Computationally efficient simulations of stochastically perturbed nonlinear dynamical systems

Thomas Breunung*and Balakumar Balachandran†

Department of Mechanical Engineering, University of Maryland, College Park, Maryland 20742, USA

June 1, 2022

Abstract

A probabilistic approach is needed to address systems with uncertainties arising in natural processes and engineering applications. For computational convenience, however, the stochastic effects are often ignored. Thus, numerical integration routines for stochastic dynamical systems are rudimentary compared to those for the deterministic case. In this work, the authors present a method to carry out stochastic simulations by using methods developed for the deterministic case. Thereby, the well-developed numerical integration routines developed for deterministic systems become available for studies of stochastic systems. The convergence of the developed method is shown and the method's performance is demonstrated through illustrative examples.

1 Introduction

Natural processes are inevitably uncertain, and systems in engineering commonly have uncertainties. To capture this stochasticity, one needs a probabilistic approach. Thus, stochastic models are widely used in physics (e.g., Gardiner [13]), engineering (e.g., Wirsching *et al.* [46]), and biology (e.g., Wilkinson [45]). Uncertain parameters are classically related to manufacturing imperfections, finite measurement resolution, or incomplete data, whereas random loads arise in complex environments due to wind, ocean waves, seismic excitations, or road roughness. However, for computational convenience, stochasticity is often ignored in studies of associated systems. Thus, a variety of methods exist to simulate deterministic systems, while the same is not true for carrying out simulations of stochastic systems. In particular, for high dimensional systems, methods are still in their infancy.

The response of stochastic dynamical systems is fully prescribed by the time evolution of the probability density function (PDF). If the stochasticity is generated by a Wiener process, then the PDF is governed by the Fokker-Planck equation, a partial differential equation (e.g., Risken [32]). Exact solutions have been obtained by, for example, Soize [35], Caughey [7], and Lin and Cai [22]. To obtain these results one requires a balance between dissipative terms and white noise perturbations. This balance, however, is hardly met in realistic systems (Lin and Cai [22]).

In the absence of exact solutions to the Fokker-Planck equation, it can be solved with numerical methods. In principle, space and/or time can be discretized to obtain a finite-dimensional dynamical system. Examples include methods based on finite differences (e.g., Pichler *et al.* [30]), the path integral method (e.g., Wehner and Wolfer [44] or Yu *et al.* [47]) and a finite element discretization formulated by Spencer and Bergman [36]. However, Masud and Bergman [24] point out that the computational burden is so significant, that the applications of those methods are limited to two or three-dimensional systems. In fact, the memory requirements of such methods generally grow exponentially with the dimensions of the considered dynamical system.

*Corresponding author email: thomasbr@umd.edu
†Email: balab@umd.edu

The most common and effective approach to approximate solutions to stochastic differential equations are discrete-time approximations (cf. Kloeden and Platen [20]). With this approach, the computational costs and memory requirements grow only polynomially with respect to number of dimensions. However, for example, as remarked by Rüemelin [33], numerical time integration schemes derived to approximate solutions of deterministic ordinary differential equations do not carry over to the stochastic setting in a straightforward manner. An underlying reason is that additional terms arise when solutions to stochastic differential equations are expanded in a Taylor series. These terms lead to the so-called Wagner-Platen formula (cf. Wagner and Platen [42] or Milstein [26]). Therefore, numerical integration of stochastic differential equations requires fundamentally different algorithms.

Schurz through his chapter in [18], Kloeden and Platen [20], and the introductory treatment by Platen [31] offer an extensive list of available integration schemes. The most prominent and basic ones are the Euler-Maruyama scheme and Milstein's method (e.g. Milstein [26]). Higher order schemes include the stochastic Runge-Kutta in which a predictor-corrector scheme is employed similar to the deterministic equivalent to enhance numerical stability (Schurz in [18]). Moreover, Boyce [4] developed a scheme with adaptive step size control. Milstein and Tret'yakov [25, 27] present integration methods for stochastic differential equations with small noise terms. Such a setting is especially relevant for engineering applications, wherein the stochastic part can often be considered as small or weak. However, the computational expense of such higher-order schemes can be significant (e.g., Schurz in [18]), so much so, that these costs outweigh their benefits (cf. Mannella [23]).

In terms of the current state of the art, the available numerical methods to solve the Fokker-Plank equation suffer from the curse of dimensionality, whereas the discrete-time approximations of stochastic differential equations are often limited to the stochastic equivalent of the forward Euler scheme. However, in many applications, accurate and effective approximations of solutions to the deterministic part already require intricate numerical integration routines employing, for example, adaptive step-size control (cf. Gear [14]), predictor-corrector schemes (cf. de Jalon and Bayo [9]), and customized methods such as Newmark's method for structural dynamics (cf. Géradin and Rixen [15]). Unfortunately, these well-developed routines are not applicable in a stochastic setting.

Here, the authors propose an algorithm to extend any deterministic numerical integrator to a stochastic setting. To that end, they proceed as follows. In many applications, especially in engineering, the noise intensity is small and one can perform a parameter expansion (cf. Section 3). In this setting, the authors formulate a small noise integrator (SNI), which relies on an appropriate resampling of distributions along sample paths (cf. Section 4.1). Within this algorithm, any deterministic integrator can be used to approximate the sample paths of stochastic dynamical systems. After proving the strong convergence of the SNI, the algorithm performance is demonstrated by considering various nonlinear mechanical systems with up to one hundred degrees of freedom in Section 4.2. Furthermore, in Section 5.1, a Gaussian kernel is presented to rigorously deduce a smooth distribution from a finite number of samples and the corresponding performance is shown for nonlinear mechanical systems (cf. Section 5.2).

2 System Setting

The authors consider the general dynamical system

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t) + \sigma \mathbf{b}(\mathbf{x}, t), \qquad 0 < \sigma \ll 1, \quad \mathbf{x} \in \mathbb{R}^N, \tag{1}$$

where the vector $\mathbf{x}$ denotes the states and the integer N is the numbers of degrees of freedom. The parameter σ in equations (1) is small and scales the stochastic excitation defined in differential form as

$$d\mathbf{b}(\mathbf{x}, t) = \mathbf{B}(\mathbf{x}, t) d\mathbf{W} = \sum_{m=1}^{M} \mathbf{B}_m(\mathbf{x}, t) dW_m, \qquad \mathbf{B} : \mathbb{R}^N \times \mathbb{R} \mapsto \mathbb{R}^{N \times M}, \tag{2}$$

where $W_1, W_2, ..., W_M$ denote uncorrelated one dimensional standard Wiener processes (e.g., Kloeden and Platen [20]). The m-th column of the matrix $\mathbf{B}(\mathbf{x}, t)$ prescribes the direction in which the m-th Wiener process acts. The directions can generally depend on the states as indicated in equations (2). Such multiplicative noise commonly arises, for example, in micro-electro-mechanical devices (MEMS, cf. Vig and Yonkee [41]). The parameter σ can be considered to be representative of small uncertainities

that arise in various settings, including high precision manufacturing, measurements, and environmental modeling.

The time dependence of the deterministic part $\mathbf{f}(\mathbf{x},t)$ and the stochastic part $\mathbf{B}(\mathbf{x},t)$ is general. Thus, system (1) includes autonomous (i.e., time-invariant) systems as well as systems with time-periodic, quasi-periodic, and even more general, possibly aperiodic, time dependence. Moreover, it is assumed that the matrix $\mathbf{B}(\mathbf{x},t)$ and the vector $\mathbf{f}(\mathbf{x},t)$ are at least twice continuously differentiable with respect to the spatial coordinate $\mathbf{x}$.

This study is motivated by, but not limited to, the general nonlinear mechanical system

$$\mathbf{M}\ddot{\mathbf{q}} + \mathbf{C}\dot{\mathbf{q}} + \mathbf{K}\mathbf{q} + \mathbf{S}(\mathbf{q},\dot{\mathbf{q}},t) = \sigma\mathbf{g}(\mathbf{q},\dot{\mathbf{q}},t), \quad 0 < \sigma \ll 1, \quad \mathbf{q} \in \mathbb{R}^{N^*}, \quad d\mathbf{g}(\mathbf{q},\dot{\mathbf{q}},t) = \sum_{m=1}^{M} \mathbf{g}_m(\mathbf{q},\dot{\mathbf{q}},t)dW_m, \tag{3}$$

where $\mathbf{q}$ is the vector of coordinates and the integer N^* denotes the numbers of degrees of freedom. The mass $\mathbf{M} \in \mathbb{R}^{N^*\times N^*}$ is assumed to be invertible, whereas the damping $\mathbf{C} \in \mathbb{R}^{N^*\times N^*}$ and stiffness $\mathbf{K} \in \mathbb{R}^{N^*\times N^*}$ matrix are assumed to be general matrices. In many engineering applications the aforementioned matrices, however, are assumed to be positive definite (cf. Géradin and Rixen [15] or Balachandran and Magrab [2]). The vector $\mathbf{S}(\mathbf{q},\dot{\mathbf{q}},t)$ includes nonlinear terms depending on positions and velocities as well as time varying terms such as external and/or parameteric excitation. Introducing the state-space coordinate $\mathbf{x} := \left[\mathbf{q}^\top, \dot{\mathbf{q}}^\top\right]^\top$, the notation

$$\mathbf{A} := \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{M}^{-1}\mathbf{K} & \mathbf{M}^{-1}\mathbf{C} \end{bmatrix} \in \mathbb{R}^{2N^*\times 2N^*}, \quad \mathbf{G}(\mathbf{q},\dot{\mathbf{q}},t) := \begin{bmatrix} \mathbf{0} \\ \mathbf{M}^{-1}\mathbf{S}(\mathbf{q},\dot{\mathbf{q}},t) \end{bmatrix} \in \mathbb{R}^{2N^*}, \tag{4}$$

and defining the columns of the matrix $\mathbf{B}$ as

$$\mathbf{B}_m(\mathbf{q},\dot{\mathbf{q}},t) := \begin{bmatrix} \mathbf{0} \\ \mathbf{M}^{-1}\mathbf{g}_m(\mathbf{q},\dot{\mathbf{q}},t) \end{bmatrix}, \quad m = 1,...,M, \tag{5}$$

the mechanical system (3) can be reformulated in the form of equations (1) with $\mathbf{f}(\mathbf{x},t) = \mathbf{A}\mathbf{x} + \mathbf{G}(\mathbf{x},t)$.

As customary with the treatment of stochastic differential equations, for the forthcoming development, the stochastic dynamical system is most conveniently formulated as the differential equivalent of system (1); that is,

$$d\mathbf{x} = \mathbf{f}(\mathbf{x},t)\ dt + \sigma\mathbf{B}(\mathbf{x},t)\ d\mathbf{W}, \qquad \mathbf{x}(t=t_0) = \mathbf{x}_0, \qquad 0 < \sigma \ll 1, \tag{6}$$

where $\mathbf{x}_0$ denotes a deterministic initial condition.

3 Small Noise Expansion

Similar to approximate solutions obtained through perturbation analysis in the deterministic case (e.g., Verhulst [40] or Nayfeh [28]), approximate solutions to system (6) can be obtained through an expansion in the small parameter σ. The convergence of such a series can be guaranteed by the following result due to Blagoveshchenskii [3]:

Theorem 3.1. *Assume that* $\mathbf{f}(\mathbf{x},t)$ *and* $\mathbf{B}(\mathbf{x},t)$ *and their partial derivatives with respect to the coordinate* $\mathbf{x}$ *up to order two are bounded and Lipschitz continuous for* $t_0 < t < t_1$*. Then, there exists* $\sigma_0 > 0$ *such that*

$$\mathbf{x}(t) = \mathbf{x}^0(t) + \sigma\mathbf{x}^1(t) + \sigma^2\mathbf{r}(t,\sigma), \qquad \textit{for all} \ \ 0 \le \sigma \le \sigma_0, \tag{7}$$

holds, whereby the remainder $\mathbf{r}(t,\sigma)$ *is bounded in the mean square sense; that is,*

$$E\left[\sup_{\substack{0\le\sigma\le\sigma_0 \\ t_0<t<t_1}} |\mathbf{r}(t,\sigma)|^2\right] \le K < \infty. \tag{8}$$

Proof. This is a restatement of a Theorem by Blagoveshchenskii [3] for the current setting of the paper. Related versions have alo been stated by Freidlin and Wentzell [11] and Kaszás and Haller [19]. □

Remark 3.1. Polynomial nonlinearities appearing in common nonlinear oscillator prototypes such as Duffing's equation (cf. Kovacic and Brennan [21]) or the van der Pol oscillator (e.g., Guckenheimer and Holmes [16]) are neither bounded nor globally Lipschitz continuous. These nonlinearities generally grow unbounded for infinite states. Hence, Theorem (3.1) is not applicable in those cases. On the other hand, realistic systems, do not allow for arbitrarily large states. Thus, equations (1) are necessarily only valid for states inside a bounded domain; that is, the model's domain of validity. Thus, for example, one can follow an argument of Breunung [5] and modify equations (1) outside its domain of validity such that the requirements of Theorem 3.1 are met.

After substituting the expansion (7) into equations (6) and equating terms of equal order in σ, the result is the following hierarchy of equations for the first two orders:

$$\mathcal{O}(1): \qquad d\mathbf{x}^0 = \mathbf{f}(\mathbf{x}^0, t)\, dt, \qquad \mathbf{x}^0(t_0) = \mathbf{x}_0, \tag{9}$$

$$\mathcal{O}(\sigma): \qquad d\mathbf{x}^1 = \partial_{\mathbf{x}}\mathbf{f}(\mathbf{x}^0, t)\mathbf{x}^1\, dt + \mathbf{B}(\mathbf{x}^0, t)\, d\mathbf{W}, \qquad \mathbf{x}^1(t_0) = \mathbf{0}, \tag{10}$$

The zeroth order (9) is simply the deterministic limit of equation (6) ($\sigma \to 0$). It can be solved or approximated by standard methods available for deterministic systems such as numerical time integration. Moreover, the stochastic process $\mathbf{x}^0 + \sigma\mathbf{x}^1$ is Gaussian (cf. Blagoveshchenskii [3] or Freidlin and Wentzell [11]) and the linear stochastic differential equation (10) can be solved in closed form. To this end, the flow map; that is, the map mapping the initial condition $\mathbf{x}_0$ at the time $t = t_0$ to their position at time t is denoted by $\mathbf{F}^t_{t_0}(\mathbf{x}_0)$. The gradient of $\mathbf{F}^t_{t_0}(\mathbf{x}_0)$ with respect to the initial condition is denoted by $\mathbf{DF}^t_{t_0}(\mathbf{x}_0)$. The linearized flow map $\mathbf{DF}^t_{t_0}(\mathbf{x}_0)$ is also the fundamental solution to the deterministic part of the first order equations (10) , i.e. it satisfies

$$\frac{d}{dt}\mathbf{DF}^t_{t_0}(\mathbf{x}_0) = \partial_{\mathbf{x}}\mathbf{f}(\mathbf{x}^0(t), t)\mathbf{DF}^t_{t_0}(\mathbf{x}_0), \qquad \mathbf{DF}^{t_0}_{t_0}(\mathbf{x}_0) = \mathbf{I}. \tag{11}$$

With this notation, the first two moments of the stochastic process (10) are given by the explicit formulae

$$\mathbb{E}\left[\mathbf{x}^1(t)\right] = \mathbf{0}, \qquad \mathbb{E}\left[\mathbf{x}^1(t)(\mathbf{x}^1(t))^\top\right] = \int_{t_0}^{t} \mathbf{DF}^s_{t_0}(\mathbf{x}_0)\mathbf{B}(\mathbf{x}^0, s)\left[\mathbf{DF}^s_{t_0}(\mathbf{x}_0)\mathbf{B}(\mathbf{x}^0, s)\right]^\top ds =: \mathbf{\Sigma}(t; \mathbf{x}_0, t_0), \tag{12}$$

which have been obtained by van Kampen [39]. Caughey [7] and Risken [32] present solutions for the special case of a constant Jacobian $\partial_{\mathbf{x}}\mathbf{f}(\mathbf{x}^0, t) = \mathbf{A}_1$ in equations (10).

In the following, the probability density of the stochastic process (10) is obtained. The matrix $\mathbf{\Sigma}(t; \mathbf{x}_0, t_0)$ defined in equation (12) is positive semi-definite. Thus, it admits the decomposition $\mathbf{\Sigma}(t; \mathbf{x}_0, t_0) = \mathbf{U}^\top\mathbf{\Lambda}\mathbf{U}$, where the matrix $\mathbf{U}$ is orthogonal and the matrix $\mathbf{\Lambda}$ is diagonal containing the $L \leq N$ positive eigenvalues of $\mathbf{\Sigma}(t; t_0, \mathbf{x}_0)$. The pseudo inverse of $\mathbf{\Lambda}$ is defined elementwise by $\tilde{\Lambda}^{-1}_{jl} := \delta_{jl}/\Lambda_{ll}$ for $1 \leq j, l \leq L$ and $\tilde{\Lambda}^{-1}_{jl} := 0$ otherwise. With this notation, the pseudo inverse of $\mathbf{\Sigma}(t; t_0, \mathbf{x}_0)$ is given by

$$\tilde{\mathbf{\Sigma}}(t; \mathbf{x}_0, t_0)^{-1} := \mathbf{U}^\top \tilde{\mathbf{\Lambda}}^{-1}\mathbf{U}. \tag{13}$$

Moreover, the pseudo determinant of $\mathbf{\Sigma}(t; t_0, \mathbf{x}_0)$ is defined as

$$|\mathbf{\Sigma}(t; \mathbf{x}_0, t_0)| := \begin{cases} \Lambda_{11}\Lambda_{22}...\Lambda_{LL}, & L \geq 1, \\ 1, & \text{otherwise.} \end{cases} \tag{14}$$

With the definition of the pseudo-inverse (13) and determinant (14), the time varying probability density of the stochastic process (10) is given by

$$p^1(\mathbf{x}, t; \mathbf{x}_0, t_0) = \frac{1}{\sqrt{(2\pi)^L |\mathbf{\Sigma}(t; \mathbf{x}_0, t_0)|}} \exp\left(-\frac{1}{2}\mathbf{x}^\top \tilde{\mathbf{\Sigma}}(t; \mathbf{x}_0, t_0)^{-1}\mathbf{x}\right), \tag{15}$$

which generally depends on the solution to the first order $\mathbf{x}^0$.

In summary, the solutions to system (6) can be approximated as

$$\mathbf{x}(t) = \mathbf{F}^t_{t_0}(\mathbf{x}_0) + \sigma\mathbf{x}^1(t; \mathbf{x}_0, t_0) + \mathcal{O}(\sigma^2), \qquad \mathbf{x}^1(t; \mathbf{x}_0, t_0) \sim p^1(\mathbf{x}, t; \mathbf{x}_0, t_0), \tag{16}$$

where the symbol $\sim$ indicates that the random variable $\mathbf{x}^1(t;\mathbf{x}_0,t_0)$ has the distribution $p^1(\mathbf{x},t;\mathbf{x}_0,t_0)$ (cf. equation (15)). Moreover, the probability density function of $\mathbf{x}^0+\sigma\mathbf{x}^1$; that is, the first order approximation of solutions to system (6), is given by

$$p^\sigma(\mathbf{x},t;\mathbf{x}_0,t_0)=\frac{1}{\sigma\sqrt{(2\pi)^L|\boldsymbol{\Sigma}(t;\mathbf{x}_0,t_0)|}}\exp\left(-\frac{1}{2\sigma^2}(\mathbf{x}-\mathbf{F}_{t_0}^t(\mathbf{x}_0))^\top\tilde{\boldsymbol{\Sigma}}(t;\mathbf{x}_0,t_0)^{-1}(\mathbf{x}-\mathbf{F}_{t_0}^t(\mathbf{x}_0))\right). \quad (17)$$

Thus, the PDF can be approximated by a Gaussian distribution whereby the mean is determined by solution to the deterministic limit $\mathbf{F}_{t_0}^t(\mathbf{x}_0)$ and the variance is given in equation (12). The authors emphasize that both, the mean and variance, are defined through purely deterministic equations and hence, they can be computed without relying on stochastic integration methods.

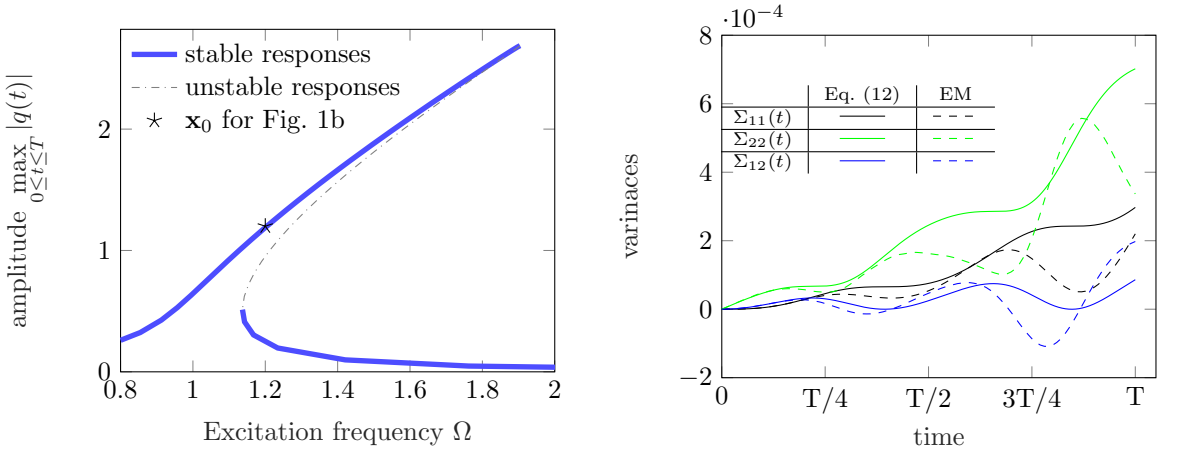
Before continuing with the development, it is revealing to evaluate the validity of approximation (16) based on Theorem 3.1. To this end, the classical Duffing equation

$$\ddot{q}+c\dot{q}+kq+\kappa q^3=a\sin(\Omega t)+\sigma f, \qquad df=dW, \quad (18)$$

with the dimensionless parameters

$$c=0.02, \quad k=1, \quad \kappa=0.5, \quad a=0.1, \quad (19)$$

is considered. First, the deterministic limit ($\sigma\to 0$) is analyzed. Varying the excitation frequency Ω in the vicinity of the natural frequency, the frequency response curve is computed with the numerical continuation package COCO [8] and shown in Figure 1a. For the forcing frequency $\Omega=1.2$, two stable periodic responses exist for system (18). After selecting the initial condition $\mathbf{x}_0$ to be at the orbit with higher amplitude, the variances (12) are computed and depicted in Figure 1b. For comparison, 10^4 approximations of solutions to the stochastic differential equation (18) are calculated with the Euler-Maruyama scheme (e.g., Kloeden and Platen [20]) with the noise intensity $\sigma=0.01$. Subsequently, the variances are calculated from the Euler-Maruyama samples.



(a) Frequency response of the deterministic limit of system (18)

(b) Comparison of the computed variances (12) with variances computed from 10^4 Euler-Maruyama approximations of equation (18) ($\sigma=0.01$).

Figure 1: Deterministic and stochastic solutions to Duffing's equation (18) with parameters (19).

From Figure 1b, it is evident that the approximation (16) is only accurate for about a quarter of the period T. After half a period, the computed variances (12) differ significantly from the results obtained through Monte Carlo sampling. This discrepancy is not an artifact of the obtained numerical approximation, as increasing the sample size or decreasing the step size in the Euler-Maruyama approximation does not alter Figure 1b. Furthermore, Fig. 1b is not in contradiction with Theorem 3.1, according to which, for any final time t_1 there exists some σ_0 such that the series (7) converges and the remainder is bounded in the mean square sense (cf. estimate (8)). It can be said that Figure 1b solely indicates

that for $t_1 = T$ and $\sigma = 0.01$, the series (7) either does not converge or the remainder is unacceptably large. However, for $t_1 = T/4$, the approximation (16) is found to be acceptable. This observation is in line with experience from deterministic dynamical systems. In this setting, Nayfeh [28] has noted that the straightforward expansions in the form of equation (7) are often limited to small time intervals and more sophisticated methods need to be employed to obtain solutions valid over larger time intervals.

4 Approximation of Sample Paths

In the previous section, it was demonstrated that the explicit formula (16) works well for small enough times, but the formula's accuracy deteriorates significantly for longer time spans. To address this shortcoming, the small noise integrator (SNI) is proposed in Section 4.1. Subsequently, the integrator's performance is evaluated on nonlinear mechanical systems with up to one hundred degrees of freedom.

4.1 Small Noise Integrator

To overcome the small time horizon of the approximation (16), the Algorithm 1 is proposed next. After fixing a time step τ and providing an initial condition $\mathbf{x}_0$ at t_0, the distribution (15) for $t = \tau$ is computed and then this distribution is sampled to generate the new initial condition $\mathbf{x}_\tau$ according to equation (16). Then, the algorithm is restarted with the initial condition $\mathbf{x}_\tau$ to generate a new sample $\mathbf{x}_{2\tau}$. By repeating these two steps, samples are generated at later time instances $\mathbf{x}_{k\tau}$.

Algorithm 1: Small noise integrator (SNI)

Result: $\mathbf{x}(K\tau; \mathbf{x}_0, t_0)$
Set time step τ and provide initial condition $\mathbf{x}_0$;
k=1;
while $k < K$ **do**
  Solve equation (9) to obtain the deterministic solution $\mathbf{F}^{k\tau}_{(k-1)\tau}(\mathbf{x}_{k-1})$;
  Compute the variance (12) ;
  Sample the distribution $p^1(\mathbf{x}, k\tau; \mathbf{x}_{k-1}, (k-1)\tau)$ (cf. equation (15)) to obtain $\mathbf{x}^1(k\tau; \mathbf{x}_{k-1}, (k-1)\tau)$;
  $\mathbf{x}_k = \mathbf{F}^{k\tau}_{(k-1)\tau}(\mathbf{x}_{k-1}) + \sigma \mathbf{x}^1(k\tau; \mathbf{x}_{k-1}, (k-1)\tau)$;
  $k = k + 1$;
end

The SNI is motivated by the fact that the small noise expansion is accurate only for a limited time span. Thus, by selecting τ to be within this time span, one can ensure that the approximation (16) is accurate. Repeated sampling removes the limited time horizon of the small noise expansion and the following Lemma guarantees the accuracy of the SNI 1 for longer time spans:

Lemma 4.1. *Assume that the flow map of the deterministic limit* (9) *is approximated with an accuracy not less than* $\mathcal{O}(\tau)$*. Then the SNI 1 strongly converges to the sample paths of the stochastic system* (6) *with order* $\mathcal{O}(\tau^{\frac{1}{2}})$*; that is, the following holds*

$$\mathbb{E}\left[|\mathbf{x}(k\tau) - \mathbf{x}_k^{SNI}|\right] < C\sqrt{\tau}, \qquad 0 \leq k \leq K, \tag{20}$$

where $\mathbf{x}(k\tau)$ *denotes a solution to system* (6)*,* $\mathbf{x}_k^{SNI}$ *is an approximation obtained by the SNI 1, and* C *is a finite constant.*

Proof. The above Lemma 4.1 is proven by showing that for system (6) the SNI 1 converges to the Euler-Maruyama approximation for a small enough time step τ. Then, the convergence follows from well-established convergence results (e.g., Kloeden and Platen [20] or Schurz's chapter in [18]). The details are presented in Appendix A. □

Remark 4.1. Notably, the convergence result of Lemma 4.1 does not depend on the size of the noise intensity σ. More specifically, in Appendix A only a parameter expansion in the time step τ is performed and no restrictions on the noise intensity σ are imposed. For small enough time step τ, the SNI 1 resembles the Euler-Maruyama scheme, which is valid for arbitrarily large σ. Hence, the SNI 1 will also converge for large σ.

Remark 4.2. As utilized in the Euler-Maruyama scheme and also shown by Risken [32], solutions to equation (6) can be approximated by a Gaussian process on small time scales regardless of the size of the noise intensity. With the SNI 1 one arrives at a similar conclusion and exploits the convergence of the expansion (7) for small noise intensities. Within the SNI (1) a non-Gaussian distribution arises on longer time scales when multiple sample paths are computed.

If the directions along which the Gaussian white noise act are constant; that is, $\mathbf{B}_m(\mathbf{x}, t) = \mathbf{B}_m$ for $m = 1, ..., M < N$, one can reduce the computational costs of obtaining the variance (12). Instead of computing the linearized flow map $\mathbf{DF}^t_{t_0}(\mathbf{x}_0)$, which requires N integrations of the equations of variation, the new variable $\mathbf{V}(t) := \mathbf{DF}^t_{t_0}(\mathbf{x}_0)\mathbf{B}$ is introduced. Then, the variance (12) can be reformulated as

$$\mathbf{\Sigma}(t; t_0, \mathbf{x}_0) = \int_{t_0}^{t} \mathbf{V}(t)\mathbf{V}^\top(t)\ ds, \tag{21}$$

and differentiation of $\mathbf{V}(t)$ yields

$$\dot{\mathbf{V}} = \frac{d}{dt}\left(\mathbf{DF}^t_{t_0}(\mathbf{x}_0)\mathbf{B}\right) = \frac{d}{dt}\left(\mathbf{DF}^t_{t_0}(\mathbf{x}_0)\right)\mathbf{B} = \partial_{\mathbf{x}}\mathbf{f}(\mathbf{x}, t)\mathbf{DF}^t_{t_0}(\mathbf{x}_0)\mathbf{B} = \partial_{\mathbf{x}}\mathbf{f}(\mathbf{x}, t)\mathbf{V}, \qquad \mathbf{V}(t_0) = \mathbf{B}. \tag{22}$$

Thus, it suffices to compute the matrix $\mathbf{V}$ which requires only M integrations of the linearized flow.

The fact, that for small times solutions of system (6) can be approximated by a Gaussian process is well-known (see, e.g., Risken [32] or Freidlin and Wentzell [11]) and has also been exploited in the construction of solution schemes that repeatedly resample an arising Gaussian distribution in a similar manner as the SNI 1. For example, Sun and Hsu [37] introduce a short-time Gaussian approximation for the cell mapping method or Yu et al. [47] and Wehner and Wolfer [43] use a Gaussian transition probability density function in the formulation of a path integral solution to the Fokker-Planck equation. In these schemes, however, the mean and variance of the Gaussian approximation are obtained differently from the SNI 1. Wehner and Wolfer [43] utilize the short time propagator also used in the Euler-Maruyama scheme and Sun and Hsu [37] and Yu et al. [47] rely on moment equations with heuristic closure schemes. Contrarily, the SNI 1 utilizes the rigorously deduced mean and variance (16) for the arising Gaussian processes.

For further understanding, the SNI 1 can be compared with the classical Euler-Maruyama scheme which applied to system (6) yields

$$\mathbf{x}^e_k = \mathbf{x}^e_{k-1} + \tau\mathbf{f}(\mathbf{x}^e_{k-1}, (k-1)\tau) + \sigma\sqrt{\tau}\mathbf{B}(\mathbf{x}^e_{k-1}, (k-1)\tau)\Delta\mathbf{W}, \tag{23}$$

where $\Delta\mathbf{W}$ denotes the increments of the m-dimensional of the normalized Wiener process. In this setting, the SNI 1 replaces the forward Euler scheme of the deterministic part in the Euler-Maruyama scheme $(\mathbf{x}^e_{k-1} + \tau\mathbf{f}(\mathbf{x}^e_{k-1}, (k-1)\tau))$ by the deterministic flow map $\mathbf{F}^{k\tau}_{(k-1)\tau}(\mathbf{x}_{k-1})$ at each time step. Moreover, instead of sampling the normal distribution with zero mean and variance $\tau\mathbf{B}(\mathbf{x}^e_{k-1}, (k-1)\tau)\left[\mathbf{B}(\mathbf{x}^e_{k-1}, (k-1)\tau)\right]^\top$ the variance (12) is sampled for the stochastic increment in the SNI 1 at each time step.

Although the order of convergence of the SNI is only one-half, which is the order of the classic Euler-Maruyama scheme, the SNI 1 has inherent advantages making it an appealing alternative to existing techniques to solve equation (6) numerically. First, the SNI requires only sampling of normal distributions and hence it does not require sampling of intricate distributions as some competing method (e.g., Schurz's chapter in [18]). Second, the SNI allows for the use of deterministic integration routines. More specifically, the flow-map $\mathbf{F}^t_{t_0}(\mathbf{x}_0)$ can be approximated by any deterministic numerical integrator. This is especially appealing in applications, where accurate and effective approximations of solutions to the deterministic limit of equation (6) already require sophisticated numerical integration routines. For example, the forward Euler approximation of the deterministic flow map in the Euler-Maruyama scheme or Milstein's scheme (e.g., Kloeden and Platen [20]) can be unstable for simple linear equations (see, e.g., Butcher [6]).

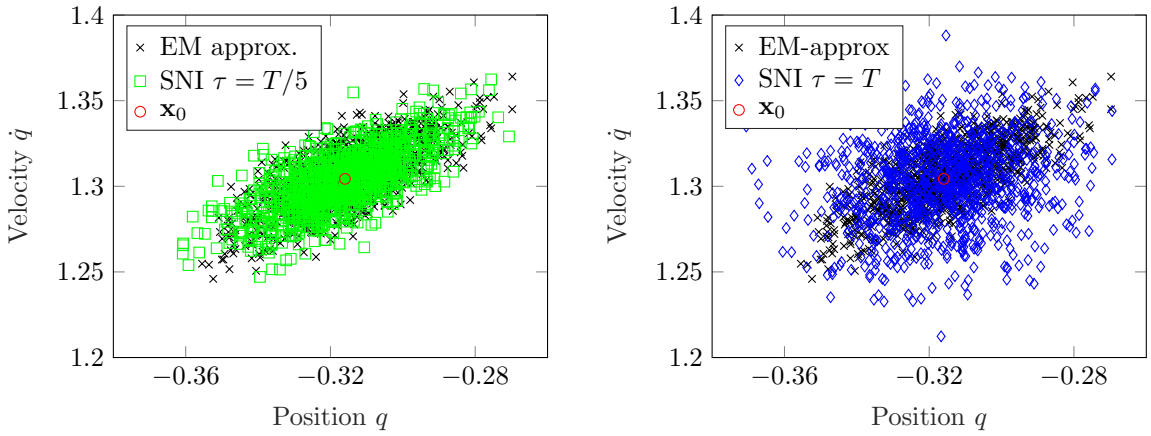
Most importantly, the SNI's performance is drawn from not only the limit of an asymptotically small time step, but this algorithm is also accurate for large time steps, given that the noise terms are sufficiently small. Thus, it is anticipated that the time step τ can be chosen considerably larger than in the available numerical integration routines for stochastic differential equations.

Overall, a possible increase in speed and accuracy gained by using a deterministic integrator for system (1) is counteracted by the necessity of computing the linearized flow map $\mathbf{DF}_{t_0}^t(\mathbf{x}_0)$ and the variance (12). In special cases, this burden can be reduced by computing the matrix $\mathbf{V}$ instead (cf. equation (22)). These observations lead to the conclusion that the SNI 1 is superior to other stochastic integration methods if the deterministic limit is challenging enough such that the speed up gained by the deterministic numerical integration routines is larger than the computational effort of computing the linearized flow map $\mathbf{DF}_{t_0}^t(\mathbf{x}_0)$ and the variance (12). In the next section, it is shown that this can be the case, through many examples.

4.2 Numerical Investigations

As a benchmarking study, the SNI 1 is compared to Euler-Maruyama (EM) approximations of solutions to equation (6) (e.g., Kloeden and Platen [20]). Although many higher-order schemes are available (cf. Schurz's overview in [18]), as reported by Mannella [23], the performance gain can be insignificant at the expense of additional computational expense. Thus, for a first comparison, the widely-used Euler-Maruyama method is used[1] [2]. Within the SNI 1 the following numerical routines are used. Equations (9) and (10), and in turn equation (22), are solved using MATLAB's ODE45 routine. Moreover, the integral to compute the variance (12), respectively equation (21), is approximated by using the trapezoidal rule.

For an initial demonstration of the accuracy of the SNI, the Duffing system (18) with parameter values (19) and excitation frequency $\Omega = 1.2$ is investigated for two different choices of the time step τ. For the first choice, $\tau = T/5$ is selected, and subsequently, for the second choice, it is increased to T for comparison. The results of the SNI are shown in Figure 2. While the distribution obtained with the SNI 1 with time step $\tau = T/5$ shown in Figure 2a matches the distribution from the EM-approximations, the SNI does not yield an accurate approximation for the time step $\tau = T$ (cf. Figure 2b). The accuracy can be quantified by computing the first two statistical moments for all three distributions. In Figure 2a, the mean and standard deviation of the EM-approximation and the SNI agree up to an accuracy of 10^{-3}, whereas the standard deviations differ by more than 0.01 in Figure (2b). With the SNI 1, one computes the 10^3 samples shown in Figure 2 about 20-times faster than the Euler-Maruyama approximation.



(a) SNI-algorithm 1 with $\tau = T/5$

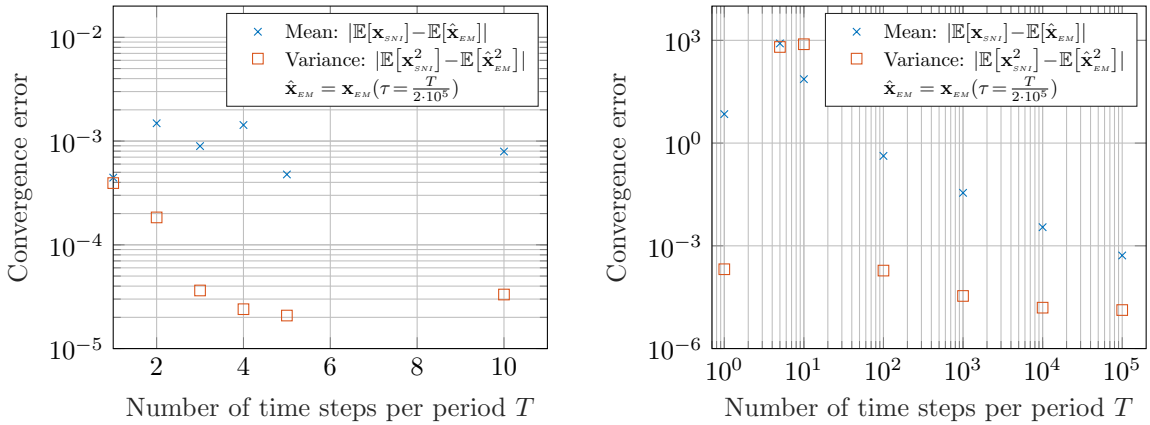
(b) SNI-algorithm 1 with $\tau = T$

Figure 2: Deterministic and stochastic solutions to Duffing's equation (18) with parameters (19).

[1] The MATLAB provided Euler-Maruyama approximation is not utilized. Instead, a self-written code, freely available at an appropriate repository, is employed. For the cases tested the self-written code outperforms the MATLAB provided code. Moreover, various optimizations of the Euler-Maruyama scheme summarized by Higham [17] (especially vectorization) have been tested and a version was selected that seemed optimal in terms of speed and memory requirements for the purposes of this study.

[2] It is noted that for the investigated systems, the first order equivalent of systems (18) and (24), the matrix $\mathbf{B}$ (cf. equation (6)) does not depend on the coordinates. Hence, the Euler-Maruyama scheme is equivalent to Milstein's scheme (e.g., Kloeden and Platen [20]). Thus, it convergences to the sample path of system (24) with order one.

In Figure 3a, the authors show the convergence of the SNI relative to the step size τ. The errors in the first two statistical moments, that is the mean and variance, computed with the SNI 1 are compared to those associated with from an Euler-Maruyama approximation with step size of $\tau = T/(2 \cdot 10^5)$. For the selected step sizes, the mean computed with the SNI 1 remains accurate with an error of the order of about 10^{-3}. This accuracy is also recognizable in Fig. 2. For both step sizes $\tau = T$ and $\tau = T/5$, the sample populations are centered at the same phase space location indicating that both distributions have a similar mean. However, the accuracy of sample variance computed with the SNI 1 increases significantly between $\tau = T/2$ and $\tau = T/3$. This observation is also discernible in Fig. 2. The sample population computed with the SNI 1 for the step size $\tau = T$ spans an evidently larger area than the converged sample distribution (cf. Fig. 2b). The same numerical convergence analysis is also performed for the Euler-Maruyama scheme and shown in Fig. 3a. A converged result from the Euler-Maruyama scheme requires a significantly higher number of time steps per period compared to that for SNI 1.



(a) Convergence of the SNI 1 for varying step size τ

(b) Convergence of the Euler-Maruyama approximation for varying step size τ

Figure 3: Convergence analysis for the Duffing equation (18) with parameters (19).

The performance of the SNI 1 is further demonstrated in a series of numerical experiments on an array of N^* coupled oscillators shown in Fig. 4. The equation of motion of the nth mass is

$$\begin{aligned} m_n \ddot{q}_n + c_n \dot{q}_n + k_n q_n + \kappa_n q_n^3 + s_{n-1}(q_n - q_{n-1}) + s_n(q_n - q_{n+1}) = f_n, \\ f_n = a_n \sin(\Omega t) + \sigma b_n, \quad db_n = g_n \, dW_1, \quad n = 1, ..., N^*, \end{aligned} \tag{24}$$

where $q_0 \equiv q_{N^*+1} \equiv 0$ and $s_0 = s_{N^*} = 0$ holds. The springs with stiffness s_n induce coupling between the individual degrees of freedom of system (24). The quantity a_n is the amplitude of the deterministic harmonic component in the nth forcing, and the quantity g_n is the level of the noise term W_1 in the nth coordinate. It is noted that considering the oscillator array (24) with a single mass ($N^* = 1$) yields the Duffing equation (18) investigated in Figs. (2) and (3). Next, identical parameters for each oscillator are chosen with the following dimensionless values

$$m_n = 1, \quad k_n = 1, \quad s_n = 0.1, \quad c_n = 0.02, \quad \kappa_n = 0.5, \quad a_n = 0.1, \quad g_n = 1, \quad n = 1, .., N^*, \tag{25}$$

and the noise intensity is set to $\sigma = 0.01$.

The excitation frequency Ω is set to 1.2 and for each oscillator, the same initial condition $\mathbf{x}_0$ shown in Figure 1a is selected. Then, the sample paths of the oscillator array (24) are approximated for 1, 10, and 100 forcing periods. To ensure that the Euler approximation yields the correct deterministic limit, the step size of the EM-approximation is decreased until it matches the solution of MATLAB's ODE45 with an error less than 10^{-3}. For the SNI 1, $\tau = T/5$ is selected and MATLAB's ODE45 algorithm is used to solve equations (6) and (22). For both approximations, 10^3 sample paths are obtained and the first two statistical moments are compared. It is observed that in all cases there is an agreement with an accuracy of 10^{-3}. The run-time comparison of the EM-approximation and the SNI 1 is shown in Table 1.

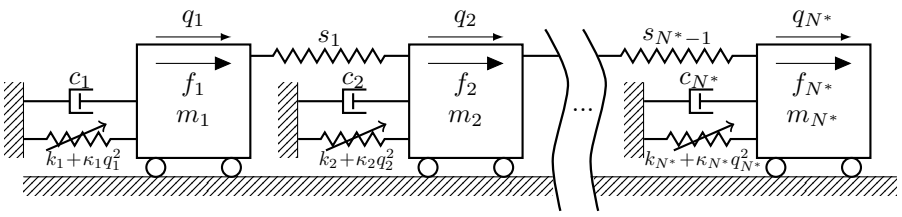


Figure 4: Oscillator array with cubic nonlinearities κ and linear coupling s_n.

Table 1: Run-time comparison for the SNI 1 with $\tau = T/5$ and the Euler-Maruyama approximation for the oscillator array (24). The sample size is 10^3 and the convergence of the two methods is ensured by computing the first two statistical moments. The excitation frequency Ω is 1.2.

	final time $t = T$		final time $t = 10T$		final time $t = 100T$	
# dof (N^*)	EM: Steps per T	Speed-up	EM: Steps per T	Speed-up	EM: Steps per T	Speed-up
1	$4 \cdot 10^4$	**21**	$6 \cdot 10^4$	**32**	$1 \cdot 10^5$	**47**
2	$6 \cdot 10^4$	**32**	$8 \cdot 10^4$	**45**	$2 \cdot 10^5$	**95**
5	$8 \cdot 10^4$	**37**	$1 \cdot 10^5$	**52**	$4 \cdot 10^5$	**194**
10	$1 \cdot 10^5$	**46**	$2 \cdot 10^5$	**195**	$6 \cdot 10^5$	**259**
20	$2 \cdot 10^5$	**72**	$3 \cdot 10^5$	**115**	$6 \cdot 10^5$	**175**[1]
50	$3 \cdot 10^5$	**46**	$4 \cdot 10^5$	**45**	$8 \cdot 10^5$	**102**[1]
100	$4 \cdot 10^5$	**21**	$6 \cdot 10^5$	**32**	$1 \cdot 10^6$	**62**[1]

The computations have been performed using MATLAB 2020a installed on a Windows PC with Intel Xeon CPU E5-2687 W @ 3.1 GHz and 64 GB RAM.

[1] Due to the excessive computation time of the Euler-Maruyama scheme (more than one day), the run-time for the full sample size is estimated by extrapolating the run-time for one sample.

Overall, from Table 1, it can be discerned that a significant speed-up of the SNI 1 is obtained in comparison to the classical Euler-Maruyama scheme. In all cases, the speed-up is at least one order of magnitude. Especially large is the speed-up for oscillator arrays with 5 to 50 oscillators. Moreover, the efficiency of the SNI 1 increases with the time span, which makes it an appealing choice for long time horizons.

To further demonstrate the versatility of the SNI 1, the excitation frequency is increased to $\Omega = 1.9$ and for each oscillator, the initial condition is selected to be on the upper stable branch depicted in Figure 1a. At this frequency, MATLAB's ODE45 algorithm can be used to compute the stable periodic orbit effectively, whereas the step size of the primitive EM-approximation needs to be decreased significantly to converge to a stable periodic orbit of the deterministic limit. Proceeding as previously described for the stochastic simulations, the SNI 1 yields a significant computational gain compared to the Euler-Maruyama scheme (cf. Table 2).

To verify that the performance of the SNI 1 is independent of specific parameter values (25), the parameter values of system (24) are assigned randomly as described in Appendix B. Also for this numerical experiment the SNI 1 is found to outperform the Euler-Maruyama approximation (cf. Table 2). Overall, from Table 2, one can discern a speed-up of about two orders in magnitude. Moreover, for longer simulation times an even higher speed-up of the SNI 1 is expected, since the small step size to be used for the EM-approximation will increase the computational burden excessively.

Although the SNI 1 relies on a repeated sampling of Gaussian distributions, the resulting sample distribution is not necessarily Gaussian. To demonstrate this, the oscillator array (24) with two masses $N^* = 2$ and a coupling spring stiffness $k_c = 0.025$ is considered. For these parameter values, four stable periodic orbits exist. For one of there orbits, both masses oscillate with a high amplitude (i.e., the high amplitude orbit), whereas along the low amplitude orbit, the displacements of both masses are low. Moreover, two *localized modes* exist. Along one of the localized modes, the first mass oscillates considerably, whereas the amplitude of the second mass is low. For the other localized mode, the energy distribution is reversed; that is, the amplitude of the first mass is low and the second mass oscillates with a high amplitude. Such energy localization phenomena have been previously investigated, for example,

Sievers and Takeno [34], Vakakis and Cetinkaya [38] and Dick *et al.* [10]. Recently, there has been an interest to study the effects of noise on such energy localization (e.g., Perkins *et al.* [29] and Balachandran *et al.* [1]). Selecting the high amplitude orbit as an initial condition, 10^3 sample paths for 100 periods have been computed with the SNI 1. The sample locations are shown in Figure 5. The projections show that some samples remained in the neighborhood of the high amplitude orbit, whereas other realizations have escaped towards the low amplitude periodic orbit. Overall, the distribution shown in Figure 5 is clearly non-Gaussian.

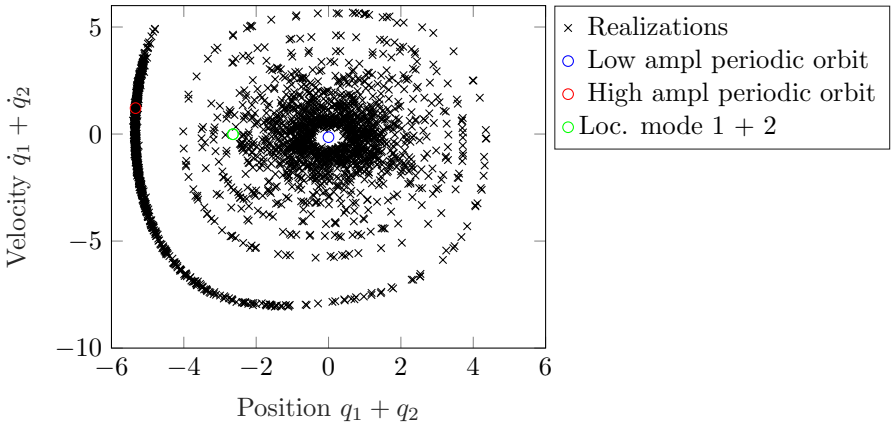


Figure 5: Samples of the oscillator array (24) with localized modes after 100 periods.

5 Approximation of the Probability Density Function

A straightforward method to compute probability densities is to discretize the state space into volumes, compute sample paths (e.g., with the SNI 1 or the EM-approximation) and then count the number of realizations within each volume. The result of this Monte Carlo approach is an invariably non-smooth approximation of the probability density function. This requires a high number of samples to accurately represent the PDF. To overcome these shortcomings a more effective method is proposed in

Table 2: Run-time comparison for the SNI 1 with $\tau = T/5$ and the Euler-Maruyama approximation for the oscillator array (24): i) $\Omega = 1.9$ and uniform parameters (25) and ii) random parameters (cf. also Appendix B). The sample size is 10^3 and the convergence of the two methods is ensured by computing the first two statistical moments.

	$\Omega = 1.9$ final time $t = T$		random parameters final time $t = T$		random parameters final time $t = 10T$	
# dof (N^*)	EM: Steps per T	Speed-up	EM: Steps per T	Speed-up	EM: Steps per T	Speed-up
1	$2 \cdot 10^5$	**87**	$6 \cdot 10^4$	**21**	$3 \cdot 10^5$	**89**
2	$3 \cdot 10^5$	**130**	$2 \cdot 10^4$	**16**	$2 \cdot 10^5$	**161**
5	$5 \cdot 10^5$	**156**	$2 \cdot 10^4$	**13**	$2 \cdot 10^5$	**163**
10	$8 \cdot 10^5$	**303**	$5 \cdot 10^4$	**27**	$5 \cdot 10^5$	**270**
20	$1 \cdot 10^6$	**257**	$5 \cdot 10^5$	**60**	$4 \cdot 10^6$	**488**[1]
50	$2 \cdot 10^6$	**270**	$5 \cdot 10^6$	**22**	$1 \cdot 10^7$	**250**[1]
100	$3 \cdot 10^6$	**153**	$5 \cdot 10^5$	**7**	$5 \cdot 10^6$	**201**[1]

The computations have been performed using MATLAB 2020a installed on a Windows PC with Intel Xeon CPU E5-2687 W @ 3.1 GHz and 64 GB RAM.

[1] Due to the excessive computation time of the Euler-Maruyama scheme (more than one day), the run-time for the full sample size is estimated by extrapolating the run-time for one sample.

Section 5.1. This approach relies on the rigorously deduced Gaussian kernel (12) to smoothly approximate the probability density function. Next, the performance of the proposed method is examined through examples in Section 5.2.

5.1 Gaussian Kernel Approximation

To approximate the probability density function of system (6) at $t = K\tau$, it is assumed that R realizations $\mathbf{x}_r(K\tau) = \mathbf{x}_r(K\tau; \mathbf{x}_0, t_0)$ of the stochastic process (6) have been computed. At time $t = (K-1)\tau$, the sample locations are given by $\mathbf{x}_r((K-1)\tau) = \mathbf{x}_r((K-1)\tau; \mathbf{x}_0, t_0)$. Now advancing the sample population from time $t = (K-1)\tau$ to the final time $t = K\tau$, each individual sample results in the generation of the Gaussian distribution

$$p_r(\mathbf{x}, K\tau; \mathbf{x}_r((K-1)\tau), (K-1)\tau) := p^\sigma(\mathbf{x}, K\tau; \mathbf{x}_r((K-1)\tau), (K-1)\tau), \tag{26}$$

where the distribution p^σ is defined in equation (17). Averaging the distributions (26) over the samples yields

$$p(\mathbf{x}, K\tau; \mathbf{x}_0, t_0) \approx \frac{1}{R} \sum_{r=1}^{R} p_r(\mathbf{x}, K\tau; \mathbf{x}_r((K-1)\tau), (K-1)\tau), \tag{27}$$

that is, a *Gaussian kernel approximation* (GKA) to the probability density at the final time. Equation (27) is an approximation of the probability density $p(\mathbf{x}, K\tau; \mathbf{x}_0, t_0)$ obtained by a sum of Gaussian distributions centered at $\mathbf{F}^{N\tau}_{(N-1)\tau}(\mathbf{x}_r((K-1)\tau))$ with variance (12), which generally differ for each realization r. Hence, it can also be viewed as kernel smoothening method (e.g., Friedman *et al.* [12]). Compared to the usual adhoc chosen kernel distributions, the kernel (27) is rigorously chosen based on the Gaussian kernel (12). Moreover, since (27) is the standard Monte-Carlo estimator, the corresponding convergence can be guaranteed under appropriate assumptions; that is, small enough τ and small enough σ.

5.2 Numerical Investigations

In the following, the time-varying probability density function of the stochastic process (6) is computed with the Gaussian kernel approximation (27) and compared with the probability density obtained by using the straight forward Monte-Carlo simulations, wherein the state space is discretized into volumes and the number of realizations in each volume is counted.

For a first comparison, the single degree of freedom oscillator (18) with the parameters (19) is reinvestigated. The initial condition is depicted in Figure 1a and the solutions of the stochastic process (18) are approximated via the SNI 1 for one period. In Figure 6, the authors show the obtained probability density functions for various sample sizes. For 10^5 samples, the probability density function of both methods, Monte Carlo and GKA (27) match very well. Reducing the sample size in the GKA to include only 10^3 samples, still yields a PDF, which matches the PDF obtained at convergence with 10^5 samples. However, with a reduced number of samples in the Monte-Carlo method, there is a significant deviation from the converged results.

To quantify the superior convergence of the GKA, the L_2-error relative to Monte-Carlo simulations with 10^6 samples is visualized for both approximations and various sample sizes. The relative error of the GKA shown in Figure 7 is about two orders of magnitude smaller than that of the Monte-Carlo approximation. Only for the largest sample size ($N_s = 10^5$), the error of the GKA is one order of magnitude less than that of the Monte-Carlo simulations. However, the error with the Monte-Carlo simulations with 10^5 samples is still larger than the error with the GKA with only 10^3 samples. Thus, in this example, the sample size of the GKA-approximation can be reduced by two orders of magnitude compared to the Monte-Carlo method.

Before proceeding with a high dimensional oscillator array (24), it is insightful to emphasize the enormous sample size needed for the Monte Carlo simulations to obtain a converged PDF. For the one degree-of-freedom oscillator considered in Figures 6 and 7, about 10^5 samples were necessary to compute a converged PDF with the Monte Carlo method. Thus, an estimated number of 10^{5N^*} samples would be required to obtain a converged PDF for an N^*-oscillator array (24). Even for three oscilla-

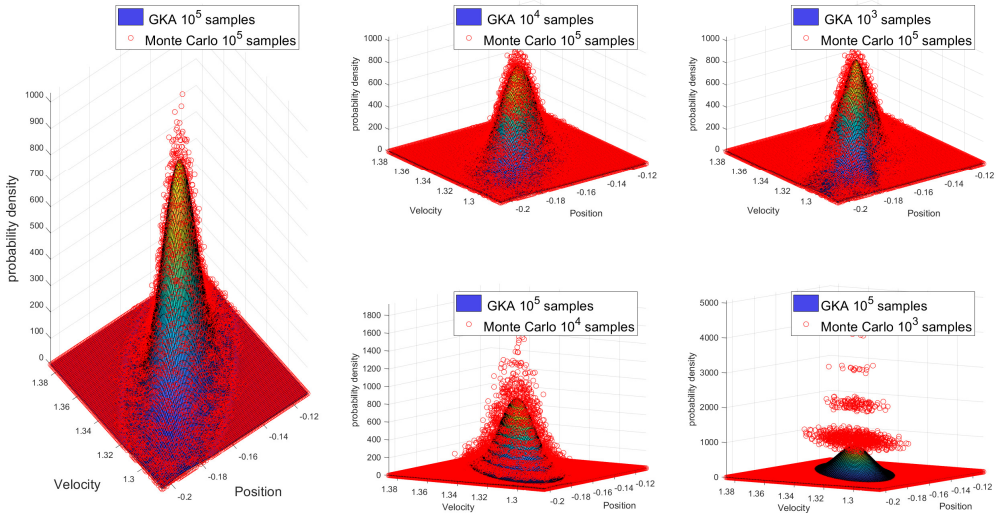


Figure 6: Comparisons of computed probability density function with the proposed Gaussian kernel approximation (27) and Monte-Carlo sampling for different sample sizes for Duffing's equation (18) with parameters (19).

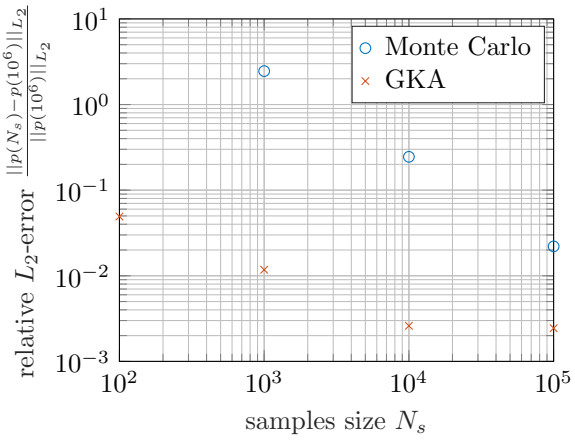


Figure 7: Relative error of Gaussian kernel approximation (27) and Monte Carlo approximation for Duffing's equation (18) with parameters (19).

tors in system (24), the computational effort necessary to generate such a high number of samples is overwhelming[3].

To verify whether one obtains the correct results with the GKA-approximation (27) in higher dimensions, an perturbation scheme is employed. For the weak coupling spring stiffness s_n, the time evolution the oscillator array (24) can be estimated with N uncoupled oscillators; that is, the *uncoupled* limit $s_n \to 0$. More precisely, it is assumed that the coupling spring stiffness s_n is of order $\mathcal{O}(\sigma^2)$. Then, for the uniform parameters (25) each oscillator q_j and $\dot{q}_j$ are identical up to order $\mathcal{O}(\sigma^2)$. Thus, in this limit, it is sufficient to simulate a single degree-of-freedom oscillator to infer about the full oscillator array (24).

[3]To give an idea of the computational effort, the following is mentioned: the storage required to store 10^{15} samples with double precision floating point numbers is 48 petabytes ($= 10^{15} \cdot 3 \cdot 2 \cdot 8$ bytes). With the current advances in high-performance computing and data storage systems, such simulations may not be impossible but one requires a serious commitment in computing power and data storing capabilities, which is beyond the scope of this paper.

In Figure 8, the authors depict the PDF of the oscillator array (24) with three masses $N = 3$ along the two dimensional plane where the positions and velocities of the second and third mass are at their sample mean; that is, $q_2 = \bar{q}_2$, $q_3 = \bar{q}_3$, $\dot{q}_2 = \bar{\dot{q}}_2$ and $\dot{q}_2 = \bar{\dot{q}}_2$, where the bar indicates the mean values. Both approximations match with an relative L_2-error of less than 0.01. It is emphasized, that for the Monte Carlo simulations approximate symmetries of system (24) are explicitly exploited, whereas no such reduction technique has been employed for the GKA[4]. Similar to the single degree-of-freedom case, with the GKA (27), one converges to a significantly lower number of samples than with the Monte-Carlo sampling.

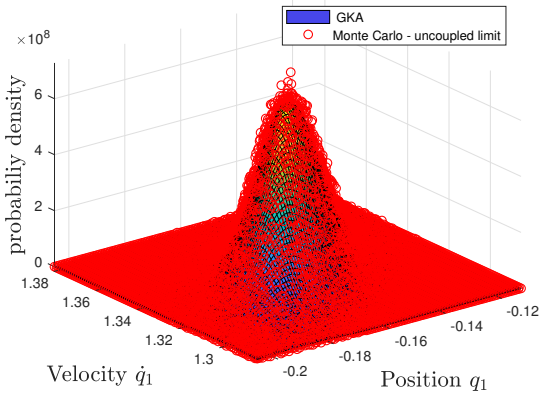


Figure 8: Comparison of the computed probability density function via the proposed Gaussian kernel approximation (GKA) (27) and Monte-Carlo sampling at the uncoupled limit for the oscillator array (24) with parameters (25) and a sample size of 10^5.

Finally, the two-degree-of-freedom system with localized modes from Section 4.2 is reinvestigated to verify whether samples escaped from the high amplitude orbit to the other stable periodic orbits. With the two-dimensional projection shown in Figure 5, one invariably ignores the other directions and hence such projections are of limited use to infer about an escape from the high amplitude orbit. Accordingly, measuring distances of samples to the periodic orbits is more appropriate. For each periodic orbit, the distance is a random variable, whose PDF can be approximated with the GKA. The arising four PDFs are shown in Figure 9 for 10^5 samples. The peak in Figure 9 is due to the initial condition, whose distance to the localized modes and the low amplitude orbit is about 2.7. The distances to the high amplitude orbit and the low amplitude orbit (blue and red lines in Figure 9) confirm the impression from Figure 5 that realizations do not stay close to the high amplitude orbit and escape towards the low energy orbit.

6 Conclusions

In this work, the small noise expansion (cf. equation (7)) is exploited to propose a method to obtain approximations for the following: i) the sample paths of the stochastic dynamical system (1) and ii) the associated time-varying probability density function. With the formulated small noise integrator 1, one removes the limited time horizon of the small noise expansion, by using an appropriate resampling of the Gaussian distribution (12) along sample paths. For the SNI 1, one only requires the solution to deterministic differential equations, which means that deterministic integration routines can be used in the stochastic setting. Additionally, the authors have proven convergence of the proposed algorithm, which notably also holds for arbitrarily large noise intensities.

The computational benefit of the proposed SNI is examined with a series of coupled oscillator arrays with up to one hundred degrees of freedom, including a randomly parameterized array. Compared to

[4]Without using the approximate symmetries, not a single realization out of the 10^5 samples ended up in the discretized state space volume used to compute the results for Figure 8. Hence, the Monte-Carlo approximation for the PDF is identically zero.

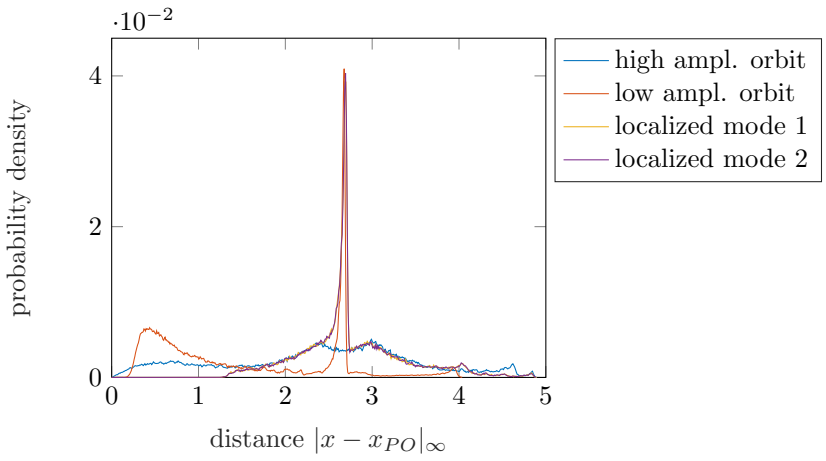


Figure 9: Distances of samples to the four stable periodic orbits of the oscillator array (24) with localized modes after 100 periods.

the standard Euler-Maruyama scheme, a speed-up of two orders of magnitude is observed for a wide range of investigated systems. Especially for longer simulation times, the computational gain of the SNI is significant.

Moreover, the Gaussian kernel approximation (27) yields a justified approach to recover a smooth probability density function from samples, without relying on adhoc kernel choices or interpolation. With this method, the necessary sample size to accurately approximate the probability density function can be reduced drastically. For a single degree-of-freedom oscillator, the number of samples can be reduced by three orders of magnitude compared to established Monte-Carlo methods. For a three degree-of-freedom system, the GKA (27) is found to yield an accurate PDF, while a computation with Monte-Carlo methods is simply infeasible with reasonable computational resources. Thus, the GKA (27) opens up a new horizon to compute PDFs for higher dimensional systems beyond the current limitation to two or three dimensions.

While the SNI 1 can be used to extend the validity of the straightforward expansion (7) to longer time intervals, it does not yield a steady-state distribution, which can be of interest in applications. The underlying theory 3.1, unfortunately, is fundamentally restricted to finite time intervals. Thus, an extension of the underlying theory as well as computational algorithms is desirable to compute a statistical steady state, if it exists.

The probability density function, a time-varying scalar quantity in a usually high dimensional space, is often difficult to understand, visualize or access. In many applications, the quantities of interest, for example, the exceedance probability, escape times, or frequency of occurrence, can, in principle be answered by computing the PDF, but this might not always be the most efficient approach. Thus, it would be of interest to tailor the method presented in such specific contexts.

Acknowledgements. The authors gratefully acknowledge the support of the National Science Foundation through Grant No. CMMI 1760366 and the associated data science supplements. Additionally, T.B. is thankful to Lautaro Clienti and Abdulrahman Alofi for encouraging discussions throughout this work.

Data Availability Statement

The data that support the findings of this study will be openly available in an appropriate repository.

Conflict of Interest

The authors have no conflicts to disclose.

References

[1] B. Balachandran, T. Breunung, G. Acar, A. Abdulrahman, and J. Yorke. Dynamics of circular oscillator arrays subjected to noise. *Nonlinear Dynamics*, pages 1–14, 2022.

[2] B. Balachandran and E. B. Magrab. *Vibrations*. Cambridge University Press, 2018.

[3] Y. N. Blagoveshchenskii. Diffusion processes depending on a small parameter. *Theory of Probability & Its Applications*, 7(2):130–146, 1962.

[4] W. E. Boyce. Approximate solution of random ordinary differential equations. *Advances in Applied Probability*, 10(1):172–184, 1978.

[5] T. Breunung. Existence of quasi-periodic responses in quasi-periodically forced nonlinear mechanical systems. *Nonlinear Dynamics*, 105(3):1977–2004, 2021.

[6] J. C. Butcher. *Numerical methods for ordinary differential equations*. John Wiley & Sons, 2016.

[7] T. Caughey. Nonlinear theory of random vibrations. In *Advances in applied mechanics*, volume 11, pages 209–253. Elsevier, 1971.

[8] H. Dankowicz and F. Schilder. *Recipes for continuation*. SIAM, 2013.

[9] J. G. De Jalon and E. Bayo. *Kinematic and dynamic simulation of multibody systems: the real-time challenge*. Springer Science & Business Media, 2012.

[10] A. Dick, B. Balachandran, and C. Mote. Intrinsic localized modes in microresonator arrays and their relationship to nonlinear vibration modes. *Nonlinear Dynamics*, 54(1):13–29, 2008.

[11] M. I. Freidlin and A. D. Wentzell. Random perturbations. In *Random perturbations of dynamical systems*, pages 15–43. Springer, 1998.

[12] J. Friedman, T. Hastie, R. Tibshirani, et al. *The elements of statistical learning*, volume 1. Springer series in statistics New York, 2001.

[13] C. W. Gardiner. *Handbook of stochastic methods for physics, chemistry and the natural sciences*. Springer series in synergetics 13. Springer, Berlin, 3rd edition edition, 2004.

[14] C. W. Gear. Numerical initial value problems in ordinary differential equations. *Prentice-Hall series in automatic computation*, 1971.

[15] M. Géradin and D. J. Rixen. *Mechanical vibrations: theory and application to structural dynamics*. John Wiley & Sons, 2014.

[16] J. Guckenheimer and P. Holmes. *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*, volume vol. 42, Ed. 2002 of *Applied mathematical sciences*. Springer, New York, corr. 7th printing edition, 2002.

[17] D. J. Higham. An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM review*, 43(3):525–546, 2001.

[18] D. Kannan and V. Lakshmikantham. *Handbook of stochastic analysis and applications*. CRC Press, 2001.

[19] B. Kaszás and G. Haller. Universal upper estimate for prediction errors under moderate model uncertainty. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 30(11):113144, 2020.

[20] P. E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*. Applications of mathematics 23. Springer, Berlin [etc, [3rd corrected printing] edition, 1999.

[21] I. Kovacic and M. J. Brennan. *The Duffing equation: nonlinear oscillators and their behaviour*. John Wiley & Sons, 2011.

[22] Y. Lin and G. Cai. *Probabilistic structural dynamics : advanced theory and applications*. McGraw-Hill engineering reference. McGraw-Hill, New York, 2004.

[23] R. Mannella. Integration of stochastic differential equations on a computer. *International Journal of Modern Physics C*, 13(09):1177–1194, 2002.

[24] A. Masud and L. Bergman. Solution of the four dimensional fokker-planck equation: Still a challenge. In *Icossar*, volume 2005, pages 1911–1916. Citeseer, 2005.

[25] G. Milstein and M. Tret'yakov. Numerical methods in the weak sense for stochastic differential equations with small noise. *SIAM journal on numerical analysis*, 34(6):2142–2167, 1997.

[26] G. N. Milstein. *Numerical integration of stochastic differential equations*, volume 313. Springer Science & Business Media, 1994.

[27] G. N. Milstein and M. V. Tretyakov. *Stochastic numerics for mathematical physics*. Springer Science & Business Media, 2013.

[28] A. H. Nayfeh. *Perturbation methods*. John Wiley & Sons, 2008.

[29] E. Perkins, M. Kimura, T. Hikihara, and B. Balachandran. Effects of noise on symmetric intrinsic localized modes. *Nonlinear Dynamics*, 85(1):333–341, 2016.

[30] L. Pichler, A. Masud, and L. A. Bergman. Numerical solution of the fokker–planck equation by finite difference and finite element methods-a comparative study. In *Computational Methods in Stochastic Dynamics*, pages 69–85. Springer, 2013.

[31] E. Platen. An introduction to numerical methods for stochastic differential equations. *Acta numerica*, 8:197–246, 1999.

[32] H. Risken. Fokker-planck equation. In *The Fokker-Planck Equation*, pages 63–95. Springer, 1996.

[33] W. Rüemelin. Numerical treatment of stochastic differential equations. *SIAM Journal on Numerical Analysis*, 19(3):604–613, 1982.

[34] A. Sievers and S. Takeno. Intrinsic localized modes in anharmonic crystals. *Physical Review Letters*, 61(8):970, 1988.

[35] C. Soize. *The Fokker-Planck equation for stochastic dynamical systems and its explicit steady state solutions*, volume 17. World Scientific, 1994.

[36] B. Spencer and L. Bergman. On the numerical solution of the fokker-planck equation for nonlinear stochastic systems. *Nonlinear Dynamics*, 4(4):357–372, 1993.

[37] J. Q. Sun and C. S. Hsu. The Generalized Cell Mapping Method in Nonlinear Random Vibration Based Upon Short-Time Gaussian Approximation. *Journal of Applied Mechanics*, 57(4):1018–1025, 12 1990.

[38] A. F. Vakais and C. Cetinkaya. Mode localization in a class of multidegree-of-freedom nonlinear systems with cyclic symmetry. *SIAM Journal on Applied Mathematics*, 53(1):265–282, 1993.

[39] N. G. Van Kampen. The expansion of the master equation. *Adv. Chem. Phys*, 34:245–309, 1976.

[40] F. Verhulst. *Methods and applications of singular perturbations*. Springer, 2005.

[41] J. R. Vig and Y. Kim. Noise in microelectromechanical system resonators. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 46(6):1558–1565, 1999.

[42] W. Wagner and E. Platen. *Approximation of Ito integral equations*. Akademie der Wissenschaften der DDR. Zentralinstitut für Mathematik und Mechanik, 1978.

[43] M. Wehner and W. Wolfer. Numerical evaluation of path-integral solutions to fokker-planck equations. iii. time and functionally dependent coefficients. *Physical Review A*, 35(4):1795, 1987.

[44] M. F. Wehner and W. Wolfer. Numerical evaluation of path-integral solutions to fokker-planck equations. *Physical Review A*, 27(5):2663, 1983.

[45] D. J. Wilkinson. Stochastic modelling for quantitative description of heterogeneous biological systems. *Nature Reviews Genetics*, 10(2):122–133, 2009.

[46] P. H. Wirsching, T. L. Paez, and K. Ortiz. *Random vibrations: theory and practice*. Courier Corporation, 2006.

[47] J. Yu, G. Cai, and Y. Lin. A new path integration procedure based on gauss-legendre scheme. *International journal of non-linear mechanics*, 32(4):759–768, 1997.

A Convergence Proof of the SNI 1

In the following, it is shown that the SNI 1 has the same order of convergence as the Euler-Maruyama scheme. The Euler-Maruyama approximation to system (10) with step size τ is given by

$$\mathbf{x}_k^e = \mathbf{x}_{k-1}^e + \tau \mathbf{f}(\mathbf{x}_{k-1}^e, (k-1)\tau) + \sigma\sqrt{\tau}\mathbf{B}(\mathbf{x}_{k-1}^e, (k-1)\tau)\Delta\mathbf{W}, \tag{28}$$

where $\Delta\mathbf{W}$ denotes the increments of the m-dimensional of the normalized Wiener process; that is, they are independent and identically normal-distributed random variables with zero mean and variance one. First, the variance (12) is expanded for small τ yielding

$$\begin{aligned}
&\mathbf{\Sigma}(k\tau; \mathbf{x}_{k-1}, (k-1)\tau) \\
=&\int_{(k-1)\tau}^{k\tau} \mathbf{DF}_{(k-1)\tau}^{s}(\mathbf{x}_{k-1})\mathbf{B}(\mathbf{F}_{(k-1)\tau}^{s}(\mathbf{x}_{k-1}), s)\left[\mathbf{DF}_{(k-1)\tau}^{s}(\mathbf{x}_{k-1})\mathbf{B}(\mathbf{F}_{(k-1)\tau}^{s}(\mathbf{x}_{k-1}), s)\right]^\top ds \\
=&\tau \mathbf{DF}_{(k-1)\tau}^{(k-1)\tau}(\mathbf{x}_{k-1})\mathbf{B}(\mathbf{F}_{(k-1)\tau}^{(k-1)\tau}(\mathbf{x}_{k-1}), (k-1)\tau)\left[\mathbf{DF}_{(k-1)\tau}^{(k-1)\tau}(\mathbf{x}_{k-1})\mathbf{B}(\mathbf{F}_{(k-1)\tau}^{(k-1)\tau}(\mathbf{x}_{k-1}), (k-1)\tau)\right]^\top + \mathcal{O}(\tau^2) \\
=&\tau \mathbf{B}(\mathbf{x}_{k-1}, (k-1)\tau)\left[\mathbf{B}(\mathbf{x}_{k-1}, (k-1)\tau)\right]^\top + \mathcal{O}(\tau^2),
\end{aligned} \tag{29}$$

where the integral is approximated by the values of the integrand at time $s = (k-1)\tau$. Equation (29) reveals that the variance of distribution (15) is $\tau\mathbf{B}(\mathbf{x}_{k-1}, (k-1)\tau)\mathbf{B}^\top(\mathbf{x}_{k-1}, (k-1)\tau)$. The sampling of the normal distribution with variance $\tau\mathbf{B}(\mathbf{x}_{k-1}, (k-1)\tau)\mathbf{B}^\top(\mathbf{x}_{k-1}, (k-1)\tau)$ is equivalent to scaling the M columns of $\sqrt{\tau}\mathbf{B}(\mathbf{x}_{k-1}, (k-1)\tau)$ by samples drawn from the standard distribution. Thus, one time step of the SNI can be written as

$$\mathbf{x}_k^{SNI} = \tilde{\mathbf{F}}_{(k-1)\tau}^{k\tau}(\mathbf{x}_{k-1}^{SNI}) + \sigma\sqrt{\tau}\mathbf{B}(\mathbf{x}_{k-1}^{SNI}, (k-1)\tau)\Delta\mathbf{W} + \mathcal{O}(\tau), \tag{30}$$

where $\tilde{\mathbf{F}}_{n\tau}^{(k-1)\tau}(\mathbf{x}_k)$ denotes an appropriate approximation to the flow map of the deterministic system (9). The error of the SNI has the following upper bound

$$\begin{aligned}
\mathbb{E}\left[|\mathbf{x}(k\tau) - \mathbf{x}_k^{SNI}|\right] &\leq \mathbb{E}\left[|\mathbf{x}(k\tau) - \mathbf{x}_k^e|\right] + \mathbb{E}\left[|\mathbf{x}_k^e - \mathbf{x}_k^{SNI}|\right] \\
&\leq \mathbb{E}\left[|\mathbf{x}(k\tau) - \mathbf{x}_k^e|\right] + |\mathbf{x}_{k-1}^e + \tau\mathbf{f}(\mathbf{x}_{k-1}^e, (k-1)\tau) - \tilde{\mathbf{F}}_{k\tau}^{(k-1)\tau}(\mathbf{x}_{k-1}^{SNI})| + \mathcal{O}(\tau),
\end{aligned} \tag{31}$$

wherein equations (28) and (30) have been used. By adding and subtracting the flow map $\mathbf{F}_{(k-1)\tau}^{k\tau}(\mathbf{x}((k-1)\tau))$, the arising error can be split into three parts

$$\begin{aligned}
\mathbb{E}\left[|\mathbf{x}(k\tau) - \mathbf{x}_k^{SNI}|\right] \leq\ & \underbrace{\mathbb{E}\left[|\mathbf{x}(k\tau) - \mathbf{x}_k^e|\right]}_{\text{error of stochastic Euler approximation}} + \underbrace{|\tilde{\mathbf{F}}_{(k-1)\tau}^{k\tau}(\mathbf{x}_{k-1}^{SNI}) - \mathbf{F}_{k\tau}^{(k-1)\tau}(\mathbf{x}((k-1)\tau))|}_{\text{approximation error of deterministic flow map}} \\
&+ \underbrace{|\mathbf{F}_{(k-1)\tau}^{k\tau}(\mathbf{x}((k-1)\tau)) - (\mathbf{x}_{k-1}^e + \tau\mathbf{f}(\mathbf{x}_{k-1}^e, (k-1)\tau))|}_{\text{error of deterministic explicit Euler scheme}} + \mathcal{O}(\tau) \\
\leq\ & C\sqrt{\tau},
\end{aligned} \tag{32}$$

where the following is used: i) the convergence result of the stochastic Euler-Maruyama approximation (28) to the solutions of equation (10) (e.g., Kloeden and Platen [20]), ii) the assumption that the

approximated flow map $\tilde{\mathbf{F}}_{k\tau}^{(n-1)\tau}(\mathbf{x}_k^{SNI})$ is $\mathcal{O}(\tau)$ close to the true flow map $\tilde{\mathbf{F}}_{k\tau}^{(n-1)\tau}(\mathbf{x}_k^{SNI})$ (cf. assumption in Lemma 4.1), and iii) the convergence of the deterministic forward Euler scheme. The required estimate for Lemma 4.1 is given in equation (31).

B Random Parameter Selection for Oscillator Array (24)

To assign values for the parameters of the oscillator array (24) either the standard distribution $\mathcal{N}(\mu, \sigma_{\mathcal{N}})$ with mean μ and variance $\sigma_{\mathcal{N}}$ or the uniform distribution $\mathcal{U}[a, b]$ with a denoting the minimum and b the maximum value are sampled. The parameter values are drawn from the following distributions:

$$\begin{aligned} m_n &\sim \mathcal{U}[0.5, 2], \quad k_n \sim \mathcal{U}[0.5, 2], \quad s_n \sim \mathcal{U}[0.05, 0.2], \quad c_n \sim \mathcal{U}[0.01, 0.03], \\ \kappa_n &\sim \mathcal{U}[0, 0.1], \quad f_j \sim \mathcal{N}(0, 0.1), \quad g_j \sim \mathcal{N}(0, 1), \quad \Omega \sim \mathcal{U}[1, 2]. \end{aligned} \tag{33}$$

The distributions (33) are selected such that the linear unforced limit of the oscillator array (24) ($\kappa \to 0$ and $\sigma \to 0$) is a weakly damped oscillator customary in the structural dynamics literature (cf. Géradin and Rixen [15]). The parameter values for the numerical experiments presented in Table 2 are freely available from an appropriate repository .