Weierstraß-Institut
für Angewandte Analysis und Stochastik
Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint ISSN 0946 – 8633

Simulation of conditional diffusions via forward–reverse stochastic representations

Christian Bayer, John G. M. Schoenmakers

submitted: January 30, 2013

Weierstrass Institute
Mohrenstr. 39
10117 Berlin
Germany
christian.bayer@wias-berlin.de
john.schoenmakers@wias-berlin.de

No. 1764
Berlin 2013

2010 *Mathematics Subject Classification.* 65C05, 65C30.

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

Partially supported by the DFG Research Center MATHEON "Mathematics for Key Technologies" in Berlin.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 20372-303
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Abstract. 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 et al. [Bernoulli 10(2):281–312, 2004] in the context of a forward-reverse transition density estimator. The corresponding Monte Carlo estimators have essentially root-N accuracy, 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 at simulation of the functional

$$\mathbb{E}\left[g(X_{s_1}, ..., X_{s_R}) \middle| X_T \in A,\ X_0 = x \right], \tag{1.1}$$

where $0 \le s_1 < s_2 < \cdots < s_R < T$, A is some set that eventually may consist of only one point, and g is an arbitrarily given suitable test function, and $x \in \mathbb{R}^d$ is a given state. In the recent years, the problem of computing terms such as (1.1) has attracted a lot of attention in the literature, sparked by several applications. As a most prominent one we mention the problem of statistical inference based on discrete observations of trajectories of a stochastic process, see Bladt and Sørensen [2012] for more information. We also refer to Lyons and Zheng [1990] for useful analytical properties related to diffusion bridges.

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 for instance 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 et al. [2010].

On the one hand let us mention the work by Beskos et al. [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 et al. [2010] use a sequential weighted Monte Carlo framework, including re-sampling with optimal priority scores.

Another general technique used for simulation of diffusion bridges is the Markov chain Monte Carlo method. Indeed, Stuart et al. [2004] and Hairer et al. [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, i.e., of a Langevin-type stochastic partial differential equation (SPDE). Thus, in principle MCMC methods are applicable as explored by Stuart et al. [2004] and Beskos et al. [2008]. However, this requires the numerical solution of the SPDE involved. It should be noted that in Hairer et al. [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.

The approach of Bladt and Sørensen [2012] seems to be closest to our approach in spirit, even though they are restricted to a one-dimensional setting. 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.

Finally, a very general approach is given by Delyon and Hu [2006] which relies on an explicit Radon-Nikodym derivative of the diffusion X conditioned on its initial and terminal values and another diffusion Y, which is modeled by 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 the most general regime, it is mainly required that the diffusion coefficient $\sigma = \sigma(t, x)$ is $C^{1,2}$, bounded with bounded derivatives and bounded inverse , together with a Lipschitz condition on the drift – Delyon and Hu [2006] provide a Girsanov type theorem, which leads to a representation of the form

$$\mathbb{E}\left[g(X) \,|\, X_0 = x,\ X_T = y\right] = \mathbb{E}\left[g(Y)Z(Y)\right]$$

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 – note that $Z(Y)$ explicitly depends on σ^{-1}. Consequently, this approach allows for efficient Monte Carlo based computations of (1.1) under some rather mild regularity conditions, provided that $Z(Y)$ and Y can be efficiently calculated/simulated. The latter requirement is non-trivial, as the exploding drift term in the stochastic differential equation for Y poses severe difficulties to many standard techniques.

Our new method is inspired by the forward-reverse estimator for the transition density $p(0, x, T, Y)$ constructed by Milstein et al. [2004]. Thus, the story in a nut-shell goes like this: We simulate the diffusion process X started at $X_0 = x$ until some (deterministic, fixed) time $0 < t^* < T$. Moreover, we simulate the *reverse process* Y started at $Y_{t^*} = y$ until time T. (Note that Y is different from the time-reversed diffusion in the sense of Haussmann and Pardoux [1986]. Indeed, the dynamics of Y are explicitly given below in terms of the dynamics of X and, usually, share the same regularity properties, see (2.2) and (2.3).) Next, we weight the trajectories according to the distance between X_{t^*} and Y_T using a kernel K with bandwidth ϵ. Here, it is important that $t^* < T$, as then Y induces smoothing on the kernel (representing $\delta(X_{t^*} - Y_T)$), thereby leading to an efficient Monte Carlo algorithm. In contrast, if we chose $t^* = T$, we would obtain the trivial algorithm of simulating the diffusion X until time T while discarding all paths with X_T too far away from y, an algorithm with much worse performance than the forward-reverse algorithm suggested in this paper. Notice, however, that the forward-reverse algorithm for (1.1) presented here is not a trivial extension of the forward-reverse algorithm for transition densities of Milstein et al. [2004]. The main difficulty lies in the extension of the representation from just one intermediate time $0 < t^* < T$ to an arbitrary time grid $0 < s_1 < \cdots < s_R < T$ (with t^* included as interior grid point). This issue is further complicated as the reverse process as defined by Milstein et al. [2004] depends explicitly on its domain $[t^*, T]$, so that the issue of how to connect the reverse processes defined on the sub-intervals $[s_k, s_{k+1}]$ to one reverse process defined on the full interval $[t^*, T]$ becomes a delicate issue, in particular in the time-inhomogeneous case, see Theorem 3.3. Moreover, the forward-reverse algorithm for conditional expectations of path dependent functionals of diffusion processes presented here requires a more precise error analysis then the forward-reverse density estimator due to Milstein et al. [2004], as bounds here generally need to be (Lebesgue) integrable as compared to merely uniformly bounded.

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 exponential 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 (1.1) is basically root-N consistent, that is the mean square estimation accuracy is of order $O(N^{-1/2})$ 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 non-trivial. 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 other ones – is only one-dimensional. The methods of Stuart et al. [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, i.e., either A has positive measure both under the Lebesgue measure and the distribution of X_T or A is an affine plane of dimension $0 \leq d' \leq d$. 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 et al. [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 Condition 4.1, 4.4 and 4.5. Section 5 provides a numerical study involving a Heston type stochastic volatility model. Some technical parts of the main theorems are deferred to the Appendix.

2. Recap of forward-reverse representations for diffusions

Let us consider the SDE

$$dX(s) = a(s, X(s))ds + \sigma(s, X(s))dW(s), \qquad 0 \leq s \leq T, \tag{2.1}$$

where $X \in \mathbb{R}^d$, $a : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d\times m}$, W is an m-dimensional standard Wiener process. At this stage, we only assume that X admits a C^2 transition density p and that the coefficients of (2.1) are C^2 as well.

Along with the (forward) process X given by (2.1), Milstein et al. [2004] introduced an associated (from X independent) process $(Y_{t,y}(s), \mathcal{Y}_{t,y,1}(s))$ in $\mathbb{R}^d \times \mathbb{R}$, $t \leq s \leq T$, termed *reverse* process on the interval $[t, T]$, that solves the SDE system

$$\begin{aligned} dY &= \alpha(s, Y)ds + \tilde{\sigma}(s, Y)d\widetilde{W}(s), \quad & Y(t) &= Y_{t,y}(t) = y, \\ d\mathcal{Y} &= c(s, Y)\mathcal{Y}ds, & \mathcal{Y}(t) &= \mathcal{Y}_{t,y,1} = 1, \end{aligned} \tag{2.2}$$

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

$$\begin{aligned} \alpha^i(s, y) &:= \sum_{j=1}^{d} \frac{\partial}{\partial y^j} b^{ij}(T + t - s, y) - a^i(T + t - s, y), \quad b := \sigma\sigma^\top \\ \widetilde{\sigma}(s, y) &:= \sigma(T + t - s, y), \\ c(s, y) &:= \frac{1}{2} \sum_{i,j=1}^{d} \frac{\partial^2 b^{ij}}{\partial y^i \partial y^j}(T + t - s, y) - \sum_{i=1}^{d} \frac{\partial a^i}{\partial y^i}(T + t - s, y). \end{aligned} \tag{2.3}$$

Despite its name, we stress that $(Y, \mathcal{Y})$ is the solution of an ordinary SDE *forward* in time on the interval $[t, T]$. In particular, One of the central results in Milstein et al. [2004] is the following theorem.

Theorem 2.1. *(M.S.S. (2004)) For fixed t, x, y and $t < t^* < T$, and any bi-variate test function f we have*

(2.4) $$\mathbb{E}\left[f(X_{t,x}(t^*), Y_{t^*,y}(T))\,\mathcal{Y}_{t^*,y}(T)\right] = \iint p(t,x,t^*,x')p(t^*,y',T,y)f(x',y')dx'dy',$$

where $X_{t,x}(s)$ satisfies the forward equation (2.1) and $(Y_{t^*,y}(s), \mathcal{Y}_{t^*,y}(s))$, $s \geq t^*$, is the solution of the reverse system (2.2) with $\mathcal{Y}_{t,y} := \mathcal{Y}_{t,y,1}$.

Corollary 2.2. *By taking $f \equiv 1$, (2.4) yields*

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

which obviously extends to $t^* = t$. By next taking $f(x',y') = f(x')$ (while abusing notation slightly) we obtain from (2.4), using (2.5) and the independence of X and $(Y,\mathcal{Y})$,

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

which obviously extends to $t^* = T$, i.e. 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

(2.6) $$\mathbb{E}\left[f(Y_{t^*,y}(T))\,\mathcal{Y}_{t^*,y}(T)\right] = \int p(t^*,y',T,y)f(y')dy',$$

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

3. Forward-reverse representations for conditional diffusions

The following simple lemma is a key for generalizing (2.6).

Lemma 3.1. *For any $t < u < T$ it holds that*

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

Proof. Due to (2.2) we have

$$\begin{aligned}\mathcal{Y}_{t,y}(T) = \mathcal{Y}_{t,y,1}(T) = e^{\int_0^T c(s,Y_{t,y}(s))ds} &= e^{\int_0^u c(s,Y_{t_1,y}(s))}e^{\int_u^T c(s,Y_{u,Y_{t,y}(u)}(s))ds} \\ &= \mathcal{Y}_{t,y,1}(u)\mathcal{Y}_{u,Y_{t,y}(u),1}(T) = \mathcal{Y}_{t,y}(u)\mathcal{Y}_{u,Y_{t,y}(u)}(T).\end{aligned}$$ □

In the autonomous case we have the following useful generalization of the reverse representation (2.6).

Lemma 3.2. *Let $y \in \mathbb{R}^d$ be fixed and $L \in \mathbb{N}_+$, $f : \mathbb{R}^{d\times L} \to \mathbb{R}$, $0 < t^* < t_1 \cdots < t_L := T$. If X is autonomous (hence Y, too), it holds that*

(3.1) $$\mathbb{E}\left[f(Y_{t^*,y}(T), Y_{t^*,y}(t_{L-1}), ..., Y_{t^*,y}(t_1))\mathcal{Y}_{t^*,y}(T)\right] = \int_{\mathbb{R}^{d\times L}} f(y_1,y_2,...,y_L)\prod_{i=1}^{L} p(t_{L-i},y_i,t_{L-i+1},y_{i+1})dy_i$$

with $y_i \in \mathbb{R}^d$, $y_{L+1} =: y$, $t_0 := t^*$.

Proof. We use induction on L. For $L = 1$ the statement boils down to (2.6) (and the autonomy is not needed in fact). Suppose the statement is proved for some $L \geq 1$. We then have with $0 < t^* < t_1 \cdots < t_{L+1} := T$, $f : \mathbb{R}^{d\times(L+1)} \to \mathbb{R}$,

$$\begin{aligned}(*) &:= \mathbb{E}\left[f(Y_{t^*,y}(t_{L+1}), Y_{t^*,y}(t_L), ..., Y_{t^*,y}(t_1))\mathcal{Y}_{t^*,y}(t_{L+1})\right] \\ &= \mathbb{E}\Big[\mathbb{E}^{Y_{t^*,y}(t_L),...,Y_{t^*,y}(t_1);\mathcal{Y}_{t^*,y}(t_L)}\Big[f(Y_{t_L,Y_{t^*,y}(t_L)}(t_{L+1}), Y_{t^*,y}(t_L), \ldots, Y_{t^*,y}(t_1))\mathcal{Y}_{t^*,y}(t_L)\mathcal{Y}_{t_L,Y_{t^*,y}(t_L)}(t_{L+1})\Big]\Big].\end{aligned}$$

Since the process is autonomous, the dependence of the reverse drift and diffusion coefficients on T and t in (2.3) disappears. We so have by Lemma 3.1 that for any h,

$$\mathbb{E}\left[h(Y_{t_L,u}(t_{L+1}))\mathcal{Y}_{t_L,u}(t_{L+1})\right] = \int_{\mathbb{R}^d} h(z)p(t_L,z,t_{L+1},u)dz,$$

and hence $(*)$ becomes

$$\mathbb{E}\left[\mathbb{E}^{Y_{t^*,y}(t_L),\ldots,Y_{t^*,y}(t_1);\mathcal{Y}_{t^*,y}(t_L)}\left[\int_{\mathbb{R}^d} p(t_L,z,t_{L+1},Y_{t^*,y}(t_L))dzf(z,Y_{t^*,y}(t_L),...,Y_{t^*,y}(t_1))\mathcal{Y}_{t^*,y}(t_L)\right]\right]$$

$$= \int_{\mathbb{R}^d} \mathbb{E}\left[p(t_L,z,t_{L+1},Y_{t^*,y}(t_L))f(z,Y_{t^*,y}(t_L),\ldots,Y_{t^*,y}(t_1))\mathcal{Y}_{t^*,y}(t_L)\right]dz.$$

We now apply the induction hypothesis for L to the function

$$\mathbb{R}^{d\times L} \ni (y_1,...,y_L) \longrightarrow p(t_L,z,t_{L+1},y_1)f(z,y_1,...,y_L)$$

in order to obtain,

$$(*) = \int_{\mathbb{R}^d} dz \int_{\mathbb{R}^{d\times L}} p(t_L,z,t_{L+1},y_1)f(z,y_1,...,y_L)\prod_{i=1}^{L} p(t_{L-i},y_i,t_{L-i+1},y_{i+1})dy_i$$

$$= \int_{\mathbb{R}^{d\times(L+1)}} f(y'_1,y'_2,...,y'_{L+1})\prod_{i=1}^{L+1} p(t_{L+1-i},y'_i,t_{L+1-i+1},y'_{i+1})dy'_i.$$

□

We are now ready to state the following key theorem.

Theorem 3.3. *Given a grid $\mathcal{D} = \{t^* < t_1 < \cdots < t_L = T\}$, we introduce the modified time grid $\widehat{\mathcal{D}} = \{t^* < \hat{t}_1 < \cdots < \hat{t}_L = T\}$ defined by*

(3.2) $$\hat{t}_i := T + t^* - t_{L-i}, \quad i = 1,\ldots,L, \quad \text{with} \quad \hat{t}_0 := t_0 := t^*,$$

and the notation $y_{L+1} := y$. Then we have

$$\mathbb{E}\left[f(Y_{t^*,y}(T),Y_{t^*,y}(\widehat{t_{L-1}}),\ldots,Y_{t^*,y}(\widehat{t_1}))\mathcal{Y}_{t^*,y}(T)\right] = \int_{\mathbb{R}^{d\times L}} f(y_1,y_2,...,y_L)\prod_{i=1}^{L} p(t_{i-1},y_i,t_i,y_{i+1})dy_i$$

where X is not necessarily autonomous!

Proof. Part I, autonomous case: This case is a direct corollary of Lemma 3.2 as we have,

$$\mathbb{E}\left[f(Y_{t^*,y}(T),Y_{t^*,y}(\widehat{t_{L-1}}),\ldots,Y_{t^*,y}(\widehat{t_1}))\mathcal{Y}_{t^*,y}(T)\right] = \int f(y_1,y_2,...,y_L)\prod_{i=1}^{L} p(\widehat{t_{L-i}},y_i,\widehat{t_{L-i+1}},y_{i+1})dy_i$$

$$= \int f(y_1,y_2,...,y_L)\prod_{i=1}^{L} p(t_{i-1},y_i,t_i,y_{i+1})dy_i,$$

due to the autonomous transition kernels. The proof of the non-autonomous case is carried out by lifting the process X to an autonomous one in $\mathbb{R}^{d+1}$ in a standard way. After adding to the lifted process a small noise term we may apply Theorem 3.3 to it. By next letting the noise term go to zero we obtain the statement of Theorem 3.3 for the autonomous case. The details are spelled out in the Appendix. □

Remark 3.4. It should be noted that unlike Theorem 3.3, Lemma 3.2 is generally not true in the non-autonomous case.

By considering a further time grid $0 \le s_0 < s_1 < \cdots < s_K \le t^*$, and a fixed starting point $x \in \mathbb{R}^d$, we may as well formulate the following theorem.

Theorem 3.5. *For any $f : \mathbb{R}^{d\times(K+L)} \to \mathbb{R}$ and $\mathcal{D}$ and $\widehat{\mathcal{D}}$ as in Theorem 3.3, we have*

$$\mathbb{E}\left[f(X_{s_0,x}(s_1),\ldots,X_{s_0,x}(s_K),Y_{t^*,y}(T),Y_{t^*,y}(\widehat{t_{L-1}}),\ldots,Y_{t^*,y}(\widehat{t_1}))\mathcal{Y}_{t^*,y}(T)\right]$$

$$= \int_{\mathbb{R}^{d\times(K+L)}} f(x_1,...,x_K,y_1,y_2,\ldots,y_L)\prod_{i=1}^{K} p(s_{i-1},x_{i-1},s_i,x_i)dx_i\prod_{i=1}^{L} p(t_{i-1},y_i,t_i,y_{i+1})dy_i$$

with $x_0 := x$, $y_{L+1} := y$, and the processes X and $(Y,\mathcal{Y})$ being independent.

Theorem 3.5 follows directly from Theorem 3.3 by a standard pre-conditioning argument and the Chapman-Kolmogorov equation. Note that for $K = L = 1$, Theorem 3.5 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 $0 \le s_0 < T$, and fixed $x, y \in \mathbb{R}^d$. To this end we henceforth assume that

$$p(s_0, x, T, y) > 0. \tag{3.3}$$

We also need to assume continuity of p. Let us take an arbitrary but fixed time grid

$$0 \le s_0 < s_1 < \cdots < s_K = t^* < t_1 \cdots < t_L := T, \tag{3.4}$$

with $K, L \in \mathbb{N}_+$, and a bounded measurable test function

$$g(x_1, ..., x_K, y_2, ..., y_L) : \mathbb{R}^{d\times(K+L-1)} \to \mathbb{R},$$

and consider the conditional expectation

$$\mathbb{E}\left[g(X_{s_0,x}(s_1), ..., X_{s_0,x}(s_{K-1}), X_{s_0,x}(t^*), X_{s_0,x}(t_1), ..., X_{s_0,x}(t_{L-1})) \middle| X_{s_0,x}(T) = y \right]. \tag{3.5}$$

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 (3.5). These conditional expectations may be obtained due to Theorem 3.6 below.

Theorem 3.6. *Consider the forward process X and its reverse process $(Y, \mathcal{Y})$ as before and the grids $\mathcal{D}$ and $\widehat{\mathcal{D}}$ as specified in Theorem 3.3. Let*

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

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 $\epsilon \downarrow 0$. 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

$$\begin{aligned}\mathbb{E}\Big[g\left(X_{s_0,x}(s_1), \ldots, X_{s_0,x}(t^*), X_{s_0,x}(t_1), \ldots, X_{s_0,x}(t_{L-1})\right) \Big| X_{s_0,x}(T) = y \Big] = \\ \frac{1}{p(s_0, x, T, y)} \lim_{\epsilon\downarrow 0} \mathbb{E}\Big[g(X_{s_0,x}(s_1), \ldots, X_{s_0,x}(t^*), Y_{t^*,y}(\hat{t}_{L-1}), \ldots, Y_{t^*,y}(\hat{t}_1)) \times \\ \times K_\epsilon\left(Y_{t^*,y}(T) - X_{s_0,x}(t^*)\right) \mathcal{Y}_{t^*,y}(T)\Big],\end{aligned} \tag{3.6}$$

Proof. By applying Theorem 3.5 to

$$f(x_1, \ldots, x_K, y_1, y_2, \ldots, y_L) := g(x_1, \ldots, x_K, y_2, \ldots, y_L) K_\epsilon(y_1 - x_K),$$

we obtain,

$$\begin{aligned}&\mathbb{E}\Big[g(X_{s_0,x}(s_1), \ldots, X_{s_0,x}(t^*), Y_{t^*,y}(\hat{t}_{L-1}), \ldots, Y_{t^*,y}(\hat{t}_1)) K_\epsilon\left(Y_{t^*,y}(T) - X_{s_0,x}(t^*)\right) \mathcal{Y}_{t^*,y}(T)\Big] \\ &= \int_{\mathbb{R}^{d\times(K+L)}} g(x_1, \ldots, x_K, y_2, \ldots, y_L) K_\epsilon(y_1 - x_K) \times \\ &\quad \times \prod_{i=1}^{K} p(s_{i-1}, x_{i-1}, s_i, x_i) dx_i \prod_{i=1}^{L} p(t_{i-1}, y_i, t_i, y_{i+1}) dy_i \\ &= \int_{\mathbb{R}^{d\times(K+L)}} g(x_1, \ldots, x_K, y_2, \ldots, y_L) K(v) dv\, p(t^*, x_k + \epsilon v, t_1, y_2) \times \qquad (3.7) \\ &\quad \times \prod_{i=1}^{K} p(s_{i-1}, x_{i-1}, s_i, x_i) dx_i \prod_{i=2}^{L} p(t_{i-1}, y_i, t_l, y_{i+1}) dy_i.\end{aligned}$$

By sending ϵ to zero, (3.7) clearly converges to

$$\int_{\mathbb{R}^{d\times(K+L-1)}} g(x_1, ..., x_K, y_2, ..., y_L) \prod_{i=1}^{K} p(s_{i-1}, x_{i-1}, s_i, x_i) dx_i \cdot p(t^*, x_K, t_1, y_2) \cdot \prod_{i=2}^{L} p(t_{i-1}, y_i, t_l, y_{i+1}) dy_i,$$

from which (3.6) easily follows. □

If the original grid $\mathcal{D} = \{t^* = t_0 < \cdots < t_L = T\}$ is equidistant, then the transformed grid $\widehat{\mathcal{D}}$ is actually equal to $\mathcal{D}$, which leads to the following corollary.

Corollary 3.7. *If the time grid $\mathcal{D}$ is equidistant, we have*

$$\mathbb{E}\left[g\left(X_{s_0,x}(s_1), \ldots, X_{s_0,x}(t^*), X_{s_0,x}(t_1), \ldots, X_{s_0,x}(t_{L-1})\right) \middle| \, X_{s_0,x}(T) = y\right] = \frac{1}{p(s_0, x, T, y)} \lim_{\epsilon \downarrow 0} \mathbb{E}\Big[g(X_{s_0,x}(s_1), \ldots, X_{s_0,x}(t^*), Y_{t^*,y}(t_{L-1}), \ldots, Y_{t^*,y}(t_1)) \times \\ \times K_\epsilon\left(Y_{t^*,y}(T) - X_{s_0,x}(t^*)\right) \mathcal{Y}_{t^*,y}(T)\Big].$$

Moreover, by setting $g \equiv 1$, we retrieve the forward-reverse representation of the transition density in Milstein et al. [2004],

$$p(s_0, x, T, y) = \lim_{\epsilon \downarrow 0} \mathbb{E}\Big[K_\epsilon\left(Y_{t^*,y}(T) - X_{s_0,x}(t^*)\right) \mathcal{Y}_{t^*,y}(T)\Big]. \tag{3.8}$$

Remark 3.8. For fixed $x, y \in \mathbb{R}^d$ and $s_0 < t^* < T$ as before, let us define a process Z by

$$Z(t) := Z_{t^*, Y_{t^*,y}(T)}(t) := Y_{t^*,y}(T - t + t^*), \quad t^* \le t \le T.$$

The idea is that we run along the reverse diffusion Y backwards in time. Then (3.6) reads

$$\mathbb{E}\left[g\left(X_{s_0,x}(s_1), \ldots, X_{s_0,x}(t^*), X_{s_0,x}(t_1), \ldots, X_{s_0,x}(t_{L-1})\right) \middle| \, X_{s_0,x}(T) = y\right] = \frac{1}{p(s_0, x, T, y)} \lim_{\epsilon \downarrow 0} \mathbb{E}\Big[g(X_{s_0,x}(s_1), \ldots, X_{s_0,x}(t^*), Z(t_1), ..., Z(t_{L-1})) \times \\ \times K_\epsilon\left(Z(t^*) - X_{s_0,x}(t^*)\right) \mathcal{Y}_{t^*,y}(T)\Big].$$

Now let us assume that we are interested in the conditional expectation of a functional $g\left(X_{s_0,x}(s_1), \ldots, X_{s_0,x}(t_{L-1})\right)$ 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

$$A = \left\{x \in \mathbb{R}^d : x^1 = c^1, ..., x^{d-d'} = c^{d-d'}\right\}. \tag{3.9}$$

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

$$\lambda_A(dx) = \delta_{\{c^1\}}(dx^1) \cdots \delta_{\{c^{d-d'}\}}(dx^{d-d'}) \cdot dx^{d-d'+1} \cdots dx^d, \tag{3.10}$$

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 $(Y_{t^*,\xi}, \mathcal{Y}_{t^*,\xi})$ denote the reverse process starting at the random location $(\xi, 1)$ at time t^*. Here, we replace the condition (3.3) on the positivity of the transition density by

$$\int_A p(s_0, x, T, z) \lambda_A(dz) > 0. \tag{3.11}$$

Theorem 3.9. *Let the kernel function K be as in Theorem 3.6, and let there be given a time grid of the form (3.4). The conditional expectation of*

$$g\left(X_{s_0,x}(s_1), \ldots, X_{s_0,x}(t^*), X_{s_0,x}(t_1), ..., X_{s_0,x}(t_{L-1})\right)$$

given $X_{s_0,x}(T) \in A$ with A being a Borel set, either with positive probability or a hyperplane of the form (3.9), and g being a bounded measurable test function, has the stochastic representation

$$\int_A p(s_0, x, T, y) \lambda_A(dy) \cdot \mathbb{E}\left[g\left(X_{s_0,x}(s_1), \ldots, X_{s_0,x}(t_{L-1})\right) \middle| X_{s_0,x}(T) \in A\right] \\ = \lim_{\epsilon \downarrow 0} \mathbb{E}\left[g\left(X_{s_0,x}(s_1), \ldots, X_{s_0,x}(t^*), Y_{t^*,\xi}(\hat{t}_{L-1}), \ldots, Y_{t^*,\xi}(\hat{t}_i)\right) K_\epsilon\left(Y_{t^*,\xi}(T) - X_{s_0,x}(t^*)\right) \frac{\mathcal{Y}_{t^*,\xi,1}(T)}{\varphi(\xi)}\right].$$

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

$$\int_A p(s_0, x, T, y)\lambda_A(dy) = \lim_{\epsilon\downarrow 0} \mathbb{E}\left[K_\epsilon\left(Y_{t^*,\xi}(T) - X_{s_0,x}(t^*)\right)\frac{\mathcal{Y}_{t^*,\xi,1}(T)}{\varphi(\xi)}\right].$$

Proof. Let us abbreviate

$$H_A := \mathbb{E}\left[g\left(X_{s_0,x}(s_1), \ldots, X_{s_0,x}(t_{L-1})\right)\middle| X_{s_0,x}(T) \in A\right],$$
$$H(y) := \mathbb{E}\left[g\left(X_{s_0,x}(s_1), \ldots, X_{s_0,x}(t_{L-1})\right)\middle| X_{s_0,x}(T) = y\right],$$

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, i.e.,

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

Recall (3.11) and the construction (3.10) of λ_A. Then, we have

$$\begin{aligned}
H_A &= \int_A \mathbb{E}\left[g\left(X_{s_0,x}(s_1), \ldots, X_{s_0,x}(t_{L-1})\right)\middle| X_{s_0,x}(T) = y\right] q(y)\lambda_A(dy) \\
&= \int_A H(y)q(y)\lambda_A(dy) \\
&= \mathbb{E}\left[\frac{H(\xi)q(\xi)}{\varphi(\xi)}\right] \\
&= \frac{1}{\int_A p(s_0, x, T, z)\lambda_A(dz)} \mathbb{E}\left[\frac{p(s_0, x, T, \xi)}{\varphi(\xi)} \mathbb{E}\left[g\left(X_{s_0,x}(s_1), \ldots, X_{s_0,x}(t_{L-1})\right)\middle| X_{s_0,x}(T) = \xi\right]\right].
\end{aligned}$$

Hence,

$$\begin{aligned}
H_A \times \int_A p(s_0, x, T, y)\lambda_A(dy) &= \mathbb{E}\Bigg[\frac{1}{\varphi(\xi)} \lim_{\epsilon\downarrow 0} \mathbb{E}^{\xi}\Big[g\left(X_{s_0,x}(s_1), \ldots, X_{s_0,x}(t^*), Y_{t^*,\xi}(\hat{t}_{L-1}), \ldots, Y_{t^*,\xi}(\hat{t}_i)\right) \times \\
&\qquad \times K_\epsilon\left(Y_{t^*,\xi}(T) - X_{s_0,x}(t^*)\right)\mathcal{Y}_{t^*,\xi,1}(T)\Big]\Bigg] \\
&= \lim_{\epsilon\downarrow 0} \mathbb{E}\Big[g\left(X_{s_0,x}(s_1), \ldots, X_{s_0,x}(t^*), Y_{t^*,\xi}(\hat{t}_{L-1}), \ldots, Y_{t^*,\xi}(\hat{t}_i)\right) \times \\
&\qquad \times K_\epsilon\left(Y_{t^*,\xi}(T) - X_{s_0,x}(t^*)\right)\frac{\mathcal{Y}_{t^*,\xi,1}(T)}{\varphi(\xi)}\Big].
\end{aligned}$$

□

Corollary 3.10. *The conditional expectation of $g\left(X_{s_0,x}(s_1), \ldots, X_{s_0,x}(t_{L-1})\right)$ given $X^1_{s_0,x}(T) = c^1 \in \mathbb{R}$ has the stochastic representation*

$$\begin{aligned}
&\lim_{\epsilon\downarrow 0} \mathbb{E}\left[g\left(X_{s_0,x}(s_1), \ldots, X_{s_0,x}(t^*), Y_{t^*,\xi}(\hat{t}_{L-1}), \ldots, Y_{t^*,\xi}(\hat{t}_i)\right) K_\epsilon\left(Y_{t^*,\xi}(T) - X_{s_0,x}(t^*)\right)\frac{\mathcal{Y}_{t^*,\xi,1}(T)}{\varphi(\xi)}\right] = \\
&\qquad \int_{\mathbb{R}^{d-1}} p(s_0, x, T, c^1, y^2, \ldots, y^d)dy^2 \cdots dy^d \mathbb{E}\left[g\left(X_{s_0,x}(s_1), \ldots, X_{s_0,x}(t_{L-1})\right)\middle| X^1_{s_0,x}(T) = c^1\right]
\end{aligned}$$

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

$$\lim_{\epsilon\downarrow 0} \mathbb{E}\left[K_\epsilon\left(Y_{t^*,\xi}(T) - X_{s_0,x}(t^*)\right)\frac{\mathcal{Y}_{t^*,\xi,1}(T)}{\varphi(\xi)}\right] = \int_{\mathbb{R}^{d-1}} p(s_0, x, T, c^1, y^2, \ldots, y^d)dy^2 \cdots dy^d.$$

Remark 3.11. Note that we have never really used that X is a continuous-time diffusion process. In fact, our approach also works for discrete-time Markov chains, following the ideas of [Milstein et al., 2007].

4. Forward-reverse estimators and their analysis

The stochastic representations for the conditional diffusion problem (1.1), 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 (2.2) below 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 4.19.

Remark 4.3. By the results of Kusuoka and Stroock [1985], Cor. (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, i.e., all the derivatives are bounded as well. We know of no similar study for non-autonomous 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 is non-negative and satisfies $\int_{\mathbb{R}^d} K(v)dv = 1$ and $K(-v) = K(v)$ for $v \in \mathbb{R}^d$. 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

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

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 Hesse matrix are bounded. Moreover, the coefficient c in (2.2) is bounded.

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

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

$$(4.1) \quad h_\epsilon \quad := \quad \mathbb{E}\Big[g\left(X_{s_0,x}(s_1), \ldots, X_{s_0,x}(t^*), Y_{t^*,y}(\hat{t}_{L-1}), \ldots, Y_{t^*,y}(\hat{t}_1)\right) \epsilon^{-d} K\left(\frac{Y_{t^*,y}(T) - X_{s_0,x}(t^*)}{\epsilon}\right) \mathcal{Y}_{t^*,y}(T)\Big],$$

which can – and will – be computed using Monte Carlo simulation. By Theorem 3.6, h_ϵ converges to

$$(4.2) \qquad h := p(s_0, x, T, y) \mathbb{E}\left[g\left(X_{s_0,x}(s_1), \ldots, X_{s_0,x}(t_{L-1})\right) \,\middle|\, X_{s_0,x}(T) = y\right].$$

Theorem 4.7. *Assuming Conditions 4.1, 4.4 and 4.5, there are constants $C, \epsilon_0 > 0$ such that the bias of the approximation h_ϵ can be bounded by*

$$|h - h_\epsilon| \le C\epsilon^2, \quad 0 < \epsilon < \epsilon_0.$$

Proof. Changing variables $y_1 \to v := \frac{y_1 - x_K}{\epsilon}$ in Theorem 3.5, we arrive at

$$h_\epsilon = \int g(x_1, \dots, x_K, y_2, \dots, y_L) K(v) \prod_{i=1}^{K} p(s_{i-1}, x_{i-1}, s_i, x_i) \times$$
$$\times p(t^*, x_K + \delta v, t_1, y_2) \prod_{i=2}^{L} p(t_{i-1}, y_i, t_i, y_{i+1}) dx_1 \cdots dx_K dv dy_2 \cdots dy_L.$$

In particular, we have that $h = \lim_{\epsilon \downarrow 0} h_\epsilon$. Consider

$$r_\epsilon(x_K, y_2) := \int k(v) p(t^*, x_K + \epsilon v, t_1, y_2) dv - p(t^*, x_K, t_1, y_2).$$

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, Condition 4.4 and Condition 4.1, we get

$$\begin{aligned} r_\epsilon &= \int K(v) \left[p(t^*, x_K + \epsilon v, t_1, y_2) - p(t^*, x_K, t_1, y_2) \right] dv \\ &= \int K(v) \left[\epsilon \partial_x p(t^*, x_K, t_1, y_2) \cdot v \right] dv + \\ &\quad + \sum_{|\beta|=2} \frac{2}{\beta!} \epsilon^2 \int \int_0^1 (1-t) \partial_x^\beta p(t^*, x_K + t\epsilon v, t_1, y_2) \cdot v^\beta dt K(v) dv \end{aligned}$$

implying that

$$\begin{aligned} |r_\epsilon(x_K, y_2)| &\le \sum_{|\beta|=2} \frac{2}{\beta!} \epsilon^2 \int_0^1 (1-t) \int \left| \partial_x^\beta p(t^*, x_K + t\epsilon v, t_1, y_2) \right| |v|^2 K(v) dv dt \\ &\le \sum_{|\beta|=2} \frac{2}{\beta!} \epsilon^2 C_1 \int_0^1 (1-t) \int e^{-\gamma |y_2 - x_K - t\epsilon v|^2} |v|^2 K(v) dv dt \\ &\le \sum_{|\beta|=2} \frac{2}{\beta!} \epsilon^2 C_1 C_\eta \int_0^1 (1-t) \int e^{-\gamma |y_2 - x_K - t\epsilon v|^2} e^{-\eta |v|^2} dv dt, \end{aligned}$$

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

$$|y_2 - x_K - t\epsilon v|^2 = |y_2 - x_K|^2 - 2t\epsilon \langle y_2 - x_K\,,\, v \rangle + t^2 \epsilon^2 |v|^2,$$

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

$$\begin{aligned} \int e^{-\gamma |y_2 - x_K - t\epsilon v|^2} e^{-\eta |v|^2} dv &= e^{-\gamma |y_2 - x_K|^2} \int e^{2t\gamma\epsilon \langle y_2 - x_K\,,\, v \rangle} e^{-\frac{|v|^2}{2\sigma^2}} dv \\ &= \left(\frac{\eta + \gamma t^2 \epsilon^2}{\pi} \right)^{-d/2} \exp\left(\epsilon^2 \frac{t^2 \gamma^2}{\eta} |y_2 - x_K|^2 \right) e^{-\gamma |y_2 - x_K|^2} \\ &\le \left(\frac{\pi}{\eta} \right)^{d/2} e^{\epsilon^2 \frac{\gamma^2}{\eta} |y_2 - x_K|^2} e^{-\gamma |y_2 - x_K|^2}. \end{aligned}$$

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

$$\begin{aligned} |r_\epsilon(x_K, y_2)| &\le 2\widetilde{C} \epsilon^2 e^{-\gamma |y_2 - x_K|^2} \int_0^1 (1-t) e^{\epsilon^2 \frac{\gamma^2}{\eta} |y_2 - x_K|^2} dt \\ &\le \widetilde{C} \epsilon^2 e^{-\gamma' |y_2 - x_K|^2}, \end{aligned}$$

with $\gamma' = \gamma - \frac{\gamma^2}{\eta} \epsilon^2$, which is positive for $0 < \epsilon < \epsilon_0 := (\eta/\gamma)^{1/2}$. Consequently, for $0 < \epsilon < \epsilon_0$, we can interpret $s_\epsilon(x_K, y_2) := |r_\epsilon(x_K, y_2)|/(C_2 \epsilon^2)$ 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 obtain

$$
\begin{aligned}
|h_\epsilon - h| &\le \int |g(x_1,\dots,x_K,y_2,\dots,y_L)| \prod_{i=1}^{K} p(s_{i-1},x_{i-1},s_i,x_i)\times \\
&\quad \times |r_\epsilon(x_K,y_2)| \prod_{i=2}^{L} p(t_{i-1},y_i,t_i,y_{i+1}) dx_1\cdots dx_K dy_2\cdots dy_L \\
&\le C_2\epsilon^2 \int |g(x_1,\dots,x_K,y_2,\dots,y_L)| \prod_{i=1}^{K} p(s_{i-1},x_{i-1},s_i,x_i)\times \\
&\quad \times s_\epsilon(x_K,y_2) \prod_{i=2}^{L} p(t_{i-1},y_i,t_i,y_{i+1}) dx_1\cdots dx_K dy_2\cdots dy_L \qquad (4.3)\\
&=: C\epsilon^2 < \infty,
\end{aligned}
$$

provided that $0 < \epsilon < \epsilon_0$, as the last expression can be interpreted as

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

for a Markov process Z with transition densities $p(s_{i-1},x_{i-1},s_i,x_i)$, $1 \le i \le K$, $s_\epsilon(x_K,y_2)$, $p(t_{i-1},y_i,t_i,y_{i+1})$, $2 \le i \le L$, 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.

Remark 4.9. If we are in the autonomous setting of [Kusuoka and Stroock, 1985, Corollary (3.25)], we actually only need boundedness of the vector fields together with polynomial bounds for their derivatives. In this case, the bound in Condition 4.1 continues to hold with an extra polynomial term in x, which does not interfere with the above calculations.

In the spirit of Milstein et al. [2004], we now introduce a Monte Carlo estimator $\widehat{h_\epsilon}$ for the quantity h_ϵ introduced in (4.1) by

$$
\begin{aligned}
(4.4)\quad \widehat{h}_{\epsilon,M,N} := \frac{1}{\epsilon^d NM}\sum_{n=1}^{N}\sum_{m=1}^{M} g\left(X^n_{s_0,x}(s_1),\dots,X^n_{s_0,x}(s_K),Y^m_{t^*,y}(\hat{t}_{L-1}),\dots,Y^m_{t^*,y}(\hat{t}_1)\right)\times \\
\times K\left(\frac{Y^m_{t^*,y}(T) - X^n_{s_0,x}(t^*)}{\epsilon}\right)\mathcal{Y}^m_{t^*,y}(T),
\end{aligned}
$$

where the superscripts n and m denote different, independent realizations of the corresponding processes. Moreover, we denote

$$(4.5)\quad Z_{nm} := \frac{1}{\epsilon^d} g\left(X^n_{s_0,x}(s_1),\dots,X^n_{s_0,x}(s_K),Y^m_{t^*,y}(\hat{t}_{L-1}),\dots,Y^m_{t^*,y}(\hat{t}_1)\right) K\left(\frac{Y^m_{t^*,y}(T) - X^n_{s_0,x}(t^*)}{\epsilon}\right)\mathcal{Y}^m_{t^*,y}(T).$$

Note that $\mathbb{E}[Z_{nm}] = h_\epsilon$. We are left to analyze the variance of the estimator $\widehat{h}_{\epsilon,M,N}$. To this end, we consider the expectation $\mathbb{E}[Z_{nm}Z_{n'm'}]$ for various combinations of n, m, n', and m'.

Remark 4.10. 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 4.11. *For $m \neq m'$ we have*

$$
\begin{aligned}
\mathbb{E}[Z_{nm}Z_{nm'}]\big|_{\epsilon=0} = \int g(x_1,\dots,x_K,y_2,\dots,y_L) g(x_1,\dots,x_K,y'_2,\dots,y'_L) p(t^*,x_K,t_1,y_2) p(t^*,x_K,t_1,y'_2)\times \\
\times \prod_{i=1}^{K} p(s_{i-1},x_{i-1},s_i,x_i)dx_i \prod_{i=2}^{L} p(t_{i-1},y_i,t_i,y_{i+1})dy_i \prod_{i=2}^{L} p(t_{i-1},y'_i,t_i,y'_{i+1})dy'_i.
\end{aligned}
$$

Moreover, we have

$$|\mathbb{E}[Z_{nm}Z_{nm'}] - \mathbb{E}\,[Z_{nm}Z_{nm'}]|_{\epsilon=0}| \le C\epsilon^2.$$

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

$$\begin{aligned}
E\left[Z_{nm}Z_{nm'}\right] &= \epsilon^{-2d} E\Big[g\left(X^n_{s_1},\dots,X^n_{s_K},Y^m_{\hat t_{L-1}},\dots,Y^m_{\hat t_1}\right) g\left(X^n_{s_1},\dots,X^n_{s_K},Y^{m'}_{\hat t_{L-1}},\dots,Y^{m'}_{\hat t_1}\right)\times \\
&\times K\left(\frac{Y^m_T - X^n_{t^*}}{\epsilon}\right) K\left(\frac{Y^{m'}_T - X^n_{t^*}}{\epsilon}\right)\mathcal{Y}^m_T\mathcal{Y}^{m'}_T\Big] \\
&= \epsilon^{-2d}\int g(x_1,\dots,x_K,y_2,\dots,y_L) g(x_1,\dots,x_K,y'_2,\dots,y'_L)\times \\
&\times K\left(\frac{y_1-x_K}{\epsilon}\right) K\left(\frac{y'_1-x_K}{\epsilon}\right)\prod_{i=1}^K p(s_{i-1},x_{i-1},s_i,x_i)dx_i\times \\
&\times \prod_{i=1}^L p(t_{i-1},y_i,t_i,y_{i+1})dy_i \prod_{i=1}^L p(t_{i-1},y'_i,t_i,y'_{i+1})dy'_i \\
&= \int g(x_1,\dots,x_K,y_2,\dots,y_L) g(x_1,\dots,x_K,y'_2,\dots,y'_L)\times \\
&\times K(v)K(v')p(t^*,x_K+\epsilon v,t_1,y_2)dv\,p(t^*,x_K+\epsilon v',t_1,y'_2)dv'\times \\
&\times \prod_{i=1}^K p(s_{i-1},x_{i-1},s_i,x_i)dx_i \prod_{i=2}^L p(t_{i-1},y_i,t_i,y_{i+1})dy_i \prod_{i=2}^L p(t_{i-1},y'_i,t_i,y'_{i+1})dy'_i,
\end{aligned}$$

where we changed variables $v := (y_1 - x_K)/\epsilon$ and $v' := (y'_1 - x_K)/\epsilon$. Thus, for $\epsilon = 0$, we arrive at the above expression, which is treated as a problem-dependent constant.

Using Condition 4.4, we now consider

$$\begin{aligned}
r^{(1,2)}_\epsilon(x_K,y_2,y'_2) &:= \int K(v)K(v')[p(t^*,x_K+\epsilon v,t_1,y_2)p(t^*,x_K+\epsilon v',t_1,y'_2)- \\
&- p(t^*,x_K,t_1,y_2)p(t^*,x_K,t_1,y'_2)\big]dvdv' \\
&= \epsilon^2\int_0^1(1-t)\Big[\sum_{i=1}^d\int K(v)K(v')\partial_x^{2e_i}p(t^*,x_K+t\epsilon v,t_1,y_2)p(t^*,x_K+t\epsilon v',t_1,y'_2)v_i^2dvdv'+ \\
&+\sum_{i=1}^d\int K(v)K(v'^*,x_K+t\epsilon v,t_1,y_2)\partial_x^{2e_i}p(t^*,x_K+t\epsilon v',t_1,y'_2)(v'_i)^2dvdv'+ \\
&+2\sum_{i,j=1}^d\int K(v)K(v')\partial_x^{e_i}p(t^*,x_K+t\epsilon v,t_1,y_2)\partial_x^{e_j}p(t^*,x_K+t\epsilon v',t_1,y'_2)v_iv'_jdvdv'\Big]dt,
\end{aligned}$$

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

$$\left|r^{(1,2)}_\epsilon(x_K,y_2,y'_2)\right| \le C s^{(1,2)}_\epsilon(x_K,y_2)s^{(1,2)}_\epsilon(x_K,y'_2)$$

for a transition density $s^{(1,2)}_\epsilon(x_K,y_2)$ with Gaussian bounds. Consequently, we obtain

$$\begin{aligned}
|\mathbb{E}[Z_{nm}Z_{nm'}] - \mathbb{E}\,[Z_{nm}Z_{nm'}]|_{\epsilon=0}| &\le C\epsilon^2\int |g(x_1,\dots,x_K,y_2,\dots,y_L)|\times \\
&\times|g(x_1,\dots,x_K,y'_2,\dots,y'_L)|\prod_{i=1}^K p(s_{i-1},x_{i-1},s_i,x_i)dx_i s^{(1,2)}_\epsilon(x_K,y_2)dy_2\times \\
&\times\prod_{i=3}^L p(t_{i-1},y_i,t_i,y_{i+1})dy_i\times s^{(1,2)}_\epsilon(x_K,y'_2)dy'_2\prod_{i=3}^L p(t_{i-1},y'_i,t_i,y'_{i+1})dy'_i,
\end{aligned}$$

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

$$|\mathbb{E}[Z_{nm}Z_{nm'}] - \mathbb{E}\,[Z_{nm}Z_{nm'}]|_{\epsilon=0}| \le C\epsilon^2\int\widetilde{p}(s_0,x,t^*,x_K)\widetilde{q}(t^*,x_K,T,y)^2dx_K. \tag{4.6}$$

□

When we consider $\mathbb{E}\left[Z_{nm}Z_{n'm}\right]$, we have to take care of terms $\mathcal{Y}_T^2$ appearing in the expectation. To this end, let us introduce

$$\mu(y_1,\ldots,y_L) := \mathbb{E}\left[\mathcal{Y}_T \middle| Y_T = y_1,\ldots,Y_{\hat{t}_1} = y_L\right],$$
$$\mu_2(y_1,\ldots,y_L) := \mathbb{E}\left[\mathcal{Y}_T^2 \middle| Y_T = y_1,\ldots,Y_{\hat{t}_1} = y_L\right].$$

We then have two choices: we could replace $\mathcal{Y}_T^2$ by its conditional expectation $\mu_2\left(Y_{\hat{t}_L},\ldots,Y_{\hat{t}_1}\right)$ and rewrite the expectation as an integral w.r.t. the transition density of the reverse diffusion Y as was done in Milstein et al. [2004] for the case $L = K = 1$, or we can replace $\mu_2\left(Y_{\hat{t}_L},\ldots,Y_{\hat{t}_1}\right)$ by $\frac{\mu_2\left(Y_{\hat{t}_L},\ldots,Y_{\hat{t}_1}\right)}{\mu\left(Y_{\hat{t}_L},\ldots,Y_{\hat{t}_1}\right)}\mu\left(Y_{\hat{t}_L},\ldots,Y_{\hat{t}_1}\right)$ and then write the expectation as an integral w.r.t. the densities p as usual. In the following, we opt for the former approach, and 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 4.12. *For $n \neq n'$ we have*

$$\mathbb{E}\left[Z_{nm}Z_{n'm}\right]|_{\epsilon=0} = \int g\left(x_1,\ldots,x_{K-1},y_1,\ldots,y_L\right) g\left(x'_1,\ldots,x'_{K-1},y_1,\ldots,y_L\right)\times$$
$$\times\, \mu_2(y_1,\ldots,y_L)\prod_{i=1}^{K-1} p(s_{i-1},x_{i-1},s_i,x_i)dx_i \prod_{i=1}^{K-1} p(s_{i-1},x'_{i-1},s_i,x'_i)dx'_i\times$$
$$\times\, p(s_{K-1},x_{K-1},s_K,y_1)p(s_{K-1},x'_{K-1},s_K,y_1)\prod_{i=1}^{L} q(t_{i-1},y_{i+1},t_i,y_i)dy_i.$$

Moreover, there is a constant C such that

$$|\mathbb{E}[Z_{nm}Z_{n'm}] - \mathbb{E}\left[Z_{nm}Z_{n'm}\right]|_{\epsilon=0}| \leq \epsilon^2 C.$$

Proof. By a similar approach as in Lemma 4.11, but changing variables $x_K \to v := (y_1 - x_K)/\epsilon$ and $x'_K \to v' := (y_1 - x'_K)/\epsilon$, we arrive at

$$\mathbb{E}\left[Z_{nm}Z_{n'm}\right] = \int g\left(x_1,\ldots,x_{K-1},y_1-\epsilon v,y_2,\ldots,y_L\right) g\left(x'_1,\ldots,x'_{K-1},y_1-\epsilon v',y_2,\ldots,y_L\right)\times$$
$$\times\, K(v)\,K(v')\,\mu_2(y_1,\ldots,y_L)\prod_{i=1}^{K-1} p(s_{i-1},x_{i-1},s_i,x_i)dx_i \prod_{i=1}^{K-1} p(s_{i-1},x'_{i-1},s_i,x'_i)dx'_i\times$$
$$\times\, p(s_{K-1},x_{K-1},s_K,y_1-\epsilon v)dv\,p(s_{K-1},x'_{K-1},s_K,y_1-\epsilon v')dv'\prod_{i=1}^{L} q(t_{i-1},y_{i+1},t_i,y_i)dy_i.$$

For $\epsilon = 0$, Condition 4.4 implies

$$\mathbb{E}\left[Z_{nm}Z_{n'm}\right]|_{\epsilon=0} = \int g\left(x_1,\ldots,x_{K-1},y_1,\ldots,y_L\right) g\left(x'_1,\ldots,x'_{K-1},y_1,\ldots,y_L\right)\times$$
$$\times\, \mu_2(y_1,\ldots,y_L)\prod_{i=1}^{K-1} p(s_{i-1},x_{i-1},s_i,x_i)dx_i \prod_{i=1}^{K-1} p(s_{i-1},x'_{i-1},s_i,x'_i)dx'_i\times$$
$$\times\, p(s_{K-1},x_{K-1},s_K,y_1)p(s_{K-1},x'_{K-1},s_K,y_1)\prod_{i=1}^{L} q(t_{i-1},y_{i+1},t_i,y_i)dy_i,$$

which gives the formula from the statement of the lemma.

For the bound on the difference, note once again that

$$r_\epsilon^{(2,1)} := \int\Big[g(x_1,\dots,x_{K-1},y_1-\epsilon v,y_2,\dots,y_L)g(x_1',\dots,x_{K-1}',y_1-\epsilon v',y_2,\dots,y_L)\times$$
$$\times\, p(s_{K-1},x_{K-1},s_K,y_1-\epsilon v)p(s_{K-1},x_{K-1}',s_K,y_1-\epsilon v')-$$
$$-\,g(x_1,\dots,x_{K-1},y_1,\dots,y_L)g(x_1',\dots,x_{K-1}',y_1,\dots,y_L)\times$$
$$p(s_{K-1},x_{K-1},s_K,y_1)p(s_{K-1},x_{K-1}',s_K,y_1)\Big]K(v)K(v')dvdv'$$

can be bounded in the sense that $|r_\epsilon^{(2,1)}| \le C s_\epsilon^{(2,1)}(x_{K-1},y_1)s_\epsilon^{(2,1)}(x_{K-1}',y_1)$ for transition densities $s_\epsilon^{(2,1)}$ with Gaussian tails, so that

$$|\mathbb{E}[Z_{nm}Z_{n'm}] - \mathbb{E}\,[Z_{nm}Z_{n'm}]|_{\epsilon=0}| \le C\epsilon^2\int \mu_2(y_1,\dots,y_L)\prod_{i=1}^{K-1}p(s_{i-1},x_{i-1},s_i,x_i)dx_i\times$$
$$\times\prod_{i=1}^{K-1}p(s_{i-1},x_{i-1}',s_i,x_i')dx_i's_\epsilon^{(2,1)}(x_{K-1},y_1)s_\epsilon^{(2,1)}(x_{K-1}',y_1)\prod_{i=1}^{L}q(t_{i-1},y_{i+1},t_i,y_i)dy_i.$$

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

$$q(t_{i-1},y_{i+1},t_i,y_i) \le C^i\exp\left(-\gamma^i|y_{i+1}-y_i|^2\right) =: s_i(y_i,y_{i+1})$$

by a Gaussian transition density s_i which is naturally symmetric. Absorbing $\|\mu_2\|_\infty$ and $\prod_{i=1}^L C^i$ into the constant C and denoting

$$\widetilde{p}(s_0,x,t^*,y_1) := \int\prod_{i=1}^{K-1}p(s_{i-1},x_{i-1},s_i,x_i)dx_i s_\epsilon^{(2,1)}(x_{K-1},y_1),$$
$$\widetilde{q}(t^*,y_1,T,y) := \int\prod_{i=1}^{L}s_i(y_i,y_{i+1})dy_2\cdots dy_L,$$

the Chapman-Kolmogorov equation implies that

$$\begin{aligned}|\mathbb{E}[Z_{nm}Z_{n'm}] - \mathbb{E}\,[Z_{nm}Z_{n'm}]|_{\epsilon=0}| &\le C\epsilon^2\int\widetilde{p}(s_0,x,t^*,y_1)^2\widetilde{q}(t^*,y_1,T,y)dy_1\\ &\le C\epsilon^2\int\widetilde{p}(s_0,x,t^*,y_1)\widetilde{q}(t^*,y_1,T,y)dy_1.\end{aligned}\tag{4.7}$$

□

Remark 4.13. A closer look at the proof of Lemma 4.12 actually reveals that it would have been enough to assume g and its first and second derivatives to be polynomially bounded.

Lemma 4.14. *We have*

$$\epsilon^d E\left[Z_{nm}^2\right] = \int K(v)^2dv\int g(x_1,\dots,x_{K-1},y_1,y_2,\dots,y_L)\mu_2(y_1,y_2,\dots,y_L)\times$$
$$\times\prod_{i=1}^{K-1}p(s_{i-1},x_{i-1},s_i,x_i)p(s_{K-1},x_{K-1},s_K,y_1)\prod_{i=1}^{L}q(t_{i-1},y_{i+1},t_i,y_i)dx_1\cdots dx_{K-1}dy_1dy_2\cdots dy_L.$$

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

$$\left|\epsilon^d\mathbb{E}[Z_{nm}^2] - \lim_{\epsilon\to 0}\epsilon^d\mathbb{E}\left[Z_{nm}^2\right]\right| \le C\epsilon^2.$$

Proof. Substituting $x_K \to v := (y_1 - x_K)/\epsilon$, we obtain

$$\epsilon^d E\left[Z_{nm}^2\right] = \int g(x_1,\dots,x_{K-1},y_1-\epsilon v,y_2,\dots,y_L)\mu_2(y_1,y_2,\dots,y_L)\times \\ \times K(v)^2 \prod_{i=1}^{K-1} p(s_{i-1},x_{i-1},s_i,x_i)p(s_{K-1},x_{K-1},s_K,y_1-\epsilon v)\times \\ \times \prod_{i=1}^{L} q(t_{i-1},y_{i+1},t_i,y_i)\times dx_1\cdots dx_{K-1}dvdy_1dy_2\cdots dy_L.$$

For $\epsilon \to 0$ the right hand side gives the statement from the Lemma.

For the difference, consider

$$r_\epsilon^{(1,1)} := \int K(v)^2\Big[g(x_1,\dots,x_{K-1},y_1-\epsilon v,y_2,\dots,y_L)^2 p(s_{K-1},x_{K-1},s_K,y_1-\epsilon v)- \\ - g(x_1,\dots,x_{K-1},y_1,y_2,\dots,y_L)^2 p(s_{K-1},x_{K-1},s_K,y_1)\Big]dv.$$

Following the procedure established in the previous lemmas, we obtain

$$|r_\epsilon^{(1,1)}| \le C s_\epsilon^{(1,1)}(x_{K-1},y_1),$$

and by the argument used in the proof of Lemma 4.12, we obtain transition densities function $\widetilde{p}(s_0,x,t^*,y_1)$ and $\widetilde{q}(t^*,y_1,T,y)$ such that

$$\left|\epsilon^d\mathbb{E}[Z_{nm}^2] - \lim_{\epsilon\to 0}\epsilon^d\mathbb{E}\left[Z_{nm}^2\right]\right| \le C\epsilon^2\int \widetilde{p}(s_0,x,t^*,y_1)\widetilde{q}(t^*,y_1,T,y)dy_1. \tag{4.8}$$ □

Remark 4.15. In fact, it is enough to assume polynomial bounds for g and its first and second derivatives.

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

- The constant in Theorem 4.7 is denoted by C, i.e., $|h_\epsilon - h| \le C\epsilon^2$;
- for $m \neq m'$, we set $\mathbb{E}[Z_{nm}Z_{nm'}] =: h_\epsilon^{(1,2)}$ and denote the constant for the difference by $C_{1,2}$, i.e., $\left|h_\epsilon^{(1,2)} - h_0^{(1,2)}\right| \le C_{1,2}\epsilon^2$;
- for $n \neq n'$, we set $\mathbb{E}[Z_{nm}Z_{n'm}] =: h_\epsilon^{(2,1)}$ and denote the constant for the difference by $C_{2,1}$, i.e., $\left|h_\epsilon^{(2,1)} - h_0^{(2,1)}\right| \le C_{2,1}\epsilon^2$;
- we set $\epsilon^d\mathbb{E}\left[Z_{nm}^2\right] =: h_\epsilon^{(1,1)}$ and denote the constant for the difference by $C_{1,1}$, i.e., $\left|h_\epsilon^{(1,1)} - h_0^{(1,1)}\right| \le C_{1,1}\epsilon^2$.

Lemma 4.16. *The variance of the estimator is given by*

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

Proof. The result follows immediately by (4.5), independence of Z_{nm} and $Z_{n'm'}$ when both $n \neq n'$ and $m \neq m'$ and the notations introduced above, noting that $E[Z_{nm}] = h_\epsilon$. □

Lemma 4.17. *We assume Conditions 4.1, 4.4 and 4.5 to hold. Then the mean square error of the estimator* $\widehat{h}_{\epsilon,M,N}$ *introduced in* (4.4) *for the term* h *defined in* (4.2) *satisfies*

$$\mathbb{E}\left[\left(\widehat{h}_{\epsilon,M,N} - h\right)^2\right] \le \frac{1-N-M}{NM}h^2 + \frac{M-1}{NM}h_0^{(1,2)} + \frac{N-1}{NM}h_0^{(2,1)} + \frac{\epsilon^{-d}}{NM}h_0^{(1,1)} + \\ + \frac{\epsilon^{-d+2}}{NM}C_{1,1} + \epsilon^2\left[2\frac{1-N-M}{NM}Ch + \frac{M-1}{NM}C_{1,2} + \frac{N-1}{NM}C_{2,1}\right] + \frac{(N-1)(M-1)}{NM}C^2\epsilon^4.$$

Proof. Follows immediately. □

Similarly to Milstein et al. [2004], we can now choose $N = M$ and the bandwith ϵ so as to obtain convergence proportional to $N^{-1/2}$.

Theorem 4.18. *Assume Conditions 4.1, 4.4 and 4.5 and set $M = N$ and $\epsilon = \epsilon_N$ dependent on N.*

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

Proof. Insert $M = N$ and the respective choice of ϵ_N in Lemma 4.17. □

Remark 4.19. By replacing the kernel K by *higher order* kernels[1], 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 4.11, 4.12 and 4.14.

So far, we have only computed the quantity h as given in (4.2). However, finally we want to compute the conditional expectation

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

As $H = \frac{h}{p(s_0,x,T,y)}$ with h defined in (4.2), 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

$$\text{(4.9)} \quad \widehat{H}_{\epsilon,M,N} \coloneqq \frac{\sum_{n=1}^N \sum_{m=1}^M g\left(X^n_{s_1}, \dots, X^n_{s_K}, Y^m_{\hat{t}_{L-1}}, \dots, Y^m_{\hat{t}_1}\right) K\left(\frac{Y^m_T - X^n_{t^*}}{\epsilon}\right) \mathcal{Y}^m_T}{\sum_{n=1}^N \sum_{m=1}^M K\left(\frac{Y^m_T - X^n_{t^*}}{\epsilon}\right) \mathcal{Y}^m_T} \mathbf{1}_{\frac{1}{NM}\epsilon^{-d} \sum_{n=1}^N \sum_{m=1}^M K\left(\frac{Y^m_T - X^n_{t^*}}{\epsilon}\right) \mathcal{Y}^m_T > \overline{p}/2},$$

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.[2]

Theorem 4.20. *Assume Conditions 4.1, 4.4 and 4.5 and set $M = N$ and $\epsilon = \epsilon_N$ dependent on N.*

- *If $d \le 4$, choose $\epsilon_N = CN^{-1/4}$. Then we have $\mathbb{E}\left[\left(\widehat{H}_{\epsilon_N,N,N} - H\right)^2\right] = O(N^{-1})$, so we achieve the optimal convergence rate $1/2$.*
- *For $d > 4$, choose $\epsilon_N = CN^{-2/(4+d)}$, and obtain $\mathbb{E}\left[\left(\widehat{H}_{\epsilon_N,N,N} - H\right)^2\right] = O(N^{-8/(4+d)})$.*

Proof. Let $X_N \coloneqq \widehat{h}_{\epsilon_N,N,N}$, and, similarly, let

$$Y_N \coloneqq \frac{1}{N^2} \epsilon_N^{-d} \sum_{n=1}^N \sum_{m=1}^N K\left(\frac{Y^m_T - X^n_{t^*}}{\epsilon_N}\right) \mathcal{Y}^m_T$$

denote the estimator in the denominator – including the normalization factor. Moreover, let $X \coloneqq h$ as defined in (4.2) and let $Y \coloneqq p(s_0, x, T, y)$. Then we have already established in Theorem 4.18 that

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

where $p = 1$ for $d \le 4$ and $p = \frac{8}{4+d}$ when $d > 4$. Moreover, we have obtained in Lemma 4.16 that $\operatorname{Var} X_N = O(N^{-p})$ and $\operatorname{Var} Y_N = O(N^{-p})$.

[1] Recall that the order of a kernel K is the order of the lowest order (non-constant) monomial f such that $\int f(v)K(v)dv \neq 0$.

[2] 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 4.21 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.

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

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{X_N}{Y_N} - \frac{X}{Y}\right)^2 \mathbf{1}_{Y_N > D_N}\right] &= \mathbb{E}\left[\frac{(X_N Y - Y_N X)^2}{(Y_N Y)^2}\mathbf{1}_{Y_N > D_N}\right] \\
&\le \frac{\mathbb{E}\left[(Y(X_N - X) + X(Y - Y_N))^2\right]}{Y^2 D_N^2} \\
&\le 2\frac{Y^2\mathbb{E}[(X_N - X)^2] + X^2\mathbb{E}\left[(Y - Y_N)^2\right]}{Y^2 D_N^2} \\
&\le \frac{C^1_{X,Y}}{D_N^2 N^p},
\end{aligned} \tag{4.10}$$

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,

$$\begin{aligned}
\mathbb{P}(Y_N \le D_N) &= \mathbb{P}(Y_N - \mathbb{E}Y_N \le D_N - \mathbb{E}Y_N; Y_N \le \mathbb{E}Y_N) \\
&\le \mathbb{P}(|Y_N - \mathbb{E}Y_N| \ge \mathbb{E}Y_N - D_N) \\
&\le \frac{\operatorname{Var} Y_N}{(\mathbb{E}Y_N - D_N)^2} \\
&\le \frac{C^2_Y}{(\mathbb{E}Y_N - D_N)^2 N^p}.
\end{aligned} \tag{4.11}$$

Finally, consider

$$\begin{aligned}
\mathbb{E}\left[\left(\zeta_N - \frac{X}{Y}\right)^2\right] &= \mathbb{E}\left[\left(\zeta_N - \frac{X}{Y}\mathbf{1}_{Y_N > D_N} - \frac{X}{Y}\mathbf{1}_{Y_N \le D_N}\right)^2\right] \\
&= \mathbb{E}\left[\left(\zeta_N - \frac{X}{Y}\mathbf{1}_{Y_N > D_N}\right)^2\right] + \frac{X^2}{Y^2}\mathbb{P}(Y_N \le D_N) \\
&\le \frac{C^1_{X,Y}}{D_N^2 N^p} + \frac{C^2_Y X^2}{(\mathbb{E}Y_N - D_N)^2 Y^2 N^p},
\end{aligned} \tag{4.12}$$

where we have combined (4.10) and (4.11). Now choose $D_N := \overline{p}/2$ for N large enough. As $\mathbb{E}Y_N \xrightarrow{N\to\infty} Y$, (4.12) implies that

$$\mathbb{E}\left[\left(\zeta_N - \frac{X}{Y}\right)^2\right] = O(N^{-p}). \tag{4.13}$$

□

Remark 4.21. Alternatively, we could replace the cut-off $\overline{p}/2$ in (4.9) by some sequence $D_N \xrightarrow{N\to\infty} 0$. In that case, the MSE of the estimator is of order $O(N^{-p}/D_N^2)$, which can be chosen as close to $O(N^{-p})$ 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 4.20, as $\mathbb{E}Y_N \xrightarrow{N\to\infty} p(s_0, x, T, y) > 0$ by assumption.

Forward-reverse estimators for conditioning on a set. In Theorem 3.9 and Corollary 3.10 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 $X_T^1 = y^1, \ldots, X_T^{d'} = y^{d'}$. 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, i.e., we treat both cases distinguished above together.

Recall that we represented the conditional expectation as

$$\lim_{\epsilon\downarrow 0}\mathbb{E}\Big[g\left(X_{s_0,x}(s_1),\dots,X_{s_0,x}(s_K),Y_{t^*,\xi}(\hat t_{L-1},\dots Y_{t^*,\xi}(\hat t_1)\right)\epsilon^{-d}K\left(\frac{Y_{t^*,\xi}(T)-X_{s_0,x}(t^*)}{\epsilon}\right)\frac{\mathcal{Y}_{t^*,\xi,1}(T)}{\varphi(\xi)}\Big]=$$
$$=\int_A p(s_0,x,T,y)\lambda_A(dy)\mathbb{E}\left[g\left(X_{s_0,x}(s_1),\dots,X_{s_0,x}(t_{L-1})\right)\,\middle|\,X_{s_0,x}(T)\in A\right],$$

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

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

We define the following Monte Carlo estimator for the conditional expectation:

$$\text{(4.14)}\quad \widehat{H}^{\xi}_{\epsilon,M,N}:=\frac{\sum_{n=1}^N\sum_{m=1}^M g\left(X^n_{s_1},\dots,X^n_{s_K},Y^m_{\hat t_{L-1}},\dots,Y^m_{\hat t_1}\right)K\left(\frac{Y^m_T-X^n_{t^*}}{\epsilon}\right)\frac{\mathcal{Y}^m_T}{\varphi(\xi^m)}}{\sum_{n=1}^N\sum_{m=1}^M K\left(\frac{Y^m_T-X^n_{t^*}}{\epsilon}\right)\frac{\mathcal{Y}^m_T}{\varphi(\xi^m)}}\times$$
$$\times\mathbf{1}_{\frac{1}{NM}\epsilon^{-d}\sum_{n=1}^N\sum_{m=1}^M K\left(\frac{Y^m_T-X^n_{t^*}}{\epsilon}\right)\frac{\mathcal{Y}^m_T}{\varphi(\xi^m)}>\overline{p}/2},$$

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 $(Y^m_{\hat t_{L-1}},\dots,Y^m_{\hat t_1})$ together with $\mathcal{Y}^m_T$, $1\le m,\le M$, are independent samples from the reverse process $(Y,\mathcal{Y})$ started at $Y^m_{t^*}=\xi^m$, $\mathcal{Y}^m_{t^*}=1$, for an independent sequence of samples ξ^m from the distribution ξ. Apart from the term $\varphi(\xi^m)$, the difference to estimator (4.9) is the randomness of the initial values of the reverse process. Again, $p(s_0,x,T,y)>\overline{p}>0$, and Remark 4.21 applies. The analysis of (4.14), however, works along the lines of the analysis of (4.9). Indeed, in all the expectations considered in Theorem 4.7 and in Lemma 4.11–4.14, 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, with constants depending on the value of ξ;
3. Move $\varphi(\xi)^{-1}$ back in and take the expectation in ξ.

Theorem 4.23. *Set $M=N$ and assume Condition 4.22 and, as usual, Condition 4.1, 4.4 and 4.5.*

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

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

$$h^{\xi}:=\int_A p(s_0,x,T,y)\lambda_A(dy)\cdot\mathbb{E}\left[g\left(X_{s_0,x}(s_1),\dots,X_{s_0,x}(t_{L-1})\right)\middle|X_{s_0,x}(T)\in A\right]$$
$$h^{\xi}_{\epsilon}:=\mathbb{E}\Big[g\left(X_{s_0,x}(s_1),\dots,X_{s_0,x}(t^*),Y_{t^*,\xi}(\hat t_{L-1}),\dots,Y_{t^*,\xi}(\hat t_i)\right)K_\epsilon\left(Y_{t^*,\xi}(T)-X_{s_0,x}(t^*)\right)\frac{\mathcal{Y}_{t^*,\xi,1}(T)}{\varphi(\xi)}\Big],$$
$$Z^{\xi}_{nm}:=\frac{1}{\epsilon^d}g\left(X^n_{s_0,x}(s_1),\dots,X^n_{s_0,x}(s_K),Y^m_{t^*,\xi^m}(\hat t_{L-1}),\dots,Y^m_{t^*,\xi^m}(\hat t_1)\right)K\left(\frac{Y^m_{t^*,\xi^m}(T)-X^n_{s_0,x}(t^*)}{\epsilon}\right)\frac{\mathcal{Y}^m_{t^*,\xi^m}(T)}{\varphi(\xi^m)},$$

and notice that the result will follow if we can establish the bounds of Theorem 4.7 and Lemma 4.11, 4.12 and 4.14 for h, h_ϵ and Z_{nm} replaced by h^ξ, h^ξ_ϵ and Z^ξ_{nm}, respectively.

For the bias, (4.3) implies a bound $|h(y) - h_\epsilon(y)| \leq C\epsilon^2 \widetilde{p}(s_0, x, T, y)$ for some density $\widetilde{p}$ in y, where we make the dependence of h and h_ϵ on y explicit. Consequently, conditioning on ξ first, we have

$$
\begin{aligned}
|h^\xi - h^\xi_\epsilon| &= \left| \mathbb{E}\left[\frac{h(\xi) - h_\epsilon(\xi)}{\varphi(\xi)} \right] \right| \\
&\leq \mathbb{E}\left[\frac{|h(\xi) - h_\epsilon(\xi)|}{\varphi(\xi)} \right] \\
&\leq C\epsilon^2 \int \frac{\widetilde{p}(s_0, x, T, \xi)}{\varphi(\xi)} \varphi(\xi) d\xi \\
&\leq C\epsilon^2. \qquad (4.15)
\end{aligned}
$$

By the same approach, using the estimate from Lemma 4.11, denoting $z_{n,m,m'}(y, y') := \mathbb{E}\left[Z_{nm} Z_{nm'}\right]$, where we assume $Y^m = Y^m_{t^*,y}$ and $Y^{m'} = Y^{m'}_{t^*,y'}$, we get, using a simple adaptation of (4.6) for different terminal values y and y',

$$
\begin{aligned}
\left| \mathbb{E}\left[Z^\xi_{nm} Z^\xi_{nm'} \right] - \mathbb{E}\left[Z^\xi_{nm} Z^\xi_{nm'} \right] \Big|_{\epsilon=0} \right| &\leq \mathbb{E}\left[\frac{\left| z_{n,m,m'}(\xi^m, \xi^{m'}) - z_{n,m,m'}(\xi^m, \xi^{m'}) \big|_{\epsilon=0} \right|}{\varphi(\xi^m)\varphi(\xi^{m'})} \right] \\
&\leq C\epsilon^2 \mathbb{E}\left[\frac{\int \widetilde{p}(s_0, x, t^*, x_K) \widetilde{q}(t^*, x_K, T, \xi^m) \widetilde{q}(t^*, x_k, T, \xi^{m'}) dx_K}{\varphi(\xi^m)\varphi(\xi^{m'})} \right] \\
&= C\epsilon^2 \int \widetilde{p}(s_0, x, t^*, x_K) \widetilde{q}(t^*, x_K, T, y) \widetilde{q}(t^*, x_k, T, y') dx_K \lambda_A(dy) \lambda_A(dy') \\
&\leq C\epsilon^2. \qquad (4.16)
\end{aligned}
$$

Adopting the above notation for the case $n \neq n'$ covered in Lemma 4.12 and using (4.7), we get

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

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

$$
\left| \mathbb{E}\left[Z^\xi_{nm} Z^\xi_{n'm} \right] - \mathbb{E}\left[Z^\xi_{nm} Z^\xi_{n'm} \right] \Big|_{\epsilon=0} \right| \leq C\epsilon^2. \qquad (4.17)
$$

In a similar way, using (4.8), we get the bound

$$
\left| \epsilon^d \mathbb{E}\left[(Z^\xi_{nm})^2 \right] - \lim_{\epsilon \to 0} \epsilon^d \mathbb{E}\left[(Z^\xi_{nm})^2 \right] \right| \leq C\epsilon^2. \qquad (4.18)
$$

The respective versions of Lemma 4.16, Lemma 4.17 and Theorem 4.18 follow immediately from the bounds (4.15), (4.16), (4.17) and (4.18), and we can repeat the proof of Theorem 4.20, arriving at the conclusion. □

5. Numerical study

5.1. **Implementation.** Some care is necessary when implementing the forward reverse estimators (4.9) and (4.14) 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 et al. [2004] for a similar discussion.

We here assume that the kernel $K(x)$ used in (4.9) and (4.14), 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×TOL outside the ball.[3] Then, due to the small support of K, the following Monte Carlo algorithm for the kernel estimator is feasible. For simplicity, we take $N = M$ and assume that the time-grid $s_0 < \cdots < s_K = t^* = t_0 < t_1 < \cdots < t_L = T$ is uniform. (We present the algorithm only for the case of (4.9), the analysis being virtually equal for (4.14).) The complexity of the simulation steps (2) and (3) in Algorithm 1 is $O(KNd)$ and $O(LNd)$

Algorithm 1 Forward-reverse algorithm

1: **procedure** ForRev(N, ϵ_N, a, σ, x, y, $\mathcal{D}$, t^*, g)
2: Simulate N trajectories $(X^n_{s_0,x})^N_{n=1}$ of the forward process on $\mathcal{D} \cap [s_0, t^*]$.
3: Simulate N trajectories $(Y^m_{t^*,y}, \mathcal{Y}^m_{t^*,y})^N_{m=1}$ of the reverse process on $\widehat{\mathcal{D}} \cap [t^*, T]$.[4]
4: **for** $m \leftarrow 1, N$ **do**
5: Find the sub-sample

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

6: **end for**
7: Evaluate (4.9) by

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

8: **end procedure**

elementary computations, respectively. The size l_m of the intersection in (5) is, on average, proportional to $N\epsilon_N^d \times p\left(s_0, x, t^*, Y^m_{t^*,y}(T)\right)$. The search procedure itself can be done at a cost of order $O(N \log N)$, see for instance Greengard and Strain [1991] where this is proved in the context of the Gauss transform. Hence, in the case $d \le 4$ we may achieve root-N accuracy[5] by choosing $\epsilon_N = (N/\log N)^{-1/d}$, implying a total cost of the forward-reverse algorithm of $O(N \log N)$ for an accuracy of $\mathrm{TOL} \sim N^{-1/2}$. I.e., the complexity C needed for an error tolerance TOL is $O(\mathrm{TOL}^{-2})$.

In the case $d > 4$, we can achieve the same complexity estimate relying on higher order kernels, see Remark 4.19. Otherwise, the choice of $\epsilon_N = N^{-\frac{2}{4+d}}$ leads to an accuracy of $\mathrm{TOL} \sim N^{-\frac{2}{4+d}}$ at a work $O(N \log N)$. Put differently, the complexity C needed for an error tolerance TOL is $O\left(|\log(\mathrm{TOL})|\,\mathrm{TOL}^{-(4+d)/4}\right)$.

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

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

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 $L + K = l$.

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

[4] Note that for a uniform grid $\mathcal{D}$, we have $\widehat{\mathcal{D}} = \mathcal{D}$.

[5] We assume that we can simulate from the forward and reverse processes exactly at constant cost. 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.

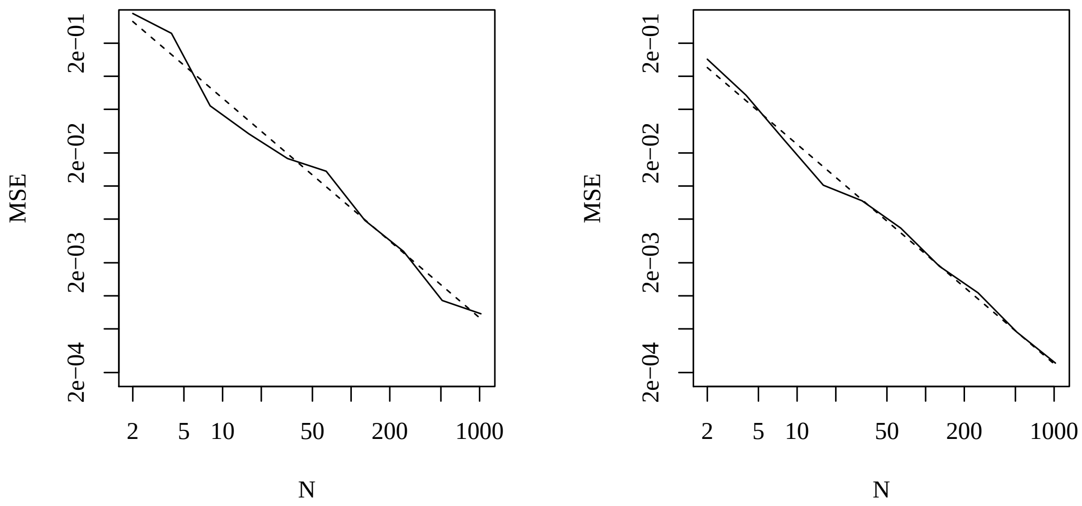
Example 5.1. We consider $X_t = B_t$, a two-dimensional standard Brownian motion, which we condition on starting at $X_0 = 0$ and ending at $X_1 = 0$, i.e., the conditioned diffusion is a classical two-dimensional Brownian bridge. In particular, the reverse process Y_t is also a standard Brownian motion, and $\mathcal{Y} \equiv 1$. We consider the functional

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

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

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

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 $\epsilon = \epsilon_N = N^{-0.4}$, which still gives the rate of convergence obtained in Theorem 4.20.



(A) $t^* = 1/10$ (B) $t^* = 4/10$

Figure 1. MSE for Example 5.1. Dashed lines are reference lines proportional to N^{-1}.

In Figure 1, we show the results for $l = 10$, with the choices $K = 1$ and $K = 4$, i.e., with $t^* = 1/10$ and $t^* = 4/10$, respectively. In both case, we observe the asymptotic relation MSE $\sim N^{-1}$ predicted by Theorem 4.20. 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.[6]

Example 5.2. Let us consider the stock price S_t in a Heston model: $X_t \coloneqq (S_t, v_t)$, i.e., the stock price together with its (stochastic) volatility satisfies the stochastic differential equation

$$\begin{aligned} dS_t &= \mu S_t dt + \sqrt{v_t} S_t dB_t^1, \\ dv_t &= (\gamma v_t + \beta) dt + \xi \sqrt{v_t} \left(\rho dB_t^1 + \sqrt{1 - \rho^2} dB_t^2 \right). \end{aligned}$$

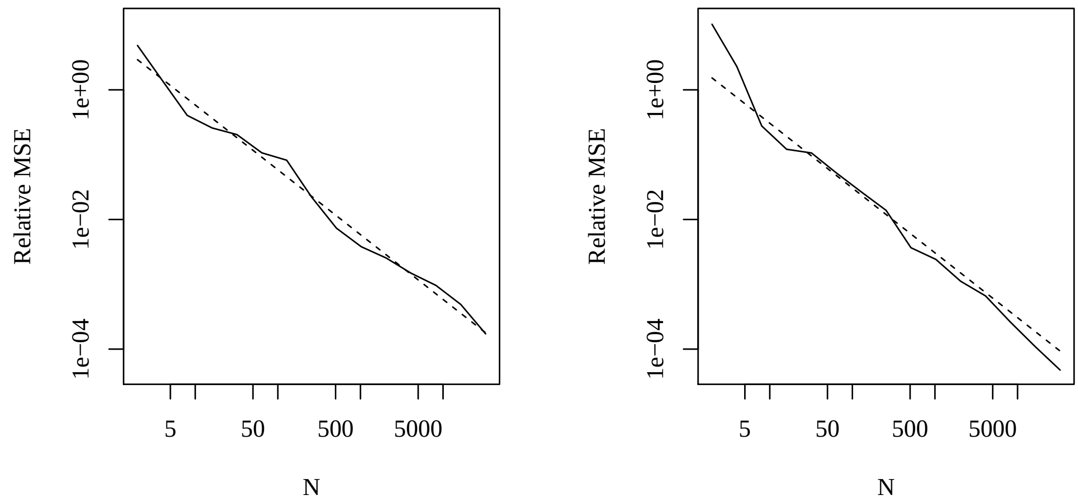
[6] Cf. Milstein et al. [2004], where it was noted that the variance of the forward-reverse density estimator explodes when $t^* \to T$ or $t^* \to 0$. Mathematically, this is a consequence of the transition densities getting singular. For our estimator, the analogous effect would happen when the mesh of the grid tends to 0, the position of t^* within the grid does not matter. More precisely, the estimates in Lemma 4.11, 4.12 and 4.14 explode when $|s_K - t^*| \to 0$ and/or $|t_1 - t^*| \to 0$.

We have

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

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

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



(A) $t^* = 15/360$ (B) $t^* = 29/360$

Figure 2. Relative MSE for Example 5.2. Dashed lines are reference lines proportional to N^{-1}.

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

$$g(x_1, \ldots, x_{l-1}, x_l) := \sum_{i=1}^{l-1} \left(\log(x^1_{i+1}) - \log(x^1_i)\right)^2.$$

(Dependence of the functional g on the final value y obviously changes nothing in the theorems of Section 3 or Section 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 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 4.22.

Contrary to Example 5.1, 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 $h = \min\left(1/360, \sqrt{0.05/N}\right)$, so that the MSE for the solution of the corresponding SDE is itself $O(N^{-1})$, implying that the asymptotic order of the MSE of our quantity of interest is not effected 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 $\epsilon_N = (4N)^{-0.4}$, so that we are in the regime of Theorem 4.23.

The numerical results in Figure 2 confirm the rate of convergence for the MSE established in Theorem 4.23. 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", i.e., the MSE normalized by the squared reference value.

Appendix A. The non-autonomous case in Theorem 3.3

We here prove Theorem 3.3 for the non-autonomous case based on the statement already proved for the autonomous case. Let us consider for each $\delta \neq 0$ the process $\mathcal{U}_\delta = (\Lambda_\delta, \mathcal{X}_\delta)$ in $\mathbb{R} \times \mathbb{R}^d$, governed by the SDE

$$d\mathcal{U}_\delta = \begin{pmatrix} 1 \\ a\left(\Lambda^\delta, \mathcal{X}^\delta\right) \end{pmatrix} dt + \begin{pmatrix} \delta & \emptyset \\ \emptyset & \sigma(\Lambda^\delta, \mathcal{X}^\delta) \end{pmatrix} d\mathcal{W}$$

$$=: \mathfrak{a}(\mathcal{U}^\delta)dt + \mathfrak{s}_\delta(\mathcal{U}^\delta)d\mathcal{W} \tag{A.1}$$

with $\mathcal{U}^\delta(0) = u = (\lambda, x)$, where $\mathcal{W} = [w, W]^\top$ with w being a new independent scalar Brownian motion, $\mathfrak{a}(z) = [1, a(s,x')]^\top$ for $z = (z^0, z^1, ..., z^d) = (s, x') \in \mathbb{R} \times \mathbb{R}^d$, and a similar definition for $\mathfrak{s}_\delta(z)$. The unique strong solution of (A.1) is in full denoted by

$$\mathcal{U}^\delta_{0,u}(t) = \mathcal{U}^\delta_{0,\lambda,x}(t) = \left(\Lambda^\delta_{0,\lambda}(t), \mathcal{X}^\delta_{0,\lambda,x}(t)\right), \quad t \geq 0,$$

where obviously $\Lambda^\delta_{0,\lambda}(t) = \lambda + t + \delta w_t$. Clearly the process $\mathcal{U}^\delta$ is autonomous and satisfies Condition 4.1 for any $\delta \neq 0$. In order to apply Theorem 3.3 for the autonomous case to (A.1) we now build its (autonomous) reverse process,

$$dZ^\delta = \mathfrak{a}_*(Z^\delta)dt + \mathfrak{s}_\delta(Z^\delta)d\mathcal{W}, \quad Z^\delta(t^*) = v = (\mu, y) \in \mathbb{R} \times \mathbb{R}^d,$$
$$d\mathcal{Z}^\delta = \mathfrak{c}(Z^\delta)\mathcal{Z}^\delta dt, \quad \mathcal{Z}^\delta(t^*) = 1,$$

where

$$\mathfrak{a}^i_*(z) = -\mathfrak{a}^i(z) + \sum_{j=0}^d \frac{\partial}{\partial z^j}\mathfrak{b}^{ij}_\delta(z), \quad i = 0, \ldots, d, \quad \mathfrak{b}_\delta := \mathfrak{s}_\delta \mathfrak{s}_\delta^\top$$
$$\mathfrak{c}(z) = \frac{1}{2}\sum_{i,j=0}^d \frac{\partial^2}{\partial z^i \partial z^j}\mathfrak{b}^{ij}_\delta(z) - \sum_{i=0}^d \frac{\partial}{\partial z^i}\mathfrak{a}^i(z),$$

and independent standard Brownian motion $\widetilde{\mathcal{W}} = \left(\widetilde{w}, \widetilde{W}\right)^\top$ in $\mathbb{R} \times \mathbb{R}^d$. Spelled out we have,

$$\mathfrak{a}^0_*(z) = -1,$$
$$\mathfrak{a}^i_*(z) = -a^i(z) + \sum_{j=1}^d \frac{\partial}{\partial z^j}b^{ij}(z) = \alpha^i(z), \quad i = 1, \ldots, d,$$
$$\mathfrak{c}(z) = \frac{1}{2}\sum_{i,j=1}^d \frac{\partial^2}{\partial z^i \partial z^j}b^{ij}(z) - \sum_{i=1}^d \frac{\partial}{\partial z^i}\mathfrak{a}^i(z) = c(z)$$

with $z = (z^0, z^1, ..., z^d) = (s, x')$, $\alpha(z) = \alpha(s, x')$ and $c(z) = c(s, x')$ defined in (2.2). Note that $\mathfrak{a}_*$ and $\mathfrak{c}$ do not depend on δ. It thus holds that

$$Z^{\delta;0}_{t^*,\mu,y}(t) = \mu + t^* - t + \delta(\widetilde{w}_t - \widetilde{w}_{t^*})$$
$$Z^{\delta;i}_{t^*,\mu,y}(t) = y + \int_{t^*}^t \alpha^i(Z^\delta_{t^*,\mu,y}(s))ds + \int_{t^*}^t \sigma^i(Z^\delta_{t^*,\mu,y}(s))d\widetilde{W}, \ \ i = 1, ..., d,$$
$$\mathcal{Z}^\delta_{t^*,\mu,y}(t) = \exp\left[\int_{t^*}^t c(Z^\delta_{t^*,\mu,y}(s))ds\right],$$

where σ^i denotes the i-th row of σ.

Now let $\widetilde{f}:\mathbb{R}^{(d+1)\times L}\to\mathbb{R}$ be a function of the form

$$\widetilde{f}(z_0,\dots,z_{L-1})=f(y_0,\dots,y_{L-1})$$

with $z_j=(z_j^0,y_j)$ in $\mathbb{R}\times\mathbb{R}^d$. Let $\mathfrak{p}^\delta(t,u,s,v)$ be the (autonomous) transition density of the process $\mathcal{U}^\delta$. Let us further denote $\mathbf{Y}^{\delta;i}_{t^*,\mu,y}:=Z^{\delta;i}_{t^*,\mu,y}$, for $i=1,\dots,d$. Then by taking $\lambda=0$, $\mu=T$, application of Theorem 3.3 for the autonomous case yields

$$\begin{aligned}(*)_\delta &:= \mathbb{E}\left[f(\mathbf{Y}^\delta_{t^*,T,y}(T),\mathbf{Y}^\delta_{t^*,T,y}(\widehat{t}_{L-1}),\dots,\mathbf{Y}^\delta_{t^*,T,y}(\widehat{t}_1))\mathcal{Z}^\delta_{t^*,T,y}(T)\right]\\ &=\mathbb{E}\left[\widetilde{f}(Z^\delta_{t^*,y}(T),Z^\delta_{t^*,y}(\widehat{t}_{L-1}),\dots,Z^\delta_{t^*,y}(\widehat{t}_1))\mathcal{Z}^\delta_{t^*,T,y}(T)\right]\\ &=\int_{\mathbb{R}^{(d+1)\times L}}\widetilde{f}(z_0,...,z_{L-1})\prod_{i=1}^{L}\mathfrak{p}^\delta(t_{i-1},z_{i-1},t_i,z_i)dz_{i-1}\\ &=\int_{\mathbb{R}^{(d+1)\times L}}f(y_0,\dots,y_{L-1})\prod_{i=1}^{L}\mathfrak{p}^\delta(t_{i-1},(z^0_{i-1},y_{i-1}),t_i,(z_i^0,y_i))d(z^0_{i-1},y_{i-1})\end{aligned}$$

We are going to show that

$$\lim_{\delta\to0}(*)_\delta=\int_{\mathbb{R}^{d\times L}}f(y_0,...,y_{L-1})dy_0\cdots dy_{L-1}\prod_{i=1}^{L}p(t_{i-1},y_{i-1},t_i,y_i).$$

Consider with $z^0_L=T=t_L$, $y_L=y$,

$$\begin{aligned}(*)_\delta &=\int_{\mathbb{R}^{(d+1)}}\mathfrak{p}^\delta(t_{L-1},(z^0_{L-1},y_{L-1}),T,(T,y))d(z^0_{L-1},y_{L-1})\times\\ &\quad\times\int_{\mathbb{R}^{(d+1)\times(L-1)}}f(y_0,...,y_{L-1})\prod_{i=1}^{L-1}\mathfrak{p}^\delta(t_{i-1},(z^0_{i-1},y_{i-1}),t_i,(z^0_i,y_i))d(z^0_{i-1},y_{i-1})\\ &=:\int_{\mathbb{R}^{(d+1)}}g^\delta_{L-1}(z^0_{L-1},y_{L-1})\mathfrak{p}^\delta(t_{L-1},(z^0_{L-1},y_{L-1}),T,(T,y))d(z^0_{L-1},y_{L-1})\\ &=\mathbb{E}\left[g^\delta_{L-1}(Z^{\delta,0}_{t_{L-1},T,y}(T),\mathbf{Y}^\delta_{t_{L-1},T,y}(T))\mathcal{Z}^\delta_{t_{L-1},T,y}(T)\right],\end{aligned}$$

where reverse process $(Z^{\delta,0}_{t_{L-1},T,y},\mathbf{Y}^\delta_{t_{L-1},T,y},\mathcal{Z}^\delta_{t_{L-1},T,y})$ is defined according to (2.2) for the interval $[t_{L-1},T]$. Note that

$$\begin{aligned}\lim_{\delta\downarrow0}\mathbb{E}\left[g^\delta_{L-1}(Z^{\delta,0}_{t_{L-1},T,y}(T),\mathbf{Y}^\delta_{t_{L-1},T,y}(T))\mathcal{Z}^\delta_{t_{L-1},T,y}(T)\right]&=\lim_{\delta\downarrow0}\lim_{\delta'\downarrow0}\mathbb{E}\left[g^\delta_{L-1}(Z^{\delta',0}_{t_{L-1},T,y}(T),\mathbf{Y}^{\delta'}_{t_{L-1},T,y}(T))\mathcal{Z}^{\delta'}_{t_{L-1},T,y}(T)\right]\\ &=\lim_{\delta\downarrow0}\mathbb{E}\left[g^\delta_{L-1}(t_{L-1},Y_{t_{L-1},y}(T))\mathcal{Y}_{t_{L-1},y}(T)\right]\end{aligned}$$

since

$$\mathbb{E}\left[f(\mathbf{Y}^\delta_{t^*,z^0_{L-1},y_{L-1}}(t_{L-1}),\mathbf{Y}^\delta_{t^*,z^0_{L-1},y_{L-1}}(\widehat{t'}_{L-2}),\dots,\mathbf{Y}^\delta_{t^*,z^0_{L-1},y_{L-1}}(\widehat{t'}_1),y_{L-1})\mathcal{Z}^\delta_{t^*,z^0_{L-1},y_{L-1}}(t_{L-1})\right]=g^\delta_{L-1}(z^0_{L-1},y_{L-1})$$

is (jointly) continuous in δ, y_{L-1} and z^0_{L-1}, where $\widehat{t'_i}=t_{L-1}+t^*-t_{L-i-1}$, $i=1,\dots,L-1$. So

$$\lim_{\delta\downarrow0}(*)_\delta=\lim_{\delta\downarrow0}\int g^\delta_{L-1}(t_{L-1},x_{L-1})p(t_{L-1},x_{L-1},T,y)dx_{L-1}.$$

We conclude the proof by a (backward) induction argument: assume that we have already proved that

$$\begin{aligned}\lim_{\delta\downarrow0}(*)_\delta&=\int\prod_{i=k+1}^{L-1}p(t_i,y_i,t_{i+1},y_{i+1})dy_i\times\\ &\qquad\times\lim_{\delta\to0}\int g^\delta_k(z^0_k,y_k,y_{k+1},\dots,y_{L-1})\mathfrak{p}^\delta(t_k,(z^0_k,y_k),t_{k+1},(t_{k+1},y_{k+1}))dz^0_kdy_k\end{aligned}$$

for $1\le k\le L-1$ and using the notation that $y_L=y$, where

$$g^\delta_k(z^0_k,y_k,y_{k+1},\dots,y_{L-1}):=\int f(y_0,\dots,y_{L-1})\prod_{i=1}^{k}\mathfrak{p}^\delta(t_{i-1},(z^0_{i-1},y_{i-1}),t_i,(z^0_i,y_i))dz^0_{i-1}dy_{i-1}.$$

Induction will give us the final result of Theorem 3.3, if we can now prove the corresponding formula for k replaced by $k-1$. As g_k^δ has the stochastic (reverse) representation

$$g_k^\delta(z_k^0, y_k, y_{k+1}, \ldots, y_{L-1}) = \mathbb{E}\Big[f\Big(\mathbf{Y}^\delta_{t^*,(z_k^0,y_k)}(t_k), \mathbf{Y}^\delta_{t^*,(z_k^0,y_k)}(t_k - (t_1 - t^*)), \ldots, \mathbf{Y}^\delta_{t^*,(z_k^0,y_k)}(t_k - (t_{k-1} - t^*)), y_k, \ldots, y_{L-1}\Big)\mathcal{Z}^\delta_{t^*,(z_k^0,y_k)}(t_k)\Big]$$

due to the reverse process defined for the interval $[t^*, t_k]$, we see that $g_k^\delta(z_k^0, y_k, y_{k+1}, \ldots, y_{L-1})$ is continuous in (δ, z_k^0, y_k), uniformly in $y_{k+1}, \ldots, y_{L-1}$. Consequently, we may again split up the limit of $\delta \to 0$ and use the reverse representation to get

$$\begin{aligned}
\lim_{\delta\downarrow 0} (*)_\delta &= \int \prod_{i=k+1}^{L-1} p(t_i, y_i, t_{i+1}, y_{i+1}) dy_i \times \\
&\quad \times \lim_{\delta\to 0} \lim_{\delta'\to 0} \mathbb{E}\Big[g_k^\delta\Big(Z^{\delta',0}_{t_k,(t_{k+1},y_{k+1})}(t_{k+1}), \mathbf{Y}^{\delta'}_{t_k,(t_{k+1},y_{k+1})}(t_{k+1}), y_{k+1}, \ldots, y_{L-1}\Big)\mathcal{Z}^{\delta'}_{t_k,(t_{k+1},y_{k+1})}\Big] \\
&= \int \prod_{i=k+1}^{L-1} p(t_i, y_i, t_{i+1}, y_{i+1}) dy_i \lim_{\delta\to 0} \mathbb{E}\Big[g_k^\delta\Big(t_k, Y_{t_k,y_{k+1}}(t_{k+1}), y_{k+1}, \ldots, y_{L-1}\Big)\mathcal{Y}_{t_k,y_{k+1}}\Big] \\
&= \int \prod_{i=k+1}^{L-1} p(t_i, y_i, t_{i+1}, y_{i+1}) dy_i \lim_{\delta\to 0} \int g_k^\delta\Big(t_k, y_k, y_{k+1}, \ldots, y_{L-1}\Big) p(t_k, y_k, t_{k+1}, y_{k+1}) dy_k \\
&= \int \prod_{i=k}^{L-1} p(t_i, y_i, t_{i+1}, y_{i+1}) dy_i \times \\
&\quad \times \lim_{\delta\to 0} \int g_{k-1}^\delta(z_{k-1}^0, y_{k-1}, y_k, \ldots, y_{L-1}) \mathfrak{p}^\delta(t_{k-1}, (z_{k-1}^0, y_{k-1}), t_k, (t_k, y_k)) dz_{k-1}^0 dy_{k-1},
\end{aligned}$$

which concludes the induction step, and thus the proof of the theorem.

References

D. G. Aronson. Bounds for the fundamental solution of a parabolic equation. *Bull. Amer. Math. Soc.*, 73:890–896, 1967.

A. Beskos, O. Papaspiliopoulos, and G. O. Roberts. Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli*, 12(6):1077–1098, 2006.

A. Beskos, G. Roberts, A. Stuart, and J. Voss. MCMC methods for diffusion bridges. *Stoch. Dyn.*, 8 (3):319–350, 2008.

M. Bladt and M. Sørensen. Simple simulation of diffusion bridges with application to likelihood inference for diffusions. Preprint, 2012. URL `http://www.math.ku.dk`.

P. Cattiaux and L. Mesnager. Hypoelliptic non-homogeneous diffusions. *Probab. Theory Related Fields*, 123(4):453–483, 2002.

B. Delyon and Y. Hu. Simulation of conditioned diffusion and application to parameter estimation. *Stochastic Process. Appl.*, 116(11):1660–1675, 2006.

L. Greengard and J. Strain. The fast Gauss transform. *SIAM J. Sci. Statist. Comput.*, 12(1):79–94, 1991.

M. Hairer, A. Stuart, and J. Voß. Sampling conditioned diffusions. In *Trends in stochastic analysis*, volume 353 of *London Math. Soc. Lecture Note Ser.*, pages 159–185. Cambridge Univ. Press, Cambridge, 2009.

M. Hairer, A. M. Stuart, and J. Voss. Sampling conditioned hypoelliptic diffusions. *Ann. Appl. Probab.*, 21(2):669–698, 2011.

U. G. Haussmann and É. Pardoux. Time reversal of diffusions. *Ann. Probab.*, 14(4):1188–1205, 1986.

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

M. Lin, R. Chen, and P. Mykland. On generating Monte Carlo samples of continuous diffusion bridges. *J. Amer. Statist. Assoc.*, 105(490):820–838, 2010.

T. J. Lyons and W. A. Zheng. On conditional diffusion processes. *Proc. Roy. Soc. Edinburgh Sect. A*, 115(3-4):243–255, 1990.

G. N. Milstein and M. V. Tretyakov. Evaluation of conditional Wiener integrals by numerical integration of stochastic differential equations. *J. Comput. Phys.*, 197(1):275–298, 2004.

G. N. Milstein, J. Schoenmakers, and V. Spokoiny. Transition density estimation for stochastic differential equations via forward-reverse representations. *Bernoulli*, 10(2):281–312, 2004.

G. N. Milstein, J. Schoenmakers, and V. Spokoiny. Forward and reverse representations for Markov chains. *Stochastic Process. Appl.*, 117(8):1052–1075, 2007.

O. Papaspiliopoulos and G. Roberts. Importance sampling techniques for estimation of diffusion models. In *Statistical methods for stochastic differential equations*, volume 124 of *Monogr. Statist. Appl. Probab.*, pages 311–340. CRC Press, Boca Raton, FL, 2012.

L. C. G. Rogers and D. Williams. *Diffusions, Markov processes, and martingales. Vol. 2*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2000.

P. Stinis. Conditional path sampling for stochastic differential equations through drift relaxation. *Commun. Appl. Math. Comput. Sci.*, 6(1):63–78, 2011.

A. M. Stuart, J. Voss, and P. Wiberg. Fast communication conditional path sampling of SDEs and the Langevin MCMC method. *Commun. Math. Sci.*, 2(4):685–697, 2004.