ÖAW Austrian Academy of Sciences

Doctoral Program
Computational Mathematics
Numerical Analysis and Symbolic Computation

JKU JOHANNES KEPLER UNIVERSITY LINZ

Importance Sampling Techniques for Stochastic Partial Differential Equations

Evelyn Buckwar Andreas Thalhammer

DK-Report No. 2017-01 01 2017

A–4040 LINZ, ALTENBERGERSTRASSE 69, AUSTRIA

Supported by

Austrian Science Fund (FWF) Upper Austria

FWF Der Wissenschaftsfonds.

LAND OBERÖSTERREICH

DK sponsors:

- **Johannes Kepler University Linz** (JKU)
- **Austrian Science Fund** (FWF)
- **Upper Austria**

IMPORTANCE SAMPLING TECHNIQUES FOR STOCHASTIC PARTIAL DIFFERENTIAL EQUATIONS

EVELYN BUCKWAR AND ANDREAS THALHAMMER

ABSTRACT. In this work we consider importance sampling techniques for stochastic partial differential equations (SPDE) based on an infinite dimensional version of the well-known Girsanov theorem. For this we develop construction techniques for measure transformations that can be used in numerical experiments for variance reduction purposes of the infinite dimensional equations. The key advantage of the proposed methods is that these techniques are independent of the (spatial) discretisation of the SPDE and that the difference in the computational effort between simulating the numerical trajectories for the standard Monte Carlo estimator and for the importance sampling methods is only the approximation of a linear, one-dimensional SODE. Besides the analysis of the infinite dimensional framework, various approximation and implementation issues are discussed. We conclude by presenting numerical experiments showing the effectiveness of the proposed techniques due to a remarkable reduction of the Monte Carlo error.

1. INTRODUCTION

In many fields of science, stochastic partial differential equations (SPDEs) are used to model problems with uncertainties. Since the importance of such SPDE-based models increases constantly, it is necessary to develop efficient numerical methods in order to be able to cope with the computational complexity of the numerical experiments. In many of these applications, the quantities of interest are the statistics of the underlying solution process at a given time T, i.e. we want to estimate the quantity $\mathbb{E}[\varphi(X(T))]$ for a sufficiently smooth mapping $\varphi : H \to B$, where H and B are separable Hilbert spaces and $X(T)$ denotes the H-valued mild solution of a semilinear SPDE. Due to the fact that it is in general not possible to compute the expectation of $\varphi(X(T))$ explicitly, we have to rely on suitable estimators. Thus, besides issues concerning the spatial and temporal discretisation of the SPDE and the approximation of the noise, we have to additionally *discretise the probability space*, when we want to estimate the expectation.

The goal of this work is to derive and implement numerical methods that enhance the efficiency of the standard Monte Carlo estimator given by

$$\mathbb{E}[\varphi(X(T))] \approx E_M[\varphi(X(T))] := \frac{1}{M}\sum_{i=1}^{M} \varphi(X^{(i)}(T)),$$

2000 *Mathematics Subject Classification.* 60H15, 35R60, 65C05.

Key words and phrases. Importance sampling, Variance reduction techniques, Stochastic partial differential equations, Monte Carlo methods.

The second author was supported by the Austrian Science Fund (FWF): W1214-N15, project DK14.

where $X^{(i)}(T)$ are M independent realisations of the mild solution $X(T)$ defined in (2.4) below. For the resulting approximation error, also called the Monte Carlo error, the following result can be found in [4, Lemma 4.1]: For a random variable $Y \in L^2(\Omega; B)$, i.e. $\mathbb{E}[\|Y\|_B^2] < \infty$, the Monte Carlo error can be expressed by

$$\|\mathbb{E}[Y] - E_M[Y]\|_{L^2(\Omega;B)} = \frac{1}{\sqrt{M}} \operatorname{Var}_B[Y]^{1/2}, \tag{1.1}$$

where the variance in the Hilbert space B is defined as $\operatorname{Var}_B[Y] := \mathbb{E}[\|Y - \mathbb{E}[Y]\|_B^2]$. Thus, increasing the number of samples M obviously leads to a decrease in the Monte Carlo error. However, this is not always suitable due to the high computational cost of the involved numerical realisations and the rather slow convergence rate of the Monte Carlo estimator (with respect to the number of samples M) attaining only order $\mathcal{O}(M^{-1/2})$.

Especially, if we are dealing with approximations of mild solutions of SPDEs, an increase of the sample size is severely limited because of the complexity of the simulations. Thus, in order to make Monte Carlo estimators more efficient, we apply appropriate variance reduction techniques. These are Monte Carlo-type methods using instead of the original quantity of interest Y independent realisations of a random variable $\widetilde{Y}$ satisfying $\mathbb{E}[Y] = \mathbb{E}[\widetilde{Y}]$ and $\operatorname{Var}_B[\widetilde{Y}] \ll \operatorname{Var}_B[Y]$, see e.g. [16, 9] for an overview of variance reduction techniques in the finite dimensional setting. In this article, we develop measure transformations of the underlying probability measure $\mathbb{P}$ based on an infinite dimensional version of the Girsanov theorem in order to reduce the variance of the transformed random variables $\widetilde{Y}$. This variance reduction technique is called *importance sampling* and has already been successfully applied to SODEs, see [16, 14, 1].

Note that importance sampling in the context of variance reduction techniques for SPDEs has already been treated in the literature, see e.g. [1, 15]. There are also optimal measure transformations, where optimality has to be understood in the sense that the corresponding measure transformation leads to a vanishing Monte Carlo error, see e.g. Section 4.1 of this article and [15] for SPDEs driven by a finite number of standard Wiener processes. However, in both cases it is highly non-trivial to compute these optimal measure transformations or at least to find a sufficiently good approximation of it. Thus, such techniques are not optimal from a computational point of view.

In this work, we construct families of infinite dimensional measure transformations that can be used for variance reduction purposes in numerical simulations of $\mathbb{E}[\varphi(X(T))]$. These measure transformations are constructed in a way such that they can easily be implemented and that the computational effort to simulate the numerical trajectories of the transformed process is just slightly higher than the cost for simulating the original solution trajectories. Theoretical optimality results such as the one presented in Section 4.1 are used as guidance for the design of such families of measure transformations and in particular motivate the use of path-dependent transformations. As numerical experiments in Section 6 show, the variance can be significantly reduced by using the considered transformations and thus, the Monte Carlo error based on the proposed importance sampling techniques is considerably reduced.

The authors in [1] presented an importance sampling technique based on finite dimensional considerations: First, the underlying SPDE (the one-dimensional stochastic heat equation) is spatially discretised by using finite differences on a fixed

spatial mesh and afterwards importance sampling is applied to the resulting system of SODEs. In contrast, we apply the proposed importance sampling techniques directly to the infinite dimensional SPDE to reduce the variance of the original quantity of interest. Afterwards any spatial discretisation scheme can be applied to the transformed equation. This has the big advantage that the importance sampling method is independent of the chosen spatial discretisation scheme and consequently we do not have to exploit special structures of the finite dimensional SODE systems.

The flow chart in Figure 1 summarises the importance sampling techniques for SPDEs driven by infinitely many standard Wiener processes. In this work, we develop the approach following the bold arrows. The dashed arrows indicate alternative methods that have already been treated in the literature.

Besides importance sampling there are also other variance reduction techniques for SPDEs, e.g. control variates or multilevel Monte Carlo methods. Especially the latter methods, see e.g. [2, 4], reduce the complexity of estimating $\mathbb{E}[\varphi(X(T))]$ remarkably and we want to emphasize that these methods could be further enhanced by a coupling with the proposed importance sampling techniques from this article.

The outline of this paper is as follows: In Section 2 we briefly discuss the general framework and in Section 3, we recall basic results for measure transformations for semilinear SPDEs based on the infinite dimensional Girsanov theorem for Q-Wiener processes. In Section 3.1, we present a method how such measure transformations can be constructed. We are discussing in Section 4 how these measure transformations can be used in order to reduce the Monte Carlo error. In an optimal benchmark setting, we are able to prove the existence of an optimal measure transformation in Section 4.1, for which the Monte Carlo error vanishes completely. In Section 5, we provide a framework that enables us to couple the importance sampling techniques with spatial and temporal discretisation schemes. Finally, we conclude this article by presenting numerical experiments in Section 6, where we could achieve a remarkable improvement in the efficiency of the standard Monte Carlo estimator.

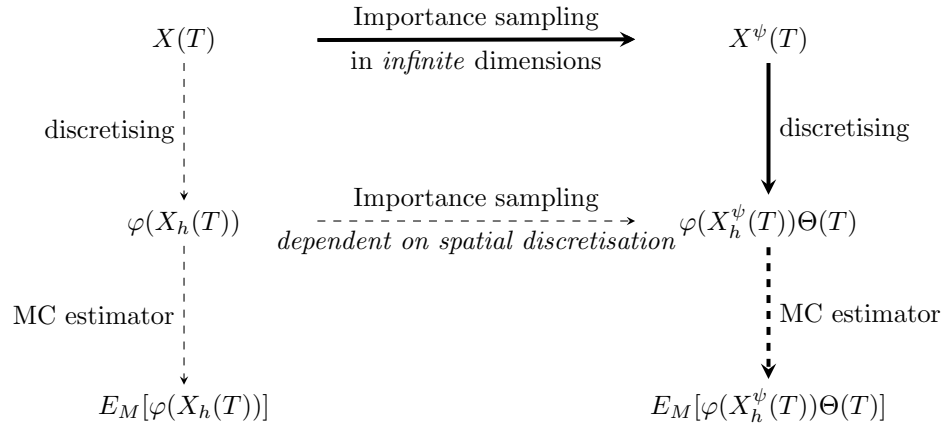

FIGURE 1. Overview of importance sampling techniques for stochastic partial differential equations driven by Q-Wiener processes.

2. Framework

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P})$ denote a complete probability space with a filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$ satisfying the usual conditions. In this article we consider semilinear stochastic partial differential equations (SPDEs) on a separable Hilbert space $(H, \langle\cdot,\cdot\rangle_H)$ given by

$$\mathrm{d}\,X(t) = [AX(t) + F(X(t))]\,\mathrm{d}\,t + G(X(t))\,\mathrm{d}\,W(t), \qquad X(0) = X_0. \tag{2.1}$$

Here A is assumed to be a densely defined, linear, symmetric and positive definite operator $A : \mathrm{dom}(A) \subset H \to H$ and it is assumed to be the generator of an analytic semigroup $S(t), t \in [0,T]$. The domain of the fractional powers of the operator A, which we denote by $\dot{H}^r := \mathrm{dom}((-A)^{r/2})$, $r \in (0,\infty)$, endowed with the inner product

$$\langle\cdot,\cdot\rangle_r := \langle(-A)^{r/2}\cdot, (-A)^{r/2}\cdot\rangle_H$$

also form separable Hilbert spaces, for details see [11] and the references therein. Using this notation, we assume that the initial value satisfies $X_0 \in L^2(\Omega, \dot{H}^1)$, where for any Hilbert space H

$$L^2(\Omega, H) := \{v : \Omega \to H \mid v \text{ is strongly measurable, } \|v\|^2_{L^2(\Omega;H)} := \mathbb{E}[\|v\|^2_H] < \infty\}.$$

Furthermore, let W be an H-valued Q-Wiener process with covariance operator $Q \in L(H)$ being a symmetric, non-negative definite, trace class operator. We denote by $\{e_k\}_{k\in\mathbb{N}}$ the eigenfunctions of Q, which form an orthonormal basis of H, see e.g. [11]. Due to the trace class property of the covariance operator, the corresponding eigenvalues $\{\lambda_k\}_{k\in\mathbb{N}}$ are summable, i.e. $\mathrm{tr}(Q) = \sum_{k=1}^{\infty}\lambda_k < \infty$. Based on these properties the Q-Wiener process can be represented by (see e.g. [7, 18, 13] and the references therein)

$$W(t) = \sum_{k=1}^{\infty}\sqrt{\lambda_k}e_k\beta_k(t), \tag{2.2}$$

where $\{\beta_k(t), t \in [0,T]\}_{k\in\mathbb{N}}$ is a sequence of independent, real-valued Wiener processes. Additionally, since Q is non-negative definite, the square-root of the operator Q is well-defined, i.e. for all $\phi \in H$

$$Q^{1/2}\phi = \sum_{k=1}^{\infty}\sqrt{\lambda_k}\langle\phi, e_k\rangle_H e_k \tag{2.3}$$

exists. In the following part, we denote by $Q^{-1/2}$ the pseudo-inverse of $Q^{1/2}$, see [18, Appendix C]. Using the representation of the square-root operator $Q^{1/2}$ given in Equation (2.3), we define the space $\mathcal{H}_0 := Q^{1/2}(H)$, which together with the norm $|\cdot|_0$ induced by the inner product

$$\langle\phi_1,\phi_2\rangle_0 = \langle Q^{-1/2}\phi_1, Q^{-1/2}\phi_2\rangle_H = \sum_{k=1}^{\infty}\frac{1}{\lambda_k}\langle\phi_1,e_k\rangle_H\langle\phi_2,e_k\rangle_H, \qquad \phi_1,\phi_2 \in \mathcal{H}_0,$$

forms a separable Hilbert space, see [8], with orthonormal basis $\{\sqrt{\lambda_k}e_k\}_{k\in\mathbb{N}}$.

Moreover, we denote by $L_{HS}(\mathcal{H}_0, H)$ the space of all Hilbert-Schmidt operators mapping from $\mathcal{H}_0$ to H. Finally, we assume that the drift operator $F : H \to H$ and the diffusion operator $G : H \to L_{HS}(\mathcal{H}_0, H)$ satisfy the following Lipschitz continuity and linear growth condition, see [4, Assumption 2.2]:

Assumption 2.1. Let $Z = H, \dot{H}^1$. Assume that there exist constants $C_1, C_2 > 0$ such that for all $\phi \in Z, \phi_1, \phi_2 \in H$ it holds that

$$\|F(\phi)\|_Z + \|G(\phi)\|_{L_{HS}(\mathcal{H}_0,Z)} \leq C_1(1+\|\phi\|_Z),$$
$$\|F(\phi_1) - F(\phi_2)\|_H + \|G(\phi_1) - G(\phi_2)\|_{L_{HS}(\mathcal{H}_0,H)} \leq C_1\|\phi_1 - \phi_2\|_H.$$

If Assumption 2.1 is fulfilled, then Equation (2.1) has a unique H-valued mild solution $X(t)$ (see [11, Theorem 2.25]), i.e. there exists an $\mathcal{F}_t$-predictable, H-valued stochastic process $X : [0,T] \to H$ such that for all $t \in [0,T]$ it holds $\mathbb{P}$-a.s that

$$\sup_{t\in[0,T]} \|X(t)\|_{L^2(\Omega;H)} < \infty$$

and

$$X(t) = S(t)X_0 + \int_0^t S(t-s)F(X(s))\ \mathrm{d}\, s + \int_0^t S(t-s)G(X(s))\ \mathrm{d}\, W(s). \tag{2.4}$$

3. Measure transformations for SPDEs

In this section we first recall basic results on measure transformations for SPDEs. Afterwards we present in Section 3.1 a technique how these measure transformations can be constructed such that approximations of the involved transformation operators can be efficiently implemented for numerical simulations. Motivated by examples of finite dimensional measure transformations for SODE systems, we derive explicit representations of two transformations in infinite dimensions for which numerical experiments are presented in Section 6.

The considered measure transformations are based on the following infinite dimensional version of the Girsanov theorem, see [7, Theorem 10.14]:

Theorem 3.1. *Assume that $\psi(t), t \in [0,T]$, is an $\mathcal{H}_0$-valued $\mathcal{F}_t$-predictable process such that the density process*

$$\Theta(t) = \exp\left(\int_0^t \langle\psi(s),\ \mathrm{d}\, W(s)\rangle_0 - \frac{1}{2}\int_0^t |\psi(s)|_0^2\ \mathrm{d}\, s\right)$$

is a martingale. Then, the process

$$\widehat{W}(t) = W(t) - \int_0^t \psi(s)\ \mathrm{d}\, s, \qquad t \in [0,T] \tag{3.1}$$

is a Q-Wiener process with respect to $\{\mathcal{F}_t\}_{t\in[0,T]}$ on the probability space $(\Omega, \mathcal{F}, \widehat{\mathbb{P}})$, where $\mathrm{d}\,\widehat{\mathbb{P}} = \Theta(T)\ \mathrm{d}\,\mathbb{P}$.

If the *weight operator* $\psi(t), t \in [0,T]$, is chosen appropriately according to Theorem 3.1, then one can apply the corresponding measure transformation to SPDE (2.1). Thus, we obtain with respect to the (transformed) probability space $(\Omega, \mathcal{F}, \widehat{\mathbb{P}})$ the following SPDE

$$\begin{aligned} X^\psi(t) - X_0 &= \int_0^t [AX^\psi(s) + F(X^\psi(s))]\ \mathrm{d}\, s + \int_0^t G(X^\psi(s))\ \mathrm{d}\,\widehat{W}(s) \\ &= \int_0^t [AX^\psi(s) + F(X^\psi(s)) - G(X^\psi(s))\psi(s)]\ \mathrm{d}\, s + \int_0^t G(X^\psi(s))\mathrm{dW(s)}. \end{aligned} \tag{3.2}$$

Since the coefficients F and G satisfy Assumption 2.1, there exists a unique mild solution $X^{\psi}(t)$ of the transformed SPDE (3.2) (with respect to the new probability measure $\widehat{\mathbb{P}}$) given by

$$\begin{aligned} X^{\psi}(t) = S(t)X_0 &+ \int_0^t S(t-s)[F(X^{\psi}(s)) - G(X^{\psi}(s))\psi(s)] \, \mathrm{d}\, s \\ &+ \int_0^t S(t-s)G(X^{\psi}(s)) \, \mathrm{d}\, W(s). \end{aligned} \tag{3.3}$$

Note that $\Theta(t)$ can be represented in terms of the sequence of independent, one-dimensional Wiener processes from the Karhunen-Loève expansion (2.2), i.e.

$$\Theta(t) = \exp\left(\sum_{k=1}^{\infty} \int_0^t \frac{1}{\sqrt{\lambda_k}} \langle \psi(s), e_k \rangle_H \, \mathrm{d}\, \beta_k(s) - \frac{1}{2} \sum_{k=1}^{\infty} \int_0^t \frac{1}{\lambda_k} \langle \psi(s), e_k \rangle_H^2 \, \mathrm{d}\, s \right).$$

Thus, the density process $\Theta(t)$ can be seen as the solution of the one-dimensional SDE driven by infinitely many independent one-dimensional Wiener processes, i.e.

$$\begin{aligned} \mathrm{d}\, \Theta(t) &= \Theta(t) \langle \psi(t), \, \mathrm{d}\, W(t) \rangle_0 = \Theta(t) \left(\sum_{k=1}^{\infty} \frac{1}{\sqrt{\lambda_k}} \langle \psi(s), e_k \rangle_H \, \mathrm{d}\, \beta_k(t) \right), \\ \Theta(0) &= 1. \end{aligned} \tag{3.4}$$

Unfortunately, the martingale property of the density process $\Theta(t)$ in Theorem 3.1 is not fulfilled for arbitrary choices of $\psi(t)$ since in general $\Theta(t)$ is only a supermartingale and for this reason $\mathbb{E}[\Theta(T)] \leq 1$, see [12, Appendix I]. However, under the condition

$$\mathbb{P}\left(\int_0^T |\psi(s)|_0^2 \, \mathrm{d}\, s < \infty \right) = 1 \tag{3.5}$$

the density process can be represented by a one-dimensional stochastic exponential, see [7, Lemma 10.15], given by

$$\Theta(t) = \exp\left(\int_0^t |\psi(s)|_0 \, \mathrm{d}\, \beta(s) - \frac{1}{2} \int_0^t |\psi(s)|_0^2 \, \mathrm{d}\, s \right) \quad \text{for } t \in [0,T], \tag{3.6}$$

where $\beta(t), t \in [0,T]$, denotes a one-dimensional Wiener process. If condition (3.5) is fulfilled, then it is sufficient to show that the stochastic exponential (3.6) is a martingale. Due to the fact that the considered stochastic process from Equation (3.6) is now one-dimensional and real-valued, standard results from finite dimensional stochastic calculus can be applied. Hence, the martingale property of (3.6) can be verified e.g. by Novikov's condition, see e.g. [17, 19],

$$\mathbb{E}\left[\exp\left(\frac{1}{2} \int_0^t |\psi(s)|_0^2 \, \mathrm{d}\, s \right) \right] < \infty. \tag{3.7}$$

3.1. **Constructing measure transformations.** In this section, we present a method to construct weight operators $\psi(t), t \in [0,T]$, that fulfil the rather strong assumptions of the Girsanov theorem for infinite dimensional SDEs driven by Q-Wiener processes as stated in Theorem 3.1. Recall that a suitable weight operator ψ has to fulfil that it is $\mathcal{H}_0$-valued, it is $\mathcal{F}_t$-predictable and the induced density process $\Theta(t)$ has to be a martingale. The proposed construction method looks as follows:

At first, we start by choosing an $\mathcal{F}_t$-predictable process $\Psi(t)$ taking values in H as an initial guess for the weight operator, where the particular choice of Ψ might be motivated by already existing finite dimensional examples. In the next step Ψ is projected onto $\mathcal{H}_0$ by using the orthogonal projector $P_{\mathcal{H}_0}$ given by

$$\psi(s) = P_{\mathcal{H}_0}\Psi(s) = \sum_{k=1}^{\infty} \lambda_k \langle \Psi(s), e_k\rangle_H e_k \quad \text{for all } s \in [0,T]. \tag{3.8}$$

Since the density process can be represented by the one-dimensional stochastic exponential (3.6) under Condition (3.5), we want to construct ψ in a way such that this condition is fulfilled. A sufficient condition for (3.5) to hold is that the weight operator is uniformly bounded in the $\mathcal{H}_0$-norm, i.e. there exists a constant $C_0 > 0$ such that $\sup_{s\in[0,T]} |\psi(s)|_0 \leq C_0 < \infty$. If the weight operator is uniformly bounded then the martingale property of $\Theta(t)$ follows directly from Novikov's condition (3.7). If we cannot show the uniform boundedness of the weight operator ψ (with respect to the $\mathcal{H}_0$-norm) we have to find a suitable truncation of the initial process Ψ denoted by $\bar{\Psi}$ such that there exists a constant $C_1 > 0$ with

$$\sup_{k\in\mathbb{N}} \sup_{s\in[0,T]} |\langle \bar{\Psi}(s), e_k\rangle_H| \leq C_1 < \infty.$$

Then, we obtain

$$|\psi(s)|_0^2 = \sum_{k=1}^{\infty} \lambda_k \langle \bar{\Psi}(s), e_k\rangle_H^2 \leq C_1^2 \operatorname{tr}(Q) < \infty.$$

In the next part we present two examples of such measure transformations that are not only of theoretical interest but can also been implemented for numerical experiments, see Section 6. For this, we specify for these two examples the Hilbert space H to be $L^2(D)$, D being a bounded domain. Furthermore, we assume that the eigenfunctions $\{e_k\}_{k\in\mathbb{N}}$ of the covariance operator Q are uniformly bounded by a constant $C_e > 0$, i.e. on the bounded spatial domain D there exists a constant $C_e > 0$ such that for all $j \in \mathbb{N}$ it holds that $\|e_j\|_{L^\infty(D)} \leq C_e < \infty$. An example of an SPDE that fulfils all these assumptions is given e.g. by the one-dimensional stochastic heat equation with multiplicative noise defined in Section 6.

3.1.1. *Time-constant weight operator.* The idea of this weight operator is based on choosing ψ constant in time, which is motivated by the importance sampling technique presented in [1] in a finite dimensional setting. For this let $\{\psi_k\}_{k\in\mathbb{N}}$ be a bounded, real-valued sequence, i.e. there exists a constant $M > 0$ such that $|\psi_k| \leq M$ for all $k \in \mathbb{N}$. Then, we define

$$\psi(s) = \sum_{k=1}^{\infty} \lambda_k \psi_k \langle 1, e_k\rangle_H e_k \tag{3.9}$$

for all $s \in [0,T]$, which corresponds to the generalisation of the case $\psi_k = \bar{\psi} \in \mathbb{R}$ for all $k \in \mathbb{N}$, i.e. the projection of the constant function (that is generally not in $\mathcal{H}_0$) onto $\mathcal{H}_0$. The weight operator (3.9) is uniformly bounded in the $\mathcal{H}_0$-norm, since it holds for all $k \in \mathbb{N}$ that

$$|\langle \psi_k, e_k\rangle_H| = |\langle \psi_k, e_k\rangle_{L^2(D)}| \leq C_e M |D|,$$

where $|D|$ denotes the volume of the bounded spatial domain D. For this reason, Condition (3.5) is automatically fulfilled and therefore, the corresponding density process

$$\Theta(t) = \exp\left(\sum_{k=1}^{\infty} \sqrt{\lambda_k}\psi_k \langle 1, e_k\rangle_H \beta_k(t) - \frac{t}{2}\sum_{k=1}^{\infty} \lambda_k \psi_k^2 \langle 1, e_k\rangle_H^2\right)$$

is a martingale by Novikov's condition (3.7). Thus, all conditions of Theorem 3.1 are fulfilled and the measure transformation based on the weight operator from Equation (3.9) can be applied.

3.1.2. *Truncated path-dependent weight operator.* We want to construct a weight operator $\psi : [0,T] \times H \to \mathcal{H}_0$ that also depends on the pathwise behaviour of the mild solution of the underlying SPDE. Note that such path-dependent weight operators and the corresponding measure transformations are commonly used for finite dimensional problems as it is indicated e.g. in [10, 15].

For constructing such path-dependent weight operators we specify in this example that the diffusion operator is chosen to be a Nemytskii operator induced by a Lipschitz continuous function $\gamma : \mathbb{R} \to \mathbb{R}$, i.e. $(G(u)v)[x] := \gamma(u(x))v(x)$ for all $v \in \mathcal{H}_0$ and $x \in D$.

For the choice of the initial process $\bar{\Psi}$, we consider a bounded approximation $\bar{G}$ of the Nemytskii operator G constructed by a bounded Lipschitz continuous approximation $\bar{\gamma}$ (e.g. by truncation) of the corresponding Lipschitz continuous function γ, i.e. there exists a constant $C_G > 0$ such that $\bar{\gamma} : \mathbb{R} \to [-C_G, C_G]$. Note that for diffusion operators that are already induced by a bounded, Lipschitz continuous function γ, e.g. $\gamma(x) = \sin(x)$, $\bar{G}$ could also be chosen as G. Examples of the construction of such truncated operators $\bar{G}$ are presented in Section 6.

Note that the operator $\bar{G}(\phi)$ is for all $\phi \in H$ a Hilbert-Schmidt operator mapping from $\mathcal{H}_0$ to H, since

$$\begin{aligned}\|\bar{G}(\phi)\|^2_{L_{HS}(\mathcal{H}_0,H)} &= \sum_{j=1}^{\infty} \|\bar{G}(\phi)\sqrt{\lambda_j}e_j\|_H^2 = \sum_{j=1}^{\infty} \lambda_j \int_D |\bar{\gamma}(\phi(x))e_j(x)|^2 \,\mathrm{d}\,x \\ &\leq C_G^2 \sum_{j=1}^{\infty} \lambda_j \|e_j\|_H^2 = C_G^2 \,\mathrm{tr}(Q) < \infty.\end{aligned}$$

In the next step, we choose an element $\eta \in \mathcal{H}_0$ such that there exists a constant $C_\eta > 0$ with $\|\eta\|_H \leq C_\eta$ and set $\bar{\Psi}(s) := G(X^\psi(s))\eta$. Since $\bar{\Psi}(s)$ is H-valued, we have to project $G(X^\psi(s))\eta$ onto $\mathcal{H}_0$. Thus, we define the truncated path-dependent weight operator as

$$\psi(s) = \psi(s, X^\psi(s)) := c\sum_{k=1}^{\infty} \lambda_k \langle \bar{G}(X^\psi(s))\eta, e_k\rangle_H e_k, \tag{3.10}$$

where the scaling parameter $c \in \mathbb{R}$ is introduced in order to control the intensity of the measure transformation in numerical experiments, see Section 6. Note that instead of using a single scaling parameter c one could also introduce a sequence $\{c_k\}_{k\in\mathbb{N}}$, where each of the elements c_k denotes a separate scaling for each of the summands in (3.10). However, for notational convenience we restrict ourselves to

the single parameter setting in this article. With this choice, we obtain

$$\begin{aligned}|\psi(s,X^{\psi}(s))|_0^2 &= \sum_{k=1}^{\infty}\frac{c^2}{\lambda_k}\Big\langle \sum_{l=1}^{\infty}\langle \bar{G}(X^{\psi}(s))\eta,\sqrt{\lambda_l}e_l\rangle_H\sqrt{\lambda}e_l,e_k\Big\rangle_H^2\\ &= c^2\sum_{k=1}^{\infty}\lambda_k\langle \bar{G}(X^{\psi}(s))\eta,e_k\rangle_H^2\\ &\leq (c\,C_G C_\eta)^2\,\mathrm{tr}(Q)<\infty.\end{aligned}$$

Hence, Conditions (3.5) and (3.7) are both fulfilled and for this reason the density process is a martingale. For the $\mathcal{F}_t$-predictability of the weight operator $\psi(t)$ it is sufficient to show that the transformed SPDE (3.2) has a unique mild solution $X^{\psi}(t)$ with respect to the transformed probability measure $\widehat{\mathbb{P}}$. This is fulfilled since F and G both satisfy Assumption 2.1 and for this reason, there exists a unique $\mathcal{F}_t$-predictable mild solution $X^{\psi}(t)$ to the transformed SPDE (3.2). Thus, all conditions from Theorem 3.1 are fulfilled and the measure transformation based on the path-dependent weight operator (3.10) can be applied.

4. Variance reduction by importance sampling

In this section, we consider how we can exploit the measure transformation introduced by Theorem 3.1 for variance reduction purposes. In the last section, we presented a technique how to construct the weight operator $\psi(t), t\in[0,T]$, to obtain a measure transformation satisfying all conditions of Theorem 3.1.

Based on the notation above, let $X(T)$ denote the (mild) solution of SPDE (2.1) and let $X^{\psi}(T)$ be the corresponding (mild) solution of the transformed SPDE (3.2) based on an appropriate weight operator ψ. Then, due to the Girsanov theorem, we get

$$\mathbb{E}_{\mathbb{P}}[\varphi(X(T))]=\mathbb{E}_{\widehat{\mathbb{P}}}[\varphi(X^{\psi}(T))]=\mathbb{E}_{\mathbb{P}}[\varphi(X^{\psi}(T))\Theta(T)],$$

where $\varphi: H\to B$ and $(B,\langle\cdot,\cdot\rangle_B)$ denotes a separable Hilbert space. Thus, by using a Monte Carlo estimator based on independent realisations of $\varphi(X^{\psi}(T))\Theta(T)$ instead of simulating independent samples of $\varphi(X(T))$ we obtain an unbiased estimator for $\mathbb{E}[\varphi(X(T))]$. The goal is now to find a weight operator $\psi(t), t\in[0,T]$, such that

$$\mathrm{Var}_B[\varphi(X^{\psi}(T))\Theta(T)]\ll\mathrm{Var}_B[\varphi(X(T))],$$

which by Equation (1.1) implies that the Monte Carlo error is reduced.

As it is also indicated in [9] for the finite dimensional setting, importance sampling is a strong tool for variance reduction and the choice of the weight operator $\psi(s), s\in[0,T]$, is crucial for a successful application of importance sampling. In particular, one has to be aware of the fact that it is also possible to enlarge the variance of the underlying random variable dramatically by choosing an inappropriate weight operator ψ.

For this reason, we proceed as follows: First, we present in Section 4.1 an optimal measure transformation (for a theoretical benchmark problem) such that the variance of $\varphi(X^{\psi}(T))\Theta(T)$ vanishes. Afterwards, we discuss how we can use this optimal measure transformation as guidance for setting up measure transformations

leading to variance reduction techniques that can be efficiently implemented. Numerical experiments in Section 6 show the effectiveness of the resulting importance sampling methods.

From a computational point of view, it is also very important to find a balance between the computational cost of computing the importance sampling technique and its reduction of the variance. Thus, optimal importance sampling techniques or at least approximations of such measure transformations might not be suitable for implementations if the computational cost for computing these measure transformations exceeds the computational complexity of solving the original problem.

4.1. **Optimal variance reduction.** In this section, we want to optimise the above importance sampling technique with respect to the weight operator $\psi(t), t \in [0,T]$. The term *optimal* has to be understood in the sense that the variance of $\varphi(X^{\psi}(T))\Theta(T)$ vanishes for a fixed time point T with respect to the mapping φ. This is summarised in the following definition:

Definition 4.1. The importance sampling weight operator $\widetilde{\psi} : [0,T] \to U_0$ is called *φ-optimal* with respect to time T if

$$\mathrm{Var}_B[\varphi(X^{\widetilde{\psi}}(T))\Theta(T)] = 0.$$

Such optimal variance reduction techniques have already been developed for different types of SDEs. In [15, Chapter 4], the authors proposed an importance sampling technique combined with a control variate approach for (strong solutions of) parabolic SPDEs that would eliminate the Monte Carlo error completely. However, this is only possible if the SPDE (2.1) is driven by finitely many standard Wiener processes. Moreover, the resulting methods require the simulation of an additional SDE system and of additional Wiener processes.

In this section, we want to find such an optimal importance sampling method for SPDEs driven by general Q-Wiener processes. The proof for this can be seen as the infinite dimensional counterpart of the proof presented in [16, Theorem 4.1], where the optimality result in finite dimensions is shown by using the solution of the corresponding Kolmogorov backward equation and the Itô formula. However, the techniques used in the proof of the finite dimensional result cannot be applied in the full generality to the SPDE setting, since the Itô formula is only applicable to SPDEs, for which a strong solution exists. Results extending the Itô formula to mild solutions as presented in [6] cannot be applied in the same manner since the proof of the optimality result relies on a specific interplay of the solution of the underlying Kolmogorov backward equation and the Itô formula that is not present if one applies the mild Itô formula from [6].

In order to prove that there exists a φ-optimal measure transformation in a (theoretical) benchmark setting we assume here that SPDE (2.1) has a unique strong solution. For the optimal weight operator $\psi(t), t \in [0,T]$, we have to ensure additionally that there exists a unique solution to the corresponding Kolmogorov backward equation of SPDE (2.1). For this reason, we have to introduce the following assumption within this section:

Assumption 4.2. The first and second Fréchet derivatives of F and G are continuous and bounded and satisfy the two conditions:

(1) there exists a constant $M_1 > 0$ such that for all $x, y \in H$
$$\|DF(x)y\|_H + \|DG(x)y\|_{L_{HS}(\mathcal{H}_0,H)} \leq M_1\|y\|_H,$$
(2) there exists a constant $M_2 > 0$ such that for all $x, y, z \in H$
$$\|D^2F(x)(y,z)\|_H + \|D^2G(x)(y,z)\|_{L_{HS}(\mathcal{H}_0,H)} \leq M_2\|y\|_H\|z\|_H.$$

Under this additional assumption, we can find a weight operator $\psi(t), t \in [0,T]$, and therefore also a measure transformation such that the variance of the random variable $\varphi(X^{\widetilde{\psi}}(T))\Theta(T)$ vanishes:

Theorem 4.3. *Let $X(t)$ denote the strong solution of SPDE (2.1) with deterministic initial value $X_0 \in H$. Furthermore, let $\varphi \in C_b^2(H;\mathbb{R}^+)$, i.e. the first and second Fréchet derivatives are bounded and $\varphi(h) > 0$ for all $h \in H \setminus \{0\}$. Additionally assume that the coefficients of SPDE (2.1) satisfy Assumptions 2.1 and 4.2. For all $t \in [0,T]$, let $X^{\psi}(t)$ denote the strong solution of the transformed SPDE (3.2) based on weight operator $\psi(t) \in \mathcal{H}_0$ and let $u : [0,T] \times H \to \mathbb{R}$ denote the solution of the Kolmogorov backward equation (of SPDE (2.1)) given in Equation (4.2) below.*

Then, the choice

$$\begin{aligned}\widetilde{\psi}(s) &= \sum_{k=1}^{\infty} \widetilde{\psi}_k(s, X^{\widetilde{\psi}}(s))e_k \\ &= \sum_{k=1}^{\infty} -\frac{\lambda_k}{u(s, X^{\widetilde{\psi}}(s))}\left\langle \frac{\partial u}{\partial x}(s, X^{\widetilde{\psi}}(s)), G(X^{\widetilde{\psi}}(s))e_k \right\rangle_H e_k\end{aligned} \tag{4.1}$$

is φ-optimal if the resulting density process $\Theta(t), t \in [0,T]$, is a martingale.

Proof. Due to Assumptions 2.1 and 4.2, there exists a unique solution u to the Kolmogorov Backward Equation of SPDE (2.1), see [8, Theorem 3.11]: For all $t \in [0,T]$ and $x \in H$, u solves

$$\begin{aligned}\mathcal{L}u(t,x) = \frac{\partial u(t,x)}{\partial t} &+ \left\langle Ax + F(x), \frac{\partial u(t,x)}{\partial x} \right\rangle_H \\ &+ \frac{1}{2}\operatorname{tr}\left(\frac{\partial^2 u(t,x)}{\partial x^2}(G(x)Q^{1/2})(G(x)Q^{1/2})^*\right) = 0, \\ u(T,x) &= \phi(x).\end{aligned} \tag{4.2}$$

In the next step, we apply the integration by parts formula from [5, Corollary 2.6.] to $u(t, X^{\psi}(t))\Theta(t)$, i.e.

$$\begin{aligned}u(t, X^{\psi}(t))\Theta(t) = \; & u(0,X_0)\Theta_0 \\ &+ \int_0^t \left(\mathcal{L}u(s, X^{\psi}(s)) - \left\langle \frac{\partial u}{\partial x}(s, X^{\psi}(s)), G(X^{\psi})(s)\psi(s) \right\rangle_H\right)\Theta(s)\,\mathrm{d}\,s \\ &+ \int_0^t \sum_{n=1}^{\infty} \left\langle \frac{\partial u}{\partial x}(s, X^{\psi}(s)), G(X^{\psi}(s))e_n \right\rangle_H \Theta(s)\langle\psi(s), e_n\rangle_0\,\mathrm{d}\,s \\ &+ \int_0^t u(s, X^{\psi}(s))\Theta(s)\langle\psi(s),\,\mathrm{d}\,W(s)\rangle_0 \\ &+ \int_0^t \Theta(s)\left\langle \frac{\partial u}{\partial x}(s, X^{\psi}(s)), G(X^{\psi}(s))\,\mathrm{d}\,W(s) \right\rangle_H.\end{aligned} \tag{4.3}$$

Note that $\mathcal{L}u = 0$ since u solves Equation (4.2). In the following part, we have to show that the remaining deterministic integrals in Equation (4.3) are equal. For this, recall that $e_n, n \in \mathbb{N}$, is an ONB of H. Since $\psi(s), s \in [0,T]$, is $\mathcal{H}_0$-valued, we get that for all $s \in [0,T]$

$$\begin{aligned}\langle \psi(s), e_n\rangle_0 &= \sum_{k=1}^{\infty} \frac{1}{\lambda_k}\langle \psi(s), e_k\rangle_H \langle e_k, e_n\rangle_H \\ &= \frac{1}{\lambda_n}\left\langle \sum_{\ell=1}^{\infty}\langle \psi(s), \sqrt{\lambda_\ell} e_\ell\rangle \sqrt{\lambda_\ell} e_\ell, e_n \right\rangle_H = \langle \psi(s), e_n\rangle_H.\end{aligned}$$

Then, by using that $\mathcal{H}_0 \subset H$,

$$\begin{aligned}&\int_0^t \sum_{n=1}^{\infty} \left\langle \frac{\partial u}{\partial x}(s, X^\psi(s)), G(X^\psi(s)) e_n \right\rangle_H \Theta(s)\langle \psi(s), e_n\rangle_0 \,\mathrm{d}\, s \\ &\qquad = \int_0^t \sum_{n=1}^{\infty} \left\langle \frac{\partial u}{\partial x}(s, X^\psi(s)), G(X^\psi(s))\langle \psi(s), e_n\rangle e_n \right\rangle_H \Theta(s) \,\mathrm{d}\, s \\ &\qquad = \int_0^t \left\langle \frac{\partial u}{\partial x}(s, X^\psi(s)), G(X^\psi(s)) \sum_{n=1}^{\infty}\langle \psi(s), e_n\rangle e_n \right\rangle_H \Theta(s) \,\mathrm{d}\, s \\ &\qquad = \int_0^t \left\langle \frac{\partial u}{\partial x}(s, X^\psi(s)), G(X^\psi(s))\psi(s) \right\rangle_H \Theta(s) \,\mathrm{d}\, s.\end{aligned}$$

Thus, the two deterministic integrals in Equation (4.3) are equal and therefore, the difference between these two integrals vanishes. Consequently, Equation (4.3) reduces to

$$\begin{aligned}u(t, X^\psi(t))\Theta(t) &= u(0, X_0)\Theta_0 + \int_0^t u(s, X^\psi(s))\Theta(s)\langle \psi(s), \,\mathrm{d}\, W(s)\rangle_0 \\ &\qquad + \int_0^t \Theta(s)\left\langle \frac{\partial u}{\partial x}(s, X^\psi(s)), G(X^\psi(s)) \,\mathrm{d}\, W(s) \right\rangle_H \\ &=: I + II + III.\end{aligned}$$

Note that I is deterministic,

$$\begin{aligned}II &= \sum_{k=1}^{\infty} \frac{1}{\sqrt{\lambda_k}} \int_0^t u(s, X^\psi(s))\Theta(s)\, \langle \psi(s), e_k\rangle_H \,\mathrm{d}\, \beta_k(s), \\ III &= \sum_{k=1}^{\infty} \int_0^t \Theta(s)\sqrt{\lambda_k}\left\langle \frac{\partial u}{\partial x}(s, X^\psi(s)), G(X^\psi(s)) e_k \right\rangle_H \,\mathrm{d}\, \beta_k(s).\end{aligned}$$

Thus, by defining for all $k \in \mathbb{N}$ and $s \in [0,T]$

$$\widetilde{\psi}_k(s, X^{\widetilde{\psi}}(s)) := -\lambda_k u(s, X^{\widetilde{\psi}}(s))^{-1}\left\langle \frac{\partial u}{\partial x}(s, X^{\widetilde{\psi}}(s)), G(X^{\widetilde{\psi}}(s)) e_k \right\rangle_H,$$

the importance sampling weight function

$$\widetilde{\psi}(s, X^{\widetilde{\psi}}(s)) = \sum_{k=1}^{\infty} \widetilde{\psi}_k(s, X^{\widetilde{\psi}}(s)) e_k$$

is φ-optimal at time $t = T$ since for all $t \in [0,T]$ it holds that

$$u(t, X^{\widetilde{\psi}}(t))\Theta(t) = u(0, X_0)\Theta_0 = u(0, X_0)$$

is deterministic. This is in particular also valid for time point $t = T$, at which by Equation (4.2) $u(T, X^{\widetilde{\psi}}(T))\Theta(T)$ coincides with $\varphi(X^{\widetilde{\psi}}(T))\Theta(T)$. Consequently, we get $\mathrm{Var}[\varphi(X^{\widetilde{\psi}}(T))\Theta(T)] = 0$. □

This optimal measure transformation can now be used as guidance to adjust the proposed measure transformations from Section 3.1 such that the used importance sampling technique results in a significant variance reduction. The path-dependent structure of the optimal weight operator $\widetilde{\psi}$ particularly motivates the use of the path-dependent weight operator defined in Equation (3.10) since this can be interpreted as a crude approximation of the optimal choice given by (4.1). In numerical experiments in Section 6, we examine how to choose the parameter c of the weight operator (3.10) in order to achieve the best possible variance reduction.

5. Approximation of measure transformations

In this section, we recall approximation properties of spatial and temporal discretisation techniques and discuss how the importance sampling methods based on the measure transformations form Section 3.1 can be implemented for numerical experiments using fully discrete schemes. The advantage of proposed methods is that the computational complexity of simulating the transformed random variables and the simulation of the original mild solution process is of the same order. This is due to the fact that for the final implementation the computation of the density process $\Theta(t)$ reduces for the considered measure transformations to the approximation of the solution of a one-dimensional SODE driven by finitely many standard Wiener processes. Hence, the difference in the computational cost of the standard approach compared to the proposed importance sampling methods is limited to the (numerical) approximation of a one-dimensional linear SODE.

Since the explicit formula of the mild solution of an SPDE is not known in general, we have to additionally approximate its mild solution $X(T)$ by a (spatially and temporally) discretised approximation $X_{h,N}$, see e.g. [13, 11] for details on strong and weak numerical approximations of stochastic equations in infinite dimensions. By using this additional approximation, the overall error of the Monte-Carlo method can be estimated by, see e.g. [16],

$$\begin{aligned}
&\|\mathbb{E}[\varphi(X(T))] - E_M[\varphi(X_{h,N})]\|_{L^2(\Omega;B)} \\
&\quad \leq \underbrace{\|\mathbb{E}[\varphi(X(T))] - \mathbb{E}[\varphi(X_{h,N})]\|_B}_{\text{systematic error}} + \underbrace{\|\mathbb{E}[\varphi(X_{h,N})] - E_M[\varphi(X_h(T))]\|_{L^2(\Omega;B)}}_{\text{Monte Carlo error}} \\
&\quad = \|\mathbb{E}[\varphi(X(T))] - \mathbb{E}[\varphi(X_{h,N})]\|_B + \frac{\mathrm{Var}_B[\varphi(X_{h,N})]^{1/2}}{\sqrt{M}}.
\end{aligned}$$

The following lemma shows that the variance of the approximated mild solution (2.4) can be estimated by the strong error and the variance of the original quantity of interest. For the analysis of $\mathrm{Var}[\varphi(X_h(T))]$, we assume that the mapping $\varphi : H \to B$ satisfies a Lipschitz continuity condition, i.e. there exists a constant $C > 0$ such that for all $\phi_1, \phi_2 \in H$ it holds that $\|\varphi(\phi_1) - \varphi(\phi_2)\|_B \leq C\|\phi_1 - \phi_2\|$.

Lemma 5.1. *Let $\varphi : H \to B$ be Lipschitz continuous. Then, there exists a constant $C > 0$ independent of h such that*

$$\mathrm{Var}_B[\varphi(X_{h,N})] \leq C\|X(T) - X_{h,N}\|^2_{L^2(\Omega;H)} + 3\,\mathrm{Var}_B[\varphi(X(T))].$$

Proof. By using the triangle inequality and the Lipschitz continuity of φ, we get

$$\begin{aligned}\mathrm{Var}_B[\varphi(X_{h,N})] &= \mathbb{E}[\|\varphi(X_{h,N}) - \mathbb{E}[\varphi(X_{h,N})]\|_B^2] \\ &= \mathbb{E}[\|\varphi(X_{h,N}) \pm \varphi(X_{h,N}) \pm \mathbb{E}[\varphi(X(T))] - \mathbb{E}[\varphi(X_{h,N})]\|_B^2] \\ &\le 3\Big(\mathbb{E}[\|\varphi(X(T)) - \varphi(X_{h,N})\|_B^2] + \|\mathbb{E}[\varphi(X(T))] - \mathbb{E}[\varphi(X_{h,N})]\|_B^2 \\ &\qquad + \mathbb{E}[\|\varphi(X(T)) - \mathbb{E}[\varphi(X(T))]\|_B^2]\Big) \\ &\le C\|X(T) - X_{h,N}\|_{L^2(\Omega;H)}^2 + 3\,\mathrm{Var}_B[\varphi(X(T))].\end{aligned}$$

Thus, the statement is shown. □

Hence the variance of $\varphi(X_{h,N})$ can be estimated in terms of the strong approximation error and the variance of the original quantity of interest $\varphi(X(T))$.

In the following two sections, we derive importance sampling techniques based on the infinite dimensional measure transformations from Section 3.1 combined with different spatial and temporal discretisation, i.e. for the space discretisation we consider a finite difference method (in Section 5.1) and a Galerkin finite element method (in Section 5.2). For both spatial discretisation methods, we consider the backward Euler scheme for the time integration.

5.1. **Approximation of the mild solution by finite differences.** In this section, we consider a finite difference method for approximating the mild solution of the transformed SPDE (3.2). For simplicity, we consider a one-dimensional setting on a bounded domain, where we fix a spatial mesh $\mathcal{T}_h$ with $N = 1/h, h > 0$, spatial grid points, i.e.

$$\mathcal{T}_h := \{x_0 < x_1 < \cdots < x_N\}.$$

The operator A is assumed to be a differential operator and its approximation $\mathbf{A}_h \in \mathbb{R}^{(N+1)\times(N+1)}$ is computed by using discrete difference quotients on $\mathcal{T}_h$. Then, we obtain for the finite difference approximation $\mathbf{x}(t) = [X(t,x_0), \dots X(t,x_N)]^T$ the following finite dimensional SODE system

$$\begin{aligned}\mathrm{d}\,\mathbf{x}(t) &= [\mathbf{A}_h\mathbf{x}(t) + \mathbf{F}(\mathbf{x}(t)) + \mathbf{G}(\mathbf{x}(t))\psi(t)]\;\mathrm{d}\,t + \mathbf{G}(\mathbf{x}(t))d\mathbf{W}(t) \quad \text{for } t \in [0,T], \\ \mathbf{x}(0) &= [X_0(x_0), \dots, X_0(x_N)]^T,\end{aligned}$$

where $\mathbf{W}(t) = [W(t,x_0), \dots, W(t,x_N)]^T$, $\mathbf{F}(\mathbf{x}(t)) = [F(X(t,x_0)), \dots, F(X(t,x_N))]^T$, and $\mathbf{G}(\mathbf{x}(t))\psi(t) = [G(X(t,x_0))\psi(t,x_0), \dots, G(X(t,x_N))\psi(t,x_N)]^T$. The resulting SODE system can be approximated by using any appropriate time integration method. For this, we fix a partition θ_K of the time interval $[0,T]$ consisting of $(K+1)$ time points with not necessarily equidistant time step sizes $\delta t_j = t_j - t_{j-1}$, i.e.

$$\theta_K := \{0 = t_0 < t_1 < \cdots < t_K = T\}. \tag{5.1}$$

As an example of a time integration method we consider the backward Euler scheme (or the linearly implicit Euler-Maruyama scheme) which is given by $\mathbf{x}_0 = \mathbf{x}(0)$ and

$$\mathbf{x}_j = \mathbf{x}_{j-1} + \delta t_j(\mathbf{A}_h\mathbf{x}_j + \mathbf{F}(\mathbf{x}_{j-1}) + \mathbf{G}(\mathbf{x}_{j-1})\psi(t_{j-1})) + \mathbf{G}(\mathbf{x}_{j-1})\Delta\mathbf{W}_j$$

for $j = 1, \dots, K$, where $\mathbf{x}_j$ is an approximation of $[X(t_j, x_0), \dots, X(t_j, x_N)]^T$ and $\Delta \mathbf{W}_j = \mathbf{W}(t_j) - \mathbf{W}(t_{j-1})$.

As already mentioned in the introduction, the authors in [1] already considered the coupling of finite difference approximations of SPDEs and importance sampling. At first, they fixed the spatial discretisation of the linear, one-dimensional stochastic heat equation (see Equation (6.1) below) based on finite differences with N spatial grid points and afterwards an importance sampling technique was applied to the resulting finite dimensional SODE system. However, the approach in [1] is equivalent to the truncated infinite dimensional importance sampling technique from above using the time-constant weight operator from Equation (3.9), i.e. we use the weight operator $\psi^J(s), s \in [0, T]$, defined for all $s \in [0, T]$ by

$$\psi^J(s) := P_J(\psi(s)) = \sum_{k=1}^{J} \lambda_k \psi_k \langle 1, e_k \rangle e_k,$$

where $J \in \mathbb{N}$ is sufficiently large and P_J denotes the orthogonal projector onto $\operatorname{span}(e_1, \dots, e_J)$. Thus, for numerical experiments concerning finite difference approximations of SPDE (2.1) and importance sampling based on the weight operator from Section 3.1.1 we refer to [1].

5.2. **Approximation of the mild solution by Galerkin methods.** For simulations of the mild solution (2.4), we use now an approximation based on a Galerkin finite element method in space. For this, we choose a nested sequence $\mathcal{V} = (V_h)_{h \in (0,1]}$ of finite dimensional subspaces of H satisfying $V_h \subset \dot{H}^1 \subset H$.

Following the approach from [11], we consider the Ritz projection $R_h : \dot{H}^1 \to V_h$ and the orthogonal projection $P_h : H \to V_h$ defined by

$$\begin{aligned} \langle R_h v, v_h \rangle_1 &= \langle v, v_h \rangle_1 \qquad \text{for all } v \in \dot{H}^1, v_h \in V_h, \\ \langle P_h v, v_h \rangle_H &= \langle v, v_h \rangle_H \qquad \text{for all } v \in H, v_h \in V_h \end{aligned}$$

For the convergence analysis of the Galerkin finite element method we need the following assumption on the orthogonal projectors R_h and P_h:

Assumption 5.2. For the family of finite dimensional subspaces $\mathcal{V} = (V_h)_{h \in (0,1]}$, there exists a constant $C > 0$ such that for all refinement parameters $h \in (0, 1]$ it holds that

(1) $\|P_h v\|_1 \leq C \|v\|_1$ for all $v \in \dot{H}^1$,
(2) $\|R_h v - v\|_H \leq C h^s \|v\|_s$ for all $v \in \dot{H}^s$ with $s \in \{1, 2\}$.

According to the considered finite dimensional subspace V_h, we define the discrete operator A_h as the unique operator satisfying

$$\langle -A v_h, w_h \rangle_H = \langle v_h, w_h \rangle_1 = \langle -A_h v_h, w_h \rangle_H$$

for all $v_h, w_h \in V_h$. Note that A_h is a symmetric and positive definite operator on V_h and thus it is also the generator of an analytic semigroup. For this reason, there exists a unique mild solution $X_h(t)$ of the spatially discretised SPDE (2.1) given by

$$\begin{aligned} X_h(t) = S_h(t) P_h X_0 &+ \int_0^t S_h(t-s) P_h f(X_h(s)) \, \mathrm{d}\, s \\ &+ \int_0^t S_h(t-s) P_h G(X_h(s)) \, \mathrm{d}\, W(s). \end{aligned} \tag{5.2}$$

For a given weight operator $\psi(s), s \in [0,T]$, satisfying the conditions of Theorem 3.1, we can rewrite the transformed mild solution of the semi-discrete problem (5.2) as

$$\begin{aligned} X_h^\psi(t) &= S_h(t)P_hX_0 + \int_0^t S_h(t-s)P_hF(X_h^\psi(s))\,\mathrm{d}\,s \\ &\qquad + \int_0^t S_h(t-s)P_hG(X_h^\psi(s))\,\mathrm{d}\,\widehat{W}(s) \\ &= S_h(t)P_hX_0 + \int_0^t S_h(t-s)P_h[F(X_h^\psi(s)) - G(X_h^\psi(s))\psi(s)]\,\mathrm{d}\,s \\ &\qquad + \int_0^t S_h(t-s)P_hG(X_h^\psi(s))\,\mathrm{d}\,W(s). \end{aligned}$$

Thus, by using the density process $\Theta(t), t \in [0,T]$, we can also deduce for the semi-discrete approximation that

$$\mathbb{E}[\varphi(X_h(T))] = \mathbb{E}_{\widehat{\mathbb{P}}}[\varphi(X_h^\psi(T))] = \mathbb{E}[\varphi(X_h^\psi(T))\Theta(T)]. \tag{5.3}$$

For a fully discrete approximation of the mild solution we have to additionally consider a discretisation in time. For this , we choose again the backward Euler scheme on the time grid θ_K given in Equation (5.1)

$$\begin{aligned} &X_{h,j} = X_{h,j-1} + \delta t_j\, A_hX_{h,j} + \delta t_j P_hF(X_{h,j-1}) + P_hG(X_{h,j-1})\Delta W^j \quad j = 1,\dots,K, \\ &X_h^0 = P_hX_0, \end{aligned}$$

where $\delta t_j = t_j - t_{j-1}$ and $\Delta W^j = W(t_j) - W(t_{j-1})$ are the Wiener increments. Now, let $\delta t = \max_i \delta t_i$. Then, by Assumption 2.1 and 5.2, it follows that there exists a constant independent of h and δt such that, see [11, Theorem 3.14],

$$\|X(T) - X_{h,N}\|_{L^2(\Omega;H)} \le C(h + \delta t^{1/2}). \tag{5.4}$$

The following Lemma provides an estimate of the overall error of the standard Monte Carlo estimator in terms of spatial and temporal refinement parameters and the variance of the original quantity of interest.

Lemma 5.3. *Let $\varphi : H \to B$ be Lipschitz continuous. Then, there exists a constant $C > 0$ independent of h and δt such that*

$$\|\mathbb{E}[\varphi(X(T))] - E_M[\varphi(X_{h,N})]\|_{L^2(\Omega;B)} \le C(h + \delta t^{1/2}) + \frac{\sqrt{3}\,\mathrm{Var}_B[\varphi(X(T))]^{1/2}}{\sqrt{M}}.$$

Proof. Note that

$$\begin{aligned} &\|\mathbb{E}[\varphi(X(T))] - E_M[\varphi(X_{h,N})]\|_{L^2(\Omega;B)} \le \\ &\qquad \|\mathbb{E}[\varphi(X(T))] - \mathbb{E}[\varphi(X_{h,N})]\|_B + \frac{1}{\sqrt{M}}\,\mathrm{Var}_B[\varphi(X_{h,N})]^{1/2}. \end{aligned}$$

Since φ is Lipschitz, the first summand can be estimated by the strong approximation error from (5.4). Finally, by applying Lemma 5.1 and using again the strong convergence properties from (5.4), the statement is shown. □

Note that in cases where the systematic error is dominated by the Monte Carlo error a further reduction of the spatial and temporal grid width does not lead to an improvement in the error bound of Lemma 5.3. Thus, we propose to enhance the performance of the Monte Carlo simulation by using importance sampling, for which

we show numerous numerical experiments in Section 6 using the path-dependent weight operator from Section 3.1.2.

For the implementation of the transformed SPDE, note that the temporal discretisation of transformed SPDE (3.2) by the backward Euler scheme is given by

$$\begin{aligned} X_{h,j}^{\psi} &= X_{h,j-1}^{\psi} + \delta t_j (A_h X_{h,j}^{\psi}) + \delta t_j P_h [F(X_{h,j-1}^{\psi}) + G(X_{h,j-1}^{\psi})\psi(t_{j-1})] \\ &\qquad + P_h G(X_{h,j-1}^{\psi}) \Delta W^j \quad j = 1, \dots, K, \\ X_h^0 &= P_h X_0. \end{aligned} \tag{5.5}$$

Until now, we have only discussed the discretisation of SPDE (2.1) with respect to space and time. For an implementation, we have to consider in addition an appropriate noise approximation, where we truncate the Karhunen-Loève expansion by using

$$G^{J_0}(u) = G(u) P_{J_0}, \tag{5.6}$$

where P_J denotes the orthogonal projector onto $\operatorname{span}(e_1, \dots, e_{J_0})$. Note that in order to preserve the convergence rates of the backward Euler scheme, we have to choose the truncation parameter $J_0 \in \mathbb{N}$ carefully, see e.g. [13, Lemma 10.33] and [3, Lemma 3.1].

5.3. **Approximation of the density process.** In this section, we discuss how to approximate the density process $\Theta(t), t \in [0,T]$, derived in Section 3.1. By the construction of the corresponding weight operators $\psi(t), t \in [0,T]$, the resulting density processes are the strong solutions of SODEs of the following type:

$$\mathrm{d}\,\Theta(t) = \Theta(t) \left(\sum_{k=1}^{\infty} \sqrt{\lambda_k}\, \kappa_k(t)\, \mathrm{d}\,\beta_k(t) \right), \qquad \Theta(0) = 1, \tag{5.7}$$

where the kernel functions $\kappa_k(s), s \in [0,T]$, are bounded, i.e. for all $s \in [0,T]$ and $k \in \mathbb{N}$ there exists a constant $C > 0$ such that $\kappa_k(s) \leq C < \infty$. Since in general we cannot compute the infinite series appearing in the density process explicitly, we introduce a truncation of the corresponding series that preserves the convergence properties of the used numerical method. Thus, we look for an appropriate truncation index $J_1 \in \mathbb{N}$ such that the truncated version of the SDE (5.7), i.e. for $t \in [0,T]$ let

$$\mathrm{d}\,\Theta^{J_1}(t) = \Theta^{J_1}(t) \left(\sum_{k=1}^{J_1} \sqrt{\lambda_k}\, \kappa_k(t)\, \mathrm{d}\,\beta_k(t) \right), \qquad \Theta^{J_1}(0) = 1, \tag{5.8}$$

reduces to a one-dimensional linear SDE with autonomous coefficients (for the transformation based on the time-independent weight operator from Equation (3.9)) or with non-autonomous coefficients (for the path-dependent choice of $\psi(t), t \in [0,T]$ from Equation (3.10)).

The goal is to choose truncation parameter J_1 in a way such that the corresponding weak convergence rates of the involved numerical methods are preserved. In the following lemma we discuss conditions on how to find J_1 for an arbitrary numerical method approximating the mild solution of SPDE (2.1) with given weak convergence order.

Lemma 5.4. *Assume that for all $t \in [0,T]$ there exists a constant $C > 0$ such that $\|\varphi(X_h^\psi(t))\|_{L^2(\Omega;\mathbb{R})} \leq C < \infty$ and that $\varphi : H \to \mathbb{R}$ is sufficiently smooth (according to the considered weak convergence order result). Furthermore, the numerical scheme approximating SPDE (2.1) has given weak convergence order $\mathcal{O}(h^\alpha)$ (under the assumption that the time step size δt and the spatial refinement parameter h are appropriately coupled). Then it holds that the weak convergence order is preserved if $J_1 \in \mathbb{N}$ is chosen such that*

$$\sum_{k=J_1+1}^{\infty} \lambda_k = \mathcal{O}(h^{2\alpha}).$$

Proof. First note that

$$\begin{aligned}
&|\mathbb{E}[\varphi(X(T))] - \mathbb{E}[\varphi(X_h^\psi(T))\Theta^{J_1}(T)]| \\
&\qquad = |\mathbb{E}[\varphi(X(T))] - \mathbb{E}[\varphi(X_h^\psi(T))\Theta^{J_1}(T)] \pm \mathbb{E}[\varphi(X_h(T))]| \\
&\qquad \leq |\mathbb{E}[\varphi(X(T))] - \mathbb{E}[\varphi(X_h(T))]| + |\mathbb{E}[\varphi(X_h(T)) - \mathbb{E}[\varphi(X_h^\psi(T))\Theta^{J_1}(T)]|.
\end{aligned}$$

The first part of the sum corresponds to the weak approximation error and for the second summand, we obtain by Equation (5.3) that

$$|\mathbb{E}[\varphi(X_h(T))] - \mathbb{E}[\varphi(X_h^\psi(T))\Theta^{J_1}(T)]| = |\mathbb{E}[\varphi(X_h^\psi(T))(\Theta(T) - \Theta^{J_1}(T))]|.$$

By applying Cauchy-Schwarz inequality and using the assumption that the second moment of $\varphi(X_h^\psi(T))$ is bounded, we get

$$\begin{aligned}
|\mathbb{E}[\varphi(X_h^\psi(T))(\Theta(T) - \Theta^{J_1}(T))]| &\leq \|\varphi(X_h^\psi(T))\|_{L^2(\Omega;\mathbb{R})} \|\Theta(T) - \Theta^{J_1}(T)\|_{L^2(\Omega;\mathbb{R})} \\
&\leq C\|\Theta(T) - \Theta^{J_1}(T)\|_{L^2(\Omega;\mathbb{R})}
\end{aligned}$$

Due to the independence of the Wiener processes $\beta_k(t)$, Itô's isometry and the boundedness of $\kappa(s)$ we obtain

$$\begin{aligned}
\|\Theta(T) - \Theta^{J_1}(T)\|^2_{L^2(\Omega;\mathbb{R})} &= \mathbb{E}\left[\left(\sum_{k=J_1+1}^{\infty} \int_0^T \sqrt{\lambda_k}\kappa_k(s)\, \mathrm{d}\,\beta_k(s)\right)^2\right] \\
&= \sum_{k=J_1+1}^{\infty} \lambda_k\, \mathbb{E}\left[\left(\int_0^T \kappa_k(s)\, \mathrm{d}\,\beta_k(s)\right)^2\right] \\
&= \sum_{k=J_1+1}^{\infty} \lambda_k\, \mathbb{E}\left[\int_0^T \kappa_k(s)^2\, \mathrm{d}\, s\right] \\
&\leq C \sum_{k=J_1+1}^{\infty} \lambda_k.
\end{aligned}$$

Thus, we get

$$|\mathbb{E}[\varphi(X(T))] - \mathbb{E}[\varphi(X_h^\psi(T))\Theta^{J_1}(T)]|^2 \leq |\mathbb{E}[\varphi(X(T))] - \mathbb{E}[\varphi(X_h(T))]|^2 + C \sum_{k=J_1+1}^{\infty} \lambda_k.$$

Finally, the weak convergence order $\mathcal{O}(h^\alpha)$ is preserved if $\sum_{k=J_1+1}^{\infty} \lambda_k = \mathcal{O}(h^{2\alpha})$. □

Until now we have discussed how to choose two different truncation parameters: First, the truncation parameter J_0 for the Karhunen-Loève expansion in Equation (5.6) that is chosen in a way such that certain convergence properties of the numerical methods are preserved, and second, the truncation parameter J_1 for the approximation of the density process, for which we found a specific selection criteria in terms of the weak convergence properties. Note that both parameters might be in general different depending on the numerical method of which the properties have to be preserved. However, we subsequently choose $J = \max\{J_0, J_1\}$ for both truncations which essentially means that our simulation results are based on J independent Wiener processes $\beta_k(t), k = 1, \dots, J$.

Concerning the implementation of the density process of the two measure transformation examples from Section 3.1, there is a slight difference. The time-constant weight operator defined in Equation (3.9) leads to an explicitly solvable SODE (5.8) with solution

$$\Theta^J(t) = \exp\left(\sum_{k=1}^{J} \sqrt{\lambda_k}\psi_k\langle 1, e_k\rangle_H \beta_k(t) - \frac{t}{2}\sum_{k=1}^{J}\lambda_k\psi_k^2\langle 1, e_k\rangle_H^2\right).$$

In contrast, the density process $\Theta(t)$ induced by the path-dependent weight operator given in Equation (3.10) is now represented as the solution of a linear one-dimensional SODE driven by J independent Wiener processes, i.e.

$$\mathrm{d}\,\Theta^J(t) = \Theta^J(t)\left(c\sum_{k=1}^{J}\sqrt{\lambda_k}\langle \bar{G}(X^\psi(t))\eta, e_k\rangle_H \,\mathrm{d}\,\beta_k(t)\right), \quad \Theta^J(0) = 1, \tag{5.9}$$

which can be approximated by using a standard numerical scheme for SDEs such as e.g. the Euler-Maruyama scheme or the Milstein scheme.

6. Numerical experiments

In this section, we present numerical experiments for the one-dimensional stochastic heat equation with homogeneous Dirichlet boundary conditions on the spatial domain $D = [0, 1]$ given by

$$\begin{aligned} \mathrm{d}\,X(t,x) &= \Delta_x X(t,x)\,\mathrm{d}\,t + G(X(t,x))\,\mathrm{d}\,W(t,x), \\ X(0,x) &= \sin(\pi x), \qquad \text{for } x \in D, \\ X(t,0) &= X(t,1) = 0, \qquad \text{for } t \in [0,T], \end{aligned} \tag{6.1}$$

where the Nemytskii operator $G : H \to L_{HS}(\mathcal{H}_0, H)$ is induced by either $\gamma(x) = \sin(x)$ (a globally bounded and Lipschitz continuous function) or by $\gamma(x) = \sigma x, \sigma \in \mathbb{R}$ (a globally unbounded but Lipschitz continuous function). Note that the underlying Hilbert space is $H = L^2([0, 1])$. Furthermore, we assume that the eigenvalues of the covariance operator Q (of the Q-Wiener process) $W(t), t \in [0, T]$, are defined by $\lambda_k = k^{-(2r+1+\epsilon)}$ for given regularity parameter $r > 0$ and $\epsilon > 0$. Finally, we assume that Q and the operator A (i.e. the Laplacian) commute.

Here, we consider importance sampling techniques for SPDEs based on the path-dependent measure transformation from Equation (3.10) with $\eta = e_1 \in \mathcal{H}_0$. For the discretisation of space and time, we follow the methods proposed in Section 5.2, i.e. a standard finite element method using piecewise linear test functions as basis functions of V_h and for the time integration we use the backward Euler scheme on an equidistant partition θ_K of the time interval $[0, T]$ with time step size $\delta t > 0$.

In Figure 2, we compare the performance of the importance sampling method with respect to different values of the scaling parameter $c \in \mathbb{R}$, where the diffusion operator is defined as the bounded Nemytskii operator G induced by $\gamma(x) = \sin(x)$. The functional $\varphi : H \to \mathbb{R}$ is chosen to be $\varphi(h) = \|h\|_H^2$ for all $h \in H$. Note that $c = 0$ corresponds to the standard Monte Carlo estimator without using importance sampling. The approximation of the solution is computed on a coarse spatial grid consisting of $N = 2^3$ grid points and on a partition θ_K of the time interval $[0, 1]$ consisting of $K = 2^6$ time steps. The regularity parameter controlling the decay of the eigenvalues of the covariance operator of the Q-Wiener process is chosen to be $r = 2$ and $\epsilon = 10^{-5}$.

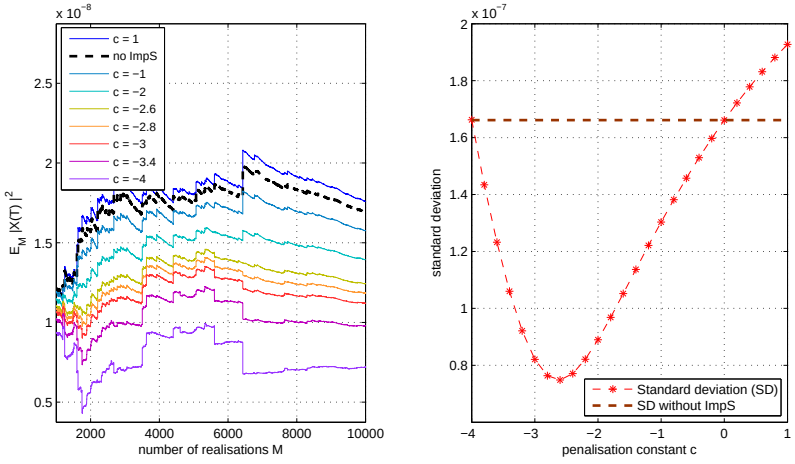

FIGURE 2. Simulation of $\mathbb{E}[\|X(1)\|_H^2]$ of the one-dimensional stochastic heat equation (6.1) with diffusion operator G defined as a Nemytskii operator based on $\gamma(x) = \sin(x)$. Left: Evolution of Monte Carlo estimators over M using importance sampling based on path-dependent weight operator from (3.10) with different scaling parameter c. Right: Standard deviation of the simulated realisations ($M = 10^5$) over different values of scaling parameter c.

As we can see on the left hand side of Figure 2, there are jumps appearing in the evolution of the standard Monte Carlo estimator (dashed line) due to realisations that are remarkably larger than the average. This causes severe difficulties for estimating $\mathbb{E}[\|X(1)\|_H^2]$ since we are likely to overestimate the mean-square process right after such jumps. For scaling parameter $c < 0$ we damp these jumps and for this reason we are able to reduce the standard deviation (and therefore also the variance) of the simulated realisations. However if we choose c too small, then *negative* jumps in the evolution of the Monte Carlo estimator appear. These jumps are caused by numerical instabilities in the simulation of $\Theta(t)$, which result in negative realisations of the simulated density process. As a consequence, one has to choose a scaling parameter c that leads to variance reduction but still guarantees numerical stability of the involved approximation schemes.

In Figure 3, we consider a realisation of the density process $\Theta(T)$ based on the path-dependent weight operator given in Equation (3.10) for fixed scaling parameter

$c = -2$. As proposed in Section 5 we truncate the infinite series appropriately (due to the regularity parameter $r = 2$ it is sufficient to choose the truncation parameter $J = N = 2^3$ according to Lemma 5.4), which leads to the one-dimensional SDE (5.9) that is approximated by using the standard Euler-Maruyama scheme.

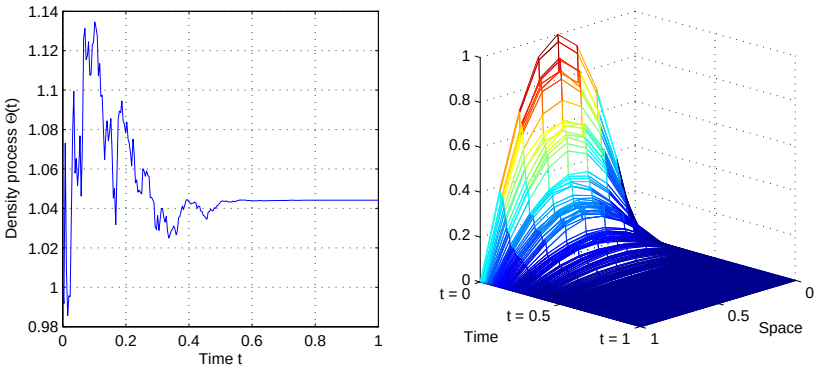

FIGURE 3. Left: A realisation of the approximated density process $\Theta^J(t), t \in [0,1]$, given in Equation (5.9) approximated by the standard Euler-Maruyama scheme. Right: The underlying realisation of the approximation of the transformed solution $X^\psi(t), t \in [0,1]$.

All the simulations in Figure 3 are performed by using $K = 2^8$ time steps in order to observe the path-dependence of $\Theta(t), t \in [0,1]$. After a short time horizon (here $t \approx 0.5$) the solution of SDE (5.9) stays almost constant due to the fact that the values of the underlying trajectory of $X^\psi(t)$ are very small. Thus, the realisation of the transformed process at the end time point is weighted by a constant that is mainly dependent on time intervals, where the norm of the transformed process is large or at least significantly larger than 0. As a consequence of the chosen scaling of the measure transformation introduced by the parameter c in (3.10), we are forcing (above $c > 0$) or damping (above $c < 0$) the importance of such events for the Monte Carlo estimator.

In the following part, we consider simulations of SPDE (2.1), where the diffusion operator G is defined as a Nemytskii operator based on a globally unbounded, but Lipschitz continuous function γ, i.e. $\gamma(x) = \sigma x, \sigma \in \mathbb{R}$. Thus, as proposed in Section 3.1 for the path-dependent measure transformation, we have to derive a representation of the truncated operator $\bar{G}$ in order to guarantee the martingale property of the corresponding density process $\Theta(t), t \in [0,T]$. This could be done for Nemytskii operators by approximating the underlying function $\gamma(x)$ by a bounded function $\bar{\gamma} : \mathbb{R} \to [-C_G, C_G]$ as we showed in Section 3.1.2. There are different possibilities how to achieve this. First, one can truncate γ if its absolute value exceeds a certain threshold $C_G > 0$, i.e. choose $\bar{\gamma}$ as

$$\gamma_c(C_G; x) = \operatorname{sign}(\gamma(x)) \min(C_G, |\gamma(x)|) \qquad \text{for all } x \in \mathbb{R}. \tag{6.2}$$

The choice of C_G depends on typical solution properties, e.g. one could choose C_G such that the majority of the discretised solution is smaller in absolute value than the chosen constant C_G. In Figure 4, we consider the case $\gamma(x) = x$ and the bounded function γ_c is truncated at $C_G = 1$. Further details concerning the

numerical outcome of the Monte Carlo simulations using the truncated function γ_c can be found in Table 1.

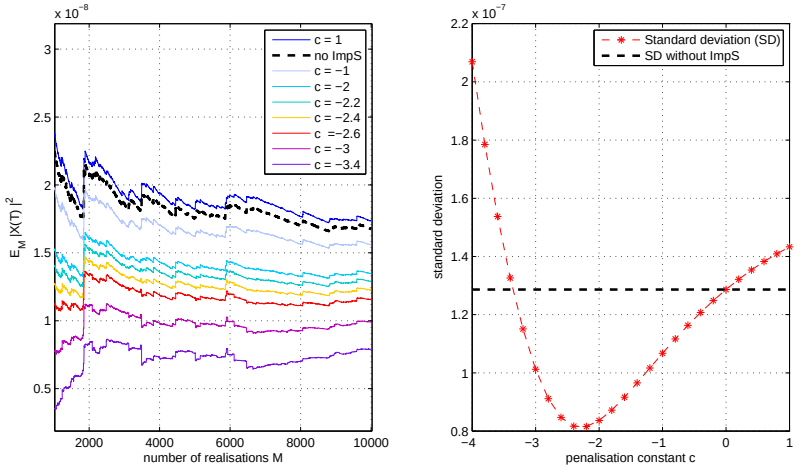

FIGURE 4. Simulation of $\mathbb{E}[\|X(1)\|_H^2]$ of the one-dimensional stochastic heat equation (6.1) with diffusion operator G defined as a Nemytskii operator based on $\gamma(x) = x$. Left: Evolution of Monte Carlo estimators over M using importance sampling based on path-dependent weight operator from (3.10) with different scaling parameter c. Right: Standard deviation of the simulated realisations ($M = 10^5$) over different values of scaling parameter c.

According to Figure 4, we are also able to reduce the variance in the case of a globally unbounded diffusion operator. However, for large and negative scaling parameter c, i.e. for c being approximately smaller than -2.4, the performance of the importance sampling method is worse compared to the case of the bounded diffusion ($\gamma(x) = \sin(x)$), since the same numerical instabilities as in Figure 2 already appear for choices $c \approx -2.4$.

By the same arguments as they were used to derive γ_c in Equation (6.2), one could choose $\bar{\gamma}$ as a bounded function such that the function γ is approximated well by $\bar{\gamma}$ at typical values of the discretised solution. For instance for linear $\gamma(x) = x$ one could consider an approximation by a sigmoid function, e.g. $\bar{\gamma}(x) = \mathrm{erf}(x)$ (error function).

In Table 1, we compare different numerical tests based on different importance sampling techniques using the path-dependent density operator from Section 3.1.2. Each of the 4 tests is based on $M = 10^5$ independent realisations. As a measure of the effectiveness of the importance sampling methods we consider the ratio between the standard deviation of the transformed processes (SD_c) over the standard deviation (SD_0) of the original realisations of $\|X(1)\|_H^2$. This ratio measures the relative change of the Monte Carlo error using importance sampling (based on scaling parameter c) with respect to the approximation error of the standard Monte Carlo estimator, i.e. for values $SD_c/SD_0 < 1$ the importance sampling method performs better than the original Monte Carlo simulation and for values c where

$SD_c/SD_0 > 1$ the performance is worse. In Table 1, we see that the proposed importance sampling techniques can reduce the Monte Carlo error significantly for specific scaling parameters c. Thus, these results show the efficiency of the proposed methods, since in order to obtain the same accuracy as for the standard Monte Carlo estimator, one needs $M_c = (\frac{SD_c}{SD_0})^2 M$ samples.

TABLE 1. Performance of the importance sampling techniques applied to SPDE (2.1) with path-dependent measure transformation from Section 3.1.2 using different scaling parameter c for test function $\varphi(h) = \|h\|_H^2, h \in H$.

	Test 1: $\gamma(x) = \sin(x)$		Test 2: $\gamma(x) = x$ with $\bar{\gamma}(x) = \gamma_c(1; x)$	
c	SD_c	SD_c/SD_0	SD_c	SD_c/SD_0
1	1.928E-07	1.1607	1.433E-07	1.1143
0	1.661E-07	1	1.286E-07	1
−1	1.303E-07	0.7845	1.067E-07	0.8297
−2	8.897E-08	0.5356	8.362E-08	0.6502
−3	8.210E-08	0.4943	9.111E-08	0.7085
−4	1.662E-07	1.0006	2.069E-07	1.6089
	Test 3: $\gamma(x) = x$ with $\bar{\gamma}(x) = \gamma_c(0.5; x)$		Test 4: $\gamma(x) = x$ with $\bar{\gamma}(x) = \mathrm{erf}(x)$	
c	SD_c	SD_c/SD_0	SD_c	SD_c/SD_0
1	2.467E-07	1.1989	2.516E-07	1.1770
0	2.060E-07	1	2.138E-07	1
−1	1.510E-07	0.7332	1.581E-07	0.7395
−2	1.209E-07	0.5870	1.242E-07	0.5810
−3	2.261E-07	1.0977	2.555E-07	1.1948
−4	4.630E-07	2.2481	5.690E-07	2.6611

References

[1] M. Ableidinger, E. Buckwar, and A. Thalhammer. "An importance sampling technique in Monte Carlo methods for SDEs with a.s. stable and mean-square unstable equilibrium". In: *Journal of Applied and Computational Mathematics* (2016). DOI: 10.1016/j.cam.2016.08.043.

[2] A. Barth and A. Lang. "Multilevel Monte Carlo method with applications to stochastic partial differential equations". In: *Int. J. Comput. Math.* 89.18 (2012), pp. 2479–2498. ISSN: 0020-7160. DOI: 10.1080/00207160.2012.701735.

[3] A. Barth and A. Lang. "Simulation of stochastic partial differential equations using finite element methods". In: *Stochastics An International Journal of Probability and Stochastic Processes* 84.2-3 (2012), pp. 217–231.

[4] A. Barth, A. Lang, and C. Schwab. "Multilevel Monte Carlo method for parabolic stochastic partial differential equations". In: *BIT* 53.1 (2013), pp. 3–27. ISSN: 0006-3835.

[5] Z. Brzeźniak et al. "Itô's formula in UMD Banach spaces and regularity of solutions of the Zakai equation". In: *J. Differential Equations* 245.1 (2008), pp. 30–58. ISSN: 0022-0396.

[6] G. Da Prato, A. Jentzen, and M. Röckner. "A mild Ito formula for SPDEs". In: *arXiv:1009.3526* (2011), 31 pages.

[7] G. Da Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimensions*. Encyclopedia of Mathematics and its Applications, 44. Cambridge University Press, Cambridge, 1992.

[8] L. Gawarecki and V. Mandrekar. *Stochastic differential equations in infinite dimensions with applications to stochastic partial differential equations*. Probability and its Applications (New York). Springer, Heidelberg, 2011.

[9] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Vol. 53. Springer Science & Business Media, 2003.

[10] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Applications of Mathematics, 23. Springer-Verlag, Berlin, 1992.

[11] R. Kruse. *Strong and Weak Approximation of Semilinear Stochastic Evolution Equations*. Lecture Notes in Mathematics, 2093. Springer, Cham, 2014.

[12] W. Liu and M. Röckner. *Stochastic partial differential equations: an introduction*. Universitext. Springer, Cham, 2015. ISBN: 978-3-319-22353-7; 978-3-319-22354-4. DOI: 10.1007/978-3-319-22354-4. URL: http://dx.doi.org/10.1007/978-3-319-22354-4.

[13] G. J. Lord, C. E. Powell, and T. Shardlow. *An Introduction to Computational Stochastic PDEs*. Cambridge Texts in Applied Mathematics. Cambridge University Press, New York, 2014.

[14] G. N. Milstein and M. V. Tretyakov. "Practical variance reduction via regression for simulating diffusions". In: *SIAM Journal on Numerical Analysis* 47.2 (2009), pp. 887–910.

[15] G. N. Milstein and M. V. Tretyakov. "Solving parabolic stochastic partial differential equations via averaging over characteristics". In: *Math. Comp.* 78.268 (2009), pp. 2075–2106. ISSN: 0025-5718.

[16] G. N. Milstein and M. V. Tretyakov. *Stochastic Numerics for Mathematical Physics*. Scientific Computation. Springer-Verlag, Berlin, 2004.

[17] B. Øksendal. *Stochastic differential equations*. Fourth. Universitext. An introduction with applications. Springer-Verlag, Berlin, 1995. ISBN: 3-540-60243-7. DOI: 10.1007/978-3-662-03185-8. URL: http://dx.doi.org/10.1007/978-3-662-03185-8.

[18] C. Prévôt and M. Röckner. *A concise course on stochastic partial differential equations*. Vol. 1905. Lecture Notes in Mathematics. Springer, Berlin, 2007. ISBN: 978-3-540-70780-6; 3-540-70780-8.

[19] P. E. Protter. *Stochastic integration and differential equations*. Vol. 21. Stochastic Modelling and Applied Probability. Second edition. Version 2.1, Corrected third printing. Springer-Verlag, Berlin, 2005. ISBN: 3-540-00313-4. DOI: 10.1007/978-3-662-10061-5. URL: http://dx.doi.org/10.1007/978-3-662-10061-5.

(Evelyn Buckwar)
Institute for Stochastics
Johannes Kepler University Linz
A–4040 Linz, Austria.
E-mail address: evelyn.buckwar@jku.at

(Andreas Thalhammer)
Doktoratskolleg Computational Mathematics & Institute for Stochastics
Johannes Kepler University Linz
A–4040 Linz, Austria.
E-mail address: andreas.thalhammer@jku.at

Technical Reports of the Doctoral Program "Computational Mathematics"

2017

2017-01 E. Buckwar, A. Thalhammer: *Importance Sampling Techniques for Stochastic Partial Differential Equations* January 2017. Eds.: U. Langer, R. Ramlau

2016

2016-01 P. Gangl, U. Langer: *A Local Mesh Modification Strategy for Interface Problems with Application to Shape and Topology Optimization* November 2016. Eds.: B. Jüttler, R. Ramlau

2016-02 C. Hofer: *Parallelization of Continuous and Discontinuous Galerkin Dual-Primal Isogeometric Tearing and Interconnecting Methods* November 2016. Eds.: U. Langer, W. Zulehner

2016-03 C. Hofer: *Analysis of Discontinuous Galerkin Dual-Primal Isogeometric Tearing and Interconnecting Methods* November 2016. Eds.: U. Langer, B. Jüttler

2016-04 A. Seiler, B. Jüttler: *Adaptive Numerical Quadrature for the Isogeometric Discontinuous Galerkin Method* November 2016. Eds.: U. Langer, J. Schicho

2016-05 S. Hubmer, A. Neubauer, R. Ramlau, H. U. Voss: *On the Parameter Estimation Problem of Magnetic Resonance Advection Imaging* December 2016. Eds.: B. Jüttler, U. Langer

2016-06 S. Hubmer, R. Ramlau: *Convergence Analysis of a Two-Point Gradient Method for Nonlinear Ill-Posed Problems* December 2016. Eds.: B. Jüttler, U. Langer

2015

2015-01 G. Grasegger, A. Lastra, J. Rafael Sendra, F. Winkler: *A Solution Method for Autonomous First-Order Algebraic Partial Differential Equations in Several Variables* January 2015. Eds.: U. Langer, J. Schicho

2015-02 P. Gangl, U. Langer, A. Laurain, H. Meftahi, K. Sturm: *Shape Optimization of an Electric Motor Subject to Nonlinear Magnetostatics* January 2015. Eds.: B. Jüttler, R. Ramlau

2015-03 C. Fürst, G. Landsmann: *Computation of Dimension in Filtered Free Modules by Gröbner Reduction* May 2015. Eds.: P. Paule, F. Winkler

2015-04 A. Mantzaflaris, H. Rahkooy, Z. Zafeirakopoulos: *Efficient Computation of Multiplicity and Directional Multiplicity of an Isolated Point* July 2015. Eds.: B. Buchberger, J. Schicho

2015-05 P. Gangl, S. Amstutz, U. Langer: *Topology Optimization of Electric Motor Using Topological Derivative for Nonlinear Magnetostatics* July 2015. Eds.: B. Jüttler, R. Ramlau

2015-06 A. Maletzky: *Exploring Reduction Ring Theory in Theorema* August 2015. Eds.: B. Buchberger, W. Schreiner

The complete list since 2009 can be found at
`https://www.dk-compmath.jku.at/publications/`

Doctoral Program
"Computational Mathematics"

Director:

Prof. Dr. Peter Paule
Research Institute for Symbolic Computation

Deputy Director:

Prof. Dr. Bert Jüttler
Institute of Applied Geometry

Address:

Johannes Kepler University Linz
Doctoral Program "Computational Mathematics"
Altenbergerstr. 69
A-4040 Linz
Austria
Tel.: ++43 732-2468-6840

E-Mail:

office@dk-compmath.jku.at

Homepage:

http://www.dk-compmath.jku.at

Submissions to the DK-Report Series are sent to two members of the Editorial Board who communicate their decision to the Managing Editor.