POLYNOMIAL PROPAGATION OF MOMENTS IN STOCHASTIC DIFFERENTIAL EQUATIONS*

ALBERTO LÓPEZ YELA[†] AND JOAQUÍN MÍGUEZ[‡]

Abstract. We address the problem of approximating the moments of the solution, $\boldsymbol{X}(t)$, of an Itô stochastic differential equation (SDE) with drift and diffusion terms over a time-grid $t_0, t_1, \ldots, t_n$. In particular, we assume an explicit numerical scheme for the generation of sample paths $\hat{\boldsymbol{X}}(t_0), \hat{\boldsymbol{X}}(t_1), \ldots, \hat{\boldsymbol{X}}(t_n), \ldots$ and then obtain recursive equations that yield any desired non-central moment of $\hat{\boldsymbol{X}}(t_n)$ as a function of the initial condition $\hat{\boldsymbol{X}}(t_0) = \boldsymbol{X}_0$. The core of the methodology is the decomposition of the numerical solution $\boldsymbol{X}(t_n)$ into a "central part" and an "effective noise" term. The central term is computed deterministically from the ordinary differential equation (ODE) that results from eliminating the diffusion term in the SDE, while the effective noise accounts for the stochastic deviation from the numerical solution of the ODE. For simplicity, we describe the proposed methodology based on an Euler-Maruyama integrator, but other explicit numerical schemes can be exploited in the same way. We also apply the moment approximations to construct estimates of the 1-dimensional marginal probability density functions of $\hat{\boldsymbol{X}}(t_n)$ based on a Gram-Charlier expansion. Both for the approximation of moments and 1-dimensional densities, we describe how to handle the cases in which the initial condition is fixed (i.e., $\boldsymbol{X}_0 = \boldsymbol{x}_0$ for some deterministic and known $\boldsymbol{x}_0$) or random. In the latter case, we resort to polynomial chaos expansion (PCE) schemes in order to approximate the target moments. The methodology has been inspired by the PCE and differential algebra (DA) methods used for uncertainty propagation in astrodynamics problems. Hence, we illustrate its application for the quantification of uncertainty in a 2-dimensional Keplerian orbit perturbed by a Wiener noise process.

Key words. Uncertainty propagation; moment approximation; density estimation; Euler-Maruyama; polynomial chaos expansion; Gram-Charlier expansion

AMS subject classifications. 65C30, 41A58, 41A10

1. Introduction. Let us consider the stochastic differential equation (SDE) in Itô form [18]

$$(1.1) \qquad \begin{cases} \mathrm{d}\boldsymbol{X}(t) &= \boldsymbol{u}(\boldsymbol{X}, t)\mathrm{d}t + \boldsymbol{G}(\boldsymbol{X}, t)\mathrm{d}\boldsymbol{W}(t) \\ \boldsymbol{X}(0) &= \boldsymbol{X}_0, \end{cases}$$

where $t \geqslant 0$ denotes continuous time, $\boldsymbol{X}(t)$ is a real v-dimensional random process representing the solution of the SDE, $\boldsymbol{X}_0$ is a real v-dimensional random variable that describes the initial condition of the process, functions $\boldsymbol{u} : \mathbb{R}^v \times [0, \infty) \to \mathbb{R}^v$ and $\boldsymbol{G} : \mathbb{R}^v \times [0, \infty) \to \mathbb{R}^{v \times d}$ are the the drift coefficient and the diffusion coefficient, respectively, and $\boldsymbol{W}(t)$ is a d-dimensional stochastic process with independent increments.

When $\boldsymbol{W}(t)$ is a Wiener process and the drift and diffusion coefficients satisfy some standard differentiability assumptions, it can be shown that the solution $\boldsymbol{X}(t)$ to Eq. (1.1) can be characterised by a time-varying probability density function (pdf)

†Department of Signal Theory and Communications, Network Systems and Computing. Universidad Rey Juan Carlos. E-mail: alberto.lopez.yela@urjc.es.

‡Department of Signal Theory & Communications, Universidad Carlos III de Madrid. E-mail: joaquin.miguez@uc3m.es.

1

that we denote as $f_{\boldsymbol{X}}(\cdot, t)$ and satisfies the Fokker–Planck equation [24]

$$(1.2) \quad \frac{\partial f_{\boldsymbol{X}}(\boldsymbol{x},t)}{\partial t} + \sum_{k=1}^{v} \frac{\partial}{\partial x^{(k)}}[\mathrm{f}^{(k)}(\boldsymbol{x},t) f_{\boldsymbol{X}}(\boldsymbol{x},t)]$$

$$- \frac{1}{2} \sum_{k=1}^{v} \sum_{j=1}^{v} \frac{\partial^2}{\partial x^{(k)} \partial x^{(j)}}[D^{(k,j)}(\boldsymbol{x},t) f_{\boldsymbol{X}}(\boldsymbol{x},t)] = 0,$$

with initial condition $f_{\boldsymbol{X}}(\boldsymbol{x},0) = f_{\boldsymbol{X}_0}(\boldsymbol{x})$, where $u^{(k)}$, $k = 1, \ldots, v$, are the components of the drift coefficient $\boldsymbol{u}(\boldsymbol{X},t)$ in Eq. (1.1) and $D^{(k,j)}(\boldsymbol{x},t)$ is the entry in the k-th row and j-th column of the diffusion tensor $\boldsymbol{D}(\boldsymbol{x},t) = \boldsymbol{G}(\boldsymbol{x},t)\boldsymbol{G}(\boldsymbol{x},t)^\top$. In principle, we may completely characterise the solution of Eq. (1.1) by solving the partial differential equation (PDE) (1.2). However, this cannot be done exactly except for special (simple) cases [24]. On the other hand, the computational cost of numerical schemes for PDEs, based on finite differences [26] or finite elements [5, 15], quickly becomes prohibitive as the dimensions v and d increase.

Because of the difficulties in solving the Fokker-Planck equation (1.2), most authors have focused on the study of time-discretised numerical schemes to simulate realisations of the random process $\boldsymbol{X}(t)$. Such schemes are extensions of classical algorithms for the numerical solution of ordinary differential equations (ODEs) and they include the classical Euler-Maruyama, Milstein or stochastic Runge-Kutta methods [8, 13], as well as their implicit and semi-implicit variants [27, 17, 32]. When the noise process $\boldsymbol{W}(t)$ is Wiener, the convergence and stability of these numerical algorithms can be studied using a variety of techniques [8, 12, 10], although Taylor approximations have become the standard approach in the past years [13]. Let us remark, however, the fundamental difference between simulating a realisation $\boldsymbol{X}(t) = \boldsymbol{x}(t)$ for a discrete-time grid, $t \in \{t_0, t_1, \ldots, t_N\}$, and the probabilistic characterisation that would be obtained by computing the pdf's $f_{\boldsymbol{X}}(\boldsymbol{x}, t_i)$, even if just approximately. While one can certainly generate many trajectories $\boldsymbol{X}(t) = \boldsymbol{x}_i(t)$, $i = 1, ..., N$, in order to construct a standard Monte Carlo estimator over the grid $t \in \{t_0, t_1, \ldots, t_N\}$, the computational cost of such an approach becomes intractable, again, as the dimension v of the process increases. More sophisticated Monte Carlo methods, specifically designed for high-dimensional systems, exist. For example, [4] applies multi-level Monte Carlo to approximate the probability distribution associated to the solution of a PDE, while the authors of [3] prove the stability of a sequential Monte Carlo sampler as the dimension of a target probability distribution goes to infinity. Efficient methods for Monte Carlo filtering in high-dimensional settings have also been proposed [21, 25]. These techniques are often used to tackle Bayesian inference problems (where data are available for *a posteriori* estimation), and they involve elaborate sampling schemes.

In this paper we introduce a new approach to the probabilistic characterisation of the solution $\boldsymbol{X}(t)$ to the SDE (1.1). Choose a time grid $\{t_0, \ldots, t_n, \ldots\}$, an initial condition $\boldsymbol{X}_0 = \boldsymbol{x}_0$ and let $\hat{\boldsymbol{X}}_n$ be the random sequence generated by the Euler-Maruyama scheme applied to the SDE (1.1). The proposed method builds upon:
   (a) The classical Euler scheme applied to the ordinary differential equation (ODE) $\dot{\boldsymbol{x}} = \boldsymbol{u}(\boldsymbol{x},t)$ with initial condition $\boldsymbol{x}_0$, that yields a deterministic sequence $\hat{\boldsymbol{x}}_n^C \approx \boldsymbol{x}(t_n)$, $n = 0, 1, \ldots$. We refer to this sequence as the *central part* of $\hat{\boldsymbol{X}}_n$.
   (b) The construction of an *effective noise* sequence, denoted $\Delta\hat{\boldsymbol{W}}_n$, that relates the central component and the Euler-Maruyama realisation as $\hat{\boldsymbol{X}}_n = \hat{\boldsymbol{x}}_n^C + $

$\Delta \hat{\boldsymbol{W}}_n$.
Specifically, we show how the moments of the effective noise process $\Delta \hat{\boldsymbol{W}}_n$ can be approximated recursively using a polynomial (Taylor) expansion. The moments of the random vectors $\hat{\boldsymbol{X}}_n$ are then obtained in a straightforward way via the binomial theorem. When the initial condition $\boldsymbol{X}_0$ is random, the method can be combined in a straightforward way with a polynomial chaos expansion (PCE) scheme [31, 16] to account for the initial uncertainty. Finally, we also show how to approximate the marginal pdf of each component $\hat{X}_n^{(k)}$ in the vector $\hat{\boldsymbol{X}}_n = \left(\hat{X}_n^{(1)}, \ldots, \hat{X}_n^{(v)} \right)$ by combining its moment estimates with a Gram–Charlier expansion of type A. The practical performance of the proposed scheme is illustrated with two examples related to astrodynamics, namely the propagation of uncertainty for a Keplerian orbit in two dimensions perturbed by a Wiener process.

While in this manuscript we have restricted the analysis to the Euler and Euler-Maruyama schemes for the sake of clarity, our arguments can be extended to other numerical algorithms.

The rest of the paper is organised as follows. In Section 2, we introduce the methodology and outline the recursive algorithms for the approximation of the moments of $\hat{\boldsymbol{X}}_n$ with both fixed $(\boldsymbol{x}_0)$ and random $(\boldsymbol{X}_0)$ initial condition, as well as the scheme to estimate the marginal pdf's of $\hat{X}_n^{(k)}$, $k = 1, \ldots, v$, from the approximate moments. In Sections 3 and 4 we present the analysis that supports the proposed algorithms. In particular, in Section 3 we establish the convergence of the estimates of the moments of the effective noise terms (when the order of their polynomial approximations increases), while in Section 4 we provide conditions for the convergence of the Gram-Charlier expansion of the marginal pdf's. In Section 5 we apply the proposed numerical schemes to the characterisation of the uncertainty in a 2-dimensional Keplerian orbit perturbed by a Wiener process. Finally, a brief discussion of the theoretical and numerical results is presented in Section 6.

2. The algorithm. In this Section we introduce the proposed algorithms for the approximation of the moments and the 1-dimensional marginal pdf's of the solution of Eq. (1.1) over a time grid. These schemes are the main contribution of the paper. We provide the general argument for their derivation and a summary aimed at facilitating their implementation, but postpone the proof of the key theoretical results to Sections 3 and 4 for clarity. We start with a brief summary of the key notation used in this section (and the rest of the paper).

2.1. Notation. Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is the sample space, $\mathcal{F}$ denotes a σ-algebra of subsets of Ω and $\mathbb{P}$ is a reference probability measure. We denote real random variables (r.v.'s) and random processes (r.p.'s) on $(\Omega, \mathcal{F}, \mathbb{P})$ with capital letters, e.g., X and $X(t)$, respectively, and use lower-case letters to indicate specific realisations. For example, x is a realisation of the r.v. X and $x(t)$ denotes a sample path of $X(t)$, $t \in [0, \infty)$.

Vectors are denoted with bold-face letters while we use regular-face for scalars, e.g., $\boldsymbol{x}$ and x, respectively. For a vector $\boldsymbol{x}$, a superscript $^{(k)}$ indicates the k-th component of the vector, i.e., if $\boldsymbol{x}$ is a v-dimensional vector then $\boldsymbol{x} = \left(x^{(1)}, \ldots, x^{(v)} \right)$. A multi-index $\boldsymbol{r} = \left(r^{(1)}, \ldots, r^{(v)} \right)$ is a vector of non-negative integers, i.e., $x^{(i)} \in \mathbb{N} \cup \{0\}$ for every i. We define the following shorthands for typical operations on multi-indices:

130

$$|\boldsymbol{r}| := \sum_{k=1}^{v} \boldsymbol{r}^{(k)},$$

131

$$\boldsymbol{r}! := \prod_{k=1}^{v} r^{(k)}!,$$

132

$$\frac{\partial^{|\boldsymbol{r}|}}{\partial \boldsymbol{a}^{\boldsymbol{r}}} := \frac{\partial^{|\boldsymbol{r}|}}{\partial a^{(1)^{r^{(1)}}} \cdots \ \partial a^{(v)^{r^{(v)}}}}$$

133

$$\boldsymbol{a}^{\boldsymbol{r}} := \prod_{k=1}^{v} a^{r^{(k)}}$$

134

$$\sum_{\boldsymbol{r}'=\boldsymbol{0}}^{\boldsymbol{r}} := \sum_{r'^{(1)}=0}^{r^{(1)}} \cdots \sum_{r'^{(v)}=0}^{r^{(v)}},$$

135

$$\binom{\boldsymbol{r}}{\boldsymbol{r}'} := \prod_{k=1}^{v} \binom{r^{(k)}}{r'^{(k)}}.$$

136

We adopt the convention $0! = 1$ (hence, $(0,\ldots,0)! = 1$ as well).

2.2. Euler–Maruyama discretisation and the effective noise process.
The discretisation of the SDE (1.1) using the explicit Euler–Maruyama scheme yields

(2.1) $\hat{\boldsymbol{X}}_n = \hat{\boldsymbol{X}}_{n-1} + h\boldsymbol{u}(\hat{\boldsymbol{X}}_{n-1}, t_{n-1}) + \boldsymbol{G}(\hat{\boldsymbol{X}}_{n-1}, t_{n-1})\Delta\boldsymbol{W}_n,$ $n = 1, 2, \ldots,$

where $\hat{\boldsymbol{X}}_n \approx \boldsymbol{X}(t_n)$ is the approximation of the solution at time t_n, with $t_n = t_0 + nh$, the subscript n denotes discrete time, h is the step-size and $\Delta\boldsymbol{W}_n = \boldsymbol{W}(t_n) - \boldsymbol{W}(t_{n-1})$ is the increment of the r.p. $\boldsymbol{W}(t)$ in the interval (t_{n-1}, t_n). The key of the proposed method is to decompose the random sequence $\boldsymbol{X}_n$ into two parts: a *central part*, that results from the integration of an ODE, and an *effective noise* sequence that accounts for the randomness in $\boldsymbol{X}_n$. These two notions are explicitly introduced below.

DEFINITION 2.1. *The random sequence in Eq.* (2.1) *can be written as*

$$\hat{\boldsymbol{X}}_n = \hat{\boldsymbol{x}}_n^C + \Delta\hat{\boldsymbol{W}}_n,$$

where the deterministic sequence $\hat{\boldsymbol{x}}_n^C \approx \boldsymbol{x}(t_n)$ is the central part that results from the explicit Euler integration of the ODE $\dot{\boldsymbol{x}} = \boldsymbol{u}(\boldsymbol{x}, t)$ with a prescribed initial condition $\boldsymbol{x}^C(t_0) = \boldsymbol{x}_0$; specifically

(2.2) $\hat{\boldsymbol{x}}_n^C = \hat{\boldsymbol{x}}_{n-1}^C + h\boldsymbol{u}(\hat{\boldsymbol{x}}_{n-1}^C, t_{n-1}),$ $n \in \mathbb{N},$

and $\Delta\hat{\boldsymbol{W}}_n = \hat{\boldsymbol{X}}_n - \hat{\boldsymbol{x}}_n^C$ is the effective noise r.p.

The central part is easily computed as in Eq. (2.2). However, the characterisation of the effective noise is not straightforward. The gist of our approach is to perform a Taylor expansion of $\Delta\hat{\boldsymbol{W}}_n$ around $\hat{\boldsymbol{x}}_{n-1}^C$ at each time step. Such expansion is convenient because it naturally provides a probabilistic description of the effective noise (and, as a consequence, of the numerical solution $\hat{\boldsymbol{X}}_n$) and it can be carried out recursively over time.

2.3. Polynomial expansion of the effective noise. For the analysis of the effective noise process it is convenient to handle separately the uncertainty in Eq. (2.1) due to the random initial condition $\boldsymbol{X}_0$ and the uncertainty due to the sequence of independent noise increments $\Delta\boldsymbol{W}_n$ (this separation is already implicit in the definition of the effective noise). Consequently, let us first assume that the initial condition is deterministic and fixed, i.e., $\boldsymbol{X}_0 = \boldsymbol{x}_0$. The probability distributions and statistical moments of the r.p.'s $\hat{\boldsymbol{X}}_n$ and $\Delta\hat{\boldsymbol{W}}_n$ can then be computed conditionally on $\boldsymbol{X}_0 = \boldsymbol{x}_0$. In particular, in Section 3, we prove that the polynomial expansion of order N for the effective noise at time n with initial condition $\boldsymbol{X}_0 = \boldsymbol{x}_0$ can be recursively written as

$$\Delta\hat{W}_{n,N}^{(k)}(\boldsymbol{x}_0) = \Delta\hat{W}_{n-1,N}^{(k)}(\boldsymbol{x}_0)$$

$$+h\sum_{|\boldsymbol{r}|=1}^{N}\frac{1}{\boldsymbol{r}!}\frac{\partial^{|\boldsymbol{r}|}}{\partial\boldsymbol{x}_{n-1}^{\boldsymbol{r}}}u^{(k)}\big(\hat{\boldsymbol{x}}_{n-1}^{C}(\boldsymbol{x}_0),t_{n-1}\big)\Delta\hat{\boldsymbol{W}}_{n-1,N}^{\boldsymbol{r}}(\boldsymbol{x}_0)$$

$$(2.3) \qquad +\sum_{|\boldsymbol{r}|=0}^{N-1}\sum_{j=1}^{d}\frac{1}{\boldsymbol{r}!}\frac{\partial^{|\boldsymbol{r}|}}{\partial\boldsymbol{x}_{n-1}^{\boldsymbol{r}}}G^{(k,j)}\big(\hat{\boldsymbol{x}}_{n-1}^{C}(\boldsymbol{x}_0),t_{n-1}\big)\Delta W_n^{(j)}\Delta\hat{\boldsymbol{W}}_{n-1,N}^{\boldsymbol{r}}(\boldsymbol{x}_0),$$

for $k = 1, \dots, v$ and, hence, we denote $\Delta\hat{\boldsymbol{W}}_{n,N}(\boldsymbol{x}_0) = \big(\Delta\hat{W}_{n,N}^{(1)}(\boldsymbol{x}_0), \dots, \Delta\hat{W}_{n,N}^{(v)}(\boldsymbol{x}_0)\big)$. Note that we explicitly indicate the dependance on the initial condition of the central component $\hat{\boldsymbol{x}}_n^{C}(\boldsymbol{x}_0)$ and the effective noise increments $\Delta\hat{\boldsymbol{W}}_{n,N}(\boldsymbol{x}_0)$. Besides, since we have assumed that the initial condition is fixed, then, $\Delta\hat{\boldsymbol{W}}_{0,N}(\boldsymbol{x}_0) = \boldsymbol{0}$. Let us also notice that the multi-index $\boldsymbol{r}$ in the summations is v-dimensional. The subscript N in $\Delta\hat{\boldsymbol{W}}_{n,N}(\boldsymbol{x}_0)$ indicates that we construct a polynomial approximation of order N with no remainder term.

From Eq. (2.3), it is straightforward to obtain the expansion of $\Delta\hat{\boldsymbol{W}}_{n,N}^{\boldsymbol{r}}(\boldsymbol{x}_0)$ (using combinatorics) for any v-dimensional multi-index $\boldsymbol{r}$ such that $|\boldsymbol{r}| > 1$. In particular, the conditional moments of the effective noise truncated to order N can be written as

$$(2.4) \quad \mathbb{E}\big[\Delta\hat{\boldsymbol{W}}_{n,N}^{\boldsymbol{r}}(\boldsymbol{x}_0)\big] = \sum_{|\boldsymbol{s}|+|\boldsymbol{r}'|=1}^{N}a_{\boldsymbol{r},n,N}^{\boldsymbol{s},\boldsymbol{r}'}\big(\hat{\boldsymbol{x}}_{n-1}^{C}(\boldsymbol{x}_0),t_{n-1}\big)\mathbb{E}\big[\Delta\boldsymbol{W}_n^{\boldsymbol{s}}\big]\mathbb{E}\big[\Delta\hat{\boldsymbol{W}}_{n-1,N}^{\boldsymbol{r}'}(\boldsymbol{x}_0)\big],$$

where $a_{\boldsymbol{r},n,N}^{\boldsymbol{s},\boldsymbol{r}'}\big(\hat{\boldsymbol{x}}_{n-1}^{C}(\boldsymbol{x}_0),t_{n-1}\big)$ are the coefficients of the expansion obtained from the polynomial coefficients of Eq. (2.3). Note that the multi-index $\boldsymbol{r}'$ is v-dimensional and the multi-index $\boldsymbol{s}$ is d-dimensional.

To obtain the identity (2.4), we have assumed that the noise increments $\Delta\boldsymbol{W}_n$ form an independent random sequence, which implies that the effective noise $\Delta\hat{\boldsymbol{W}}_m$ is itself independent of $\Delta\boldsymbol{W}_n$ for every $m < n$.

Finally, using the binomial theorem, we arrive at a the formula of the conditional moments of $\hat{\boldsymbol{X}}_n$ given the initial condition $\boldsymbol{X}_0 = \boldsymbol{x}_0$ in terms of the conditional moments of the effective noise in Eq. (2.4) and the central part,

$$(2.5) \qquad \mathbb{E}\big[\hat{\boldsymbol{X}}_{n,N}^{\boldsymbol{r}}(\boldsymbol{x}_0)\big] = \sum_{\boldsymbol{r}'=0}^{\boldsymbol{r}}\binom{\boldsymbol{r}}{\boldsymbol{r}'}\hat{\boldsymbol{x}}_n^{C}(\boldsymbol{x}_0)^{(\boldsymbol{r}-\boldsymbol{r}')}\mathbb{E}\big[\Delta\hat{\boldsymbol{W}}_{n,N}^{\boldsymbol{r}'}(\boldsymbol{x}_0)\big],$$

for any multi-index $\boldsymbol{r}$. Because of the truncation of order N, the approximation is accurate for moments of order $k \leqslant N$. For example, if we choose $N = 1$, the approximation is truncated to order 1 and the polynomial is linear in the noise and hence, not dependent on the second or higher moments of the r.p. $\boldsymbol{W}_n$.

The convergence of the polynomial expansions presented above is rigorously established in Section 3.

2.4. Initial uncertainty. In general, the initial condition for the SDE (1.1) is unknown and $\boldsymbol{X}_0$ is modelled as a random vector with a given probability distribution. It is tempting to handle this uncertainty as an initial effective noise, i.e., to assume that $\Delta\hat{\boldsymbol{W}}_{0,N}(\boldsymbol{x}_0) = \boldsymbol{X}_0$ and $\boldsymbol{x}_0 = \boldsymbol{0}$ in Eqs. (2.3) and (2.4). However, this approach turns out naive. Since Eqs. (2.3) and (2.4) are obtained from a Taylor expansion of $\Delta\hat{\boldsymbol{W}}_{0,N}(\boldsymbol{x}_0)$, when the higher order moments of the effective noise are significant we need to increase the order N of the approximation in order to maintain a prescribed (sufficiently good) accuracy. A larger N implies the computation of higher-order derivatives of functions $\boldsymbol{u}$ and $\boldsymbol{G}$ and, as a consequence, an increased computational effort. In general, the uncertainty of the initial conditions can be expected to be independent of the dynamical perturbation $\boldsymbol{W}(t)$ and, possibly, to have a larger power and more significant higher-order moments compared to the process $\boldsymbol{W}(t)$. For these reasons, it is more convenient to handle the initial uncertainty using a specific expansion of order possibly higher than N.

The polynomial chaos expansion (PCE) method [31, 16] is a technique that provides a polynomial expansion of a r.v. propagated through a deterministic dynamical system. The standard PCE scheme cannot be used in a SDE like Eq. (1.1). However, the argument in Section 2.3 enables us to circumvent this problem, as we have already obtained a deterministic recursion for the moments of the effective noise in Eq. (2.4).

In order to compute a PCE of the conditional moments of $\hat{\boldsymbol{X}}_n$ we take a set of N_p polynomials $\{\Phi_i : \mathbb{R}^v \to \mathbb{R}\}_{i=1}^{N_p}$, selected to be orthogonal with respect to the pdf $f_{\boldsymbol{X}_0}$ of the initial condition $\boldsymbol{X}_0$. Then, we construct the approximation

$$(2.6) \qquad \mathbb{E}\big[\hat{\boldsymbol{X}}_{n,N}^{\boldsymbol{r}}\big|\mathcal{X}_0\big] \approx \sum_{i=1}^{N_p} c_{i,n,N}^{(\boldsymbol{r})}\Phi_i(\boldsymbol{X}_0)\,,$$

where $\mathcal{X}_0$ is the σ-algebra generated by $\boldsymbol{X}_0$ and the $c_{i,n,N}^{(\boldsymbol{r})}$'s are the PCE coefficients (note thet the superscript $^{(\boldsymbol{r})}$ simply indicates dependence on the multi-index $\boldsymbol{r}$ on the left-hand side). A simple way to compute these coefficients is the so-called non-intrusive method [16], for which

$$(2.7) \qquad \Big\{c_{i,n,N}^{(\boldsymbol{r})}\Big\}_{i=1}^{N_p} = \operatorname*{argmin}_{\{c_k\}_{k=1}^{N_p}} \int \left(\mathbb{E}\big[\hat{\boldsymbol{X}}_{n,N}^{\boldsymbol{r}}(\boldsymbol{x}_0)\big] - \sum_{k=1}^{N_p} c_k\Phi_k(\boldsymbol{x}_0)\right)^2 f_{\boldsymbol{X}_0}(\boldsymbol{x}_0)\mathrm{d}\boldsymbol{x}_0\,.$$

While the optimisation problem (2.7) above cannot be solved exactly in general, for most practical applications it is possible to approximate the integral using Monte Carlo. If we draw N_s samples from the pdf $f_{\boldsymbol{X}_0}$, denoted by $\boldsymbol{X}_{0,j}$, $j = 1,\ldots,N_s$, it is straightforward to compute an approximation of the PCE coefficients by solving the linear least-squares problem

$$(2.8) \qquad \Big\{\hat{c}_{i,n,N}^{(\boldsymbol{r})}\Big\}_{i=1}^{N_p} = \operatorname*{argmin}_{\{c_k\}_{k=1}^{N_p}} \sum_{j=1}^{N_s} \left(\mathbb{E}\big[\hat{\boldsymbol{X}}_{n,N}^{\boldsymbol{r}}(\boldsymbol{X}_{0,j})\big] - \sum_{k=1}^{N_p} c_k\Phi_k(\boldsymbol{X}_{0,j})\right)^2,$$

which, in turn, yields the approximate conditional moments

$$(2.9) \qquad \mathbb{E}\big[\hat{\boldsymbol{X}}_{n,N}^{\boldsymbol{r}}\big|\mathcal{X}_0\big] \approx \mathbb{E}\big[\hat{\boldsymbol{X}}_{n,N}^{\boldsymbol{r}}\big|\mathcal{X}_0\big]_{N_p} := \sum_{i=1}^{N_p} \hat{c}_{i,n,N}^{(\boldsymbol{r})}\Phi_i(\boldsymbol{X}_0)\,.$$

Some remarks are in order regarding the validity of Eq. (2.9):

- When the dimension of the state space is v, the PCE approximation is of order N_{PCE} where the number of orthogonal polynomials is [2, Chapter 2]

$$(2.10) \qquad N_p = \binom{N_{\mathrm{PCE}} + v}{v}.$$

- The least-squares problem in (2.8) can be solved when the correlation matrix $\boldsymbol{\Phi}^\top \boldsymbol{\Phi}$ has full rank, where

$$(2.11) \qquad \boldsymbol{\Phi} := \begin{pmatrix} \Phi_1(\boldsymbol{X}_{0,1}) & \cdots & \Phi_{N_p}(\boldsymbol{X}_{0,1}) \\ \vdots & \ddots & \vdots \\ \Phi_1(\boldsymbol{X}_{0,N_s}) & \cdots & \Phi_{N_p}(\boldsymbol{X}_{0,N_s}) \end{pmatrix}.$$

  This implies that $N_s \geqslant N_p$ (in practice, $N_s > N_p$ and sufficiently large). The numerical computation of (2.8) is typically more stable when the polynomials $\{\Phi_i\}_{i=1}^{N_p}$ are orthonormal [7], i.e., when $\|\Phi_i\| = 1$.
- The polynomial expansions (2.6) and (2.9) converge in mean square error (MSE) when the L^2-norm of the conditional moments with respect to the measure $f_{\boldsymbol{X}_0}(\boldsymbol{x}_0)\mathrm{d}\boldsymbol{x}_0$ are finite [6], i.e.,

$$(2.12) \qquad \int \mathbb{E}\big[\hat{\boldsymbol{X}}_{n,N}^{\boldsymbol{r}}(\boldsymbol{x}_0)\big]^2 f_{\boldsymbol{X}_0}(\boldsymbol{x}_0)\mathrm{d}\boldsymbol{x}_0 < \infty.$$

  for the selected multi-index $\boldsymbol{r}$.

Finally, we recall the rule of iterated expectations [30, Theorem 3.4]), which yields

$$(2.13) \qquad \mathbb{E}\big[\hat{\boldsymbol{X}}_{n,N}^{\boldsymbol{r}}\big] = \mathbb{E}\Big[\mathbb{E}\big[\hat{\boldsymbol{X}}_{n,N}^{\boldsymbol{r}}|\mathcal{X}_0\big]\Big] \approx \sum_{i=1}^{N_p} \hat{c}_{i,n,N}^{(\boldsymbol{r})} \mathbb{E}\big[\Phi_i(\boldsymbol{X}_0)\big].$$

When the polynomials Φ_i are orthonormal with respect to $f_{\boldsymbol{X}_0}$ it follows that

$$(2.14) \qquad \int_{\mathbb{R}} \Phi_i(\boldsymbol{x}_0)\Phi_j(\boldsymbol{x}_0)f_{\boldsymbol{X}_0}(\boldsymbol{x}_0)\mathrm{d}\boldsymbol{x}_0 = \delta_{ij} := \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{otherwise,} \end{cases}$$

and in particular,

$$(2.15) \qquad \mathbb{E}\big[\Phi_i(\boldsymbol{X}_0)\big] = \delta_{i1}.$$

Therefore, Eq. (2.13) readily yields

$$(2.16) \qquad \mathbb{E}\big[\hat{\boldsymbol{X}}_{n,N}^{\boldsymbol{r}}\big] \approx \mathbb{E}\big[\hat{\boldsymbol{X}}_{n,N}^{\boldsymbol{r}}\big]_{N_p} = \hat{c}_{1,n,N}^{(\boldsymbol{r})}$$

when the expansion is based on an orthonormal set of polynomials and $\mathbb{E}[\cdot]_{N_p}$ is constructed as in (2.9).

2.5. 1-Dimensional marginal densities. The approximate moments in Eq. (2.16) yield a partial description of the probability distribution of $\hat{\boldsymbol{X}}_n$. However, in many problems, the uncertainty associated to the random sequence $\hat{\boldsymbol{X}}_n$ is easier to interpret in terms of the probability density function (pdf) of the r.v.'s of interest. In

this section, we describe a procedure to approximate the marginal pdf of each variable $\hat{X}_n^{(k)}$ using a Gram–Charlier expansion [14].

We introduce some notation first. Let X be a real r.v. The pdf of X is denoted by f_X, while $\Psi_X(t) \coloneqq \mathbb{E}[e^{iXt}]$ is the characteristic function of X (where i is the imaginary unit and $t \in \mathbb{R}$). The cumulant generating function of X is [14]

$$(2.17) \qquad \log\big(\Psi_X(t)\big) = \sum_{r=1}^{\infty} \frac{\kappa_r(X)}{r!}(it)^r,$$

where κ_r denotes the r-th order cumulant. The cumulants $\kappa_r(X)$ can be computed in terms of the moments of X using lookup tables [14].

The general Gram–Charlier expansion of the marginal pdf of $\hat{X}_{n,N}^{(k)}$ conditional on a fixed initialization $\boldsymbol{X}_0 = \boldsymbol{x}_0$ can be written as

$$(2.18) \qquad f_{\hat{X}_{n,N}^{(k)}|\boldsymbol{x}_0}(x|\boldsymbol{x}_0) \approx \left[1 + \sum_{r=1}^{N} \frac{(-1)^r}{r!} C_r\big[\hat{X}_{n,N}^{(k)}(\boldsymbol{x}_0), Z_\varphi\big] \frac{\mathrm{d}^r}{\mathrm{d}x^r}\right]\varphi(x)$$

for any $x \in \mathbb{R}$ (such that the expansion converges), where

$$(2.19) \quad C_r\big[\hat{X}_{n,N}^{(k)}(\boldsymbol{x}_0), Z_\varphi\big] \coloneqq B_r\Big(\kappa_1\big(\hat{X}_{n,N}^{(k)}(\boldsymbol{x}_0)\big) - \kappa_1(Z_\varphi), \dots$$

$$\dots, \kappa_r\big(\hat{X}_{n,N}^{(k)}(\boldsymbol{x}_0)\big) - \kappa_r(Z_\varphi)\Big),$$

and B_r is the r-th Bell polynomial [1], φ is an auxiliary pdf and Z_φ is a r.v. with density φ.

In the case at hand, we note that for a fixed $\boldsymbol{X}_0 = \boldsymbol{x}_0$ the distribution of the solution $\hat{\boldsymbol{X}}_n(\boldsymbol{x}_0) = \hat{\boldsymbol{x}}_n^C(\boldsymbol{x}_0) + \Delta\hat{\boldsymbol{W}}_n(\boldsymbol{x}_0)$ depends essentially on the distribution of the effective noise $\Delta\hat{\boldsymbol{W}}_n(\boldsymbol{x}_0)$, as $\hat{\boldsymbol{x}}_n^C(\boldsymbol{x}_0)$ is the numerical approximation of the deterministic solution to the ODE $\dot{\boldsymbol{x}}^C(t) = \boldsymbol{u}(\boldsymbol{x}^C, t)$ with initial condition $\boldsymbol{x}^C(t_0) = \boldsymbol{x}_0$. Using Eq. (2.3), we can expand the effective noise in terms of the noise increments $\Delta\boldsymbol{W}_n$. Specifically, if we apply a truncation of order $N = 1$, the k-th effective noise coordinate becomes

$$(2.20) \qquad \Delta\hat{W}_{n,1}^{(k)}(\boldsymbol{x}_0) = \sum_{j=1}^{d}\sum_{m=1}^{n} b_{n,m,j}^{(k)}(\boldsymbol{x}_0)\Delta W_m^{(j)},$$

where the $b_{n,m,j}^{(k)}(\boldsymbol{x}_0)$'s are deterministic coefficients. Hence, $\Delta\hat{W}_{n,1}^{(k)}(\boldsymbol{x}_0)$ is a linear combination of independent r.v.'s. If $\boldsymbol{W}(t)$ is a Wiener process, then $\Delta\hat{W}_{n,1}^{(k)}(\boldsymbol{x}_0)$ is Gaussian and, even for more general processes, recent results on Berry–Essen bounds [9, 11] suggest that a Gaussian approximation for $\Delta\hat{W}_{n,1}^{(k)}(\boldsymbol{x}_0)$ is a plausible choice. Therefore, we let the auxiliary pdf $\varphi(x)$ in Eq. (2.18) be a normal pdf depending on $\boldsymbol{x}_0$, denoted by $\varphi_n^{G^{(k)}}(x|\boldsymbol{x}_0)$ with mean $\mu_n^{(k)}(\boldsymbol{x}_0)$ and standard deviation $\sigma_n^{(k)}(\boldsymbol{x}_0)$. We write $Z_\varphi^{G^{(k)}}(\boldsymbol{x}_0)$ to denote a r.v. with pdf precisely $\varphi_n^{G^{(k)}}(\cdot|\boldsymbol{x}_0)$.

The Gram–Charlier expansion with a Gaussian auxiliary density is well studied and known as Gram–Charlier expansion of type A. In particular, Eq. (2.18) can be

rewritten as [14]

$$(2.21) \quad f_{\hat{X}_{n,N}^{(k)}|\boldsymbol{x}_0}(x|\boldsymbol{x}_0) \approx \left[1 + \sum_{r=1}^{N} \frac{1}{r!\,\sigma_n^{(k)}(\boldsymbol{x}_0)^r} C_r\left[\hat{X}_{n,N}^{(k)}(\boldsymbol{x}_0), Z_\varphi^{G^{(k)}}(\boldsymbol{x}_0)\right]\right.$$

$$\left. \times H_{e_r}\left(\frac{x - \mu_n^{(k)}(\boldsymbol{x}_0)}{\sigma_n^{(k)}(\boldsymbol{x}_0)}\right)\right] \varphi_n^{G^{(k)}}(x|\boldsymbol{x}_0),$$

where $H_{e_r}(x)$ is the r-th Hermite polynomial that satisfies the Rodrigues formula [20]

$$(2.22) \qquad\qquad H_{e_r}(x) = (-1)^r\, \mathrm{e}^{x^2/2}\, \frac{\mathrm{d}^r}{\mathrm{d}x^r}\, \mathrm{e}^{-x^2/2}\,.$$

For simplicity, we propose to compute the mean $\mu_n^{(k)}(\boldsymbol{x}_0)$ and standard deviation $\sigma_n^{(k)}(\boldsymbol{x}_0)$ of the auxiliary Gaussian density $\varphi_n^{G^{(k)}}(x|\boldsymbol{x}_0)$ using the truncations of order $N = 1$ of the effective noise $\Delta\hat{W}_{n,1}^{(k)}(\boldsymbol{x}_0)$. This yields

$$(2.23) \qquad\qquad \mu_n^{(k)}(\boldsymbol{x}_0) = \hat{\boldsymbol{x}}_n^{C^{(k)}}(\boldsymbol{x}_0) + \mathbb{E}\left[\Delta\hat{W}_{n,1}^{(k)}(\boldsymbol{x}_0)\right],$$

and

$$(2.24) \qquad\qquad \sigma_n^{(k)}(\boldsymbol{x}_0) = \sqrt{\mathbb{E}\left[\left(\Delta\hat{W}_{n,1}^{(k)}(\boldsymbol{x}_0)\right)^2\right] - \mathbb{E}\left[\Delta\hat{W}_{n,1}^{(k)}(\boldsymbol{x}_0)\right]^2},$$

for $k = 1,\ldots,v$ where $\mathbb{E}\left[\Delta\hat{W}_{n,1}^{(k)}(\boldsymbol{x}_0)\right]$ and $\mathbb{E}\left[\left(\Delta\hat{W}_{n,1}^{(k)}(\boldsymbol{x}_0)\right)^2\right]$ can be computed recursively from Eq. (2.3).

The convergence of the expansion in Eq. (2.21), i.e., the analysis of the approximation error when the series is truncated to some finite order is addressed in Section 4.

When the initial condition $\boldsymbol{X}_0$ is random, it is possible to construct PCE approximations of $\mu_n^{(k)}$, $\sigma_n^{(k)}$ and C_r in a similar way as we computed the approximations of conditional moments of $\hat{\boldsymbol{X}}_{n,N}^r$ in Section 2.4. In particular,

$$(2.25) \qquad \mu_n^{(k)}(\boldsymbol{X}_0)_{N_p} = \sum_{i=1}^{N_p} \hat{c}_{i,n}^{\mu(k)} \Phi_i(\boldsymbol{X}_0), \qquad \sigma_n^{(k)}(\boldsymbol{X}_0)_{N_p} = \sum_{i=1}^{N_p} \hat{c}_{i,n}^{\sigma(k)} \Phi_i(\boldsymbol{X}_0),$$

and

$$(2.26) \qquad\qquad C_r\left[\hat{X}_{n,N}^{(k)}(\boldsymbol{X}_0), Z_\varphi^{G^{(k)}}(\boldsymbol{X}_0)\right]_{N_p} = \sum_{i=1}^{N_p} \hat{c}_{i,n}^{C_r(k)} \Phi_i(\boldsymbol{X}_0),$$

where, the same as in Section 2.4, the coefficients of the expansion are obtained by solving the least-squares problem

$$(2.27) \quad \left\{\hat{c}_{i,n}^{[s](k)}\right\}_{i=1}^{N_p} = \operatorname*{argmin}_{\{c_k\}_{k=1}^{N_p}} \sum_{j=1}^{N_s} \left(u_n^{[s](k)}(\boldsymbol{X}_{0,j}) - \sum_{k=1}^{N_p} c_k \Phi_k(\boldsymbol{X}_{0,j})\right)^2, \qquad s = 1, 2, 3,$$

where
- $\hat{c}_{i,n}^{[1](k)} = \hat{c}_{i,n}^{\mu(k)}$ and $u_n^{[1](k)} = \mu_n^{(k)}$;

- $\hat{c}_{i,n}^{[2](k)} = \hat{c}_{i,n}^{\sigma(k)}$ and $u_n^{[2](k)} = \sigma_n^{(k)}$;
- $\hat{c}_{i,n}^{[3](k)} = \hat{c}_{i,n}^{C_r}$ and $u_n^{[3](k)} = C_r\left[\hat{X}_{n,N}^{(k)}, Z_\varphi^{G^{(k)}}\right]$.

Finally, if we draw N_s' i.i.d. samples from the random initial condition $\boldsymbol{X}_0$, denoted $\boldsymbol{X}_{0,j}'$, $j = 1, \ldots, N_s'$, then we can use Eq. (2.21) to approximate the pdf of $\hat{X}_{n,N}^{(k)}$ as

$$(2.28) \qquad f_{\hat{X}_{n,N}^{(k)}}(x) = \int_{\mathbb{R}^v} f_{\hat{X}_{n,N}^{(k)}|\boldsymbol{X}_0}(x|\boldsymbol{x}_0) f_{\boldsymbol{X}_0}(\boldsymbol{x}_0)\mathrm{d}\boldsymbol{x}_0 \approx \frac{1}{N_s'}\sum_{j=1}^{N_s'} f_{\hat{X}_{n,N}^{(k)}|\boldsymbol{X}_0}(x|\boldsymbol{X}_{0,j}').$$

2.6. Outline of the algorithms. In this section we provide a summary of the proposed algorithms for the approximate computation of the moments $\mathbb{E}\big[\hat{\boldsymbol{X}}_{n,N}^{\boldsymbol{r}}\big]$ and the 1-dimensional marginal densities $f_{\hat{X}_{n,N}^{(k)}}(x)$, $k = 1, \ldots, v$.

Table 2.1 provides a list, with brief descriptions, of the inputs and outputs of the two proposed approximation schemes. Algorithm 2.1 displays a pseudocode, with cross-references to Sections 2.3 and 2.4, of the numerical scheme for the computation of moments assuming a random initial condition $\boldsymbol{X}_0$. If $\boldsymbol{X}_0 = \boldsymbol{x}_0$ the algorithm is simply run with $N_s = 1$. Algorithm 2.2 shows a pseudocode for the approximation of marginal densities, with cross-references to Section 2.5.

Inputs	Description
h	Step-size.
t_0	Initial time.
t_n	Final time.
N	Order of the polynomial expansions.
v	Dimension of $\boldsymbol{X}(t)$.
d	Dimension of $\boldsymbol{W}(t)$.
$\boldsymbol{X}_0$	Initial condition.
$\mathbb{E}\big[\Delta\boldsymbol{W}_m^{\boldsymbol{r}}\big]$	Moments of the noise increments for $m \geqslant 1$ and $|\boldsymbol{r}| \leqslant N$.
N_{PCE}	Truncation order of the PCE scheme (when $\boldsymbol{X}_0 = \boldsymbol{x}_0$ is fixed, this is not needed).
N_s	Number of samples (if $\boldsymbol{X}_0 = \boldsymbol{x}_0$ is fixed, $N_s = 1$).
N_s'	Number of i.i.d. samples of $\boldsymbol{X}_0$ to approximate the pdf $f_{\hat{X}_{n,N}^{(k)}}(x)$ in Eq. (2.28).
$\boldsymbol{u}$	Drift coefficient in Eq. (1.1).
$\boldsymbol{G}$	Diffusion coefficient in Eq. (1.1).

Outputs	Description
$\mathbb{E}\big[\hat{\boldsymbol{X}}_{n,N}^{\boldsymbol{r}}\big]_{N_p}$	Moments of the numerical solution of Eq. (1.1), for $|\boldsymbol{r}| \leqslant N$, computed with a basis of N_p orthogonal polynomials, where N_p is given by Eq. (2.10).
$f_{\hat{X}_{n,N}^{(k)}}$	Estimate of the pdf $f_{\hat{X}_n}$ where $\hat{\boldsymbol{X}}_n$ is the numerical approximation of the r.v. $\boldsymbol{X}(t_n)$

Table 2.1: Inputs and outputs of the algorithms for moment computation and estimation of 1-dimensional marginal pdf's.

Algorithm 2.1 Computation of moments

1: Generate N_s samples of $\boldsymbol{X}_0$, denoted $\boldsymbol{X}_{0,j}$, $j = 1, \ldots, N_s$.
2: Compute N_p using Eq. (2.10) and matrix $\boldsymbol{\Phi}$ using Eq. (2.11)) such that $\boldsymbol{\Phi}^\top \boldsymbol{\Phi}$. We assume $\boldsymbol{\Phi}$ is full-rank.
3: Set $\hat{\boldsymbol{x}}_0^C(\boldsymbol{X}_{0,j}) = \boldsymbol{X}_{0,j}$, $\Delta \hat{\boldsymbol{W}}_{0,N}(\boldsymbol{X}_{0,j}) = 0$ and $\Delta \hat{\boldsymbol{W}}_{0,1}(\boldsymbol{X}_{0,j}) = 0$ for $j = 1, \ldots, N_s$.

4: Set $n = \lceil (t_n - t_0)/h \rceil$, where $\lceil \cdot \rceil$ denotes the ceiling function.
5: **for** $m = 1, \ldots, n$ **do**
6:   **for** $j = 1, \ldots, N_s$ **do**
7:     Evaluate the central part $\hat{\boldsymbol{x}}_m^C(\boldsymbol{X}_{0,j}) = \hat{\boldsymbol{x}}_{m-1}^C(\boldsymbol{X}_{0,j}) + h\boldsymbol{u}(\hat{\boldsymbol{x}}_{m-1}^C(\boldsymbol{X}_{0,j}), t_{m-1})$, where $t_m = t_0 + mh$.
8:     Evaluate $\mathbb{E}[\Delta \hat{\boldsymbol{W}}_{m,N}^{\boldsymbol{r}}(\boldsymbol{X}_{0,j})]$ for $1 \leqslant |\boldsymbol{r}| \leqslant N$ using Eq. (2.4).
9:     Evaluate $\mathbb{E}[\Delta \hat{W}_{m,1}^{(k)}(\boldsymbol{X}_{0,j})]$ and $\mathbb{E}[(\Delta \hat{W}_{m,1}^{(k)}(\boldsymbol{X}_{0,j}))^2]$ for $k = 1, \ldots, v$.
10:   **end for**
11: **end for**
12: Compute $\mathbb{E}[\hat{\boldsymbol{X}}_{n,N}^{\boldsymbol{r}}(\boldsymbol{X}_{0,j})]$ for $j = 1, \ldots, N_s$ using Eq. (2.5).
13: Solve the least-squares problem (2.8) and compute $\mathbb{E}[\hat{\boldsymbol{X}}_{n,N}^{\boldsymbol{r}}]_{N_p}$ for $1 \leqslant |\boldsymbol{r}| \leqslant N$ with Eq. (2.16).

Algorithm 2.2 Computation of 1-dimensional marginal pdf's

1: Generate N_s samples of $\boldsymbol{X}_0$, denoted $\boldsymbol{X}_{0,j}$, $j = 1, \ldots, N_s$.
2: **for** $j = 1, \ldots, N_s$ **do**
3:   **for** $k = 1, \ldots, v$ **do**
4:     Compute $\mu_n^{(k)}(\boldsymbol{X}_{0,j})$ and $\sigma_n^{(k)}(\boldsymbol{X}_{0,j})$ using Eqs. (2.23) and (2.24) respectively.
5:     Compute $C_r[\hat{X}_{n,N}^{(k)}(\boldsymbol{X}_{0,j}), Z_\varphi^{G^{(k)}}(\boldsymbol{X}_{0,j})]$ for $r = 1, \ldots, N$.
6:   **end for**
7: **end for**
8: Solve the least-squares problem to compute the PCE coefficients of Eqs. (2.25) and (2.26).
9: Generate N_s' samples of $\boldsymbol{X}_0$, denoted $\boldsymbol{X}_{0,j}'$, $j = 1, \ldots, N_s'$.
10: **for** $j = 1, \ldots, N_s'$ **do**
11:   **for** $k = 1, \ldots, v$ **do**
12:     Compute $\mu_n^{(k)}(\boldsymbol{X}_{0,j}')$ and $\sigma_n^{(k)}(\boldsymbol{X}_{0,j}')$ using Eq. (2.25).
13:     Compute $C_r[\hat{X}_{n,N}^{(k)}(\boldsymbol{X}_{0,j}'), Z_\varphi^{G^{(k)}}(\boldsymbol{X}_{0,j}')]$, for $r = 1, \ldots, N$, using Eq. (2.26).
14:   **end for**
15: **end for**
16: Compute the coefficients of the Hermite polynomials H_{e_r} for $r = 0, \ldots, N$.
17: Apply Eq. (2.28), combined with Eq. (2.21), to compute $f_{\hat{X}_{n,N}^{(k)}}(x)$ for any $x \in \mathbb{R}$ and $k = 1, \ldots, v$.

---

3. Variational solution based on a polynomial expansion over the noise process. In this section we provide the analysis needed to support the results in Section 2.3 and, specifically, Algorithm 2.1 for the approximate computation of moment of the random sequence $\hat{\boldsymbol{X}}_n$. Our analysis relies on the notion of convergence

region for a Taylor expansion as defined below.

DEFINITION 3.1. *Let* **g** *be a smooth function,* $\mathbf{g} : \mathbb{R}^v \times [0, \infty) \to \mathbb{R}$. *The convergence region of the Taylor expansion of* g, *centred around* $\boldsymbol{x}_0 \in \mathbb{R}^v$ *at time* $t \in [0, \infty)$, *is the set*

$$\rho_{\mathrm{g}}(\boldsymbol{x}_0, t) := \left\{ \boldsymbol{x} \in \mathbb{R}^v : \sum_{|\boldsymbol{r}|=0}^{\infty} \frac{1}{\boldsymbol{r}!} \frac{\partial^{|\boldsymbol{r}|}}{\partial \boldsymbol{x}^{\boldsymbol{r}}} \mathrm{g}(\boldsymbol{x}_0, t) \boldsymbol{x}^{\boldsymbol{r}} < \infty \right\} \subseteq \mathbb{R}^v.$$

Let us assume a fixed initial condition $\boldsymbol{X}_0 = \boldsymbol{x}_0$. The moments of the sequence $\hat{X}_n$ follow readily from the statistics of the effective noise sequence $\Delta \hat{W}_n^{(k)}(\boldsymbol{x}_0)$. Therefore, we start with the expansion formula for the effective noise in Eq. (2.3).

THEOREM 3.2. *Assume that the functions* $\boldsymbol{u}$ *and* $\boldsymbol{G}$ *in Eq.* (1.1) *are real and smooth. For any positive integers* N *and* n, *the effective noise given an initial condition* $\boldsymbol{X}_0 = \boldsymbol{x}_0$ *can be written as*

$$\Delta \hat{W}_n^{(k)}(\boldsymbol{x}_0) = \Delta \hat{W}_{n-1}^{(k)}(\boldsymbol{x}_0) + h \sum_{|\boldsymbol{r}|=1}^{N} \frac{1}{\boldsymbol{r}!} \frac{\partial^{|\boldsymbol{r}|}}{\partial \boldsymbol{x}_{n-1}^{\boldsymbol{r}}} u^{(k)}\big(\hat{\boldsymbol{x}}_{n-1}^C(\boldsymbol{x}_0), t_{n-1}\big) \Delta \hat{\boldsymbol{W}}_{n-1}^{\boldsymbol{r}}(\boldsymbol{x}_0)$$

$$+ \sum_{|\boldsymbol{r}|=0}^{N-1} \sum_{j=1}^{d} \frac{1}{\boldsymbol{r}!} \frac{\partial^{|\boldsymbol{r}|}}{\partial \boldsymbol{x}_{n-1}^{\boldsymbol{r}}} G^{(k,j)}\big(\hat{\boldsymbol{x}}_{n-1}^C(\boldsymbol{x}_0), t_{n-1}\big) \Delta W_n^{(j)} \Delta \hat{\boldsymbol{W}}_{n-1}^{\boldsymbol{r}}(\boldsymbol{x}_0)$$

$$(3.1) \qquad + R_{n,N}^{(k)}\big(\Delta \boldsymbol{W}_n, \Delta \hat{\boldsymbol{W}}_{n-1}(\boldsymbol{x}_0)\big),$$

where $\Delta \hat{\boldsymbol{W}}_0(\boldsymbol{x}_0) = 0$ *and* $R_{n,N}^{(k)}$ *is the remainder term of the polynomial expansion at step* n *with truncation order* N. *If*

$$\Delta \hat{\boldsymbol{W}}_{n-1}(\boldsymbol{x}_0) \in \rho_{u^{(k)}}\big(\hat{\boldsymbol{x}}_{n-1}^C(\boldsymbol{x}_0), t_{n-1}\big) \cap_{j=1}^{d} \rho_{G^{(k,j)}}\big(\hat{\boldsymbol{x}}_{n-1}^C(\boldsymbol{x}_0), t_{n-1}\big),$$

then

$$\lim_{N \to \infty} R_{n,N}^{(k)}\big(\Delta \boldsymbol{W}_n, \Delta \hat{\boldsymbol{W}}_{n-1}(\boldsymbol{x}_0)\big) = 0.$$

Remark 3.3. Note that $\frac{\Delta \hat{\boldsymbol{W}}_{n-1}^{\boldsymbol{r}}(\boldsymbol{x}_0)}{\boldsymbol{r}!} = 1$ for $|\boldsymbol{r}| = 0$.

PROOF: Recall the decomposition of the sequence $\hat{\boldsymbol{X}}_n(\boldsymbol{x}_0)$ into its central part and the effective noise,

$$(3.2) \qquad \hat{\boldsymbol{X}}_{n-1}(\boldsymbol{x}_0) = \hat{\boldsymbol{x}}_{n-1}^C(\boldsymbol{x}_0) + \Delta \hat{\boldsymbol{W}}_{n-1}(\boldsymbol{x}_0).$$

Using the relationship above, the the Taylor expansions of $u^{(k)}$ (of order N) and $G^{(k,j)}$ (of order $N - 1$) with respect to $\hat{\boldsymbol{X}}_{n-1}(\boldsymbol{x}_0)$ and centred at $\hat{\boldsymbol{x}}_{n-1}^C(\boldsymbol{x}_0)$ can be written as

$$u^{(k)}\big(\hat{\boldsymbol{X}}_{n-1}, t_{n-1}\big) = \sum_{|\boldsymbol{r}|=0}^{N} \frac{1}{\boldsymbol{r}!} \frac{\partial^{|\boldsymbol{r}|}}{\partial \boldsymbol{x}_{n-1}^{\boldsymbol{r}}} u^{(k)}\big(\hat{\boldsymbol{x}}_{n-1}^C(\boldsymbol{x}_0), t_{n-1}\big) \Delta \hat{\boldsymbol{W}}_{n-1}^{\boldsymbol{r}}(\boldsymbol{x}_0)$$

$$(3.3) \qquad + R_{n-1,N}^{u^{(k)}}\big(\Delta \hat{\boldsymbol{W}}_{n-1}(\boldsymbol{x}_0)\big)$$

and

$$G^{(k,j)}\big(\hat{\boldsymbol{X}}_{n-1},t_{n-1}\big) = \sum_{|\boldsymbol{r}|=0}^{N-1}\frac{1}{\boldsymbol{r}!}\frac{\partial^{|\boldsymbol{r}|}}{\partial\boldsymbol{x}_{n-1}^{\boldsymbol{r}}}G^{(k,j)}\big(\hat{\boldsymbol{x}}_{n-1}^{C}(\boldsymbol{x}_0),t_{n-1}\big)\Delta\hat{\boldsymbol{W}}_{n-1}^{\boldsymbol{r}}(\boldsymbol{x}_0)$$

(3.4)
$$+R_{n-1,N-1}^{G^{(k,j)}}\big(\Delta\hat{\boldsymbol{W}}_{n-1}(\boldsymbol{x}_0)\big),$$

respectively, where $R_{n-1,N}^{u^{(k)}}\big(\Delta\hat{\boldsymbol{W}}_{n-1}(\boldsymbol{x}_0)\big)$ and $R_{n-1,N-1}^{G^{(k,j)}}\big(\Delta\hat{\boldsymbol{W}}_{n-1}(\boldsymbol{x}_0)\big)$ are remainder terms. If we substitute Eqs. (3.2)–(3.4) into the Euler–Maruyama scheme of (2.1), we obtain the expansion

$$\hat{X}_n(\boldsymbol{x}_0) = \hat{x}_{n-1}^{C^{(k)}}(\boldsymbol{x}_0) + \Delta\hat{W}_{n-1}^{(k)}(\boldsymbol{x}_0)$$

$$+h\sum_{|\boldsymbol{r}|=0}^{N}\frac{1}{\boldsymbol{r}!}\frac{\partial^{|\boldsymbol{r}|}}{\partial\boldsymbol{x}_{n-1}^{\boldsymbol{r}}}u^{(k)}\big(\hat{\boldsymbol{x}}_{n-1}^{C}(\boldsymbol{x}_0),t_{n-1}\big)\Delta\hat{\boldsymbol{W}}_{n-1}^{\boldsymbol{r}}(\boldsymbol{x}_0)$$

$$+\sum_{|\boldsymbol{r}|=0}^{N-1}\sum_{j=1}^{d}\frac{1}{\boldsymbol{r}!}\frac{\partial^{|\boldsymbol{r}|}}{\partial\boldsymbol{x}_{n-1}^{\boldsymbol{r}}}G^{(k,j)}\big(\hat{\boldsymbol{x}}_{n-1}^{C}(\boldsymbol{x}_0),t_{n-1}\big)\Delta W_n^{(j)}\Delta\hat{\boldsymbol{W}}_{n-1}^{\boldsymbol{r}}(\boldsymbol{x}_0)$$

(3.5)
$$+R_{n,N}^{(k)}\big(\Delta\boldsymbol{W}_n,\Delta\hat{\boldsymbol{W}}_{n-1}(\boldsymbol{x}_0)\big),$$

where the new remainder term is

$$R_{n,N}^{(k)}\big(\Delta\boldsymbol{W}_n,\Delta\hat{\boldsymbol{W}}_{n-1}(\boldsymbol{x}_0)\big) := hR_{n-1,N}^{u^{(k)}}\big(\Delta\hat{\boldsymbol{W}}_{n-1}(\boldsymbol{x}_0)\big)$$

$$+\sum_{j=1}^{d}\Delta W_n^{(j)}R_{n-1,N-1}^{G^{(k,j)}}\big(\Delta\hat{\boldsymbol{W}}_{n-1}(\boldsymbol{x}_0)\big).$$

If we decompose $\hat{X}_n^{(k)}(\boldsymbol{x}_0) = \hat{x}_n^{C^{(k)}}(\boldsymbol{x}_0) + \Delta\hat{W}_n^{(k)}(\boldsymbol{x}_0)$ and then substitute

$$\hat{\boldsymbol{x}}_n^{C}(\boldsymbol{x}_0) = \hat{\boldsymbol{x}}_{n-1}^{C}(\boldsymbol{x}_0) + h\boldsymbol{u}\big(\hat{\boldsymbol{x}}_{n-1}^{C}(\boldsymbol{x}_0,t_{n-1})\big),$$

on the left-hand side of Eq. (3.5), then we arrive at the identity (3.1) in the statement of Theorem 3.2. The convergence condition of the expansion is straightforward from Definition 3.1. ∎

Let us remark that the polynomial approximation given in Theorem 3.2 can be written as a polynomial *exclusively dependent* on the subsequence of independent noise increments $\Delta\boldsymbol{W}_m$, for $m = 1,\ldots,n$, i.e., it is possible to write

(3.6) $$\Delta\hat{W}_n^{(k)} = \mathrm{pol}_N\big(\Delta\boldsymbol{W}_n,\ldots,\Delta\boldsymbol{W}_1\big) + R_{n,N}^{(k)}\big(\Delta\boldsymbol{W}_n,\ldots,\Delta\boldsymbol{W}_1\big),$$

where $\mathrm{pol}_N(\cdots)$ denotes a polynomial of order N. This fact can be easily verified by induction. Specifically, expression (3.6) shows that the convergence of the expansion at step n depends only on the noise increments $\Delta\boldsymbol{W}_m$, $m = 1,\ldots,n$. Therefore, in order to apply Theorem 3.2 in the analysis of Algorithm 2.1, we need to establish the conditions that $\Delta\boldsymbol{W}_m$ should satisfy in order to guarantee the convergence of the polynomial expansions of $\hat{\boldsymbol{X}}_n$ or $\Delta\hat{\boldsymbol{W}}_n(\boldsymbol{x}_0)$.

From Eq.(3.6) it can be seen that if the increments of the original noise process, $\Delta\boldsymbol{W}_m$, $m = 1,\ldots,n$, are bounded, then the increments of the effective noise process, $\Delta\hat{\boldsymbol{W}}_n$, are bounded too. Lemma 3.4 below yields explicit bounds for the effective noise $\Delta\hat{\boldsymbol{W}}_n$ in terms of any available bound on $\Delta\boldsymbol{W}_m$.

LEMMA 3.4. *If there are finite constants $A_n^{(j)}$ such that $\left|\Delta W_n^{(j)}\right| < A_n^{(j)}$ for every $n \geqslant 1$ and $j = 1, \ldots, d$, then the constants recursively computed as*

$$\widetilde{A}_{n,N}^{(k)}(\boldsymbol{x}_0) = \widetilde{A}_{n-1,N}^{(k)}(\boldsymbol{x}_0) + h \sum_{|\boldsymbol{r}|=1}^{N} \frac{1}{\boldsymbol{r}!} \left| \frac{\partial^{|\boldsymbol{r}|}}{\partial \boldsymbol{x}_{n-1}^{\boldsymbol{r}}} u^{(k)}\left(\hat{\boldsymbol{x}}_{n-1}^C(\boldsymbol{x}_0), t_{n-1}\right) \right| \widetilde{\boldsymbol{A}}_{n-1,N}^{\boldsymbol{r}}(\boldsymbol{x}_0)$$

$$+ \sum_{|\boldsymbol{r}|=0}^{N-1} \sum_{j=1}^{d} \frac{1}{\boldsymbol{r}!} \left| \frac{\partial^{|\boldsymbol{r}|}}{\partial \boldsymbol{x}_{n-1}^{\boldsymbol{r}}} G^{(kj)}\left(\hat{\boldsymbol{x}}_{n-1}^C(\boldsymbol{x}_0), t_{n-1}\right) \right| A_n^{(j)} \widetilde{\boldsymbol{A}}_{n-1,N}^{\boldsymbol{r}}(\boldsymbol{x}_0),$$

where $\widetilde{\boldsymbol{A}}_{0,N}(\boldsymbol{x}_0) = 0$, are finite and satisfy the inequalities

$$\left|\Delta \hat{W}_{n,N}^{(k)}\right| < \widetilde{A}_{n,N}^{(k)}$$

for every $n \geqslant 1$ and $k = 1, \ldots, d$. Moreover, if

$$\widetilde{\boldsymbol{A}}_{n-1,N}(\boldsymbol{x}_0) \in \rho_{u^{(k)}}\left(\hat{\boldsymbol{x}}_{n-1}^C(\boldsymbol{x}_0), t_{n-1}\right) \cap_{j=1}^{d} \rho_{G^{(k,j)}}\left(\hat{\boldsymbol{x}}_{n-1}^C(\boldsymbol{x}_0), t_{n-1}\right),$$

then

$$\lim_{N \to \infty} R_{n,N}^{(k)}\left(\Delta \boldsymbol{W}_n, \Delta \hat{\boldsymbol{W}}_{n-1}(\boldsymbol{x}_0)\right) = 0.$$

PROOF: It is straightforward from Theorem 3.2.   ∎

We can now apply the results above to provide a convergence condition for the recursive approximation of moments in Eq. (2.4).

THEOREM 3.5. *Assume that the functions $\boldsymbol{u}$ and $\boldsymbol{G}$ in Eq. (1.1) are real and smooth. For any positive integers N and n, and any fixed initial condition $\boldsymbol{X}_0 = \boldsymbol{x}_0$, we have the identity*

$$\mathbb{E}\left[\Delta \hat{\boldsymbol{W}}_n^{\boldsymbol{r}}(\boldsymbol{x}_0)\right] = \sum_{|\boldsymbol{s}|+|\boldsymbol{r}'|=1}^{N} a_{\boldsymbol{r},n,N}^{(\boldsymbol{s},\boldsymbol{r}')}\left(\hat{\boldsymbol{x}}_{n-1}^C(\boldsymbol{x}_0), t_{n-1}\right) \mathbb{E}\left[\Delta \boldsymbol{W}_n^{\boldsymbol{s}}\right] \mathbb{E}\left[\Delta \hat{\boldsymbol{W}}_{n-1,N}^{\boldsymbol{r}'}(\boldsymbol{x}_0)\right]$$

(3.7) $$+ \mathbb{E}\left[\boldsymbol{R}_{n,N}^{(\boldsymbol{r})}\left(\Delta \boldsymbol{W}_n, \Delta \hat{\boldsymbol{W}}_{n-1}(\boldsymbol{x}_0)\right)\right],$$

where $\boldsymbol{R}_{n,N}^{(\boldsymbol{r})}$ is the remainder term of the expansion and $a_{\boldsymbol{r},n,N}^{(\boldsymbol{s},\boldsymbol{r}')}\left(\hat{\boldsymbol{x}}_{n-1}^C(\boldsymbol{x}_0), t_{n-1}\right)$ are the constant coefficients of the expansion of the effective noise in Theorem 3.2. Moreover, if there are finite constants $A_n^{(j)}$ such that $\left|\Delta W_n^{(j)}\right| < A_n^{(j)}$ for every $n \geqslant 1$ and $j = 1, \ldots, d$, then

(3.8) $$\lim_{N \to \infty} \mathbb{E}\left[\boldsymbol{R}_{n,N}^{(\boldsymbol{r})}\left(\Delta \boldsymbol{W}_n, \Delta \hat{\boldsymbol{W}}_{n-1}(\boldsymbol{x}_0)\right)\right] = 0.$$

PROOF: Note that the effective noise monomial $\Delta \hat{\boldsymbol{W}}_n^{\boldsymbol{r}}(\boldsymbol{x}_0)$ can be written as

$$\Delta \hat{\boldsymbol{W}}_n^{\boldsymbol{r}}(\boldsymbol{x}_0) = \prod_i \Delta \hat{W}_n^{(r_i)}(\boldsymbol{x}_0),$$

where the factors $\Delta \hat{W}_n^{(r_i)}(\boldsymbol{x}_0)$ are expanded using Theorem 3.2 and then truncated to order N. We arrive at the identity (3.7), after straightforward manipulations, by taking expectations and realising that

$$\mathbb{E}\left[\Delta \boldsymbol{W}_n^{\boldsymbol{s}} \Delta \hat{\boldsymbol{W}}_{n-1}^{\boldsymbol{r}'}(\boldsymbol{x}_0)\right] = \mathbb{E}\left[\Delta \boldsymbol{W}_n^{\boldsymbol{s}}\right] \mathbb{E}\left[\Delta \hat{\boldsymbol{W}}_{n-1}^{\boldsymbol{r}'}(\boldsymbol{x}_0)\right],$$

which is a consequence of Eq. (3.6) and the independence of the noise increments. As for the convergence of the expansion (3.7), Lemma 3.4 yields

$$\lim_{N\to\infty} R_{n,N}^{(k)}\big(\Delta\boldsymbol{W}_n, \Delta\hat{\boldsymbol{W}}_{n-1}(\boldsymbol{x}_0)\big) = 0, \qquad k = 1, \ldots, v$$

and, since the remainder term of the expansion of $\Delta\hat{\boldsymbol{W}}_n^{\boldsymbol{r}}(\boldsymbol{x}_0)$ is the addition of a finite sum of products involving $R_{n,N}^{(k)}\big(\Delta\boldsymbol{W}_n, \Delta\hat{\boldsymbol{W}}_{n-1}(\boldsymbol{x}_0)\big)$ for all $k = 1, \ldots, v$, we obtain that

$$(3.9) \qquad \lim_{N\to\infty} \boldsymbol{R}_{n,N}^{(\boldsymbol{r})}\big(\Delta\boldsymbol{W}_n, \Delta\hat{\boldsymbol{W}}_{n-1}(\boldsymbol{x}_0)\big) = 0.$$

Finally, if we take the expectation of (3.9) and apply the dominated convergence Theorem [22] we arrive at (3.8) and complete the proof.
∎

The results we have obtained are useful to guarantee convergence when the support of the dynamical noise $\boldsymbol{W}_n$ is bounded but, in general, this is not the case. However, even if in the most common models (Gaussian distributions, Gamma distributions, etc.) the support is not actually bounded, when the tails of a distribution decrease rapidly enough the support can be treated as bounded for numerical purposes. For example, if $W_n^{(k)} \sim \mathcal{N}(0, \sigma)$ then $\mathbb{P}(|W_n^{(k)}| < 3\sigma) > 0.9973$, i.e., $W_n^{(k)}$ is bounded with high probability.
For a prescribed probability $P \in (0, 1)$, let us choose the quantities

$$(3.10) \qquad A_n^{(j)}(P) := \inf\left\{a \in \mathbb{R}_0^+ : \mathbb{P}\big(|W_n^{(j)}| < a\big) > P\right\}, \qquad j = 1, \ldots, d,$$

i.e., $A_n^{(j)}(P)$ is an upper bound for $|W_n^{(j)}|$ with probability P. One can combine bounds that hold with some probability P and Lemma 3.4 to assess the convergence of the polynomial expansions of the moments of the effective noise in Algorithm 2.1.

4. Approximation of 1-dimensional marginal densities. In this section we prove that the approximate 1-dimensional pdf's computed using Algorithm 2.2 converge as the order of the Gram-Charlier expansion, N, increases, provided that the initial condition is fixed, $\boldsymbol{X}_0 = \boldsymbol{x}_0$. When the initial condition is random, we further extend the latter result with the convergence of the Monte Carlo estimator in Eq. (2.28) as the number of samples N_s' increases.

THEOREM 4.1. *Let X be a real random variable with pdf f_X and characteristic function Ψ_X; then choose an auxiliary random variable Z_φ with smooth pdf φ and characteristic function Ψ_φ such that $\left|\frac{\Psi_X(t)}{\Psi_\varphi(t)}\right| < \infty$ for all finite t. The density f_X can be expanded with respect to the derivatives of φ as*

$$(4.1) \qquad f_X(x) = \left[1 + \sum_{r=1}^{N} \frac{(-1)^r}{r!} C_r\big[X, Z_\varphi\big]\frac{d^r}{dx^r}\right]\varphi(x) + R_N\big(f_X, \varphi; x\big),$$

where $C_r\big[X, Z_\varphi\big]$ are the coefficients of the expansion defined in Eq. (2.19) and $R_N\big(f_X, \varphi; x\big)$ is a remainder term.

PROOF: We write Ψ_X as

$$\Psi_X(t) = \frac{\Psi_X(t)}{\Psi_\varphi(t)}\Psi_\varphi(t)$$

$$= \exp\Big(\log\big(\Psi_X(t)\big) - \log\big(\Psi_\varphi(t)\big)\Big)\Psi_\varphi(t)$$

(4.2) $$= \left[\exp\left(\sum_{r=0}^{N}\frac{\kappa_r(X) - \kappa_r(Z_\varphi)}{r!}(it)^r\right) + \mathrm{R}_N\left(\frac{\Psi_X}{\Psi_\varphi};t\right)\right]\Psi_\varphi(t),$$

where $\mathrm{R}_N\left(\frac{\Psi_X}{\Psi_\varphi};t\right)$ is the remainder of the Taylor expansion of function Ψ_X/Ψ_φ. If we expand the exponential function in (4.2) in terms of Bell polynomials [1] and then compute the inverse Fourier transform on both sides of the equation we arrive at

$$f_X(x) = \left[1 + \sum_{r=1}^{N}\frac{(-1)^r}{r!}C_r\big[X, Z_\varphi\big]\frac{\mathrm{d}^r}{\mathrm{d}x^r}\right]\varphi(x) + \frac{1}{2\pi}\int_{\mathbb{R}}\mathrm{R}_N\left(\frac{\Psi_X}{\Psi_\varphi};t\right)\Psi_\varphi(t)\,\mathrm{e}^{-ixt}\,\mathrm{d}t,$$

where the second term on the r.h.s. is the remainder in Eq. (4.1). ■

Many families of orthogonal polynomials are related to specific probability distributions [6] in the sense that there are formulas to generate the polynomials from the derivatives of probability densities (the so-called Rodrigues formulas [20]). In particular, the class of probabilistic Hermite polynomials are orthogonal w.r.t. the Gaussian distribution and the Rodrigues formula for them is given by Eq. (2.22).

If we let the auxiliary pdf φ be Gaussian, the Gram–Charlier expansion of f_X in Theorem 4.1 reduces to a series of Hermite polynomials multiplied by φ. Hence, the convergence of expression (4.1) becomes a standard problem, similar to the convergence of the PCE (2.6) in Section 2.4. Indeed, if φ is Gaussian, the series in (4.1) is termed a Gram–Charlier expansion of type A and it can be expected to converge when[1] $\frac{f_X}{\varphi} \in L^2\left(\mathbb{R}, \varphi\right)$ (see [6]).

In the sequel, we restrict our attention to the Gram-Charlier expansion of type A and hence assume that the auxiliary pdf φ used to approximate the k-th 1-dimensional marginal pdf $f_{\hat{X}_n^{(k)}|\boldsymbol{x}_0}$ is Gaussian, with mean $\mu_n^{(k)}(\boldsymbol{x}_0)$ and standard deviation $\sigma_n^{(k)}(\boldsymbol{x}_0)$. We specifically denote it as $\varphi_{n,\boldsymbol{x}_0}^{(k)}(x)$ (note the dependence on the initial condition $\boldsymbol{x}_0$).

Below, we establish some regularity assumptions and then use them to provide an explicit convergence theorem for the approximations of $f_{\hat{X}_n^{(k)}|\boldsymbol{x}_0}$.

ASSUMPTION 4.2. *Let* $\mathrm{supp}(f)$ *denote the support of function* f *and let* $f_{\Delta\boldsymbol{W}_m}$ *denote the pdf of the random vector of noise increments at time* m, $\boldsymbol{W}_m$. *There are bounded sets* $D_m \subset \mathbb{R}^d$, $m = 1, \ldots, n$, *such that*

$$\mathrm{supp}(f_{\Delta\boldsymbol{W}_m}) \subseteq D_m.$$

Moreover, there is a sequence of finite constants M_m, $m = 1, \ldots, n$, *that satisfy the inequalities*

$$\sup_{x\in\mathbb{R}}\big(f_{\Delta\boldsymbol{W}_m}(x)\big) \leqslant M_m.$$

---

[1] We construct the class of real L^2 functions w.r.t. a density $\varphi : S \mapsto (0, \infty)$ as

$$L^2(S, \varphi) := \left\{h : S \mapsto \mathbb{R} \text{ such that } \int_S h(x)\varphi(x)\mathrm{d}x < \infty\right\}.$$

480

481 *Remark* 4.3. When the SDE (1.1) is driven by, e.g., a Wiener process $\boldsymbol{W}(t)$ and
482 we define $\Delta\boldsymbol{W}_m = \boldsymbol{W}(t_m) - \boldsymbol{W}(t_{m-1})$, the support of $f_{\Delta\boldsymbol{W}_m}$ is $\mathbb{R}^v$ and Assumption
483 4.2 does not hold. It is well known, however, that weak Euler-Maruyama schemes can
484 be designed with simplified noise increments [13]. To be specific, weak convergence
485 of the Euler-Maruyama scheme (2.1) can be guaranteed when the noise increments
486 $\Delta W_m^{(j)}$, $j = 1, \ldots, v$, are independent and satisfy the set of the inequalities

487 (4.3) $\quad \left| \mathbb{E}\left[\Delta W_m^{(j)} \right] \right| + \left| \mathbb{E}\left[\left(\Delta W_m^{(j)} \right)^3 \right] \right| + \left| \mathbb{E}\left[\left(\Delta W_m^{(j)} \right)^2 - h \right] \right| \leqslant Ch^2, \quad 1 \leqslant j \leqslant v,$

488 for some constant $C < \infty$ –see [13, Section 14.1] for details. Hence, if weak convergence
489 is sufficient, the noise increments $\Delta W_m^{(j)}$ can be selected in many ways. For example,
490 choosing the $\Delta W_m^{(j)}$'s to be i.i.d. with common uniform distribution $\mathcal{U}(-ah, +ah)$,
491 for some constant $a > 0$, guarantees that (4.3) is satisfied for any $C \geqslant \frac{1}{3}a^2$, while
492 Assumption 4.2 holds with $D_m = [-ah, +ah]^v$.

493 ASSUMPTION 4.4. *There are finite constants* $\left\{ A_m^{(j)} : m = 1, \ldots, n; \quad j = 1, \ldots, v \right\}$
494 *such that* $|\Delta W_m^{(j)}| < A_m^{(j)}$, *for* $1 \leqslant j \leqslant v$ *and* $1 \leqslant m \leqslant n$, *and*

495 (4.4) $\qquad \widetilde{\boldsymbol{A}}_{m,N}(\boldsymbol{x}_0) \in \rho_{u^{(k)}}\left(\hat{\boldsymbol{x}}_m^C(\boldsymbol{x}_0), t_m \right) \cap_{j=1}^d \rho_{G^{(k,j)}}\left(\hat{\boldsymbol{x}}_m^C(\boldsymbol{x}_0), t_m \right)$

496 *where* k*-th entry of the* v*-dimensional vector* $\widetilde{\boldsymbol{A}}_{m,N}(\boldsymbol{x}_0)$ *is constructed as*

497 $\widetilde{A}_{m,N}^{(k)}(\boldsymbol{x}_0) = \widetilde{A}_{m-1,N}^{(k)}(\boldsymbol{x}_0) + h \sum_{|\boldsymbol{r}|=1}^N \frac{1}{\boldsymbol{r}!} \left| \frac{\partial^{|\boldsymbol{r}|}}{\partial\boldsymbol{x}_{m-1}^{\boldsymbol{r}}} u^{(k)}\left(\hat{\boldsymbol{x}}_{m-1}^C(\boldsymbol{x}_0), t_{m-1} \right) \right| \widetilde{\boldsymbol{A}}_{m-1,N}^{\boldsymbol{r}}(\boldsymbol{x}_0)$

498 (4.5) $\qquad + \sum_{|\boldsymbol{r}|=0}^{N-1} \sum_{j=1}^d \frac{1}{\boldsymbol{r}!} \left| \frac{\partial^{|\boldsymbol{r}|}}{\partial\boldsymbol{x}_{m-1}^{\boldsymbol{r}}} G^{(k,j)}\left(\hat{\boldsymbol{x}}_{m-1}^C(\boldsymbol{x}_0), t_{m-1} \right) \right| A_m^{(j)} \widetilde{\boldsymbol{A}}_{m-1,N}^{\boldsymbol{r}}(\boldsymbol{x}_0),$

499 *with initial condition* $\widetilde{\boldsymbol{A}}_{0,N}(\boldsymbol{x}_0) = 0$.

500 While Assumption 4.2 states that the support of the noise components is bounded,
501 Assumption 4.4 guarantees that finite noise increments $\Delta W_m^{(k)}$ yield finite effective
502 noise terms $\Delta \hat{W}_m^{(k)}$ and enables us to apply Lemma 3.4. Given the above regularity
503 assumptions we can provide guarantees on the approximation of the marginal densities
504 $f_{\hat{X}_n^{(k)} | \boldsymbol{x}_0}(x)$.

505

506 THEOREM 4.5. *Let the functions* $\boldsymbol{u}$ *and* $\boldsymbol{G}$ *in the SDE* (1.1) *be smooth, let*
507 *Assumptions* 4.2 *and* 4.4 *hold and let* $\boldsymbol{x}_0$ *be a fixed initial condition. Then, the*
508 *type A Gram-Charlier expansion of the 1-dimensional marginal pdf of* $\hat{X}_n^{(k)}(\boldsymbol{x}_0)$,
509 $k \in \{1, \ldots, v\}$, *can be written as*

510 $f_{\hat{X}_n^{(k)} | \boldsymbol{x}_0}(x) = \left[1 + \sum_{r=1}^N \frac{1}{r! \, \sigma_n^{(k)}(\boldsymbol{x}_0)^r} C_r\left[\hat{X}_{n,N}^{(k)}(\boldsymbol{x}_0), Z^{(k)}(\boldsymbol{x}_0) \right] H_{e_r}\left(\frac{x - \mu_n^{(k)}(\boldsymbol{x}_0)}{\sigma_n^{(k)}(\boldsymbol{x}_0)} \right) \right] \times$

511 $\qquad \times \varphi_{n,\boldsymbol{x}_0}^{(k)}(x) + R_{n,N}^{(k)}\left(f_{\hat{X}_n^{(k)} | \boldsymbol{x}_0}, \varphi_{n,\boldsymbol{x}_0}^{(k)}(\cdot); x \right),$

512 *where* $\left\{ H_{e_r} \right\}_{r=0}^\infty$ *are the probabilistic Hermite polynomials given by Eq.* (2.22),
513 $Z^{(k)}(\boldsymbol{x}_0)$ *is a random variable with pdf* $\varphi_{n,\boldsymbol{x}_0}^{(k)}$ *and the remainder term vanishes as*

the truncation order N is increased, i.e.,

(4.6) $$\lim_{N \to \infty} R_{n,N}^{(k)} \big(f_{\hat{X}_n^{(k)}|\boldsymbol{x}_0}, \varphi_{n,\boldsymbol{x}_0}^{(k)}(\cdot); x \big) = 0.$$

PROOF: The type A Gram–Charlier expansion of $f_{\hat{X}_n^{(k)}|\boldsymbol{x}_0}(x)$ is immediately obtained from Eq. (2.18) when the auxiliary pdf is Gaussian (namely, $\varphi = \varphi_{n,\boldsymbol{x}_0}^{(k)}$). Additionally, we need to prove that Eq. (4.6) holds, which takes more effort. Specifically, hereafter we prove that the function $\frac{f_{\hat{X}_n^{(k)}|\boldsymbol{x}_0}}{\varphi_{n,\boldsymbol{x}_0}^{(k)}}$ belongs to $L^2(\mathbb{R}, \varphi_{n,\boldsymbol{x}_0}^{(k)})$, which, in turn, implies that $R_{n,N}^{(k)} \big(f_{\hat{X}_n^{(k)}|\boldsymbol{x}_0}, \varphi_{n,\boldsymbol{x}_0}^{(k)}(\cdot); x \big) \xrightarrow{N \to \infty} 0$ in L^2 (see [6]).

First, we prove using an induction argument that the pdf $f_{\hat{\boldsymbol{X}}_{n-1}|\boldsymbol{x}_0}$ is bounded and it has a bounded support. Let us assume that at time $n-1$ there are a bounded set $\hat{D}_{n-1} \subset \mathbb{R}^v$ and a finite constant $\hat{M}_{n-1}$ such that

(4.7) $$\mathrm{supp}\big(f_{\hat{\boldsymbol{X}}_{n-1}|\boldsymbol{x}_0}\big) \subseteq \hat{D}_{n-1} \subset \mathbb{R}^v \qquad \text{and} \qquad \sup_{x \in \mathbb{R}} \big(f_{\hat{\boldsymbol{X}}_{n-1}|\boldsymbol{x}_0}(x)\big) \leqslant \hat{M}_{n-1} < \infty.$$

From the expression of the Euler–Maruyama integrator in Eq. (2.1), we can write the pdf of $\hat{\boldsymbol{X}}_n$ in terms of the densities of $\hat{\boldsymbol{X}}_{n-1}$ and $\Delta \boldsymbol{W}_n$ as

$$f_{\hat{\boldsymbol{X}}_n|\boldsymbol{x}_0}(\boldsymbol{x}_n) = \int_{\mathbb{R}^{v+d}} f_{\hat{\boldsymbol{X}}_{n-1}|\boldsymbol{x}_0}(\boldsymbol{x}_{n-1}) f_{\Delta \boldsymbol{W}_n}(\boldsymbol{w}_n) \times$$

(4.8) $$\times \delta \big(\boldsymbol{x}_n - \boldsymbol{x}_{n-1} - h\boldsymbol{u}(\boldsymbol{x}_{n-1}, t_{n-1}) - \boldsymbol{G}(\boldsymbol{x}_{n-1}, t_{n-1})\boldsymbol{w}_n \big) \mathrm{d}\boldsymbol{x}_{n-1}\mathrm{d}\boldsymbol{w}_n,$$

where $\delta(\cdot)$ denotes the Dirac delta function (see Eq. 4.34 in [23]). Using Assumption 4.2 and the induction hypothesis (4.7) we obtain an upper bound for the pdf $f_{\hat{\boldsymbol{X}}_n|\boldsymbol{x}_0}(\boldsymbol{x}_n)$ in Eq. (4.8) of the form

$$f_{\hat{\boldsymbol{X}}_n|\boldsymbol{x}_0}(\boldsymbol{x}_n) \leqslant$$

$$\hat{M}_{n-1}M_n \int_{\hat{D}_{n-1} \times D_n} \delta \big(\boldsymbol{x}_n - \boldsymbol{x}_{n-1} - h\boldsymbol{u}(\boldsymbol{x}_{n-1}, t_{n-1}) - \boldsymbol{G}(\boldsymbol{x}_{n-1}, t_{n-1})\boldsymbol{w}_n \big) \mathrm{d}\boldsymbol{x}_{n-1}\mathrm{d}\boldsymbol{w}_n \leqslant$$

$$\hat{M}_{n-1}M_n,$$

hence

(4.9) $$\sup_{x \in \mathbb{R}} \Big(f_{\hat{\boldsymbol{X}}_n|\boldsymbol{x}_0}(x) \Big) \leqslant \hat{M}_n < \infty, \quad \text{where} \quad \hat{M}_n = \hat{M}_{n-1}M_n.$$

Moreover, since $\boldsymbol{u}$ and $\boldsymbol{G}$ are smooth and $\hat{D}_{n-1} \times D_n$ is bounded, all the solutions of the equation

$$\boldsymbol{x}_n - \boldsymbol{x}_{n-1} - h\boldsymbol{u}(\boldsymbol{x}_{n-1}, t_{n-1}) - \boldsymbol{G}(\boldsymbol{x}_{n-1}, t_{n-1})\boldsymbol{w}_n = 0$$

necessarily lie in a bounded set $\hat{D}_n \subset \mathbb{R}^v$, which implies that

(4.10) $$\mathrm{supp}\big(f_{\hat{\boldsymbol{X}}_n|\boldsymbol{x}_0}\big) \subseteq \hat{D}_n \subset \mathbb{R}^v.$$

To complete the induction argument, we need to prove that

$$\mathrm{supp}\big(f_{\hat{\boldsymbol{X}}_1|\boldsymbol{x}_0}\big) \subseteq \hat{D}_1 \subset \mathbb{R}^v \qquad \text{and} \qquad \sup_{x \in \mathbb{R}} \big(f_{\hat{\boldsymbol{X}}_1|\boldsymbol{x}_0}(x)\big) \leqslant \hat{M}_1 < \infty$$

539 for some bounded set $\hat{D}_1$ and some finite constant $\hat{M}_1$. Resorting again to the
540 expression of the Euler–Maruyama integrator (2.1) and Assumption 4.2 we obtain
541 the inequalities

542 $$f_{\hat{\boldsymbol{X}}_1|\boldsymbol{x}_0}(\boldsymbol{x}_1) = \int_{\mathbb{R}^d} f_{\Delta\boldsymbol{W}_1}(\boldsymbol{w}_1)\delta\Big(\boldsymbol{x}_1 - \boldsymbol{x}_0 - h\boldsymbol{u}(\boldsymbol{x}_0,t_0) - \boldsymbol{G}(\boldsymbol{x}_0,t_0)\boldsymbol{w}_1\Big)\mathrm{d}\boldsymbol{w}_1$$

543 $$\leqslant M_1 \int_{\mathbb{R}^d} \delta\Big(\boldsymbol{x}_1 - \boldsymbol{x}_0 - h\boldsymbol{u}(\boldsymbol{x}_0,t_0) - \boldsymbol{G}(\boldsymbol{x}_0,t_0)\boldsymbol{w}_1\Big)\mathrm{d}\boldsymbol{w}_1$$

544 $$\leqslant M_1 < \infty,$$

545 hence $\hat{M}_1 = M_1$ and, by the same reasoning as in the induction step, the solutions of
546 the equation $\boldsymbol{x}_1 - \boldsymbol{x}_0 - h\boldsymbol{u}(\boldsymbol{x}_0,t_0) - \boldsymbol{G}(\boldsymbol{x}_0,t_0)\boldsymbol{w}_1 = 0$ lie in a bounded set $\hat{D}_1 \subset \mathbb{R}^v$
547 which contains the support of $f_{\hat{\boldsymbol{X}}_1|\boldsymbol{x}_0}$.

548 The bounds in (4.9) and (4.10) imply that $f_{\hat{\boldsymbol{X}}_n|\boldsymbol{x}_0} \in L^2\left(\mathbb{R}, \varphi^{(k)}_{n,\boldsymbol{x}_0}(x)\mathrm{d}x\right)$. To show
549 it, let $\hat{D}_n^{(k)}$ be the projection of the bounded set $\hat{D}_n \subset \mathbb{R}^v$ along the k-th dimension
550 and then note that

551 (4.11) $$\int_{\mathbb{R}} \left(\frac{f_{\hat{X}_n^{(k)}|\boldsymbol{x}_0}(x)}{\varphi^{(k)}_{n,\boldsymbol{x}_0}(x)}a\right)^2 \varphi^{(k)}_{n,\boldsymbol{x}_0}(x)\mathrm{d}x \leqslant \frac{\int_{\hat{D}_n^{(k)}}\left(f_{\hat{X}_n^{(k)}|\boldsymbol{x}_0}(x)\right)^2\mathrm{d}x}{\inf_{x\in\hat{D}_n^{(k)}}\varphi^{(k)}_{n,\boldsymbol{x}_0}(x)} < \infty,$$

552 where the second inequality holds because
553   • $\inf_{x\in\hat{D}_n^{(k)}}\varphi^{(k)}_{n,\boldsymbol{x}_0}(x) > 0$, since $\varphi^{(k)}_{n,\boldsymbol{x}_0}$ is Gaussian and $\hat{D}_n^{(k)}$ is bounded, and
554   • the marginal density $f_{\hat{X}_n^{(k)}|\boldsymbol{x}_0}$ is bounded because the joint density $f_{\hat{\boldsymbol{X}}_n|\boldsymbol{x}_0}$ is
555   bounded.
556 The inequality (4.11) yields $f_{\hat{\boldsymbol{X}}_n|\boldsymbol{x}_0} \in L^2\left(\mathbb{R}, \varphi^{(k)}_{n,\boldsymbol{x}_0}(x)\mathrm{d}x\right)$ which, in turn, implies that
557 (4.6) holds [6]. ∎

Theorem 4.5 states that the estimates of the 1-dimensional marginal pdf's
$f_{\hat{X}_n^{(k)}|\boldsymbol{x}_0}(x)$ converge pointwise, for any fixed $\boldsymbol{x}_0$ and $x \in D_n$, as the truncation order
N increases. When the initial condition is random, the natural estimate to compute is
the Monte Carlo approximation in Eq. (2.28). The proposition below guarantees that,
under similar assumptions as in Theorem 4.5, the Monte Carlo estimator converges
to
$$f_{\hat{X}_n^{(k)}}(x) = \mathbb{E}\left[f_{\hat{X}_n^{(k)}|\boldsymbol{X}_0}(x)\right]$$
558 almost surely (a.s.) for any $x \in D_n$.
559

560 PROPOSITION 4.6. *Let* $\boldsymbol{X}_{0,j}$, $j = 1,\ldots,N_s'$, *be i.i.d. samples form the initial*
561 *distribution with pdf* $f_{\boldsymbol{X}_0}$. *If Assumptions 4.2 and 4.4 hold and* $f_{\boldsymbol{X}_0}$ *is bounded with*
562 *bounded support, then*

563 (4.12) $$\lim_{N_s'\to\infty}\left[\lim_{N\to\infty}\frac{1}{N_s'}\sum_{j=1}^{N_s'}f_{\hat{X}_{n,N}^{(k)}|\boldsymbol{X}_{0,j}'}(x)\right] = f_{\hat{X}_n^{(k)}}(x) \quad a.s.,$$

564 *for* $k = 1,\ldots,v$.

PROOF: For any $N_s' \in \mathbb{N}$, Theorem 4.5 yields
$$\lim_{N\to\infty}\frac{1}{N_s'}\sum_{j=1}^{N_s'}f_{\hat{X}_{n,N}^{(k)}|\boldsymbol{X}_{0,j}'}(x) = \frac{1}{N_s'}\sum_{j=1}^{N_s'}f_{\hat{X}_n^{(k)}|\boldsymbol{X}_{0,j}'}(x).$$

Moreover, since $f_{\boldsymbol{X}_0}$ is bounded and has a bounded support, the same argument as in the proof of Theorem 4.5 shows that the pdf's $f_{\hat{X}_n^{(k)}|\boldsymbol{X}_{0,j}'}$ are uniformly bounded[2],

hence $\mathbb{E}\left[\left(f_{\hat{X}_n^{(k)}|\boldsymbol{X}_{0,j}'}(x)\right)^2\right] < \infty$ and the strong law of large numbers yields Eq. (4.12).

■

5. Numerical examples. In order to illustrate the performance of the proposed uncertainty quantification scheme we provide a numerical example in which we compare the solution of the dynamics of a Keplerian orbit in two-dimensional space, perturbed by a diffusion term, using Algorithm 2.1 for the approximation of moments and a Monte Carlo simulation with 10^4 samples as a baseline. We present two sets of simulation results that differ essentially in the choice of initial condition, which is deterministic for the first set (Section 5.1) while we assume it random, with a Gaussian distribution, for the second one (Section 5.2).

We start from the general equation (1.1). The state vector of the orbiting object has dimension $d = 4$ and we denote it as $\boldsymbol{X}(t) = (x(t), y(t), v_x(t), v_y(t))^\top$, where $(x(t), y(t))$ is the object position in km and $(v_x(t), v_y(t))$ is its velocity in km/s, respectively, in 2-dimensional space. For the Keplerian dynamics, the drift coefficient $\boldsymbol{u}(\boldsymbol{X}, t)$ can be written as [28]

$$(5.1) \qquad \boldsymbol{u}(\boldsymbol{X}, t) = \begin{pmatrix} v_x(t) \\ v_y(t) \\ \dfrac{-\mu x(t)}{\left[x(t)^2 + y(t)^2\right]^{3/2}} \\ \dfrac{-\mu y(t)}{\left[x(t)^2 + y(t)^2\right]^{3/2}} \end{pmatrix},$$

where μ is the standard gravitational parameter, and we set the diffusion coefficient as the 4×4 diagonal matrix $\boldsymbol{G}(\boldsymbol{X}, t) = \text{diag}\left(0, 0, \sigma_w/\|\boldsymbol{v}(t)\|, \sigma_w/\|\boldsymbol{v}(t)\|\right)$, where σ_w is a known positive constants and $\|\boldsymbol{v}(t)\| = \sqrt{v_x(t)^2 + v_y(t)^2}$ is the Euclidean norm of the velocity. The noise process $\boldsymbol{W}(t)$ is a standard 4×1 Wiener process. Physically, the diffusion term $\boldsymbol{G}(\boldsymbol{X}, t)\mathrm{d}\boldsymbol{W}(t)$ represents a stochastic perturbation in the acceleration of the orbiting object (which depends on the object velocity $\boldsymbol{v}(t)$). The numerical values used for the simulation are summarised in Table 5.1.

All computer experiments have been performed using Matlab R2018b running on a Mac Book Pro computer equipped with a 2.3 GHz Intel Core i5 CPU and 16 GB of RAM.

5.1. Deterministic initial condition. For the first experiment, we fix the initial condition as

$$(5.2) \qquad \boldsymbol{x}_0 = \begin{pmatrix} 200 + R_T \\ 0 \\ 0 \\ \sqrt{\dfrac{\mu}{200 + R_T}} \end{pmatrix},$$

---

[2]The bounds $\hat{M}_n$ in the proof of Theorem 4.5 depend on the initialization, i.e., $\hat{M}_n = \hat{M}_n(\boldsymbol{X}_0)$. However, the bounds $\hat{M}_n(\boldsymbol{X}_0)$ are continuous by construction and, since the support of $\boldsymbol{X}_0$ is bounded, $\sup_{\boldsymbol{X}_0} \hat{M}_n(\boldsymbol{X}_0) < \infty$.

Parameters	Value	Description
μ	$3.986 \ \mathrm{km}^3/\mathrm{s}^2$	Standard gravitational parameter.
h	$0.1\,\mathrm{s}$	Step-size for time discretisation.
t_0	$0\,\mathrm{days}$	Initial time.
t_n	$1.51\,\mathrm{days}$	Final time.
n	$1,304,640$	Number of discrete-time steps in the simulations, namely, $n = \lceil (t_n - t_0)/h \rceil$ where $\lceil \cdot \rceil$ denotes the ceiling function.
N	2	Order of the polynomial expansions.
N_{PCE}	4	Truncation order of PCE (for the second example only).
N_s	140	Number of samples (for the second example only).
N_s'	10^6	Number of samples of $\boldsymbol{X}_0$ generated to reconstruct the marginal pdf's $f_{\hat{X}_{n,N}^{(k)}}(x)$.
σ_w	2×10^{-4}	Scale parameter in the diffusion term.

Table 5.1: Simulation parameters.

where $R_T = 6.378 \times 10^3$ km is the Earth radius. By taking a known initial condition, we can asses the moment and density approximations when the only source of uncertainty is the dynamical noise $\boldsymbol{W}(t)$ and, therefore, we avoid any PCE approximation.

The initial state $\boldsymbol{x}_0$ has been chosen to simulate the evolution of a nearly circular orbit at 200 km above the Earth surface. At this low altitude, it is relevant to use a SDE to represent the orbital dynamics because the object motion depends on the atmosphere drag which, in turn, depends on several parameters (atmosphere density, mass, volume, shape of the object, etc.) which are often difficult to determine in practice [28]. The diffusion term $\boldsymbol{G}(\boldsymbol{X},t)\mathrm{d}\boldsymbol{W}(t)$ may account for these uncertainties.

Table 5.2 shows a comparison between the outcomes, at the final time t_n, of Algorithm 2.1 and the baseline Monte Carlo method with 10^4 independent trajectories generated using the Euler-Maruyama scheme (2.1). The first column in the table displays the expected values of x, y, v_x and v_y computed with Algorithm 2.1, while the second column shows the Monte Carlo estimates for each state variable. The third column displays the absolute differences between the first and second columns, and the fourth column shows the relative difference (with the Monte Carlo estimates taken as reference). We can observe that both methods yield very similar outputs, with relative differences between 0.4% and 0.8% for all state variables.

Table 5.3 shows a comparison between the estimates of the second order moments of $\boldsymbol{X}(t_n)$ computed via Algorithm 2.1 and the standard Monte Carlo method that runs the Euler-Maruyama scheme 10^4 times. The first row shows the covariance matrix of $\boldsymbol{X}(t_n)$ as output by Algorithm 2.1, while the second row shows the Monte Carlo estimate. The entry-wise absolute and relative differences between the two matrices are displayed in the third and fourth rows of the table, respectively. The differences are larger than for the first-order moments, yet the two methods still yield similar outputs (with relative differences between 10% and 20% for all entries of the covariance

	Algorithm 2.1	Monte Carlo, 10^4 samples	Absolute difference	Relative difference
x	$5.95 \times 10^3 \,\mathrm{km}$	$5.92 \times 10^3 \,\mathrm{km}$	$24.34 \,\mathrm{km}$	4.09×10^{-3}
y	$-2.67 \times 10^3 \,\mathrm{km}$	$-2.69 \times 10^3 \,\mathrm{km}$	$20.64 \,\mathrm{km}$	7.74×10^{-3}
v_x	$3.01 \,\mathrm{km/s}$	$2.99 \,\mathrm{km/s}$	$2.27 \times 10^{-2} \,\mathrm{km/s}$	7.58×10^{-3}
v_y	$6.66 \,\mathrm{km/s}$	$6.68 \,\mathrm{km/s}$	$2.82 \times 10^{-2} \,\mathrm{km/s}$	4.21×10^{-3}

Table 5.2: Estimate of $\mathbb{E}[\boldsymbol{X}(t_n)]$ with the moment-computation Algorithm 2.1, compared with standard Monte Carlo estimates. The initial condition is deterministic.

Algorithm 2.1	$\begin{pmatrix} 6.08 \times 10^5 & 1.58 \times 10^6 & -1.77 \times 10^3 & 7.01 \times 10^2 \\ 1.58 \times 10^6 & 3.46 \times 10^6 & -3.87 \times 10^3 & 1.81 \times 10^3 \\ -1.77 \times 10^3 & -3.87 \times 10^3 & 4.32 & -2.03 \\ 7.01 \times 10^2 & 1.81 \times 10^3 & -2.03 & 8.09 \times 10^{-1} \end{pmatrix}$
Monte Carlo	$\begin{pmatrix} 7.41 \times 10^5 & 1.32 \times 10^6 & -1.47 \times 10^3 & 8.45 \times 10^2 \\ 1.32 \times 10^6 & 3.14 \times 10^6 & -3.52 \times 10^3 & 1.51 \times 10^3 \\ -1.47 \times 10^3 & -3.52 \times 10^3 & 3.93 & -1.69 \\ 8.45 \times 10^2 & 1.51 \times 10^3 & -1.69 & 9.65 \times 10^{-1} \end{pmatrix}$
Absolute differences	$\begin{pmatrix} 1.33 \times 10^5 & 2.61 \times 10^5 & 2.97 \times 10^2 & 1.44 \times 10^2 \\ 2.61 \times 10^5 & 3.18 \times 10^5 & 3.52 \times 10^2 & 2.98 \times 10^2 \\ 2.97 \times 10^2 & 3.52 \times 10^2 & 3.89 \times 10^{-1} & 3.39 \times 10^{-1} \\ 1.44 \times 10^2 & 2.98 \times 10^2 & 3.39 \times 10^{-1} & 1.56 \times 10^{-1} \end{pmatrix}$
Relative differences	$\begin{pmatrix} 1.79 \times 10^{-1} & 1.99 \times 10^{-1} & 2.02 \times 10^{-1} & 1.71 \times 10^{-1} \\ 1.99 \times 10^{-1} & 1.01 \times 10^{-1} & 1.00 \times 10^{-1} & 1.97 \times 10^{-1} \\ 2.02 \times 10^{-1} & 1.00 \times 10^{-1} & 9.90 \times 10^{-2} & 2.01 \times 10^{-1} \\ 1.71 \times 10^{-1} & 1.97 \times 10^{-1} & 2.01 \times 10^{-1} & 1.62 \times 10^{-1} \end{pmatrix}$

Table 5.3: Estimates of the covariance matrix of $\boldsymbol{X}(t_n)$ computed via Algorithm 2.1 and standard Monte Carlo, with 10^4 independent samples, with deterministic initial condition.

matrix).

Since Algorithm 2.1 yields outputs which are close to the baseline Monte Carlo estimates, it is of interest to compare the computational cost of the two procedures. This is done in Table 5.5, which displays the mean run-time per discrete time step (first row) and the total run-time up to time t_n (second row)

- for Algorithm 2.1,
- for the Monte Carlo method with 10^4 samples and
- for a single run of the Euler-Maruyama scheme (2.1).

We see that the cost of running the Algorithm 2.1 (computation of moments) is roughly of the same order as running the standard Euler-Maruyama scheme once, and

632 two orders of magnitude less expensive than computing the Monte Carlo estimators.

	Algorithm 2.1	Monte Carlo, 10^4 samples	Euler-Maruyama, single run
Mean run-time per time step	8.66×10^{-6} s	1.62×10^{-3} s	5.39×10^{-6} s
Total run-time	11.30 s	2.11×10^3 s	7.04 s

Table 5.4: Run-times in seconds (s) with deterministic initial condition. The total number of discrete time steps is $n = 1,304,640$.

x	y	v_x	v_y
1.71×10^{-5}	2.09×10^{-6}	1.64×10^{-3}	1.82×10^{-2}

Table 5.5: TVD between the estimates of the marginal densities computed with Algorithm 2.2 and the KDEs computed from 10^4 Monte Carlo samples.

633   Next, we turn attention to the performance of Algorithm 2.2, which yields
634 estimates of the marginal densities of the state variables x, y, v_x and v_y. Figure
635 5.1 shows corresponding pdf's as generated by Algorithm 2.2 (in red colour) and
636 the kernel density estimators[3] (KDEs) computed from the independent samples
637 generated by running the Euler-Maruyama scheme (2.1) 10^4 times. We see that
638 the KDEs are clearly non-Gaussian for x and v_y and there is a clear mismatch with
639 the approximations computed using Algorithm 2.2. The reason for this discrepancy
640 is that Algorithm 2.2 uses only the moments up to second order (in this example) to
641 construct the Gram-Charlier approximations. This means that, effectively, we obtain
642 a Gaussian-like estimate of the density. Performance can be improved by increasing
643 the order of the polynomial approximation, at the expense of a higher computational
644 cost.
645   Despite the visual discrepancy in Figure 5.1, Table 5.5 shows that the total
646 variation distance (TVD) between the marginal densities estimated using Algorithm
647 2.2 and Monte Carlo (KDEs with 10^4 samples generated via Euler-Maruyama) is
648 small. Let us recall that the TVD between two probability distributions with pdf's
649 f and g can be computed as $\|f - g\|_{TV} = \frac{1}{2} \int_{\mathbb{R}} |f(x) - g(x)| \mathrm{d}x \leqslant 1$. We can see that
650 the TVD is particularly small for the estimators of the pdf's of the position variables
651 x and y –of order 10^{-5} and 10^{-6}, respectively. This can be expected from Figure
652 5.1 because both densities (of x and y) are very spread, with maximum values of
653 order 10^{-3} (for x) and 10^{-4} (for y). Hence the tails accumulate a large fraction of
654 the probability mass and this is well approximated by Algorithm 2.2, despite the
655 discrepancy around the mode for the pdf of x. A similar argument can be made for
656 the densities of v_x and v_y. While there is more probability mass around the modes
657 (the maximum values of the pdfs are ≈ 0.2 and ≈ 0.8 for v_x and v_y, respectively),
658 which leads to higher discrepancies in TVD ($\approx 10^{-3}$ and $\approx 10^{-2}$, respectively), the

---

[3]We use the `ksdensity` function available in Matlab, which determines the kernel bandwidth for the estimator automatically from the samples.
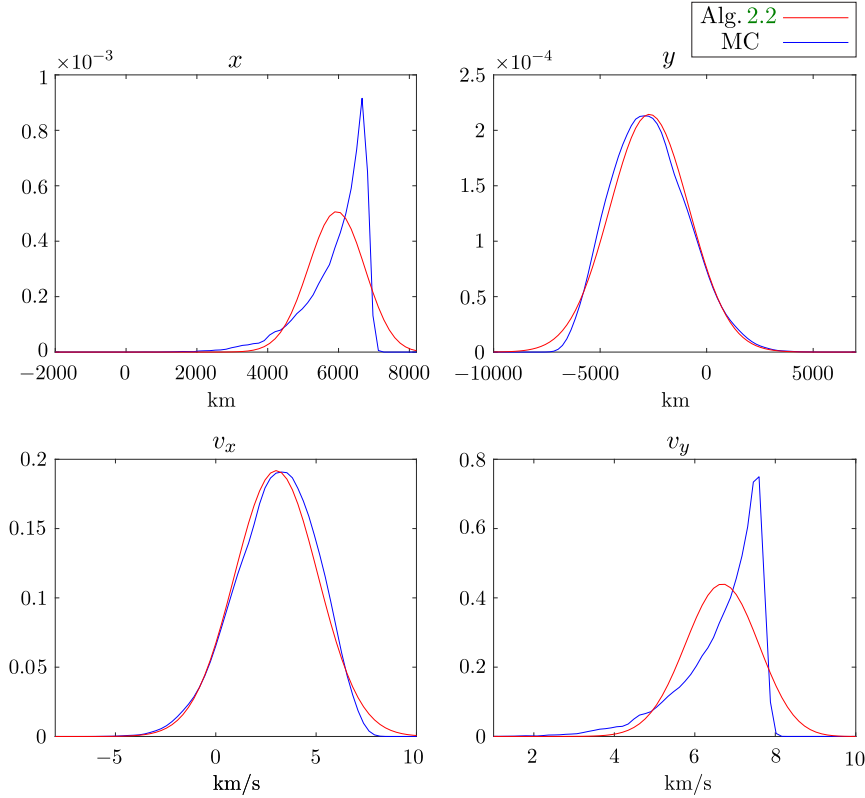
Fig. 5.1: Marginal pdf's at the final time t_n when the initial condition is fixed. The red curves are the estimates obtained with Algorithm 2.2 and the blue curves are KDEs computed from 10^4 independent samples.

tails still accumulate a large fraction of the probability mass and this is captured by Algorithm 2.2.

5.2. Random initial condition. We illustrate the performance of the proposed approximation methods for the same dynamical model in Section 5.1, except that we now assume a random initial condition $\boldsymbol{X}_0$. This is modelled as a Gaussian random vector with mean $\boldsymbol{x}_0$, the same as in Eq. (5.2), and covariance matrix

$$(5.3) \qquad \boldsymbol{\Sigma}_0 = \begin{pmatrix} 10^{-1} & 0 & 0 & 0 \\ 0 & 10^{-1} & 0 & 0 \\ 0 & 0 & 10^{-4} & 0 \\ 0 & 0 & 0 & 10^{-4} \end{pmatrix}.$$

The computation of moments is carried out using Algorithm 2.1, with a PCE of order $N_{\mathrm{PCE}} = 4$ and $N_s = 140$ samples. We emphasise that the PCE is not needed when the initial condition is deterministic.

The computer experiments are similar to Section 5.1. In particular:
- Table 5.6 shows a comparison of the expected values of the state variables (x, y, v_x and v_y) as obtained though Algorithm 2.1 and the baseline Monte Carlo method with 10^4 independent trajectories. Both the (approximate)

expectations, computed by the two methods, and the absolute and relative differences are displayed. We observe small relative differences, the largest ones $\approx 1\%$ for variables y and v_x, while for x and v_y they are less than 0.02%.

- Table 5.7 shows a comparison of the second-order moment estimates, also at time t_n, using Algorithm 2.1 and standard Monte Carlo with 10^4 independent runs. The entry-wise relative differences between the two matrices, displayed in the fourth row of the table, shows similar discrepancies for all entries of the covariance matrix, ranging between 1% and 8%.

- Table 5.8 displays a comparison of the computational cost of Algorithm 2.1 (with $N_{PCE} = 4$ and $N_s = 140$) and the baseline Monte Carlo procedure (with 10^4 independent runs). We see that the overall run-time of Algorithm 2.1 for this example is one order of magnitude smaller than the Monte Carlo scheme, while both methods yield very similar approximations of the first and second order moments.

- Table 5.9 shows the TVD between the marginal densities estimated using Algorithm 2.2 and KDEs computed from the independent samples generated by running the Euler-Maruyama scheme (2.1) 10^4 times with random and independent initialisations. The results are similar to those obtained in Section 5.1. The discrepancies are very small for the position variables and larger, yet very moderate ($< 3 \times 10^{-2}$), for the velocity variables.

- Figure 5.2 displays a graphical comparison of the marginal densities approximated via Algorithm 2.2 and Monte Carlo KDEs. We observe that all densities are clearly non-Gaussian. Since the initial condition $\boldsymbol{X}_0$ is random, the density approximations of Algorithm 2.2 are computed using Eq. (2.28), which enables the approximation of non-Gaussian pdf's by averaging many Gram–Charlier estimates.

  However, we observe that the resulting estimates are not necessarily proper densities, as they can take negative values. This is a drawback of the Gram–Charlier expansion compared, e.g., to KDEs. We have chosen the final time, t_n, of the simulation to make sure that the marginal densities depart significantly from the initial Gaussian distribution at time t_0 and show the artefacts in the tails of the densities of x and v_y. The accuracy of the approximation can be improved by increasing the order of the polynomial expansion in Algorithm 2.1 ($N = 2$ for this example) and the order of the PCE ($N_{\mathrm{PCE}} = 4$ here) at the expense of an increased computational complexity.

	Algorithm 2.1	Monte Carlo, 10^4 samples	Absolute difference	Relative error difference
x	$5.09 \times 10^3 \,\mathrm{km}$	$5.09 \times 10^3 \,\mathrm{km}$	$6.47 \times 10^{-1} \,\mathrm{km}$	1.27×10^{-4}
y	$-2.32 \times 10^3 \,\mathrm{km}$	$-2.34 \times 10^3 \,\mathrm{km}$	$2.42 \times 10^1 \,\mathrm{km}$	1.03×10^{-2}
v_x	$2.59 \,\mathrm{km/s}$	$2.62 \,\mathrm{km/s}$	$2.75 \times 10^{-2} \,\mathrm{km/s}$	1.05×10^{-2}
v_y	$5.72 \,\mathrm{km/s}$	$5.72 \,\mathrm{km/s}$	$2.22 \times 10^{-4} \,\mathrm{km/s}$	3.89×10^{-5}

Table 5.6: Estimate of $\mathbb{E}[\boldsymbol{X}(t_n)]$ with the moment-computation Algorithm 2.1, compared with standard Monte Carlo estimates. The initial condition $\boldsymbol{X}_0$ is a Gaussian random vector.

Algorithm 2.1	$\begin{pmatrix} 4.46 \times 10^6 & 4.17 \times 10^6 & -4.65 \times 10^3 & 4.99 \times 10^3 \\ 4.17 \times 10^6 & 1.07 \times 10^7 & -1.20 \times 10^4 & 4.67 \times 10^3 \\ -4.65 \times 10^3 & -1.20 \times 10^4 & 1.35 \times 10^1 & -5.20 \\ 4.99 \times 10^3 & 4.67 \times 10^3 & -5.20 & 5.59 \end{pmatrix}$
Monte Carlo	$\begin{pmatrix} 4.24 \times 10^6 & 3.89 \times 10^6 & -4.32 \times 10^3 & 4.74 \times 10^3 \\ 3.89 \times 10^6 & 1.09 \times 10^7 & -1.22 \times 10^4 & 4.35 \times 10^3 \\ -4.32 \times 10^3 & -1.22 \times 10^4 & 1.36 \times 10^1 & -4.84 \\ 4.74 \times 10^3 & 4.35 \times 10^3 & -4.84 & 5.29 \end{pmatrix}$
Absolute differences	$\begin{pmatrix} 2.26 \times 10^5 & 2.89 \times 10^5 & 3.23 \times 10^2 & 2.59 \times 10^2 \\ 2.89 \times 10^5 & 1.14 \times 10^5 & 1.30 \times 10^2 & 3.27 \times 10^2 \\ 3.23 \times 10^2 & 1.30 \times 10^2 & 1.48 \times 10^{-1} & 3.65 \times 10^{-1} \\ 2.59 \times 10^2 & 3.27 \times 10^2 & 3.65 \times 10^{-1} & 2.96 \times 10^{-1} \end{pmatrix}$
Relative differences	$\begin{pmatrix} 5.34 \times 10^{-2} & 7.44 \times 10^{-2} & 7.48 \times 10^{-2} & 5.47 \times 10^{-2} \\ 7.44 \times 10^{-2} & 1.05 \times 10^{-2} & 1.07 \times 10^{-2} & 7.51 \times 10^{-2} \\ 7.48 \times 10^{-2} & 1.07 \times 10^{-2} & 1.08 \times 10^{-2} & 7.55 \times 10^{-2} \\ 5.47 \times 10^{-2} & 7.51 \times 10^{-2} & 7.55 \times 10^{-2} & 5.60 \times 10^{-2} \end{pmatrix}$

Table 5.7: Estimates of the covariance matrix of $\boldsymbol{X}(t_n)$ computed via Algorithm 2.1 and standard Monte Carlo, with 10^4 independent samples. The initial condition $\boldsymbol{X}_0$ is random.

	Algorithm 2.1	Monte Carlo, 10^4 samples	Euler-Maruyama, single run
Mean run-time per step	2.12×10^{-4} s	1.75×10^{-3} s	4.16×10^{-6} s
Total run-time	2.76×10^2 s	2.29×10^3 s	5.43 s

Table 5.8: Run-times in seconds (s) with random initial condition ($N_{\mathrm{PCE}} = 4$, $N_s = 140$). The total number of discrete time steps is $n = 1,304,640$.

x	y	v_x	v_y
1.97×10^{-5}	2.75×10^{-6}	1.23×10^{-3}	2.62×10^{-2}

Table 5.9: Total variation distance when initial condition is a Gaussian random vector. Algorithm 2.2 has truncation order $N_{\mathrm{PCE}} = 4$ for the PCE and $N_s = 140$ samples for the approximation. The Monte Carlo baseline KDEs are constructed from 10^4 independent samples.
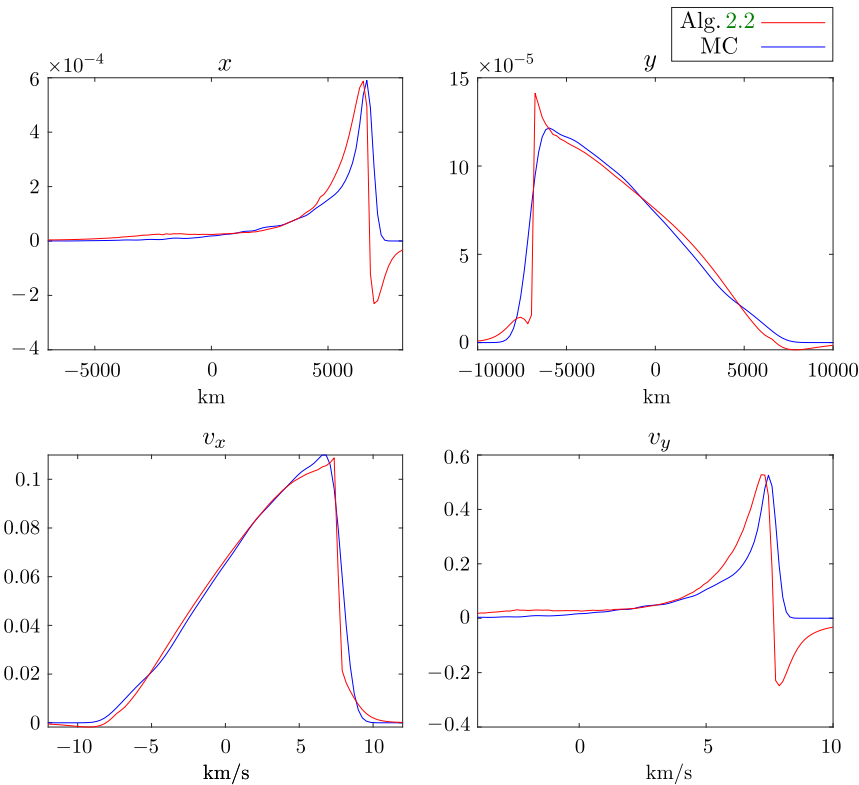
Fig. 5.2: Marginal pdf's at the final time t_n when the initial condition is Gaussian. The red curves are the estimates obtained with Algorithm 2.2 ($N_{\mathrm{PCE}} = 4$, $N_s = 140$) and the blue curves are KDEs computed from 10^4 independent samples.

6. Conclusions. We have introduced a methodology for the computation of the moments of the numerical solution of a multidimensional SDE, denoted $\hat{\boldsymbol{X}}_n$, using truncated Taylor polynomial approximations. The core of the method is the decomposition of the solution $\hat{\boldsymbol{X}}_n$ into a central part that can be computed deterministically from an ODE using an explicit numerical scheme and an effective noise process, whose moments determine the characterisation of $\hat{\boldsymbol{X}}_n$.

While we have derived the algorithm based on an Euler-Maruyama numerical scheme, the same ideas can be extended in a rather straightforward way to other explicit schemes, such as stochastic Runge-Kutta methods. When the initial condition is deterministic, the proposed algorithm involves a single run of the Euler-Maruyama numerical scheme (plus minor additional computations for the moments) and attains approximately the same performance as a Monte Carlo scheme with 10^4 independent runs of the Euler-Maruyama scheme. When the initial condition is random, we resort to a PCE scheme and still attain the same performance as the standard Monte Carlo estimators of the mean and second order moments with just a fraction (approximately 10%) of the run-time for a problem involving the propagation of uncertainty in a 2-dimensional Keplerian orbit.

We have also shown how to use the approximate moments of the numerical solution to compute type A Gram–Charlier estimates of the 1-dimensional marginal

pdf's of the dynamical variables. When the initial condition is random, a PCE scheme combined with a simple averaging enables the approximation of non-Gaussian densities. The Gram–Charlier expansions, however, are not guaranteed to yield proper probability densities and we have shown, numerically, that they can take negative values on the tails of the estimated function. This can be an important drawback. Improving the accuracy of the density estimators is conceptually straightforward (by computing higher order moments of the distribution) but implies an increase in the computational complexity of the numerical procedure.

The implementation of the algorithms as they have been presented demands the a priori calculation of the partial derivatives of the drift and diffusion coefficients of the SDE. Although it has not been explored in this paper, such calculations can be implemented automatically in the numerical scheme resorting to the tools of Taylor differential algebra [29]. As future work, we intend to explore the numerical performance of the method for a more realistic model of the orbital dynamics in 3-dimensional space and including the effect of the atmospheric drag in the drift. A realistic representation of this drag can be obtained from the NRLMSISE-00 model of the atmospheric density [19]. Such model brings in two relevant features that are not present in the simpler example of Section 5: first, the atmospheric density output by the NRLMSISE-00 model is time-inhomogeneous (while the drift in the current example is time-homogeneous) and, second, the partial derivatives of the atmospheric density have to be computed numerically (NRLMSISE-00 is a computer, non-algebraic, model). The latter approximation is not accounted for by our analysis, hence a numerical study is of interest.

REFERENCES

[1] M. ABBAS AND S. BOUROUBI, *On new identities for Bell's polynomials*, Discrete Mathematics, 293 (2005), pp. 5–10, https://doi.org/10.1016/j.disc.2004.08.023.
[2] M. BERZ, *Advances in Imaging and Electron Physics: Modern map methods in particle beam physics, Volume 108*, Academic Press, USA, 1999.
[3] A. BESKOS, D. CRISAN, A. JASRA, K. KAMATANI, AND Y. ZHOU, *On the stability of sequential Monte Carlo methods in high dimensions*, The Annals of Applied Probability, 24 (2014), pp. 1396–1445.
[4] A. BESKOS, A. JASRA, K. LAW, R. TEMPONE, AND Y. ZHOU, *Multilevel sequential Monte Carlo samplers*, Stochastic Processes and their Applications, 127 (2017), pp. 1417–1440.
[5] S. BRENNER AND L. SCOTT, *The Mathematical Theory of Finite Element Methods*, Springer–Verlag, New York, 2008.
[6] T. CHIHARA, *An introduction to Orthogonal Polynomials*, Dover., New York, 2011.
[7] J. DEMMEL, *The Geometry of Ill-Conditioning*, Journal of Complexity, 3 (1987), pp. 201–229, https://doi.org/10.1016/0885-064X(87)90027-6.
[8] T. C. GARD, *Introduction to stochastic differential equations*, M. Dekker, 1988.
[9] L. GOLDSTEIN, *Bounds on the constant in the mean central limit theorem*, Annals of Probability, 38 (2010), pp. 1672–1689, https://doi.org/10.1214/10-AOP527.
[10] M. HUTZENTHALER, A. JENTZEN, P. E. KLOEDEN, ET AL., *Strong convergence of an explicit numerical method for sdes with nonglobally lipschitz continuous coefficients*, Annals of Applied Probability, 22 (2012), pp. 1611–1641.
[11] B. KLARTAG AND S. SODIN, *Variations on the Berry–Esseen Theorem*, Theory of Probability and its Applications, 56 (2012), pp. 403–419, https://doi.org/10.1137/S0040585X97985522.
[12] P. E. KLOEDEN AND A. NEUENKIRCH, *The pathwise convergence of approximation schemes for stochastic differential equations*, LMS journal of Computation and Mathematics, 10 (2007), pp. 235–253.
[13] P. E. KLOEDEN AND E. PLATEN, *Numerical Solution of Stochastic Differential Equations*, Springer, 1992.
[14] J. E. KOLASSA, *Series Approximation Methods in Statistics*, Springer–Verlag, New York, 1997.
[15] A. LÓPEZ-YELA AND J. M. PÉREZ-PARDO, *Finite element method to solve spectral problem for arbitrary self-adjoint extensions of the Laplace–Beltrami operator on manifolds with a*

boundary, Journal of Computational Physics, 347 (2017), pp. 235–260, https://doi.org/10.1016/j.jcp.2017.06.043.

[16] Y. Luo and Z. Yang, *A review of uncertainty propagation in orbital mechanics*, Progress in Aerospace Sciences, 89 (2017), pp. 23–29, https://doi.org/10.1016/j.paerosci.2016.12.002.

[17] X. Mao and L. Szpruch, *Strong convergence rates for backward Euler–Maruyama method for non-linear dissipative-type stochastic differential equations with super-linear diffusion coefficients*, Stochastics An International Journal of Probability and Stochastic Processes, 85 (2013), pp. 144–171.

[18] B. Oksendal, *Stochastic differential equations: an introduction with applications*, Springer Science & Business Media, 2013.

[19] J. Picone, A. Hedin, D. P. Drob, and A. Aikin, *NRLMSISE-00 empirical model of the atmosphere: Statistical comparisons and scientific issues*, Journal of Geophysical Research: Space Physics, 107 (2002), pp. SIA–15.

[20] R. Rasala, *The Rodrigues Formula and Polynomial Differential Operators*, Journal of Mathematical Analysis and Applications, 84 (1981), pp. 443–482, https://doi.org/10.1016/0022-247X(81)90180-3.

[21] P. Rebeschini and R. V. Handel, *Can local particle filters beat the curse of dimensionality?*, The Annals of Applied Probability, 25 (2015), pp. 2809–2866.

[22] M. Reed and B. Simon, *Functional Analysis*, Academic Press, USA, 1980.

[23] L. E. Reichl, *A modern course in statistical physics. 2nd edition*, John Wiley and Sons, USA, 1998.

[24] H. Risken, *The Fokker–Planck equation: Methods of Solution and Applications*, Springer–Verlag, Berlin, 1989.

[25] H. Ruzayqat, A. Beskos, D. Crisan, A. Jasra, and N. Kantas, *Unbiased estimation using a class of diffusion processes*, arXiv preprint arXiv:2203.03013, (2022).

[26] G. Smith, *Numerical Solution of Partial Differential Equations: Finite Difference Methods*, Oxford University Press, New York, 1985.

[27] T. Tian and K. Burrage, *Implicit taylor methods for stiff stochastic differential equations*, Applied Numerical Mathematics, 38 (2001), pp. 167–185.

[28] D. Vallado, *Fundamentals of astrodynamics and applications*, Microscosm Press, California, 2007.

[29] M. Valli, R. Armellin, P. Di Lizia, and M. Lavagna, *Nonlinear mapping of uncertainties in celestial mechanics*, Journal of Guidance, Control, and Dynamics, 36 (2013), pp. 48–63.

[30] L. Wasserman, *All of Statistics: A Concise Course in Statistical Inference*, Springer–Verlag, New York, 2004.

[31] D. Xiu and G. E. Karniadakis, *The Wiener–Askey polynomial chaos for stochastic differential equations*, SIAM journal on scientific computing, 24 (2002), pp. 619–644.

[32] J. Yao and S. Gan, *Stability of the drift-implicit and double-implicit Milstein schemes for nonlinear SDEs*, Applied Mathematics and Computation, 339 (2018), pp. 294–301.