A Control Variate Method Driven by Diffusion Approximation

JOSSELIN GARNIER
Ecole polytechnique

LAURENT MERTZ
NYU-ECNU Institute of Mathematical Sciences at NYU Shanghai

Abstract

In this paper we introduce a control variate estimator for a quantity of interest that can be expressed as the expectation of a function of a random process, that is itself the solution of a differential equation driven by fast mean-reverting ergodic random forces. The control variate is built with the same function and with the limit diffusion process that approximates the original random process when the mean reversion time of the driving forces goes to 0. We propose a coupling of the original process and the limit diffusion process that gives a control variate estimator with small variance. We show that the correlation between the two processes indeed goes to 1 when the mean reversion time goes to 0 and we quantify the convergence rate, which allows us to characterize the variance reduction of the proposed control variate estimator. The efficiency of the method is illustrated on a few examples. © 2020 Wiley Periodicals LLC.

1 Introduction

In this paper we consider a system driven by external time-dependent random forces, and we aim to compute a quantity of interest that is the expectation of a function of the system. The system state is the solution of an ordinary differential equation (or a system of ordinary differential equations) driven by external forces which are stationary random processes. The random processes may have complicated spectra that have to be taken into account to compute the quantity of interest. This happens for instance in seismic probabilistic risk assessment studies or in the analysis of the structural performance of installations under seismic excitations [31] or under other loading sources such as wind or waves [17, 33].

For instance, the reliability of complex systems such as fixed or floating offshore wind turbines depends on its resistance against fatigue damage. Fatigue damage can be assessed by time-domain simulations in which the structure is subjected to wind, wave, and current loads [5]. The different loads can be described by (locally) stationary Gaussian processes with tabulated power spectral densities (such as the JONSWAP spectrum [15]). We may then wish to estimate the mean cumulative

Communications on Pure and Applied Mathematics, 0001–0038 (PREPRINT)

fatigue damage or a probability of failure that corresponds to the exceeding of a threshold value.

Monte Carlo simulations are standard methods for estimating the quantities of interest but they may be very time consuming. We look for an efficient variance reduction technique in this framework. It is known from the diffusion approximation theory [6, 8, 29] that the driving forces can often be approximated by white noises, and the responses of the system can then be modeled by stochastic differential equations. This allows us to implement a partial differential equation approach to compute the quantity of interest.

However, the bias due to the approximation of the original driving force by a white noise may be significant. A control variate method can compensate for such a bias [12]. Such a strategy has already been implemented in a Markov chain Monte Carlo context, where the goal is to sample from a complex invariant probability distribution of a Markov chain for which an approximate distribution has a known expression. The expectation of the approximate distribution then provides an initial guess, which can be corrected by simulating the two coupled processes to estimate the difference (in expected values) between the true distribution and the approximate distribution [13]. The implementation of a control variate method in our framework requires being able to simulate the system driven by the original driving force with its complicated spectrum and the limit system driven by the white noise in such a way that both systems are strongly correlated.

Unfortunately, most diffusion approximation results are established in a weak sense [6, 8]. Some strong results have been obtained but only when the drift is a term of order 1 [11, 19, 24], not when it is a zero-mean large term as we deal with in this paper. In this paper we build an efficient coupling between the original and limit systems, we establish a strong convergence result by quantifying the mean square distance between the original and limit processes, and we characterize the variance reduction of the control variate method. We show by our theoretical results and numerical simulations that the variance reduction can be dramatic.

Our method is relevant when the quality of the approximation of the driving forces by a white noise is moderate. If it is very accurate, then the quantity of interest can be estimated (up to a very small and negligible bias) by resolution of a Kolmogorov equation based on the limit diffusion system (or by a brute force Monte Carlo method applied to the limit system), so there is no need to apply a control variate method. If it is very poor, then the limit diffusion system is not correlated to the original system and the control variate method is not efficient. If it is moderate, then the bias of the estimation method that consists in replacing the original system by the limit one is nonnegligible, and the two systems are correlated, and the control variate method turns out to be very efficient.

The paper is organized as follows. In Section 2 we introduce the random ordinary differential equations addressed in this paper and state the main results of the paper. Motivated by applications in engineering mechanics and physics such as the study of the risk analysis of failure for mechanical structures subjected to random

vibrations [1,2,7,22] or the modeling of the stochastic dynamics of fluid-structure interaction in turbulent thermal convection [16], we also consider the case of multivalued ordinary differential equations. Sections 3–4 consider random ordinary differential equations. In Section 3 we state the diffusion approximation theorem that gives the convergence in probability of the original process to the limit process. In Section 4 we apply the control variate method to a few examples. The results are extended to the multivalued case in Sections 5–7. In particular, Section 7 reports numerical results for relevant engineering mechanics problems. The concluding remark of Section 8 connects our findings to the multilevel Monte Carlo literature.

2 Main Results

We consider the $\mathbb{R}^n$-valued process $\boldsymbol{X}^\varepsilon = (\boldsymbol{X}^\varepsilon_t)_{t\in[0,T]}$ solution of the ordinary differential equation (ODE)[1]

$$\frac{\mathrm{d}\boldsymbol{X}^\varepsilon}{\mathrm{d}t} = \boldsymbol{b}(\boldsymbol{X}^\varepsilon) + \frac{1}{\varepsilon}\boldsymbol{\sigma}(\boldsymbol{X}^\varepsilon)\boldsymbol{\eta}^\varepsilon, \quad \boldsymbol{X}^\varepsilon_0 = \boldsymbol{x}_0, \tag{2.1}$$

where $\boldsymbol{b}(\boldsymbol{x})$ is a Lipschitz function from $\mathbb{R}^n$ to $\mathbb{R}^n$, $\boldsymbol{\sigma}(\boldsymbol{x})$ is a function of class $\mathcal{C}^2$ with bounded derivatives from $\mathbb{R}^n$ to $\mathcal{M}_{n,d}(\mathbb{R})$, and $\boldsymbol{\eta}^\varepsilon$ is a $\mathbb{R}^d$-valued rapidly varying mean-reverting process, with a mean equal to 0, a unique invariant probability distribution, and a mean reversion time of the order of ε^2. More exactly, in this paper we address the case when $\boldsymbol{\eta}^\varepsilon$ is a multivariate d-dimensional Ornstein-Uhlenbeck process

$$\mathrm{d}\boldsymbol{\eta}^\varepsilon = \frac{\mathbf{K}}{\varepsilon}\,\mathrm{d}\boldsymbol{W}_t - \frac{\mathbf{A}}{\varepsilon^2}\,\boldsymbol{\eta}^\varepsilon\,\mathrm{d}t, \tag{2.2}$$

where $\mathbf{A}$ is a $d\times d$ matrix, whose eigenvalues have positive real parts, $\mathbf{K}$ is a $d\times d'$ matrix, and $\boldsymbol{W}$ is a d'-dimensional Brownian motion. This model is classical. It can be encountered in earthquake engineering [23] and also in finance [35]. It can model stationary Gaussian processes with very general spectra (see Section 3).

Our main motivation is to estimate a quantity of the form

$$I^\varepsilon \triangleq \mathbb{E}[F(\boldsymbol{X}^\varepsilon)] \tag{2.3}$$

for a fixed, small or moderate, parameter ε for a smooth real-valued function F defined on the space of continuous functions over $[0,T]$. We may think of $F(\boldsymbol{X}) = f(\boldsymbol{X}_T)$ where f is smooth with polynomial growth, or $F(\boldsymbol{X}) = \int_0^\mathsf{T} h(\boldsymbol{X}_s)\mathrm{d}s + f(\boldsymbol{X}_T)$. By the Feynman-Kac formula it is possible to get the value of I^ε for the model (2.1)–(2.2) by solving a parabolic equation, but this equation is formulated in a $d+n$-dimensional space, and it possesses large terms (of order ε^{-2}) that give rapid fluctuations. These rapid fluctuations need to be resolved by the numerical scheme, which imposes to take a time step smaller than ε^2. The numerical resolution (with a finite difference method) is, therefore, challenging, if not

[1] Throughout the paper, symbols of scalar quantities are printed in italic type, symbols of vectors are printed in bold italic type, and symbols of matrices are printed in bold type.

impossible, and we look for other resolution methods. It is also possible to estimate I^ε by a brute force Monte Carlo method. The Monte Carlo method, however, requires many simulations to get an accurate estimation, and each simulation requires resolving the rapid fluctuations at the scale ε^2, so we would like to propose an efficient variance reduction method. The main idea is to find a limiting process $\boldsymbol{X}^0$ that approximates $\boldsymbol{X}^\varepsilon$ in a strong sense when $\varepsilon \to 0$ and for which the value

$$I^0 = \mathbb{E}[F(\boldsymbol{X}^0)] \tag{2.4}$$

is known or can be estimated efficiently. It is then possible to propose a control variate method to estimate I^ε for a fixed ε.

We consider the limiting $\mathbb{R}^n$-valued process $\boldsymbol{X}^0$ solution of the stochastic differential equation (SDE)

$$\mathrm{d}\boldsymbol{X}^0 = \widetilde{\boldsymbol{b}}(\boldsymbol{X}^0)\mathrm{d}t + \boldsymbol{\Gamma}(\boldsymbol{X}^0)\mathrm{d}\boldsymbol{W}_t, \tag{2.5}$$

where $\boldsymbol{X}^0$ shares the same driving Brownian motion as $\boldsymbol{\eta}$, with the functions $\widetilde{\boldsymbol{b}}(\boldsymbol{x})$ from $\mathbb{R}^n$ to $\mathbb{R}^n$ and $\boldsymbol{\Gamma}(\boldsymbol{x})$ from $\mathbb{R}^n$ to $\mathcal{M}_{n,d'}(\mathbb{R})$ given by

$$\widetilde{b}_j(\boldsymbol{x}) \triangleq b_j(\boldsymbol{x}) + \sum_{i=1}^{n} \left((\partial_{x_i}\boldsymbol{\sigma}(\boldsymbol{x}))\mathbf{A}^{-1}\mathbf{C}\boldsymbol{\sigma}(\boldsymbol{x})^{\mathsf{T}}\right)_{ji}, \tag{2.6}$$

$$\boldsymbol{\Gamma}(\boldsymbol{x}) \triangleq \boldsymbol{\sigma}(\boldsymbol{x})\mathbf{A}^{-1}\mathbf{K}, \tag{2.7}$$

and $\mathbf{C}$ is the $d \times d$ matrix defined by

$$\mathbf{C} \triangleq \int_0^\infty e^{-\mathbf{A}s}\mathbf{K}\mathbf{K}^{\mathsf{T}}e^{-\mathbf{A}^{\mathsf{T}}s}\,\mathrm{d}s, \tag{2.8}$$

where the superscript T stands for "transpose." The matrix $\mathbf{C}$ is the covariance matrix of the stationary distribution of the process $\boldsymbol{\eta}^\varepsilon$. We show in Proposition 3.5 that the continuous process $(\boldsymbol{X}^\varepsilon - \boldsymbol{X}^0)$ converges in probability to 0 as $\varepsilon \to 0$. The fact that $\boldsymbol{X}^\varepsilon$ converges in distribution to $\boldsymbol{X}^0$ is well-known [8, chap. 6], but here we get a stronger result with a particular coupling between the two processes $\boldsymbol{X}^\varepsilon$ and $\boldsymbol{X}^0$, which is needed to implement the control variate method that we have in mind.

The form of the limiting equation (2.5) is not surprising. Indeed, by (2.2), we can anticipate that $\frac{1}{\varepsilon}\boldsymbol{\eta}^\varepsilon \mathrm{d}t \simeq \mathbf{A}^{-1}\mathbf{K}\,\mathrm{d}\boldsymbol{W}_t$ + corrections, which explains the form (2.7) of the diffusion $\boldsymbol{\Gamma}$ by substitution into (2.1). The form (2.6) of the drift $\widetilde{\boldsymbol{b}}$ is a manifestation of the Itô-versus-Stratonovich problem [32]. This problem is whether one should interpret the stochastic integral in the limiting equation in Itô's sense, the Stratonovich sense, or another sense. The Wong-Zakai theory [36] claims that the limiting diffusion should be a Stratonovich equation when $d = 1$. Indeed,

equations (2.6)–(2.8) then reduce to $\boldsymbol{\Gamma}(\boldsymbol{x}) = \frac{1}{A}\boldsymbol{\sigma}(\boldsymbol{x})\mathbf{K}$, $\mathbf{C} = \frac{1}{2A}\mathbf{K}\mathbf{K}^{\mathsf{T}}$,

$$\begin{aligned}\widetilde{b}_j(\boldsymbol{x}) - b_j(\boldsymbol{x}) &= \frac{1}{2A^2}\sum_{i=1}^{n}\big(\partial_{x_i}(\boldsymbol{\sigma}(\boldsymbol{x})\mathbf{K})(\boldsymbol{\sigma}(\boldsymbol{x})\mathbf{K})^{\mathsf{T}}\big)_{ji}\\ &= \frac{1}{2}\sum_{i=1}^{n}\big(\partial_{x_i}\boldsymbol{\Gamma}(\boldsymbol{x})\boldsymbol{\Gamma}(\boldsymbol{x})^{\mathsf{T}}\big)_{ji},\end{aligned}$$

so that (2.5) can be written as

$$\mathrm{d}\boldsymbol{X}^0 = \boldsymbol{b}(\boldsymbol{X}^0)\mathrm{d}t + \boldsymbol{\Gamma}(\boldsymbol{X}^0)\circ\mathrm{d}\boldsymbol{W}_t, \tag{2.9}$$

where $\circ$ stands for the Stratonovich integral, because

$$\begin{aligned}\big(\boldsymbol{\Gamma}(\boldsymbol{X}^0)\circ\mathrm{d}\boldsymbol{W}_t\big)_j &= \big(\boldsymbol{\Gamma}(\boldsymbol{X}^0)\mathrm{d}\boldsymbol{W}_t\big)_j + \frac{1}{2}\sum_{i=1}^{n}\sum_{j'=1}^{d'}\partial_{x_i}\Gamma_{jj'}(\boldsymbol{X}^0)\mathrm{d}\big\langle X_i^0, W_{j'}\big\rangle_t\\ &= \big(\boldsymbol{\Gamma}(\boldsymbol{X}^0)\mathrm{d}\boldsymbol{W}_t\big)_j + \frac{1}{2}\sum_{i=1}^{n}\big(\partial_{x_i}\boldsymbol{\Gamma}(\boldsymbol{X}^0)\boldsymbol{\Gamma}(\boldsymbol{X}^0)^{\mathsf{T}}\big)_{ji}\,\mathrm{d}t.\end{aligned}$$

The form (2.9) is valid when $d = 1$, and it looks simpler than (2.5), but we have chosen to write the stochastic integral in (2.5) in Itô's sense and to add the appropriate Itô-Stratonovich drift correction $\widetilde{\boldsymbol{b}} - \boldsymbol{b}$, because it is the natural starting point for numerical schemes [20] and is the appropriate form to express the martingale problems used in the proofs (see Appendix and [9]). When $d > 1$ the difference between $\widetilde{\boldsymbol{b}}$ and $\boldsymbol{b}$ is an Itô-Stratonovich correction that is more complex, and the limiting equation (2.5) cannot be reduced to (2.9).

We can now introduce the Monte Carlo method for the estimation of I^ε. Let $\boldsymbol{W}^k$, $k = 1, \ldots, N$, be N independent and identically distributed d'-dimensional Brownian motions. We consider three Monte Carlo-type estimators of I^ε:

(1) The brute force Monte Carlo estimator is

$$\widehat{I}_N^\varepsilon \triangleq \frac{1}{N}\sum_{k=1}^{N} F(\boldsymbol{X}^\varepsilon(\boldsymbol{W}^k)), \tag{2.10}$$

where $\boldsymbol{X}^\varepsilon(\boldsymbol{W}^k)$ is the solution of (2.1)–(2.2) with $\boldsymbol{W}^k$. The estimator $\widehat{I}_N^\varepsilon$ is unbiased and its variance is

$$\mathrm{Var}(\widehat{I}_N^\varepsilon) = \frac{1}{N}\mathrm{Var}(F(\boldsymbol{X}^\varepsilon)). \tag{2.11}$$

It is asymptotically normal as $N \to +\infty$:

$$\sqrt{N}\big(\widehat{I}_N^\varepsilon - I^\varepsilon\big) \xrightarrow{\text{dist}} \mathcal{N}\big(0, \sigma_{I^\varepsilon}^2\big), \tag{2.12}$$

with the asymptotic variance

$$\sigma_{I^\varepsilon}^2 = \mathrm{Var}(F(\boldsymbol{X}^\varepsilon)), \tag{2.13}$$

which has the following behavior as $\varepsilon \to 0$ when F is continuous and bounded (because $\boldsymbol{X}^\varepsilon$ weakly converges to $\boldsymbol{X}^0$):

$$\sigma^2_{I^\varepsilon} = \mathrm{Var}(F(\boldsymbol{X}^0)) + o(1). \tag{2.14}$$

(2) The control variate estimator [12] is

$$\widehat{J}^\varepsilon_N \triangleq I^0 + \frac{1}{N}\sum_{k=1}^N F(\boldsymbol{X}^\varepsilon(\boldsymbol{W}^k)) - F(\boldsymbol{X}^0(\boldsymbol{W}^k)), \tag{2.15}$$

where $I^0 = \mathbb{E}[F(\boldsymbol{X}^0)]$ is supposed to be known exactly (or with high accuracy). The value I^0 can be obtained by solving a Kolmogorov equation in an n-dimensional framework and without a large term; if this is not possible (because n is too large for instance), then the value I^0 can be obtained by a brute force Monte Carlo method, which is easier than for I^ε because there is no large term of order ε^{-2}, so that a standard Euler scheme for stochastic differential equations can be used [20]. The control variate estimator $\widehat{J}^\varepsilon_N$ is unbiased, and its variance is

$$\mathrm{Var}(\widehat{J}^\varepsilon_N) = \frac{1}{N}\mathrm{Var}(F(\boldsymbol{X}^\varepsilon) - F(\boldsymbol{X}^0)). \tag{2.16}$$

It is asymptotically normal as $N \to +\infty$:

$$\sqrt{N}\big(\widehat{J}^\varepsilon_N - I^\varepsilon\big) \xrightarrow{\text{dist}} \mathcal{N}\big(0, \sigma^2_{J^\varepsilon}\big), \tag{2.17}$$

with the asymptotic variance

$$\sigma^2_{J^\varepsilon} = \mathrm{Var}(F(\boldsymbol{X}^\varepsilon) - F(\boldsymbol{X}^0)). \tag{2.18}$$

When F is continuous and bounded, we have by Proposition 3.5 that $\sigma^2_{J^\varepsilon}$ goes to 0 as $\varepsilon \to 0$. More quantitatively, if $F(\boldsymbol{X}) = f(\boldsymbol{X}_T)$ for a smooth f with bounded derivatives, then the asymptotic variance has the following behavior as $\varepsilon \to 0$ (by Lemma 3.8):

$$\sigma^2_{J^\varepsilon} \leq C\,\varepsilon^2. \tag{2.19}$$

The order of magnitude ε^2 of the asymptotic variance of $\widehat{J}^\varepsilon_N$ is confirmed by the numerical simulations that we report in Section 4.

(3) The theoretical optimal control variate estimator is

$$\widehat{O}^\varepsilon_N \triangleq \rho^\varepsilon I^0 + \frac{1}{N}\sum_{k=1}^N F(\boldsymbol{X}^\varepsilon(\boldsymbol{W}^k)) - \rho^\varepsilon F(\boldsymbol{X}^0(\boldsymbol{W}^k)), \tag{2.20}$$

with

$$\rho^\varepsilon = \mathrm{Cov}(F(\boldsymbol{X}^\varepsilon), F(\boldsymbol{X}^0))/\mathrm{Var}(F(\boldsymbol{X}^0)). \tag{2.21}$$

This estimator is unbiased and has the minimal variance

$$\mathrm{Var}\big(\widehat{O}^\varepsilon_N\big) = \frac{1}{N}\mathrm{Var}(F(\boldsymbol{X}^\varepsilon) - \rho^\varepsilon F(\boldsymbol{X}^0)), \tag{2.22}$$

amongst all control variate estimators of the form

$$\rho I^0 + \frac{1}{N}\sum_{k=1}^{N} F(\boldsymbol{X}^\varepsilon(\boldsymbol{W}^k)) - \rho F(\boldsymbol{X}^0(\boldsymbol{W}^k)).$$

Note that $\rho = 0$ corresponds to the brute force Monte Carlo estimator $\widehat{I}_N^\varepsilon$, $\rho = 1$ corresponds to the control variate estimator $\widehat{J}_N^\varepsilon$, and $\rho = \rho^\varepsilon$ corresponds to the optimal control variate estimator $\widehat{O}_N^\varepsilon$. The estimator $\widehat{O}_N^\varepsilon$ is asymptotically normal as $N \to +\infty$:

$$\sqrt{N}\big(\widehat{O}_N^\varepsilon - I^\varepsilon\big) \xrightarrow{\text{dist}} \mathcal{N}\big(0, \sigma_{O^\varepsilon}^2\big), \tag{2.23}$$

with the asymptotic variance

$$\sigma_{O^\varepsilon}^2 = \mathrm{Var}(F(\boldsymbol{X}^\varepsilon) - \rho^\varepsilon F(\boldsymbol{X}^0)). \tag{2.24}$$

The estimator $\widehat{O}_N^\varepsilon$ is, however, not practical as it depends on ρ^ε, which is unknown. The practical optimal control variate estimator [12] is

$$\widehat{K}_N^\varepsilon \triangleq \widehat{\rho}_N^\varepsilon I^0 + \frac{1}{N}\sum_{k=1}^{N} F(\boldsymbol{X}^\varepsilon(\boldsymbol{W}^k)) - \widehat{\rho}_N^\varepsilon F(\boldsymbol{X}^0(\boldsymbol{W}^k)), \tag{2.25}$$

where $\widehat{\rho}_N^\varepsilon$ is the empirical correlation

$$\widehat{\rho}_N^\varepsilon = \frac{\sum_{k=1}^N (F(\boldsymbol{X}^\varepsilon(\boldsymbol{W}^k)) - \widehat{I}_N^\varepsilon)(F(\boldsymbol{X}^0(\boldsymbol{W}^k)) - \widehat{I}_N^0)}{\sum_{k=1}^N (F(\boldsymbol{X}^0(\boldsymbol{W}^k)) - \widehat{I}_N^0)^2}, \tag{2.26}$$

with $\widehat{I}_N^\varepsilon = \frac{1}{N}\sum_{k=1}^N F(\boldsymbol{X}^\varepsilon(\boldsymbol{W}^k))$ as in (2.10) and

$$\widehat{I}_N^0 = \frac{1}{N}\sum_{k=1}^{N} F(\boldsymbol{X}^0(\boldsymbol{W}^k)).$$

This estimator is a practical and approximate version of the theoretical optimal control variate estimator $\widehat{O}_N^\varepsilon$ in which the unknown correlation coefficient ρ^ε has been replaced by its empirical estimator $\widehat{\rho}_N^\varepsilon$.

The estimator $\widehat{K}_N^\varepsilon$ may be slightly biased and may have a variance slightly larger than (2.22) because of the empirical estimation of ρ^ε. $\widehat{K}_N^\varepsilon$ is, however, asymptotically normal with an asymptotic variance that is the same as that of the optimal estimator $\widehat{O}_N^\varepsilon$, as shown by the following proposition.

PROPOSITION 2.1. *As $N \to +\infty$,*

$$\sqrt{N}\big(\widehat{K}_N^\varepsilon - I^\varepsilon\big) \xrightarrow{\text{dist}} \mathcal{N}\big(0, \sigma_{K^\varepsilon}^2\big), \tag{2.27}$$

with

$$\sigma_{K^\varepsilon}^2 = \sigma_{O^\varepsilon}^2 = \mathrm{Var}(F(\boldsymbol{X}^\varepsilon) - \rho^\varepsilon F(\boldsymbol{X}^0)). \tag{2.28}$$

Furthermore, if F is continuous and bounded, then $\sigma^2_{K^\varepsilon}$ goes to 0 as $\varepsilon \to 0$. If $F(\boldsymbol{X}) = f(\boldsymbol{X}_T)$, with f with bounded derivatives, then there exists $C > 0$ such that

$$\sigma^2_{K^\varepsilon} \leq C\varepsilon^2, \quad 0 \leq \sigma^2_{J^\varepsilon} - \sigma^2_{K^\varepsilon} \leq C\varepsilon^4. \tag{2.29}$$

PROOF. By the law of large numbers, $\widehat{\rho}^\varepsilon_N$ converges to ρ^ε as $N \to +\infty$. The convergence holds almost surely, hence in probability. We have

$$\widehat{K}^\varepsilon_N - I^\varepsilon = \big(\widehat{O}^\varepsilon_N - I^\varepsilon\big) - \big(\widehat{\rho}^\varepsilon_N - \rho^\varepsilon\big)\big(\widehat{I}^0_N - I^0\big),$$

so we get (2.27)–(2.28) from Slutsky's theorem.

Furthermore, we have

$$\sigma^2_{J^\varepsilon} - \sigma^2_{K^\varepsilon} = \mathrm{Var}(F(\boldsymbol{X}^0))(1-\rho^\varepsilon)^2.$$

If F is continuous and bounded, then ρ^ε goes to 1 and $\sigma^2_{J^\varepsilon}$ goes to 0 as $\varepsilon \to 0$ by Proposition 3.5. If $F(\boldsymbol{X}) = f(\boldsymbol{X}_T)$, then, by Lemma 3.8, $1-\rho^\varepsilon$ and $\sigma^2_{J^\varepsilon}$ are of order $O(\varepsilon^2)$ for small ε. This shows the desired result (2.29). □

Proposition 2.1 shows that the asymptotic variances of the estimators $\widehat{K}^\varepsilon_N$ and $\widehat{J}^\varepsilon_N$ are equivalent for vanishingly small ε and of the order of $O(\varepsilon^2)$, and that the asymptotic variance of the estimator $\widehat{K}^\varepsilon_N$ is slightly smaller than that of $\widehat{J}^\varepsilon_N$ for moderately small ε. These statements are confirmed by the numerical simulations that we report in Section 4.

In addition, motivated by the examples that we address in Section 7, we consider the case where the $\mathbb{R}^n$-valued process $\boldsymbol{X}^\varepsilon$ satisfies a multivalued ODE of the form

$$\frac{\mathrm{d}\boldsymbol{X}^\varepsilon}{\mathrm{d}t} + \partial\varphi(\boldsymbol{X}^\varepsilon) \ni \boldsymbol{b}(\boldsymbol{X}^\varepsilon) + \frac{1}{\varepsilon}\boldsymbol{\sigma}\,\boldsymbol{\eta}^\varepsilon, \quad \boldsymbol{X}^\varepsilon_0 = \boldsymbol{x}_0, \tag{2.30}$$

and the case where $\boldsymbol{X}^\varepsilon$ together with an $\mathbb{R}^m$-valued process $\boldsymbol{Z}^\varepsilon$ satisfy the multivalued ODE

$$\begin{cases} \dfrac{\mathrm{d}\boldsymbol{X}^\varepsilon}{\mathrm{d}t} + \partial\varphi(\boldsymbol{X}^\varepsilon) \ni \boldsymbol{b}^X(\boldsymbol{X}^\varepsilon, \boldsymbol{Z}^\varepsilon) + \dfrac{1}{\varepsilon}\boldsymbol{\sigma}\,\boldsymbol{\eta}^\varepsilon, & \boldsymbol{X}^\varepsilon_0 = \boldsymbol{x}_0, \\ \dfrac{\mathrm{d}\boldsymbol{Z}^\varepsilon}{\mathrm{d}t} + \partial\psi(\boldsymbol{Z}^\varepsilon) \ni \boldsymbol{b}^Z(\boldsymbol{X}^\varepsilon, \boldsymbol{Z}^\varepsilon), & \boldsymbol{Z}^\varepsilon_0 = z_0. \end{cases} \tag{2.31}$$

Here $\boldsymbol{\sigma} \in \mathcal{M}_{n,d}(\mathbb{R})$ is constant, $\boldsymbol{b}(\boldsymbol{x})$ from $\mathbb{R}^n$ to $\mathbb{R}^n$, $\boldsymbol{b}^Z(\boldsymbol{x}, z)$ from $\mathbb{R}^{n+m}$ to $\mathbb{R}^m$, and $\boldsymbol{b}^X(\boldsymbol{x}, z)$ from $\mathbb{R}^{n+m}$ to $\mathbb{R}^n$ are Lipschitz functions. The operators $\partial\varphi$ and $\partial\psi$ are the subdifferentials of the lower semicontinuous (l.s.c.) convex functions φ from $\mathbb{R}^n$ to $[0, +\infty]$ and ψ from $\mathbb{R}^m$ to $[0, +\infty]$. Stronger hypotheses will be assumed on φ compared to ψ as explained in Section 5 and important examples motivate the two situations as shown in Section 7. We observe that:

(1) The multivalued operators that appear in the differential inclusions (2.30) and (2.31) are subdifferential of convex functions, therefore existence and uniqueness are guaranteed [4, p. 72]. For the reader's convenience, existence and uniqueness results are given in Appendix C.

(2) There is an alternative formulation using the language of variational inequalities that is equivalent to differential inclusions. Equation (2.30) is equivalent to

$$\forall \boldsymbol{\xi} \in \mathbb{R}^n,\ \forall t > 0, \quad \left(\boldsymbol{b}(\boldsymbol{X}^\varepsilon) + \frac{1}{\varepsilon}\boldsymbol{\sigma}\,\boldsymbol{\eta}^\varepsilon - \frac{\mathrm{d}\boldsymbol{X}^\varepsilon}{\mathrm{d}t} \right) \cdot \left(\boldsymbol{\xi} - \boldsymbol{X}^\varepsilon \right) + \varphi(\boldsymbol{X}^\varepsilon) \leq \varphi(\boldsymbol{\xi}),$$

with $\boldsymbol{X}_0^\varepsilon = \boldsymbol{x}_0$, and equation (2.31) is equivalent to

$$\begin{aligned} &\forall \boldsymbol{\xi} \in \mathbb{R}^n,\ \forall \boldsymbol{\zeta} \in \mathbb{R}^m,\ \forall t > 0, \\ &\left(\boldsymbol{b}^X(\boldsymbol{X}^\varepsilon, \boldsymbol{Z}^\varepsilon) + \frac{1}{\varepsilon}\boldsymbol{\sigma}\,\boldsymbol{\eta}^\varepsilon - \frac{\mathrm{d}\boldsymbol{X}^\varepsilon}{\mathrm{d}t} \right) \cdot \left(\boldsymbol{\xi} - \boldsymbol{X}^\varepsilon \right) + \varphi(\boldsymbol{X}^\varepsilon) \leq \varphi(\boldsymbol{\xi}), \\ &\left(\boldsymbol{b}^Z(\boldsymbol{X}^\varepsilon, \boldsymbol{Z}^\varepsilon) - \frac{\mathrm{d}\boldsymbol{Z}^\varepsilon}{\mathrm{d}t} \right) \cdot \left(\boldsymbol{\zeta} - \boldsymbol{Z}^\varepsilon \right) + \psi(\boldsymbol{Z}^\varepsilon) \leq \psi(\boldsymbol{\zeta}), \end{aligned}$$

with $\boldsymbol{X}_0^\varepsilon = \boldsymbol{x}_0$ and $\boldsymbol{Z}_0^\varepsilon = z_0$.

Propositions 5.1 and 5.2 show that the multivalued process $\boldsymbol{X}^\varepsilon$ strongly converges to a limiting process solution of a multivalued SDE. Equations (6.8) and (6.9) show that the control variate estimators have asymptotic variances of order ε^2 for (2.30) and ε for (2.31).

To demonstrate the efficiency of our method on a practical problem, we consider a two degrees of freedom (TDOF) system as shown in Figure 2.1. It describes a broad class of TDOF structures, including a two-story building as presented in [34, figure 4.6(a)].

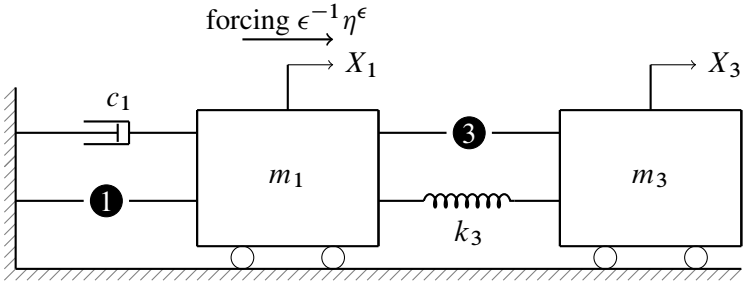

FIGURE 2.1. A rheological model of a two degrees of freedom (TDOF) system. Two masses m_1 and m_3 are associated in series with elements that are themselves an association of dampers and springs. Elements ① and ③ represent a spring and a damper respectively, both possibly nonlinear or hysteretic. Here c_1 is the damping coefficient associated to the linear damper connecting the mass m_1 to the foundation and k_3 is the stiffness coefficient of the linear spring linking the masses m_1 and m_3. A random forcing $\epsilon^{-1}\eta^\epsilon$ is applied to the mass m_1 (e.g., wind forces on a two-story building).

When the external force η^ε is a colored noise such as an Ornstein-Uhlenbeck process, the equation of motion can be written in the form of equation (2.1) with

$n = 4$, where $(X_1^\varepsilon, X_2^\varepsilon)$, resp., $(X_3^\varepsilon, X_4^\varepsilon)$, represents the position and the velocity of the mass m_1, resp., m_3, shown in Figure 2.1. Many nonlinear behaviors enter into this framework; we have in mind a nonlinear spring of the linear-plus-quadratic cubic type and a nonlinear damper of the linear-plus-quadratic type (see Example 4.3). Similarly, Equations (2.30) and (2.31) arise in the description of nonlinear behaviors with hysteresis such as elasto-plasticity and friction, see [34, chap. 8] and Section 7 (Example 7.3). In Figures 2.2-2.3 we compare the behaviors of the brute force Monte Carlo estimator $\widehat{I}_N^\varepsilon$ with the ones of the control variate estimators $\widehat{J}_N^\varepsilon$ and $\widehat{K}_N^\varepsilon$. We also plot the empirical estimators of the asymptotic variances of the estimators $\widehat{I}_N^\varepsilon$, $\widehat{J}_N^\varepsilon$, and $\widehat{K}_N^\varepsilon$ as described in Section 7.3. In addition, for each value of $\varepsilon \in \{0.1, 0.5, 0.9\}$, error bars (95% confidence interval) are shown for each of the estimators (in $\widehat{I}_N^\varepsilon$, $\widehat{J}_N^\varepsilon$, $\widehat{K}_N^\varepsilon$ order from left to right). Here I^0 is obtained by a massive Monte Carlo estimation of the limit process, which is possible with a coarse grid step as there is no large term involved. We can observe that the control variate estimator $\widehat{K}_N^\varepsilon$ has always the minimal variance. When ε is large and the original system and the limit system are poorly correlated $\rho^\varepsilon \simeq 0$, it behaves as the standard Monte Carlo estimator $\widehat{I}_N^\varepsilon$. When ε is small and the original system and the limit system are strongly correlated $\rho^\varepsilon \simeq 1$, it behaves as the control variate estimator $\widehat{J}_N^\varepsilon$. We can also observe that the variance reduction is by a factor of order ε^2 when the quantity to be estimated is the expectation of a smooth function, while it is of order ε when the quantity to be estimated is the expectation of an indicator function.

3 Diffusion Approximation for a Driving Multivariate Ornstein-Uhlenbeck Process

We consider the $\mathbb{R}^n$-valued process $\boldsymbol{X}^\varepsilon$ solution of the ODE (2.1) when $\boldsymbol{\eta}^\varepsilon$ is the multivariate d-dimensional Ornstein-Uhlenbeck process (2.2). We give several explicit examples.

EXAMPLE 3.1. *η^ε is a one-dimensional Ornstein-Uhlenbeck process, $d = d' = 1$, $A, K > 0$,*

$$\mathrm{d}\eta^\varepsilon = -\frac{A}{\varepsilon^2}\eta^\varepsilon \,\mathrm{d}t + \frac{K}{\varepsilon}\,\mathrm{d}W_t. \tag{3.1}$$

EXAMPLE 3.2. *$\boldsymbol{\eta}^\varepsilon$ is a Langevin process:*

$$\mathrm{d}\eta_1^\varepsilon = \frac{1}{\varepsilon^2}\eta_2^\varepsilon \,\mathrm{d}t, \tag{3.2}$$

$$\mathrm{d}\eta_2^\varepsilon = -\frac{1}{\varepsilon^2}\big[\mu\eta_1^\varepsilon + \gamma\eta_2^\varepsilon\big]\mathrm{d}t + \frac{K}{\varepsilon}\,\mathrm{d}W_t, \tag{3.3}$$

which corresponds to $d = 2$, $d' = 1$,

$$\mathbf{A} = \begin{pmatrix} 0 & -1 \\ \mu & \gamma \end{pmatrix} \quad \textit{and} \quad \mathbf{K} = \begin{pmatrix} 0 \\ K \end{pmatrix}.$$

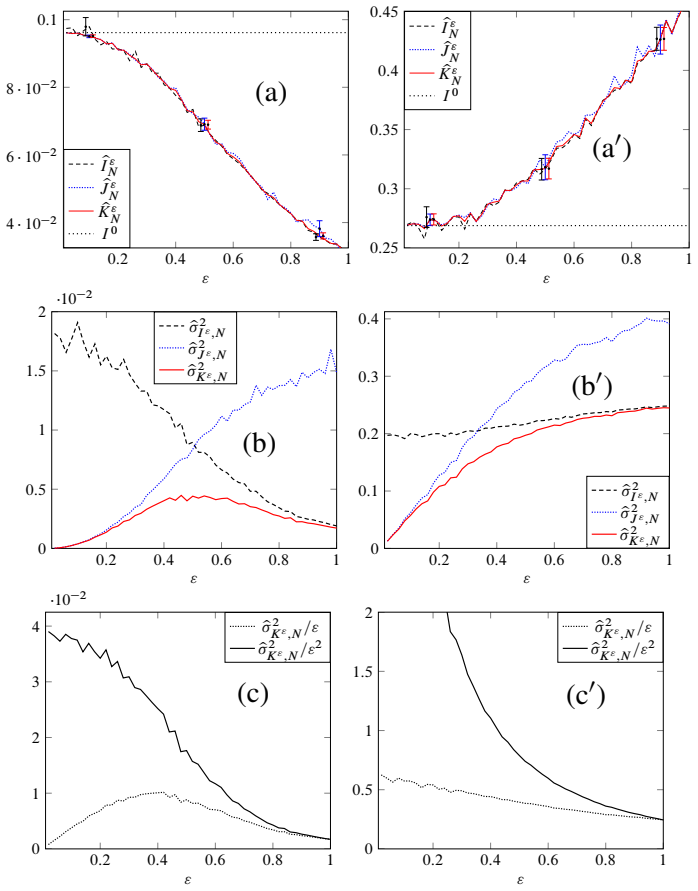

FIGURE 2.2. Example 4.3 of a TDOF system modeling a two-story building type and nonlinear damper of the linear-plus-quadratic type, driven by an Ornstein-Uhlenbeck noise. In the left column (subfigures a–c), resp., in the right column (subfigures a′–c′), the aim is to estimate $I^\varepsilon = \mathbb{E}[(X^\varepsilon_{1,T})^2 + (X^\varepsilon_{3,T})^2]$ for $T = 1$, resp., $I^\varepsilon = \mathbb{P}(|X^\varepsilon_{1,T}| \le a, |X^\varepsilon_{3,T}| \le b)$ for $T = 1$, $a = 0.1$, $b = 0.1$. The expectation of the control variate $I^0 = \mathbb{E}[(X^0_{1,T})^2 + (X^0_{3,T})^2]$, resp., $I^0 = \mathbb{P}(|X^0_{1,T}| \le a, |X^0_{3,T}| \le b)$, is obtained by a massive Monte Carlo computation with coarse time step. The numerical procedure is described in Section 4 (Euler-Maruyama time discretization with time step $\delta t = 10^{-5}$). The number of Monte Carlo samples is $N = 10^4$ and $m_1 = m_3 = c_1 = k_3 = 1$.

The process η^ε_1 is a white-noise driven linear oscillator with stiffness $\mu > 0$ and damping $\gamma > 0$. It can be encountered in earthquake engineering because it is considered to be a realistic type of random forcing to represent seismic excitation (it is the so-called Kanai-Tajimi model [23]).

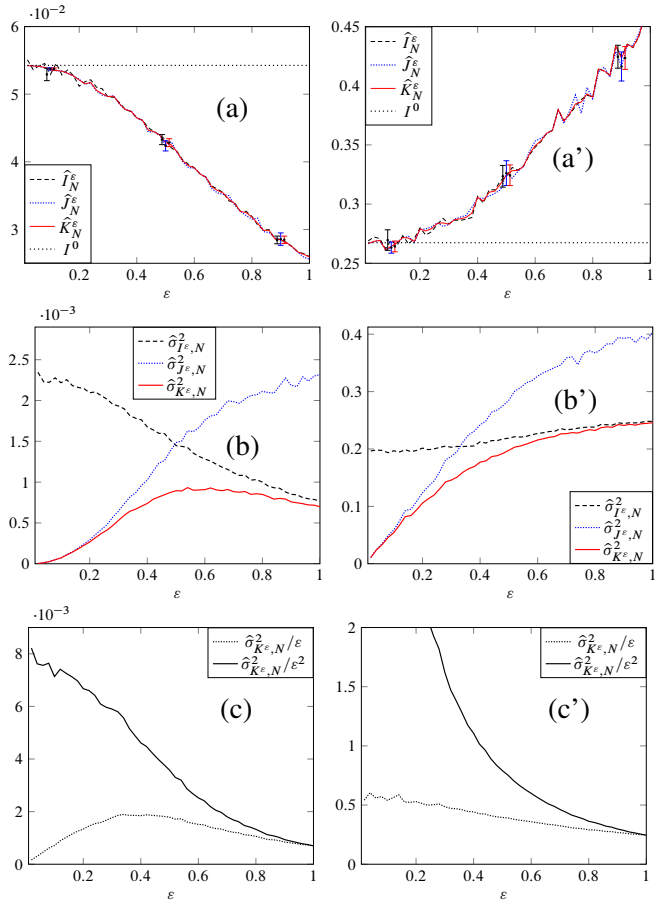

FIGURE 2.3. Example 7.3 of a TDOF system modeling a two-story building with an hysteretic spring of elasto-plastic type, driven by an Ornstein-Uhlenbeck noise. The quantities presented here are similar to those presented in Figure 2.2. The details of the elements ① and ③ with nonlinear and hysteretic behaviors can be found in Example 7.3.

EXAMPLE 3.3. *If $\widetilde{\eta}^\varepsilon$ is a real-valued zero-mean stationary Gaussian process with power spectral density*

$$\mathrm{PSD}^\varepsilon(\omega) = \varepsilon^2\,\mathrm{PSD}(\varepsilon^2\omega), \quad \mathrm{PSD}(\omega) = \sum_{k=1}^{q} \frac{\sigma_k^2}{1+\omega^2/\Delta\Omega_k^2},$$

then it has the same distribution as the process $\sum_{k=1}^q \sigma_k \eta_k^\varepsilon$ where $\boldsymbol{\eta}^\varepsilon$ is solution of (2.2) *with $d = d' = q$ and*

$$\mathbf{A} = \mathbf{K} = \mathrm{diag}(\Delta\Omega_k, k = 1, \ldots, q).$$

This shows that any zero-mean stationary Gaussian process with power spectral density that can be decomposed as a sum of centered Lorentzians belongs to the model (2.2).

EXAMPLE 3.4. *If $\widetilde{\eta}^\varepsilon$ is a real-valued zero-mean stationary Gaussian process with power spectral density*

$$\mathrm{PSD}^\varepsilon(\omega) = \varepsilon^2 \mathrm{PSD}(\varepsilon^2\omega),$$
$$\mathrm{PSD}(\omega) = \frac{1}{2}\sum_{k=1}^{q} \frac{\sigma_k^2}{1+(\omega-\omega_k)^2/\Delta\Omega_k^2} + \frac{\sigma_k^2}{1+(\omega+\omega_k)^2/\Delta\Omega_k^2},$$

then it has the same distribution as the process $\sum_{k=1}^q \sigma_k \eta^\varepsilon_{2k-1}$ where $\boldsymbol{\eta}^\varepsilon$ is solution of (2.2) *with $d = d' = 2q$ and*

$$\mathbf{A} = \oplus_{k=1}^q \begin{pmatrix} \Delta\Omega_k & -\omega_k \\ \omega_k & \Delta\Omega_k \end{pmatrix}, \quad \mathbf{K} = \oplus_{k=1}^q \begin{pmatrix} \Delta\Omega_k & 0 \\ 0 & \Delta\Omega_k \end{pmatrix}.$$

This shows that any zero-mean stationary Gaussian process with power spectral density that can be decomposed as a sum of noncentered Lorentzian functions belongs to the model (2.2).

We also consider the limiting $\mathbb{R}^n$-valued process $\boldsymbol{X}^0$ solution of the SDE (2.5). The continuous process $(\boldsymbol{X}^\varepsilon - \boldsymbol{X}^0)$ converges in probability to 0 as $\varepsilon \to 0$ as stated in the following proposition.

PROPOSITION 3.5. *If $\boldsymbol{X}_0^\varepsilon = \boldsymbol{X}_0^0$, then the continuous process $(\boldsymbol{X}^\varepsilon - \boldsymbol{X}^0)$ converges in probability to* 0 *as $\varepsilon \to 0$. The convergence holds in the space of continuous functions equipped with the topology associated to the uniform norm over compact intervals.*

The proof of Proposition 3.5 is based on the perturbed test function method as described first in [21, chap. 7] or in [8, chap. 6]. It is given in Appendix A.

EXAMPLE 3.6. *We consider the process $\boldsymbol{X}^\varepsilon$ solution of the ODE* (2.1) *where $\boldsymbol{\eta}^\varepsilon$ is the rapidly varying mean-reverting process* (3.1). *We also consider the limiting process*

$$\mathrm{d}\boldsymbol{X}^0 = \boldsymbol{b}(\boldsymbol{X}^0)\mathrm{d}t + \frac{K}{A}\boldsymbol{\sigma}(\boldsymbol{X}^0)\mathrm{d}W_t + \frac{K^2}{2A^2}(\boldsymbol{\sigma}(\boldsymbol{X}^0)\cdot\nabla_{\boldsymbol{x}^0})\boldsymbol{\sigma}(\boldsymbol{X}^0)\mathrm{d}t,$$

driven by the same Brownian motion. The continuous process $(\boldsymbol{X}^\varepsilon - \boldsymbol{X}^0)$ converges in probability to 0 *as $\varepsilon \to 0$.*

EXAMPLE 3.7. *We consider the process $\boldsymbol{X}^\varepsilon$ solution of the ODE* (2.1) *where $\boldsymbol{\eta}^\varepsilon$ is the rapidly varying mean-reverting process* (3.2)–(3.3). *We also consider the limiting process*

$$\mathrm{d}\boldsymbol{X}^0 = \boldsymbol{b}(\boldsymbol{X}^0)\mathrm{d}t + \frac{K}{\mu}\boldsymbol{\sigma}(\boldsymbol{X}^0)\mathrm{d}W_t + \frac{K^2}{2\mu^2}(\boldsymbol{\sigma}(\boldsymbol{X}^0)\cdot\nabla_{\boldsymbol{x}^0})\boldsymbol{\sigma}(\boldsymbol{X}^0)\mathrm{d}t,$$

driven by the same Brownian motion. The continuous process $(\boldsymbol{X}^\varepsilon - \boldsymbol{X}^0)$ *converges in probability to* 0 *as* $\varepsilon \to 0$.

The proof that the optimal control variate estimator $\widehat{K}_N^\varepsilon$ and the control variate estimator $\widehat{J}_N^\varepsilon$ have asymptotic variances of the order of ε^2 as stated in Proposition 2.1 follows from the following lemma.

LEMMA 3.8. *Let* f, g *be smooth functions from* $\mathbb{R}^n$ *to* $\mathbb{R}$ *with bounded derivatives. Let* $T > 0$. *There exists* $C > 0$ *such that, for any* $t \in [\varepsilon, T]$,

$$\begin{aligned} \left|\mathbb{E}\left[g\left(\boldsymbol{X}_t^0\right)\left(f\left(\boldsymbol{X}_t^\varepsilon\right) - f\left(\boldsymbol{X}_t^0\right)\right)\right]\right| &\le C\varepsilon^2, \\ \mathbb{E}\left[\left(f\left(\boldsymbol{X}_t^\varepsilon\right) - f\left(\boldsymbol{X}_t^0\right)\right)^2\right] &\le C\varepsilon^2. \end{aligned} \tag{3.4}$$

The important hypothesis is that f should be smooth. We could certainly relax the hypothesis on the bounded derivatives by using uniform estimates of high-order moments of the process $\boldsymbol{X}^\varepsilon$. Lemma 3.8 is proved in Appendix B.

4 Numerical Simulations

In this section, we investigate our control variate method and report numerical results on different types of dynamical systems driven by colored noises. The two examples are smooth oscillators that can be described by equation (2.1) (one being linear with time-dependent coefficients and the other being of Van der Pol type). Other examples with nonsmooth dynamical systems will be addressed in Section 7.

We use the Euler-Maruyama approximation method to compute the approximate numerical solution of an SDE [20]. In Section 4.1, we recall the two types of colored noise that we consider and provide their time discretization. In Section 4.2, some details and discretization of the dynamical systems under consideration are given. In Section 4.3, numerical experiments are carried out in each case.

4.1 Colored noise models and their discretization

The two models of noise are shown in equation (3.1) (OU) and in the system of equations (3.2)–(3.3) (Langevin). The OU noise has two parameters, $A, K_{\mathrm{ou}} > 0$, whereas the Langevin noise has three parameters, $\mu, \gamma, K_{\mathrm{lan}} > 0$. Their discretization works as follows. Let $T > 0$ and $N_T \in \mathbb{N}$ be the number of time steps such that $T = N_T \delta t$. Let N be the number of Monte Carlo samples. Consider a sequence of independent and identically distributed standard Gaussian variables

$$\left\{\Delta W_n^k \sim \mathcal{N}(0,1),\ 0 \le n \le N_T - 1,\ 1 \le k \le N\right\}.$$

Let $\varepsilon > 0$. For each $1 \le k \le N$, we overload the notation by denoting the discretized noise in both cases by $\{\widehat{\eta}_n^{\varepsilon,k},\ 0 \le n \le N_T\}$.

- **Ornstein-Uhlenbeck noise:** $\widehat{\eta}_0^{\varepsilon,k} \sim \mathcal{N}\left(0, \dfrac{K_{\mathrm{ou}}^2}{2A}\right)$ and for $0 \le n \le N_T - 1$,

$$\widehat{\eta}_{n+1}^{\varepsilon,k} = \widehat{\eta}_n^{\varepsilon,k}\left(1 - \delta t \frac{A}{\varepsilon^2}\right) + \sqrt{\delta t}\frac{K_{\mathrm{ou}}}{\varepsilon}\Delta W_n^k.$$

- **Langevin noise:** $\widehat{\eta}_0^{\varepsilon,k}$ and $\widehat{\eta}_{2,0}^{\varepsilon,k}$ are independent variables with

$$\widehat{\eta}_0^{\varepsilon,k} \sim \mathcal{N}\Big(0, \frac{K_{\text{lan}}^2}{2\gamma}\Big), \quad \widehat{\eta}_{2,0}^{\varepsilon,k} \sim \mathcal{N}\Big(0, \frac{K_{\text{lan}}^2}{2\gamma\mu}\Big),$$

and for $0 \le n \le N_T - 1$,

$$\widehat{\eta}_{n+1}^{\varepsilon,k} = \widehat{\eta}_n^{\varepsilon,k} + \frac{\delta t}{\varepsilon^2}\widehat{\eta}_{2,n}^{\varepsilon,k}, \quad \widehat{\eta}_{2,n+1}^{\varepsilon,k} = \widehat{\eta}_{2,n}^{\varepsilon,k} - \frac{\delta t}{\varepsilon^2}\big[\mu\widehat{\eta}_n^{\varepsilon,k} + \gamma\widehat{\eta}_{2,n}^{\varepsilon,k}\big] + \sqrt{\delta t}\,\frac{K_{\text{lan}}}{\varepsilon}\Delta W_n^k.$$

4.2 Details and discretization of the illustrative dynamical systems

We consider systems of the form (2.1). We first consider the case of smooth systems that can have time-dependent coefficients:

$$\frac{\mathrm{d}X_1^\varepsilon}{\mathrm{d}t} = X_2^\varepsilon, \quad \frac{\mathrm{d}X_2^\varepsilon}{\mathrm{d}t} = -h(X_1^\varepsilon, X_2^\varepsilon, t) + \frac{1}{\varepsilon}\eta^\varepsilon. \tag{4.1}$$

Here we are interested in $\mathbb{E}[\|\boldsymbol{X}_T^\varepsilon\|^2]$ and in $\mathbb{P}(|X_{1,T}^\varepsilon| \le 1)$ for $T = 1$. Note that the second case corresponds to an expectation $\mathbb{P}(|X_{1,T}^\varepsilon| \le 1) = \mathbb{E}[f(\boldsymbol{X}_T^\varepsilon)]$ with a nonsmooth function $f(\boldsymbol{x}) = \mathbf{1}_{|x_1|\le 1}$. As $\varepsilon \to 0$, $\boldsymbol{X}^\varepsilon = (X_1^\varepsilon, X_2^\varepsilon)$ converges to $\boldsymbol{X}^0 = (X_1^0, X_2^0)$ where

$$\mathrm{d}X_1^0 = X_2^0\,\mathrm{d}t, \quad \mathrm{d}X_2^0 = -h\big(X_1^0, X_2^0, t\big)\mathrm{d}t + C\,\mathrm{d}W, \tag{4.2}$$

$C = K_{\text{ou}}A^{-1}$ for an OU noise, and $C = K_{\text{lan}}\mu^{-1}$ for a Langevin noise. For the stochastic simulation of (4.1) and (4.2), we proceed as follows:

- $\widehat{X}_{1,0}^{\varepsilon,k} = x_{1,0}$, $\widehat{X}_{2,0}^{\varepsilon,k} = x_{2,0}$, and for $0 \le n \le N_T - 1$,

$$\begin{cases} \widehat{X}_{1,n+1}^{\varepsilon,k} = \widehat{X}_{1,n}^{\varepsilon,k} + \delta t\,\widehat{X}_{2,n}^{\varepsilon,k}, \\ \widehat{X}_{2,n+1}^{\varepsilon,k} = \widehat{X}_{2,n}^{\varepsilon,k} - \delta t\,h(\widehat{X}_{1,n}^{\varepsilon,k}, \widehat{X}_{2,n}^{\varepsilon,k}, n\delta t) + \frac{\delta t}{\varepsilon}\widehat{\eta}_n^{\varepsilon,k}. \end{cases}$$

- $\widehat{X}_{1,0}^{0,k} = x_{1,0}$, $\widehat{X}_{2,0}^{0,k} = x_{2,0}$, and for $0 \le n \le N_T - 1$,

$$\begin{cases} \widehat{X}_{1,n+1}^{0,k} = \widehat{X}_{1,n}^{0,k} + \delta t\,\widehat{X}_{2,n}^{0,k}, \\ \widehat{X}_{2,n+1}^{0,k} = \widehat{X}_{2,n}^{0,k} - \delta t\,h(\widehat{X}_{1,n}^{0,k}, \widehat{X}_{2,n}^{0,k}, n\delta t) + C\sqrt{\delta t}\,\Delta W_n^k. \end{cases}$$

$\widehat{\boldsymbol{X}}_n^{\varepsilon,k}$ and $\widehat{\boldsymbol{X}}_n^{0,k}$ are independent (in k) copies that are meant to approximate $\boldsymbol{X}_{n\delta t}^\varepsilon$ and $\boldsymbol{X}_{n\delta t}^0$.

EXAMPLE 4.1 (LINEAR OSCILLATOR WITH TIME-DEPENDENT COEFFICIENTS). *We take* $h(x_1, x_2, t) \triangleq p(t)x_1 + q(t)x_2$ *where* $p(t) \triangleq 1 + \cos(t)$ *and* $q(t) \triangleq 1 + \sin(t)$ (*the choice is purely arbitrary*). *Here, in both the OU and Langevin cases, the limiting process* $\boldsymbol{X}^0 = (X_1^0, X_2^0)$ *is a Gaussian process provided that the initial condition is deterministic or Gaussian. This is useful to derive the expectation of the control variate. The distribution of* $\boldsymbol{X}_t^0 = (X_{1,t}^0, X_{2,t}^0)$ *is characterized by its first-order moment* $\boldsymbol{m}(t) \triangleq \mathbb{E}[\boldsymbol{X}_t^0] \in \mathbb{R}^2$ *and second-order moment*

$$\mathbf{M}(t) \triangleq (\mathbb{E}[X_{i,t}^0 X_{j,t}^0])_{i,j=1}^2 \in \mathcal{M}_{2,2}(\mathbb{R}),$$

which satisfy the following systems of differential equations:

- FIRST-ORDER MOMENT
  - $(m_1(0), m_2(0)) = (x_0, \dot{x}_0)$,
  - $\dot{m}_1(t) = m_2(t)$,
  - $\dot{m}_2(t) = -p(t)m_1(t) - q(t)m_2(t)$.
- SECOND-ORDER MOMENT

$$\begin{cases} (M_{11}(0), M_{22}(0), M_{12}(0)) = (x_0^2, \dot{x}_0^2, x_0\dot{x}_0), \\ \dot{M}_{11}(t) = 2M_{12}(t), \\ \dot{M}_{22}(t) = -2p(t)M_{12}(t) - 2q(t)M_{22}(t) + C^2, \\ \dot{M}_{12}(t) = M_{22}(t) - p(t)M_{11}(t) - q(t)M_{12}(t). \end{cases} \tag{4.3}$$

The expectation of the control variate $\mathbb{E}[\|\boldsymbol{X}_T^0\|^2]$ *with* $T = 1$ *is estimated by solving numerically, with an Euler method, the differential equations for the first- and second-order moments.*

EXAMPLE 4.2 (VAN DER POL OSCILLATOR). *We take*

$$h(x_1, x_2) = x_1 - \nu(1 - x_1^2)x_2 \quad \textit{where } \nu > 0.$$

The expectation of the control variate can be represented by $\mathbb{E}[\|\boldsymbol{X}_T^0\|^2] = \mathfrak{c}(\boldsymbol{x}_0, 0)$ *with* $T = 1$, *where* $\mathfrak{c}$ *satisfies the following backward-in-time PDE*

$$\begin{cases} \partial_t \mathfrak{c} + \frac{C^2}{2}\partial_{x_2}^2 \mathfrak{c} - h(x_1, x_2)\partial_{x_2}\mathfrak{c} + x_2\partial_{x_1}\mathfrak{c} = 0, \quad \text{in } \mathbb{R}^2 \times [0, 1), \\ \mathfrak{c}(\boldsymbol{x}, 1) = \|\boldsymbol{x}\|^2 \text{ in } \mathbb{R}^2. \end{cases} \tag{4.4}$$

The expectation $\mathbb{E}[\|\boldsymbol{X}_T^0\|^2]$ *is estimated by solving the PDE* (4.4) *with a finite difference method.*

EXAMPLE 4.3 (NONLINEAR TDOF). *A TDOF modeling a nonlinear spring of the linear-plus-quadratic cubic type and a nonlinear damper of the linear-plus-quadratic type can be seen as a coupling between two systems of the form* (4.1)

$$\begin{cases} \dfrac{\mathrm{d}X_1^\varepsilon}{\mathrm{d}t} = X_2^\varepsilon, \quad \dfrac{\mathrm{d}X_2^\varepsilon}{\mathrm{d}t} = -g_2\big(X_1^\varepsilon, X_2^\varepsilon, X_3^\varepsilon, X_4^\varepsilon\big) + \dfrac{1}{\varepsilon}\eta^\varepsilon, \end{cases} \tag{4.5}$$

In addition to $\mathbb{E}[(X_{1,T}^\varepsilon)^2 + (X_{3,T}^\varepsilon)^2]$ *for* $T = 1$, *we are interested in* $\mathbb{P}(|X_{1,T}^\varepsilon| \le a, |X_{3,T}^\varepsilon| \le b)$. *As* $\varepsilon \to 0$, $\boldsymbol{X}^\varepsilon \to \boldsymbol{X}^0$ *where*

$$\begin{cases} \mathrm{d}X_1^0 = X_2^0\mathrm{d}t, \quad \mathrm{d}X_2^0 = -g_2\big(X_1^0, X_2^0, X_3^0, X_4^0\big)\mathrm{d}t + C\,\mathrm{d}W, \\ \mathrm{d}X_3^0 = X_4^0\mathrm{d}t, \quad \mathrm{d}X_4^0 = -g_4\big(X_1^0, X_2^0, X_3^0, X_4^0\big)\mathrm{d}t. \end{cases} \tag{4.6}$$

Here

$$\begin{aligned} g_2(x_1, x_2, x_3, x_4) \triangleq\ & k_1 x_1(1 + \min(x_1^2, \widetilde{L})) + c_1 x_2 - k_3(x_3 - x_1) \\ & - c_3(x_4 - x_2)(1 + \min(|x_4 - x_2|, \widetilde{L})) \end{aligned}$$

and

$$g_4(x_1, x_2, x_3, x_4) \triangleq k_3(x_3 - x_1) + c_3(x_4 - x_2)(1 + \min(|x_4 - x_2|, \widetilde{L})).$$

In the original model of Spanos $\widetilde{L} = \infty$*, see* [34, pp. 189–190]. *For any positive finite value of* $\widetilde{L}$*, the system above enters into the scope of our results. The simulation of* (4.5) *and* (4.6) *is similar to what is done for* (4.1) *and* (4.2). *We take* $\widetilde{L} = 1000$*,* $c_1 = c_3 = k_1 = k_3 = 1$.

4.3 Numerical experiments

We report our numerical results for the two systems mentioned above. In each of the two figures below, Figures 4.1 and 4.2, there are four subfigures (a)–(d). For subfigures (a) and (b), the driving force is an Ornstein-Uhlenbeck noise (3.1) with $A = K = 1$. In subfigure (a), the dashed black, dotted blue, and solid red lines represent the standard MC estimator $\widehat{I}_N^\varepsilon$ and the control variate estimators $\widehat{J}_N^\varepsilon$ and $\widehat{K}_N^\varepsilon$, respectively. For each value of $\varepsilon \in \{0.1, 0.5, 0.9\}$, error bars (95% confidence interval) are shown for each of the estimators (in $\widehat{I}_N^\varepsilon$, $\widehat{J}_N^\varepsilon$, $\widehat{K}_N^\varepsilon$ order from the left to the right). The black dotted line represents the expectation of the control variate. The objective of the subfigure (b) is to illustrate the bound (2.19) and to show that the ε^2-behavior is actually sharp. The same description applies to (c) and (d), except they correspond to the case of a Langevin noise (3.2)–(3.3) with $\mu = \gamma = K = 1$. In the figures the asymptotic variance of the standard MC estimator $\widehat{I}_N^\varepsilon$ is estimated by

$$\widehat{\sigma}^2_{I^\varepsilon,N} = \frac{1}{N}\sum_{k=1}^{N}\left(F(\boldsymbol{X}^\varepsilon(\boldsymbol{W}^k))\right)^2 - \left(\widehat{I}_N^\varepsilon\right)^2, \tag{4.7}$$

the asymptotic variance of the control variate estimator $\widehat{J}_N^\varepsilon$ is estimated by

$$\widehat{\sigma}^2_{J^\varepsilon,N} = \frac{1}{N}\sum_{k=1}^{N}\left(F(\boldsymbol{X}^\varepsilon(\boldsymbol{W}^k)) - F(\boldsymbol{X}^0(\boldsymbol{W}^k)) + I^0\right)^2 - \left(\widehat{J}_N^\varepsilon\right)^2, \tag{4.8}$$

and the asymptotic variance of the optimal control variate estimator $\widehat{K}_N^\varepsilon$ is estimated by

$$\widehat{\sigma}^2_{K^\varepsilon,N} = \frac{1}{N}\sum_{k=1}^{N}\left(F(\boldsymbol{X}^\varepsilon(\boldsymbol{W}^k)) - \widehat{\rho}_N^\varepsilon F(\boldsymbol{X}^0(\boldsymbol{W}^k)) + \widehat{\rho}_N^\varepsilon I^0\right)^2 - \left(\widehat{K}_N^\varepsilon\right)^2, \tag{4.9}$$

with $\widehat{\rho}_N^\varepsilon$ defined by (2.26). $\widehat{\sigma}^2_{I^\varepsilon,N}$, $\widehat{\sigma}^2_{J^\varepsilon,N}$, and $\widehat{\sigma}^2_{K^\varepsilon,N}$ are consistent estimators of $\sigma^2_{I^\varepsilon}$, $\sigma^2_{J^\varepsilon}$, and $\sigma^2_{K^\varepsilon}$, respectively.

We use $N = 10^4$ samples with a time step of $\delta t = 10^{-5}$ (note that $\delta t/\varepsilon^2 = 0.1$ for the smallest $\varepsilon = 10^{-2}$ used in the numerical experiments). We report the numerical results for the linear oscillator with time-dependent coefficients in Figure 4.1 and for the Van der Pol oscillator in Figure 4.2. The numerical results concern

the estimation of $I^\varepsilon = \mathbb{E}[\|\boldsymbol{X}_T^\varepsilon\|^2]$ or $I^\varepsilon = \mathbb{P}(|X_{1,T}^\varepsilon| \leq 1)$ with $T = 1$ where $\boldsymbol{X}^\varepsilon$ satisfies (4.1) and thus the expectation of the control variate is $I^0 = \mathbb{E}[\|\boldsymbol{X}_T^0\|^2]$ or $\mathbb{P}(|X_{1,T}^0| \leq 1)$ where $\boldsymbol{X}^0$ satisfies (4.2).

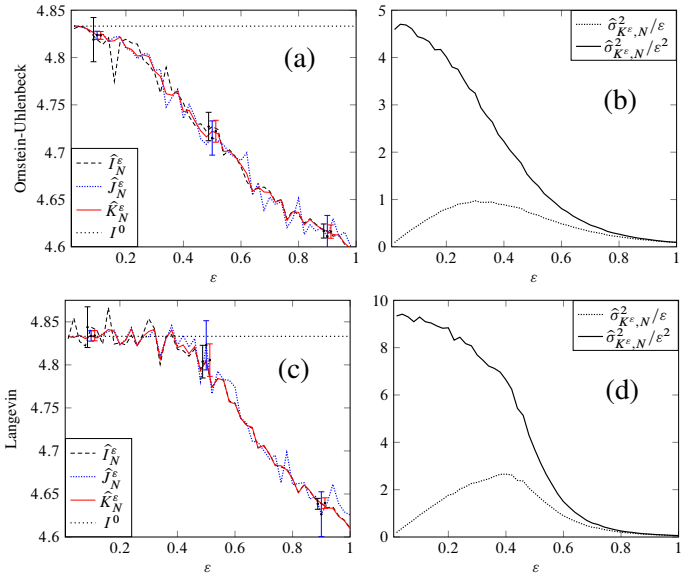

FIGURE 4.1. Example 4.1 (linear oscillator with time-dependent coefficients) with $h(x_1, x_2, t) \triangleq p(t)x_1 + q(t)x_2$, $p(t) \triangleq 1 + \cos(t)$ and $q(t) \triangleq 1 + \sin(t)$. The aim is to estimate $I^\varepsilon = \mathbb{E}[\|\boldsymbol{X}_T^\varepsilon\|^2]$ for $T = 1$ and the expectation of the control variate $I^0 = \mathbb{E}[\|\boldsymbol{X}_T^0\|^2]$ is obtained by solving the set of differential equations (4.3). In subfigures (a) and (b), resp., in subfigures (c) and (d), the driving process is an OU, resp., a Langevin, process.

The theoretical predictions provided by equation (2.19) and Proposition 2.1 are based on the condition that f has bounded derivatives. As we have discussed above, the assumption that f is smooth is important but the hypothesis on the boundedness of the derivatives can certainly be relaxed. The numerical results shown in Figures 4.1 and 4.2 are actually in good agreement with the theoretical predictions: the asymptotic variances $\sigma_{J^\varepsilon}^2$ and $\sigma_{K^\varepsilon}^2$ behave as $O(\varepsilon^2)$. The only cases where the behavior is $O(\varepsilon)$, and not $O(\varepsilon^2)$, are when the quantity of interest is of the form $\mathbb{E}[f(\boldsymbol{X}_T^\varepsilon)]$ with a function f that is not smooth, which is not surprising.

5 Diffusion Approximation for the Multivalued Case

In this section we consider multivalued ODEs of the form (2.30) or (2.31).

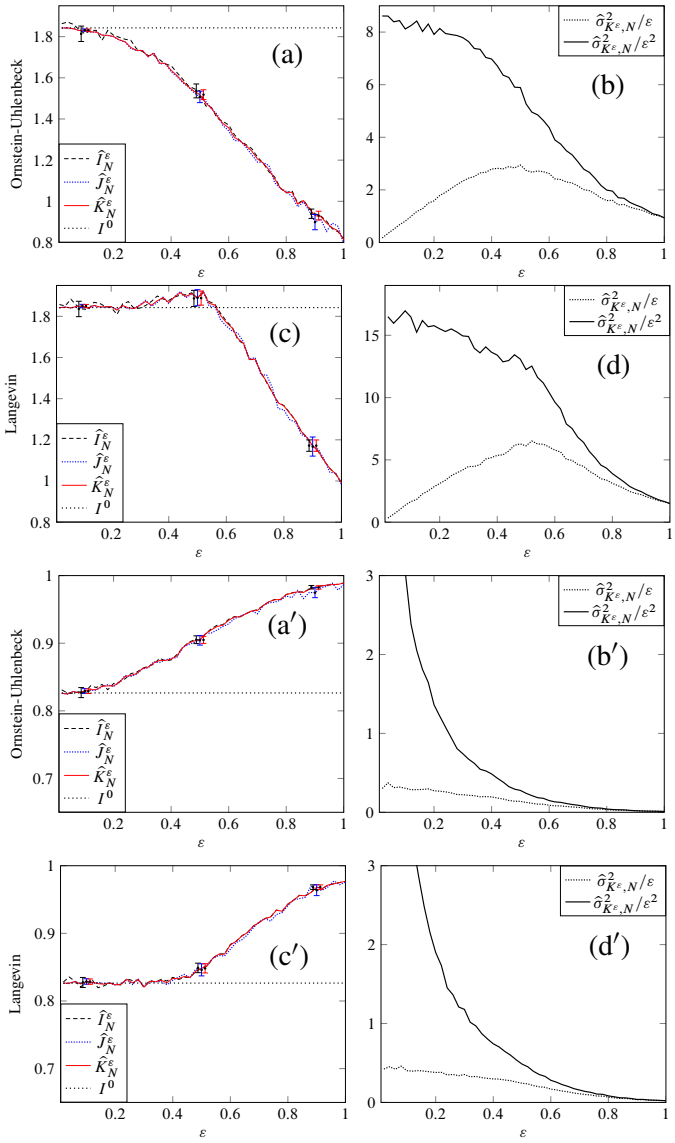

FIGURE 4.2. Example 4.2 (Van der Pol oscillator) with $h(x_1, x_2) \triangleq x_1 - (1 - x_1^2)x_2$. In subfigures (a)–(d), resp., in subfigures (a′)–(d′), the aim is to estimate $I^{\varepsilon} = \mathbb{E}[\|\boldsymbol{X}^{\varepsilon}_T\|^2]$, resp., $I^{\varepsilon} = \mathbb{P}(|X^{\varepsilon}_{1,T}| \leq 1)$, for $T = 1$. The expectation of the control variate $I^0 = \mathbb{E}[\|\boldsymbol{X}^0_T\|^2]$, resp., $I^0 = \mathbb{P}(|X^0_{1,T}| \leq 1)$, is obtained by solving the PDE (4.4) with the suitable final condition.

5.1 Basic properties on differential inclusions

We recall that the subdifferential of a convex function $F : \mathbb{R}^r \to (-\infty, \infty]$ such that $\mathrm{Dom}(F) \triangleq \{\boldsymbol{x} \in \mathbb{R}^r,\ F(\boldsymbol{x}) < \infty\}$ is not empty, is the map from $\mathbb{R}^r$ to $\mathcal{P}(\mathbb{R}^r)$ (the set of subsets of $\mathbb{R}^r$) defined by

$$\partial F(\boldsymbol{x}) \triangleq \{\boldsymbol{\xi} \in \mathbb{R}^r,\ \forall \boldsymbol{z} \in \mathbb{R}^r,\ \boldsymbol{\xi} \cdot (\boldsymbol{z} - \boldsymbol{x}) + F(\boldsymbol{x}) \leq F(\boldsymbol{z})\}$$

for $\boldsymbol{x} \in \mathrm{Dom}(F)$ and $\partial F(\boldsymbol{x}) = \varnothing$ for $\boldsymbol{x} \notin \mathrm{Dom}(F)$. To grasp quickly the idea when $r = 1$, $\partial F(x)$ can be seen as the set of subslopes of F at the point x and when F is differentiable at the point x, $\partial F(x) = \{F'(x)\}$. See [4] for more details.

One way to construct a solution to a multivalued ODE of the form (2.30) or (2.31) is to proceed by penalization. The inclusion is replaced by an equality involving the Moreau-Yosida regularization of $F : \mathbb{R}^r \to (-\infty, +\infty]$ (with $F = \varphi$, $r = n$ or $F = \psi$, $r = m$ in our cases), that is,

$$\forall p \geq 1,\ \forall \boldsymbol{x} \in \mathbb{R}^r,\ F_p(\boldsymbol{x}) \triangleq \inf_{\boldsymbol{z} \in \mathbb{R}^r} \left\{ F(\boldsymbol{z}) + \frac{p}{2} \|\boldsymbol{x} - \boldsymbol{z}\|^2 \right\}. \tag{5.1}$$

We recall from annex B in [30] some properties of F_p:

(1) $F_p : \mathbb{R}^r \mapsto \mathbb{R}$ is a convex differentiable function.

(2) $\forall \boldsymbol{x} \in \mathbb{R}^r$, $\partial F_p(\boldsymbol{x}) = \{\nabla F_p(\boldsymbol{x})\}$, and $\nabla F_p(\boldsymbol{x}) \in \partial F(J_p \boldsymbol{x})$ where $J_p \boldsymbol{x} \triangleq \boldsymbol{x} - \frac{1}{p} \nabla F_p(\boldsymbol{x})$.

(3) $\exists C > 0,\ \forall \boldsymbol{x} \in \mathbb{R}^r,\ \forall p,\ \|J_p \boldsymbol{x}\| \leq \|\boldsymbol{x}\| + C$.

(4) $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^r, \big(\nabla F_p(\boldsymbol{x}) - \nabla F_p(\boldsymbol{y})\big) \cdot (\boldsymbol{x} - \boldsymbol{y}) \geq 0$.

(5) $\forall \boldsymbol{x} \in \mathbb{R}^r$,

$$\boldsymbol{x} \cdot \nabla F_p(\boldsymbol{x}) \geq 0. \tag{5.2}$$

(6) $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^r$,

$$(\nabla F_p(\boldsymbol{x}) - \nabla F_{p'}(\boldsymbol{y})) \cdot (\boldsymbol{x} - \boldsymbol{y}) \geq -\left(\frac{1}{p} + \frac{1}{p'}\right) \nabla F_p(\boldsymbol{x}) \cdot \nabla F_{p'}(\boldsymbol{y}). \tag{5.3}$$

(7) As a consequence of properties 2 and 3 above, we also have

$$\sup_{p \geq 1} \sup_{\boldsymbol{x} \in \mathbb{R}^r} \frac{\|\nabla F_p(\boldsymbol{x})\|}{p(1 + \|\boldsymbol{x}\|)} < \infty. \tag{5.4}$$

Thus, the penalized versions of (2.30) and (2.31) are

$$\frac{\mathrm{d}\boldsymbol{X}^{\varepsilon,p}}{\mathrm{d}t} + \nabla \varphi_p(\boldsymbol{X}^{\varepsilon,p}) = \boldsymbol{b}(\boldsymbol{X}^{\varepsilon,p}) + \frac{1}{\varepsilon} \boldsymbol{\sigma}\, \boldsymbol{\eta}^{\varepsilon}, \quad \boldsymbol{X}_0^{\varepsilon,p} = \boldsymbol{x}_0, \tag{5.5}$$

and

$$\begin{cases} \dfrac{\mathrm{d}\boldsymbol{X}^{\varepsilon,p}}{\mathrm{d}t} + \nabla \varphi_p(\boldsymbol{X}^{\varepsilon,p}) = \boldsymbol{b}^X(\boldsymbol{X}^{\varepsilon,p}, \boldsymbol{Z}^{\varepsilon,p}) + \dfrac{1}{\varepsilon} \boldsymbol{\sigma}\, \boldsymbol{\eta}^{\varepsilon}, & \boldsymbol{X}_0^{\varepsilon,p} = \boldsymbol{x}_0, \\ \dfrac{\mathrm{d}\boldsymbol{Z}^{\varepsilon,p}}{\mathrm{d}t} + \nabla \psi_p(\boldsymbol{Z}^{\varepsilon,p}) = \boldsymbol{b}^Z(\boldsymbol{X}^{\varepsilon,p}, \boldsymbol{Z}^{\varepsilon,p}), & \boldsymbol{Z}_0^{\varepsilon,p} = z_0. \end{cases} \tag{5.6}$$

It can be shown [4] that, if φ satisfies the condition

$$\sup_{p\geq 1}\sup_{\boldsymbol{x}\in\mathbb{R}^n}\|\nabla\varphi_p(\boldsymbol{x})\|<\infty, \tag{5.7}$$

where φ_p is the Yosida approximation (5.1) of φ, then the sequence of solutions of (5.5) $\{\boldsymbol{X}^{\varepsilon,p},\ p\geq 1\}$ is a Cauchy sequence in $\mathcal{C}([0,T];\mathbb{R}^n)$, the limit $\boldsymbol{X}^{\varepsilon}$ satisfies the differential inclusion (2.30), and its solution is unique.

A similar statement that uses the sequence of solutions of (5.6) $\{(\boldsymbol{X}^{\varepsilon,p},\boldsymbol{Z}^{\varepsilon,p}),\ p\geq 1\}$ in $\mathcal{C}([0,T];\mathbb{R}^n\times\mathbb{R}^m)$ holds for the existence and uniqueness of a solution for (2.31) when φ (but not necessarily ψ) satisfies the condition (5.7), while ψ satisfies the assumption

$$\sup_{p\geq 1}\psi_p(z_0)<\infty. \tag{5.8}$$

For the convenience of the reader we give the proofs of these results in Appendix C.

5.2 Diffusion approximation for equation (2.30)

We consider the $\mathbb{R}^n$-valued process $\boldsymbol{X}^{\varepsilon}$ solution of the multivalued ODE (2.30) when $\boldsymbol{\eta}^{\varepsilon}$ is given by (2.2). We assume that $\boldsymbol{b}$ is Lipschitz and that φ satisfies the condition (5.7). We also consider the limiting $\mathbb{R}^n$-valued process $\boldsymbol{X}^0$ solution of the multivalued SDE

$$\mathrm{d}\boldsymbol{X}^0+\partial\varphi(\boldsymbol{X}^0)\mathrm{d}t\ni\boldsymbol{b}(\boldsymbol{X}^0)\mathrm{d}t+\boldsymbol{\Gamma}\,\mathrm{d}\boldsymbol{W}_t, \tag{5.9}$$

driven by the same Brownian motion, with $\boldsymbol{\Gamma}=\boldsymbol{\sigma}\mathbf{A}^{-1}\mathbf{K}$. Existence and uniqueness of the solution of (5.9) is the same one as in Proposition C.1 and is discussed in Appendix D. The following proposition gives the convergence of the process $(\boldsymbol{X}^{\varepsilon}-\boldsymbol{X}^0)$ to 0. It is proved in [9, app. E].

PROPOSITION 5.1.

(1) *We have for all* $p\geq 1$

$$\sup_{\varepsilon}\mathbb{E}\Big[\sup_{t\leq T}\|\boldsymbol{X}_t^{\varepsilon,p}-\boldsymbol{X}_t^{\varepsilon}\|^2\Big]\leq\frac{C_T}{p}, \tag{5.10}$$

where $\boldsymbol{X}^{\varepsilon,p}$ *is the approximation* (5.5) *of* $\boldsymbol{X}^{\varepsilon}$.

(2) *The continuous process* $(\boldsymbol{X}^{\varepsilon}-\boldsymbol{X}^0)$ *converges in probability to* 0 *as* $\varepsilon\to 0$.

5.3 Diffusion approximation for equation (2.31)

We consider the $\mathbb{R}^n\times\mathbb{R}^m$-valued process $(\boldsymbol{X}^{\varepsilon},\boldsymbol{Z}^{\varepsilon})$ solution of the multivalued ODE (2.31) when $\boldsymbol{\eta}^{\varepsilon}$ is given by (2.2). We assume that $\boldsymbol{b}^X$ and $\boldsymbol{b}^Z$ are Lipschitz, that φ satisfies (5.7), and that ψ satisfies (5.8). We also consider the limiting $\mathbb{R}^n\times\mathbb{R}^m$-valued process $(\boldsymbol{X}^0,\boldsymbol{Z}^0)$ solution of the multivalued SDE

$$\begin{aligned}\mathrm{d}\boldsymbol{X}^0+\partial\varphi(\boldsymbol{X}^0)\mathrm{d}t&\ni\boldsymbol{b}^X(\boldsymbol{X}^0,\boldsymbol{Z}^0)\mathrm{d}t+\boldsymbol{\Gamma}\,\mathrm{d}\boldsymbol{W}_t,\\ \mathrm{d}\boldsymbol{Z}^0+\partial\psi(\boldsymbol{Z}^0)\mathrm{d}t&\ni\boldsymbol{b}^Z(\boldsymbol{X}^0,\boldsymbol{Z}^0)\mathrm{d}t,\end{aligned} \tag{5.11}$$

driven by the same Brownian motion, with $\mathbf{\Gamma} = \boldsymbol{\sigma}\mathbf{A}^{-1}\mathbf{K}$. Existence and uniqueness of the solution of (5.11) is the same as in Proposition C.2.

The convergence of the continuous process $(\boldsymbol{X}^\varepsilon - \boldsymbol{X}^0, \boldsymbol{Z}^\varepsilon - \boldsymbol{Z}^0)$ to 0 is stated in the next proposition. It is proved in [9, app. F].

PROPOSITION 5.2.

(1) *For all $p \geq 1$, we have*

$$\sup_\varepsilon \mathbb{E}\Big[\sup_{t\leq T}\big\{\big\|\boldsymbol{X}_t^{\varepsilon,p} - \boldsymbol{X}_t^\varepsilon\big\|^2 + \big\|\boldsymbol{Z}_t^{\varepsilon,p} - \boldsymbol{Z}_t^\varepsilon\big\|^2\big\}\Big] \leq \frac{C_T}{p}, \tag{5.12}$$

where $(\boldsymbol{X}^{\varepsilon,p}, \boldsymbol{Z}^{\varepsilon,p})$ is the approximation (5.6) *of $(\boldsymbol{X}^\varepsilon, \boldsymbol{Z}^\varepsilon)$.*

(2) *The continuous process $(\boldsymbol{X}^\varepsilon - \boldsymbol{X}^0, \boldsymbol{Z}^\varepsilon - \boldsymbol{Z}^0)$ converges in probability to* 0 *as $\varepsilon \to 0$.*

6 Control Variate Method in the Multivalued Case

We here consider the multivalued case. Let $\boldsymbol{X}^\varepsilon$ satisfy (2.30) or $(\boldsymbol{X}^\varepsilon, \boldsymbol{Z}^\varepsilon)$ satisfy (2.31). We want to estimate I^ε defined by (2.3) when $\boldsymbol{X}^\varepsilon$ satisfies (2.30) or $I^\varepsilon = \mathbb{E}[F(\boldsymbol{X}^\varepsilon, \boldsymbol{Z}^\varepsilon)]$ when $(\boldsymbol{X}^\varepsilon, \boldsymbol{Z}^\varepsilon)$ satisfies (2.31). The control variate method can be applied in this framework as in the ODE case addressed in Section 2. The control variate estimator $\widehat{J}_N^\varepsilon$ and the optimal control variate estimator $\widehat{K}_N^\varepsilon$ are defined by (2.15) and (2.25), respectively, for $\boldsymbol{X}^\varepsilon$ satisfying (2.30); they are asymptotically normal and their asymptotic variances are (2.18) and (2.28), respectively. For $(\boldsymbol{X}^\varepsilon, \boldsymbol{Z}^\varepsilon)$ satisfying (2.31) the control variate estimator

$$\widehat{J}_N^\varepsilon \triangleq \frac{1}{N}\sum_{k=1}^N F(\boldsymbol{X}^\varepsilon(\boldsymbol{W}^k), \boldsymbol{Z}^\varepsilon(\boldsymbol{W}^k)) - F(\boldsymbol{X}^0(\boldsymbol{W}^k), \boldsymbol{Z}^0(\boldsymbol{W}^k)) + I^0, \tag{6.1}$$

with $I^0 = \mathbb{E}[F(\boldsymbol{X}^0, \boldsymbol{Z}^0)]$, is asymptotically normal with an asymptotic variance given by

$$\sigma_{J^\varepsilon}^2 = \mathrm{Var}\big(F(\boldsymbol{X}^\varepsilon, \boldsymbol{Z}^\varepsilon) - F(\boldsymbol{X}^0, \boldsymbol{Z}^0)\big). \tag{6.2}$$

The practical optimal control variate estimator is

$$\begin{aligned}\widehat{K}_N^\varepsilon \triangleq \widehat{\rho}_N^\varepsilon I^0 &+ \frac{1}{N}\sum_{k=1}^N F(\boldsymbol{X}^\varepsilon(\boldsymbol{W}^k), \boldsymbol{Z}^\varepsilon(\boldsymbol{W}^k)) \\ &- \widehat{\rho}_N^\varepsilon F(\boldsymbol{X}^0(\boldsymbol{W}^k), \boldsymbol{Z}^0(\boldsymbol{W}^k)),\end{aligned} \tag{6.3}$$

where $\widehat{\rho}_N^\varepsilon$ is the empirical correlation

$$\widehat{\rho}_N^\varepsilon = \frac{\sum_{k=1}^N (F(\boldsymbol{X}^\varepsilon(\boldsymbol{W}^k), \boldsymbol{Z}^\varepsilon(\boldsymbol{W}^k)) - \widehat{I}_N^\varepsilon)(F(\boldsymbol{X}^0(\boldsymbol{W}^k), \boldsymbol{Z}^0(\boldsymbol{W}^k)) - \widehat{I}_N^0)}{\sum_{k=1}^N (F(\boldsymbol{X}^0(\boldsymbol{W}^k), \boldsymbol{Z}^0(\boldsymbol{W}^k)) - \widehat{I}_N^0)^2}, \tag{6.4}$$

$\widehat{I}_N^\varepsilon$ is the standard Monte Carlo estimator

$$\widehat{I}_N^\varepsilon \triangleq \frac{1}{N}\sum_{k=1}^{N} F(\boldsymbol{X}^\varepsilon(\boldsymbol{W}^k), \boldsymbol{Z}^\varepsilon(\boldsymbol{W}^k)), \tag{6.5}$$

and $\widehat{I}_N^0 = \frac{1}{N}\sum_{k=1}^{N} F(\boldsymbol{X}^0(\boldsymbol{W}^k), \boldsymbol{Z}^0(\boldsymbol{W}^k))$. The estimator $\widehat{K}_N^\varepsilon$ is asymptotically normal with an asymptotic variance given by

$$\sigma_{K^\varepsilon}^2 = \operatorname{Var}\big(F(\boldsymbol{X}^\varepsilon, \boldsymbol{Z}^\varepsilon) - \rho^\varepsilon F(\boldsymbol{X}^0, \boldsymbol{Z}^0)\big), \tag{6.6}$$

with $\rho^\varepsilon = \operatorname{Cov}(F(\boldsymbol{X}^\varepsilon, \boldsymbol{Z}^\varepsilon), F(\boldsymbol{X}^0, \boldsymbol{Z}^0))/\operatorname{Var}(F(\boldsymbol{X}^0, \boldsymbol{Z}^0))$. The asymptotic variance of the estimator $\widehat{I}_N^\varepsilon$ is

$$\sigma_{I^\varepsilon}^2 = \operatorname{Var}\big(F(\boldsymbol{X}^\varepsilon, \boldsymbol{Z}^\varepsilon)\big). \tag{6.7}$$

For small ε the asymptotic variance $\sigma_{I^\varepsilon}^2$ is approximately equal to $\operatorname{Var}(F(\boldsymbol{X}^0, \boldsymbol{Z}^0))$ and the asymptotic variances $\sigma_{J^\varepsilon}^2$ and $\sigma_{K^\varepsilon}^2$ are small, by Propositions 5.1 and 5.2. More quantitatively, if $F(\boldsymbol{X}, \boldsymbol{Z}) = f(\boldsymbol{X}_T, \boldsymbol{Z}_T)$ and f is a smooth function with bounded derivatives, then the asymptotic variances $\sigma_{J^\varepsilon}^2$ and $\sigma_{K^\varepsilon}^2$ are of order ε^2 when $\boldsymbol{X}^\varepsilon$ satisfies (2.30):

$$\sigma_{K^\varepsilon}^2 \le C\varepsilon^2, \quad 0 \le \sigma_{J^\varepsilon}^2 - \sigma_{K^\varepsilon}^2 \le C\varepsilon^4, \tag{6.8}$$

or of order ε when $(\boldsymbol{X}^\varepsilon, \boldsymbol{Z}^\varepsilon)$ satisfies (2.31):

$$\sigma_{K^\varepsilon}^2 \le C\varepsilon, \quad 0 \le \sigma_{J^\varepsilon}^2 - \sigma_{K^\varepsilon}^2 \le C\varepsilon^2. \tag{6.9}$$

Equations (6.8)–(6.9) are consequences of the following lemma, which is proved in [9, app. G].

LEMMA 6.1.

(1) *Let $\boldsymbol{X}^\varepsilon$ satisfy* (2.30). *If $f, g : \mathbb{R}^n \to \mathbb{R}$ are smooth functions with bounded derivatives and $T > 0$, then there exists $C > 0$ such that, for any $t \in [\varepsilon, T]$,*

$$\begin{aligned} \big|\mathbb{E}\big[g(\boldsymbol{Z}_t^0)\big(f(\boldsymbol{X}_t^\varepsilon) - f(\boldsymbol{X}_t^0)\big)\big]\big| &\le C\varepsilon^2, \\ \mathbb{E}\big[\big(f(\boldsymbol{X}_t^\varepsilon) - f(\boldsymbol{X}_t^0)\big)^2\big] &\le C\varepsilon^2. \end{aligned} \tag{6.10}$$

(2) *Let $(\boldsymbol{X}^\varepsilon, \boldsymbol{Z}^\varepsilon)$ satisfy* (2.31). *If $f, g : \mathbb{R}^{n+m} \to \mathbb{R}$ are smooth functions with bounded derivatives and $T > 0$, then there exists $C > 0$ such that, for any $t \in [\varepsilon, T]$,*

$$\begin{aligned} \big|\mathbb{E}\big[g(\boldsymbol{X}_t^0, \boldsymbol{Z}_t^0)\big(f(\boldsymbol{X}_t^\varepsilon, \boldsymbol{Z}_t^\varepsilon) - f(\boldsymbol{X}_t^0, \boldsymbol{Z}_t^0)\big)\big]\big| &\le C\varepsilon, \\ \mathbb{E}\big[\big(f(\boldsymbol{X}_t^\varepsilon, \boldsymbol{Z}_t^\varepsilon) - f(\boldsymbol{X}_t^0, \boldsymbol{Z}_t^0)\big)^2\big] &\le C\varepsilon. \end{aligned} \tag{6.11}$$

7 Numerical Simulations in the Multivalued Case

We present examples that are nonsmooth dynamical systems which are prevalent in engineering mechanics. Examples 7.1 and 7.2 are oscillators involving friction or/and elasto-plastic behaviors; Example 7.3 is a nonlinear and nonsmooth two-degrees-of-freedom (TDOF) system. All three of these examples can be described by equations (2.30) and (2.31). Examples 7.4 and 7.5, which do not fall within the scope of any aforementioned case, correspond to an obstacle problem and to the reflection of the integral of a colored noise, respectively.

7.1 Nonsmooth systems in the form of equations (2.30 and (2.31)

EXAMPLE 7.1 (FRICTION BEHAVIOR). *Consider equation* (2.30) *with* $\varphi(x) \triangleq c_{\rm f}|x|\ \forall x \in \mathbb{R}$, *where* $c_{\rm f} > 0$ *is a friction coefficient. The* $\mathbb{R}$*-valued process* X^ε *represents the velocity of a material point* (*stick-slip motion*) *subjected to friction and colored noise. See, for instance,* [28] *for an explanation of the physics behind the friction force, and* [3] *for the use of SDEs with multivalued drift for modeling. As* $\varepsilon \to 0$, $X^\varepsilon \to X^0$ *where* X^0 *satisfies equation* (5.9). *For the stochastic simulation, we proceed as follows*:

- $\widehat{X}_0^{\varepsilon,k} = x_0$ *and for* $0 \le n \le N_T - 1$,

$$\widehat{X}_{n+1}^{\varepsilon,k} = \widehat{X}_n^{\varepsilon,k} + \frac{\delta t}{\varepsilon}\widehat{\eta}_n^{\varepsilon,k} - \delta t\,\mathrm{proj}_{[-c_{\rm f},c_{\rm f}]}\left(\frac{\widehat{X}_n^{\varepsilon,k}}{\delta t} + \frac{1}{\varepsilon}\widehat{\eta}_n^{\varepsilon,k}\right),$$

  with $\widehat{\eta}_n^{\varepsilon,k}$ *described in Section* 4.1.
- $\widehat{X}_0^{0,k} = x_0$ *and for* $0 \le n \le N_T - 1$,

$$\widehat{X}_{n+1}^{0,k} = \widehat{X}_n^{0,k} + (\delta t)^{1/2} C\Delta W_n^k - \delta t\,\mathrm{proj}_{[-c_{\rm f},c_{\rm f}]}\left(\frac{\widehat{X}_n^{0,k}}{\delta t} + \frac{C}{\sqrt{\delta t}}\Delta W_n^k\right).$$

We are interested in $\mathbb{E}[\big(X_T^\varepsilon\big)^2]$ *for* $T = 1$. *The expectation of the control variate is* $\mathbb{E}[(X_T^0)^2]$. *The latter can be represented as* $\mathfrak{c}(x_0, 0)$ *where* $\mathfrak{c}$ *satisfies the following backward-in-time partial differential inclusion*

$$\begin{cases} \partial_t \mathfrak{c}(x,t) + \frac{C^2}{2}\partial_x^2 \mathfrak{c}(x,t) \in \partial\varphi(x)\partial_x \mathfrak{c}(x,t) & \text{for } (x,t) \in \mathbb{R}\times[0,1), \\ \mathfrak{c}(x,1) = x^2 & \text{for } x \in \mathbb{R}. \end{cases} \tag{7.1}$$

It can be estimated by solving this partial differential inclusion with a finite difference method. We proceed as follows. For every $t > 0$, *the function* $x \mapsto \mathfrak{c}(x,t)$ *is smooth and even, provided that the initial condition is smooth and even. Indeed, this comes from the probabilistic representation and the fact that, for any starting point* $x \in \mathbb{R}$, $\{X_t^x, t \ge 0\}$ *and* $\{X_t^{-x}, t \ge 0\}$ *have the same distribution because* φ *is even. Therefore we must have* $\forall t > 0$, $\partial_x \mathfrak{c}(0,t) = 0$. *The solution of*

(7.1) *is thus estimated by applying a finite difference method to*

$$\begin{cases} \partial_t \mathfrak{c}(x,t) + \frac{C^2}{2}\partial_x^2 \mathfrak{c}(x,t) - c_{\mathrm{f}}\partial_x \mathfrak{c}(x,t) = 0 \quad \text{for } (x,t) \in (0,\infty)\times[0,1), \\ \partial_x \mathfrak{c}(0,t) = 0 \quad \text{for } t \in [0,1), \\ \mathfrak{c}(x,1) = x^2 \quad \text{for } x \in [0,+\infty). \end{cases} \tag{7.2}$$

The whole function $x \mapsto \mathfrak{c}(x,t)$ *can be recovered by using the symmetry property.*

EXAMPLE 7.2 (ELASTO-PLASTIC BEHAVIOR). *Consider equation* (2.31) *for* $\varphi \triangleq 0$, $\psi \triangleq \chi_D$ *the indicator function of* $D \triangleq [-c_{\mathrm{ep}}, c_{\mathrm{ep}}]$ *in the sense of convex analysis, that is,* $\chi_D(x) = 0$ *if* $x \in D$ *and* $+\infty$ *otherwise. Here* $c_{\mathrm{ep}} > 0$ *is an elasto-plastic coefficient. The real-valued process* X^ε *represents the velocity of a material point subjected to an elasto-plastic restoring force and colored noise. The process* Z^ε *taking values in* $[-c_{\mathrm{ep}}, c_{\mathrm{ep}}]$ *represents the restoring force. See, for instance, [26] for an explanation of the physics and the use of SDEs with multivalued drift for modeling. Here we are interested in* $\mathbb{E}[(X_T^\varepsilon)^2 + (Z_T^\varepsilon)^2]$ *and in* $\mathbb{P}(|Z_T^\varepsilon| = c_{\mathrm{ep}})$ *for* $T = 1$. *As* $\varepsilon \to 0$, $(X^\varepsilon, Z^\varepsilon) \to (X^0, Z^0)$ *where* (X^0, Z^0) *satisfies* (5.11). *For the stochastic simulation, we proceed as follows*:

- $\widehat{X}_0^{\varepsilon,k} = x_0$ *and* $\widehat{Z}_0^{\varepsilon,k} = z_0$ *and for* $0 \le n \le N_T - 1$,

$$\begin{cases} \widehat{Z}_{n+1}^{\varepsilon,k} = \mathrm{proj}_{[-c_{\mathrm{ep}},c_{\mathrm{ep}}]}\big(\widehat{Z}_n^{\varepsilon,k} + \delta t\, \widehat{X}_n^{\varepsilon,k}\big), \\ \widehat{X}_{n+1}^{\varepsilon,k} = \widehat{X}_n^{\varepsilon,k} - \delta t\, \widehat{Z}_n^{\varepsilon,k} + \frac{\delta t}{\varepsilon}\widehat{\eta}_n^{\varepsilon,k}. \end{cases}$$

- $\widehat{X}_0^{0,k} = x_0$ *and* $\widehat{Z}_0^{0,k} = z_0$ *and for* $0 \le n \le N_T - 1$,

$$\begin{cases} \widehat{Z}_{n+1}^{0,k} = \mathrm{proj}_{[-c_{\mathrm{ep}},c_{\mathrm{ep}}]}\big(\widehat{Z}_n^{0,k} + \delta t\, \widehat{X}_n^{0,k}\big), \\ \widehat{X}_{n+1}^{0,k} = \widehat{X}_n^{0,k} - \delta t\, \widehat{Z}_n^{0,k} + \sqrt{\delta t}\, C\, \Delta W_n^k. \end{cases}$$

The expectations of the control variates are

$$\mathbb{E}\big[\big(X_T^0\big)^2 + \big(Z_T^0\big)^2\big] \quad \textit{and} \quad \mathbb{P}\big(\big|Z_T^0\big| = c_{\mathrm{ep}}\big) \quad \textit{for } T = 1.$$

They are estimated using the PDE method of [27].

EXAMPLE 7.3 (NONLINEAR AND NONSMOOTH TDOF). *A TDOF with an elasto-plastic element can be represented as a system of the form* (2.30), *which becomes*

$$\begin{cases} \dfrac{\mathrm{d}X_1^\varepsilon}{\mathrm{d}t} + \partial\chi_D(X_1^\varepsilon) = X_2^\varepsilon, \quad \dfrac{\mathrm{d}X_2^\varepsilon}{\mathrm{d}t} = -\widetilde{g}_2(X_1^\varepsilon, X_2^\varepsilon, X_3^\varepsilon, X_4^\varepsilon) + \dfrac{1}{\varepsilon}\eta^\varepsilon, \\ \dfrac{\mathrm{d}X_3^\varepsilon}{\mathrm{d}t} = X_4^\varepsilon, \quad \dfrac{\mathrm{d}X_4^\varepsilon}{\mathrm{d}t} = -\widetilde{g}_4(X_1^\varepsilon, X_2^\varepsilon, X_3^\varepsilon, X_4^\varepsilon). \end{cases} \tag{7.3}$$

Here $D = [-c_p, c_p]$, $c_p = 0.25$. *In addition to* $\mathbb{E}[(X^\varepsilon_{1,T})^2 + (X^\varepsilon_{3,T})^2]$ *for* $T = 1$, *we are interested in* $\mathbb{P}(|X^\varepsilon_{1,T}| \le a, |X^\varepsilon_{3,T}| \le b)$. *As* $\varepsilon \to 0$, $\boldsymbol{X}^\varepsilon \to \boldsymbol{X}^0$ *where*

$$\begin{cases} \mathrm{d}X^0_1 + \partial\chi_D(X^0_1)\mathrm{d}t = X^0_2\,\mathrm{d}t, \quad \mathrm{d}X^0_2 = -\widetilde{g}_2(X^0_1, X^0_2, X^0_3, X^0_4)\mathrm{d}t + C\,\mathrm{d}W, \\ \mathrm{d}X^0_3 = X^0_4\,\mathrm{d}t, \quad \mathrm{d}X^0_4 = -\widetilde{g}_4(X^0_1, X^0_2, X^0_3, X^0_4)\mathrm{d}t. \end{cases} \tag{7.4}$$

Here, for simplicity the other elements are linear:

$$\widetilde{g}_2(x_1, x_2, x_3, x_4) \triangleq k_1 x_1 + c_1 x_2 - k_3(x_3 - x_1) - c_3(x_4 - x_2)$$

and

$$\widetilde{g}_4(x_1, x_2, x_3, x_4) \triangleq k_3(x_3 - x_1) + c_3(x_4 - x_2).$$

The simulation of (7.3) *and* (7.4) *is similar to what is explained above. We take* $c_1 = c_3 = k_1 = k_3 = 1$.

7.2 Nonsmooth systems: beyond equations (2.30) and (2.31)

The two models presented in this subsection do not fall in the scope of our theoretical results, though they are not too far off. The presentation of the impact problem remains formal. The behavior of the control variate estimator is investigated via numerical experiments.

EXAMPLE 7.4 (IMPACT PROBLEM). *The pair displacement-velocity* $\boldsymbol{X}^\varepsilon = (X^\varepsilon_1, X^\varepsilon_2)$ (*taking values in* $\mathbb{R}^2$) *of a colored noise-driven oscillator constrained by an obstacle can be formulated in terms of an equation of the form* (4.1) *when* $|X^\varepsilon_{1,t}| < P_{\mathrm{O}}$ *with the condition* (*that expresses the switch of the velocity at collision*): *for all* t,

$$|X^\varepsilon_{1,t}| = P_{\mathrm{O}} \implies X^\varepsilon_{2,t+} = -\mathfrak{e} X^\varepsilon_{2,t-}$$

where P_{O} *is the location of the obstacle and* $\mathfrak{e} \in [0, 1]$ *is the coefficient of restitution of energy. The notations* $X^\varepsilon_{2,t\pm}$ *stand for the velocity immediately before and after the collision. Here we are interested in* $\mathbb{E}[(X^\varepsilon_{2,T})^2]$ *for* $T = 1$. *Formally, as* $\varepsilon \to 0$, *the* $\mathbb{R}^2$*-valued limit process* $\boldsymbol{X}^0 = (X^0_1, X^0_2)$ *is a white-noise-driven oscillator constrained by an obstacle that can be formulated similarly to the former case except that we replace* (4.1) *by* (4.2). *When* $\mathfrak{e} = 1$ (*resp.,* $0 \le \mathfrak{e} < 1$), *we say that the collisions are elastic* (*resp., inelastic*). *It is important to stress that obstacle problems with inelastic collisions deserve more attention for practical purposes, since in real world phenomena kinetic energy is dissipated through heat or plastic deformation. With elastic collisions, there is no loss of kinetic energy. For the stochastic simulation, we use the same numerical procedure as for* (4.1) *and* (4.2) *except that if we find out that the* $(n+1)^{st}$ *point does not satisfy the obstacle condition, i.e.,* $|\widehat{X}^{\varepsilon,k}_{1,n+1}| > P_{\mathrm{O}}$, *we adjust the time step length to* $\theta_{n+1}\delta t$ *with*

$$\theta_{n+1} \triangleq \frac{\pm P_{\mathrm{O}} - \widehat{X}^{\varepsilon,k}_{1,n}}{\widehat{X}^{\varepsilon,k}_{1,n+1} - \widehat{X}^{\varepsilon,k}_{1,n}} \quad \textit{and set} \quad t_{n+1} \triangleq t_n + \theta_{n+1}\delta t,\ \widehat{X}^{\varepsilon,k}_{1,n+1} \triangleq P_{\mathrm{O}},$$

$$\widehat{X}^{\varepsilon,k}_{2,n+1} \triangleq -\mathfrak{e}\left(\widehat{X}^{\varepsilon,k}_{2,n} - \theta_{n+1}\delta t f(\widehat{X}^{\varepsilon,k}_{1,n}, \widehat{X}^{\varepsilon,k}_{2,n}) + \theta_{n+1}\frac{\delta t}{\varepsilon}\widehat{\eta}^{\varepsilon,k}_n\right),$$
$$\widehat{\eta}^{\varepsilon,k}_{2,n+1} \triangleq \widehat{\eta}^{\varepsilon,k}_n\left(1 - \theta_{n+1}\delta t\,\frac{A}{\varepsilon^2}\right) + \sqrt{\theta_{n+1}\delta t}\,\frac{K_{\mathrm{ou}}}{\varepsilon}\Delta W^k_n.$$

A similar adjustment is done in the other cases with Langevin and white noises. The expectation of the control variate is $\mathbb{E}[(X^0_{2,T})^2]$ *for* $T = 1$*, which is estimated using the PDE method of* [27].

EXAMPLE 7.5 (REFLECTION OF AN INTEGRATED COLORED NOISE). *Define* $E \triangleq [0,\infty)$ *and consider the indicator function of* E*, that is,* $\chi_E(x) = 0$ *if* $x \in E$ *and* $+\infty$ *otherwise. The reflection of an integrated colored noise corresponds to the case where* X^ε *satisfies*

$$\frac{\mathrm{d}X^\varepsilon}{\mathrm{d}t} + \partial\chi_E(X^\varepsilon) \ni \frac{1}{\varepsilon}\eta^\varepsilon, \tag{7.5}$$

and X^0*, the limit process as* $\varepsilon \to 0$*, is a reflected Brownian motion*

$$\mathrm{d}X^0 + \partial\chi_E(X^0)\mathrm{d}t \ni C\,\mathrm{d}W. \tag{7.6}$$

We are interested in $\mathbb{E}[X^\varepsilon_T]$ *for* $T = 1$*. For the stochastic simulation of* (7.5) *and* (7.6)*, we use the following scheme:* $\widehat{X}^{\varepsilon,k}_0 = x_0$*,* $\widehat{X}^{0,k}_0 = x_0$*, and for* $0 \le n \le N_T - 1$*,*

- $\widehat{X}^{\varepsilon,k}_{n+1} = \mathrm{proj}_E\left(\widehat{X}^{\varepsilon,k}_n + \frac{\delta t}{\varepsilon}\widehat{\eta}^{\varepsilon,k}_n\right)$,
- $\widehat{X}^{0,k}_{n+1} = \mathrm{proj}_E\left(\widehat{X}^{0,k}_n + \sqrt{\delta t}C\Delta W^k_n\right)$.

The expectation of the control variate is given by an explicit formula $\mathbb{E}[X^0_1] = \sqrt{2/\pi}$*. Indeed, the backward Kolmogorov equation for the reflected Brownian motion in* (7.6) *is*

$$\partial_t w = C\partial^2_x w, \quad x > 0,\ t > 0,$$
$$w(x, t = 0) = x,\ x > 0, \quad \partial_x w(0,t) = 0,\ t > 0.$$

It has an explicit solution

$$w(x,t) = \frac{1}{\sqrt{4C\pi t}}\int_0^\infty y\left(\exp\left(-\frac{(x-y)^2}{4Ct}\right) + \exp\left(-\frac{(x+y)^2}{4Ct}\right)\right)\mathrm{d}y,$$

which gives $\mathbb{E}[X^0_T] = w(0,T) = \sqrt{2/\pi}$ *for* $T = 1$*. In this case, we can provide an ad hoc proof to get an estimate similar to* (3.4) (*see [9, app. H]*)*:*

$$\mathbb{E}\left[(X^\varepsilon_T - X^0_T)^2\right] \le C\varepsilon^2|\log\varepsilon|. \tag{7.7}$$

The $\log\varepsilon$ *correction comes from a maximal inequality for the OU process* [14] *and a standard result on the maxima of Gaussian processes* [25].

7.3 Numerical experiments

We report our numerical results for the four systems mentioned above. The convention is as in Section 4.3. In each of the four figures below (Figures 7.1–7.4), there are four subfigures (a)–(d). For subfigures (a) and (b), the driving force is an Ornstein-Uhlenbeck noise (3.1) with $A = K = 1$. In subfigure (a), the dashed black lines, the dotted blue lines, and the solid red lines represent the standard MC estimator $\widehat{I}_N^\varepsilon$ and the control variate estimators $\widehat{J}_N^\varepsilon$ and $\widehat{K}_N^\varepsilon$, respectively. The dotted black line represents the expectation of the control variate I^0. The objective of the subfigure (b) is to illustrate the bounds (6.8) and (6.9). The same description applies to (c) and (d), except that they correspond to the case of a Langevin noise (3.2)–(3.3) with $\mu = \gamma = K = 1$. For the examples in which $\boldsymbol{X}^\varepsilon$ satisfies (2.30), the asymptotic variance of the standard MC estimator $\widehat{I}_N^\varepsilon$ is estimated by (4.7), the asymptotic variance of the control variate estimator $\widehat{J}_N^\varepsilon$ is estimated by (4.8), and the asymptotic variance of the optimal control variate estimator $\widehat{K}_N^\varepsilon$ is estimated by (4.9). For the examples in which $(\boldsymbol{X}^\varepsilon, \boldsymbol{Z}^\varepsilon)$ satisfies (2.31), the asymptotic variance of the standard MC estimator $\widehat{I}_N^\varepsilon$ is estimated by

$$\widehat{\sigma}^2_{I^\varepsilon,N} = \frac{1}{N}\sum_{k=1}^N \big(F(\boldsymbol{X}^\varepsilon(\boldsymbol{W}^k), \boldsymbol{Z}^\varepsilon(\boldsymbol{W}^k)))\big)^2 - \big(\widehat{I}_N^\varepsilon\big)^2, \tag{7.8}$$

the asymptotic variance of the control variate estimator $\widehat{J}_N^\varepsilon$ is estimated by

$$\begin{aligned}\widehat{\sigma}^2_{J^\varepsilon,N} = \frac{1}{N}\sum_{k=1}^N \big(&F(\boldsymbol{X}^\varepsilon(\boldsymbol{W}^k), \boldsymbol{Z}^\varepsilon(\boldsymbol{W}^k)) \\ &- F(\boldsymbol{X}^0(\boldsymbol{W}^k), \boldsymbol{Z}^0(\boldsymbol{W}^k)) + I^0\big)^2 - \big(\widehat{J}_N^\varepsilon\big)^2,\end{aligned} \tag{7.9}$$

and the asymptotic variance of the optimal control variate estimator $\widehat{K}_N^\varepsilon$ is estimated by

$$\begin{aligned}\widehat{\sigma}^2_{K^\varepsilon,N} = \frac{1}{N}\sum_{k=1}^N \big(&F(\boldsymbol{X}^\varepsilon(\boldsymbol{W}^k), \boldsymbol{Z}^\varepsilon(\boldsymbol{W}^k)) \\ &- \widehat{\rho}_N^\varepsilon F(\boldsymbol{X}^0(\boldsymbol{W}^k), \boldsymbol{Z}^0(\boldsymbol{W}^k)) + \widehat{\rho}_N^\varepsilon I^0\big)^2 - \big(\widehat{K}_N^\varepsilon\big)^2,\end{aligned} \tag{7.10}$$

with $\widehat{\rho}_N^\varepsilon$ defined by (6.4). $\widehat{\sigma}^2_{I^\varepsilon,N}$, $\widehat{\sigma}^2_{J^\varepsilon,N}$, and $\widehat{\sigma}^2_{K^\varepsilon,N}$ are consistent estimators of $\sigma^2_{I^\varepsilon}$, $\sigma^2_{J^\varepsilon}$, and $\sigma^2_{K^\varepsilon}$, respectively.

Similarly to what was presented in Section 4, we use $N = 10^4$ samples with a time step of $\delta t = 10^{-5}$. In Figures 7.1 and 7.2, we report the numerical results for the friction and elasto-plastic problems, which are of the form (2.30) and (2.31), respectively. In Figures 7.3 and 7.4, we report the numerical results for the obstacle problem and for the reflection of the integral of a colored noise. The numerical results include errors bars on the estimators for each value of $\varepsilon \in \{0.1, 0.5, 0.9\}$ in

the $\widehat{I}_N^\varepsilon, \widehat{J}_N^\varepsilon, \widehat{K}_N^\varepsilon$ order.

The theoretical predictions provided by (6.8) and (6.9) are based on the condition that f has bounded derivatives. The assumption that f is smooth is important, but the hypothesis on the boundedness of the derivatives can certainly be relaxed. The numerical results shown in Figures 7.1 and 7.2 are in good agreement with the theoretical predictions: the asymptotic variances $\sigma^2_{J^\varepsilon}$ and $\sigma^2_{K^\varepsilon}$ behave as $O(\varepsilon^2)$. The only cases where the behavior is $O(\varepsilon)$, and not $O(\varepsilon^2)$, are when the quantity of interest is of the form $\mathbb{E}[f(\boldsymbol{X}_T^\varepsilon)]$ or $\mathbb{E}[f(\boldsymbol{X}_T^\varepsilon, \boldsymbol{Z}_T^\varepsilon)]$ with a function f that is not smooth, which is not surprising. In Figure 7.2, we also observe that $\sigma^2_{J^\varepsilon}$ and $\sigma^2_{K^\varepsilon}$ behave as $O(\varepsilon^2)$, which is better than the behavior $O(\varepsilon)$ expected from (6.9) (which is an upper bound).

In Figures 7.3 and 7.4, the numerical results concern two problems that do not fall within the scope of our theoretical predictions. The first one (Figure 7.3) is the impact problem that cannot be formulated in the form of a differential inclusion similar to (2.30) or (2.31). The function f is smooth but the behavior of $\sigma^2_{J^\varepsilon}$ and $\sigma^2_{K^\varepsilon}$ is not of order $O(\varepsilon^2)$, only of order $O(\varepsilon)$. The second one (Figure 7.4) is the reflection of an integrated colored noise that can be formulated with a differential inclusion that is similar to (2.30) but the multivalued drift does not satisfy the condition (5.7). However, $\sigma^2_{J^\varepsilon}$ and $\sigma^2_{K^\varepsilon}$ behave as $O(\varepsilon^2)$. To summarize, the numerical simulations indicate that the $O(\varepsilon^2)$ behavior of the asymptotic variance is observed in the cases predicted by the theory and also slightly beyond. The smoothness of the function f that appears in the quantity of interest is, however, an important condition to ensure the $O(\varepsilon^2)$-behavior; otherwise one only observes an $O(\varepsilon)$-behavior.

8 Concluding Remark

When the expectation of the limit process $\mathbb{E}[f(\boldsymbol{X}_T^0)]$ cannot be computed by a PDE method but is estimated by a massive Monte Carlo method, the control variate method with $\widehat{J}_N^\varepsilon$ (or $\widehat{K}_N^\varepsilon$) shares an important similarity with a two-level Monte Carlo method [10] in the sense that many but cheap simulations are performed (samples of $\boldsymbol{X}^0$ used to estimate $\mathbb{E}[f(\boldsymbol{X}_T^0)]$) together with a few expensive simulations (samples of $(\boldsymbol{X}^\varepsilon, \boldsymbol{X}^0)$ used to estimate $\mathbb{E}[f(\boldsymbol{X}_T^\varepsilon) - f(\boldsymbol{X}_T^0)]$). More generally, multilevel Monte Carlo (MLMC) methods rely on random samples taken on different levels of accuracy, when several approximations with different costs and accuracies are available. The overall idea of MLMC methods is to reduce the computational cost of standard Monte Carlo methods by taking most samples with a low accuracy and corresponding low cost, and by taking only few samples with a high accuracy and corresponding high cost [10].

In this two-level Monte Carlo framework, the total cost of computing the control variate estimator $\widehat{J}_N^\varepsilon$ (or $\widehat{K}_N^\varepsilon$) is $C_J^\varepsilon = N_0 C_0 + N_1 C_1$ where C_0 is the cost of computing one realization of $f(\boldsymbol{X}_T^0)$, C_1 is the cost of computing one realization

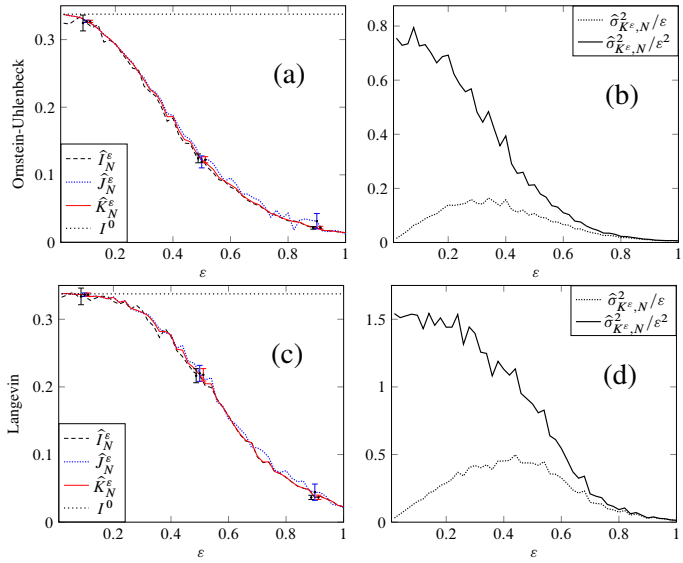

FIGURE 7.1. Example 7.1 (friction problem). The aim is to estimate $I^\varepsilon = \mathbb{E}[(X_T^\varepsilon)^2]$ for $T = 1$ where X^ε satisfies (2.30) with $\varphi(x) \triangleq c_f|x|$ with $c_f > 0$. The expectation of the control variate $I^0 = \mathbb{E}[(X_T^0)^2]$, where X^0 satisfies (5.9), is obtained by solving the partial differential inclusion (7.1).

of $(f(\boldsymbol{X}_T^\varepsilon) - f(\boldsymbol{X}_T^0))$, N_0 is the number of samples of $f(\boldsymbol{X}_T^0)$, and N_1 is the number of samples of $(f(\boldsymbol{X}_T^\varepsilon) - f(\boldsymbol{X}_T^0))$. The variance of the control variate estimator $\widehat{J}_N^\varepsilon$ is $V_J^\varepsilon = N_0^{-1}V_0 + N_1^{-1}V_1$, where V_0 is the variance of $f(\boldsymbol{X}_T^0)$ and V_1 is the variance of $(f(\boldsymbol{X}_T^\varepsilon) - f(\boldsymbol{X}_T^0))$. For a fixed total budget $C_{\rm tot}$, the variance is minimized when $(N_1/N_0)^2 = (V_1/V_0)(C_0/C_1)$ and it is then equal to $V_J^\varepsilon = (\sqrt{V_0C_0} + \sqrt{V_1C_1})^2/C_{\rm tot}$. This can be compared to the brute force Monte Carlo method: the cost is $C_I^\varepsilon = N_I C_1$ where N_I is the number of samples of $f(\boldsymbol{X}_T^\varepsilon)$ (we neglect the difference of cost between $(f(\boldsymbol{X}_T^\varepsilon) - f(\boldsymbol{X}_T^0))$ and $f(\boldsymbol{X}_T^\varepsilon)$, which is very small because $\boldsymbol{X}^\varepsilon$ is more difficult to simulate than $\boldsymbol{X}^0$) and the variance is $V_I^\varepsilon = V_0/N_I$ (by (2.11)), so that for the total budget $C_{\rm tot}$, we have $V_I^\varepsilon = V_0C_1/C_{\rm tot}$. If $V_1/V_0 = O(\varepsilon^2)$ (by Proposition 2.1) and $C_0/C_1 = O(\varepsilon^2)$ (because the time step used to simulate $\boldsymbol{X}^\varepsilon$ should be $O(\varepsilon^2)$ smaller than the one used to simulate $\boldsymbol{X}^0$), then we find that $N_1/N_0 = O(\varepsilon^2)$ (hence the "massive Monte Carlo" strategy for the control variate) and the ratio of the variance of the control variate estimator over the one of the brute force Monte Carlo estimator is finally $V_J^\varepsilon/V_I^\varepsilon = O(\varepsilon^2)$. The control variate method is very advantageous in this context.

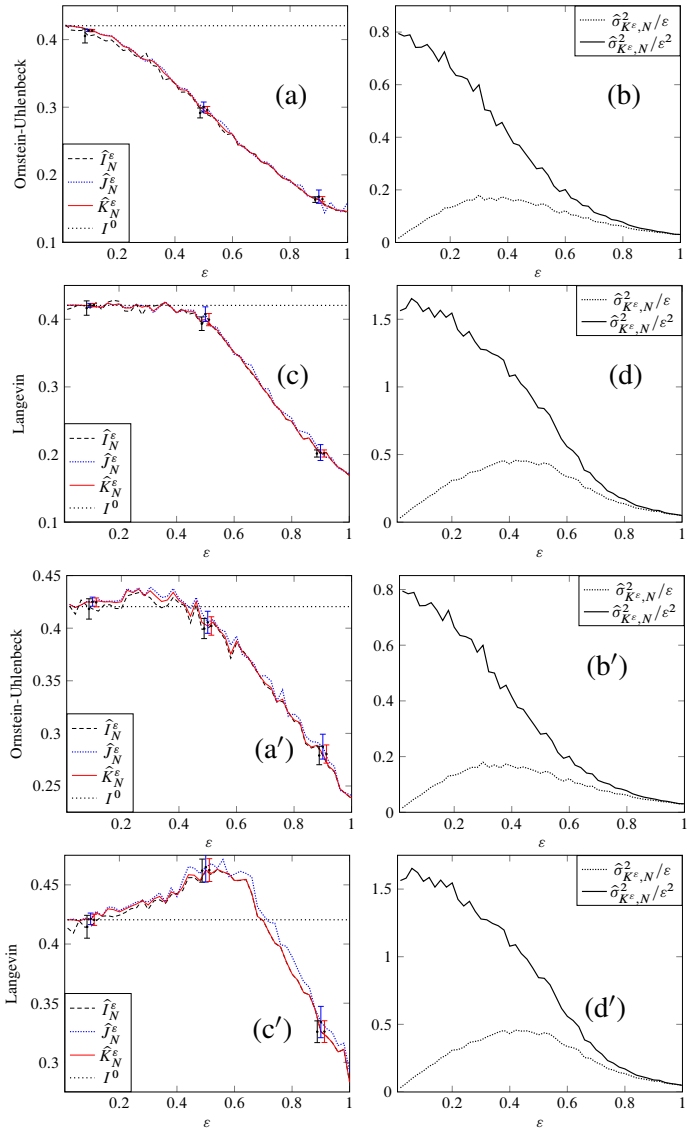

FIGURE 7.2. Example 7.2 (elasto-plastic problem). In subfigures (a)–(d) the aim is to estimate $I^\varepsilon = \mathbb{E}[(X^\varepsilon_T)^2 + (Z^\varepsilon_T)^2]$ with $T = 1$ where $(X^\varepsilon, Z^\varepsilon)$ satisfies (2.31) with $\varphi(x) \triangleq 0$ and $\psi(x) \triangleq 0$ if $|x| \le c_{\rm ep}$ and $+\infty$ otherwise. Here $c_{\rm ep} = 0.25$. The expectation of the control variate $I^0 = \mathbb{E}[(X^0_T)^2 + (Z^0_T)^2]$, where (X^0, Z^0) satisfies (5.11), is obtained by using the PDE method of [27]. In subfigures (a′)–(d′) the aim is to estimate $I^\varepsilon = \mathbb{P}(|Z^\varepsilon_T| = c_{\rm ep})$.

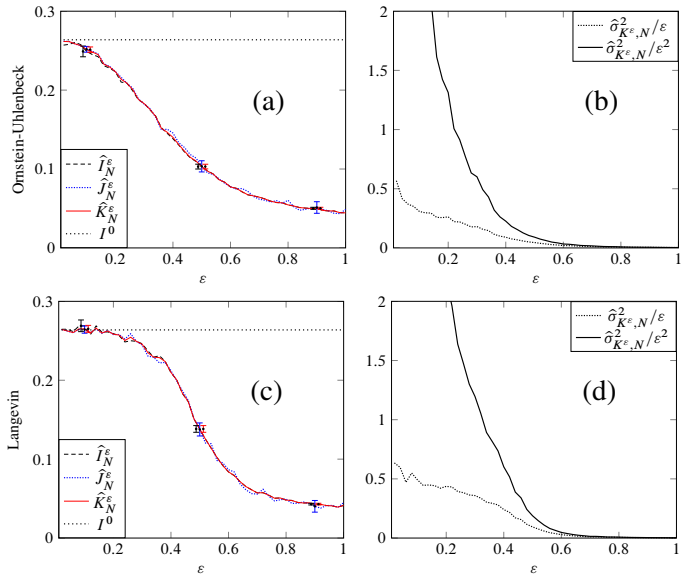

FIGURE 7.3. Example 7.4 (impact problem). The aim is to estimate $I^\varepsilon = \mathbb{E}[(X^\varepsilon_{2,T})^2]$ with $T = 1$ where $\boldsymbol{X}^\varepsilon$ satisfies the impact problem with a colored noise forcing. Here $P_O = 0.25$. The expectation of the control variate $I^0 = \mathbb{E}[(X^0_{2,T})^2]$, where $\boldsymbol{X}^0$ satisfies the impact problem with a colored noise forcing, is obtained by using the PDE method of [27].

Appendix A Proof of Proposition 3.5

Let us first study the driving noise. The process $\boldsymbol{\eta}^1$ is a Gaussian, Markov process. It has the form

$$\boldsymbol{\eta}^1_t = e^{-\mathbf{A}t}\boldsymbol{\eta}^1_0 + \int_0^t e^{-\mathbf{A}(t-s)}\mathbf{K}\,\mathrm{d}\boldsymbol{W}_s.$$

Its infinitesimal generator is

$$Q = \frac{1}{2}\sum_{k,k'=1}^{d}\sum_{k''=1}^{d'} K_{kk''}K_{k'k''}\partial^2_{\eta_k\eta_{k'}} - \sum_{k,k'=1}^{d} A_{kk'}\eta_{k'}\partial_{\eta_k}. \tag{A.1}$$

The properties of the matrix $\mathbf{A}$ show that the process $\boldsymbol{\eta}^1$ is stationary and ergodic; its unique invariant probability measure is the Gaussian measure with mean zero and variance $\mathbf{C}$ given by (2.8).

The process $(\boldsymbol{\eta}^\varepsilon_t)_{t\geq 0}$ has the same distribution as $(\boldsymbol{\eta}^1_{t/\varepsilon^2})_{t\geq 0}$ since $(\varepsilon^{-1}W_t)_{t\geq 0}$ has the same distribution as $(W_{t/\varepsilon^2})_{t\geq 0}$. Therefore it is a Markov process with generator $\varepsilon^{-2}Q$.

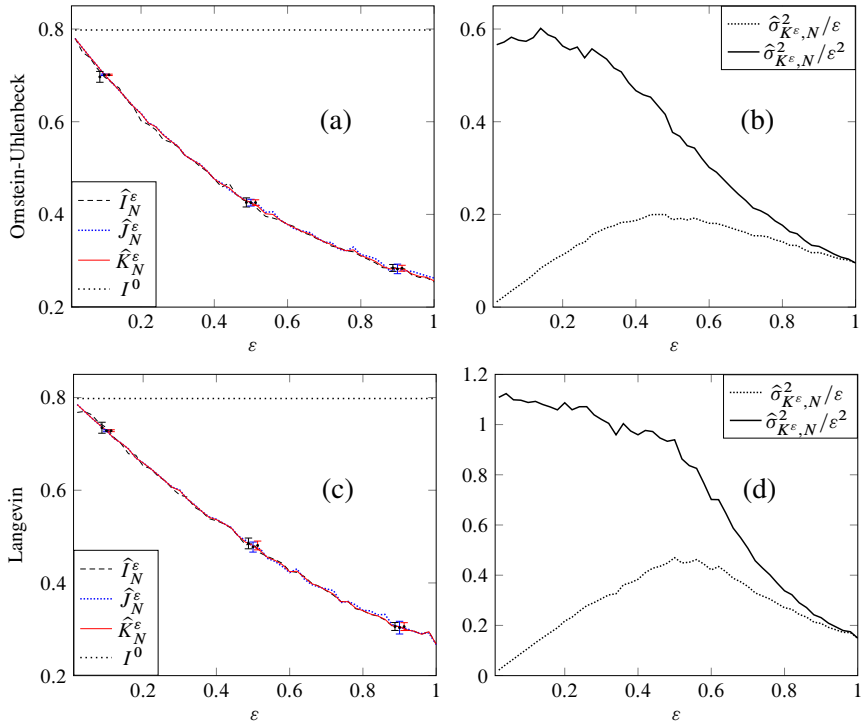

FIGURE 7.4. Example 7.5 (reflection of an integrated colored noise). The aim is to estimate $I^\varepsilon = \mathbb{E}[X^\varepsilon_T]$ for $T = 1$ where X^ε satisfies (7.5). The expectation of the control variate is $I^0 = \mathbb{E}[X^0_T] = \sqrt{2/\pi}$.

The process $(\boldsymbol{X}^\varepsilon, \boldsymbol{X}^0, \boldsymbol{\eta}^\varepsilon)$ is Markov with generator $\mathcal{L}^\varepsilon$ given by

$$\begin{aligned}\mathcal{L}^\varepsilon &= \frac{1}{\varepsilon^2}Q + \frac{1}{\varepsilon}\Big[\sum_{j=1}^{n}\sum_{i=1}^{d}\sigma_{ji}(\boldsymbol{x})\eta_i\partial_{x_j} + \sum_{j=1}^{n}\sum_{i=1}^{d}\sum_{k=1}^{d'}\Gamma_{jk}(\boldsymbol{x}^0)K_{ik}\partial^2_{\eta_i x^0_j}\Big] \\ &+ \Big[\sum_{j=1}^{n} b_j(\boldsymbol{x})\partial_{x_j} + \frac{1}{2}\sum_{i,j=1}^{n}\sum_{k=1}^{d'}\Gamma_{ik}(\boldsymbol{x}^0)\Gamma_{jk}(\boldsymbol{x}^0)\partial^2_{x^0_i x^0_j} \\ &\quad + \sum_{j=1}^{n}\widetilde{b}_j(\boldsymbol{x}^0)\partial_{x^0_j}\Big], \end{aligned} \tag{A.2}$$

where Q is the generator (A.1).

LEMMA A.1. *For any smooth and bounded test function $\phi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, for any compact subset K of $\mathbb{R}^{2n}$, there exists a test function ϕ^ε such that*

$$\begin{aligned} &\sup_{(\boldsymbol{x},\boldsymbol{x}^0)\in K} |\phi^\varepsilon(\boldsymbol{x},\boldsymbol{x}^0,\boldsymbol{\eta}) - \phi(\boldsymbol{x},\boldsymbol{x}^0)| \le C\varepsilon(1+\|\boldsymbol{\eta}\|^2), \\ &\sup_{(\boldsymbol{x},\boldsymbol{x}^0)\in K} |\mathcal{L}^\varepsilon\phi^\varepsilon(\boldsymbol{x},\boldsymbol{x}^0,\boldsymbol{\eta}) - \mathcal{L}\phi(\boldsymbol{x},\boldsymbol{x}^0)| \le C\varepsilon(1+\|\boldsymbol{\eta}\|^3), \end{aligned} \tag{A.3}$$

for any $\varepsilon \in (0,1)$, where $\mathcal{L}$ is the generator defined by

$$\begin{aligned} \mathcal{L} = &\sum_{j=1}^n \widetilde{b}_j(\boldsymbol{x})\partial_{x_j} + \sum_{j=1}^n \widetilde{b}_j(\boldsymbol{x}^0)\partial_{x_j^0} + \frac{1}{2}\sum_{i,j=1}^n [\boldsymbol{\Gamma}(\boldsymbol{x})\boldsymbol{\Gamma}(\boldsymbol{x})^\mathsf{T}]_{ij}\partial^2_{x_i x_j} \\ &+ \frac{1}{2}\sum_{i,j=1}^n [\boldsymbol{\Gamma}(\boldsymbol{x}^0)\boldsymbol{\Gamma}(\boldsymbol{x}^0)^\mathsf{T}]_{ij}\partial^2_{x_i^0 x_j^0} + \sum_{i,j=1}^n [\boldsymbol{\Gamma}(\boldsymbol{x})\boldsymbol{\Gamma}(\boldsymbol{x}^0)^\mathsf{T}]_{ij}\partial^2_{x_i x_j^0}. \end{aligned} \tag{A.4}$$

PROOF OF LEMMA A.1. The proof is an application of the perturbed test function method (see the extended online version [9]). □

Appendix B Proof of Lemma 3.8

We define

$$\phi(\boldsymbol{x},\boldsymbol{x}^0) = g(\boldsymbol{x}^0)\big(f(\boldsymbol{x}) - f(\boldsymbol{x}^0)\big) \quad \text{or} \quad \phi(\boldsymbol{x},\boldsymbol{x}^0) = (f(\boldsymbol{x}) - f(\boldsymbol{x}^0))^2. \tag{B.1}$$

Lemma A.1 applied to ϕ gives an estimate for (3.4) of order ε, but the particular form of ϕ allows us to get ε^2. Lemma 3.8 can be proved in four steps; see the extended online version [9].

Appendix C Existence and Uniqueness of (2.30) and (2.31)

PROPOSITION C.1. *Fix $T > 0$, $n \in \mathbb{N}^\star$. Suppose that $\boldsymbol{f} \in \mathcal{C}([0,T];\mathbb{R}^n)$, $\boldsymbol{b}$ is Lipschitz, and φ is a l.s.c. convex function satisfying* (5.7). *Then there exists a unique solution $\boldsymbol{x} \in \mathcal{C}([0,T];\mathbb{R}^n)$ to the differential inclusion*

$$\dot{\boldsymbol{x}}(t) + \partial\varphi(\boldsymbol{x}(t)) \ni \boldsymbol{b}(\boldsymbol{x}(t)) + \boldsymbol{f}(t), \quad t > 0, \tag{C.1}$$

with $\boldsymbol{x}(0) = \boldsymbol{x}_0 \in \mathbb{R}^n$.

PROPOSITION C.2. *Fix $T > 0$, $n, m \in \mathbb{N}^\star$. Suppose that $\boldsymbol{f} \in \mathcal{C}([0,T];\mathbb{R}^n)$, $\boldsymbol{b}^x : \mathbb{R}^n \to \mathbb{R}$ and $\boldsymbol{b}^z : \mathbb{R}^m \to \mathbb{R}$ are Lipschitz, and $\varphi : \mathbb{R}^n \to \mathbb{R}$, $\psi : \mathbb{R}^m \to \mathbb{R}$ are l.s.c. convex functions, with φ satisfying* (5.7) *and ψ satisfying* (5.8). *Then there exists a unique solution $(\boldsymbol{x},z) \in \mathcal{C}([0,T];\mathbb{R}^n \times \mathbb{R}^m)$ to the following differential inclusion*:

$$\begin{aligned} &\dot{\boldsymbol{x}}(t) + \partial\varphi(\boldsymbol{x}(t)) \ni \boldsymbol{b}^x(\boldsymbol{x}(t), z(t)) + \boldsymbol{f}(t), \\ &\dot{z}(t) + \partial\psi(z(t)) \ni \boldsymbol{b}^z(\boldsymbol{x}(t), z(t)), \end{aligned} \quad t > 0, \tag{C.2}$$

with $(\boldsymbol{x}(0), z(0)) = (\boldsymbol{x}_0, z_0) \in \mathbb{R}^n \times \mathbb{R}^m$.

For the proofs of Propositions C.1 and C.2, see the extended online version [9].

Appendix D Existence and Uniqueness for (5.9)

Let us first take a look at the case where we remove the multivalued operator $\partial\varphi$ from the drift in (5.9). The problem becomes the same as (2.5) where $\boldsymbol{\sigma}$ is constant and in particular does not involve a stochastic integral. Thus, as pointed out in [18, p. 294], the proof of existence and uniqueness of a solution (still in [18, theorem 2.9, p. 289]) can be simplified in a way that makes no use of probabilistic tools. We consider the Wiener space

$$\big(\Omega \triangleq C([0,T];\mathbb{R}^d), \mathcal{F} \triangleq \mathcal{B}\big(C([0,T];\mathbb{R}^d)\big), \mathbb{P}\big)$$

where Ω is the space of $\mathbb{R}^d$-valued continuous functions on $[0,T]$ endowed with the norm

$$\forall \boldsymbol{\omega} \in \Omega, \quad \|\boldsymbol{\omega}\| \triangleq \sup_{0\le t\le T} \|\boldsymbol{\omega}(t)\|,$$

$\mathcal{B}(C([0,T];\mathbb{R}^d))$ is the Borel σ-algebra on Ω and $\mathbb{P}$ is the Wiener measure. We introduce the mappings indexed by $t \in [0,T]$, $\boldsymbol{W}_t(\cdot)\colon \Omega \to \mathbb{R}^d$, $\boldsymbol{\omega} \mapsto \boldsymbol{W}_t(\boldsymbol{\omega}) \triangleq \boldsymbol{\omega}(t)$, the sequence of σ-algebras $\mathcal{F}^t \triangleq \sigma\{\boldsymbol{W}_s,\ 0 \le s \le t\}$, and the map $\boldsymbol{X}\colon \Omega \to C([0,T];\mathbb{R}^n)$, $\boldsymbol{\omega} \mapsto \boldsymbol{X}(\boldsymbol{\omega}) \triangleq \boldsymbol{x}$ where $\forall 0 \le t \le T$, $\boldsymbol{x}(t) = \boldsymbol{x}(0) + \int_0^t \boldsymbol{b}(\boldsymbol{x}(s))\mathrm{d}s + \boldsymbol{\Gamma}\boldsymbol{\omega}(t)$. Under $\mathbb{P}$, $\boldsymbol{W}$ is a Wiener process and $\boldsymbol{X}(\boldsymbol{W})$ solves (2.5) where $\boldsymbol{\sigma}$ is constant. In this approach, the key ingredient is the mapping $\boldsymbol{X}$. For obtaining the existence and uniqueness of the solution to (5.9) with the multivalued operator $\partial\varphi$, we discuss below the properties of a similar mapping to $\boldsymbol{X}$ that involves the multivalued operator. This is done via the so-called "generalized Skorokhod problem." The discussion follows [30, pp. 245–252]. We use the notation $BV[0,T]$ for the space of functions with bounded variation on $[0,T]$.

DEFINITION D.1 (Generalized convex Skorokhod problem). If a pair of functions $(\boldsymbol{x}, \boldsymbol{\Delta})$ satisfies the conditions

(1) $\boldsymbol{x}, \boldsymbol{\Delta} : [0,T] \to \mathbb{R}^n$ are continuous, $\boldsymbol{x}(0) = \boldsymbol{x}_0$ and $\boldsymbol{\Delta}(0) = 0$,
(2) $\forall 0 \le t \le T,\ \boldsymbol{x}(t) \in \overline{\mathrm{Dom}(\partial\varphi)},\ \boldsymbol{\Delta} \in BV([0,T];\mathbb{R}^n)$,
(3) $\forall 0 \le t \le T,\ \boldsymbol{x}(t) + \boldsymbol{\Delta}(t) = x_0 + \int_0^t \boldsymbol{b}(\boldsymbol{x}(s))\mathrm{d}s + \boldsymbol{\Gamma}\boldsymbol{\omega}(t)$,
(4) $\forall 0 \le s \le t \le T,\ \forall \mathfrak{z} \in \mathbb{R}^n,\ \int_s^t (\boldsymbol{x}(r) - \mathfrak{z}) \cdot \mathrm{d}\boldsymbol{\Delta}(r) + \int_s^t \varphi(\boldsymbol{x}(r))\mathrm{d}r \le (t-s)\varphi(\mathfrak{z})$,

then we say that $\boldsymbol{x}$ solves the generalized Skorokhod problem with parameters $\partial\varphi$, $\boldsymbol{x}_0$, $\boldsymbol{b}$, and $\boldsymbol{\omega}$, and we use the notation $\boldsymbol{x} = \mathcal{GSP}(\partial\varphi, \boldsymbol{x}_0, \boldsymbol{b}, \boldsymbol{\omega})$.

Existence and uniqueness of a solution for the generalized Skorokhod problem can be found in [30, theorem 4.17, p. 252]. This is obtained under the following conditions : φ is an l.s.c. convex function and $\mathrm{int}(\mathrm{Dom}(\varphi)) \neq \varnothing$; $\boldsymbol{b}$ is Lipschitz, $\boldsymbol{x}_0 \in \overline{\mathrm{Dom}(\partial\varphi)}$, and $\boldsymbol{\omega} : [0,T] \mapsto \mathbb{R}^n$ is continuous with $\boldsymbol{\omega}(0) = 0$. The continuity of the mapping $\boldsymbol{X}\colon \Omega \to C([0,T];\mathbb{R}^n)$, $\boldsymbol{\omega} \mapsto \boldsymbol{X}(\boldsymbol{\omega}) \triangleq \boldsymbol{x}$, where $\boldsymbol{x} = \mathcal{GSP}(\partial\varphi, \boldsymbol{x}_0, \boldsymbol{b}, \boldsymbol{\omega})$ is shown in [30, prop. 4.16, p. 247]. We use the notation

$\mathcal{S}_n^0[0,T]$ for the space of progressively measurable continuous stochastic processes (p.m.c.s.p.) from $\Omega \times [0,T]$ to $\mathbb{R}^n$,

$$\mathcal{S}_n^2[0,T] \triangleq \Big\{ \boldsymbol{Z} \in \mathcal{S}_n^0[0,T],\ \mathbb{E}\Big[\sup_{0\le t\le T} \|\boldsymbol{Z}(t)\|^2 \Big] < \infty \Big\}.$$

Within the framework of the aforementioned Wiener space, $\boldsymbol{X}(\boldsymbol{W}) \in \mathcal{S}_n^0[0,T]$ solves (5.9). Furthermore, it can be shown that $\boldsymbol{X}(\boldsymbol{W}) \in \mathcal{S}_n^2[0,T]$.

Acknowledgments. LM expresses his sincere gratitude to Prof. Jean Michel Coron and the Sino-French International Associated Laboratory for Applied Mathematics for being supported for travels and housing at Ecole Polytechnique.

Bibliography

[1] Bensoussan, A.; Féau, C.; Mertz, L.; Yam, S. C. P. An analytical approach for the growth rate of the variance of the deformation related to an elasto-plastic oscillator excited by a white noise. *Appl. Math. Res. Express. AMRX* **2015** (2015), no. 1, 99–128. doi:10.1093/amrx/abu008

[2] Bensoussan, A.; Mertz, L.; Yam, S. C. P. Long cycle behavior of the plastic deformation of an elasto-perfectly-plastic oscillator with noise. *C. R. Math. Acad. Sci. Paris* **350** (2012), no. 17–18, 853–859. doi:10.1016/j.crma.2012.09.020

[3] Bernardin, F. Equations différentielles multivoques : aspects théoriques et numériques - Applications. Ph.D. thesis, Université Claude Bernard - Lyon I, 2004.

[4] Brézis, H. *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*. North-Holland Mathematics Studies, 5. Notas de Matemática (50), North-Holland, Amsterdam-London; Elsevier, New York, 1973.

[5] Cordle, A.; Jonkman, J. State of the art design tools for floating offshore wind turbines. 21st International Offshore and Polar Engineering Conference, Maui, Hawaii, June 19-24, 2011. Available at: `https://www.nrel.gov/docs/fy12osti/50543.pdf`

[6] E, W.; Liu, D.; Vanden-Eijnden, E. Analysis of multiscale methods for stochastic differential equations. *Comm. Pure Appl. Math.* **58** (2005), no. 11, 1544–1585. doi:10.1002/cpa.20088

[7] Féau, C.; Laurière, M.; Mertz, L. Asymptotic formulae for the risk of failure related to an elasto-plastic problem with noise. *Asymptot. Anal.* **106** (2018), no. 1, 47–60. doi:10.3233/asy-171445

[8] Fouque, J.-P.; Garnier, J.; Papanicolaou, G.; Sølna, K. *Wave propagation and time reversal in randomly layered media.* Stochastic Modelling and Applied Probability, 56. Springer, New York, 2007.

[9] Garnier, J; Mertz, L. A control variate method driven by diffusion approximation. [Online extension of the present paper] Preprint, 2019. arXiv:1906.01225 [math.PR]

[10] Giles, M. B. Multilevel Monte Carlo methods. *Acta Numer.* **24** (2015), 259–328. doi:10.1017/S096249291500001X

[11] Givon, D.; Kevrekidis, I. G.; Kupferman, R. Strong convergence of projective integration schemes for singularly perturbed stochastic differential systems. *Commun. Math. Sci.* **4** (2006), no. 4, 707–729.

[12] Glasserman, P. *Monte Carlo methods in financial engineering.* Applications of Mathematics (New York), 53. Stochastic Modelling and Applied Probability. Springer, New York, 2004.

[13] Goodman, J. B.; Lin, K. K. Coupling control variates for Markov chain Monte Carlo. *J. Comput. Phys.* **228** (2009), no. 19, 7127–7136. doi:10.1016/j.jcp.2009.03.043

[14] Graversen, S. E.; Peskir, G. Maximal inequalities for the Ornstein-Uhlenbeck process. *Proc. Amer. Math. Soc.* **128** (2000), no. 10, 3035–3041. doi:10.1090/S0002-9939-00-05345-4

[15] Hasselmann, K.; Barnett, T.; Bouws, E.; Carlson, H.; Cartwright, D.; Enke, K.; Ewing, J.; Gienapp, H.; Hasselmann, D.; Kruseman, P.; Meerburg, A.; Müller, P.; Olbers, D. J.; Richter, K.;

Sell, W.; Walden, H. Measurements of wind-wave growth and swell decay during the Joint North Sea Wave Project (JONSWAP). Tech. rep., Deutsches Hydrographisches Institut, 1973.

[16] Huang, J. M.; Zhong, J.-Q.; Zhang, J.; Mertz, L. Stochastic dynamics of fluid-structure interaction in turbulent thermal convection. *J. Fluid Mech.* **854** (2018), R5, 13 pp. doi:10.1017/jfm.2018.683

[17] Kameshwar, S.; Padgett, J. E. Storm surge fragility assessment of above ground storage tanks. *Structural Safety* **70** (2018), 48–58. doi:10.1016/j.strusafe.2017.10.002

[18] Karatzas, I.; Shreve, S. E. *Brownian motion and stochastic calculus.* Second edition. Graduate Texts in Mathematics, 113. Springer, New York, 1991. doi:10.1007/978-1-4612-0949-2

[19] Kifer, Y. Stochastic versions of Anosov's and Neistadt's theorems on averaging. *Stoch. Dyn.* **1** (2001), no. 1, 1–21. doi:10.1142/S0219493701000023

[20] Kloeden, P. E.; Platen, E. *Numerical solution of stochastic differential equations.* Applications of Mathematics (New York), 23. Springer, Berlin, 1992. doi:10.1007/978-3-662-12616-5

[21] Kushner, H. J. *Weak convergence methods and singularly perturbed stochastic control and filtering problems.* Systems & Control: Foundations & Applications, 3. Birkhäuser, Boston, 1990. doi:10.1007/978-1-4612-4482-0

[22] Laurière, M.; Mertz, L. Penalization of nonsmooth dynamical systems with noise: ergodicity and asymptotic formulae for threshold crossings probabilities. *SIAM J. Appl. Dyn. Syst.* **18** (2019), no. 2, 853-–880. doi:10.1137/18M1173423

[23] Lin, Y. K.; Yong, Y. Evolutionary Kanai-Tajimi earthquake models. *J. Eng. Mech.* **113** (1987), no. 8, 1119–1137. doi:10.1061/(ASCE)0733-9399(1987)113:8(1119)

[24] Liu, D. Strong convergence of principle of averaging for multiscale stochastic dynamical systems. *Commun. Math. Sci.* **8** (2010), no. 4, 999–1020.

[25] Massart, P. *Concentration inequalities and model selection.* Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003. Lecture Notes in Mathematics, 1896. Springer, Berlin, 2007.

[26] Mertz, L. Stochastic variational inequalities for random mechanics, CIRM, Audiovisual resource. doi:10.24350/CIRM.V.19217703

[27] Mertz, L.; Stadler, G.; Wylie, J. A backward Kolmogorov equation approach to compute means, moments and correlations of non-smooth stochastic dynamical systems. *Physica D* **397** (2019), 25–38.

[28] MIT open courseware. Available at: `https://ocw.mit.edu/courses/physics/8-01sc-classical-mechanics-fall-2016/week-2-newtons-laws/6.1-contact-forces/`; `https://ocw.mit.edu/courses/physics/8-01sc-classical-mechanics-fall-2016/week-2-newtons-laws/6.2-static-friction-lesson/`

[29] Papanicolaou, G. C.; Kohler, W. Asymptotic theory of mixing stochastic ordinary differential equations. *Comm. Pure Appl. Math.* **27** (1974), 641–668. doi:10.1002/cpa.3160270503

[30] Pardoux, E.; Răşcanu, A. *Stochastic differential equations, backward SDEs, partial differential equations.* Stochastic Modelling and Applied Probability, 69. Springer, Cham, 2014. doi:10.1007/978-3-319-05714-9

[31] Patil, A.; Jung, S.; Kown, O.-S. Structural performance of a parked wind turbine tower subjected to strong ground motions. *Eng. Struct.* **120** (2016), no. 1, 92–102. doi:10.1016/j.engstruct.2016.04.020

[32] Pavliotis, G. A.; Stuart, A. M. *Multiscale methods: averaging and homogenization*. Springer, New York, 2008.

[33] Quilligan, A.; O'Connor, A.; Pakrashi, V. Fragility analysis of steel and concrete wind turbine towers. *Eng. Struct.* **36** (2012), 270–282. doi:10.1016/j.engstruct.2011.12.013

[34] Roberts, J. B.; Spanos, P. D. *Random vibration and statistical linearization*. Dover, Mineola, N.Y., 2003.

[35] Vasicek, O. An equilibrium characterisation of the term structure. *J. Financ. Econ.* **5** (1977), no. 2, 177–188. doi:10.1016/0304-405X(77)90016-2

[36] Wong, E.; Zakai, M. On the convergence of ordinary integrals to stochastic integrals. *Ann. Math. Statist.* **36** (1965), 1560–1564. doi:10.1214/aoms/1177699916

JOSSELIN GARNIER
CMAP, CNRS, Ecole Polytechnique
Institut Polytechnique de Paris
Palaiseau
FRANCE
E-mail: josselin.garnier@polytechnique.edu

LAURENT MERTZ
NYU-ECNU Institute of Mathematical Sciences at NYU Shanghai
3663 Zhongshan Road North
Shanghai 200062
P.R. CHINA
E-mail: laurent.mertz@nyu.edu

Received June 2019.