achieves the optimal rate of convergence for Monte Carlo algorithms.

We underline that the forward-reverse algorithm for (1) presented here is not a straightforward extension of the forward-reverse algorithm for transition densities of Milstein, Schoenmakers and Spokoiny (2004). The main difficulty lies in the extension of the representation from just one intermediate time $0 < t^* < T$ to an arbitrary time grid $0 \le s_0 < s_1 < \cdots < s_K = t^* < t_1 < \cdots < t_L = T$ with $L > 1$. In the nonautonomous case this issue is further complicated due to the fact that the dynamics of the reverse process as defined in Milstein, Schoenmakers and Spokoiny (2004) depends explicitly on both t^* and T. Obviously, the different structure also requires a different error analysis. In particular, we need sharper error bounds than Milstein, Schoenmakers and Spokoiny (2004).

In comparison to the other methods mentioned above, our new procedure has the following main features:

(i) The method applies to multidimensional diffusions.

(ii) It is based on simulation of *unconditional* diffusions only, hence technical simulation problems due to exploding drifts in SDEs that govern particular diffusion bridges are avoided.

(iii) The vector fields determining the (forward) SDE that governs X only need to satisfy a Hörmander-type condition guaranteeing sufficient regularity and ex-

[2] In fact, this restriction can be lifted by use of higher order kernels.

[3] The constant will increase in the dimension. Moreover, we ignore the cost of checking equality of two integers.

ponential decay of the transition densities. In particular, the diffusion matrix of X may be degenerate.

(iv) The estimator corresponding to the developed stochastic representation for (1) is root-N consistent, that is the mean square estimation accuracy is of order $\mathcal{O}(N^{-1/2})$ with N being the number of trajectories that need to be simulated.

As a matter of fact, the methods for simulating diffusion bridges known in the literature so far, do not cover all the features (i)–(iv) simultaneously. For example, Delyon and Hu (2006) require that either the diffusion matrix is invertible, or impose some very specific structural conditions on the drift and diffusion matrix of the process X. Moreover, the exploding drift terms in their process Y makes simulation of the auxiliary process Y nontrivial. On the other hand, the method of Bladt and Sørensen (2012) in germ carries some ideas related to our approach, but they need to impose balance restrictions on the transition density of X, and moreover their method—together with several others—is only one dimensional. The methods of Stuart, Voss and Wiberg (2004) and the related papers mentioned above also involve some further structural assumptions and, in addition, require numerical solutions of SPDEs.

Moreover, we complement our algorithm by an adaptation, which allows us to treat the more general problem of conditioning at final time T not on all, but just on some components of the vector X_T. More precisely, we present a variant of the algorithm for computing conditional expectations where X_T is conditioned to lie in a "simple" set A, that is, either A has positive measure both under the Lebesgue measure and the distribution of X_T, or A is an affine plane of dimension $0 \le d' \le d$. In order to achieve this extension, we need to prove (Lebesgue) integrable error bounds for the forward-reverse algorithm for the case where X_T is conditioned to a value y.

The structure of the paper is as follows. In Section 2 we recap the essential facts concerning the reverse diffusion system of Milstein, Schoenmakers and Spokoiny (2004). The main representation theorems for the diffusion conditioned on reaching a fixed state, or conditioned on reaching some Borel set, are derived in Section 3. A detailed accuracy analysis concerning the Monte Carlo estimators for the respective conditioned diffusions is provided in Section 4, including the precise required regularity assumptions given in Conditions 4.1, 4.4 and 4.5. Limitations of the method are discussed in Section 4.3, while Section 5 provides a numerical study involving a Heston-type stochastic volatility model.

2. Recap of forward-reverse representations for diffusions. In this section we recapitulate shortly the main ingredients in the approach by Milstein, Schoenmakers and Spokoiny (2004). Let us consider the SDE

$$dX_{t,x}(s) = a\big(s, X_{t,x}(s)\big)\,ds + \sigma\big(s, X_{t,x}(s)\big)\,dW(s), \qquad 0 \le s \le T,\ X_{t,x}(t) = x, \tag{3}$$

where $X_{t,x} \in \mathbb{R}^d$, $a:[t,T]\times\mathbb{R}^d \to \mathbb{R}^d$, $\sigma:[t,T]\times\mathbb{R}^d \to \mathbb{R}^{d\times m}$, W is an m-dimensional standard Wiener process and $x \in \mathbb{R}^d$. At this stage, we only assume that X admits a C^2 transition density p and that the coefficients of (3) are C^2 as well.

Along with the (forward) process X given by (3), Milstein, Schoenmakers and Spokoiny (2004) introduced an associated process $(Y_{t,y;T}(s), \mathcal{Y}_{t,y;T}(s))$ in $\mathbb{R}^d \times \mathbb{R}$, $t \le s \le T$, termed *reverse* process on the interval $[t,T]$, that solves the SDE

$$\begin{aligned} dY_{t,y;T}(s) &= \alpha_{t,T}\big(s, Y_{t,y;T}(s)\big)\,ds \\ &\quad + \widetilde{\sigma}_{t,T}\big(s, Y_{t,y;T}(s)\big)\,d\widetilde{W}(s), \qquad Y_{t,y;T}(t) = y, \\ \mathcal{Y}_{t,y;T}(s) &= \exp\Big(\int_t^s c_{t,T}\big(u, Y_{t,y;T}(u)\big)\,du\Big) \end{aligned} \tag{4}$$

with $\widetilde{W}$ being a (from W independent) m-dimensional Wiener process, and

$$\begin{aligned} \alpha^i_{t,T}(s,y) &:= \sum_{j=1}^d \frac{\partial}{\partial y^j} b^{ij}(T+t-s,y) - a^i(T+t-s,y), \qquad b := \sigma\sigma^\top, \\ \widetilde{\sigma}_{t,T}(s,y) &:= \sigma(T+t-s,y), \\ c_{t,T}(s,y) &:= \frac{1}{2}\sum_{i,j=1}^d \frac{\partial^2 b^{ij}}{\partial y^i\,\partial y^j}(T+t-s,y) - \sum_{i=1}^d \frac{\partial a^i}{\partial y^i}(T+t-s,y). \end{aligned} \tag{5}$$

Despite its name, we stress that $(Y,\mathcal{Y})$ is the solution of an ordinary SDE *forward* in time on the interval $[t,T]$.

One of the central results in Milstein, Schoenmakers and Spokoiny (2004) is the following theorem.

THEOREM 2.1 [M.S.S. (2004)]. *For fixed t, x, y and $t < t^* < T$, and any bivariate test function f we have*

$$\begin{aligned} &\mathbb{E}\big[f\big(X_{t,x}(t^*), Y_{t^*,y;T}(T)\big)\mathcal{Y}_{t^*,y;T}(T)\big] \\ &\quad = \iint p(t,x,t^*,x')\,p(t^*,y',T,y)\,f(x',y')\,dx'\,dy', \end{aligned} \tag{6}$$

where $X_{t,x}(s)$ satisfies the forward equation (3), and $(Y_{t^*,y;T}(s), \mathcal{Y}_{t^*,y;T}(s))$, $s \ge t^*$, is the solution of the reverse system (4).

COROLLARY 2.2. *By taking $f \equiv 1$, (6) yields*

$$\mathbb{E}[\mathcal{Y}_{t^*,y;T}(T)] = \int p(t^*,y',T,y)\,dy', \tag{7}$$

which obviously extends to $t^* = t$. By next taking $f(x',y') = f(x')$ (while abusing notation slightly) we obtain from (6), using (7) and the independence of X and

$(Y, \mathcal{Y})$,

$$\mathbb{E}\big[f\big(X_{t,x}(t^*)\big)\big] = \int p(t, x, t^*, x')\, f(x')\, dx',$$

which obviously extends to $t^* = T$, *that is, the standard forward stochastic representation for* $\int p(t, x, T, x')\, f(x')\, dx'$. *On the other hand, by taking* $f(x', y') = f(y')$ *we obtain the so called reverse stochastic representation*

$$\mathbb{E}\big[f\big(Y_{t^*,y;T}(T)\big)\mathcal{Y}_{t^*,y;T}(T)\big] = \int p(t^*, y', T, y)\, f(y')\, dy', \tag{8}$$

which obviously extends to $t^* = t$.

3. Forward-reverse representations for conditional diffusions. It should be noted that in Milstein, Schoenmakers and Spokoiny (2004) the time domain of the reverse process was considered fixed. For our purposes however, it turns out to be more effective (in particular regarding the proof of Theorem 3.2 below) to consider reverse processes suitably defined on different time domains. In particular it turns out be fruitful to formulate the forward-reverse representations of the previous section in terms of reverse processes defined on $[0, T]$ for suitable $T > 0$. We therefore introduce the reverse process

$$\big(Y_{y;T}(s), \mathcal{Y}_{y;T}(s)\big)_{0\le s\le T} := \big(Y_{0,y;T}(s), \mathcal{Y}_{0,y;T}(s)\big)_{0\le s\le T} \tag{9}$$

that starts at time $s = 0$ at a generic state $(y, 1)$, is defined on an interval $[0, T]$ and satisfies (4) with coefficients (5) for $t = 0$, that is, (9) solves the SDE

$$\begin{aligned} dY(s) &= \alpha_{0,T}\big(s, Y(s)\big)\, ds + \widetilde{\sigma}_{0,T}\big(s, Y(s)\big)\, d\widetilde{W}(s), \qquad Y(0) = y, \\ \mathcal{Y}(s) &= \exp\Big(\int_0^s c_{0,T}\big(u, Y(u)\big)\, du\Big). \end{aligned} \tag{10}$$

As a result, we have for any fixed t, $0 \le t \le T$, that

$$\big(Y_{y;T}(s), \mathcal{Y}_{y;T}(s)\big)_{0\le s\le T-t} = \big(Y_{t,y;T}(t+s), \mathcal{Y}_{t,y;T}(t+s)\big)_{0\le s\le T-t},$$

whence (6) and (8) may be equivalently written as

$$\begin{aligned} &\mathbb{E}\big[f\big(X_{t,x}(t^*), Y_{y;T}(T-t^*)\big)\mathcal{Y}_{y;T}(T-t^*)\big] \\ &\qquad = \iint p(t, x, t^*, x')\, p(t^*, y', T, y)\, f(x', y')\, dx'\, dy' \end{aligned} \tag{11}$$

and

$$\mathbb{E}\big[f\big(Y_{y;T}(T-t^*)\big)\mathcal{Y}_{y;T}(T-t^*)\big] = \int p(t^*, y', T, y)\, f(y')\, dy', \tag{12}$$

respectively. The main benefit is that the reverse process used in representations (6) and (8) depend on both t^* and T, while the one used in (11) and (12) depends on

T only. In particular, (12) may be considered as a reverse representation for all $0 < t^* < T$ in terms of one and the same reverse process $(Y_{y;T}, \mathcal{Y}_{y;T})$.

3.1. *Representations for conditioning on a fixed state.* Let us start with the following lemma.

LEMMA 3.1. *For any $0 < s < t \le T$ it holds that*

$$Y_{Y_{y;T}(s);T-s}(t-s) = Y_{y;T}(t),$$
$$\mathcal{Y}_{y;T}(t) = \mathcal{Y}_{y;T}(s)\mathcal{Y}_{Y_{y;T}(s);T-s}(t-s).$$

PROOF. The first statement is directly obvious from (10). From this the second statement follows by

$$\begin{aligned}\mathcal{Y}_{y;T}(t) &= e^{\int_0^s c_{0,T}(u,Y_{y;T}(u))\,du} e^{\int_s^t c_{0,T}(u,Y_{Y_{y;T}(s);T-s}(u-s))\,du} \\ &= \mathcal{Y}_{y;T}(s) e^{\int_0^{t-s} c_{0,T-s}(u,Y_{Y_{y;T}(s);T-s}(u))\,du} \\ &= \mathcal{Y}_{y;T}(s)\mathcal{Y}_{Y_{y;T}(s);T-s}(t-s).\end{aligned}$$

□

We are now ready to state the following key theorem.

THEOREM 3.2. *Given a grid $\mathcal{D}_L := \{0 \le t^* < t_1 < \cdots < t_L\}$, it holds that*

$$\mathbb{E}\big[f\big(Y_{y;t_L}(t_L - t_0), Y_{y;t_L}(t_L - t_1), \ldots, Y_{y;t_L}(t_L - t_{L-1})\big)\mathcal{Y}_{y;t_L}(t_L - t_0)\big]$$
$$= \int_{\mathbb{R}^{d\times L}} f(y_0, y_1, \ldots, y_{L-1}) \prod_{i=1}^{L} p(t_{i-1}, y_{i-1}, t_i, y_i)\,dy_{i-1}$$

with $y_L := y$ and $t_0 := t^*$.

PROOF. The following proof—much shorter than the proof in a previous version of the paper—has essentially been pointed out to us by an anonymous referee. We fix t_0 $(= t^*)$ and use induction on L. For $L = 1$ the statement boils down to (12) with $T := t_1$. Suppose the statement is proved for some $L \ge 1$. For the grid

$$\mathcal{D}_{L+1} = \{t_0 < t_1 < \cdots < t_{L+1}\}$$

we next consider for any test function $f : \mathbb{R}^{d\times(L+1)} \to \mathbb{R}$,

$$\begin{aligned}&\mathbb{E}\big[f\big(Y_{y;t_{L+1}}(t_{L+1} - t_0), Y_{y;t_{L+1}}(t_{L+1} - t_1), \ldots, Y_{y;t_{L+1}}(t_{L+1} - t_L)\big) \\ &\qquad\qquad \times \mathcal{Y}_{y;t_{L+1}}(t_{L+1} - t_0)\big] \\ &= \mathbb{E}\big[\mathbb{E}\big[f\big(Y_{Y_{y;t_{L+1}}(t_{L+1}-t_L);t_L}(t_L - t_0), Y_{Y_{y;t_{L+1}}(t_{L+1}-t_L);t_L}(t_L - t_1), \ldots, \\ &\qquad\qquad Y_{Y_{y;t_{L+1}}(t_{L+1}-t_L);t_L}(t_L - t_{L-1}), Y_{Y_{y;t_{L+1}}(t_{L+1}-t_L);t_L}(0)\big)\end{aligned}$$

$$\begin{aligned}
&\times \mathcal{Y}_{y;t_{L+1}}(t_{L+1}-t_L)\mathcal{Y}_{Y_{y;t_{L+1}}(t_{L+1}-t_L);t_L}(t_L-t_0)| \\
&\qquad Y_{y;t_{L+1}}(t_{L+1}-t_L), \mathcal{Y}_{y;t_{L+1}}(t_{L+1}-t_L)]] \\
&= \mathbb{E}[\mathcal{Y}_{y;t_{L+1}}(t_{L+1}-t_L) \\
&\quad \times \mathbb{E}[f\big(Y_{Y_{y;t_{L+1}}(t_{L+1}-t_L);t_L}(t_L-t_0), Y_{Y_{y;t_{L+1}}(t_{L+1}-t_L);t_L}(t_L-t_1), \dots, \\
&\qquad Y_{Y_{y;t_{L+1}}(t_{L+1}-t_L);t_L}(t_L-t_{L-1}), Y_{y;t_{L+1}}(t_{L+1}-t_L)\big) \\
&\qquad \times \mathcal{Y}_{Y_{y;t_{L+1}}(t_{L+1}-t_L);t_L}(t_L-t_0)|Y_{y;t_{L+1}}(t_{L+1}-t_L)]].
\end{aligned}$$

By the induction hypothesis, we have that

$$\begin{aligned}
&\mathbb{E}[f\big(Y_{Y_{y;t_{L+1}}(t_{L+1}-t_L);t_L}(t_L-t_0), Y_{Y_{y;t_{L+1}}(t_{L+1}-t_L);t_L}(t_L-t_1), \dots, \\
&\qquad Y_{Y_{y;t_{L+1}}(t_{L+1}-t_L);t_L}(t_L-t_{L-1}), Y_{y;t_{L+1}}(t_{L+1}-t_L)\big) \\
&\qquad \times \mathcal{Y}_{Y_{y;t_{L+1}}(t_{L+1}-t_L);t_L}(t_L-t_0)|Y_{y;t_{L+1}}(t_{L+1}-t_L)=z] \\
&= \mathbb{E}[f\big(Y_{z;t_L}(t_L-t_0), Y_{z;t_L}(t_L-t_1), \dots, Y_{z;t_L}(t_L-t_{L-1}), z\big)\mathcal{Y}_{z;t_L}(t_L-t_0)] \\
&= \int_{\mathbb{R}^{d\times L}} f(y_0, y_1, \dots, y_{L-1}, z) \prod_{i=1}^{L} p(t_{i-1}, y_{i-1}, t_i, y_i)\, dy_{i-1} =: F(z)
\end{aligned}$$

with $y_L := z$, and so we obtain

$$\begin{aligned}
&\mathbb{E}[f\big(Y_{y;t_{L+1}}(t_{L+1}-t_0), Y_{y;t_{L+1}}(t_{L+1}-t_1), \dots, Y_{y;t_{L+1}}(t_{L+1}-t_L)\big) \\
&\qquad \times \mathcal{Y}_{y;t_{L+1}}(t_{L+1}-t_0)] \\
&\quad = \mathbb{E}[\mathcal{Y}_{y;t_{L+1}}(t_{L+1}-t_L)F\big(Y_{y;t_{L+1}}(t_{L+1}-t_L)\big)] \\
&\quad \overset{(12)}{=} \int p(t_L, z, t_{L+1}, y)F(z)\, dz \\
&\quad = \int_{\mathbb{R}^{d\times(L+1)}} f(y_0, y_1, \dots, y_L) \prod_{i=1}^{L+1} p(t_{i-1}, y_{i-1}, t_i, y_i)\, dy_{i-1},
\end{aligned}$$

where $y_{L+1} := y$ and the integration variable z is renamed to y_L. □

For the next theorem, we consider an extended time grid

$$0 \le s_0 < s_1 < \cdots < s_K = t^* = t_0 < t_1 < \cdots < t_L = T. \tag{13}$$

For convenience, we also introduce the notation

$$\widehat{t_i} := t_L - t_{L-i}, \qquad i = 1, \dots, L. \tag{14}$$

Moreover, we fix a starting point $x \in \mathbb{R}^d$.

THEOREM 3.3. *For any* $f:\mathbb{R}^{d\times(K+L)}\to\mathbb{R}$ *and grids* (13) *together with* (14), *we have*

$$\begin{aligned}\mathbb{E}\big[f\big(X_{s_0,x}(s_1),\dots,X_{s_0,x}(s_K),Y_{y;T}(\widehat{t}_L),Y_{y;T}(\widehat{t}_{L-1}),\dots,Y_{y;T}(\widehat{t}_1)\big)\mathcal{Y}_{y;T}(\widehat{t}_L)\big]\\ =\int_{\mathbb{R}^{d\times(K+L)}} f(x_1,\dots,x_K,y_0,y_1,\dots,y_{L-1})\\ \times\prod_{i=1}^{K}p(s_{i-1},x_{i-1},s_i,x_i)\,dx_i\prod_{i=1}^{L}p(t_{i-1},y_{i-1},t_i,y_i)\,dy_{i-1}\end{aligned}$$

with $x_0:=x$, $y_L:=y$, *and the processes* X *and* $(Y,\mathcal{Y})$ *being independent*.

Theorem 3.3 follows directly from Theorem 3.2 by a standard conditioning argument and the Chapman–Kolmogorov equation. Note that for $K=L=1$, Theorem 3.3 collapses to Theorem 2.1.

We are now ready to derive a forward-reverse stochastic representation for the finite dimensional distributions of the process $X_{s_0,x}$, conditional on $X_{s_0,x}(T)=y$, for fixed $s_0<T$, and fixed $x,y\in\mathbb{R}^d$. To this end we henceforth assume that

$$p(s_0,x,T,y)>0. \tag{15}$$

We also need to assume continuity of p. Let us take a bounded measurable test function

$$g(x_1,\dots,x_K,y_1,\dots,y_{L-1}):\mathbb{R}^{d\times(K+L-1)}\to\mathbb{R},$$

and consider the conditional expectation

$$\begin{aligned}\mathbb{E}\big[g\big(X_{s_0,x}(s_1),\dots,X_{s_0,x}(s_{K-1}),X_{s_0,x}(t^*),X_{s_0,x}(t_1),\dots,X_{s_0,x}(t_{L-1})\big)\big|\\ X_{s_0,x}(T)=y\big].\end{aligned} \tag{16}$$

The distribution of the diffusion $X_{s_0,x}$ conditional on $X_{s_0,x}(T)=y$ is completely determined by the totality of conditional expectations of the form (16). These conditional expectations may be obtained due to Theorem 3.4 below.

THEOREM 3.4. *Consider the forward process* X *and its reverse process* $(Y,\mathcal{Y})$ *as before and the grids as specified in* (13) *and* (14). *Let*

$$K_\varepsilon(u):=\varepsilon^{-d}K(u/\varepsilon),\qquad y\in\mathbb{R}^d,$$

with K *being integrable on* $\mathbb{R}^d$ *and* $\int_{\mathbb{R}^d}K(u)\,du=1$. *Hence, formally* K_ε *converges to the delta function* δ_0 *on* $\mathbb{R}^d$ (*in distribution sense*) *as* $\varepsilon\downarrow 0$. *Then, since* $p(s_0,x,T,y)>0$ *by assumption, for any bounded measurable function*

$g:\mathbb{R}^{d\times(K+L-1)}\to\mathbb{R}$, *we have*

$$\begin{aligned}
&\mathbb{E}\big[g\big(X_{s_0,x}(s_1),\dots,X_{s_0,x}(t^*),X_{s_0,x}(t_1),\dots,X_{s_0,x}(t_{L-1})\big)|X_{s_0,x}(T)=y\big]\\
&\quad=\frac{1}{p(s_0,x,T,y)}\\
&\quad\quad\times\lim_{\varepsilon\downarrow 0}\mathbb{E}\big[g\big(X_{s_0,x}(s_1),\dots,X_{s_0,x}(t^*),Y_{y;T}(\widehat{t}_{L-1}),\dots,Y_{y;T}(\widehat{t}_1)\big)\\
&\quad\quad\quad\times K_\varepsilon\big(Y_{y;T}(\widehat{t}_L)-X_{s_0,x}(t^*)\big)\mathcal{Y}_{y;T}(\widehat{t}_L)\big].
\end{aligned}\tag{17}$$

PROOF. By applying Theorem 3.3 to

$$f(x_1,\dots,x_K,y_0,y_1,\dots,y_{L-1}):=g(x_1,\dots,x_K,y_1,\dots,y_{L-1})K_\varepsilon(y_0-x_K),$$

we obtain

$$\begin{aligned}
&\mathbb{E}\big[g\big(X_{s_0,x}(s_1),\dots,X_{s_0,x}(t^*),Y_{y;T}(\widehat{t}_{L-1}),\dots,Y_{y;T}(\widehat{t}_1)\big)\\
&\qquad\qquad\times K_\varepsilon\big(Y_{y;T}(\widehat{t}_L)-X_{s_0,x}(t^*)\big)\mathcal{Y}_{y;T}(\widehat{t}_L)\big]\\
&\quad=\int_{\mathbb{R}^{d\times(K+L)}}g(x_1,\dots,x_K,y_1,\dots,y_{L-1})K_\varepsilon(y_1-x_K)\\
&\qquad\qquad\times\prod_{i=1}^{K}p(s_{i-1},x_{i-1},s_i,x_i)\,dx_i\prod_{i=1}^{L}p(t_{i-1},y_{i-1},t_i,y_i)\,dy_{i-1}\\
&\quad=\int_{\mathbb{R}^{d\times(K+L)}}g(x_1,\dots,x_K,y_1,\dots,y_{L-1})K(v)\,dv\,p\big(t^*,x_k+\varepsilon v,t_1,y_2\big)\\
&\qquad\qquad\times\prod_{i=1}^{K}p(s_{i-1},x_{i-1},s_i,x_i)\,dx_i\prod_{i=2}^{L}p(t_{i-1},y_{i-1},t_l,y_i)\,dy_{i-1}.
\end{aligned}\tag{18}$$

By sending ε to zero, (18) clearly converges to

$$\begin{aligned}
&\int_{\mathbb{R}^{d\times(K+L-1)}}g(x_1,\dots,x_K,y_1,\dots,y_{L-1})\prod_{i=1}^{K}p(s_{i-1},x_{i-1},s_i,x_i)\,dx_i\\
&\qquad\times p\big(t^*,x_K,t_1,y_1\big)\prod_{i=2}^{L}p(t_{i-1},y_{i-1},t_l,y_i)\,dy_{i-1},
\end{aligned}$$

from which (17) easily follows. □

If the original grid $t^*=t_0<\cdots<t_L=T$ is equidistant, then the transformed grid $0=\widehat{t}_0<\cdots<\widehat{t}_L=T-t^*$ is obtained by a translation with $-t^*$, which leads to the following corollary.

COROLLARY 3.5. *If the time grid $t^* = t_0 < \cdots < t_L = T$ is equidistant, we have*

$$\begin{aligned}\mathbb{E}\big[g\big(X_{s_0,x}(s_1),\dots,X_{s_0,x}(t^*),X_{s_0,x}(t_1),\dots,X_{s_0,x}(t_{L-1})\big)|X_{s_0,x}(T)=y\big]\\ =\frac{1}{p(s_0,x,T,y)}\\ \times\lim_{\varepsilon\downarrow 0}\mathbb{E}\big[g\big(X_{s_0,x}(s_1),\dots,X_{s_0,x}(t^*),Y_{y;T}\big(t_{L-1}-t^*\big),\dots,Y_{y;T}\big(t_1-t^*\big)\big)\\ \times K_\varepsilon\big(Y_{y;T}\big(T-t^*\big)-X_{s_0,x}(t^*)\big)\mathcal{Y}_{y;T}\big(T-t^*\big)\big].\end{aligned}$$

Moreover, by setting $g\equiv 1$, we retrieve the forward-reverse representation of the transition density in Milstein, Schoenmakers and Spokoiny (2004),

$$p(s_0,x,T,y)=\lim_{\varepsilon\downarrow 0}\mathbb{E}\big[K_\varepsilon\big(Y_{y;T}\big(T-t^*\big)-X_{s_0,x}(t^*)\big)\mathcal{Y}_{y;T}\big(T-t^*\big)\big], \tag{19}$$

expressed with the variant of the reverse process introduced in (10) above.

REMARK 3.6. For fixed $x,y\in\mathbb{R}^d$ and $s_0<t^*<T$ as before, let us define a process Z by

$$Z(t):=Y_{y;T}(T-t),\qquad t^*\le t\le T.$$

The idea is that we run along the reverse diffusion Y backwards in time. Then (17) reads

$$\begin{aligned}\mathbb{E}\big[g\big(X_{s_0,x}(s_1),\dots,X_{s_0,x}(t^*),X_{s_0,x}(t_1),\dots,X_{s_0,x}(t_{L-1})\big)|X_{s_0,x}(T)=y\big]\\ =\frac{1}{p(s_0,x,T,y)}\lim_{\varepsilon\downarrow 0}\mathbb{E}\big[g\big(X_{s_0,x}(s_1),\dots,X_{s_0,x}(t^*),Z(t_1),\dots,Z(t_{L-1})\big)\\ \times K_\varepsilon\big(Z\big(t^*\big)-X_{s_0,x}(t^*)\big)\mathcal{Y}_{y;T}\big(T-t^*\big)\big].\end{aligned}$$

3.2. *Representations for conditioning on a set.* Now let us assume that we are interested in the conditional expectation of a functional $g(X_{s_0,x}(s_1),\dots,X_{s_0,x}(t_{L-1}))$ given $X_T\in A$ for some Borel set A. It is assumed for simplicity that either A is a subset of $\mathbb{R}^d$ with positive Lebesgue measure and with $\mathbb{P}(X_{s_0,x}(T)\in A)>0$, or A is an affine hyperplane of dimension d', $0\le d'\le d$. As a further simplification in the latter case, although without further loss of generality, we assume that A is of the form

$$A=\big\{x\in\mathbb{R}^d:x^1=c^1,\dots,x^{d-d'}=c^{d-d'}\big\}. \tag{20}$$

For $0\le d'\le d$ we consider the "restricted" Lebesgue measure

$$\lambda_A(dx)=\delta_{\{c^1\}}\big(dx^1\big)\cdots\delta_{\{c^{d-d'}\}}\big(dx^{d-d'}\big)\cdot dx^{d-d'+1}\cdots dx^d, \tag{21}$$

which coincides with the ordinary Lebesgue measure if $d'=d$, and with a Dirac point measure if $d'=0$. We next introduce a random variable ξ with support in

A independent from X and Y, whose law has a density $\varphi > 0$ with respect to λ_A. Let further $(Y_{t^*,\xi}, \mathcal{Y}_{\xi;T})$ denote the reverse process starting at the random location $(\xi, 1)$ at time t^*. Here, we replace condition (15) on the positivity of the transition density by

$$\int_A p(s_0, x, T, z)\lambda_A(dz) > 0. \tag{22}$$

THEOREM 3.7. *Let the kernel function K be as in Theorem 3.4, and let there be given a time grid of the form (13). The conditional expectation of*

$$g\big(X_{s_0,x}(s_1), \dots, X_{s_0,x}(t^*), X_{s_0,x}(t_1), \dots, X_{s_0,x}(t_{L-1})\big)$$

given $X_{s_0,x}(T) \in A$ with A being a Borel set, either with positive probability or a hyperplane of the form (20), and g being a bounded measurable test function, has the stochastic representation

$$\begin{aligned}\int_A p(s_0, x, T, y)\lambda_A(dy) \cdot \mathbb{E}\big[g\big(X_{s_0,x}(s_1), \dots, X_{s_0,x}(t_{L-1})\big) | X_{s_0,x}(T) \in A\big] \\ = \lim_{\varepsilon \downarrow 0} \mathbb{E}\Big[g\big(X_{s_0,x}(s_1), \dots, X_{s_0,x}(t^*), Y_{\xi;T}(\widehat{t}_{L-1}), \dots, Y_{\xi;T}(\widehat{t}_1)\big) \\ \times K_\varepsilon\big(Y_{\xi;T}(\widehat{t}_L) - X_{s_0,x}(t^*)\big)\frac{\mathcal{Y}_{\xi;T}(\widehat{t}_L)}{\varphi(\xi)}\Big].\end{aligned}$$

In particular, by setting $g \equiv 1$ we obtain a stochastic representation for the factor

$$\int_A p(s_0, x, T, y)\lambda_A(dy) = \lim_{\varepsilon \downarrow 0} \mathbb{E}\Big[K_\varepsilon\big(Y_{\xi;T}(\widehat{t}_L) - X_{s_0,x}(t^*)\big)\frac{\mathcal{Y}_{\xi;T}(\widehat{t}_L)}{\varphi(\xi)}\Big].$$

PROOF. Let us abbreviate

$$\begin{aligned}H_A &:= \mathbb{E}\big[g\big(X_{s_0,x}(s_1), \dots, X_{s_0,x}(t_{L-1})\big) | X_{s_0,x}(T) \in A\big], \\ H(y) &:= \mathbb{E}\big[g\big(X_{s_0,x}(s_1), \dots, X_{s_0,x}(t_{L-1})\big) | X_{s_0,x}(T) = y\big],\end{aligned}$$

and consider the density of the conditional distribution of $X_{s_0,x}(T)$ given $X_{s_0,x}(T) \in A$ with respect to the measure λ_A, that is,

$$q(y) = \frac{p(s_0, x, T, y)}{\int_A p(s_0, x, T, z)\lambda_A(dz)} 1_A(y).$$

Recall (22) and the construction (21) of λ_A. Then we have

$$\begin{aligned}H_A &= \int_A \mathbb{E}\big[g\big(X_{s_0,x}(s_1), \dots, X_{s_0,x}(t_{L-1})\big) | X_{s_0,x}(T) = y\big] q(y)\lambda_A(dy) \\ &= \int_A H(y)q(y)\lambda_A(dy)\end{aligned}$$

$$= \mathbb{E}\left[\frac{H(\xi)q(\xi)}{\varphi(\xi)}\right]$$
$$= \frac{\mathbb{E}[(p(s_0,x,T,\xi)/\varphi(\xi))\mathbb{E}[g(X_{s_0,x}(s_1),\dots,X_{s_0,x}(t_{L-1}))|X_{s_0,x}(T)=\xi]]}{\int_A p(s_0,x,T,z)\lambda_A(dz)}.$$

Hence, denoting $p_A := \int_A p(s_0,x,T,z)\lambda_A(dz)$,

$$\begin{aligned} H_A \times p_A &= \mathbb{E}\Big[\frac{1}{\varphi(\xi)}\lim_{\varepsilon\downarrow 0}\mathbb{E}^{\xi}\big[g\big(X_{s_0,x}(s_1),\dots,X_{s_0,x}(t^*),Y_{\xi;T}(\widehat{t}_{L-1}),\dots,Y_{\xi;T}(\widehat{t}_1)\big) \\ &\qquad \times K_\varepsilon\big(Y_{\xi;T}(\widehat{t}_L)-X_{s_0,x}(t^*)\big)\mathcal{Y}_{\xi;T}(\widehat{t}_L)\big]\Big] \\ &= \lim_{\varepsilon\downarrow 0}\mathbb{E}\Big[g\big(X_{s_0,x}(s_1),\dots,X_{s_0,x}(t^*),Y_{\xi;T}(\widehat{t}_{L-1}),\dots,Y_{\xi;T}(\widehat{t}_1)\big) \\ &\qquad \times K_\varepsilon\big(Y_{\xi;T}(\widehat{t}_L)-X_{s_0,x}(t^*)\big)\frac{\mathcal{Y}_{\xi;T}(\widehat{t}_L)}{\varphi(\xi)}\Big]. \end{aligned}$$

□

COROLLARY 3.8. *The conditional expectation of* $g(X_{s_0,x}(s_1),\dots,X_{s_0,x}(t_{L-1}))$ *given* $X^1_{s_0,x}(T)=c^1\in\mathbb{R}$ *has the stochastic representation*

$$\begin{aligned} &\lim_{\varepsilon\downarrow 0}\mathbb{E}\Big[g\big(X_{s_0,x}(s_1),\dots,X_{s_0,x}(t^*),Y_{\xi;T}(\widehat{t}_{L-1}),\dots,Y_{\xi;T}(\widehat{t}_1)\big) \\ &\qquad \times K_\varepsilon\big(Y_{\xi;T}(\widehat{t}_L)-X_{s_0,x}(t^*)\big)\frac{\mathcal{Y}_{\xi;T}(\widehat{t}_L)}{\varphi(\xi)}\Big] \\ &= \int_{\mathbb{R}^{d-1}} p\big(s_0,x,T,c^1,y\big)\,dy \\ &\qquad \times \mathbb{E}\big[g\big(X_{s_0,x}(s_1),\dots,X_{s_0,x}(t_{L-1})\big)|X^1_{s_0,x}(T)=c^1\big] \end{aligned}$$

for any ξ *taking values in the hyperplane* $A := \{z\in\mathbb{R}^d | z^1 = c^1\}$ *such that* $\varphi > 0$ *is the density of the law of* ξ *with respect to* λ_A *defined accordingly. In particular, by setting* $g\equiv 1$, *we obtain a stochastic representation for the marginal density*

$$\begin{aligned} &\lim_{\varepsilon\downarrow 0}\mathbb{E}\Big[K_\varepsilon\big(Y_{\xi;T}(\widehat{t}_L)-X_{s_0,x}(t^*)\big)\frac{\mathcal{Y}_{\xi;T}(\widehat{t}_L)}{\varphi(\xi)}\Big] \\ &\qquad = \int_{\mathbb{R}^{d-1}} p\big(s_0,x,T,c^1,y^2,\dots,y^d\big)\,dy^2\cdots dy^d. \end{aligned}$$

REMARK 3.9. Without doubt it is possible to construct analogous stochastic representations for conditional Markov chains in the spirit of Milstein, Schoenmakers and Spokoiny (2007). The details, however, are considered beyond the scope of the present paper.

4. Forward-reverse estimators and their analysis. The stochastic representations for the conditional diffusion problem (1), derived in the previous section, naturally lead to respective Monte Carlo estimators. In this section we analyze the accuracy of these estimators, under the following assumptions. First we need suitably regularity of the transition densities of both forward and reverse processes.

CONDITION 4.1. We assume that the diffusion X as well as the reverse diffusion Y (not including $\mathcal{Y}$) defined in (4) have C^∞ transition densities $p(t,x,s,y)$ and $q(t,x,s,y)$, respectively. Moreover, for fixed $N \in \mathbb{N}$, there are constants $m_N \in \mathbb{N}$, $\nu_N > 0$, $\lambda_N > 0$, $K_N > 0$ and $C_0 > 0$ such that for any multi-indices $\alpha, \beta \in \mathbb{N}_0^d$ with $|\alpha| + |\beta| \le N$ we have

$$\left|\partial_x^\alpha \partial_y^\beta p(t,x,s,y)\right| \le \frac{K_N}{(s-t)^{\nu_N}} \exp\left(-\lambda_N \frac{|y-x|^2}{(1+C_0^2)(s-t)}\right),$$

uniformly for $(t,x,s,y) \in (0,T] \times \mathbb{R}^d \times \mathbb{R}^d$ and similarly for q.

REMARK 4.2. In fact, for the theorems formulated as below, we only need Condition 4.1 for $N = 2$. Higher order versions only become necessary in the context of Remark 4.17.

REMARK 4.3. By the results of Kusuoka and Stroock (1985), Corollary 3.25, Condition 4.1 is satisfied *in the autonomous case* provided that (the vector fields driving) the forward diffusion X and Y satisfy a uniform Hörmander condition, and a and σ are bounded, and C^∞ bounded; that is, all the derivatives are bounded as well. We know of no similar study for nonautonomous stochastic differential equations. Of course, the seminal work by Aronson (1967) gives upper (and lower) Gaussian bounds for the transition density of time-dependent, but uniformly elliptic stochastic differential equations. Moreover, Cattiaux and Mesnager (2002) prove the existence and smoothness of transition densities for time-dependent SDEs under Hörmander conditions.

In any case, an extension of the Kusuoka–Stroock result to the time-inhomogeneous case seems entirely possible, in particular since we do not consider time-derivatives, for instance, by first considering the case of piecewise constant coefficients.

CONDITION 4.4. The kernel K satisfies $\int_{\mathbb{R}^d} K(v)\,dv = 1$ and $\int_{\mathbb{R}^d} vK(v)\,dv = 0$. Moreover, it has lighter tails than a Gaussian density in the sense that there are constants $C, \alpha > 0$ and $\beta \ge 0$ such that

$$K(v) \le C \exp\left(-\alpha |v|^{2+\beta}\right), \qquad v \in \mathbb{R}^d.$$

In many applications, one would probably choose a compactly supported kernel, which trivially satisfies the above tail-condition. Finally, we also introduce some further assumptions put forth for convenience, which could be easily relaxed.

CONDITION 4.5. The functional $g:\mathbb{R}^{(K+L-1)\times d}\to\mathbb{R}$ together with its gradient and its Hessian are bounded. Moreover, the coefficient c in (4) is bounded.

REMARK 4.6. Condition 4.5 could be replaced by a requirement of polynomial boundedness.

4.1. *Forward-reverse estimators for conditioning on a fixed state.* Let us consider

$$h_\varepsilon := \mathbb{E}\Big[g\big(X_{s_0,x}(s_1),\dots,X_{s_0,x}(t^*),Y_{y;T}(\widehat{t}_{L-1}),\dots,Y_{y;T}(\widehat{t}_1)\big) \times \varepsilon^{-d}K\Big(\frac{Y_{y;T}(\widehat{t}_L)-X_{s_0,x}(t^*)}{\varepsilon}\Big)\mathcal{Y}_{y;T}(\widehat{t}_L)\Big], \tag{23}$$

which can—and will—be computed using Monte Carlo simulation. Here, we recall the definition of $\widehat{t}_i = t_L - t_{L-i}$ given in (14). By Theorem 3.4, h_ε converges to

$$h := p(s_0,x,T,y)\mathbb{E}\big[g\big(X_{s_0,x}(s_1),\dots,X_{s_0,x}(t_{L-1})\big)|X_{s_0,x}(T)=y\big]. \tag{24}$$

THEOREM 4.7. *Assuming Conditions 4.1, 4.4 and 4.5, there are constants $C,\varepsilon_0>0$ such that the bias of the approximation h_ε can be bounded by*

$$|h-h_\varepsilon|\le C\varepsilon^2,\qquad 0<\varepsilon<\varepsilon_0.$$

PROOF. Changing variables $y_0\to v := \frac{y_0-x_K}{\varepsilon}$ in Theorem 3.3, we arrive at

$$\begin{aligned} h_\varepsilon = \int & g(x_1,\dots,x_K,y_1,\dots,y_{L-1})K(v) \\ &\times\prod_{i=1}^{K}p(s_{i-1},x_{i-1},s_i,x_i)\,p\big(t^*,x_K+\delta v,t_1,y_1\big) \\ &\times\prod_{i=2}^{L}p(t_{i-1},y_{i-1},t_i,y_i)\,dx_1\cdots dx_K\,dv\,dy_1\cdots dy_{L-1}. \end{aligned}$$

In particular, we have that $h=\lim_{\varepsilon\downarrow0}h_\varepsilon$. Consider

$$r_\varepsilon(x_K,y_1) := \int k(v)p\big(t^*,x_K+\varepsilon v,t_1,y_1\big)\,dv - p\big(t^*,x_K,t_1,y_1\big).$$

In the following, we use the notation $\partial_x^\beta := \partial_{x^1}^{\beta^1}\cdots\partial_{x^d}^{\beta^d}$, for $x\in\mathbb{R}^d$, $\beta\in\mathbb{N}^d$. By Taylor's formula, Conditions 4.4 and 4.1, we get

$$\begin{aligned} r_\varepsilon &= \int K(v)\big[p\big(t^*,x_K+\varepsilon v,t_1,y_1\big)-p\big(t^*,x_K,t_1,y_1\big)\big]\,dv \\ &= \int K(v)\big[\varepsilon\partial_x p\big(t^*,x_K,t_1,y_1\big)\cdot v\big]\,dv \\ &\quad+\sum_{|\beta|=2}\frac{2}{\beta!}\varepsilon^2\iint_0^1(1-t)\partial_x^\beta p\big(t^*,x_K+t\varepsilon v,t_1,y_1\big)\cdot v^\beta\,dt\,K(v)\,dv \end{aligned}$$

implying that

$$
\begin{aligned}
|r_\varepsilon(x_K, y_1)| &\le \sum_{|\beta|=2} \frac{2}{\beta!}\varepsilon^2 \int_0^1 (1-t) \int \left|\partial_x^\beta p\big(t^*, x_K + t\varepsilon v, t_1, y_1\big)\right| |v|^2 K(v)\, dv\, dt \\
&\le \sum_{|\beta|=2} \frac{2}{\beta!}\varepsilon^2 C_1 \int_0^1 (1-t) \int e^{-\gamma|y_1 - x_K - t\varepsilon v|^2} |v|^2 K(v)\, dv\, dt \\
&\le \sum_{|\beta|=2} \frac{2}{\beta!}\varepsilon^2 C_1 C_\eta \int_0^1 (1-t) \int e^{-\gamma|y_1 - x_K - t\varepsilon v|^2} e^{-\eta|v|^2}\, dv\, dt,
\end{aligned}
$$

where $C_1 := \frac{K_2(t_1 - t^*)}{(t_1 - t^*)^{\nu_2}}$, $\gamma := \frac{\lambda_2}{(1 + C_0^2)(t_1 - t^*)}$ as given in Condition 4.1, $\eta > 0$ and C_η is chosen such that $|v|^2 K(v) \le C_\eta e^{-\eta|v|^2}$, which is possible by Condition 4.4. Since

$$
|y_1 - x_K - t\varepsilon v|^2 = |y_1 - x_K|^2 - 2t\varepsilon\langle y_1 - x_K, v\rangle + t^2\varepsilon^2|v|^2,
$$

we can further compute, using $\sigma^2 := \frac{1}{2(\eta + \gamma t^2 \varepsilon^2)} \le \frac{1}{2\eta}$,

$$
\begin{aligned}
&\int e^{-\gamma|y_1 - x_K - t\varepsilon v|^2} e^{-\eta|v|^2}\, dv \\
&\quad = e^{-\gamma|y_1 - x_K|^2} \int e^{2t\gamma\varepsilon\langle y_1 - x_K, v\rangle} e^{-|v|^2/(2\sigma^2)}\, dv \\
&\quad = \left(\frac{\eta + \gamma t^2 \varepsilon^2}{\pi}\right)^{-d/2} \exp\left(\varepsilon^2 \frac{t^2\gamma^2}{\eta} |y_1 - x_K|^2\right) e^{-\gamma|y_1 - x_K|^2} \\
&\quad \le \left(\frac{\pi}{\eta}\right)^{d/2} e^{\varepsilon^2(\gamma^2/\eta)|y_1 - x_K|^2} e^{-\gamma|y_1 - x_K|^2}.
\end{aligned}
$$

Defining $\widetilde{C}_\eta := \sum_{|\beta|=2} \frac{1}{\beta!} C_1 C_\eta (\pi/\eta)^{d/2}$, we get the bound

$$
\begin{aligned}
|r_\varepsilon(x_K, y_1)| &\le 2\widetilde{C}\varepsilon^2 e^{-\gamma|y_1 - x_K|^2} \int_0^1 (1-t) e^{\varepsilon^2(\gamma^2/\eta)|y_1 - x_K|^2}\, dt \\
&\le \widetilde{C}\varepsilon^2 e^{-\gamma'|y_1 - x_K|^2},
\end{aligned}
$$

with $\gamma' = \gamma - \frac{\gamma^2}{\eta}\varepsilon^2$, which is positive for $0 < \varepsilon < \varepsilon_0 := (\eta/\gamma)^{1/2}$. Consequently, for $0 < \varepsilon < \varepsilon_0$, we can interpret $s_\varepsilon(x_K, y_1) := |r_\varepsilon(x_K, y_1)|/(C_2\varepsilon^2)$ as a (Gaussian) transition density, which has moments of all orders, for a suitable normalization constant C_2, for which we can derive explicit upper bounds. Thus we finally ob-

tain

$$
\begin{aligned}
|h_\varepsilon - h| &\le \int |g(x_1,\dots,x_K,y_1,\dots,y_{L-1})| \prod_{i=1}^{K} p(s_{i-1},x_{i-1},s_i,x_i) \\
&\qquad \times |r_\varepsilon(x_K,y_1)| \prod_{i=2}^{L} p(t_{i-1},y_{i-1},t_i,y_i)\,dx_1\cdots dx_K\,dy_1\cdots dy_{L-1} \\
&\le C_2\varepsilon^2 \int |g(x_1,\dots,x_K,y_1,\dots,y_{L-1})| \prod_{i=1}^{K} p(s_{i-1},x_{i-1},s_i,x_i) \qquad (25)\\
&\qquad \times s_\varepsilon(x_K,y_1) \prod_{i=2}^{L} p(t_{i-1},y_{i-1},t_i,y_i)\,dx_1\cdots dx_K\,dy_1\cdots dy_{L-1} \\
&=: C\varepsilon^2 < \infty,
\end{aligned}
$$

provided that $0 < \varepsilon < \varepsilon_0$, as the last expression can be interpreted as

$$C_2\mathbb{E}\big[|g(Z_{s_1},\dots,Z_{s_K},Z_{t_1},\dots,Z_{t_{L-1}})||Z_{s_0}=x,\,Z_T=y\big]$$

for a Markov process Z with transition densities $p(s_{i-1},x_{i-1},s_i,x_i)$, $1\le i\le K$, $s_\varepsilon(x_K,y_1)$, $p(t_{i-1},y_{i-1},t_i,y_i)$, $2\le i\le L$, which admits finite moments of all orders by construction. □

REMARK 4.8. Note that the constant C in the above statement can be explicitly bounded in terms of the bound on g, the constants appearing in Condition 4.1 and η.

In the spirit of Milstein, Schoenmakers and Spokoiny (2004) we now introduce a Monte Carlo estimator $\widehat{h}_\varepsilon$ for the quantity h_ε introduced in (23). Let us denote

$$
\begin{aligned}
Z^\varepsilon_{nm} &:= \frac{1}{\varepsilon^d} g\big(X^n_{s_0,x}(s_1),\dots,X^n_{s_0,x}(s_K),Y^m_{y;T}(\widehat{t}_{L-1}),\dots,Y^m_{y;T}(\widehat{t}_1)\big) \\
&\qquad \times K\left(\frac{Y^m_{y;T}(\widehat{t}_L) - X^n_{s_0,x}(t^*)}{\varepsilon}\right)\mathcal{Y}^m_{y;T}(\widehat{t}_L).
\end{aligned} \qquad (26)
$$

Note that $\mathbb{E}[Z^\varepsilon_{nm}] = h_\varepsilon$. The Monte Carlo estimator is now defined by

$$\widehat{h}_{\varepsilon,M,N} := \frac{1}{NM}\sum_{n=1}^{N}\sum_{m=1}^{M} Z^\varepsilon_{nm}, \qquad (27)$$

where the superscripts n and m denote different, independent realizations of the corresponding processes. We are left to analyze the variance of the estimator

$\widehat{h}_{\varepsilon,M,N}$. To this end, we consider the expectation $\mathbb{E}[Z^{\varepsilon}_{nm} Z^{\varepsilon}_{n'm'}]$ for various combinations of n, m, n' and m'.

REMARK 4.9. For the remainder of the section, we omit the sub-scripts in X, Y and $\mathcal{Y}$ as we keep the initial times and values fixed.

LEMMA 4.10. *For $m \neq m'$ we obtain*

$$\begin{aligned}
&\mathbb{E}[Z^{\varepsilon}_{nm} Z^{\varepsilon}_{nm'}]|_{\varepsilon=0} \\
&\quad = \int g(x_1, \dots, x_K, y_1, \dots, y_{L-1}) g(x_1, \dots, x_K, y'_1, \dots, y'_{L-1}) \\
&\qquad \times p(t^*, x_K, t_1, y_1) p(t^*, x_K, t_1, y'_1) \\
&\qquad \times \prod_{i=1}^{K} p(s_{i-1}, x_{i-1}, s_i, x_i)\, dx_i \prod_{i=2}^{L} p(t_{i-1}, y_{i-1}, t_i, y_i)\, dy_{i-1} \\
&\qquad \times \prod_{i=2}^{L} p(t_{i-1}, y'_{i-1}, t_i, y'_i)\, dy'_{i-1}.
\end{aligned}$$

Moreover, we can bound

$$\left|\mathbb{E}[Z^{\varepsilon}_{nm} Z^{\varepsilon}_{nm'}] - \mathbb{E}[Z^{\varepsilon}_{nm} Z^{\varepsilon}_{nm'}]|_{\varepsilon=0}\right| \leq C\varepsilon^2.$$

PROOF. In what follows, C is a positive constant, which may change from line to line. We have

$$\begin{aligned}
&E[Z^{\varepsilon}_{nm} Z^{\varepsilon}_{nm'}] \\
&\quad = \varepsilon^{-2d} E\Bigg[g(X^n_{s_1}, \dots, X^n_{s_K}, Y^m_{\widehat{t}_{L-1}}, \dots, Y^m_{\widehat{t}_1}) g(X^n_{s_1}, \dots, X^n_{s_K}, Y^{m'}_{\widehat{t}_{L-1}}, \dots, Y^{m'}_{\widehat{t}_1}) \\
&\qquad\qquad \times K\left(\frac{Y^m_{\widehat{t}_L} - X^n_{t^*}}{\varepsilon}\right) K\left(\frac{Y^{m'}_{\widehat{t}_L} - X^n_{t^*}}{\varepsilon}\right) \mathcal{Y}^m_{\widehat{t}_L} \mathcal{Y}^{m'}_{\widehat{t}_L} \Bigg] \\
&\quad = \varepsilon^{-2d} \int g(x_1, \dots, x_K, y_1, \dots, y_{L-1}) g(x_1, \dots, x_K, y'_1, \dots, y'_{L-1}) \\
&\qquad \times K\left(\frac{y_0 - x_K}{\varepsilon}\right) K\left(\frac{y'_0 - x_K}{\varepsilon}\right) \prod_{i=1}^{K} p(s_{i-1}, x_{i-1}, s_i, x_i)\, dx_i \\
&\qquad \times \prod_{i=1}^{L} p(t_{i-1}, y_{i-1}, t_i, y_i)\, dy_{i-1} \prod_{i=1}^{L} p(t_{i-1}, y'_{i-1}, t_i, y'_i)\, dy'_{i-1} \\
&\quad = \int g(x_1, \dots, x_K, y_1, \dots, y_{L-1}) g(x_1, \dots, x_K, y'_1, \dots, y'_{L-1}) \\
&\qquad \times K(v) K(v') p(t^*, x_K + \varepsilon v, t_1, y_1)\, dv p(t^*, x_K + \varepsilon v', t_1, y'_1)\, dv'
\end{aligned}$$

$$\times \prod_{i=1}^{K} p(s_{i-1}, x_{i-1}, s_i, x_i)\,dx_i$$
$$\times \prod_{i=2}^{L} p(t_{i-1}, y_{i-1}, t_i, y_i)\,dy_{i-1}$$
$$\times \prod_{i=2}^{L} p\bigl(t_{i-1}, y'_{i-1}, t_i, y'_i\bigr)\,dy'_{i-1},$$

where we changed variables $v := (y_0 - x_K)/\varepsilon$ and $v' := (y'_0 - x_K)/\varepsilon$. Thus, for $\varepsilon = 0$, we arrive at the above expression, which is treated as a problem-dependent constant.

Using Condition 4.4 [and the short-hand notation $p(x, y) := p(t^*, x, t_1, y)$], we now consider

$$\begin{aligned}
&r_\varepsilon^{(1,2)}\bigl(x_K, y_1, y'_1\bigr)\\
&\quad := \int K(v)K\bigl(v'\bigr)\\
&\qquad \times \bigl[p(x_K + \varepsilon v, y_1)p\bigl(x_K + \varepsilon v', y'_1\bigr) - p(x_K, y_1)p\bigl(x_K, y'_1\bigr)\bigr]\,dv\,dv'\\
&\quad = \varepsilon^2 \int K(v)K\bigl(v'\bigr)\\
&\qquad \times \int_0^1 (1-t)\Biggl[\sum_{i=1}^{d} \partial_x^{2e_i} p(x_K + t\varepsilon v, y_1)p\bigl(x_K + t\varepsilon v', y'_1\bigr)v_i^2\\
&\qquad\qquad + \sum_{i=1}^{d} p(x_K + t\varepsilon v, y_1)\\
&\qquad\qquad\qquad \times \partial_x^{2e_i} p\bigl(x_K + t\varepsilon v', y'_1\bigr)\bigl(v'_i\bigr)^2\\
&\qquad\qquad + 2\sum_{i,j=1}^{d} \partial_x^{e_i} p(x_K + t\varepsilon v, y_1)\\
&\qquad\qquad\qquad \times \partial_x^{e_j} p\bigl(x_K + t\varepsilon v', y'_1\bigr)v_i v'_j\,dv\,dv'\Biggr]\,dt\,dv\,dv',
\end{aligned}$$

where, for instance, $\partial_x^{e_i} \equiv \partial_{x^i}$ and $\partial_x^{2e_i} \equiv \partial_{x^i}\partial_{x^i}$. By similar techniques as in the proof of Theorem 4.7, relying once more on the uniform bounds of Condition 4.1, we arrive at an upper bound

$$\bigl|r_\varepsilon^{(1,2)}\bigl(x_K, y_1, y'_1\bigr)\bigr| \le C s_\varepsilon^{(1,2)}(x_K, y_1)s_\varepsilon^{(1,2)}\bigl(x_K, y'_1\bigr)$$

for a transition density $s_\varepsilon^{(1,2)}(x_K, y_1)$ with Gaussian bounds. Consequently, we obtain

$$\begin{aligned}
&\big|\mathbb{E}\big[Z^\varepsilon_{nm} Z^\varepsilon_{nm'}\big] - \mathbb{E}\big[Z^\varepsilon_{nm} Z^\varepsilon_{nm'}\big]|_{\varepsilon=0}\big| \\
&\quad \le C\varepsilon^2 \int \big|g(x_1, \dots, x_K, y_1, \dots, y_{L-1})\big| \\
&\qquad\qquad \times \big|g\big(x_1, \dots, x_K, y'_1, \dots, y'_{L-1}\big)\big| \\
&\qquad\qquad \times \prod_{i=1}^{K} p(s_{i-1}, x_{i-1}, s_i, x_i)\, dx_i\, s_\varepsilon^{(1,2)}(x_K, y_1)\, dy_1 \\
&\qquad\qquad \times \prod_{i=3}^{L} p(t_{i-1}, y_{i-1}, t_i, y_i)\, dy_{i-1} \times s_\varepsilon^{(1,2)}\big(x_K, y'_1\big)\, dy'_1 \\
&\qquad\qquad \times \prod_{i=3}^{L} p\big(t_{i-1}, y'_{i-1}, t_i, y'_i\big)\, dy'_{i-1},
\end{aligned}$$

which can be bounded by $C\varepsilon^2$ by boundedness of g. In fact, we can find densities $\widetilde{p}$ and $\widetilde{q}$ with Gaussian tails such that

$$\begin{aligned}
&\big|\mathbb{E}\big[Z^\varepsilon_{nm} Z^\varepsilon_{nm'}\big] - \mathbb{E}\big[Z^\varepsilon_{nm} Z^\varepsilon_{nm'}\big]|_{\varepsilon=0}\big| \\
&\quad \le C\varepsilon^2 \int \widetilde{p}(s_0, x, t^*, x_K)\widetilde{q}(t^*, x_K, T, y)^2\, dx_K.
\end{aligned} \tag{28}$$

□

When we consider $\mathbb{E}[Z^\varepsilon_{nm} Z^\varepsilon_{n'm}]$, we have to take care of terms $\mathcal{Y}^2_{\widehat{t}_L}$ appearing in the expectation. To this end, let us introduce

$$\mu_2(y_0, \dots, y_{L-1}) := \mathbb{E}\big[\mathcal{Y}^2_{\widehat{t}_L} | Y_{\widehat{t}_L} = y_0, \dots, Y_{\widehat{t}_1} = y_{L-1}\big].$$

In what follows, we replace $\mathcal{Y}^2_{\widehat{t}_L}$ by its conditional expectation $\mu_2(Y_{\widehat{t}_L}, \dots, Y_{\widehat{t}_1})$ and re-write the expectation as an integral w.r.t. the transition density of the reverse diffusion Y; by independence of X and $(Y, \mathcal{Y})$, we do not need to condition on X as well. Note that by Condition 4.5, μ_2 is a bounded function, and the transition densities q of the reverse process Y satisfy the bounds provided by Condition 4.1 as well.

LEMMA 4.11. *For $n \neq n'$ we have*

$$\begin{aligned}
&\mathbb{E}\big[Z^\varepsilon_{nm} Z^\varepsilon_{n'm}\big]|_{\varepsilon=0} \\
&\quad = \int g(x_1, \dots, x_{K-1}, y_0, \dots, y_{L-1}) g\big(x'_1, \dots, x'_{K-1}, y_0, \dots, y_{L-1}\big) \\
&\qquad\qquad \times \mu_2(y_0, \dots, y_{L-1}) \prod_{i=1}^{K-1} p(s_{i-1}, x_{i-1}, s_i, x_i)\, dx_i
\end{aligned}$$

$$\times \prod_{i=1}^{K-1} p(s_{i-1}, x'_{i-1}, s_i, x'_i)\,dx'_i \times$$
$$\times\, p(s_{K-1}, x_{K-1}, s_K, y_0)\, p(s_{K-1}, x'_{K-1}, s_K, y_0)$$
$$\times \prod_{i=1}^{L} q(\widehat{t}_{i-1}, y_i, \widehat{t}_i, y_{i-1})\,dy_{i-1}.$$

Moreover, there is a constant C such that

$$\bigl|\mathbb{E}[Z^{\varepsilon}_{nm} Z^{\varepsilon}_{n'm}] - \mathbb{E}[Z^{\varepsilon}_{nm} Z^{\varepsilon}_{n'm}]|_{\varepsilon=0}\bigr| \le \varepsilon^2 C.$$

PROOF. We first note that

$$\begin{aligned}
&\mathbb{E}[Z^{\varepsilon}_{nm} Z^{\varepsilon}_{n'm}] \\
&\quad = \varepsilon^{-2d}\mathbb{E}\Biggl[g(X^n_{s_1}, \ldots, X^n_{s_K}, Y^m_{\widehat{t}_{L-1}}, \ldots, Y^m_{\widehat{t}_1}) g(X^{n'}_{s_1}, \ldots, X^{n'}_{s_K}, Y^m_{\widehat{t}_{L-1}}, \ldots, Y^m_{\widehat{t}_1}) \\
&\qquad\qquad \times K\biggl(\frac{Y^m_{\widehat{t}_L} - X^n_{t^*}}{\varepsilon}\biggr) K\biggl(\frac{Y^m_{\widehat{t}_L} - X^{n'}_{t^*}}{\varepsilon}\biggr) (\mathcal{Y}^m_{\widehat{t}_L})^2 \Biggr] \\
&\quad = \varepsilon^{-2d}\mathbb{E}\Biggl[g(X^n_{s_1}, \ldots, X^n_{s_K}, Y^m_{\widehat{t}_{L-1}}, \ldots, Y^m_{\widehat{t}_1}) g(X^{n'}_{s_1}, \ldots, X^{n'}_{s_K}, Y^m_{\widehat{t}_{L-1}}, \ldots, Y^m_{\widehat{t}_1}) \\
&\qquad\qquad \times K\bigl(\frac{Y^m_{\widehat{t}_L} - X^n_{t^*}}{\varepsilon}\bigr) K\biggl(\frac{Y^m_{\widehat{t}_L} - X^{n'}_{t^*}}{\varepsilon}\biggr) \mu_2\Bigl(Y^m_{\widehat{t}_L}, \ldots, Y^m_{\widehat{t}_1}\Bigr) \Biggr].
\end{aligned}$$

By a similar approach as in Lemma 4.10, but changing variables $x_K \to v := (y_0 - x_K)/\varepsilon$ and $x'_K \to v' := (y_0 - x'_K)/\varepsilon$, we arrive at

$$\begin{aligned}
&\mathbb{E}[Z^{\varepsilon}_{nm} Z^{\varepsilon}_{n'm}] \\
&\quad = \int g(x_1, \ldots, x_{K-1}, y_0 - \varepsilon v, y_1, \ldots, y_{L-1}) \\
&\qquad \times g(x'_1, \ldots, x'_{K-1}, y_0 - \varepsilon v', y_1, \ldots, y_{L-1}) \\
&\qquad \times K(v) K(v') \mu_2(y_0, \ldots, y_{L-1}) \\
&\qquad \times \prod_{i=1}^{K-1} p(s_{i-1}, x_{i-1}, s_i, x_i)\,dx_i \prod_{i=1}^{K-1} p(s_{i-1}, x'_{i-1}, s_i, x'_i)\,dx'_i \\
&\qquad \times p(s_{K-1}, x_{K-1}, s_K, y_0 - \varepsilon v)\,dv \\
&\qquad \times p(s_{K-1}, x'_{K-1}, s_K, y_0 - \varepsilon v')\,dv' \\
&\qquad \times \prod_{i=1}^{L} q(\widehat{t}_{i-1}, y_i, \widehat{t}_i, y_{i-1})\,dy_{i-1}.
\end{aligned}$$

For $\varepsilon = 0$, Condition 4.4 implies

$$\begin{aligned}
&\mathbb{E}\big[Z^{\varepsilon}_{nm} Z^{\varepsilon}_{n'm}\big]|_{\varepsilon=0} \\
&\quad = \int g(x_1, \dots, x_{K-1}, y_0, \dots, y_{L-1}) g\big(x'_1, \dots, x'_{K-1}, y_0, \dots, y_{L-1}\big) \\
&\qquad \times \mu_2(y_0, \dots, y_{L-1}) \prod_{i=1}^{K-1} p(s_{i-1}, x_{i-1}, s_i, x_i)\, dx_i \\
&\qquad \times \prod_{i=1}^{K-1} p\big(s_{i-1}, x'_{i-1}, s_i, x'_i\big)\, dx'_i \\
&\qquad \times p(s_{K-1}, x_{K-1}, s_K, y_0) p\big(s_{K-1}, x'_{K-1}, s_K, y_0\big) \\
&\qquad \times \prod_{i=1}^{L} q(\widehat{t_i}-1, y_i, \widehat{t_i}, y_{i-1})\, dy_{i-1},
\end{aligned}$$

which gives the formula from the statement of the lemma.

For the bound on the difference, note once again that

$$\begin{aligned}
r^{(2,1)}_{\varepsilon} := \int \big[& g(x_1, \dots, x_{K-1}, y_0 - \varepsilon v, y_1, \dots, y_{L-1}) \\
& \times g\big(x'_1, \dots, x'_{K-1}, y_0 - \varepsilon v', y_1, \dots, y_{L-1}\big) \\
& \times p(s_{K-1}, x_{K-1}, s_K, y_0 - \varepsilon v) \\
& \times p\big(s_{K-1}, x'_{K-1}, s_K, y_0 - \varepsilon v'\big) \\
& - g(x_1, \dots, x_{K-1}, y_0, \dots, y_{L-1}) \\
& \times g\big(x'_1, \dots, x'_{K-1}, y_0, \dots, y_{L-1}\big) \\
& \times p(s_{K-1}, x_{K-1}, s_K, y_0) p\big(s_{K-1}, x'_{K-1}, s_K, y_0\big)\big] K(v) K\big(v'\big)\, dv\, dv'
\end{aligned}$$

can be bounded in the sense that $|r^{(2,1)}_{\varepsilon}| \le C s^{(2,1)}_{\varepsilon}(x_{K-1}, y_0) s^{(2,1)}_{\varepsilon}(x'_{K-1}, y_0)$ for transition densities $s^{(2,1)}_{\varepsilon}$ with Gaussian tails, so that

$$\begin{aligned}
&\big|\mathbb{E}\big[Z^{\varepsilon}_{nm} Z^{\varepsilon}_{n'm}\big] - \mathbb{E}\big[Z^{\varepsilon}_{nm} Z^{\varepsilon}_{n'm}\big]|_{\varepsilon=0}\big| \\
&\quad \le C\varepsilon^2 \int \mu_2(y_0, \dots, y_{L-1}) \prod_{i=1}^{K-1} p(s_{i-1}, x_{i-1}, s_i, x_i)\, dx_i \\
&\qquad \times \prod_{i=1}^{K-1} p\big(s_{i-1}, x'_{i-1}, s_i, x'_i\big)\, dx'_i s^{(2,1)}_{\varepsilon}(x_{K-1}, y_0) s^{(2,1)}_{\varepsilon}\big(x'_{K-1}, y_0\big) \\
&\qquad \times \prod_{i=1}^{L} q(\widehat{t_i}-1, y_i, \widehat{t_i}, y_{i-1})\, dy_{i-1}.
\end{aligned}$$

If q was symmetric, that is, $q(\widehat{t}_{i-1}, y_i, \widehat{t}_i, y_{i-1}) = q(\widehat{t}_{i-1}, y_{i-1}, \widehat{t}_i, y_i)$, then this expression would already have the desired form. While symmetry of q would be a very strong assumption, note that Condition 4.1 allows us to bound

$$q(\widehat{t}_{i-1}, y_i, \widehat{t}_i, y_{i-1}) \le C^i \exp\bigl(-\gamma^i |y_i - y_{i-1}|^2\bigr) =: \widetilde{C}^i s_{i-1}(y_{i-1}, y_i)$$

by a Gaussian transition density s_{i-1} which is naturally symmetric. Absorbing $\|\mu_2\|_\infty$ and $\prod_{i=1}^L \widetilde{C}^i$ into the constant C and denoting (by a mild abuse of notation)

$$\widetilde{p}(s_0, x, t^*, y_1) := \int \prod_{i=1}^{K-1} p(s_{i-1}, x_{i-1}, s_i, x_i)\, dx_i s_\varepsilon^{(2,1)}(x_{K-1}, y_0),$$

$$\widetilde{q}(t^*, y_0, T, y) := \int \prod_{i=1}^{L} s_{i-1}(y_{i-1}, y_i)\, dy_1 \cdots dy_{L-1},$$

the Chapman–Kolmogorov equation implies that

$$\begin{aligned} &\bigl|\mathbb{E}[Z^\varepsilon_{nm} Z^\varepsilon_{n'm}] - \mathbb{E}[Z^\varepsilon_{nm} Z^\varepsilon_{n'm}]|_{\varepsilon=0}\bigr| \\ &\quad \le C\varepsilon^2 \int \widetilde{p}(s_0, x, t^*, y_0)^2 \widetilde{q}(t^*, y_0, T, y)\, dy_0 \\ &\quad \le C\varepsilon^2 \int \widetilde{p}(s_0, x, t^*, y_0) \widetilde{q}(t^*, y_0, T, y)\, dy_0. \end{aligned} \tag{29}$$

□

LEMMA 4.12. *We have*

$$\begin{aligned} &\varepsilon^d E\bigl[(Z^\varepsilon_{nm})^2\bigr] \\ &\quad = \int K(v)^2\, dv \int g(x_1, \ldots, x_{K-1}, y_0, y_1, \ldots, y_{L-1}) \\ &\qquad \times \mu_2(y_0, y_1, \ldots, y_{L-1}) \\ &\qquad \times \prod_{i=1}^{K-1} p(s_{i-1}, x_{i-1}, s_i, x_i)\, p(s_{K-1}, x_{K-1}, s_K, y_0) \\ &\qquad \times \prod_{i=1}^{L} q(\widehat{t}_{i-1}, y_i, \widehat{t}_i, y_{i-1})\, dx_1 \cdots dx_{K-1}\, dy_0\, dy_1 \cdots dy_{L-1}. \end{aligned}$$

Moreover, there is a constant $C > 0$ such that

$$\Bigl|\varepsilon^d \mathbb{E}\bigl[(Z^\varepsilon_{nm})^2\bigr] - \lim_{\varepsilon \to 0} \varepsilon^d \mathbb{E}\bigl[(Z^\varepsilon_{nm})^2\bigr]\Bigr| \le C\varepsilon^2.$$

PROOF. Substituting $x_K \to v := (y_0 - x_K)/\varepsilon$, we obtain

$$\begin{aligned} &\varepsilon^d E\bigl[(Z^\varepsilon_{nm})^2\bigr] \\ &\quad = \int g(x_1, \ldots, x_{K-1}, y_0 - \varepsilon v, y_1, \ldots, y_{L-1}) \mu_2(y_0, y_1, \ldots, y_{L-1}) \end{aligned}$$

$$\times K(v)^2 \prod_{i=1}^{K-1} p(s_{i-1}, x_{i-1}, s_i, x_i)\, p(s_{K-1}, x_{K-1}, s_K, y_0 - \varepsilon v)$$

$$\times \prod_{i=1}^{L} q(\widehat{t}_{i-1}, y_i, \widehat{t}_i, y_{i-1}) \times dx_1 \cdots dx_{K-1}\, dv\, dy_0\, dy_1 \cdots dy_{L-1}.$$

For $\varepsilon \to 0$ the right-hand side gives the statement from the lemma.

For the difference, consider

$$\begin{aligned} r_\varepsilon^{(1,1)} := \int K(v)^2 \big[& g(x_1, \ldots, x_{K-1}, y_0 - \varepsilon v, y_1, \ldots, y_{L-1})^2 \\ & \times p(s_{K-1}, x_{K-1}, s_K, y_0 - \varepsilon v) \\ & - g(x_1, \ldots, x_{K-1}, y_0, y_1, \ldots, y_{L-1})^2 \\ & \qquad \times p(s_{K-1}, x_{K-1}, s_K, y_0) \big]\, dv. \end{aligned}$$

Following the procedure established in the previous lemmas, we obtain

$$|r_\varepsilon^{(1,1)}| \le C s_\varepsilon^{(1,1)}(x_{K-1}, y_0),$$

and by the argument used in the proof of Lemma 4.11, we obtain transition densities function $\widetilde{p}(s_0, x, t^*, y_0)$ and $\widetilde{q}(t^*, y_0, T, y)$ such that

$$\begin{aligned} & \left| \varepsilon^d \mathbb{E}\big[(Z_{nm}^\varepsilon)^2\big] - \lim_{\varepsilon \to 0} \varepsilon^d \mathbb{E}\big[(Z_{nm}^\varepsilon)^2\big] \right| \\ & \qquad \le C\varepsilon^2 \int \widetilde{p}(s_0, x, t^*, y_0) \widetilde{q}(t^*, y_0, T, y)\, dy_0. \end{aligned} \tag{30}$$

□

In what follows, we simplify the notation by the following conventions:

- the constant in Theorem 4.7 is denoted by C_0, that is, $|h_\varepsilon - h| \le C_0 \varepsilon^2$;
- for $m \ne m'$, we set $\mathbb{E}[Z_{nm}^\varepsilon Z_{nm'}^\varepsilon] =: h_\varepsilon^{(1,2)}$ and denote the constant for the difference by $C_{1,2}$, that is, $|h_\varepsilon^{(1,2)} - h_0^{(1,2)}| \le C_{1,2}\varepsilon^2$;
- for $n \ne n'$, we set $\mathbb{E}[Z_{nm}^\varepsilon Z_{n'm}^\varepsilon] =: h_\varepsilon^{(2,1)}$ and denote the constant for the difference by $C_{2,1}$, that is, $|h_\varepsilon^{(2,1)} - h_0^{(2,1)}| \le C_{2,1}\varepsilon^2$;
- we set $\varepsilon^d \mathbb{E}[(Z_{nm}^\varepsilon)^2] =: h_\varepsilon^{(1,1)}$ and denote the constant for the difference by $C_{1,1}$, that is, $|h_\varepsilon^{(1,1)} - h_0^{(1,1)}| \le C_{1,1}\varepsilon^2$.

LEMMA 4.13. *The variance of the estimator is given by*

$$\operatorname{Var} \widehat{h}_{\varepsilon,M,N} = \frac{1 - M - N}{NM} h_\varepsilon^2 + \frac{M-1}{NM} h_\varepsilon^{(1,2)} + \frac{N-1}{NM} h_\varepsilon^{(2,1)} + \frac{\varepsilon^{-d}}{NM} h_\varepsilon^{(1,1)}.$$

REMARK 4.14. Lemma 4.13 gives a clarification of the intuitive fact that the variance of $\widehat{h}_{\varepsilon,M,N}$ explodes as $\varepsilon \to 0$ (and, hence, $K_\varepsilon \to \delta_0$). Indeed, as all the $h_\varepsilon^{(\cdot)}$

terms have a finite limit, the explosion is exclusively caused by the contribution of $\mathbb{E}[(Z^{\varepsilon}_{nm})^2] = \varepsilon^{-d} h^{(1,1)}_{\varepsilon}$. Finally, the exploding term ε^{-d} will be compensated by the factor $1/(NM)$.

PROOF OF LEMMA 4.13. The result follows immediately by (26), independence of Z^{ε}_{nm} and $Z^{\varepsilon}_{n'm'}$ when both $n \neq n'$ and $m \neq m'$ and the notation introduced above, noting that $E[Z^{\varepsilon}_{nm}] = h_{\varepsilon}$. □

We immediately obtain the following:

LEMMA 4.15. *We assume Conditions 4.1, 4.4 and 4.5 hold. Then the mean square error of the estimator $\widehat{h}_{\varepsilon,M,N}$ introduced in (27) for the term h defined in (24) satisfies*

$$\begin{aligned}
&\mathbb{E}\big[(\widehat{h}_{\varepsilon,M,N} - h)^2\big] \\
&\quad \leq \frac{1-N-M}{NM} h^2 + \frac{M-1}{NM} h_0^{(1,2)} + \frac{N-1}{NM} h_0^{(2,1)} + \frac{\varepsilon^{-d}}{NM} h_0^{(1,1)} \\
&\qquad + \frac{\varepsilon^{-d+2}}{NM} C_{1,1} + \varepsilon^2 \left[2 \frac{1-N-M}{NM} Ch + \frac{M-1}{NM} C_{1,2} + \frac{N-1}{NM} C_{2,1} \right] \\
&\qquad + \frac{(N-1)(M-1)}{NM} C_0^2 \varepsilon^4.
\end{aligned}$$

Similar to Milstein, Schoenmakers and Spokoiny (2004), we can now choose $N = M$ and the bandwidth ε so as to obtain convergence proportional to $N^{-1/2}$ in RMSE-sense.

THEOREM 4.16. *Assume Conditions 4.1, 4.4 and 4.5 and set $M = N$, and $\varepsilon = \varepsilon_N$ dependent on N.*

- *If $d \leq 4$, choose $\varepsilon_N = CN^{-\alpha}$ for some $1/4 \leq \alpha \leq 1/d$. Then we have $\mathbb{E}[(\widehat{h}_{\varepsilon_N,N,N} - h)^2] = \mathcal{O}(N^{-1})$, so we achieve the optimal convergence rate $1/2$.*
- *For $d > 4$, choose $\varepsilon_N = CN^{-2/(4+d)}$, and we obtain $\mathbb{E}[(\widehat{h}_{\varepsilon_N,N,N} - h)^2] = \mathcal{O}(N^{-8/(4+d)})$.*

PROOF. Insert $M = N$ and the respective choice of ε_N in Lemma 4.15. □

REMARK 4.17. By replacing the kernel K by *higher order* kernels,[4] one could retain the convergence rate $1/2$ even in higher dimensions, as higher order kernels lead to higher order estimates (in ε) in Lemmas 4.10, 4.11 and 4.12.

[4]Recall that the order of a kernel K is the order of the lowest order (nonconstant) monomial f such that $\int f(v)K(v)\,dv \neq 0$.

So far, we have only computed the quantity h as given in (24). However, finally we want to compute the conditional expectation

$$H := \mathbb{E}\big[g\big(X_{s_0,x}(s_1), \dots, X_{s_0,x}(t_{L-1})\big) | X_{s_0,x}(T) = y\big].$$

As $H = \frac{h}{p(s_0,x,T,y)}$ with h defined in (24), we need to divide the estimator for h by an appropriate estimator for $p(s_0, x, T, y)$—in fact, we choose the forward reverse estimator with $g \equiv 1$. Note that we have assumed that $p(s_0, x, T, y) > 0$. To rule out large error contributions when the denominator is small, we will discard experiments which give too small estimates for the transition density. More precisely, we choose our final estimator to be

$$\begin{aligned}\widehat{H}_{\varepsilon,M,N} := {} & \frac{\sum_{n=1}^N \sum_{m=1}^M g(X^n_{s_1}, \dots, X^n_{s_K}, Y^m_{\widehat{t}_{L-1}}, \dots, Y^m_{\widehat{t}_1}) K((Y^m_{\widehat{t}_L} - X^n_{t^*})/\varepsilon) \mathcal{Y}^m_{\widehat{t}_L}}{\sum_{n=1}^N \sum_{m=1}^M K((Y^m_{\widehat{t}_L} - X^n_{t^*})/\varepsilon) \mathcal{Y}^m_{\widehat{t}_L}} \\ & \times \mathbf{1}_{(1/(NM))\varepsilon^{-d} \sum_{n=1}^N \sum_{m=1}^M K((Y^m_{\widehat{t}_L} - X^n_{t^*})/\varepsilon) \mathcal{Y}^m_{\widehat{t}_L} > \overline{p}/2},\end{aligned} \tag{31}$$

where $\overline{p} > 0$ is a lower bound for $p(s_0, x, T, y)$ (for fixed s_0, x, T, y), which is assumed to be known.[5]

THEOREM 4.18. *Assume Conditions 4.1, 4.4 and 4.5 and set $M = N$ and $\varepsilon = \varepsilon_N$ dependent on N.*

- *If $d \le 4$ (or $d > 4$ and higher order kernels are used), choose $\varepsilon_N = CN^{-\alpha}$, $1/4 \le \alpha \le 1/d$. Then we have $\mathbb{E}[(\widehat{H}_{\varepsilon_N,N,N} - H)^2] = \mathcal{O}(N^{-1})$, so we achieve the optimal convergence rate $1/2$.*
- *For $d > 4$, choose $\varepsilon_N = CN^{-2/(4+d)}$, and we obtain $\mathbb{E}[(\widehat{H}_{\varepsilon_N,N,N} - H)^2] = \mathcal{O}(N^{-8/(4+d)})$.*

PROOF. Let $X_N := \widehat{h}_{\varepsilon_N,N,N}$, and, similarly, let

$$Y_N := \frac{1}{N^2} \varepsilon_N^{-d} \sum_{n=1}^N \sum_{m=1}^N K\left(\frac{Y^m_{\widehat{t}_L} - X^n_{t^*}}{\varepsilon_N}\right) \mathcal{Y}^m_{\widehat{t}_L}$$

denote the estimator in the denominator, including the normalization factor. Moreover, let $X := h$ as defined in (24) and let $Y := p(s_0, x, T, y)$. Then we have already established in Theorem 4.16 that

$$\mathbb{E}\big[|X_N - X|^2\big] = \mathcal{O}\big(N^{-p}\big),$$
$$\mathbb{E}\big[|Y_N - Y|^2\big] = \mathcal{O}\big(N^{-p}\big),$$

[5] In practice, such a lower bound could be achieved by running an independent estimation for $p(s_0, x, T, y)$ and then taking a value at the lower end of a required confidence interval. See Remark 4.19 below for a different version of the theorem. In any case, our numerical experiments suggest that the cut-off can be safely omitted in practice. Keep in mind, however, that the ratio of the asymptotic distributions for numerator and denominator may not have finite moments.

where $p = 1$ for $d \leq 4$ and $p = \frac{8}{4+d}$ when $d > 4$. Moreover, we have obtained in Lemma 4.13 that $\operatorname{Var} X_N = \mathcal{O}(N^{-p})$ and $\operatorname{Var} Y_N = \mathcal{O}(N^{-p})$.

We will now estimate the mean square error for the quotient by splitting it into two contributions, depending on whether Y_N is small or large. To this end, let

$$\zeta_N := \frac{X_N}{Y_N} \mathbf{1}_{Y_N > D_N}$$

for a constant D_N to be specified below satisfying $D_N < \mathbb{E}[Y_N]$ (in fact, for N large enough, this constant may be chosen to be $\overline{p}/2$). Then we have

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{X_N}{Y_N} - \frac{X}{Y}\right)^2 \mathbf{1}_{Y_N > D_N}\right] &= \mathbb{E}\left[\frac{(X_N Y - Y_N X)^2}{(Y_N Y)^2} \mathbf{1}_{Y_N > D_N}\right] \\
&\leq \frac{\mathbb{E}[(Y(X_N - X) + X(Y - Y_N))^2]}{Y^2 D_N^2} \\
&\leq 2\frac{Y^2 \mathbb{E}[(X_N - X)^2] + X^2 \mathbb{E}[(Y - Y_N)^2]}{Y^2 D_N^2} \\
&\leq \frac{C^1_{X,Y}}{D_N^2 N^p},
\end{aligned} \tag{32}$$

where we used the estimates on the MSEs for numerator and denominator. On the other hand, we have, using that $D_N < \mathbb{E}Y_N$, Chebyshev's inequality and our estimate on the variance of Y_N,

$$\begin{aligned}
\mathbb{P}(Y_N \leq D_N) &= \mathbb{P}(Y_N - \mathbb{E}Y_N \leq D_N - \mathbb{E}Y_N; Y_N \leq \mathbb{E}Y_N) \\
&\leq \mathbb{P}\big(|Y_N - \mathbb{E}Y_N| \geq \mathbb{E}Y_N - D_N\big) \\
&\leq \frac{\operatorname{Var} Y_N}{(\mathbb{E}Y_N - D_N)^2} \\
&\leq \frac{C_Y^2}{(\mathbb{E}Y_N - D_N)^2 N^p}.
\end{aligned} \tag{33}$$

Finally, consider

$$\begin{aligned}
\mathbb{E}\left[\left(\zeta_N - \frac{X}{Y}\right)^2\right] &= \mathbb{E}\left[\left(\zeta_N - \frac{X}{Y}\mathbf{1}_{Y_N > D_N} - \frac{X}{Y}\mathbf{1}_{Y_N \leq D_N}\right)^2\right] \\
&= \mathbb{E}\left[\left(\zeta_N - \frac{X}{Y}\mathbf{1}_{Y_N > D_N}\right)^2\right] + \frac{X^2}{Y^2}\mathbb{P}(Y_N \leq D_N) \\
&\leq \frac{C^1_{X,Y}}{D_N^2 N^p} + \frac{C_Y^2 X^2}{(\mathbb{E}Y_N - D_N)^2 Y^2 N^p},
\end{aligned} \tag{34}$$

where we have combined (32) and (33). Now choose $D_N := \overline{p}/2$ for N large enough. As $\mathbb{E}Y_N \overset{N\to\infty}{\longrightarrow} Y$, (34) implies that

$$\mathbb{E}\left[\left(\zeta_N - \frac{X}{Y}\right)^2\right] = \mathcal{O}(N^{-p}). \tag{35}$$

□

REMARK 4.19. Alternatively, we could replace the cut-off $\overline{p}/2$ in (31) by some sequence $D_N \overset{N\to\infty}{\longrightarrow} 0$. In that case, the MSE of the estimator is of order $\mathcal{O}(N^{-p}/D_N^2)$, which can be chosen as close to $\mathcal{O}(N^{-p})$ as desired by proper choices of (slowly convergent) sequences D_N. Note that finally $\mathbb{E}Y_N > D_N$ in the proof of Theorem 4.18, as $\mathbb{E}Y_N \overset{N\to\infty}{\longrightarrow} p(s_0, x, T, y) > 0$ by assumption.

4.2. *Forward-reverse estimators for conditioning on a set.* In Theorem 3.7 and Corollary 3.8 we have derived a representation of the conditional expectation of a functional g of the process X given that $X_T \in A$ (for a Borel set A with positive probability) or given $X_T^1 = y^1, \dots, X_T^{d'} = y^{d'}$. In analogy to the first part of this section, one can construct Monte Carlo estimators for these conditional expectations and analyze their bias and variance. In what follows, we assume that A is either a general Borel set with positive probability or an affine surface, that is, we treat both cases distinguished above together.

Recall that we represented the conditional expectation as

$$\begin{aligned}\lim_{\varepsilon\downarrow 0}\mathbb{E}\Bigg[& g\big(X_{s_0,x}(s_1), \dots, X_{s_0,x}(s_K), Y_{\xi;T}(\widehat{t}_{L-1}, \dots Y_{t^*,\xi}(\widehat{t}_1))\big) \\ &\times \varepsilon^{-d} K\left(\frac{Y_{\xi;T}(\widehat{t}_L) - X_{s_0,x}(t^*)}{\varepsilon}\right)\frac{\mathcal{Y}_{\xi;T}(\widehat{t}_L)}{\varphi(\xi)}\Bigg] \\ &= \int_A p(s_0, x, T, y)\lambda_A(dy)\mathbb{E}\big[g\big(X_{s_0,x}(s_1), \dots, X_{s_0,x}(t_{L-1})\big)|X_{s_0,x}(T) \in A\big],\end{aligned}$$

where ξ is an independent random variable taking values in A with density φ with respect to λ_A. In order to arrive at an estimator with bounded variance, we need to restrict the choice of φ and, consequently, ξ.

CONDITION 4.20. The density φ has (strictly) super-Gaussian tails, that is, there are constants $C, \gamma, \delta > 0$ such that

$$\varphi(v)^{-1} \le C\exp\big(\gamma|v|^{2-\delta}\big), \qquad v \in A.$$

We define the following Monte Carlo estimator for the conditional expectation

$$\begin{aligned}&\widehat{H}^{\xi}_{\varepsilon,M,N} \\ &:= \frac{\sum_{n=1}^N\sum_{m=1}^M g(X^n_{s_1}, \dots, X^n_{s_K}, Y^m_{\widehat{t}_{L-1}}, \dots, Y^m_{\widehat{t}_1})K((Y^m_{\widehat{t}_L} - X^n_{t^*})/\varepsilon)(\mathcal{Y}^m_{\widehat{t}_L}/\varphi(\xi^m))}{\sum_{n=1}^N\sum_{m=1}^M K((Y^m_{\widehat{t}_L} - X^n_{t^*})/\varepsilon)(\mathcal{Y}^m_{\widehat{t}_L}/\varphi(\xi^m))} \\ &\quad\times \mathbf{1}_{(1/(NM))\varepsilon^{-d}\sum_{n=1}^N\sum_{m=1}^M K((Y^m_{\widehat{t}_L} - X^n_{t^*})/\varepsilon)(\mathcal{Y}^m_{\widehat{t}_L}/\varphi(\xi^m)) > \overline{p}/2},\end{aligned} \tag{36}$$

where $(X^n_{s_1}, \dots, X^n_{s_K})$, $1 \le n \le N$, are independent samples from the solution of the forward process X started at $X_{s_0} = x$ and $(Y^m_{\widehat{t}_L}, \dots, Y^m_{\widehat{t}_1})$ together with $\mathcal{Y}^m_{\widehat{t}_L}$, $1 \le m \le M$, are independent samples from the reverse process $(Y, \mathcal{Y})$ started at $Y^m_0 = \xi^m$, $\mathcal{Y}^m_0 = 1$, for an independent sequence of samples ξ^m from the distribution ξ. Apart from the term $\varphi(\xi^m)$, the difference to estimator (31) is the randomness of the initial values of the reverse process. Again, $p(s_0, x, T, y) > \overline{p} > 0$, and Remark 4.19 applies. The analysis of (36), however, works along the lines of the analysis of (31). Indeed, in all the expectations considered in Theorem 4.7 and in Lemmas 4.10–4.12, we obtain the same kind of results by the following steps:

(1) condition on ξ and pull out the factor $\varphi(\xi)^{-1}$ (possibly with indices m and/or m');

(2) use the results obtained in Section 4.1, with constants depending on the value of ξ;

(3) move $\varphi(\xi)^{-1}$ back in and take the expectation in ξ.

THEOREM 4.21. *Set $M = N$ and assume Condition 4.20 and, as usual, Condition 4.1, 4.4 and 4.5.*

- *If $d \le 4$, choose $\varepsilon_N = CN^{-\alpha}$, $1/4 \le \alpha \le 1/d$. Then the MSE of the forward-reverse estimator $\widehat{H}^{\xi}_{\varepsilon,M,N}$ is $\mathcal{O}(N^{-1})$.*
- *For $d > 4$, choose $\varepsilon_N = CN^{-2/(4+d)}$. Then the MSE of the forward-reverse estimator $\widehat{H}^{\xi}_{\varepsilon,M,N}$ is $\mathcal{O}(N^{-8/(4+d)})$.*

PROOF. In this proof, the constant C may change from line to line. Define

$$h^{\xi} := \int_A p(s_0, x, T, y)\lambda_A(dy) \cdot \mathbb{E}\big[g\big(X_{s_0,x}(s_1), \dots, X_{s_0,x}(t_{L-1})\big) | X_{s_0,x}(T) \in A\big],$$

$$h^{\xi}_{\varepsilon} := \mathbb{E}\Bigg[g\big(X_{s_0,x}(s_1), \dots, X_{s_0,x}(t^*), Y_{\xi;T}(\widehat{t}_{L-1}), \dots, Y_{\xi;T}(\widehat{t}_i)\big) \times K_{\varepsilon}\big(Y_{\xi;T}(\widehat{t}_L) - X_{s_0,x}(t^*)\big)\frac{\mathcal{Y}_{\xi;T}(\widehat{t}_L)}{\varphi(\xi)}\Bigg],$$

$$Z^{\varepsilon,\xi}_{nm} := \frac{1}{\varepsilon^d} g\big(X^n_{s_0,x}(s_1), \dots, X^n_{s_0,x}(s_K), Y^m_{\xi^m;T}(\widehat{t}_{L-1}), \dots, Y^m_{\xi^m;T}(\widehat{t}_1)\big) \times K\left(\frac{Y^m_{\xi^m;T}(\widehat{t}_L) - X^n_{s_0,x}(t^*)}{\varepsilon}\right)\frac{\mathcal{Y}^m_{\xi^m;T}(\widehat{t}_L)}{\varphi(\xi^m)},$$

and notice that the result will follow if we can establish the bounds of Theorem 4.7 and Lemmas 4.10, 4.11 and 4.12 for h, h_{ε} and Z^{ε}_{nm} replaced by h^{ξ}, h^{ξ}_{ε} and $Z^{\varepsilon,\xi}_{nm}$, respectively.

For the bias, (25) implies a bound $|h(y) - h_{\varepsilon}(y)| \le C\varepsilon^2 \widetilde{p}(s_0, x, T, y)$ for some density $\widetilde{p}$ in y, where we make the dependence of h and h_{ε} on y explicit. Conse-

quently, conditioning on ξ first, we have

$$
\begin{aligned}
|h^\xi - h_\varepsilon^\xi| &= \left|\mathbb{E}\left[\frac{h(\xi) - h_\varepsilon(\xi)}{\varphi(\xi)}\right]\right| \\
&\le \mathbb{E}\left[\frac{|h(\xi) - h_\varepsilon(\xi)|}{\varphi(\xi)}\right] \\
&\le C\varepsilon^2 \int \frac{\widetilde{p}(s_0, x, T, \xi)}{\varphi(\xi)} \varphi(\xi)\, d\xi \\
&\le C\varepsilon^2.
\end{aligned} \tag{37}
$$

Similarly, using the estimate from Lemma 4.10, denoting $z^\varepsilon_{n,m,m'}(y, y') := \mathbb{E}[Z^\varepsilon_{nm} Z^\varepsilon_{nm'}]$, where we assume $Y^m = Y^m_{y;T}$ and $Y^{m'} = Y^{m'}_{t^*,y'}$, we get, using a simple adaptation of (28) for different terminal values y and y',

$$
\begin{aligned}
&|\mathbb{E}[Z^{\varepsilon,\xi}_{nm} Z^{\varepsilon,\xi}_{nm'}] - \mathbb{E}[Z^{\varepsilon,\xi}_{nm} Z^{\xi}_{nm'}]|_{\varepsilon=0}| \\
&\quad\le \mathbb{E}\left[\frac{|z^\varepsilon_{n,m,m'}(\xi^m, \xi^{m'}) - z^\varepsilon_{n,m,m'}(\xi^m, \xi^{m'})|_{\varepsilon=0}|}{\varphi(\xi^m)\varphi(\xi^{m'})}\right] \\
&\quad\le C\varepsilon^2 \mathbb{E}\left[\frac{\int \widetilde{p}(s_0, x, t^*, x_K)\widetilde{q}(t^*, x_K, T, \xi^m)\widetilde{q}(t^*, x_k, T, \xi^{m'})\, dx_K}{\varphi(\xi^m)\varphi(\xi^{m'})}\right] \\
&\quad= C\varepsilon^2 \int \widetilde{p}(s_0, x, t^*, x_K)\widetilde{q}(t^*, x_K, T, y) \\
&\qquad\qquad \times \widetilde{q}(t^*, x_k, T, y')\, dx_K \lambda_A(dy)\lambda_A(dy') \\
&\quad\le C\varepsilon^2.
\end{aligned} \tag{38}
$$

Adopting the above notation for the case $n \neq n'$ covered in Lemma 4.11 and using (29), we get

$$
\begin{aligned}
&|\mathbb{E}[Z^{\varepsilon,\xi}_{nm} Z^{\varepsilon,\xi}_{n'm}] - \mathbb{E}[Z^{\varepsilon,\xi}_{nm} Z^{\varepsilon,\xi}_{n'm}]|_{\varepsilon=0}| \\
&\quad\le \mathbb{E}\left[\frac{|z^\varepsilon_{n,n',m}(\xi^m, \xi^m) - z^\varepsilon_{n,n',m}(\xi^m, \xi^m)|_{\varepsilon=0}|}{\varphi(\xi^m)\varphi(\xi^m)}\right] \\
&\quad\le C\varepsilon^2 \int \frac{\widetilde{p}(s_0, x, t^*, y_1)\widetilde{q}(t^*, y_1, T, y)}{\varphi(y)}\, dy_1 \lambda_A(dy).
\end{aligned}
$$

By assumption the density $\int \widetilde{p}(s_0, x, t^*, y_1)\widetilde{q}(t^*, y_1, T, y)\, dy_1$ has Gaussian tails, whereas φ was assumed to have strictly sub-Gaussian tails. This implies that the above integral is finite, and we get the bound

$$
|\mathbb{E}[Z^{\varepsilon,\xi}_{nm} Z^{\varepsilon,\xi}_{n'm}] - \mathbb{E}[Z^{\varepsilon,\xi}_{nm} Z^{\varepsilon,\xi}_{n'm}]|_{\varepsilon=0}| \le C\varepsilon^2. \tag{39}
$$

In a similar way, using (30), we get the bound

$$\left|\varepsilon^d \mathbb{E}\big[(Z_{nm}^{\varepsilon,\xi})^2\big] - \lim_{\varepsilon\to 0} \varepsilon^d \mathbb{E}\big[(Z_{nm}^{\varepsilon,\xi})^2\big]\right| \le C\varepsilon^2. \tag{40}$$

The respective versions of Lemmas 4.13, 4.15 and Theorem 4.16 follow immediately from the bounds (37), (38), (39) and (40), and we can repeat the proof of Theorem 4.18, arriving at the conclusion. □

We again stress that the nonoptimal complexity rate in Theorem 4.21 can be improved to the optimal one even for $d > 4$ by Remark 4.17.

4.3. *Limitations of the forward-reverse estimator.* Theorems 4.18 and 4.21 above present the asymptotic analysis of the MSE for the forward-reverse estimator. In practice, for many methods with very good asymptotic rates, limitations arise due to potentially high constants, and the forward-reverse estimator is no exception. In fact, this can be already seen in a very simple example, where all the estimates can be given explicitly.

For $s_0 = 0 < t^* < T$, consider the one-dimensional Ornstein–Uhlenbeck process

$$dX_{0,x}(t) = -\alpha X_{0,x}(t)\,dt + dB_t \tag{41}$$

for $\alpha > 0$. The corresponding reverse process satisfies

$$dY_{y;T}(t) = \alpha Y_{y;T}(t)\,dt + dW_t \tag{42}$$

for a Brownian motion W_t. Moreover, $\mathcal{Y}_{y;T}(T - t^*) = e^{\alpha(T-t^*)}$. We first discuss the estimator $\widehat{h}_{\varepsilon,N,N}$ introduced in (27) for the numerator of the forward-reverse estimator $\widehat{H}_{\varepsilon,N,N}$ for $g \equiv 1$ with $K = L = 1$. Of course, we expect that the findings for this special case carry over to situations with nonconstant g and $K, L \ge 1$.

After elementary but tedious calculations [Milstein, Schoenmakers and Spokoiny (2004), Section 4] one arrives at

$$\mathbb{E}[\widehat{h}_{\varepsilon,N,N}] = \frac{1}{\sqrt{2\pi\big(\varepsilon^2 e^{-2\alpha(T-t^*)} + \sigma_T^2\big)}} \exp\!\left(-\frac{(e^{-\alpha T}x - y)^2}{2(\varepsilon^2 e^{-2\alpha(T-t^*)} + \sigma_T^2)}\right) \tag{43}$$

and

$$\begin{aligned}\operatorname{Var}\widehat{h}_{\varepsilon,N,N} &= -\frac{2N-1}{2\pi N^2(B+\sigma_T^2)} \exp\!\left(-\frac{A}{B+\sigma_T^2}\right) \\ &\quad + \frac{N-1}{2\pi N^2\sqrt{B+\sigma_{T-t^*}^2}\sqrt{B+2\sigma_T^2-\sigma_{T-t^*}^2}} \\ &\quad \times \exp\!\left(-\frac{A}{B+2\sigma_T^2-\sigma_{T-t^*}^2}\right)\end{aligned} \tag{44}$$

$$+\frac{N-1}{2\pi N^2\sqrt{B+\sigma_T^2-\sigma_{T-t^*}^2}\sqrt{B+\sigma_T^2+\sigma_{T-t^*}^2}}\times\exp\Big(-\frac{A}{B+\sigma_T^2+\sigma_{T-t^*}^2}\Big)+\frac{e^{\alpha(T-t^*)}}{2\pi N^2\varepsilon\sqrt{B+2\sigma_T^2}}\exp\Big(-\frac{A}{B+2\sigma_T^2}\Big),$$

where

$$\sigma_s^2:=\frac{1-e^{-2\alpha s}}{2\alpha},\qquad A:=\big(e^{-\alpha T}x-y\big)^2,\qquad B:=\varepsilon^2e^{-2\alpha(T-t^*)}.$$

Thus, all the terms in the MSE [composed of the square of (43) and (44)] exhibit fairly moderate constants, except for the last term in (44). Indeed, when $\alpha\gg 0$, we have $e^{\alpha(T-t^*)}\gg 1$, unless $T-t^*\ll 1$. In other words, the constant in Theorem 4.16 will be quite large if $\alpha\gg 0$ and $T-t^*\sim 1$. That observation is quite intuitive in view of (41) and (42): $X_{0,x}$ is contracting to 0 as time increases, whereas $Y_{y;T}$ is exponentially expanding away from y. Thus, the probability of $X_{0,x}(t^*)$ and $Y_{y;T}(T-t^*)$ be close to each other is very small.

REMARK 4.22. Note that the last term in (44) is the term estimated in Lemma 4.12. The constant in the lemma depends on the constant in Condition 4.1 for the derivatives of the transition density $p(t,x',s,y')$ with respect to the y'-variable. For the Ornstein–Uhlenbeck process, the density is given by

$$p(t,x',s,y')=\frac{1}{2\pi\sigma^2(s-t)}\exp\Big(-\frac{(e^{-\alpha(s-t)}x'-y')^2}{2\sigma^2(s-t)}\Big).$$

Therefore, we see that derivatives with respect to y' (and, hence, the corresponding constants) are considerably larger than derivatives with respect to x'. This explains why the last term (and no other term) in (44) causes problems for α large.

REMARK 4.23. There is also a source of error due to the form of $\widehat{H}_{\varepsilon,N,M}$ as a fraction of two terms. The error of an approximation

$$\frac{Q}{P}\approx\frac{\widehat{Q}}{\widehat{P}}=\frac{Q+\Delta Q}{P+\Delta P}$$

of a quantity of interest Q/P by the fraction of the approximations $\widehat{Q}$ for Q and $\widehat{P}$ for P with corresponding (absolute) errors ΔQ and ΔP is controlled by the *relative* errors for Q and P. Indeed, assume for simplicity that $\Delta Q=0$ and $Q/P=\mathcal{O}(1)$, then

$$\frac{Q}{P}-\frac{\widehat{Q}}{\widehat{P}}=\mathcal{O}\Big(\frac{\Delta P/P}{1+\Delta P/P}\Big),$$

which may be close to 1 if the relative error $\Delta P/P$ for the denominator P is large.

5. Numerical study.

5.1. *Implementation.* Some care is necessary when implementing the forward reverse estimators (31) and (36) for expectations of a functional of the diffusion bridge between two points or a point and a subset. This especially concerns the evaluation of the double sum. Indeed, straightforward computation would require the cost of MN kernel evaluations which would be tremendous, for example, when $M = N = 10^5$. But, fortunately, by using kernels with an (in some sense) small support we can get around this difficulty as outlined below; see also Milstein, Schoenmakers and Spokoiny (2004) for a similar discussion.

We here assume that the kernel $K(x)$ used in (31) and (36), respectively, has bounded support contained in some ball of radius r, an assumption which is easily fulfilled in practice. For instance, even though the Gaussian kernel $K(x) = (2\pi)^{-d/2}\exp(-|x|^2/2)$ has unbounded support, in practice $K(x)$ is negligible outside a finite ball (with exponential decay of the value as function of the radius). Therefore, it is easy to choose a ball $B_r(0)$ such that K is smaller than some error tolerance const $\times$ TOL outside the ball.[6] Then, due to the small support of K, the following Monte Carlo algorithm for the kernel estimator is feasible. For simplicity, we take $N = M$. [We present the algorithm only for the case of (31), the analysis being virtually equal for (36).] Here, the input variable $\mathcal{D}$ denotes the grid (13). The complexity of the simulation steps (2) and (3) in Algorithm 1 is $\mathcal{O}(KNd)$ and $\mathcal{O}(LNd)$ elementary computations, respectively. The size l_m of the intersection in step (5) of Algorithm 1 is, on average, proportional to $N\varepsilon_N^d \times p(s_0, x, t^*, Y^m_{y;T}(\widehat{t}_L))$. The search procedure itself can be done at a cost of order $\mathcal{O}(N \log N)$ (neglecting the cost of comparison between two integers). Thus, we get the complexity bounds summarized in Theorem 5.1 below.

THEOREM 5.1. *Assume that samples from the forward process X and the reverse process $(Y, \mathcal{Y})$ can be obtained at constant cost.*[7] *Furthermore, assume that the cost of checking for equality of integers carries negligible cost. Then the following asymptotic bounds hold for the complexity of Algorithm 1:*

- *if $d \le 4$, we choose $\varepsilon_N = \mathcal{O}(N^{-1/d})$, implying that the MSE of the output of the algorithm is $\mathcal{O}(N^{-1})$ with a complexity $\mathcal{O}(N \log N)$;*
- *if $d > 4$, we choose $\varepsilon_N = \mathcal{O}(N^{-2/(4+d)})$ and obtain an MSE of $\mathcal{O}(N^{-8/(4+d)})$ with a complexity $\mathcal{O}(N \log N)$.*

[6] Obviously, the appropriate value for const depends on the size of the constants in the MSE bound.

[7] It is a straightforward exercise to adjust this calculation for the case when the corresponding stochastic differential equations need to be solved by some numerical scheme with known rate of convergence.

Algorithm 1 Forward-reverse algorithm

procedure FORREV(N, ε_N, a, σ, x, y, $\mathcal{D}$, t^*, g)
  Simulate N trajectories $(X^n_{s_0,x})_{n=1}^N$ of the forward process on $s_0 < \cdots < s_K$.
  Simulate N trajectories $(Y^m_{y;T}, \mathcal{Y}^m_{y;T})_{m=1}^N$ of the reverse process on $0 < \widehat{t}_1 < \cdots < \widehat{t}_L$.
  for $m \leftarrow 1, N$ **do**
    Find the sub-sample

$$\{X^{n_k(m)}_{s_0,x}(t^*) : k = 1, \dots, l_m\} := \{X^n_{s_0,x}(t^*) : n = 1, \dots, N\} \cap B_{r\varepsilon_N}(Y^m_{y;T}(T)).$$

  end for
  Evaluate (31) by

$$\widehat{H}_{\varepsilon,M,N} \leftarrow \frac{\sum_{m=1}^N \sum_{k=1}^{l_m} g(X^{n_k(m)}_{s_1}, \dots, X^{n_k(m)}_{s_K}, Y^m_{\widehat{t}_{L-1}}, \dots, Y^m_{\widehat{t}_1}) K((Y^m_{\widehat{t}_L} - X^{n_k(m)}_{t^*})/\varepsilon) \mathcal{Y}^m_{\widehat{t}_L}}{\sum_{m=1}^N \sum_{k=1}^{l_m} K((Y^m_{\widehat{t}_L} - X^{n_k(m)}_{t^*})/\varepsilon) \mathcal{Y}^m_{\widehat{t}_L}} \times \mathbf{1}_{1/(NM)\varepsilon^{-d} \sum_{m=1}^N \sum_{k=1}^{l_m} K((Y^m_{\widehat{t}_L} - X^{n_k(m)}_{t^*})/\varepsilon) \mathcal{Y}^m_{\widehat{t}_L} > \overline{p}/2}.$$

end procedure

5.2. *Numerical examples*. We present two numerical studies: in the first example, the forward process is a two-dimensional Brownian motion, with the standard Brownian bridge as the conditional diffusion. In the second example, we consider a Heston model whose stock price component is conditioned to end in a certain value. In both examples, we actually use a Gaussian kernel

$$K(x) = \frac{1}{(2\pi)^{d/2}} e^{-|x|^2/2},$$

and the simulation as well as the functional g of interest are defined on a uniform grid $\mathcal{D} = \{0 = s_0 < \cdots < s_K = t^* = t_0 < \cdots < t_L = T\}$ with $s_i = i/l$ and $t_j = (K + j)/l$ for $l \in \mathbb{N}$ and $L + K = l$.

EXAMPLE 5.2. We consider $X_t = B_t$, a two-dimensional standard Brownian motion, which we condition on starting at $X_0 = 0$ and ending at $X_1 = 0$, that is, the conditioned diffusion is a classical two-dimensional Brownian bridge. In particular, the reverse process Y_t is also a standard Brownian motion, and $\mathcal{Y} \equiv 1$. We consider the functional

$$g(x_1, \dots, x_{l-1}) := \sum_{j=1}^2 \left(\frac{1}{l-1} \sum_{i=1}^{l-1} x_i^j \right)^2,$$

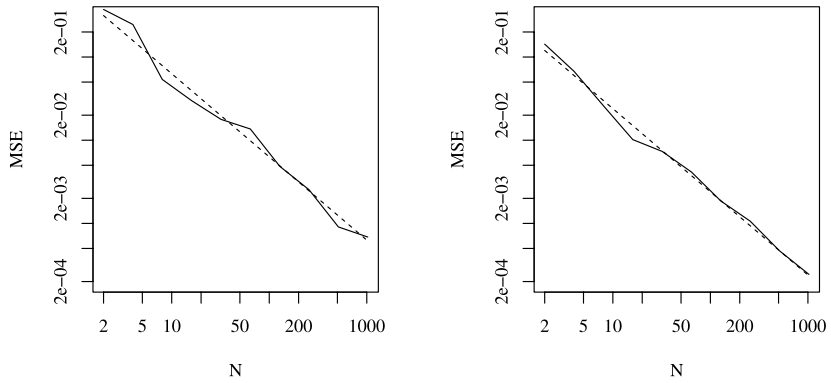


FIG. 1. *MSE for Example 5.2. Dashed lines are reference lines proportional to* N^{-1}.

where $x_i = (x_i^1, x_i^2) \in \mathbb{R}^2$. In this simple toy-example, we can actually compute the true solution

$$\mathbb{E}\big[g(X_{1/l}, \ldots, X_{(l-1)/l}) | X_0 = X_1 = 0\big] = \frac{1}{6}\frac{l+1}{l-1}.$$

As evaluation of the functional g is cheap in this case, we use a naive algorithm calculating the full double sum. We choose $M = N$ and $\varepsilon = \varepsilon_N = N^{-0.4}$, which still gives the rate of convergence obtained in Theorem 4.18.

In Figure 1, we show the results for $l = 10$, with the choices $K = 1$ and $K = 4$, that is, with $t^* = 1/10$ and $t^* = 4/10$, respectively. In both case, we observe the asymptotic relation MSE $\sim N^{-1}$ predicted by Theorem 4.18. The MSE is slightly lower when t^* is closer to the middle of the interval [0, 1] [case (b)] as compared to the situation when t^* is close to the boundary [case (a)]. Intuitively, one would expect such an effect, as in the latter case the forward process can only accumulate a considerably smaller variance as compared to the reverse process. However, it should be noted that the effect is rather small.[8]

EXAMPLE 5.3. Let us consider the stock price S_t in a Heston model: $X_t := (S_t, v_t)$, that is, the stock price together with its (stochastic) volatility satisfies the stochastic differential equation

$$dS_t = \mu S_t\, dt + \sqrt{v_t} S_t\, dB_t^1,$$
$$dv_t = (\gamma v_t + \beta)\, dt + \xi\sqrt{v_t}\big(\rho\, dB_t^1 + \sqrt{1-\rho^2}\, dB_t^2\big).$$

[8]Cf. Milstein, Schoenmakers and Spokoiny (2004), where it was noted that the variance of the forward-reverse density estimator explodes when $t^* \to T$ or $t^* \to 0$. Mathematically, this is a consequence of the transition densities getting singular.

We have

$$a(s,x) = \begin{pmatrix} \mu x^1 \\ \gamma x^2 + \beta \end{pmatrix}, \qquad \sigma(s,x) = \begin{pmatrix} \sqrt{x^2} x^1 & 0 \\ \xi \sqrt{x^2} \rho & \xi \sqrt{x^2} \sqrt{1-\rho^2} \end{pmatrix}.$$

As this process is time-homogeneous, we have $\widetilde{\sigma} = \sigma$, and the remaining coefficients of the SDE for the reverse process are given by

$$\alpha(s,x) = \begin{pmatrix} (2x^2 + \rho\xi - \mu)x^1 \\ (\rho\xi - \gamma)x^2 + \xi^2 - \beta \end{pmatrix}, \qquad c(s,x) = x^2 + \rho\xi - \mu - \gamma.$$

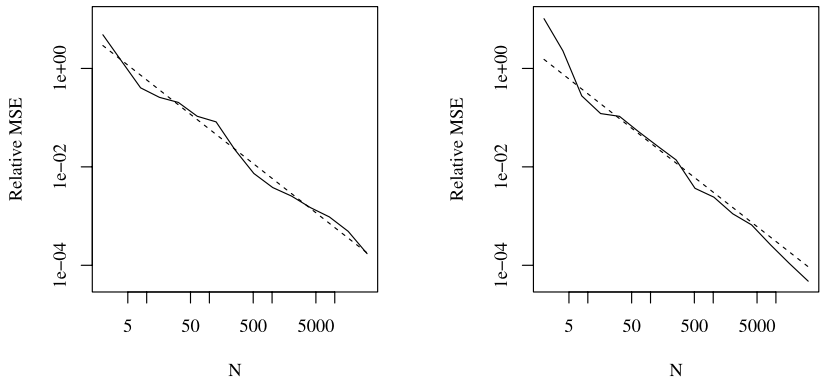
As path-dependent functional we consider the *realized variance* of the stock-price, that is, for a grid as above we consider

$$g(x_1, \ldots, x_{l-1}, x_l) := \sum_{i=1}^{l-1} \bigl(\log(x_{i+1}^1) - \log(x_i^1)\bigr)^2.$$

(Dependence of the functional g on the final value y obviously changes nothing in the theorems of Sections 3 and 4.) We choose $T = 1/12$ and $l = 30$. This time, however, we only condition on the value of the stock component at final time T. For the calculations, we use the following, arbitrarily chosen parameters: $\mu = 0.05$, $\gamma = -0.15$, $\beta = -0.045$, $\xi = 0.3$, $\rho = -0.7$. The initial stock price and the initial variance were set to $S_0 = 10$ and $v_0 = 0.25$, respectively. Moreover, the realized variance was computed conditionally on $S_T = 12$, and we choose the standard normal density for φ, despite Condition 4.20.

Contrary to Example 5.2, we cannot produce samples from the exact distributions of either the forward or the reverse processes X_t or $(Y_t, \mathcal{Y}_t)$. Thus, we approximate the corresponding paths using the Euler–Maruyama scheme on a uniform grid with mesh $h = \min(1/360, \sqrt{0.05/N})$, so that the MSE for the solution of the corresponding SDE is itself $\mathcal{O}(N^{-1})$, implying that the asymptotic order of the MSE of our quantity of interest is not affected by the numerical approximation of the forward and backward processes. Moreover, evaluation of the functional g is quite costly due to the numerous calls of the log-function. Thus, we use the cut-off procedure introduced above, so that the individual terms in the double sum are only included when the value of the kernel K_ε is larger than $\eta = 10^{-3}$. The main parameters of the forward-reverse algorithm are chosen as $M = N$ and $\varepsilon_N = (4N)^{-0.4}$, so that we are in the regime of Theorem 4.21.

The numerical results in Figure 2 confirm the rate of convergence for the MSE established in Theorem 4.21. Again, there is no significant advantage of choosing t^* in the middle of the relevant interval $[0, T]$. The "exact" reference value was computed using the forward-reverse algorithm with very large N, corresponding small ε and a very fine grid for the Euler scheme. Note that Figure 2 depicts the "relative MSE," that is, the MSE normalized by the squared reference value.



(a) $t^* = 15/360$ (b) $t^* = 29/360$

FIG. 2. *Relative MSE for Example 5.3. Dashed lines are reference lines proportional to* N^{-1}.

Acknowledgments. We are very grateful to an anonymous referee, who has pointed out to us the way to a much shorter and more transparent proof of the main Theorem 3.2. Moreover, the paper has profited from various comments made by the referee, which improved the notation and general presentation of the paper. We are also grateful to G. N. Milstein for providing us with enlightening references.

REFERENCES

ARONSON, D. G. (1967). Bounds for the fundamental solution of a parabolic equation. *Bull. Amer. Math. Soc.* (*N.S.*) **73** 890–896. MR0217444

BESKOS, A., PAPASPILIOPOULOS, O. and ROBERTS, G. O. (2006). Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli* **12** 1077–1098. MR2274855

BESKOS, A., ROBERTS, G., STUART, A. and VOSS, J. (2008). MCMC methods for diffusion bridges. *Stoch. Dyn.* **8** 319–350. MR2444507

BLADT, M. and SØRENSEN, M. (2012). Simple simulation of diffusion bridges with application to likelihood inference for diffusions. Preprint.

BOLHUIS, P. G., CHANDLER, D., DELLAGO, C. and GEISSLER, P. L. (2002). Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.* **53** 291–318.

CATTIAUX, P. and MESNAGER, L. (2002). Hypoelliptic non-homogeneous diffusions. *Probab. Theory Related Fields* **123** 453–483. MR1921010

DELYON, B. and HU, Y. (2006). Simulation of conditioned diffusion and application to parameter estimation. *Stochastic Process. Appl.* **116** 1660–1675. MR2269221

HAIRER, M., STUART, A. and VOSS, J. (2009). Sampling conditioned diffusions. In *Trends in Stochastic Analysis. London Mathematical Society Lecture Note Series* **353** 159–185. Cambridge Univ. Press, Cambridge. MR2562154

HAIRER, M., STUART, A. M. and VOSS, J. (2011). Sampling conditioned hypoelliptic diffusions. *Ann. Appl. Probab.* **21** 669–698. MR2807970

HAUSSMANN, U. G. and PARDOUX, É. (1986). Time reversal of diffusions. *Ann. Probab.* **14** 1188–1205. MR0866342

KUSUOKA, S. and STROOCK, D. (1985). Applications of the Malliavin calculus. II. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **32** 1–76. MR0783181

LIN, M., CHEN, R. and MYKLAND, P. (2010). On generating Monte Carlo samples of continuous diffusion bridges. *J. Amer. Statist. Assoc.* **105** 820–838. MR2724864

LYONS, T. J. and ZHENG, W. A. (1990). On conditional diffusion processes. *Proc. Roy. Soc. Edinburgh Sect. A* **115** 243–255. MR1069520

MILSTEIN, G. N., SCHOENMAKERS, J. G. M. and SPOKOINY, V. (2004). Transition density estimation for stochastic differential equations via forward-reverse representations. *Bernoulli* **10** 281–312. MR2046775

MILSTEIN, G. N., SCHOENMAKERS, J. G. M. and SPOKOINY, V. (2007). Forward and reverse representations for Markov chains. *Stochastic Process. Appl.* **117** 1052–1075. MR2340879

MILSTEIN, G. N. and TRETYAKOV, M. V. (2004). Evaluation of conditional Wiener integrals by numerical integration of stochastic differential equations. *J. Comput. Phys.* **197** 275–298. MR2061247

PAPASPILIOPOULOS, O. and ROBERTS, G. (2012). Importance sampling techniques for estimation of diffusion models. In *Statistical Methods for Stochastic Differential Equations*. *Monogr. Statist. Appl. Probab.* **124** 311–340. CRC Press, Boca Raton, FL. MR2976985

ROGERS, L. C. G. and WILLIAMS, D. (2000). *Diffusions, Markov Processes, and Martingales. Vol. 2: Itô Calculus*. Cambridge Univ. Press, Cambridge. MR1780932

SPIVAKOVSKAYA, D., HEEMINK, A. W. and SCHOENMAKERS, J. G. M. (2007). Two-particle models for the estimation of the mean and standard deviation of concentrations in coastal waters. *Stoch. Environ. Res. Risk Assess.* **21** 235–251. MR2339199

STINIS, P. (2011). Conditional path sampling for stochastic differential equations through drift relaxation. *Commun. Appl. Math. Comput. Sci.* **6** 63–78. MR2836693

STUART, A. M., VOSS, J. and WIBERG, P. (2004). Fast communication conditional path sampling of SDEs and the Langevin MCMC method. *Commun. Math. Sci.* **2** 685–697. MR2119934

WEIERSTRASS INSTITUTE
WIAS BERLIN
MOHRENSTR. 39
10117 BERLIN
GERMANY
E-MAIL: christian.bayer@wias-berlin.de
john.schoenmakers@wias-berlin.de