The Journal of the Astronautical Sciences (2026) 73:54
https://doi.org/10.1007/s40295-026-00593-z

RESEARCH

Sparse Approximation Method for Accurate Uncertainty Propagation through a Nonlinear System

Amit Jain[1] · **Puneet Singla**[1] · **Roshan Eapen**[1]

Received: 6 April 2025 / Accepted: 13 May 2026 / Published online: 12 June 2026
© The Author(s) 2026

Abstract
A computationally efficient approach is presented to propagate initial condition uncertainty, represented by a probability density function (PDF), through a nonlinear system. The log-PDF is assumed to be a linear combination of basis functions, and the amplitudes of these functions are computed by requiring the approximation to satisfy the Fokker-Planck-Kolmogorov equation (FPKE). Sparse approximation tools are utilized to trade off between FPKE error and the sparsity of the log-PDF model. The approach derives insights from known analytical stationary solutions of FPKE to construct multidimensional basis functions that accurately represent the state PDF. This judicious construction of basis functions helps alleviate the *curse of dimensionality* associated with the growth of basis functions in multidimensional space. Two nonlinear oscillators and a two-body problem are considered to demonstrate the efficacy of the proposed approach. Simulation results show that this approach effectively propagates uncertainty through both non-conservative and conservative systems.

Keywords Uncertainty propagation · Fokker-Planck-Kolmogorov equation · Hamiltonian systems · Probability density function

1 Introduction

The quantification and propagation of uncertainty in dynamic models are crucial for ensuring the precision and safety of engineering systems. Uncertainty propagation (UP) is a process that assesses and understands how uncertainties can affect model predictions and outcomes, playing a vital role in decision-making processes. The uncertainties associated with the state vector are typically characterized by statistical

✉ Amit Jain
amitjain@mit.edu

Puneet Singla
psingla@psu.edu

Roshan Eapen
reapen@psu.edu

[1] Department of Aerospace Engineering, Pennsylvania State University, State College, PA 16802, USA

properties such as probability distributions, moments, and probability density functions (PDFs). The state transition PDF of a nonlinear dynamical system is governed by a well-known Fokker-Planck-Kolmogorov Equation (FPKE), which is a convection-diffusion partial differential equation [1]. While closed-form solutions for the FPKE exist for linear systems under the influence of additive white Gaussian noise, finding exact solutions for a general nonlinear system is not feasible [1–4]. Therefore, approximate numerical solution procedures have been adopted for solving the FPKE.

Numerical solutions for the FPKE can be broadly categorized into sample-based, grid-based, and hybrid methods. Sample-based methods, such as the Monte Carlo (MC) method and deterministic particle-based methods, approximate the PDF by generating samples in the domain of interest and analyzing their behavior over time. The MC method involves evaluating the governing equation with a large number of random realizations of uncertain parameters and deriving statistics from the outputs [5, 6]. Other sample-based techniques known as quadrature methods, such as Gaussian Quadrature, Unscented transform (UT), and Conjugate Unscented Transform (CUT), have been developed as deterministic alternatives to random sampling schemes [7–11].

Grid-based methods discretize the spatial and temporal domains to solve the FPKE on a prescribed grid. The finite difference method, finite element method, finite volume method, and method of weighted residuals (MWR) are commonly used grid-based techniques. Finite difference methods approximate spatial derivatives of the PDF using explicit or implicit schemes, while finite element methods handle complex geometries and variable spatial resolutions in theory [12–16]. The finite volume method divides the problem domain into a grid or mesh, with grid cells representing volumes used to approximate the PDF evolution [17–19]. The class of MWR-based techniques encompasses several widely utilized grid-based approaches, including Galerkin, least-squares, method of moments, subdomain method, and pseudo-spectral/collocation-based methods [20]. The selection of weights used to satisfy the residual equation differentiates various MWR schemes. In the collocation method, the weights are chosen as Dirac delta functions, and the problem reduces to evaluating the residual to zero at the selected collocation points. Due to its ease of use, the collocation approach is a widely employed technique [21–24]. The main challenge in the implementation of the grid-based method lies in the discretization of a higher dimension state space since the number of spatial coordinates is equal to the state dimension, which is twice the degree of freedom for the system. Furthermore, the boundary conditions for the FPKE are defined at $\pm\infty$, which poses additional challenges in the discretization.

Hybrid methods combine sample and grid-based methods along with other uncategorized techniques. These include variational methods, path integration, stochastic averaging, machine learning, tensor decomposition, and Gaussian mixture model (GMM) techniques. Variational methods formulate the problem as a variational principle and find approximate solutions through optimization [25–27]. The path integral method calculates the probability distribution function of a stochastic process by summing over all possible paths that the process can take [28–32]. Stochastic averaging techniques replace the random fluctuations in the original system with their statistical averages or expectations, resulting in simplified equations that describe the system's long-term behavior [25, 33]. In machine learning approaches, researchers apply neu-

ral networks and deep learning techniques to approximate solutions of FPKE. These methods leverage the ability of neural networks to learn complex patterns and relationships from data, enabling them to approximate the solution of the equation without explicitly solving it analytically [9, 34–37]. Tensor decomposition techniques approximate the solution as a low-rank tensor, thereby reducing computational complexity and storage requirements [38–40]. Methods based on GMM, such as adaptive-GMM and adaptive-splitting based GMM [41–46], approximate the state PDF by a finite sum of Gaussian density functions. Their accuracy depends on optimally determining the number and position of mixture model components.

Sample-based methods offer several advantages, including ease of implementation, no requirement for spatial domain discretization, and natural handling of infinite boundary conditions. However, these methods have significant drawbacks. Monte Carlo methods suffer from slow convergence rates (typically $\mathcal{O}(N^{-1/2})$ where N is the number of samples), requiring a large number of samples to achieve acceptable accuracy, which becomes computationally prohibitive for complex systems [5, 6]. Deterministic sampling methods like quadrature-based approaches can improve convergence [7], but still face challenges in capturing complex, multimodal distributions efficiently [9].

Thus, hybrid methods combine elements of different numerical techniques to strike a balance between accuracy and computational efficiency. These methods leverage the strengths of both grid-based approaches (systematic coverage of the domain, deterministic accuracy) and sample-based approaches (flexibility, no grid construction, natural boundary handling), making them versatile for a wide range of problems. Blended methods such as the meshless partition of unity finite element method (PUFEM) [47], tensor decomposition with Chebyshev spectral method [40], Fourier with Chebyshev spectral methods [48], collocation with finite difference and sparse approximation method [49] for finding the evolution of PDF often offer significant computational advantages over traditional methods to solve the FPKE. However, the key limitation in computing the solution to the FPKE lies in the *curse of dimensionality*, which stems from the discretization of the PDF domain and/or the computation of multidimensional projection integrals.

The main contribution of this work is to address the computational challenge in solving the FPKE. The proposed approach is inspired by the recently developed sparse-collocation method [50–52], which involves expanding the logarithm of the PDF using polynomial basis functions. The coefficients and order of the expansion are determined by satisfying the FPKE at specific collocation points. To overcome the combinatorial growth of multidimensional basis functions, this work proposes a novel method that leverages the insights from the conditions for the existence of stationary solutions of the FPKE to construct multidimensional basis functions. By taking advantage of this structure, the developed method aims to improve both the efficiency and accuracy of uncertainty propagation.

The structure of the paper is as follows: Sect. 2 introduces the FPKE, while Sect. 3 discusses the conditions for the existence of a stationary solution of the FPKE and introduces the sparse-collocation approach. Section 4 provides the numerical results corresponding to the solution of FPKE for two nonlinear oscillators and propagating

burn uncertainty for a spacecraft in an orbit around the Earth. Section 5 concludes with the summary of main results.

2 Problem Statement and FPKE Preliminaries

Let us consider the following nonlinear dynamical system with stochastic excitation:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t) + \mathbf{G}\Gamma(t) \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^n$ is the state vector, and $\mathbf{f}$ and $\mathbf{G}$ are nonlinear functions. $\Gamma(t) \in \mathbb{R}^m$ is a vector of independent Gaussian white noise processes with zero mean, and the noise strength is characterized by the covariance matrix $\mathbf{Q}(t) \in \mathbb{R}^{m\times m}$, i.e., $\mathbb{E}[\Gamma(t)\Gamma^T(s)] = \mathbf{Q}(t)\delta(t-s)$ with $\delta(\cdot)$ denoting the Dirac delta function [53]. It should be mentioned that Eq. (1) can be written in terms of following stochastic differential equation:

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + \mathbf{G}d\beta(t) \tag{2}$$

Here $d\Gamma(t)$ is the Brownian motion process and Eq. (2) is valid under the ideal white noise forcing assumption, $\Gamma(t) \approx \frac{d\beta(t)}{dt}$. The main objective of this work is to propagate the state PDF over time through the aforementioned dynamical system, the uncertainty in the system will be captured by computing the state PDF. To propagate the uncertainty through the aforementioned nonlinear system, the initial state of the system is described by the PDF, $p(\mathbf{x}, t_0)$, and the state PDF, $p(\mathbf{x}, t_k)$ needs to be computed. In the presence of additive white Gaussian noise, the Fokker-Planck-Kolmogorov Equation (FPKE) governs the propagation of the state PDF through the nonlinear dynamical system. The FPKE is a second-order, linear parabolic partial differential equation given as:

$$\begin{aligned}\frac{\partial p(\mathbf{x}, t)}{\partial t} = &- \frac{\partial p(\mathbf{x}, t)^T}{\partial \mathbf{x}}\mathbf{f}(\mathbf{x}, t) - p(\mathbf{x}, t) \cdot Tr\left[\frac{\partial \mathbf{f}(\mathbf{x}, t)}{\partial \mathbf{x}}\right] \\ &+ \frac{1}{2}\,\mathrm{Tr}\left[\mathbf{G}\mathbf{Q}(t)\mathbf{G}^T\frac{\partial^2 p(\mathbf{x}, t)}{\partial \mathbf{x}\partial \mathbf{x}^T}\right]\end{aligned} \tag{3}$$

Here $\mathbf{Q}(t)$ is the noise strength matrix from Eq. (1). The components in the first-order partial derivative with respect to the state, commonly referred to as the drift term, capture the mean of the propagated PDF deviating from the initial PDF. The second-order partial derivatives with respect to the state variable represent the diffusion term, which accounts for the spread of the PDF over the state space. Although the underlying stochastic differential equation Eq. (1) describes individual trajectories that are non-differentiable (in the Itô or Stratonovich sense), the FPKE Eq. (3) is a deterministic partial differential equation that governs the smooth evolution of the PDF $p(\mathbf{x}, t)$ over the entire state space [1, 53, 54].

Notice that in the absence of white noise in the system, i.e., $\Gamma(t) = 0$, the governing dynamics becomes,

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t) \tag{4}$$

Thus, the sole source of uncertainty is in the initial state since the diffusion matrix becomes zero. The resulting reduced FPKE is also known as the Liouville equation, which describes the time evolution of the PDF (or equivalently phase space density) in phase space. The Liouville equation can also be expressed in terms of the Hamiltonian function using the Poisson bracket as [27, 54]:

$$\frac{\partial p}{\partial t} = -\{p, \mathcal{H}\} \tag{5}$$

where $\mathcal{H}$ is the Hamiltonian of the dynamical system and $\{p, \mathcal{H}\}$ is the Poisson bracket of $\mathcal{H}$ and p. This equation states that the rate of change of the PDF with respect to time is determined by the Poisson bracket of the Hamiltonian $\mathcal{H}$ and p, which is defined as:

$$\{p, \mathcal{H}\} = \sum_{i=1}^{\bar{n}} \left(\frac{\partial p}{\partial q_i} \frac{\partial \mathcal{H}}{\partial p_i} - \frac{\partial p}{\partial p_i} \frac{\partial \mathcal{H}}{\partial q_i} \right), \qquad \bar{n} = \frac{n}{2} \tag{6}$$

where q_i is the configuration variable and p_i is the corresponding conjugate momenta of the state variable $\mathbf{x}$, i.e., ($\mathbf{x} = [\mathbf{q}, \mathbf{p}]$). Thus, the Hamiltonian $\mathcal{H}(\mathbf{q}, \mathbf{p})$ corresponds to the total energy of the system and is governed through the Hamilton's equations of motion:

$$\frac{dq_i}{dt} = \frac{\partial \mathcal{H}}{\partial p_i}; \qquad \frac{dp_i}{dt} = -\frac{\partial \mathcal{H}}{\partial q_i} \tag{7}$$

Moreover, the characteristic solution of the Liouville equation is given by,

$$p(\mathbf{x}_t) = p(\mathbf{x}_0) \exp\left(- \int_0^t \nabla \cdot \mathbf{f}(\mathbf{x}(\tau), \tau) d\tau \right) \tag{8}$$

This provides a method for calculating the evolution of a PDF along a specific trajectory.

Although the FPKE (Eq. (3)) is linear in state PDF, it is a formidable equation to solve due to the following challenges:

1. Non-negativity of the PDF: $p(\mathbf{x}, t) \geq 0 \quad \forall \mathbf{x}, t$
2. Infinite Boundary Conditions: $p(\pm\infty, t) = 0$
3. Normality of the PDF: $\int p(\mathbf{x}, t) d\mathbf{x} = 1$
4. Number of spatial coordinates equals the state dimension n

These challenges restrict the applicability of numerical methods for solving the FPKE in engineering applications, such as the translational motion of a rigid body.

3 Proposed Methodology

In this section, a numerical method is developed while building upon our prior work [49] in this area. The main idea is to approximate the log-PDF through a linear combination of basis functions to ensure the non-negativity of the state PDF, i.e.,

$$\beta(\mathbf{x}, t) = \mathbf{c}^T(t)\mathbf{\Phi}(\mathbf{x}), \;\; p(\mathbf{x}, t) = e^{\beta(\mathbf{x},t)} \tag{9}$$

where $\beta(\mathbf{x}, t)$ denotes the log-PDF, and $\mathbf{c}(t) \in \mathbb{R}^m$ represents a vector of time-varying coefficients for the prescribed basis functions $\mathbf{\Phi}(\mathbf{x})$. To impose the infinite boundary conditions, the true PDF is regularized by a weighting function $W(\mathbf{x}, t, \theta)$ that must satisfy the following properties:

1. $W(\mathbf{x}, t, \boldsymbol{\theta}) \geq 0 \quad \forall \mathbf{x}, t, \boldsymbol{\theta}$
2. $W(-\infty, t, \theta) = W(\infty, t, \theta) = 0 \quad \forall t, \theta$

Here $\boldsymbol{\theta}$ is a parameter vector based on the chosen weighting function chosen. These two properties lead to the use of a PDF as the regularizing function. Note that the weight function has a similar form to the true PDF given in Eq. (9).

$$W(\mathbf{x}, t, \boldsymbol{\theta}) = e^{\beta_W(\mathbf{x},t,\boldsymbol{\theta})} \tag{10}$$

The weighting function is used to regularize the true PDF as follows:

$$p_A(\mathbf{x}, t) = p(\mathbf{x}, t)W(\mathbf{x}, t, \boldsymbol{\theta}) = e^{\beta(\mathbf{x},t)}e^{\beta_W(\mathbf{x},t,\boldsymbol{\theta})} = e^{(\beta(\mathbf{x},t)+\beta_W(\mathbf{x},t,\boldsymbol{\theta}))} = e^{\beta_A(\mathbf{x},t,\theta)} \tag{11}$$

where $p_A(\mathbf{x}, t)$ represents the approximate (regularized) PDF that satisfies the positivity and boundary conditions through the weighting function. The log-PDF can now be calculated using Eq. (11), satisfying both positivity and infinite boundary conditions. To handle the normality constraint, the PDF can be integrated over the domain of interest to find the normality constant. If the log-PDF is smooth with continuity and the existence of at least the first two derivatives, then $\beta_A(\mathbf{x}, t)$ can be represented as follows:

$$\beta_A(\mathbf{x}, t) = \beta(\mathbf{x}, t) + \beta_W(\mathbf{x}, \boldsymbol{\theta}) = \mathbf{c}^T(t)\mathbf{\Phi}(\mathbf{x}) + \mathbf{c}_W^T(\boldsymbol{\theta})\mathbf{\Phi}(\mathbf{x}) \tag{12}$$

where $\mathbf{c}_W(\boldsymbol{\theta})$ is a vector of coefficients for the weighting function. A key challenge in any approximation algorithm lies in choosing appropriate basis functions for accurate approximation of the state PDF. Although there are infinitely many choices for basis functions (polynomial functions, B-spline functions, wavelet functions, etc.), polynomial basis functions are most popular for representing the log-PDF [49, 55, 56]. This is because the principle of maximum entropy results in an exponential family (exponential of a polynomial function), and many known PDFs belong to this exponential family. However, as the state dimension increases, the number of polynomial basis functions grows rapidly. For instance, while a 1-D system requires only nine basis functions for eighth-order accuracy, higher dimensional systems require tensor products of polynomial basis functions to achieve the same accuracy. Figure 1 illustrates this combinatorial growth of polynomial basis functions (up to eighth-order) with system dimension. A 5-D system requires 1,287 basis functions, while a 6-D system needs 3,003 basis functions. Therefore, it is crucial to control this combinatorial growth. The following section investigates methods to prevent this explosive growth of basis functions in n-dimensional systems.

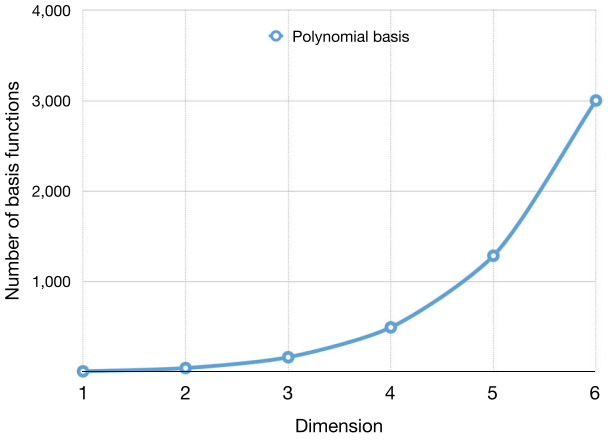


Fig. 1 Combinatorial growth of polynomial basis functions (up to eighth-order) with dimension

3.1 Selection of Basis Functions Dictionary

This section examines the stationary solution of the FPKE to construct appropriate basis functions for accurate approximation of the state PDF. To motivate the Hamiltonian formulation, both conservative and non-conservative systems are considered.

Consider an n-dimensional conservative system shown in Eq. (4), where the state variable $\mathbf{x}$ is split into a configuration variable $\mathbf{q}$ and its corresponding conjugate momenta $\mathbf{p}$ as:

$$\frac{dq_i}{dt} = f_i; \qquad \frac{dp_i}{dt} = f_{i+\bar{n}}; \qquad i = 1, \cdots, \bar{n} = \frac{n}{2} \tag{13}$$

As mentioned earlier, the system dynamics can also be described by Hamilton's equations of motion given in Eq. (7). Substituting the expressions for f_i in the FPKE of Eq. (3) in terms of the Hamiltonian function yields:

$$\begin{aligned}
\frac{\partial p(\mathbf{x},t)}{\partial t} &= -\sum_{i=1}^{\bar{n}} \frac{\partial \left(p f_i\right)}{\partial q_i} \\
&= -\sum_{i=1}^{\bar{n}} \frac{\partial}{\partial q_i}\left(p\frac{dq_i}{dt}\right) + \frac{\partial}{\partial p_i}\left(p\frac{dp_i}{dt}\right) \\
&= -\sum_{i=1}^{\bar{n}} \left(\frac{\partial}{\partial q_i}\left(p\frac{\partial \mathcal{H}}{\partial p_i}\right) - \frac{\partial}{\partial p_i}\left(p\frac{\partial \mathcal{H}}{\partial q_i}\right)\right) \\
&= -\sum_{i=1}^{\bar{n}} \left(\frac{\partial p}{\partial q_i}\frac{\partial \mathcal{H}}{\partial p_i} - \frac{\partial p}{\partial p_i}\frac{\partial \mathcal{H}}{\partial q_i} + p\frac{\partial^2 \mathcal{H}}{\partial q_i \partial p_i} - p\frac{\partial^2 \mathcal{H}}{\partial p_i \partial q_i}\right)
\end{aligned} \tag{14}$$

$$\frac{\partial p(\mathbf{x},t)}{\partial t} = -\sum_{i=1}^{\bar{n}} \left(\frac{\partial p}{\partial q_i}\frac{\partial \mathcal{H}}{\partial p_i} - \frac{\partial p}{\partial p_i}\frac{\partial \mathcal{H}}{\partial q_i}\right) \tag{15}$$

Note that Eq. (15) is equivalent to Eq. (5). When $p(\mathbf{x}, t) = p\left(\mathcal{H}\left(\mathbf{q}, \mathbf{p}\right)\right)$, Eq. (15) can be written as:

$$\begin{aligned} \frac{\partial p(\mathbf{x}, t)}{\partial t} &= -\sum_{i=1}^{\bar{n}} \left(\frac{\partial p}{\partial \mathcal{H}} \frac{\partial \mathcal{H}}{\partial q_i} \frac{\partial \mathcal{H}}{\partial p_i} - \frac{\partial p}{\partial \mathcal{H}} \frac{\partial \mathcal{H}}{\partial p_i} \frac{\partial \mathcal{H}}{\partial q_i} \right) \\ \frac{\partial p(\mathbf{x}, t)}{\partial t} &= 0 \end{aligned} \tag{16}$$

To summarize, for a conservative system ($\Gamma(t) = 0$), any function of $\mathcal{H}(\mathbf{q}, \mathbf{p})$ can serve as $p(\mathcal{H})$ provided that it satisfies normality and boundary conditions. Thus, the noise-free conservative system has infinitely many stationary PDFs, all of the form $p(\mathcal{H})$.

Next, consider a non-conservative system to demonstrate how the stationary PDF depends on the Hamiltonian. To illustrate this relationship, consider a nonlinear oscillator for which a stationary solution exists. In particular, the Duffing oscillator will be used to demonstrate the relationship between the stationary PDF and the Hamiltonian [57]. The system dynamics are given as follows:

$$\ddot{x} + \eta \dot{x} + \alpha x + \beta x^3 = Q \tag{17}$$

Notice that this system is a Hamiltonian-kind system, where external damping Q is present. (17) can be written in a state-space form as:

$$\dot{x}_1 = x_2; \qquad \dot{x}_2 = Q - \eta x_2 - \alpha x_1 - \beta x_1^3 \tag{18}$$

The equations of motion can be written using the Hamiltonian:

$$\begin{aligned} \mathcal{H} &= \frac{x_2^2}{2} + \frac{\alpha x_1^2}{2} + \frac{\beta x_1^4}{4} \\ \frac{dx_1}{dt} = \frac{\partial \mathcal{H}}{\partial x_2}; \qquad \frac{dx_2}{dt} &= -\frac{\partial \mathcal{H}}{\partial x_1} - \eta \frac{\partial \mathcal{H}}{\partial x_2} + Q \end{aligned} \tag{19}$$

The FPKE given by Eq. (3) for this system becomes:

$$-\frac{\partial}{\partial x_1}\left(p \frac{\partial \mathcal{H}}{\partial x_2}\right) + \frac{\partial}{\partial x_2}\left(p \frac{\partial \mathcal{H}}{\partial x_1}\right) + \eta \frac{\partial}{\partial x_2}\left(p \frac{\partial \mathcal{H}}{\partial x_2}\right) + \frac{1}{2} Q \frac{\partial^2 p}{\partial x_2^2} = 0 \tag{20}$$

The first two terms cancel each other out, as shown previously. Additional calculations using the remaining terms result in:

$$\begin{aligned} \eta \frac{\partial}{\partial x_2}\left(p \frac{\partial \mathcal{H}}{\partial x_2}\right) + \frac{1}{2} Q \frac{\partial^2 p}{\partial x_2^2} &= 0 \\ \eta p \frac{\partial \mathcal{H}}{\partial x_2} + \frac{1}{2} Q \frac{\partial p}{\partial x_2} &= C \text{ (constant)} \end{aligned} \tag{21}$$

Using $p(\pm\infty, t) = 0$

$$\eta p \frac{\partial \mathcal{H}}{\partial x_2} + \frac{1}{2} Q \frac{\partial p}{\partial x_2} = 0$$

Further simplifications lead to the following:

$$\begin{aligned} \eta p \frac{\partial \mathcal{H}}{\partial x_2} + \frac{1}{2} Q \frac{\partial p}{\partial \mathcal{H}} \frac{\partial \mathcal{H}}{\partial x_2} &= 0 \\ \eta p + \frac{1}{2} Q \frac{\partial p}{\partial \mathcal{H}} &= 0 \end{aligned} \tag{22}$$

$$p = p_0 \exp\left(-\frac{2\eta \mathcal{H}}{Q}\right) \tag{23}$$

This shows that the stationary PDF of a Duffing oscillator is a function of the Hamiltonian. Thus, for a non-conservative system expressed using Hamiltonian kind equations of motion, the stationary solution can be found. This motivates the inclusion of Hamiltonians, along with monomials, in the over-complete dictionary of basis functions.

3.1.1 Physics Informed Basis Functions

As demonstrated in the preceding section, the log-PDF of the non-conservative but Hamiltonian kind system, such as the Duffing oscillator, is directly proportional to the Hamiltonian function. This motivates the construction of basis functions based on the Hamiltonian, which is a multidimensional scalar function. For example, one can include powers of the Hamiltonian as basis functions, i.e.,

$$\Phi(\mathcal{H}) = [\mathcal{H}^1, \mathcal{H}^2, \mathcal{H}^3, \ldots, \mathcal{H}^{m_h}] \tag{24}$$

However, if the initial uncertainty is assumed to be Gaussian, then monomials in state variables are necessary to approximate the transient behavior of the log-PDF, while the Hamiltonian basis functions are necessary to approximate the stationary log-PDF. The complete basis function dictionary can be written as:

$$\mathbf{\Phi}(\mathbf{x}) = [\Phi(\mathbf{x}), \Phi(\mathcal{H}(\mathbf{x}))] = [\Phi(\mathbf{x}), \Phi(\mathcal{H})] \tag{25}$$

where $\Phi(\mathbf{x}) \in \mathbb{R}^{m_x}$ represents the dictionary of basis functions containing monomials, while $\Phi(\mathcal{H}(\mathbf{x})) \in \mathbb{R}^{m_h}$ denotes the dictionary of basis functions containing powers of the Hamiltonian. Additionally, $\mathbf{\Phi}(\mathbf{x}) \in \mathbb{R}^{m=m_x+m_h}$ signifies the complete dictionary of basis functions containing both monomials and powers of the Hamiltonian. Using Eq. (25), Eq. (12) can be written as:

$$\beta_A(\mathbf{x}, t) = (\mathbf{c}(t) + \mathbf{c}_W(\boldsymbol{\theta}))^T [\Phi(\mathbf{x}), \Phi(\mathcal{H})] \tag{26}$$

One can solve for unknown coefficients, $\mathbf{c}(t)$ in Eq. (26) by substituting the log-PDF into the FPKE and minimizing the resultant residual equation. Various numerical methods can be employed to minimize the global residual error when using a

series approximation for the log-PDF, including regularization techniques [58, 59], Galerkin-based approaches [60], the least squares method [61], and collocation methods [62]. The regularization methods, such as lasso and Ridge regression, offer distinct approaches to address this challenge. The lasso method promotes sparsity by penalizing the l_1-norm of coefficients, while Ridge regression uses the l_2-norm to handle multicollinearity. The Galerkin-based approach involves projecting the residual error onto a set of basis functions and minimizing the resulting weighted residuals. The least squares method seeks to minimize the sum of squared residuals between the observed data and the fitted model. Collocation-based methods aim to minimize the global residual error by reducing the problem to function evaluation at specified collocation or grid points. However, these methods can suffer from the *curse of dimensionality* when computing high-dimensional integrals and may even lead to over-fitting the data in some cases.

Thus, in this work, a sparse regression technique is utilized to solve the FPKE in a computationally attractive manner. The proposed approach exploits recent advances in non-product cubature methods in conjunction with careful selection of basis functions and sparse approximation tools to alleviate the *curse of dimensionality* to some extent. A recently developed quadrature scheme known as Conjugate Unscented Transform (CUT) provides a minimal set of cubature points. Furthermore, a minimal expansion for the log-PDF approximation is desired. If all coefficients participate in capturing the PDF, this can lead to overfitting of the data and consequently result in high testing errors. Thus, a tradeoff between sparsity and the two-norm error of FPKE is performed by utilizing sparse approximation techniques. These tools enable the automatic selection of dominant coefficients from the over-complete dictionary of basis functions, thereby preventing overfitting of the data.

In the following sections, the sparse regression method is described in detail. This approach involves a trade-off between FPKE error and sparsity of the log-PDF model to determine the optimal coefficients.

3.2 Sparse Regression

In this section, the numerical methodology for solving the FPKE using sparse regression is discussed. As mentioned in the previous section, the non-negativity and infinite boundary conditions are handled using the log-PDF and a weight function, as shown in Eq. (11). Substituting Eq. (11) into Eq. (3), the FPKE can be expressed in terms of the log-PDF as:

$$
\begin{aligned}
&\dot{\beta}_A(\mathbf{x},t,\boldsymbol{\theta}) + \frac{\partial \beta_A(\mathbf{x},t,\boldsymbol{\theta})^T}{\partial \boldsymbol{\theta}}\dot{\boldsymbol{\theta}} = -\operatorname{Tr}\left[\frac{\partial \mathbf{f}(\mathbf{x},t)}{\partial \mathbf{x}}\right] - \mathbf{f}(\mathbf{x},t)^T\left[\frac{\partial \beta_A(\mathbf{x},t,\boldsymbol{\theta})}{\partial \mathbf{x}}\right.\\
&\left. + \frac{\partial \boldsymbol{\theta}}{\partial \mathbf{x}}\frac{\partial \beta_A(\mathbf{x},t,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right] + \frac{1}{2}\operatorname{Tr}\left[\mathbf{G}\mathbf{Q}(t)\mathbf{G}^T\left(\frac{\partial^2 \beta_A(\mathbf{x},t,\boldsymbol{\theta})}{\partial \mathbf{x}\partial \mathbf{x}^T} + \frac{\partial \beta_A(\mathbf{x},t,\boldsymbol{\theta})}{\partial \mathbf{x}}\frac{\partial \beta_A(\mathbf{x},t,\boldsymbol{\theta})}{\partial \mathbf{x}^T}\right.\right.\\
&+ 2\frac{\partial \beta_A(\mathbf{x},t,\boldsymbol{\theta})}{\partial \mathbf{x}}\left(\frac{\partial \boldsymbol{\theta}}{\partial \mathbf{x}}\frac{\partial \beta_A(\mathbf{x},t,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T + \frac{\partial^2 \boldsymbol{\theta}}{\partial \mathbf{x}\partial \mathbf{x}^T}\frac{\partial \beta_A(\mathbf{x},t,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \frac{\partial^2 \beta_A(\mathbf{x},t,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \mathbf{x}^T}\frac{\partial \boldsymbol{\theta}}{\partial \mathbf{x}}\\
&\left.\left. + \left(\frac{\partial \boldsymbol{\theta}}{\partial \mathbf{x}}\frac{\partial \beta_A(\mathbf{x},t,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \boldsymbol{\theta}}{\partial \mathbf{x}}\frac{\partial \beta_A(\mathbf{x},t,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T\right)\right]
\end{aligned}
\tag{27}
$$

The time derivatives in Eq. (27) can be approximated using finite differences, and the weighting function is assumed to be a constant Gaussian kernel. Using Eq. (26), the residual error can be written as:

$$
\begin{aligned}
e(\mathbf{x},t) = \boldsymbol{\Phi}^T(\mathbf{x})\left(\frac{\mathbf{c}\left(t_{k+1}\right)-\mathbf{c}\left(t_k\right)}{\Delta t}\right) + \mathrm{Tr}\left[\frac{\partial \mathbf{f}\left(t_k,\mathbf{x}\right)}{\partial \mathbf{x}}\right] \\
+ \mathbf{f}^T\left(t_k,\mathbf{x}\right)\left[\frac{\partial \boldsymbol{\Phi}(\mathbf{x})^T}{\partial \mathbf{x}}\left(\mathbf{c}\left(t_k\right)+\mathbf{c}_W\right)\right] - \frac{1}{2}\,\mathrm{Tr}\left[\mathbf{G}(t_k)\mathbf{Q}(t_k)\mathbf{G}(t_k)^T\left(\left(\mathbf{c}\left(t_k\right)+\mathbf{c}_W\right)^T\frac{\partial \boldsymbol{\Phi}(\mathbf{x})}{\partial \mathbf{x}}\right.\right. \\
\left.\left.\frac{\partial \boldsymbol{\Phi}(\mathbf{x})}{\partial \mathbf{x}^T}\left(\mathbf{c}\left(t_k\right)+\mathbf{c}_W\right) + \frac{\partial^2}{\partial \mathbf{x}\partial \mathbf{x}^T}\left[\left(\mathbf{c}\left(t_k\right)+\mathbf{c}_W\right)^T\boldsymbol{\Phi}(\mathbf{x})\right]\right)\right]
\end{aligned}
\tag{28}
$$

where, $\dfrac{\partial \boldsymbol{\Phi}(\mathbf{x})}{\partial \mathbf{x}} = \left[\dfrac{\partial \Phi(\mathbf{x})}{\partial \mathbf{x}}, \dfrac{\partial \Phi(\mathcal{H})}{\partial \mathbf{x}}\right] = \left[\dfrac{\partial \Phi(\mathbf{x})}{\partial \mathbf{x}}, \dfrac{\partial \Phi(\mathcal{H})}{\partial \mathcal{H}}\dfrac{\partial \mathcal{H}}{\partial \mathbf{x}}\right]$. From this point forward, $\mathbf{c}\left(t_k\right)$ will be denoted as $\mathbf{c}_k$ for notational simplicity. Moreover, Eq. (28) can be expressed more compactly using the unknown coefficients $\mathbf{c}_{k+1}$ and the known quantities $\mathbf{A}$, b as:

$$
e(\mathbf{x},t) = \mathbf{A}\mathbf{c}_{k+1} - b \tag{29}
$$

where,

$$
\begin{aligned}
\mathbf{A}^j &= \boldsymbol{\Phi}_j^T(\mathbf{x})/\Delta t \\
b &= \boldsymbol{\Phi}^T(\mathbf{x})\,\mathbf{c}_k/\Delta t - \mathrm{Tr}\left(\frac{\partial \mathbf{f}\left(t_k,\mathbf{x}\right)}{\partial \mathbf{x}}\right) + \mathbf{f}^T\left(t_k,\mathbf{x}\right)\left(\frac{\partial \boldsymbol{\Phi}^T(\mathbf{x})}{\partial \mathbf{x}}\left(\mathbf{c}_k+\mathbf{c}_W\right)\right) \\
&\quad -\frac{1}{2}\,\mathrm{Tr}\left\{\mathbf{G}(t_k)\mathbf{Q}(t_k)\mathbf{G}^T(t_k)\left(\left(\mathbf{c}_k+\mathbf{c}_W\right)^T\frac{\partial \boldsymbol{\Phi}(\mathbf{x})}{\partial \mathbf{x}}\frac{\partial \boldsymbol{\Phi}(\mathbf{x})}{\partial \mathbf{x}^T}\left(\mathbf{c}_k+\mathbf{c}_W\right)\right.\right. \\
&\quad \left.\left.+\frac{\partial^2}{\partial \mathbf{x}\partial \mathbf{x}^T}\left[\left(\mathbf{c}_k+\mathbf{c}_W\right)^T\boldsymbol{\Phi}(\mathbf{x})\right]\right)\right\} \qquad j = 1,2,\ldots,m
\end{aligned}
\tag{30}
$$

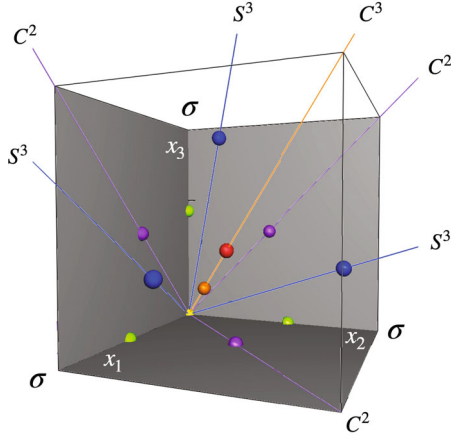
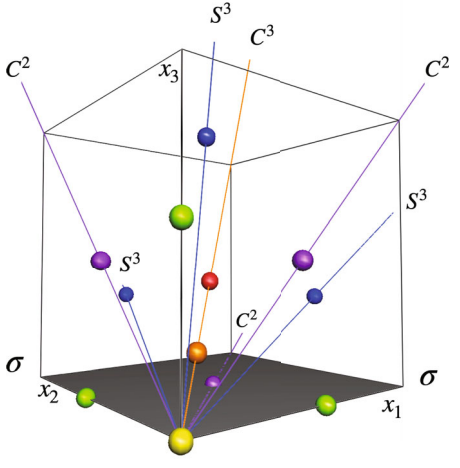
The aim is to find coefficients $\mathbf{c}_{k+1}$ that minimize the two-norm error over the given domain. Thus, the cost function can be written as:

$$
J = \frac{1}{2}\int e^2(\mathbf{x},t)\pi(\mathbf{x})d\mathbf{x} \tag{31}
$$

where $\pi(\mathbf{x})$ is the weighting function. To solve Eq. (31), a multidimensional integral needs to be evaluated. Quadrature methods, a class of numerical integration techniques, can be employed to approximate this integral. The choice of quadrature technique depends on various factors, including the dimensionality of the integral and the desired level of accuracy. Selecting an appropriate quadrature technique is crucial, as it can significantly impact the accuracy and computational efficiency of the solution. Therefore, the objective is to find the minimum number of quadrature points needed to accurately represent the entire domain while avoiding solution infeasibility in higher dimensions due to the *curse of dimensionality*.

Conventionally, Monte Carlo (MC) sampling is used to evaluate multi-dimensional integrals. However, the MC method suffers from a slow convergence rate. An alternative approach employs deterministic sampling techniques, namely quadrature

Table 1 Conjugate Unscented Transformation (CUT) points

	Position	Weights
$1 \leq i \leq 2n$	$\mathbf{x}_i = r_1 \sigma_i$	$\mathbf{w}_i = w_1$
$1 \leq i \leq 2^n$	$\mathbf{x}_{i+2n} = r_2 c_i^n$	$\mathbf{w}_{i+2n} = w_2$
Central weight	$\mathbf{x}_0 = 0$	$\mathbf{w}_0 = w_0$
		
3D 1^{st} Quadrant - view 1	3D 1^{st} Quadrant - view 2	

methods. Many methods exist for generating quadrature points, with the simplest being Gaussian quadrature. Gaussian quadrature methods provide a minimal number of points to integrate polynomial functions in one-dimensional (1-D) space. However, one needs to take a tensor product of these 1-D quadrature points to evaluate multidimensional integrals in d-D space. This results in a total of q^d quadrature points, where q is the number of quadrature points in 1-D. Thus, the number of points increases exponentially as the system's dimension increases [63, 64].

An efficient numerical sampling method known as the Conjugate Unscented Transformation (CUT) method [65] is chosen to prevent this exponential growth. This method exploits the structure of the domain by choosing specially defined axes on which quadrature points are selected. These CUT points are developed using minimal cubature rules and offer similar orders of accuracy as the Gaussian quadrature method but with fewer required points. In this method, points are chosen symmetrically at conjugate axes in addition to the principal axes. In 3-D, the conjugate axes correspond to face bisectors and principal diagonals of a cube. For an n-D system, $2n$ points of weight w_1 lie symmetrically on each principal axis at a distance of r_1, and 2^n points of weight w_2 lie symmetrically on the nth-conjugate axes (C^n) at a distance $\sqrt{n}r_2$, with a central weight placed at the origin. The configuration described above corresponds to the fourth-order accurate CUT (CUT4), where the order refers to the degree of polynomials that can be integrated exactly. Higher-order schemes such as CUT8 (eighth-order accurate) require additional points but provide improved accuracy for integrating higher-degree polynomials. Thus, for CUT4, the $(N \times n)$ set of quadrature points $\mathbf{x}$ ($N = 2n + 2^n + 1$) and the corresponding weights are given in Table 1 along with 3-D views for $n = 3$. More details about the CUT methodology and its comparison with conventional quadrature rules can be found in Ref. [8, 65–70].

Assuming there are a total of N CUT points, the n-dimensional integral of Eq. (31) can be approximated as:

$$J \approx \frac{1}{2} \sum_{i=1}^{N} \mathbf{w}_i e^2(\mathbf{x}_i, t) \tag{32}$$

where $\mathbf{x}_i$ and $\mathbf{w}_i$ are the CUT points and their corresponding weights. By satisfying the FPKE at these CUT points, the polynomial expansion ensures an accurate approximation across the domain.

Moreover, to approximate the polynomial coefficients of the log-PDF, the quadrature points must be transformed to a specific range. For example, the quadrature points can be mapped uniformly into a unit hypercube around the origin or to a domain with zero mean and identity covariance. This transformation is necessary because optimization techniques are extremely sensitive to magnitude and range fluctuations across all system dimensions, which can cause serious numerical problems if not properly managed. Assuming the global domain is known a priori, a linear transformation can be used to map the global domain of interest into a local domain within a hypercube as:

$$\mathbf{y} = \mathbf{T}_0 \mathbf{x} + \mathbf{B}_0 \tag{33}$$

where $\mathbf{T}_0$ is a scaling/transformation matrix and $\mathbf{B}_0$ is a translation vector that map the global domain $\mathbf{x}$ to the normalized local domain $\mathbf{y}$ (e.g., a unit hypercube). The

resultant system of equations will yield the coefficients of the log-PDF expansion in the local space, which can be mapped back to the global space using the following transformation of variables [53]:

$$p(\mathbf{x}, t) = p\left(\mathbf{y} = \mathbf{T}_0\mathbf{x} + \mathbf{B}_0, t\right) \left| \frac{\partial \mathbf{y}}{\partial \mathbf{x}} \right| \tag{34}$$

Another important consideration is that during problem initialization, the initial coefficients must be carefully selected to ensure that the initial PDF approximation aligns with the true initial condition PDF.

3.3 Selection of Optimal Coefficients

The optimal values of basis function amplitudes can be obtained by minimizing the l_2-norm (Euclidean norm) of the FPKE error:

$$\mathbf{c}_{k+1_{l_2}} = \min J = \frac{1}{2}\mathbf{e}^T\mathbf{we} = \frac{1}{2}\left\|\mathbf{w}(\mathbf{A}\mathbf{c}_{k+1} - \mathbf{b})\right\|_2 \tag{35}$$

Since $\mathbf{c}_{k+1_{l_2}}$ tends to utilize all possible coefficients from the dictionary of basis functions, it is not sparse. This work seeks a minimal polynomial expansion that guarantees sparsity for the log-PDF. Therefore, a weighted l_1-norm optimization problem is proposed to select the minimum possible coefficients from an over-complete dictionary of basis functions while specifying a bound on the two-norm error for the cost function. This approach allows sparse coefficients $\mathbf{c}_{k+1_s}$ to trade sparsity for approximation error, providing a more flexible option. Thus, the optimization problem can be stated in terms of the coefficients $\mathbf{c}_{k+1_s}$:

$$\begin{aligned} &\min_{\mathbf{c}_{k+1}} \left\|\mathbf{K}\mathbf{c}_{k+1}\right\|_1 \\ \text{subject to:} \quad & \\ &J = \frac{1}{2}\mathbf{e}^T\mathbf{we} \leq \epsilon_s \end{aligned} \tag{36}$$

where $\mathbf{K}$ is the coefficient weight matrix, and ϵ_s represents the soft constraint on the two-norm error. When the dictionary contains a large number of basis functions, the l_2-norm solution tends to utilize all available coefficients to achieve very high numerical accuracy (e.g., $O(10^{-18})$), which often leads to overfitting. The soft constraint ϵ_s provides flexibility in balancing approximation accuracy with sparsity, allowing the sparse solution to trade some accuracy for computational efficiency. Smaller values of ϵ_s enforce tighter accuracy but yield less sparse solutions, while larger values prioritize sparsity at the cost of slightly increased residual error. The choice of ϵ_s depends on the problem characteristics and the acceptable trade-off between accuracy and sparsity.

This iterative reweighting approach is inspired by methods in compressed sensing and sparse approximation [49, 71, 72], where iteratively reweighted l_1-norm minimization better approximates the non-convex l_0-norm (cardinality) problem by

progressively driving small coefficients toward zero. Each weighted l_1-norm subproblem is convex and can be solved efficiently using standard convex optimization solvers. However, this iterative reweighting approach is designed to approximate the non-convex l_0-norm (cardinality) minimization problem. Large coefficients (important basis functions) receive small weights and remain relatively unpenalized, while small coefficients receive increasingly large weights in subsequent iterations, promoting sparser solutions than standard l_1 minimization alone. The initial value of $\mathbf{K}$ can be set to 1 or based on any *a priori* knowledge, such as penalizing the higher-order coefficients or using information gained from the l_2-norm solution. For subsequent iterations, $\mathbf{K}$ can be updated to penalize the non-dominant coefficients using the solution of the previous iteration $\mathbf{c}^-_{k+1}$ and by introducing a small number η_s to prevent division by zero as $\mathbf{K} = \frac{1}{(\mathbf{c}^-_{k+1}+\eta_s)}$. Algorithm 1 describes the optimal procedure to select the minimal polynomial expansion of the log-PDF. The user-defined parameter Δ_s represents the desired difference between the two norms of consecutive coefficients. The algorithm converges to an optimal solution $\mathbf{c}_{k+1_s}$ when the difference between the two norms of consecutive coefficient values, δ, is less than Δ_s, typically requiring 5–10 iterations to reach the termination criterion for the problems considered in Sect. 4.

Algorithm 1 Iterative weighted l_1-norm optimization: $\mathbf{c}_{k+1_s}$ = OneNormCoefficient($\mathbf{A}, \mathbf{b}, \mathbf{w}, \mathbf{c}_{k+1_{l_2}}, \Delta_s, \epsilon_s, \eta_s$)

Input: $\mathbf{A}, \mathbf{b}, \mathbf{w}, \mathbf{c}_{k+1_{l_2}}, \Delta_s, \alpha, \epsilon_s, \eta_s$

Output: $\mathbf{c}_{k+1_s}$

1: **Initialization**: $\delta = 1$, choose: $\mathbf{K}_0 \propto \mathcal{O}(\text{basis})$ or $\mathbf{K}_0 = \frac{1}{(\mathbf{c}_{k+1_{l_2}}+\eta_s)}$ or $\mathbf{K}_0 = 1$

2: compute $\mathbf{c}^-_{k+1} = \min_{\mathbf{c}_{k+1}} \left\|\mathbf{K}\mathbf{c}_{k+1}\right\|_1$
subject to: $\left\|\mathbf{w}(\mathbf{A}\mathbf{c}_{k+1} - \mathbf{b})\right\|_2 \leq \epsilon_s$

3: **while** $\delta \geq \Delta_s$ **do**
4: Update $\mathbf{K} = \frac{1}{(\mathbf{c}^-_{k+1}+\eta_s)}$, to find $\mathbf{c}^+_{k+1} = \min_{\mathbf{c}_{k+1}} \left\|\mathbf{K}\mathbf{c}_{k+1}\right\|_1$
subject to: $\left\|\mathbf{w}(\mathbf{A}\mathbf{c}_{k+1} - \mathbf{b})\right\|_2 \leq \epsilon_s$

5: Compute $\delta = \|\mathbf{c}^+_{k+1} - \mathbf{c}^-_{k+1}\|_2$
6: $\mathbf{c}^-_{k+1} = \mathbf{c}^+_{k+1}$
7: **end while**
8: $\mathbf{c}_{k+1_s} = \mathbf{c}^-_{k+1}$

Algorithm 2 employs three coefficient computation methods at each time step: TwoNormCoefficient computes an initial l_2-norm solution by minimizing the weighted residual error, OneNormCoefficient (Algorithm 1) iteratively solves a weighted l_1-norm problem to promote sparsity, and RSCoefficient further reduces the solution to only the dominant coefficients. After obtaining the sparse coefficients $\mathbf{c}_{k+1_s}$, one can separate the dominant coefficients from the non-dominant ones using a user-defined coefficient threshold δ_{rs}. The non-dominant coefficients can be set to zero for computational efficiency. A new minimal representation of the basis functions $\mathbf{A}_{rs} \in \mathbb{R}^{N \times m_r}$ corresponding to the m_r non-zero coefficients can then be constructed.

Algorithm 2 Sparse-Regression Based Solution Methodology for the Fokker-Planck-Kolmogorov Equation

Input: $\mathbf{f}(\mathbf{x})$, $\mathbf{G}$, m basis functions $\mathbf{\Phi}(\mathbf{x})$, initial values of coefficients $\mathbf{c}\,(t_0)$, weight function coefficients $\mathbf{c}_W$, discretized time vector t, time step Δt, CUT cubature points $\mathbf{x}_i$ $i = 1, 2, \ldots, N$, weight matrix $\mathbf{w}$, initial coefficient weight matrix $\mathbf{K}_0$, update parameter η_s, sparse two-norm constraint ϵ_s, and coefficient threshold δ_{rs}

Output: $\mathbf{c}(t)$, $\beta(\mathbf{x}, t)$

1: **for** $t = 0, t < t_f, k = k + 1$ **do**
2: Compute matrix $\mathbf{A}$
3: Compute $\mathbf{b}$ using $\mathbf{c}_k$
4: $\mathbf{c}_{k+1_{l_2}} = \text{TwoNormCoefficient}(\mathbf{A}, \mathbf{b}, \mathbf{w})$
5: $\mathbf{c}_{k+1_s} = \text{OneNormCoefficient}(\mathbf{A}, \mathbf{b}, \mathbf{w}, \mathbf{c}_{k+1_{l_2}}, \Delta_s, \epsilon_s, \eta_s)$ (Algorithm 1)
6: $\mathbf{c}_{k+1} = \text{RSCoefficient}(\mathbf{c}_{k+1_s}, \mathbf{A}, \mathbf{b}, \delta_{rs})$
7: $\beta\left(\mathbf{x}, t_{k+1}\right) = \left(\mathbf{c}_{k+1} + \mathbf{c}_W\right)^T \mathbf{\Phi}(\mathbf{x})$
8: **end for**

The reduced sparse (RS) coefficients, $\mathbf{c}_{k+1} \in \mathbb{R}^{m_r}$, can be calculated by minimizing the l_2-norm error using the dominant basis functions:

$$\mathbf{c}_{k+1} = \mathbf{A}_{rs}^{\dagger}\mathbf{b} \tag{37}$$

where $\mathbf{A}_{rs}^{\dagger}$ is the pseudo-inverse. This minimal representation $\mathbf{c}_{k+1}$ is used to compute the state PDF and the marginalized state PDF.

This complete methodology calculates a minimal polynomial expression for the log-PDF up to the final time t_f. The l_1-norm minimization problem addressed in Algorithm 1 is convex and must be solved at each time instant to find the time-varying state PDF. Various efficient convex optimization solvers, including SeDuMi, SDPT3, Gurobi, MOSEK, and GLPK, are available in the CVX programming language to find the optimal solution [72, 73].

4 Numerical Simulations

In this section, numerical examples demonstrate the applicability of the proposed approach. Two nonlinear oscillators are examined as benchmark problems to evaluate the precision of the suggested methodology, as analytical solutions to their stationary FPKEs are known. Additionally, an orbit maneuver problem corresponding to satellite motion in a low-Earth orbit (LEO) is considered to demonstrate extensibility to higher-dimensional systems.

4.1 Duffing Oscillator

The first example corresponds to the noise-driven Duffing oscillator. The system dynamics are given in Table 2, where $\Gamma(t)$ is a Brownian motion process with correlation function $\mathbf{Q}\delta(t_1 - t_2)$. The Duffing oscillator with a soft spring ($\alpha\beta < 0$) and positive damping ($\eta > 0$) has a bimodal stationary solution to the FPKE. The modes

Table 2 Duffing Oscillator: Dynamics and Parameters

Dynamics	Parameters
$\ddot{x} + \eta\dot{x} + \alpha x + \beta x^3 = G\Gamma(t)$	$\mathbf{Q} = 1, G = 1, dt = 0.01s$
$p(\mathbf{x}, \infty) \propto \exp\left[-2\frac{\eta}{G^2 Q}\mathcal{H}\right]$	$p(\mathbf{x}, t_0) = \mathcal{N}(\mathbf{0}, \mathbf{I}), \mathbf{\Sigma}_w = \frac{1}{9} I_{2\times 2}$
$\mathcal{H} = \left(\frac{\alpha}{2}x^2 + \frac{\beta}{4}x^4 + \frac{1}{2}\dot{x}^2\right)$	$\epsilon_s = 1e-6, \eta_s = 1e-4, \delta_{rs} = 1e-2$
	$\Omega(\mathbf{x}) \in [-2, 2], N = 21$
	$\text{Basis}(\mathcal{O}(15)) \rightarrow m = 151 (m_x = 136, m_H = 15)$
	case - I $\eta = 10, \alpha = -1, \beta = 3\ , t_f = 100s$
	case - II $\eta = 10, \alpha = -1, \beta = 1\ , t_f = 150s$

of these bimodal solutions are concentrated at two stable equilibrium locations, and the stationary PDF is directly proportional to the exponent of the Hamiltonian, $\mathcal{H}$.

The initial PDF is assumed to be Gaussian with zero mean and identity covariance. The basis function dictionary consists of monomials and Hamiltonians up to $15th$ order, resulting in a total of $m = 151$ coefficients ($m_x = 136$ and $m_H = 15$). Furthermore, 21 CUT8 quadrature points (section 3.2) are generated to sample the domain accurately. The global domain is assumed to be $\mathbf{x} \in [-2, 2]$, and the solution domain is mapped to a hypercube centered at the origin of length 2. In the local domain, a zero-mean Gaussian weighting function is employed with a covariance of $\mathbf{\Sigma}_w = \frac{1}{9} I_{2\times 2}$ to ensure that the $\pm 3\sigma$ bounds of the weight function are at the domain boundaries. Two different cases of the Duffing oscillator are considered by varying β, as shown in Table 2.

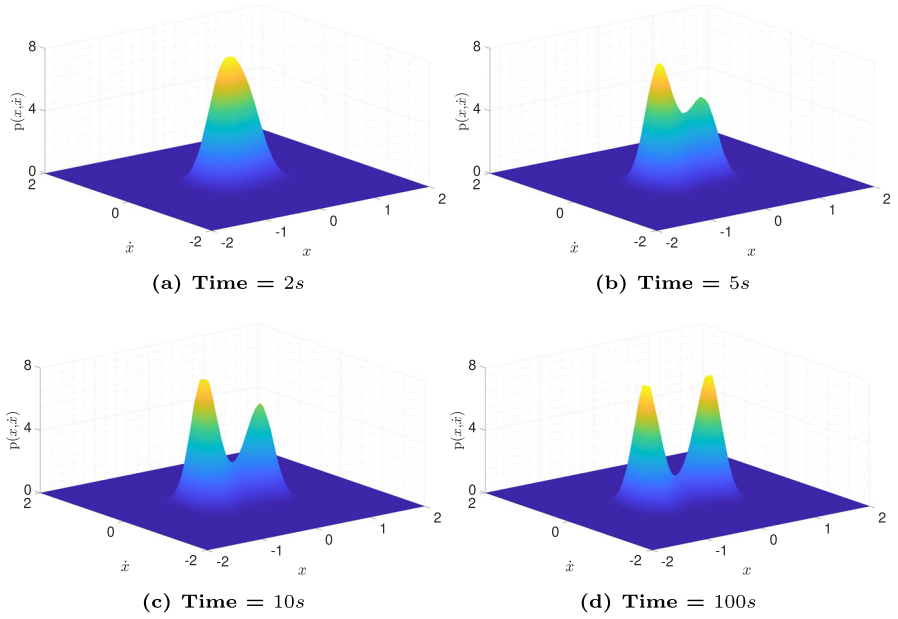
As derived in Eq. 23, the analytical stationary solution for the Duffing oscillator is:

$$\log(p_{\text{analytical}}) = \log(p_0) - \frac{2\eta}{Q}\mathcal{H} = \log(p_0) - 20\,\mathcal{H} \tag{38}$$

where the coefficient $-2\eta/Q = -20$ for the parameters $\eta = 0.1$ and $Q = 0.01$ used in both cases. This analytical form predicts that at steady state, the log-PDF should be dominated by a single Hamiltonian basis function with coefficient -20, while all monomial coefficients vanish.

For both cases, the uncertainty is propagated by examining the PDF surfaces at different time steps. Figure 2 shows the computed PDF surfaces at times $t = 2, 5, 10$ and $100s$ for the first case. As the initial PDF was assumed to be Gaussian, the PDF at time $t = 2s$ is nearly Gaussian. As time progresses, the bimodal shape begins to emerge at time $t = 5s$. Finally, at time $t = 100s$, the true bimodal shape of the Duffing oscillator is achieved. From this plot, the true stationary PDF is assumed to be given at $t = 100s$.

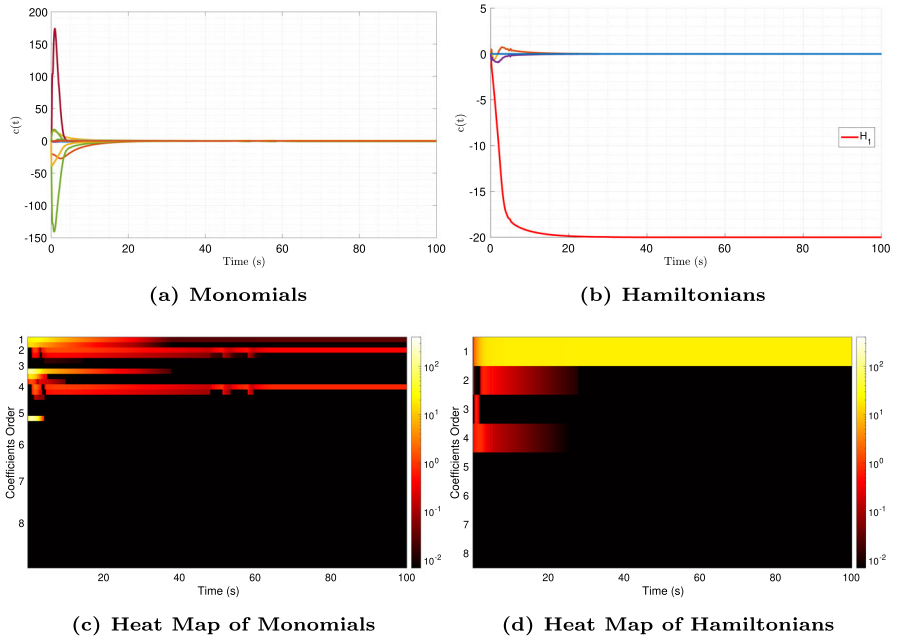
The convergence toward the analytical stationary solution can be studied by examining the evolution of coefficients over time. Figure 3 shows the evolution of RS coefficients for monomials and Hamiltonians in the first and second columns, respectively. From Fig. 3a and b, it can be observed that the transient behavior of the PDF is dictated mainly by monomials and a few higher-order Hamiltonians. However, as time progresses, the first-order Hamiltonian coefficient ($\mathcal{H}_1$) gradually becomes active and



(a) Time = $2s$ (b) Time = $5s$

(c) Time = $10s$ (d) Time = $100s$

Fig. 2 Duffing oscillator: case I - computed PDF surfaces



(a) Monomials (b) Hamiltonians

(c) Heat Map of Monomials (d) Heat Map of Hamiltonians

Fig. 3 Duffing oscillator: case I - RS coefficients evolution

Table 3 Quintic Oscillator: Dynamics and Parameters

Dynamics	Parameters
$\ddot{x} + \eta\dot{x} + x\left(\epsilon_1 + \epsilon_2 x^2 + \epsilon_3 x^4\right) = G\Gamma(t)$	$\eta = 10, \epsilon_1 = 1, \epsilon_2 = -3.065, \epsilon_3 = 1.825$
	$\mathbf{Q} = 1, G = 1$
$p(\mathbf{x}, \infty) \propto \exp\left[-2\frac{\eta}{G^2 Q}\mathcal{H}\right]$	$p(\mathbf{x}, t_0) = \mathcal{N}(\mathbf{0}, \mathbf{I}), \mathbf{\Sigma}_w = \frac{1}{9} I_{2\times 2}$
	$t_f = 150s, dt = 0.01s$
$\mathcal{H} = \left(\frac{\epsilon_1}{2}x^2 + \frac{\epsilon_2}{4}x^4 + \frac{\epsilon_3}{6}x^6 + \frac{1}{2}\dot{x}^2\right)$	$\eta_s = 1e-3, \epsilon_s = 1e-4$, and $\delta_{rs} = 1e-2$
	$\Omega(\mathbf{x}) \in [\pm 1.25, \pm 0.75], N = 21$
	Basis$(\mathcal{O}(15)) \rightarrow m = 151 (m_x = 136, m_H = 15)$

finally converges to a stationary value of -20, consistent with the analytical prediction $-2\eta/(G^2 Q)$. This behavior can also be observed in the heatmap of monomials and Hamiltonians, where the y-axis displays the coefficient order, the x-axis displays time, and the colormap shows the absolute value of the coefficient. Comparing Fig. 3c and d, the $4th$ and $6th$ order monomial coefficients dominate initially but gradually diminish as time progresses, approaching zero as predicted by the analytical solution.

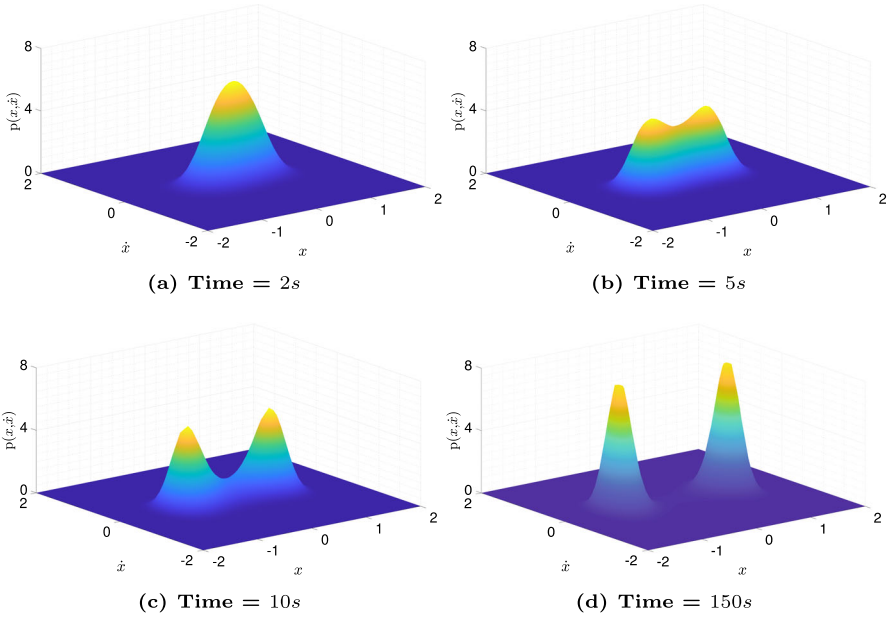
A similar comparison can be made for the second case, where $\beta = 1$ is assumed. Figure 4 depicts the computed PDF surfaces at times $t = 2, 5, 10, 150s$, with the true stationary PDF assumed at time $t = 150s$. The evolution of RS coefficients is shown in Fig. 5, where again the Hamiltonian coefficients approach the true stationary value as time increases. Notably, the solution methodology automatically selects the true Hamiltonian coefficient participating in the stationary solution from among 151 basis functions.

Finally, Fig. 6 shows the log-PDF ratio of Hamiltonians and monomials $(\beta(\mathcal{H})/\beta(\mathbf{x}))$ for both cases. This ratio is important as it indicates the relative contribution of Hamiltonian versus monomial basis functions in approximating the log-PDF. As time increases, the $\beta(\mathcal{H})/\beta(\mathbf{x})$ ratio grows, signifying the increasing dominance of Hamiltonian basis functions in capturing the log-PDF. The log-PDF ratio eventually stabilizes after $65s$ for case-I and $105s$ for case-II, indicating that the monomial coefficients are no longer active.

To quantitatively validate the convergence to the analytical solution, the numerical log-PDF at $t = 100s$ for case-I is:

$$\begin{aligned}\log(p_{\text{numerical}}) \approx & c_0 - 0.0994\, x_1 - 0.0284\, x_1^2 + 0.4569\, x_1^3 \\ & - 0.6027\, x_1^5 - 0.0107\, x_2^2 - 19.9946\, \mathcal{H}\end{aligned} \tag{39}$$

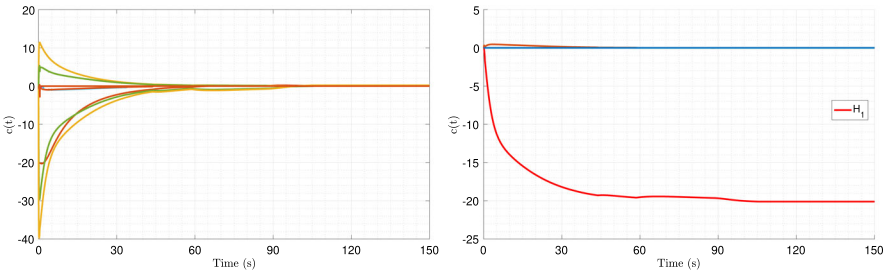
The Hamiltonian coefficient (-19.9946) agrees with the analytical value (-20) to within 0.027% error, while the remaining polynomial coefficients are small in magnitude (ranging from 0.01 to 0.6), representing residual transient effects. This confirms that the numerical solution accurately captures the analytical stationary form.



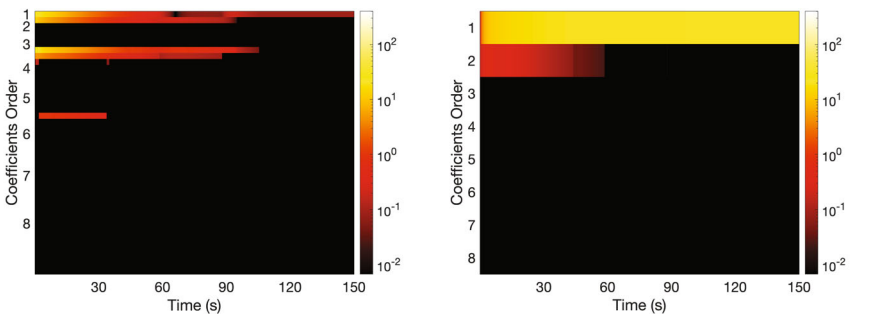
(a) Time = $2s$ (b) Time = $5s$

(c) Time = $10s$ (d) Time = $150s$

Fig. 4 Duffing oscillator: case II - computed PDF surfaces

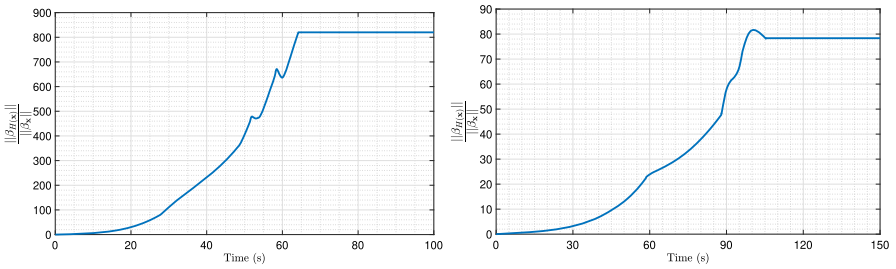


(a) Monomials (b) Hamiltonians



(c) Heat Map of Monomials (d) Heat Map of Hamiltonians

Fig. 5 Duffing oscillator: case II - RS coefficients evolution



(a) case - I (b) case - II

Fig. 6 Duffing oscillator: log-PDF ratio

4.2 Quintic Oscillator

The proposed methodology is also applied to the noise-driven Quintic oscillator [41]. The stationary PDF of the Quintic oscillator is trimodal, with modes centered at three equilibrium points. Similar to the Duffing oscillator, the stationary PDF is directly proportional to the exponent of the Hamiltonian, $\mathcal{H}$.

The global domain is chosen as $\mathbf{x} \in [-2, 2]$, and the initial PDF is assumed to be Gaussian with zero mean and identity covariance. The basis function dictionary consists of monomials and Hamiltonians up to $15th$ order. Higher-order polynomials are included to accurately capture the time-varying PDF and ensure a sparse solution. The system dynamics and parameters are given in Table 3.

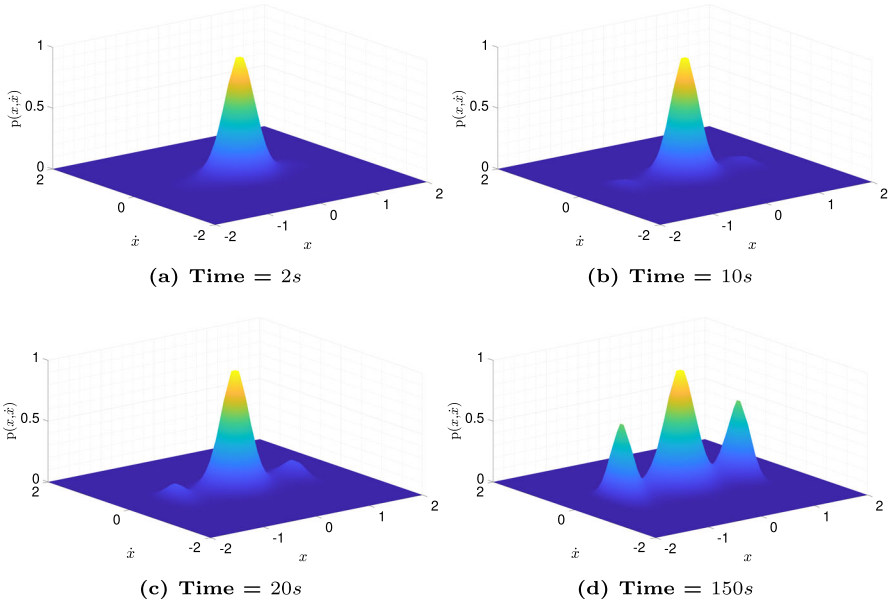
As with the Duffing oscillator, the analytical stationary solution for the Quintic oscillator has the form:

$$\log(p_{\text{analytical}}) = \log(p_0) - \frac{2\eta}{G^2 Q}\mathcal{H} = \log(p_0) - 20\,\mathcal{H} \tag{40}$$

where the coefficient is $-2\eta/(G^2 Q) = -20$ for the given parameters, predicting that the log-PDF should converge to a pure Hamiltonian form with coefficient -20 at steady state.

Figure 7 shows the computed PDF surfaces of the Quintic oscillator at times $t = 2, 10, 20$ and $150s$. Since the initial PDF was assumed to be Gaussian, the PDF at $t = 2s$ remains nearly Gaussian. As time progresses, the trimodal shape begins to emerge at $t = 10s$. Finally, at $t = 150s$, the PDF converges to its true stationary form. The convergence toward the analytical stationary solution can be studied further by examining the evolution of RS coefficients shown in Fig. 8. Similar to the previous example, the transient behavior of the PDF is governed by monomials (up to $\mathcal{O}(6)$) and a few Hamiltonians. Due to the symmetry of the true PDF, all odd-term monomials are absent in the RS coefficient dictionary, as seen in Fig. 10. As time progresses, all coefficients except the first-order Hamiltonian coefficient ($\mathcal{H}_1$) converge to zero. The first-order Hamiltonian coefficient converges to -20, consistent with the analytical value $-2\eta/G^2 Q$.

Moreover, Fig. 9 shows the log-PDF ratio of Hamiltonians and monomials ($\beta(\mathcal{H})/\beta(\mathbf{x})$). As time increases, this ratio grows, signifying the increasing dominance



(a) Time = $2s$ (b) Time = $10s$

(c) Time = $20s$ (d) Time = $150s$

Fig. 7 Quintic oscillator: computed PDF surfaces

of Hamiltonian basis functions in capturing the uncertainty. The ratio stabilizes after 135 s, indicating that all monomial coefficients have become inactive. This demonstrates that the sparse-regression scheme successfully captures the stationary log-PDF of the Quintic oscillator, which is proportional to the Hamiltonian.

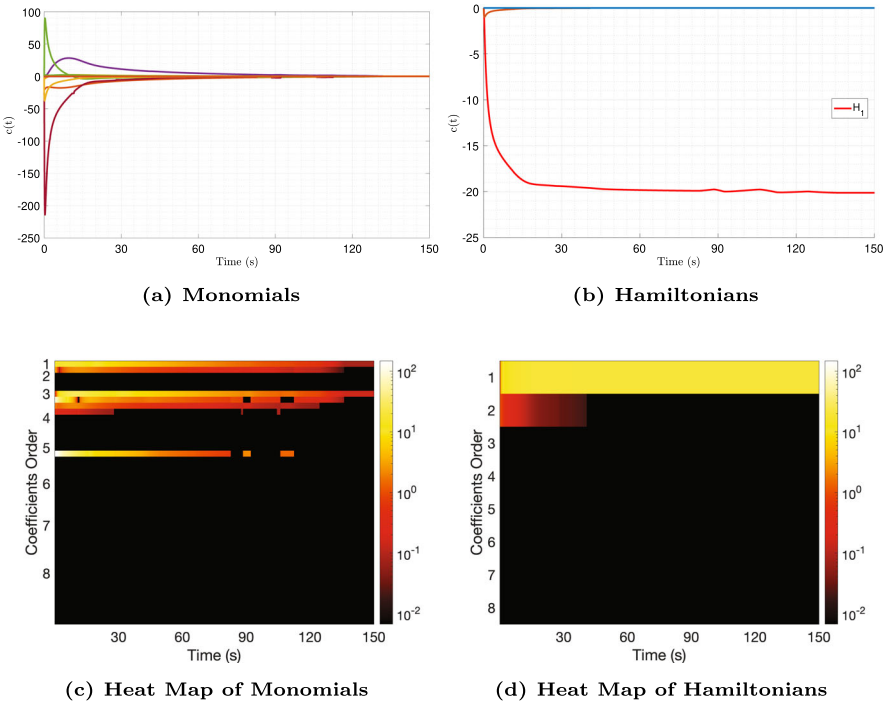
To quantitatively validate the convergence, the numerical log-PDF at $t = 150s$ is:

$$\log(p_{\text{numerical}}) \approx c_0 - 0.0824\, x_1^2 + 0.1530\, x_1^4 + 0.2501\, x_2^2 - 20.1251\, \mathcal{H} \qquad (41)$$

The Hamiltonian coefficient (-20.1251) agrees with the analytical value (-20) to within 0.63% error. Notably, only even-order polynomial terms appear, consistent with the symmetry prediction. The remaining polynomial coefficients (magnitudes 0.08 to 0.25) represent residual transient effects. This explicit comparison validates that the proposed method accurately captures the analytical stationary solutions for both oscillator examples.

4.3 Two-Body Problem

This section examines uncertainty propagation for dynamical systems governed by two-body equations of motion. Uncertainty in these systems can arise from errors in initial conditions or maneuvers. The sparse-regression method described above can be adapted for this type of nonlinear system where dynamics vary only as a function of state. In this case, the evolution of uncertainty can be found by propagating the initial state PDF through ordinary differential equations. This allows for an analytical solution to the state PDF evolution, which can be used directly to propagate the uncertainty.



(a) Monomials (b) Hamiltonians

(c) Heat Map of Monomials (d) Heat Map of Hamiltonians

Fig. 8 Quintic oscillator: RS coefficients evolution

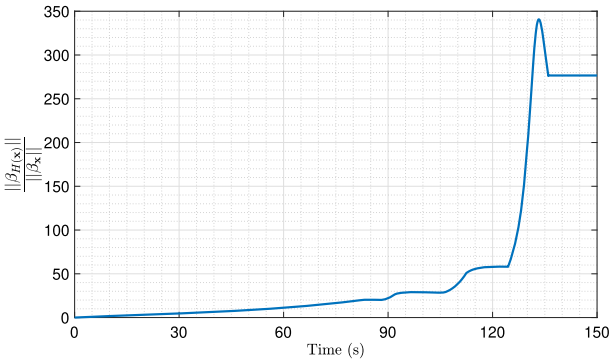


Fig. 9 Quintic oscillator: log-PDF ratio

For an arbitrary orbital state-space, the system dynamics can be expressed as:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t) \tag{42}$$

The state $\mathbf{x}_k$ at time t_k can be written as a function of the initial state $\mathbf{x}_0$ using the system flow $\mathbf{F}$:

$$\mathbf{x}_k = \mathbf{F}(\mathbf{x}_0, t_k) \tag{43}$$

Given the initial state PDF $p(\mathbf{x}_0)$, the state PDF $p(\mathbf{x}_k)$ at time t_k can be computed using the transformation as:

$$p(\mathbf{x}_k) = p\left[\mathbf{x}_0 = \mathbf{F}^{-1}(\mathbf{x}_k)\right]\left|\frac{\partial \mathbf{F}^{-1}}{\partial \mathbf{x}_k}\right| \tag{44}$$

where the flow $\mathbf{F}$ is assumed to be an invertible, continuously differentiable mapping, with inverse given by $\mathbf{x}_0 = \mathbf{F}^{-1}(\mathbf{x}_k)$. This enables determination of the true state PDF from the initial state PDF and system flow information. For precisely known dynamics, the above equation can be simplified to:

$$p(\mathbf{x}_k) = p\left[\mathbf{x}_0 = \mathbf{F}^{-1}(\mathbf{x}_k)\right]\exp\left(-\int_0^{t_k} \nabla \cdot \mathbf{f}(\mathbf{x}(\tau), \tau)d\tau\right) \tag{45}$$

This result can be obtained using the Method of Characteristics. The equation is significant because it enables determination of the discrete-time probability value $p(\mathbf{x}_k)$ from any initial PDF sample. A similar approach can be applied to propagate uncertainty in a two-body problem with maneuver uncertainties. For numerical solution, the initial quadrature points corresponding to the post-maneuver mean $\mu_{\mathbf{x}_1}$ and covariance $\Sigma_{\mathbf{x}_1}$ must be propagated through the dynamics to calculate the corresponding state PDF. Note that propagating the initial quadrature points automatically tracks the solution domain of the FPKE.

A solution approach based on Eq. (45) is proposed to compute the numerical approximation for this system. Due to the high-dimensional input space and to enforce smoothness constraints, the state PDF is assumed to have the following structure:

$$p\left(\mathbf{x}_{k+1}, t_{k+1}\right) = \delta p\left(\mathbf{x}_{k+1}, t_{k+1}\right) p\left(\mathbf{x}_{k+1}, t_k\right) \tag{46}$$

where $p\left(\mathbf{x}_{k+1}, t_{k+1}\right)$ is the PDF of state $\mathbf{x}_{k+1}$ at time t_{k+1}, $p\left(\mathbf{x}_{k+1}, t_k\right)$ is the PDF of state $\mathbf{x}_{k+1}$ at time t_k, and $\delta p\left(\mathbf{x}_{k+1}, t_{k+1}\right)$ represents the departure PDF of state $\mathbf{x}_{k+1}$ from time t_k to t_{k+1}. The state PDF at the previous time step is used as the weight function, i.e., $p_W = p\left(\mathbf{x}_{k+1}, t_k\right)$, to impose the infinite boundary conditions. Taking the logarithm of the above equation and assuming a polynomial basis function results in the following:

$$\begin{aligned} \beta\left(\mathbf{x}_{k+1}, t_{k+1}\right) &= \delta\beta\left(\mathbf{x}_{k+1}, t_{k+1}\right) + \beta\left(\mathbf{x}_{k+1}, t_k\right) \\ \ln\left[p\left(\mathbf{x}_{k+1}, t_{k+1}\right)\right] \approx \mathbf{c}_{k+1}^T \mathbf{\Phi}\left(\mathbf{x}_{k+1}\right) &= \delta\mathbf{c}_{k+1}^T \mathbf{\Phi}\left(\mathbf{x}_{k+1}\right) + \mathbf{c}_k^T \mathbf{\Phi}\left(\mathbf{x}_{k+1}\right) \end{aligned} \tag{47}$$

where $\mathbf{c}_{k+1}$ is a vector of unknown coefficients at time t_{k+1}, $\delta\mathbf{c}_{k+1}$ is a vector of unknown departure coefficients from time t_k to time t_{k+1}, $\mathbf{c}_k$ represents the coefficients at time t_k, and $\mathbf{\Phi}(\mathbf{x}_{k+1})$ is a vector of the chosen basis functions of state $\mathbf{x}_{k+1}$. The objective here is to first compute the departure coefficients $\delta\mathbf{c}_{k+1}$ that can serve as a weight function for the previous time PDF and thus enforce the smoothness constraint. Using equation Eq. (47), the coefficients $\mathbf{c}_{k+1}$ can be found as:

$$\mathbf{c}_{k+1} = \delta\mathbf{c}_{k+1} + \mathbf{c}_k \tag{48}$$

To find the optimal coefficients $\delta\mathbf{c}_{k+1}$, equation Eq. (47) can be rearranged as:

$$\delta\mathbf{c}_{k+1}^T\boldsymbol{\Phi}\left(\mathbf{x}_{k+1}^j\right) = \ln\left[p\left(\mathbf{x}_{k+1}^j\right)\right] - \mathbf{c}_k^T\boldsymbol{\Phi}\left(\mathbf{x}_{k+1}^j\right) \tag{49}$$

Further simplifying the above equation results in:

$$\mathbf{A}\delta\mathbf{c}_{k+1} = \mathbf{b} \tag{50}$$

where:

$$\begin{aligned}\mathbf{A}_j &= \boldsymbol{\Phi}^T\left(\mathbf{x}_{k+1}^j\right), \quad j = 1, 2, \ldots, N\\ \mathbf{b}_j &= \ln\left[p\left(\mathbf{x}_{k+1}^j\right)\right] - \boldsymbol{\Phi}^T\left(\mathbf{x}_{k+1}^j\right)\mathbf{c}_k, \quad j = 1, 2, \ldots, N\end{aligned} \tag{51}$$

To solve for the departure coefficients $\delta\mathbf{c}_{k+1}$, we formulate the following optimization problem:

$$\min_{\delta\mathbf{c}_{k+1}} \|\mathbf{K}\delta\mathbf{c}_{k+1}\|_1$$

subject to:

$$\|\mathbf{w}(\mathbf{A}\delta\mathbf{c}_{k+1} - \mathbf{b})\|_2 \leq \epsilon_s$$

Note that the departure from the previous time PDF must be solved by calculating the departure coefficients, $\delta\mathbf{c}_{k+1}$. The procedure for obtaining the l_2-norm, sparse, and RS coefficients remains the same as illustrated in Algorithm 1. For the two-body problem, the Hamiltonian that needs to be included in the basis function dictionary is given by:

$$\mathcal{H} = \frac{\mathbf{v}^2}{2} - \frac{\mu}{\mathbf{r}} \tag{52}$$

4.3.1 Orbit Transfer Maneuver

This section describes the numerical analysis of a cooperative maneuvering satellite as demonstrated in sect. 4.3. The example illustrates an optimal three-burn transfer orbit from low Earth orbit (LEO) to a final orbit with a specified inclination. For the analysis, it is assumed that the nominal target maneuvers of the satellite are known. For the first burn, the PDF is evaluated under assumptions of uncertain initial conditions and an uncertain burn profile. This procedure can be repeated to calculate the PDF for subsequent burns. Additionally, the initial system uncertainty is assumed to be Gaussian, with the mean given by the nominal profile, which is computed in Keplerian elements (see Table 4). The initial uncertainty is specified in Cartesian coordinates with standard deviations of $\sigma_r = 50$ m for position and $\sigma_v = 0.1$ m/s for velocity in each direction. The resultant PDF evolution after the first burn is shown in Fig. 10. As time progresses, the uncertainties grow rapidly and the PDFs become notably non-Gaussian in nature.

For computing the numerical solution of this example, two different test cases are considered to evaluate the effect of including Hamiltonian coefficients in approximating uncertainty propagation. In case-I, only monomial basis functions up to

Table 4 Orbit Transfer Maneuver: Dynamics and Parameters

Dynamics

$$\dot{\mathbf{x}} = f(\mathbf{x}, t) + g(\mathbf{u}, t) = \begin{bmatrix} \mathbf{v} \\ -\frac{\mu \mathbf{r}}{|\mathbf{r}|^3} \end{bmatrix} + \begin{bmatrix} 0 \\ \mathbf{Tu} \end{bmatrix}$$

$$\mathbf{u} = \begin{bmatrix} |\Delta \mathbf{v}| \sin\theta \cos\psi \hat{N} \\ |\Delta \mathbf{v}| \cos\theta \cos\psi \hat{T} \\ -|\Delta \mathbf{v}| \sin\psi \hat{W} \end{bmatrix}$$

Sparse Parameters

$\epsilon_s = 1e-6, \eta_s = 1e-4, \Delta_s = 1e-4$
$\delta_{rs} = 1e-4$

case - I	Basis($\mathcal{O}_x(8)$) $\rightarrow m = 3003$ ($m_x = 3003, m_H = 0$)
case - II	Basis($\mathcal{O}(8)$) $\rightarrow m = 3011$ ($m_x = 3003, m_H = 8$)

Transfer Orbit Parameters

Parameter	Parking Orbit	1st Transfer
$a^*(km)$	6667.32	22, 835.4
e^*	0	0.7080
i^* (°)	37.40	35.78
ω^* (°)	N/A	254.0

Burn Parameters

Parameter	1st Burn	Final Position
$\|\Delta v^*\|$ (m/s)	2, 383.5	N/A
θ^* (°)	−1.36	N/A
ψ^* (°)	5.99	N/A
η^* (°)	255	121.7
$t(s)$	0	6, 078.2

Training Data $N_{CUT} = 745$

Testing Data $N_{MC} = 1000$

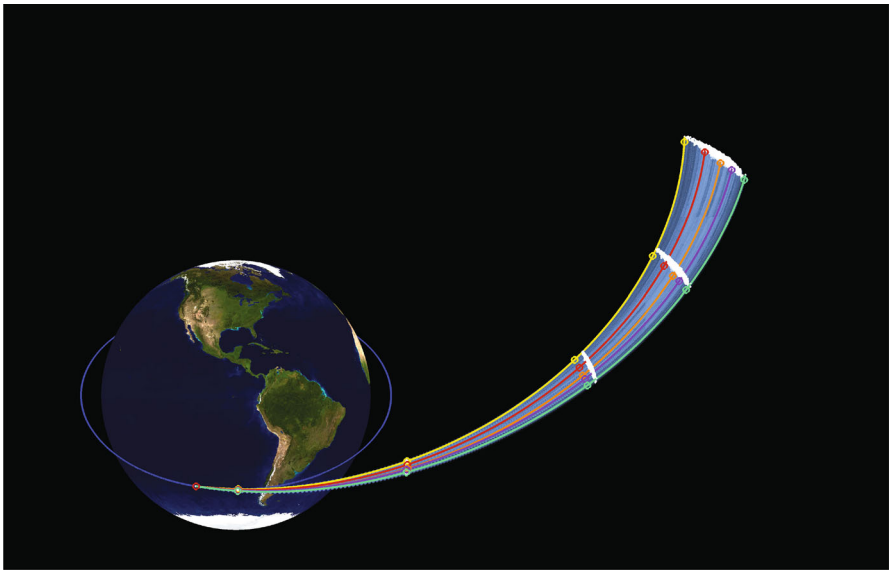


Fig. 10 Orbit transfer maneuver: evolution of PDF

eighth-order are included in the dictionary, resulting in $m = 3003$ coefficients. In case-II, both monomials and Hamiltonians up to eighth-order are included, resulting in a total of $m = 3011$ coefficients. For this 6-D system, CUT8 collocation points are used, yielding $N = 745$ quadrature points. All other parameters are kept constant between the two cases, as shown in Table 4.

Figure 11 illustrates the evolution of RS coefficients in approximating the log-PDF. The first row displays the coefficient evolution for monomials and Hamiltonians in

both cases. For case-II, the coefficient evolution is initially governed by monomials. However, after $t = 2200s$, the first-order Hamiltonian coefficient ($c(\mathcal{H}_1)$) begins to participate in approximating the log-PDF. This involvement of $c(\mathcal{H}_1)$ leads to limited participation of higher-order monomials ($\mathcal{O}(6-8)$) in the log-PDF approximation. In contrast, for case-I, the higher-order monomials actively contribute throughout. Additionally, the magnitude of higher-order coefficients in case-II is two orders lower than in case-I. The heatmap of RS coefficients in the final row provides further insight - in case-I, almost all orders of monomials actively participate, while in case-II, participation is limited to monomials up to sixth order. Including the Hamiltonian dictionary of basis functions in the log-PDF approximation clearly improves performance by requiring fewer monomial coefficients, allowing uncertainty computation with a smaller dictionary.

Figure 12 shows the enforced sparsity constraint for both cases, displaying the evolution of non-zero coefficients for the l_2-norm and RS coefficients. Non-zero coefficients are defined as those with magnitude greater than 10^{-5}. As mentioned earlier, case-I and case-II have 3003 and 3011 total coefficients, respectively. For both the l_2-norm and RS coefficients, the departure coefficients initially approach zero in the first few timesteps. However, as time progresses, non-Gaussian uncertainty enters the system due to nonlinear dynamics. Subsequently, the non-zero coefficients required by the l_2-norm solution increase exponentially with time before converging to utilize almost all dictionary coefficients. In contrast, the RS solution requires about 1500 coefficients for case-I and 200 coefficients for case-II at any given time. Including the Hamiltonian ($\mathcal{H}^1$) in the basis function dictionary reduces the required coefficients by approximately 1300.

Figure 13 depicts the normalized two-norm error of the log-PDF using training and testing data, with 1000 MC samples used as testing data. The trend of training and testing errors is similar in each instance, though the testing error grows over time relative to the training error. Furthermore, Fig. 14 shows the log-PDF ratio of Hamiltonians and monomials in case-II. As time progresses, this ratio increases, signifying growing dominance of Hamiltonian basis functions in propagating the uncertainty.

Finally, Fig. 15 depicts marginalized 2-D PDF contour plots for case-II computed using RS coefficients in the normalized zero mean and identity covariance space at times $t_f/2$ and t_f, where t_f represents the time before the second maneuver. The marginalized 2-D PDFs corresponding to $p(x, y)$, $p(x, z)$, $p(\dot{x}, \dot{y})$, and $p(\dot{x}, \dot{z})$ are displayed in rows one through four. For validation, the contour plots are superimposed with 50,000 MC points shown as blue dots. The color bar represents PDF values at various contour levels. Due to the highly non-Gaussian nature of the PDF, the contours appear non-smooth due to marginalization errors. However, compared to MC points, the proposed method accurately captures the overall PDF shape.

It is important to note that Fig. 15 displays marginalized two-dimensional projections of the six-dimensional PDF. The marginalization process integrates the computed 6-D PDF over the remaining four dimensions to obtain each 2-D view. In regions where the 6-D PDF has complex structure, this dimensional reduction can introduce artifacts in the marginalized projections, particularly at the boundaries where PDF values are small. The superimposed Monte Carlo points, which represent samples from the true 6-D distribution, are concentrated in the high-probability regions and validate that

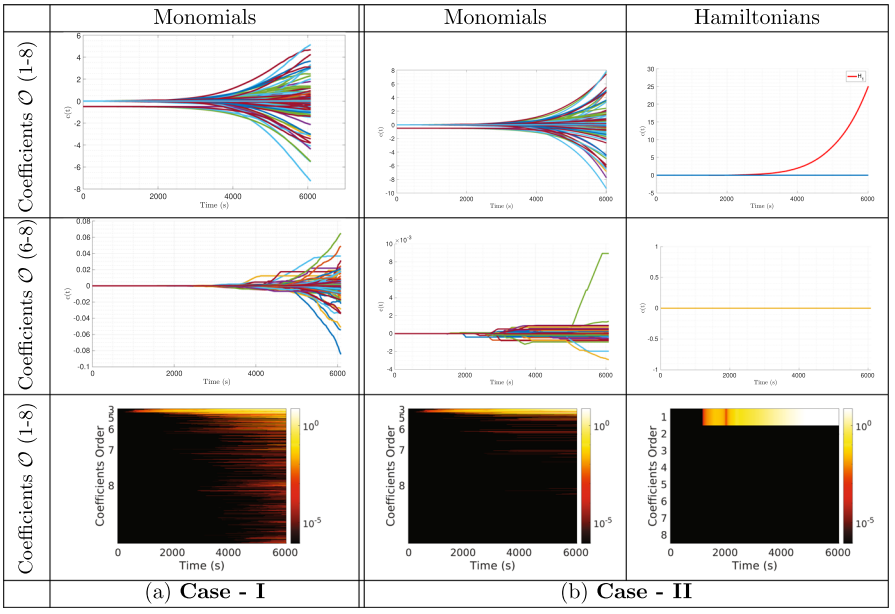


Fig. 11 Orbit transfer maneuver: evolution of RS coefficients using training data (745 CUT points)

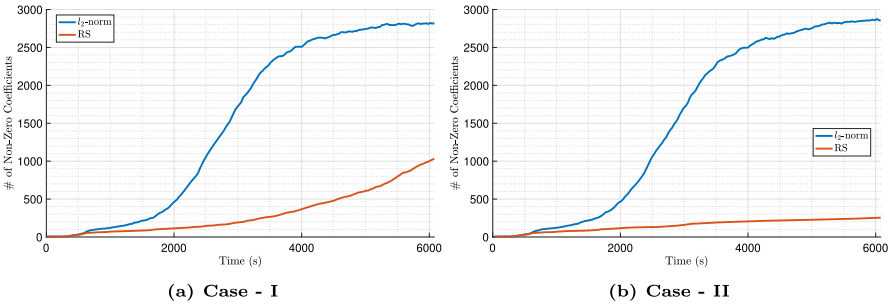


Fig. 12 Orbit transfer maneuver: evolution of non-zero coefficients using training data (745 CUT points)

the proposed method accurately captures the dominant structure of the 6-D PDF. The artifacts visible in some marginalized views (e.g., tri-modal structures in $p(\dot{x}, \dot{y})$ and $p(\dot{x}, \dot{z})$ at the boundaries) are a consequence of the marginalization process rather than inaccuracies in the underlying six-dimensional PDF representation. This is a well-known challenge when visualizing high-dimensional probability distributions through low-dimensional projections.

4.4 Discussion and Limitations

While the proposed methodology demonstrates strong performance across the numerical examples presented, it is important to discuss its limitations and cases where alternative methods may be more suitable.

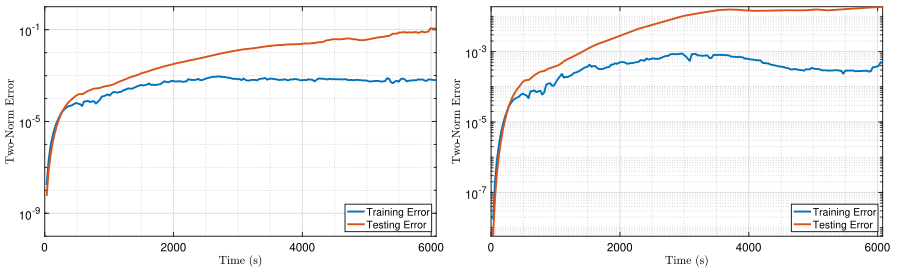


Fig. 13 Orbit transfer maneuver: training vs. testing log-PDF error using RS coefficients (745 CUT points and 1000 MC points)

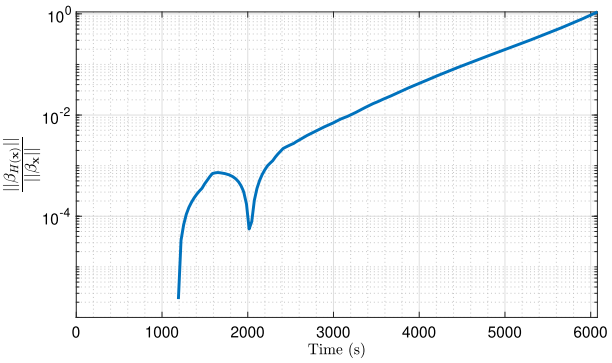


Fig. 14 Orbit transfer maneuver Case-II: log-PDF ratio using training data (745 CUT points)

Multimodal PDFs and Marginalization at Boundaries: As discussed in Sect. 4.3 with Fig. 15, visualizing high-dimensional PDFs through low-dimensional projections can introduce marginalization artifacts, particularly at domain boundaries. While the underlying high-dimensional PDF accurately captures the distribution (as validated by Monte Carlo points in high-probability regions), the dimensional reduction required for visualization can produce spurious features in the marginalized views. Enhanced validation techniques such as MC density heatmaps or univariate PDF comparisons with histograms can help distinguish between genuine multimodal behavior and marginalization artifacts.

Polynomial Basis Limitations: The method relies on polynomial basis functions to approximate the log-PDF. This works well for smooth, continuous PDFs as demonstrated in the numerical examples. However, for systems with highly irregular or discontinuous PDFs (e.g., PDFs with sharp discontinuities or delta functions), polynomial approximations may require prohibitively high orders or may fail to converge. In such cases, alternative basis functions (wavelets, piecewise polynomials) or hybrid methods may be more appropriate.

Gaussian Noise Assumption: The methodology assumes Gaussian white noise in the underlying stochastic differential equation. For systems with non-Gaussian noise (e.g., Poisson processes, Lévy flights, or colored noise), the FPKE formulation requires

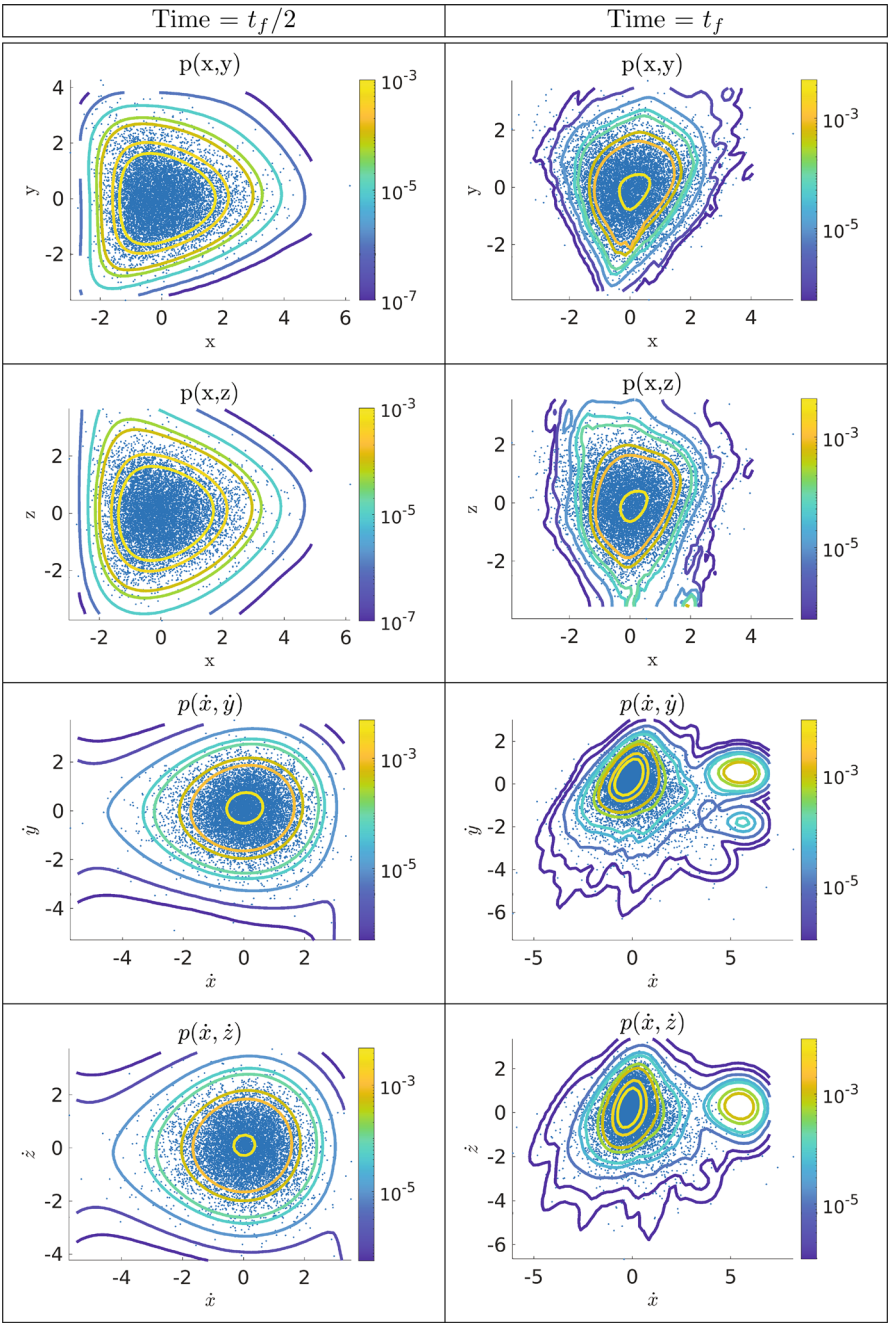


Fig. 15 Orbit transfer maneuver Case-II: PDF contours computed using RS coefficients (745 CUT points) and superimposed with 50,000 Monte Carlo points

modification, and the proposed approach would need adaptation. The performance degrades when the actual noise significantly deviates from the Gaussian assumption.

Computational Scaling: While the sparse regression approach significantly reduces the number of coefficients compared to full polynomial expansions, computational cost still grows with system dimensionality. For very high-dimensional systems (e.g., $n > 10$), even sparse methods face the curse of dimensionality. The CUT quadrature points scale as $O(2^n)$ for CUT4 and more rapidly for higher-order schemes, limiting practical applicability to moderately dimensional systems.

Parameter Sensitivity: The method involves several user-defined parameters (ϵ_s, δ_{rs}, η_s, Δ_s, domain size, basis function order). While reasonable defaults work well for the tested examples, optimal parameter selection may require problem-specific tuning. Systematic sensitivity analysis would provide guidance for practitioners.

Despite these limitations, the proposed method offers a powerful tool for uncertainty propagation in nonlinear systems with moderate dimensionality and well-behaved PDFs, as demonstrated by the numerical examples.

5 Conclusion

This paper presented a numerical framework to compute the evolution of state probability density functions (PDFs) for nonlinear dynamical systems. The state PDF is constrained to satisfy the Fokker-Planck-Kolmogorov Equation (FPKE). Sparse regression tools, in conjunction with a non-product quadrature scheme known as the Conjugate Unscented Transformation (CUT), are used to efficiently solve the FPKE through a recursive convex optimization problem. Furthermore, insights from the existence of stationary solutions of FPKE are utilized to construct appropriate basis functions for accurately representing the state PDF.

Several benchmark problems are solved to demonstrate the efficacy of the developed approach. The approximate solutions are compared with known stationary PDFs, with promising results obtained for all cases examined. Additionally, the effectiveness of the developed approach in accurately and efficiently solving the FPKE is demonstrated for an orbit transfer problem.

While this work demonstrates the method's effectiveness through several numerical examples, future research directions could further strengthen the framework. A systematic parameter sensitivity analysis examining the influence of key algorithmic parameters (e.g., ϵ_s, δ_{rs}, basis order) on solution accuracy and computational cost would provide practitioners with formal guidance for parameter selection across diverse problem classes. Additionally, comprehensive quantitative comparisons with alternative uncertainty propagation methods (e.g., Monte Carlo, grid-based PDE solvers, other sparse collocation schemes) would offer deeper insights into the proposed method's relative computational efficiency and accuracy advantages across various problem characteristics. These extensions would complement the validation approach presented herein and provide a more complete characterization of the method's performance envelope.

Author Contributions A.J. conducted the primary research, literature review, and drafted the initial manuscript. R.E. provided conceptual guidance, and advised on the research approach. P.S. assisted in refining the analytical framework, provided critical feedback on the manuscript, and helped revise the manuscript. All authors reviewed and approved the final manuscript.

Funding This material is based upon work supported jointly by the AFOSR Grants FA9550-20-1-0176 and FA9550-22-1-0092. The results and conclusions of this work do not necessarily reflect the views of the funding agencies.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Fuller, A.T.: Analysis of nonlinear stochastic systems by means of the Fokker- Planck equation. Int. J. Control **9**, 6 (1969)
2. Bogachev, V.I., Krylov, N.V., Röckner, M., Shaposhnikov, S.V.: Fokker-Planck-Kolmogorov Equations, vol. 207. American Mathematical Society, Providence (2022)
3. Caughey, T., Ma, F.: The exact steady-state solution of a class of non-linear stochastic systems. Int. J. Non-Linear Mech. **17**(3), 137–142 (1982)
4. Cai, G., Lin, Y.: Exact and approximate solutions for randomly excited MDOF non-linear systems. Int. J. Non-Linear Mech. **31**(5), 647–655 (1996)
5. Herrador, M.A., Asuero, A.G., González, A.G.: Estimation of the uncertainty of indirect measurements from the propagation of distributions by using the Monte-Carlo method: An overview. Chemom. Intell. Lab. Syst. **79**(1–2), 115–122 (2005)
6. Johnson, E., Wojtkiewicz, S., Bergman, L., Spencer, B., Jr.: Observations with regard to massively parallel computation for Monte Carlo simulation of stochastic dynamical systems. Int. J. Non-Linear Mech. **32**(4), 721–734 (1997)
7. Julier, S., Uhlmann, J., Durrant-Whyte, H.F.: A new method for the nonlinear transformation of means and covariances in filters and estimators. IEEE Trans. Autom. Control **45**(3), 477–482 (2000)
8. Adurthi, N.: The conjugate unscented transform - a method to evaluate multidimensional expectation integrals. Master's thesis. University at Buffalo (2013)
9. Maoutsa, D., Reich, S., Opper, M.: Interacting particle solutions of Fokker-Planck equations through gradient-log-density estimation. Entropy **22**(8), 802 (2020)
10. Carrillo, J.A., Craig, K., Patacchini, F.S.: A blob method for diffusion. Calc. Var. Partial. Differ. Equ. **58**, 1–53 (2019)
11. Reich, S., Weissmann, S.: Fokker-Planck particle systems for Bayesian inference: computational approaches. SIAM-ASA J. Uncertain. Quantif. **9**(2), 446–482 (2021)
12. Roberts, J.: First-passage time for randomly excited non-linear oscillators. J. Sound Vib. **109**(1), 33–50 (1986)
13. Cichon, M.G., Wedig, W.V.: On the approximation of invariant measures of stochastic dynamical systems under multiplicative noise. In: Advances in Stochastic Structural Dynamics: Proceedings of

the 5th International Conference on Stochastic Structural Dynamics-SSD'03, Hangzhou, China, May 26–28, 2003, p. 43. AMACOM Div American Mgmt Assn (2003)
14. Chang, J., Cooper, G.: A practical difference scheme for Fokker-Planck equations. J. Comput. Phys. **6**(1), 1–16 (1970)
15. Spencer, B., Bergman, L.: On the numerical solution of the Fokker-Planck equation for nonlinear stochastic systems. Nonlinear Dyn. **4**(4), 357–372 (1993)
16. Pichler, L., Masud, A., Bergman, L.A.: Numerical solution of the Fokker-Planck equation by finite difference and finite element methods-a comparative study. Comput. Methods. Stoch. Dyn **2**, 69–85 (2013)
17. Al Reda, F., Maury, B.: 15 interpretation of finite volume discretization schemes for the Fokker-Planck equation as gradient flows for the discrete wasserstein distance. In: Topological Optimization and Optimal Transport: in the Applied Sciences vol. 17, p. 400 (2017)
18. Jiang, Y., Xu, X.: A monotone finite volume method for time fractional Fokker-Planck equations. Sci. China Math. **62**, 783–794 (2019)
19. Heida, M., Kantner, M., Stephan, A.: Consistency and convergence for a family of finite volume discretizations of the Fokker-Planck operator. ESAIM: Math. Model. Numer. Anal. **55**(6), 3017–3042 (2021)
20. Finlayson, B.A., Scriven, L.E.: The method of weighted residuals-a review. Appl. Mech. Rev. **19**(9), 735–748 (1966)
21. Young, L.C.: Orthogonal collocation revisited. Comput. Methods Appl. Mech. Eng. **345**, 1033–1076 (2019)
22. Barnbrook, G., Mason, O., Krishnamurthy, R.: Collocation: Applications and Implications. Springer, London (2013)
23. Gomez, H., De Lorenzis, L.: The variational collocation method. Comput. Methods Appl. Mech. Eng. **309**, 152–181 (2016)
24. Russell, R., Shampine, L.F.: A collocation method for boundary value problems. Numer. Math. **19**, 1–28 (1972)
25. Jordan, R., Kinderlehrer, D., Otto, F.: The variational formulation of the Fokker-Planck equation. SIAM J. Math. Anal. **29**(1), 1–17 (1998)
26. Johnson, J.P., Scott, R.A.: Extension of eigenfunction-expansion solutions of a Fokker-Planck equation-II. Second order system. Int. J. Non-Linear Mech. **15**(1), 41–56 (1980)
27. Abe, S.: Hamiltonian formulation of fractional kinetics. Eur. Phys. J. Spec. Top **227**, 683–691 (2018)
28. Wehner, M.F., Wolfer, W.: Numerical evaluation of path-integral solutions to Fokker-Planck equations. Phys. Rev. A **27**(5), 2663 (1983)
29. Xu, Y., Zan, W., Jia, W., Kurths, J.: Path integral solutions of the governing equation of sdes excited by lévy white noise. J. Comput. Phys. **394**, 41–55 (2019)
30. Naess, A., Johnsen, J.M.: Response statistics of nonlinear dynamic systems by path integration. In: Nonlinear Stochastic Mechanics, pp. 401–414. Springer, Turin (1992)
31. Di Paola, M., Alotta, G.: Path integral methods for the probabilistic analysis of nonlinear systems under a white-noise process. ASCE-ASME J. Risk Uncertain. Eng. Syst. Part B: Mech. Eng. **6**(4), 040801 (2020)
32. Wei, G.: A unified approach for the solution of the Fokker-Planck equation. J. Phys. A: Math. Gen. **33**(27), 4935 (2000)
33. Hasselmann, K.: Stochastic climate models part I. Theory. Tellus **28**(6), 473–485 (1976)
34. Chacón, L., Barnes, D., Knoll, D., Miley, G.: An implicit energy-conservative 2d Fokker-Planck algorithm: I. Difference scheme. J. Comput. Phys. **157**(2), 618–653 (2000)
35. Chen, N., Majda, A.J.: Efficient statistically accurate algorithms for the Fokker-Planck equation in large dimensions. J. Comput. Phys. **354**, 242–268 (2018)
36. Shen, Z., Wang, Z., Kale, S., Ribeiro, A., Karbasi, A., Hassani, H.: Self-consistency of the Fokker Planck equation. In: Conference on Learning Theory, pp. 817–841. PMLR (2022)
37. Gu, Y., Harlim, J., Liang, S., Yang, H.: Stationary density estimation of itô diffusions using deep learning. SIAM J. Numer. Anal. **61**(1), 45–82 (2023)
38. Govaers, F.: Time evolution for dynamic probabilistic tensors in hierarchical tucker decomposition form. In: 2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM), pp. 16–20. IEEE (2018)

39. Dolgov, S.V., Khoromskij, B.N., Oseledets, I.V.: Fast solution of parabolic problems in the tensor train/quantized tensor train format with initial application to the Fokker-Planck equation. SIAM J. Sci. Comput. **34**(6), 3016–3038 (2012)
40. Sun, Y., Kumar, M.: Numerical solution of high dimensional stationary Fokker-Planck equations via tensor decomposition and Chebyshev spectral differentiation. Comput. Math. Appl. **67**(10), 1960–1977 (2014)
41. Terejanu, G., Singla, P., Singh, T., Scott, P.D.: Uncertainty propagation for nonlinear dynamic systems using Gaussian mixture models. J. Guid. Control. Dyn. **31**(6), 1623–1633 (2008)
42. Vishwajeet, K., Singla, P., Jah, M.: Nonlinear uncertainty propagation for perturbed two-body orbits. J. Guid. Control. Dyn. **37**(5), 1415–1425 (2014)
43. Vishwajeet, K., Singla, P.: Adaptive split/merge-based gaussian mixture model approach for uncertainty propagation. J. Guid. Control. Dyn. **41**(3), 603–617 (2018). https://doi.org/10.2514/1.G002801
44. Vishwajeet, K., Singla, P.: Sparse approximation based Gaussian mixture model approach for uncertainty propagation for nonlinear systems. In: 2013 American Control Conference, pp. 1213–1218. IEEE (2013)
45. Mukherjee, A., Rai, R., Singla, P., Singh, T., Patra, A.: An adaptive Gaussian mixture model approach based framework for solving Fokker-Planck Kolmogorov equation related to high dimensional dynamical systems. In: International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, vol. 50114, pp. 02–03058. American Society of Mechanical Engineers (2016)
46. Losacco, M., Fossà, A., Armellin, R.: Low-order automatic domain splitting approach for nonlinear uncertainty mapping. J. Guid. Control. Dyn. **47**(2), 291–310 (2024). https://doi.org/10.2514/1.G007271
47. Kumar, M., Chakravorty, S., Singla, P., Junkins, J.L.: The partition of unity finite element approach with hp-refinement for the stationary fokker-planck equation. J. Sound Vib. **327**(1–2), 144–162 (2009). https://doi.org/10.1016/j.jsv.2009.05.033
48. Boyd, J.P.: Chebyshev and Fourier Spectral Methods. Courier Corporation, Massachusetts (2001)
49. Mercurio, M., Singla, P.: Conjugate unscented transform based approach to solve the Fokker-Planck-Kolmogorov equation. In: AIAA/AAS Astrodynamics Specialist Conference, pp. 2016–5679. AIAA (2016)
50. Mercurio, M., Adurthi, N., Singla, P., Majji, M.: A collocation-based approach to solve the finite horizon Hamilton-Jacobi-Bellman equation. In: American Control Conference (2016)
51. Mercurio, M., Singla, P., Majji, M.: A conjugate unscented transform-based scheme for optimal control with terminal state constraints. In: Annual American Control Conference (ACC), pp. 2651–2656. IEEE (2018)
52. Adurthi, N., Singla, P., Majji, M.: Sparse approximation based collocation scheme for nonlinear optimal feedback control design. AIAA J. Guid. Control Dyn. (2017)
53. Jazwinski, A.H.: Stochastic Processes and Filtering Theory. Courier Corporation, New York (2007)
54. Risken, H.: Fokker-Planck equation. In: The Fokker-Planck Equation, pp. 63–95. Springer, Berlin (1996)
55. Jain, A., Singla, P.: Stochastic reachability analysis using sparse-collocation method. In: AIAA SCITECH 2023 Forum, p. 1852 (2023)
56. Mercurio, M.: Sparse collocation methods for solving high dimension pdes in estimation and control of dynamical systems, State University of New York at Buffalo. (PhD thesis) (2017)
57. Muscolino, G., Ricciardi, G., Vasta, M.: Stationary and non-stationary probability density function for non-linear oscillators. Int. J. Non-Linear Mech. **32**(6), 1051–1064 (1997)
58. Muthukrishnan, R., Rohini, R.: Lasso: A feature selection technique in predictive modeling for machine learning. In: 2016 IEEE International Conference on Advances in Computer Applications (ICACA), pp. 18–20. IEEE (2016)
59. McDonald, G.C.: Ridge regression. Wiley Interdiscip. Rev. Comput. Stat. **1**(1), 93–100 (2009)
60. Pesheck, E., Pierre, C., Shaw, S.W.: A new galerkin-based approach for accurate non-linear normal modes through invariant manifolds. J. Sound Vib. **249**(5), 971–993 (2002)
61. Miller, S.J.: The method of least squares. Mathematics Department Brown University, vol. 8(1) (2006)
62. Xiu, D., Hesthaven, J.S.: High-order collocation methods for differential equations with random inputs. SIAM J. Sci. Comput. **27**(3), 1118–1139 (2005)
63. Gerstner, T., Griebel, M.: Numerical integration using sparse grids. Numer. Algorithms. **18**, 209–232 (1998). https://doi.org/10.1023/A:1019129717644

64. Stroud, A.H., Secrest, D.: Gaussian Quadrature Formulas. Prentice Hall, Englewood Cliffs (1966)
65. Adurthi, N., Singla, P., Singh, T.: Conjugate unscented transformation: applications to estimation and control. J. Dyn. Syst. Meas. Contr. **140**(3), 030907 (2018)
66. Adurthi, N., Singla, P., Singh, T.: Conjugate unscented transform and its application to filtering and stochastic integral calculation. In: AIAA Guidance, Navigation, and Control Conference (2012)
67. Adurthi, N., Singla, P., Majji, M.: conjugate unscented transform based approach for dynamic sensor tasking and space situational awareness. In: American Control Conference (ACC), pp. 5218–5223. IEEE (2015)
68. Adurthi, N., Singla, P.: A Conjugate Unscented Transformation based approach for accurate conjunction analysis. AIAA J. Guid. Control Dyn. **38**(9), 1642–1658 (2015)
69. Adurthi, N., Majji, M., Singla, P.: Quadrature-based nonlinear joint probabilistic data association filter. J. Guid. Control. Dyn. **42**(11), 2369–2381 (2019)
70. Adurthi, N., Singla, P., Majji, M.: Mutual information based sensor tasking with applications to space situational awareness. J. Guid. Control. Dyn. **43**(4), 767–789 (2020)
71. Chen, S., Donoho, D.: Atomic decomposition by basis pursuit. In: SPIE International Conference on Wavelets (1995)
72. Boyd, S.P., Vandenberghe, L.: Convex Optimization. University Press, Cambridge (2004)
73. Grant, M., Boyd, S.: CVX: Matlab Software for Disciplined Convex Programming, version 2.1. http://cvxr.com/cvx (2014)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.