AN IMPORTANCE SAMPLING TECHNIQUE IN MONTE CARLO METHODS FOR SDES WITH A.S. STABLE AND MEAN-SQUARE UNSTABLE EQUILIBRIUM

MARKUS ABLEIDINGER, EVELYN BUCKWAR, AND ANDREAS THALHAMMER

Abstract. In this work we investigate the interplay of almost sure and mean-square stability for linear SDEs and the Monte Carlo method for estimating the second moment of the solution process. In the situation where the zero solution of the SDE is asymptotically stable in the almost sure sense but asymptotically mean-square unstable, the latter property is determined by rarely occurring trajectories that are sufficiently far away from the origin. The standard Monte Carlo approach for estimating higher moments essentially computes a finite number of trajectories and is bound to miss those rare events. It thus fails to reproduce the correct mean-square dynamics (under reasonable cost). A straightforward application of variance reduction techniques will typically not resolve the situation unless these methods force the rare, exploding trajectories to happen more frequently. Here we propose an appropriately tuned importance sampling technique based on Girsanov's theorem to deal with the rare event simulation. In addition further variance reduction techniques, such as multilevel Monte Carlo, can be applied to control the variance of the modified Monte Carlo estimators. As an illustrative example we discuss the numerical treatment of the stochastic heat equation with multiplicative noise and present simulation results.

1. Introduction

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ be a complete probability space, where the filtration $(\mathcal{F}_t)_{t\in[0,T]}$ satisfies the usual conditions. We consider d-dimensional systems of linear Itô stochastic differential equations

$$\mathrm{d}X(t) = AX(t)\,\mathrm{d}t + \sum_{r=1}^{m} G_r X(t)\,\mathrm{d}\beta_r(t), \qquad X(0) = X_0, \quad t \in [0,T], \tag{1}$$

where $A, G_r \in \mathbb{R}^{d\times d}$, β_r are m independent, real-valued Wiener processes, and the initial value X_0 is in $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$. Under these assumptions Eq. (1) has a unique strong solution, which, however, allows an explicit representation only for special cases such as for systems with commuting matrices A, G_r, see [1, Section 8.5, Remark 8.5.9]. For the case of a zero initial value $X_0 = 0$, Eq. (1) admits the zero solution $X_e(t) \equiv 0$ for all t.

In this article we are interested in the numerical approximation of quantities of the type $\mathbb{E}f(X(t))$ for some functional $f : \mathbb{R}^d \to \mathbb{R}$, which has at least quadratic growth in $|x|$. We first describe the setting of Monte Carlo techniques and error analysis in terms of interpreting

1991 *Mathematics Subject Classification.* 60H10, 65C05, 65M12, 60H15.

Key words and phrases. Monte Carlo estimators, Systems of linear SDEs, Asymptotic almost sure stability, Asymptotic mean-square stability, Variance reduction, Rare event simulation.

Acknowledgement. The first author was supported by the Austrian Science Fund (FWF) under P26314, and the third author was supported by the Austrian Science Fund (FWF) under W1214-N15, project DK14.

$$E_M[f(\widehat{X}(t)] := \frac{1}{M}\sum_{i=1}^{M} f(\widehat{X}^{(i)}(t)) \tag{2}$$

as an estimator for $\mathbb{E}f(X(t))$. Here, $\widehat{X}^{(i)}(t)$ denotes an independent realisation of the approximated solution $\widehat{X}(t)$, which we obtain by using a numerical integrator of weak order p. The approximation of the trajectories produces a *systematic* error, which can be expressed as ([19, 22])

$$\mathbb{E}[f(X(t))] = \mathbb{E}[f(\widehat{X}(t))] + \mathcal{O}(h^p).$$

Subsequently we will assume that the time step size is sufficiently small such that the systematic error is dominated by the *Monte Carlo* error, which can be expressed as ([19])

$$E_M[f(\widehat{X}(t))] = \frac{1}{M}\sum_{i=1}^{M} \mathbb{E}f(\widehat{X}^{(i)}(t)) \pm c\frac{\mathrm{Var}[f(\widehat{X}(t)))]^{1/2}}{M^{1/2}}.$$

Under the assumption that the variance of the approximated random variable $f(\widehat{X}(t))$ is close to the variance of $f(X(t))$, the quality of the Monte Carlo estimator depends on the variance of the underlying SDE. Thus for problems with large variances one needs either a large number of realisations M, i.e. a very fine discretisation of the underlying probability space, or estimators with a smaller variance than standard Monte Carlo estimators. This second consideration directly leads to the field of variance reduction techniques, for an overview see for example [8, 3].

In this work we analyse the impact of long time properties of linear systems of SDEs on Monte Carlo estimators, in particular how different concepts of stochastic stability, i.e. asymptotic stability in the almost sure and mean-square sense (see Definition 2.1), affect the numerical results using the standard Monte Carlo estimator defined in Eq. (2).

Ignoring any systematic error for the moment, we choose the following trivial observation as a starting point for our considerations: The estimator defined in Eq. (2) consists only of finitely many realisations of the solution process, and therefore the long time evolution is heavily influenced by the path-wise stability properties of the SDE. However, due to the (at least) quadratic growth of $f(x)$, $\mathbb{E}f(X(t))$ is governed by the corresponding mean-square stability of the zero solution of (1). Hence, the characterisation of our problem can be summarised by using the following observations:

I The zero solution of system (1) can be asymptotically stable in the almost sure sense but at the same time asymptotically mean-square unstable. This situation is well known in the literature, e.g. for the geometric Brownian motion (see [16] and Section 2).

II An immediate consequence of the above situation is that the exponential growth in time of $\mathbb{E}|X(t)|^2$ is due to very rare exploding trajectories. This leads to a prohibitively high number of realisations needed for the standard Monte Carlo estimation to obtain a decent approximation of the second or higher moments of the solution process $X(t)$.

III The problem is not purely academic: Space discretising diffusion-type SPDEs with e.g. finite differences leads to high dimensional SODE systems of type (1). The solution trajectories decrease rapidly due to the dissipative properties of the Laplacian and the stabilising structure of the diffusion matrices G_r. At the same time, the equilibrium solution will become asymptotically mean-square unstable for some, often moderate,

value of noise intensity and increasing the noise intensity will amplify this effect. Note that the computational cost for simulating such high-dimensional SDE systems automatically prohibits a substantial increase of the number of trajectories.

IV A straightforward application of multilevel Monte Carlo techniques for reducing the variance will typically not resolve the situation, unless these methods force the rare exploding trajectories to happen more frequently.

We propose the following modification of Monte Carlo estimation by importance sampling techniques: We change the drift of the system in such a way that the trajectories explode more often. This can be achieved by transforming the underlying probability measure due to Girsanov's theorem. Weighting the trajectories with the corresponding density process results in an unbiased Monte Carlo estimator for the desired quantity. An important property of the developed measure transformation is that it can be precomputed and does not change the linearity of the system. Consequently, the computational cost of our proposed method is the same as for the standard Monte Carlo estimation. Further, the variance of the modified estimators can also be reduced by applying additional variance reduction techniques such as multilevel Monte Carlo methods.

In Section 2 we will give a short review of necessary notions of stability theory of stochastic differential equations and we discuss the points I and II by providing analytical and numerical results for the geometric Brownian motion. In Section 2.1 we introduce a spatially discretised stochastic heat equation as an illustrative and non-trivial example and discuss the stability properties of the zero solution of the resulting system of SODEs. In Section 3 we discuss different variance reduction techniques (multilevel Monte Carlo methods and importance sampling) for the spatially discretised stochastic heat equation and formulate our proposed method. We will present numerical results in Section 4 and close this work with conclusions in Section 5.

2. Stability theory

In this section we summarise the main definitions and notation from stochastic stability theory and illustrate the differences between these stability concepts by considering geometric Brownian motion.

We treat the following two types of (asymptotic) stochastic stability of the equilibrium solution of Eq. (1), which is given by the zero solution $X_e(t) \equiv 0$ for all t, see [16, 13].

Definition 2.1. (i) The zero solution is called *asymptotically almost sure stable*
- (a) if for any given $0 < \varepsilon < 1, \varepsilon' > 0$, there exists $\delta > 0$ such that if $|X_0| < \delta$, then $\mathbb{P}(|X(t)| > \varepsilon') < \varepsilon$ for all $t \geq 0$ and
- (b) if there exists a $\delta' > 0$ such that for any X_0, satisfying $|X_0| < \delta'$ a.s., it holds that $\lim_{t\to\infty} |X(t)| = 0$ a.s.

(ii) The zero solution is called *asymptotically mean-square stable*
- (a) if for $\mathbb{E}|X_0|^2 \leq \delta$, then $\sup_{t\geq s} \mathbb{E}|X(t)|^2 \to 0$ as $\delta \to 0$ and
- (b) if there exists a $\delta' > 0$ such that for any X_0, satisfying $\mathbb{E}|X_0|^2 < \delta'$, it holds that $\lim_{t\to\infty} \mathbb{E}|X(t)|^2 = 0$.

In order to illustrate the difference between asymptotically almost sure stable and asymptotically mean-square stable equilibria, we recapitulate the well-known results for the geometric Brownian motion (see [1, 15]). Let

$$\mathrm{d}X(t) = \lambda X(t)\,\mathrm{d}t + \sigma X(t)\,\mathrm{d}\beta(t), \quad X(0) = X_0, \tag{3}$$

where $\lambda, \sigma, X_0 \in \mathbb{R}$. The explicit solution and the second moment of Eq. (3) are given by

$$X(t) = X_0 \exp\left(\left(\lambda - \frac{\sigma^2}{2}\right) t + \sigma\beta(t)\right), \tag{4}$$

$$\mathbb{E}X^2(t) = X_0^2 \exp\left(\left(2\lambda + \sigma^2\right) t\right). \tag{5}$$

Obviously the zero solution is asymptotically mean-square stable if and only if $2\lambda + \sigma^2 < 0$. Using the law of iterated logarithms, it can be seen that the zero solution is asymptotically stable in the almost sure sense if and only if $\lambda - \sigma^2/2 < 0$. As a consequence, the zero solution of Eq. (3) becomes unavoidably asymptotically stable in the almost sure and asymptotically unstable in the mean-square sense for sufficiently large σ (see also the discussion in [16, Example 1.4.1]).

Suppose we want to approximate $\mathbb{E}X^2(t)$ by standard Monte Carlo estimation, where we use the exact solution (4) for simulating the paths, i.e. no systematic error arises. For a given set of parameters we calculate the probability that $X^2(t)$ stays in an ϵ-neighbourhood of 0 with a certain probability $1 - \alpha$. We get

$$\mathbb{P}\left[X^2(t) < \epsilon\right] = 1 - \alpha \iff \mathbb{P}\left[\xi < \frac{\ln(\epsilon/X_0^2) - 2(\lambda - \sigma^2/2)t}{2\sigma\sqrt{t}}\right] = 1 - \alpha \quad \text{for } \xi \sim \mathcal{N}(0,1).$$

Note that for a fixed probability $1 - \alpha$ the bound ϵ *decreases* exponentially for growing σ and t, whereas the mean-square process (5) *grows* exponentially in σ and t. This compromises the results of Monte Carlo estimation and we illustrate this with the following parameter setting: Let $\lambda = -2$, $\sigma = 3$, and $X_0 = 1$, so the zero solution is asymptotically mean-square unstable. We further fix the time at $T = 10$ and obtain that the analytic solution of the mean-square process (5) has order of magnitude

$$\mathbb{E}X^2(T) \approx 10^{21}.$$

However, it is highly improbable to sample values substantially larger than zero. The probability that all paths of a standard Monte Carlo estimation with 10^4 trajectories are almost zero within machine accuracy is $1 - 10^{-4}$, since

$$\mathbb{P}\left[\left(X^{(i)}(T)\right)^2 < 10^{-14}\right] \approx 1 - 10^{-7} \quad \text{and} \quad \prod_{i=1}^{10^4} \mathbb{P}\left[\left(X^{(i)}(T)\right)^2 < 10^{-14}\right] \approx 1 - 10^{-3}.$$

As a consequence, the standard Monte Carlo estimator $E_M[X^2(T)]$ fails to approximate $\mathbb{E}X^2(T)$ (see Figure 1).

This effect is also present in the case of higher dimensional SODE systems (see Figure 2). In the next section, we consider this situation by using a space-discretised version of a stochastic heat equation with multiplicative noise.

2.1. **Linear systems of SODEs.** In this section we present the semi-discretised stochastic heat equation as a benchmark problem for spatially discretised diffusion type SPDEs and analyse the qualitative behaviour of the zero solution.

2.1.1. *Stochastic heat equation.* Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space with filtration $(\mathcal{F}_t)_{t\in[0,T]}$. We consider the one-dimensional stochastic heat equation on the spatial domain

$\mathcal{D} = [0, 1]$ with homogenous Dirichlet conditions:

$$\begin{aligned} \mathrm{d}u(x,t) &= \Delta u(x,t)\,\mathrm{d}t + \sigma u(x,t)\,\mathrm{d}W(x,t), \\ u(0,\cdot) &= u(1,\cdot) = 0, \\ u(x,0) &= u_0(x) = \sin(\pi x) \quad \text{for } x \in \mathcal{D}, \end{aligned} \tag{6}$$

with a sufficiently smooth initial function, since $u_0 \in L^2(\mathcal{D})$. The multiplicative noise is interpreted in the sense of Nemytskii operators (see [17]). For the driving noise we assume that $W(x,t)$ is an $\mathcal{F}_t$-adapted Q-Wiener process with a linear, non-negative and symmetric covariance operator Q. Furthermore Q is of trace class and has an orthonormal basis of eigenfunctions $\{\chi_j, j \in \mathbb{N}\}$ with eigenvalues $q_j \geq 0$. Under these assumptions $W(x,t)$ can be represented in the following series expansion

$$W(x,t) = \sum_{j=1}^{\infty} \sqrt{q_j}\chi_j(x)\beta_j(t), \tag{7}$$

which converges in $L^2(\Omega, \mathcal{F}, \mathbb{P})$ (see [5]). Here, $\beta_j(t)$ are independent one-dimensional Wiener processes. The existence of a unique mild solution of (6) is guaranteed by the classical framework of [5].

2.1.2. *Space discretisation and model problem.* For spatial discretisation we apply the standard finite difference scheme on an equidistant spatial mesh $\{x_0, \ldots, x_{N+1}\}$ with mesh width $h = 1/(N+1)$. Then the spatially discretised version of Eq. (6) takes the form

$$\mathrm{d}u(x_i,t) = \left(\Delta_h u(t)\right)(x_i)\,\mathrm{d}t + \sigma u(x_i,t)\,\mathrm{d}W(x_i,t),$$

where Δ_h denotes the three-point discrete Laplacian. Using the eigenfunctions of the Laplacian as a basis representation for $W(x,t)$ we define the following truncation of the series (7) (see [17]),

$$W^N(x,t) := \sum_{i=1}^{N} \sqrt{q_i}\chi_i(x)\beta_i(t)$$

with $q_i = 2i^{-(2r+1+\epsilon)}$ and $\chi_i(x) = \sqrt{2}\sin(\pi i x)$ for all $x \in \mathcal{D}$. Here $\epsilon > 0$ and r controls the regularity of the Wiener process in the sense that $W(t) \in H_0^r(0,1)$. Denoting the solution-vector at the interior points of the spatial mesh as

$$X(t) = (X_1(t), \ldots X_N(t))^T = (u(x_1,t), \ldots, u(x_N,t))^T,$$

we obtain the following N-dimensional SODE-system

$$\begin{aligned} \mathrm{d}X(t) &= A^h X(t)\,\mathrm{d}t + \sum_{i=1}^{N} G_i^h X(t)\,\mathrm{d}\beta_i(t), \\ X(0) &= (\sin(x_1\pi), \ldots, \sin(x_N\pi))^T, \end{aligned} \tag{8}$$

where

$$A^h = \frac{1}{h^2}\,\mathrm{tridiag}\left(1, -2, 1\right)_{j=1}^{N} \quad \text{and} \quad G_i^h = \mathrm{diag}\left\{\sqrt{2q_i}\sin\left(\frac{j\pi i}{N+1}\right)\right\}_{j=1}^{N}.$$

2.2. **Linear stability analysis of Eq.** (8)**.** The mean-square stability of the zero solution for linear SODE systems can be characterised via the *mean-square stability matrix* of the system. Following [4], the process $Y(t) = \mathrm{vec}(X(t)X(t)^T)$ satisfies the deterministic ODE

$$\mathrm{d}\mathbb{E}(Y(t)) = S\mathbb{E}(Y(t))\,\mathrm{d}t, \tag{9}$$

where

$$S = A^h \otimes \mathbb{I}_N + \mathbb{I}_N \otimes A^h + \sigma^2 \sum_{r=1}^{N} G_r^h \otimes G_r^h = A^h \oplus A^h + \sigma^2 \sum_{r=1}^{N} G_r^h \otimes G_r^h.$$

Here, $\otimes$ denotes the matrix Kronecker product, $\oplus$ denotes the Kronecker sum and $\mathbb{I}_N$ denotes the N-dimensional identity matrix.

Based on the following lemma, see e.g. [4], we can determine the mean-square stability properties of the zero solution of Eq. (8) by considering the spectrum of the underlying mean-square stability matrix S.

Lemma 2.2. *Let $\alpha(S)$ be the spectral abscissa of S. The zero solution of Eq.* (8) *is asymptotically mean-square stable if and only if $\alpha(S) < 0$.*

Both $A^h \oplus A^h$ and $\sum G_i^h \otimes G_i^h$ are symmetric, so we can apply Weyl's theorem, see e.g. [9, Theorem 4.3.1], to obtain a lower bound for $\alpha(S)$:

$$\max_{1\le j\le N^2} \left[\lambda_{N^2-j+1}\left(A^h \oplus A^h\right) + \sigma^2 \lambda_j \left(\sum_{i=1}^{N} G_i^h \otimes G_i^h \right) \right] \le \alpha(S).$$

Here, $\lambda_j(M)$ denotes the j-th largest eigenvalue of a matrix M. Due to basic properties of the Kronecker product (see e.g. [21]) and the form of G_i^h, the matrix $\sum G_i^h \otimes G_i^h$ has positive eigenvalues. Consequently, the zero solution of Eq. (6) is asymptotically mean-square unstable for large enough σ.

In analogy to the one dimensional case we expect that there exist values of σ for which the zero solution is asymptotic a.s. stable but asymptotic mean-square unstable. In fact, the authors in [15] showed that for linear SODE systems the asymptotic almost sure stability can be interpreted as the limiting case of p-th moment stability for $p \to 0$. However the computation of this limit is not straightforward and the p-th moments for $p < 1$ are hard to interpret. An alternative way for showing asymptotic almost sure stability of the zero solution is the analysis of the corresponding Lyapunov exponents. The *top Lyapunov exponent* is defined as

$$\Lambda = \lim_{t\to\infty} \frac{1}{t} \log |X(t)|\,.$$

As in the deterministic case, the path-wise stability of an SODE system is completely described by Λ: the zero solution of Eq. (6) is asymptotically almost sure stable iff $\Lambda < 0$ (see [1]). The actual computation of Λ requires a thorough understanding of the exact solution $X(t)$, which is typically not available, consequently analytic results are only known for low dimensional systems (see [10, 12]). For higher dimensions the analysis of the top Lyapunov exponent is still tractable if one assumes non-degeneracy of the driving noise (see [13, 2, 18] and for numerical approximations of Λ based on path simulation [22]). Let $D(\lambda) = \sum_{i=1}^{N} G_i^h \lambda\lambda^T G_i^h \in \mathbb{R}^{N\times N}$ for $\lambda \in \mathbb{R}^N$. We say the noise is non-degenerate if there is a $C > 0$ such that

$$\langle D(\lambda)\zeta, \zeta\rangle \ge C|\lambda|^2|\zeta|^2 \qquad \forall \lambda, \zeta \in \mathbb{R}^N \tag{10}$$

or equivalently, that the dimension of the linear hull of $\{G_1^h x, \dots, G_N^h x\}$ is equal to N (see [13, Appendix A]). Under Condition (10) the top Lyapunov exponent can be represented as (see [13, Theorem 6.11])

$$\Lambda = \int_{\mathcal{S}^{N-1}} \mathcal{K}(\lambda)\nu(d\lambda) \qquad \text{with} \qquad \mathcal{K}(\lambda) = \left\langle A^h\lambda, \lambda \right\rangle + \frac{1}{2}\operatorname{trace}\left[D(\lambda)\right] - \langle D(\lambda)\lambda, \lambda\rangle\,. \tag{11}$$

Here $\mathcal{S}^{N-1}$ denotes the $(N-1)$-dimensional unit sphere and ν denotes the invariant measure of the process $X(t)/|X(t)|$. The advantage of representation (11) is that even if the integral can not be calculated explicitly, one can analyse the kernel $\mathcal{K}(\lambda)$ to obtain sufficient conditions for asymptotic a.s. (in-)stability of the zero solution.

To use this representation we have to ensure that condition (10) is fulfilled. The diffusion matrices G_k^h are defined via the basis representation of the Laplacian. Consequently the noise is non-degenerate if the number of independent Wiener processes is at least equal to the dimension of the system, which is the case for Eq. (8). As the diffusion matrices G_i^h are diagonal and since

$$\operatorname{trace}\left[D(\lambda)\right] = \sigma^2 \left\langle \sum_{k=1}^{N} \left(G_k^h\right)^2 \lambda, \lambda \right\rangle,$$

we can write the kernel $\mathcal{K}(\lambda)$ as $\mathcal{K}(\lambda) = \langle K(\lambda)\lambda, \lambda\rangle$ with

$$K(\lambda) = A^h + \frac{\sigma^2}{2}\sum_{k=1}^{N}(G_k^h)^2 - D(\lambda). \tag{12}$$

By careful inspection of the eigenvalues of $K(\lambda)$, we expect $\mathcal{K}(\lambda)$ to be negative on $\mathcal{S}^{N-1}$ independently of σ. Therefore the zero solution of Eq. (8) is asymptotically a.s. stable, whereas it becomes asymptotically mean-square unstable for σ large enough. Then again the *explosion* of $\mathbb{E}|X(t)|^2$ depends on very rare trajectories and standard Monte Carlo estimation has the same difficulties to approximate $\mathbb{E}|X(t)|^2$ as in the one dimensional case, see Figure 2. In the next section we will treat the question if the mean-square stability behaviour can be recovered in Monte Carlo simulations by using variance reduction techniques.

3. Variance reduction techniques and rare event simulations

As we have already mentioned in Section 1, the Monte Carlo error can be controlled by either increasing the number of samples or by using variance reduction techniques to obtain estimators with smaller variances compared to standard Monte Carlo methods. The first approach, i.e. increasing the number of trajectories, faces the following severe limitations:

(1) A typical indicator that the number of simulated random variables might be insufficient is a substantial empirical variance of the standard Monte Carlo estimator. However, in our setting the rapid decay of the paths leads to a nearly vanishing *empirical* variance. Hence any estimation for the necessary number of simulated trajectories has to rely on the a priori knowledge of the rareness of the exploding trajectories. In general, the probability of these rare events is not known. Moreover, in situations where one can calculate the probability of these rare events, see e.g. geometric Brownian motion in Section 2, the probability of the occurrence of these trajectories is so low that the number of required trajectories is unreasonably high.

(2) A natural bound on the number of trajectories is imposed by the computational cost of the time integration method. This limits the possibility of increasing the number of numerical trajectories for high dimensional SODE systems.

Concerning variance reduction techniques we distinguish two types of methods. First, there are methods, such as Control Variates and multilevel Monte Carlo approaches, that reduce the variance by adding suitable control quantities to the standard estimators. The probability space and the distribution of the simulated random variables are unchanged. Thus problem II of the standard Monte Carlo estimator, i.e. the absence of rare events, is still present for this type of techniques. To see this, we examine the multilevel Monte Carlo estimator in Section 3.1 and analyse its disadvantages for our setting.

Second, there are methods, such as importance sampling, that change the underlying probability measure and therefore the distribution of the simulated trajectories. By being appropriately tuned, these methods are able to enforce the rare event of exploding trajectories. In Section 3.2, we present an importance sampling technique based on Girsanov's theorem where we modify the drift such that we simulate realisations of SODE systems with slower decaying trajectories. With the choice of constant weight functions in the Girsanov transformation, the resulting system is still linear and the computational cost to obtain a single realisation is of the same order as for the original system. Note that we can also apply additional variance reduction techniques of the first type to the modified estimator. In the numerical illustrations at the end of Section 4, we choose a combination of the proposed importance sampling technique (for rare event simulation purposes) and a multilevel Monte Carlo approach (for additional variance reduction).

3.1. **Multilevel Monte Carlo methods (and their failure to solve Problem II).** In this section, we consider the multilevel Monte Carlo (MLMC) estimator introduced by [11, 6] as a variance reduction technique. For this, we denote by $f(\widehat{X}_\ell(t)), \ell = 0, \dots, L$, an approximation of $f(X(t))$ using a numerical time integration method on a geometrical hierarchy of time grids with time step sizes $\Delta_\ell = 2^{-l-\kappa}$. Here, we use the parameter κ to ensure that the numerical method is stable on the coarsest level. Furthermore, let $f(\widehat{X}_{-1}(t)) = 0$.

Then, the multilevel Monte Carlo estimator is defined (see e.g. [6, 7])

$$E^L[f(\widehat{X}_L(t))] := \sum_{\ell=0}^{L} E_{N_\ell}[f(\widehat{X}_\ell(t)) - f(\widehat{X}_{\ell-1}(t))],$$

where N_ℓ independent realisations of $f(\widehat{X}_\ell(t)) - f(\widehat{X}_{\ell-1}(t))$ are used for the standard Monte Carlo estimator on each level.

There is a large literature on problems to which the MLMC approach has successfully been applied, see e.g. [7] for a survey. In many of these applications, a remarkable reduction of computational complexity compared to the standard Monte Carlo estimation is observed and a variance reduction of the underlying estimator is achieved.

In [6], a result on the optimal choice of levels L and of numbers of realisations $N_\ell, \ell = 0, \dots, L$, used on each of these levels is given for the multilevel Monte Carlo estimator. However, the number of realisations depends on the variance of the standard Monte Carlo estimators on each level, i.e. N_ℓ depends on the constant $C_\ell > 0$ in the estimate

$$\text{Var}\left[E_{N_\ell}[f(\widehat{X}_\ell(t)) - f(\widehat{X}_{\ell-1}(t))]\right] \leq C_\ell 2^{-\ell} \tag{13}$$

for all $\ell = 0, \dots, L$. By considering the variance of the Monte Carlo estimator on level 0, we obtain

$$\mathrm{Var}(E_{N_0}[f(\widehat{X}_0(t))]) = \frac{1}{N_0^2}\sum_{i=1}^{N_0}\mathrm{Var}(f(\widehat{X}_0^{(i)}(t))) = \frac{1}{N_0}\mathrm{Var}(f(\widehat{X}_0^{(1)}(t))).$$

Given an appropriately chosen initial time step size Δt_0, $\mathrm{Var}(f(\widehat{X}_0(t)))$ can be assumed to be close to $\mathrm{Var}(f(X_0(t)))$. Consequently, for SODEs where the zero solution is stable in the almost sure sense but mean-square unstable, we can deduce that the constant C_0 becomes unreasonably large by the same considerations as for the standard Monte Carlo estimator. Thus, we also have for the multilevel Monte Carlo estimator the performance problem that we need prohibitively many realisations (at least on the coarsest level) to obtain a certain accuracy. Although a reduction of computational cost is achieved by using the coarser time grids compared to the standard Monte Carlo approach, it is still not possible to simulate so many realisations that the rare trajectories for reproducing the mean-square instability occur with sufficient frequency.

In the standard MLMC algorithm proposed in [6, 7], the number of realisations on each level is computed by using the optimal choice of N_ℓ based on empirical estimators for the variance of $f(\widehat{X}_\ell(t)) - f(\widehat{X}_{\ell-1}(t))$. However, the empirical estimators for the variance are essentially zero due to the gap between almost sure and mean-square dynamics. Thus a straightforward application of multilevel Monte Carlo techniques without focusing on rare event simulation is not an appropriate approach in our setting. For this reason, we propose below a rare event simulation procedure based on importance sampling where we force the exploding trajectories to happen more frequently. Afterwards we can improve the modified estimator by further variance reduction techniques, e.g. by multilevel Monte Carlo methods as above.

3.2. **Importance sampling.** The main idea of importance sampling (see e.g. [8, 3, 14]) is to change the underlying probability measure (and therefore the distribution of $X(t)$) in order to either reduce the variance of the random variable $f(X(t))$ or in rare event simulations, to increase the frequency of rare events in the Monte Carlo simulation.

Here we concentrate on the second aspect and transform the underlying probability measure in a delicate way such that the exploding trajectories happen more often, but the zero solution is still asymptotically a.s. stable.

A convenient way to transform the underlying probability measure is given by Girsanov's theorem (see [20]) which allows us to change the drift of the system. Let

$$\mathrm{d}\widetilde{X}(t) = \widetilde{A}^h\widetilde{X}(t)\,\mathrm{d}t + \sigma\sum_{i=1}^{N} G_i^h\widetilde{X}(t)\,\mathrm{d}\beta_i(t) \tag{14}$$

be a (possibly non-linear) system of SODEs with modified drift

$$\widetilde{A}^h = A^h - \sigma\sum_{i=1}^{N} d_i(t,\widetilde{X}(t))G_i^h$$

for a set of real-valued functions $d_j : [0,T] \times \mathbb{R}^d \to \mathbb{R}$. We define a new probability measure $\widetilde{\mathbb{P}}$ via the density process

$$\mathrm{d}\Theta(t) = \sum_{j=1}^{N} d_j(t, \widetilde{X}(t))\, \Theta(t)\, \mathrm{d}\beta_j(t). \tag{15}$$

Under certain conditions on the functions d_j, e.g. they fulfill Novikov's condition (see [20]) we can apply Girsanov's theorem, see e.g. [20, Theorem 46], which states that $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ are equivalent and the processes

$$\gamma_i(t) = \int_0^t d_i(s, \widetilde{X}(s))\, \mathrm{d}s + \beta_i(t)$$

are Wiener processes under $\widetilde{\mathbb{P}}$. Consequently we have

$$\mathbb{E}\left[f(X(t))\right] = \mathbb{E}\left[f(\widetilde{X}(t))\Theta(t)\right]$$

and we can use trajectories of Eq. (14) to estimate the quantity $\mathbb{E}\left[f(X(t))\right]$.
Concerning the choice of the weight functions d_j we stipulate the following conditions:

- The rare events of exploding trajectories should happen more frequently.
- The computational effort for simulating Eq. (14) should not be larger than that for simulating Eq. (8).

It is well-known in the literature (see [19]) that there exists an optimal choice $d_j^*(t, \widetilde{X}(t))$ for the weight functions for which the quantity $f(\widetilde{X}(t))\Theta(t)$ becomes deterministic and consequently the Monte Carlo estimator has variance 0. Of course the optimal value is typically not known, and approximated solutions of the corresponding Kolmogorov backward equation (as proposed in [19]) are non-trivial to obtain. Additionally any choice of $d_j(t, \widetilde{X}(t))$ which is not constant results in a possibly high-dimensional system of non-linear SODEs leading to additional numerical challenges as e.g. to ensure the non-negativity of the density process $\Theta(t)$. To circumvent this problem we propose using only constant weight functions d_j, which implies that the system (14) is linear and the density process $\Theta(t)$ can be calculated explicitly.

Remark 3.1. In special cases, the optimal choice of weight functions $d_j^*(t, \widetilde{X}(t))$ for the second moment are constants, e.g. for the geometric Brownian motion (3), where

$$\widetilde{X}^2(t)\Theta(t) = X_0^2 \exp\left[2\left(\lambda - d\sigma - \frac{\sigma^2}{2}\right)t - \frac{d^2}{2}t + (2\sigma + d)\,\beta(t)\right] \tag{16}$$

is deterministic for $d^* = -2\sigma$.

For multi-dimensional linear systems of SODEs with non-commuting drift and diffusion matrices, a suitable choice of the weight functions is generally more challenging.

Our idea is the following: the asymptotic stability of the zero solution of the transformed system (14) can be determined by its Lyapunov exponent Λ represented by (11) where the kernel $\widetilde{\mathcal{K}}(\cdot)$ now takes the form

$$\widetilde{\mathcal{K}}(\lambda) = \left\langle \left(A^h - \sigma \sum_{i=1}^{N} d_i G_i^h\right)\lambda, \lambda \right\rangle + \frac{1}{2}\,\mathrm{trace}\left[D(\lambda)\right] - \left\langle D(\lambda)\lambda, \lambda \right\rangle.$$

As before, see Eq. (12) in Section 2.2, the definiteness of $\widetilde{\mathcal{K}}(\cdot)$ is determined by the definiteness of the matrix

$$\widetilde{K}(\lambda) := A^h - \sigma \sum_{i=1}^{N} d_i G_i^h - D(\lambda) + \frac{\sigma^2}{2} \sum_{k=1}^{N} \left(G_k^h\right)^2$$

on $\mathcal{S}^{N-1}$. In fact for $\widetilde{K}(\lambda)$ (strictly) positive or negative definite, the function $\widetilde{\mathcal{K}}(\lambda)$ is (strictly) positive or negative definite and in consequence the top Lyapunov exponent is positive or negative. As such, obtaining a positive top Lyapunov exponent is not desirable, as then almost all the trajectories of the transformed system (14) would explode, which would result in a wildly varying Monte Carlo estimator. However, the matrix $\widetilde{K}(\lambda)$ can also be indefinite, which essentially means $\widetilde{\mathcal{K}}(\lambda) \geq 0$ on some subset of $\mathcal{S}^{N-1}$, possibly yielding a larger Lyapunov exponent than in the case of strictly negative definite $\widetilde{K}(\lambda)$. As the (negative) Lyapunov exponent Λ measures the rate of exponential decay of the trajectories of system (14), such an increase of Λ implies a slower decay of the solution trajectories. The main idea now is to establish exactly this situation with judicious choices of the d_i. We propose to choose the weights d_i in such a way that the largest eigenvalue of the matrix $\widetilde{K}(\lambda)$ is positive on $\mathcal{S}^{N-1}$. As the term $\sum_{i=1}^{N} d_i G_i^h$ is a diagonal matrix, we cannot modify off-diagonal entries. Therefore we use the weights d_i to compensate the influence of the diagonal entries of $-D(\lambda) + \frac{\sigma^2}{2} \sum_{k=1}^{N} \left(G_k^h\right)^2$ which are given by

$$\left(-D(\lambda) + \frac{\sigma^2}{2} \sum_{k=1}^{N} \left(G_k^h\right)^2\right)_{ii} = \sigma^2 (\frac{1}{2} - \lambda_i^2) \left(\sum_{k=1}^{N} (G_{k,ii}^h)^2\right).$$

Measuring the distance between the diagonals with a least-squares approach we define the optimisation problem

$$\underset{d_1,\ldots,d_N \in \mathbb{R}}{\operatorname{argmin}} \left[\operatorname{trace} \left[\left(\tilde{p} \sum_{i=1}^{N} \sigma^2 (G_i^h)^2 + \sigma \sum_{i=1}^{N} d_i G_i^h \right)^2 \right] \right]. \tag{17}$$

The parameter $\tilde{p} > 0$ allows us to control the definiteness of $\widetilde{\mathcal{K}}$, in principle one could choose $\widetilde{p}$ sufficiently large such that the Lyapunov exponent Λ gets positive.

The optimisation problem (17) requires us to solve a linear system of dimension N. Alternatively we can decompose (17) into N optimisation problems of the form

$$\underset{d_i \in \mathbb{R}}{\operatorname{argmin}} \left[\operatorname{trace} \left[\left(\tilde{p}\sigma^2 (G_i^h)^2 + \sigma d_i G_i^h \right)^2 \right] \right] = -\tilde{p}\sigma \frac{\sum_{k=1}^{N} (G_{i,kk}^h)^3}{\sum_{k=1}^{N} (G_{i,kk}^h)^2} \qquad 1 \leq i \leq N, \tag{18}$$

which can be solved analytically. For both cases the weights d_i with even indices i are essentially zero, whereas for odd indices the values of $|d_i|$ decrease with growing i, see Table 1 in Section 4 for the values d_i, where $N = 10$. The corresponding density process can be calculated explicitly and is given by

$$\Theta(t) = \exp \left[\left(-\sum_{i=1}^{N} d_i^2 \right) \frac{t}{2} + \sum_{i=1}^{N} d_i \beta_i(t) \right]. \tag{19}$$

Note that for the stability analysis of the discretised stochastic heat equation (8), we required that the noise is non-degenerate in the sense of Condition (10) to obtain the representation (11) for the top Lyapunov exponent. Therefore, the number of Wiener processes

was required to be greater or equal N. However, the importance sampling technique proposed above (and thus the optimisation problems (17) and (18)) can also be modified for Eq. (8), when the number J of Wiener processes is smaller than the dimension N by appropriately truncating the corresponding sums after the first J elements.

4. Numerical results

4.1. **Monte Carlo estimators.** Using the measure transformations from Section 3.2, we define the Monte Carlo estimator

$$\widetilde{E}_N[f(\widehat{X}(T)] = \frac{1}{N}\sum_{i=1}^{N} f(\widehat{X}^{(i)}(T))\Theta^{(i)}(T), \tag{20}$$

where $\widehat{X}^{(i)}(T)$ are numerical trajectories of the transformed system (14) and $\Theta^{(i)}(T)$ are the corresponding trajectories of the density process given by (19). For calculating the numerical trajectories we use the Euler-Maruyama method with a suitable time step size for the geometric Brownian motion (3), whereas for the stochastic heat equation (6) we employ the stochastic trapezoidal rule

$$\widehat{X}(t_{n+1}) = \widehat{X}(t_n) + \frac{\Delta t}{2}\widetilde{A}^h\widehat{X}(t_{n+1}) + \frac{\Delta t}{2}\widetilde{A}^h\widehat{X}(t_n) + \sigma\sum_{r=1}^{N} G_r\widehat{X}_n\Delta\beta_r^n,$$

with $\beta_r^n \overset{iid}{\sim} \mathcal{N}(0, \sqrt{t_{n+1}-t_n})$, since the numerical treatment of the discretised stochastic heat equation (8) requires us to use a numerical method with suitable stability properties. It is straightforward to extend (20) to a multilevel Monte Carlo estimator which we denote by $\widetilde{E}^L[f(\widehat{X}(T)]$. We want to estimate the quantity $\mathbb{E}|X(T)|^2$ and compare our estimators with single and multilevel Monte Carlo estimators without modification.

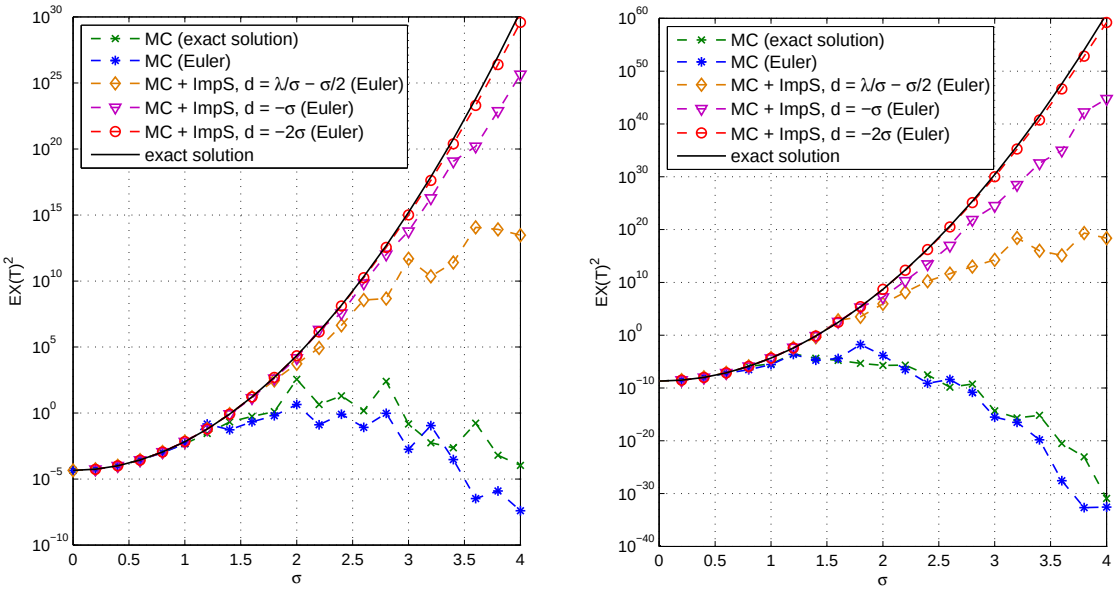
A reference solution can be obtained by using results from Section 2.2. For this, let $\mathbb{E}Y(T) = (\mathbb{E}Y_1(T), \ldots, \mathbb{E}Y_{N^2}(T))$ be the solution of the deterministic system (9) used for the mean-square stability analysis of the semi-discretised stochastic heat equation. Since $Y(t) = \mathrm{vec}(X(t)X(t)^T)$, we obtain a reference solution by

$$\mathbb{E}|X(T)|^2 = \sum_{i=1}^{N}\mathbb{E}X_i^2(T) = \sum_{i=0}^{N-1}\mathbb{E}Y_{(i\cdot N)+1}(T). \tag{21}$$

4.2. **Geometric Brownian motion.** In Figure 1, we compare the performance of the standard Monte Carlo estimator and the estimator defined in Eq. (20) applied to the geometric Brownian motion (3) with respect to different diffusion parameter σ. For these numerical experiments, we fix the drift parameter $\lambda = -1$ and the end time points $T = 5$ and $T = 10$. The time step size for the Euler-Maruyama scheme, $\Delta t = 2^{-8}$, is chosen such that the stability of the numerical method is guaranteed for all considered test cases. Furthermore, all Monte Carlo simulations are based on $M = 10^7$ independent realisations.

As we can see in Figure 1, the standard Monte Carlo estimator fails to reproduce the correct dynamics of the mean-square process. Note that there is hardly any difference in the qualitative behaviour of the standard Monte Carlo estimator using the exact solution or the numerical approximation based on the Euler-Maruyama scheme.

For the Girsanov transformation, we compare different choices of the parameter $\tilde{p}$. Here the solution of the optimisation problem (17) takes the form $d = -\sigma\tilde{p}$. First, the choice



(a) $T = 5$ (b) $T = 10$

FIGURE 1. Monte Carlo simulations for the geometric Brownian motion (3) for $\lambda = -1$, and varying σ. All Monte Carlo simulations use $M = 10^7$ trajectories.

$\tilde{p} = -\frac{\lambda}{\sigma^2} + \frac{1}{2}$ (note that $-\lambda > 0$ and $d = \lambda/\sigma - \sigma/2$) eliminates the deterministic part in the exact solution of the transformed system, i.e.

$$\widetilde{X}(t) = x_0 \exp\left(\sigma\beta(t)\right)$$

and for this reason, the zero solution is neither asymptotically a.s. stable nor unstable. The resulting estimation is already a substantial improvement compared to standard Monte Carlo estimators, however the mean-square process is still underestimated. Second, corresponding to our considerations in Section 3.2 increasing the parameter $\tilde{p}$ leads to a larger Lyapunov exponent. Since for large values of σ, it holds that $-\frac{\lambda}{\sigma^2} + \frac{1}{2} < 1$, we choose $d = -\sigma$, i.e. $\tilde{p} = 1$. For this choice the modified Monte Carlo estimator provides a better approximation of $\mathbb{E}\left[X^2(T)\right]$. Finally, the case $\tilde{p} = 2$ results in the optimal choice $d^* = -2\sigma$ (see Remark 3.1).

4.3. **Stochastic heat equation.** In Figure 2 and 3, we compare the qualitative behaviour of the standard and multilevel Monte Carlo estimators with and without measure transformation. We want to estimate the quantity $\mathbb{E}|X(T)|^2$ for different σ and time points T. For the Q-Wiener process, we choose the regularity parameter $r = 1$. To obtain reference solutions we carefully integrate the deterministic Equation (9) and use the representation (21). For the standard Monte Carlo simulations we use $M = 2^{11}$ realisations and a constant time step size $\Delta t = 2^{-11}$. For the multilevel Monte Carlo simulations we use 5 levels, where we calculate $M_\ell = 2^{11-\ell}$ trajectories with time step sizes $\Delta t_\ell = 2^{-(7+\ell)}$ on the corresponding refinement level $\ell = 1, \ldots, 5$. The Monte Carlo estimations after measure transformation

(denoted by *MC + ImpS* and *MLMC + ImpS* in the plots) are computed with the same set of parameters. Depending on the optimisation criteria of Section 3.2 we distinguish between *OPT1*, where we minimised Eq. (17), and *OPT2*, where we used Eq. (18). In Table 1 below, the computed weights $d_i, i = 1, \dots, N$, are given for $N = 10$ for both optimisation problems. In Figures 2 and 3 we see that the standard as well as the multilevel Monte Carlo estimator fail to approximate $\mathbb{E}|X(T)|^2$ correctly for $\sigma > 2$. Note that for $2 \leq \sigma \leq 3.4$ the zero solution of system (6) is still asymptotically mean-square stable, however the estimator substantially underestimates $\mathbb{E}|X(T)|^2$ due to the rapid decrease of the solution paths. For the estimators after measure transformation we see that the quality of the approximation strongly depends on the choice of the parameter $\tilde{p}$ as we have discussed in Section 3.2. For $\tilde{p} = 1/2$ the matrix $\widetilde{K}(\lambda)$ is still negative definite. The results are obviously better than for the untransformed estimators, nevertheless the approximations are not satisfying, especially for the larger time horizon $T = 5$. Increasing the parameter ($\tilde{p} = 1, \frac{3}{2}$) leads to an indefinite matrix $\widetilde{K}(\lambda)$. For both choices, we obtain good approximations of the qualitative behaviour of $\mathbb{E}|X(T)|^2$.

TABLE 1. List of computed non-zero weights $d_i, i = 1, 3, 5, 7, 9$, based on optimisation problems OPT1 (17) and OPT2 (18). All weights with even indices are essentially zero.

non-zero weights	OPT1	OPT2
d_1	$-1.201e-02$	$-1.385e-02$
d_3	$-3.877e-04$	$3.325e-03$
d_5	$-1.255e-04$	$2.316e-03$
d_7	$9.046e-05$	$1.469e-03$
d_9	$4.452e-07$	$5.832e-04$

5. Conclusions

In this paper we analysed the Monte Carlo error for estimating the second moment of the solution process of a linear system of SODEs with asymptotically a.s. stable but mean-square unstable zero solution. The mean-square instability of the equilibrium of such systems is due to very rare, exploding trajectories. Since the standard Monte Carlo estimator essentially computes the average over a finite number of realisations, the path-wise behaviour of the numerical trajectories compromises the estimation and the Monte Carlo error for estimating $\mathbb{E}f(X(t))$, where $f(x)$ is a function with (at least) quadratic growth in $|x|$, is huge.

We further remark that this situation also causes difficulties in numerical experiments testing the weak convergence properties of numerical methods for SDEs. Such numerical illustrations can be performed by simulations of the quantity $|E_M[\widehat{X}^2(T)] - \mathbb{E}X^2(T)|$. However, for problems where the path-wise behaviour of the numerical realisation is totally different to the dynamics of the mean-square process, the Monte Carlo error dominates the systematic error by several orders of magnitude. Reducing the time step size will not lead to the expected error reduction in numerical simulations.

As we showed in Section 3.1, with the straightforward application of standard variance reduction techniques it is not possible to overcome these difficulties in our setting unless they focus on rare event simulation. For this reason, we proposed an appropriately tuned

importance sampling technique which allows the trajectories to explode more frequently. We want to emphasise that the developed measure transformation does not change the linearity of the SDE system and can be precomputed. Thus, the proposed method does not increase the computational complexity of simulating a realisation of the quantity of interest. Furthermore, it is also possible to apply additional variance reduction methods to the modified estimator for further reducing its variance.

In this work we also compared standard and multilevel Monte Carlo methods with and without the proposed importance sampling technique and illustrated the improvement by the proposed method in numerical experiments for a spatially discretised stochastic heat equation.

References

[1] L. Arnold. *Stochastic Differential Equations: Theory and Applications*. Wiley-Interscience [John Wiley & Sons], New York-London-Sydney, 1974.

[2] L. Arnold. A formula connecting sample and moment stability of linear stochastic systems. *SIAM J. Appl. Math.*, 44(4):793–802, 1984.

[3] S. Asmussen and P.W. Glynn. *Stochastic Simulation. Algorithms and Analysis*. Stochastic Modelling and Applied Probability, 57. Springer, New York, 2007.

[4] E. Buckwar and T. Sickenberger. A structural analysis of asymptotic mean-square stability for multi-dimensional linear stochastic differential systems. *Applied Numerical Mathematics*, 62(7):842–859, 2012.

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

[6] M. B. Giles. Multilevel Monte Carlo path simulation. *Oper. Res.*, 56(3):607–617, 2008.

[7] M. B. Giles. Multilevel Monte Carlo methods. *Acta Numer.*, 24:259–328, 2015.

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

[9] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, second edition, 2013.

[10] P. Imkeller and C. Lederer. Some formulas for Lyapunov exponents and rotation numbers in two dimensions and the stability of the harmonic oscillator and the inverted pendulum. *Dyn. Syst.*, 16(1):29–61, 2001.

[11] A. Kebaier. Statistical Romberg extrapolation: a new variance reduction method and applications to option pricing. *Ann. Appl. Probab.*, 15(4):2681–2705, 2005.

[12] C. Kelly. Stochastic stability analysis of a reduced galactic dynamo model with perturbed α-effect. *Submitted 2015*, 2015.

[13] R. Khasminskii. *Stochastic Stability of Differential Equations*. Stochastic Modelling and Applied Probability, 66. Springer, Heidelberg, 2012.

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

[15] F. Kozin and S. Sugimoto. Relations between sample and moment stability for linear stochastic differential equations. In *Proceedings of the Conference on Stochastic Differential Equations and Applications (Park City, Utah, 1976)*, pages 145–162. Academic Press, New York, 1977.

[16] K. Liu. *Stability of Infinite Dimensional Stochastic Differential Equations with Applications*. Chapman & Hall/CRC Monographs and Surveys in Pure and Applied Mathematics 135, 2006.

[17] 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.

[18] X. Mao. *Stochastic Differential Equations and their Applications*. Horwood Publishing Series in Mathematics & Applications. Horwood Publishing Limited, Chichester, 1997.

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

[20] P. E. Protter. *Stochastic Integration and Differential Equations*, volume 21 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2005. Second edition. Version 2.1, Corrected third printing.

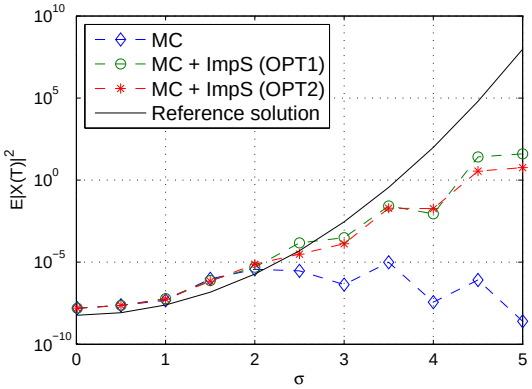
[21] W.-H. Steeb and Y. Hardy. *Matrix calculus and Kronecker product. A practical approach to linear and multilinear algebra*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, second edition, 2011.

[22] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Anal. Appl.*, 8(4):483–509 (1991), 1990.

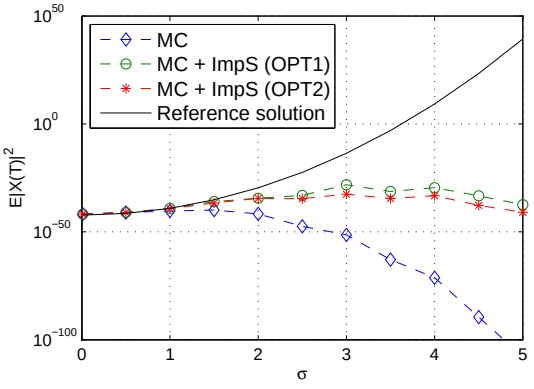
(Markus Ableidinger)
Institute for Stochastics
Johannes Kepler University Linz
Altenberger Strasse 69, 4040 Linz, Austria.
E-mail address: markus.ableidinger@jku.at

(Evelyn Buckwar)
Institute for Stochastics
Johannes Kepler University Linz
Altenberger Strasse 69, 4040 Linz, Austria.
E-mail address: evelyn.buckwar@jku.at

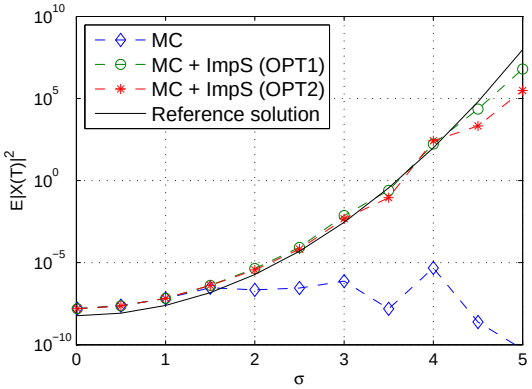
(Andreas Thalhammer)
Doktoratskolleg Computational Mathematics & Institute for Stochastics
Johannes Kepler University Linz
Altenberger Strasse 69, A–4040 Linz, Austria.
E-mail address: andreas.thalhammer@jku.at



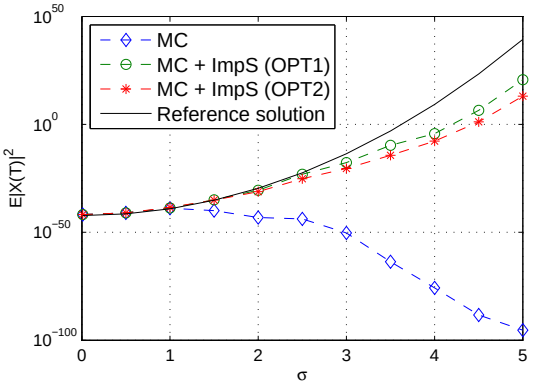
(a) $T = 1$, $\tilde{p} = \frac{1}{2}$



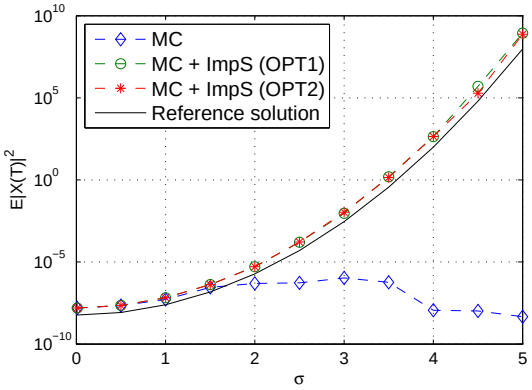
(b) $T = 5$, $\tilde{p} = \frac{1}{2}$



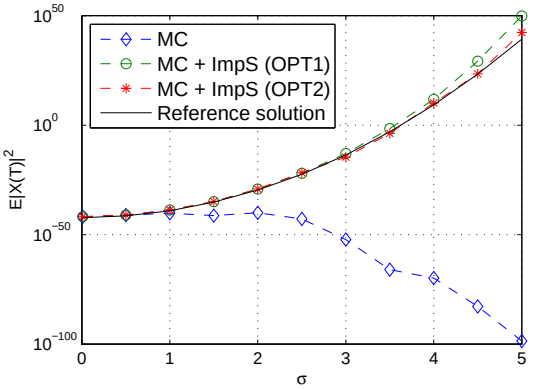
(c) $T = 1$, $\tilde{p} = 1$



(d) $T = 5$, $\tilde{p} = 1$

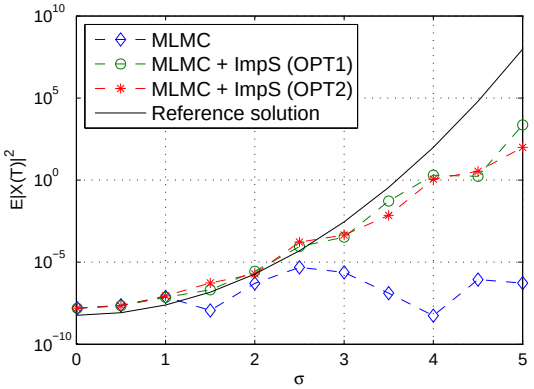


(e) $T = 1$, $\tilde{p} = \frac{3}{2}$

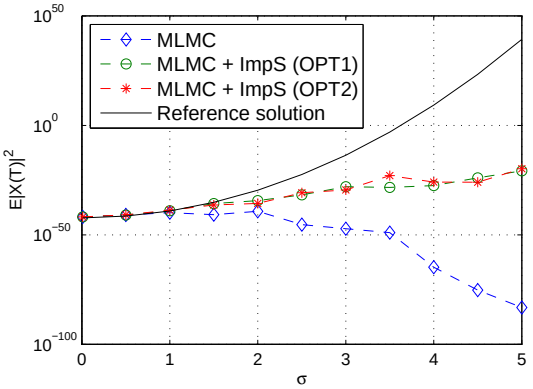


(f) $T = 5$, $\tilde{p} = \frac{3}{2}$

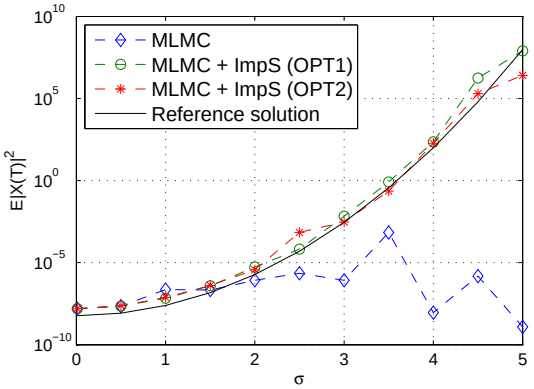
FIGURE 2. Monte Carlo methods for the spatially discretised stochastic heat equation (8) for $N = 10$ and varying σ. The reference solution is computed by Eq. (21).
Left column: Importance sampling with differently penalised weights at $T = 1$.
Right column: Importance sampling with differently penalised weights at $T = 5$.



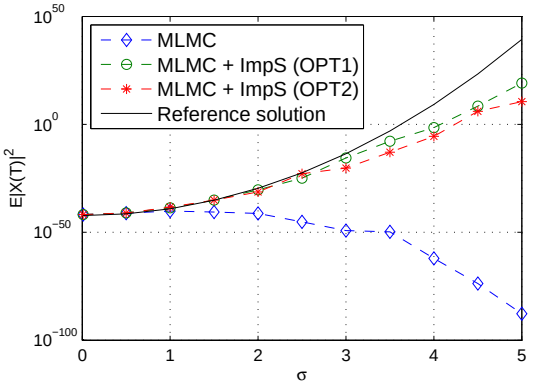
(a) $T = 1$, $\tilde{p} = \frac{1}{2}$



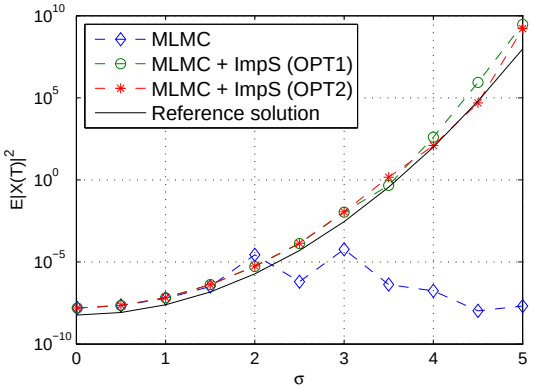
(b) $T = 5$, $\tilde{p} = \frac{1}{2}$



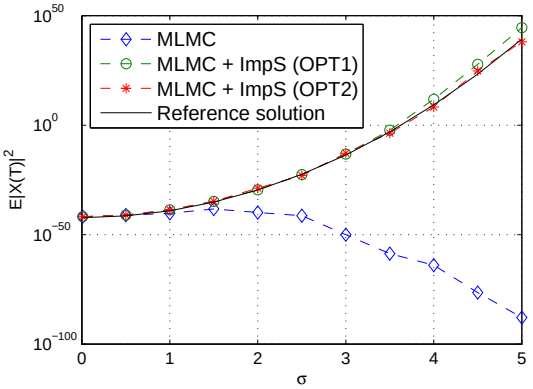
(c) $T = 1$, $\tilde{p} = 1$



(d) $T = 5$, $\tilde{p} = 1$



(e) $T = 1$, $\tilde{p} = \frac{3}{2}$



(f) $T = 5$, $\tilde{p} = \frac{3}{2}$

FIGURE 3. Multilevel Monte Carlo methods for the spatially discretised stochastic heat equation (8) for $N = 10$ and varying σ. The reference solution is computed by Eq. (21).
Left column: Importance sampling with differently penalised weights at $T = 1$.
Right column: Importance sampling with differently penalised weights at $T = 5$.