Transient anisotropic kernel for probabilistic learning on manifolds

Christian Soize, Roger Ghanem

Transient anisotropic kernel for probabilistic learning on manifolds

Christian Soize[a,*], Roger Ghanem[b]

[a] *Université Gustave Eiffel, MSME UMR 8208, 5 bd Descartes, 77454 Marne-la-Vallée, France*
[b] *University of Southern California, 210 KAP Hall, Los Angeles, CA 90089, United States*

Abstract

PLoM (Probabilistic Learning on Manifolds) is a method introduced in 2016 for handling small training datasets by projecting an Itô equation from a stochastic dissipative Hamiltonian dynamical system, acting as the MCMC generator, for which the KDE-estimated probability measure with the training dataset is the invariant measure. PLoM performs a projection on a reduced-order vector basis related to the training dataset, using the diffusion maps (DMAPS) basis constructed with a time-independent isotropic kernel. In this paper, we propose a new ISDE projection vector basis built from a transient anisotropic kernel, providing an alternative to the DMAPS basis to improve statistical surrogates for stochastic manifolds with heterogeneous data. The construction ensures that for times near the initial time, the DMAPS basis coincides with the transient basis. For larger times, the differences between the two bases are characterized by the angle of their spanned vector subspaces. The optimal instant yielding the optimal transient basis is determined using an estimation of mutual information from Information Theory, which is normalized by the entropy estimation to account for the effects of the number of realizations used in the estimations. Consequently, this new vector basis better represents statistical dependencies in the learned probability measure for any dimension. Three applications with varying levels of statistical complexity and data heterogeneity validate the proposed theory, showing that the transient anisotropic kernel improves the learned probability measure.

Keywords: Transient kernel, probabilistic learning, PLoM, diffusion maps, Fokker-Planck operator, spectrum

1. Introduction

1.1. Objectives of the paper

PLoM (Probabilistic Learning on Manifolds), introduced in 2016 [1], is a method and algorithm specifically developed for cases where the training dataset consists of a small number of data points. This method is based on projecting, along the data axis, a matrix-valued Itô equation associated with a stochastic dissipative Hamiltonian dynamical system, which acts as the MCMC generator for the probability measure estimated using the Gaussian Kernel Density Estimation method (GKDE) applied to the points of the training dataset. The projection basis is the diffusion maps (DMAPS) basis associated with a time-independent isotropic kernel, introduced in [2, 3], and will be named the reduced-order DMAPS basis (RODB) in the context of the PLoM method.

Since 2016, all extensions and applications of PLoM (see Section 1.2) have been carried out using the isotropic kernel. Through these applications, we have seen that the isotropic kernel allows for obtaining quality results, even for heterogeneous data and systems of great statistical complexity in small and large dimensions.

However, improving the construction of statistical surrogates for stochastic manifolds involving conditional statistics using PLoM based on a projection basis derived from a transient anisotropic kernel (time-dependent) is an interesting problem. Some important works on anisotropic diffusions in stochastic dynamical systems have already been published [2, 3, 4, 5, 6]. The approach presented here differs from these published works. In the present work, the stochastic system considered comes from a numerical simulation stochastic model of a generally complex physical system, from which we build a training dataset. This dataset is not considered as a structured and ordered family of

*Corresponding author: C. Soize, christian.soize@univ-eiffel.fr
Email addresses: christian.soize@univ-eiffel.fr (Christian Soize), ghanem@usc.edu (Roger Ghanem)

Published in Computer Methods in Applied Mechanics and Engineering, Elsevier, https://doi.org/10.1016/j.cma.2024.117453 October 22, 2024

points indexed by a physical parameter such as time t. The only available information is a probability measure of a random vector estimated using the training dataset. Therefore, we are not dealing with a case where a stochastic dynamic system is known and given. Both the anisotropic and transient characters are described using a completely data-driven formalism.

In this paper, we propose a novel construction of the ISDE projection vector basis, built from a transient anisotropic kernel, which enhances the representation of the statistical dependencies of the learned joint probability measure in any dimension. However, to clarify the objectives of this paper, it is necessary to address certain points right now.

The first algorithmic step of PLoM involves performing a principal component analysis (PCA) on the non-Gaussian random vector $\boldsymbol{X}$, which is generally in high dimension. This vector $\boldsymbol{X}$ results from concatenating the random vector $\boldsymbol{Q}$, representing the quantities of interest (QoI), with the random vector $\boldsymbol{W}$, representing the random control variables of the considered stochastic system. Additionally, this system depends on random latent/uncontrolled variables. The vector-valued random QoI, $\boldsymbol{Q} = \boldsymbol{F}(\boldsymbol{W})$, is connected to $\boldsymbol{W}$ by an unknown random nonlinear mapping. Denoting by $\mathscr{C}_w$ the support of the probability measure of $\boldsymbol{W}$, the random graph $\{(\boldsymbol{w}, \boldsymbol{F}(\boldsymbol{w})) \,|\, \boldsymbol{w} \in \mathscr{C}_w\}$ forms a stochastic manifold due to the latent random variables in the system implying the randomness of mapping $\boldsymbol{F}$. The PCA is performed on the training dataset of $\boldsymbol{X}$. The new coordinates resulting from the PCA form a random vector $\boldsymbol{H}$ with a reduced dimension. The training dataset of $\boldsymbol{H}$ is obtained through the PCA projection of the training dataset of $\boldsymbol{X}$. By construction, $\boldsymbol{H}$ is a non-Gaussian normalized random vector, which is centered and has a covariance matrix equal to the identity matrix.

Under these conditions, the available information is only the training dataset of $\boldsymbol{H}$. Using this dataset, we construct the probability measure of $\boldsymbol{H}$ by using a modification of the GKDE method. Thus, the only available information is represented by the probability density function of $\boldsymbol{H}$ with respect to the Lebesgue measure. Given the normalization of $\boldsymbol{H}$, the challenges in constructing the learned dataset of $\boldsymbol{H}$, with or without constraints, involve preserving the concentration of the learned probability measure for $\boldsymbol{H}$ (induced by the random manifold) and ensuring the "good quality" of the learned joint probability measure of the components of $\boldsymbol{H}$, not just the marginal probability measure of each component. The quality of the statistical surrogate model of $\boldsymbol{Q}$ given $\boldsymbol{W}$, based on conditional statistics, is directly linked to the learned joint probability measure of the components of $\boldsymbol{H}$. These aspects will be analyzed in detail in this paper. To simplify the presentation, we will construct the model directly on $\boldsymbol{H}$ rather than starting from $\boldsymbol{X} = (\boldsymbol{Q}, \boldsymbol{W})$. For numerical illustration, three applications will be formulated directly on $\boldsymbol{H}$. It should be noted that, in terms of practical applications, the PLoM method is suitable for complex systems with high-dimensional $\boldsymbol{X}$. For more on this aspect, we refer the reader to the publications referenced at the end of Section 1.2. The difficulty in accurately representing the learned joint probability measure of the components of $\boldsymbol{H}$ is not due to the dimension of $\boldsymbol{H}$ (even though its dimension is 46 for the third application), but rather to the complexity of the support of its probability measure. It is with this objective in mind that we built the three applications.

This paper proposes an analysis, using information theory, of the quality of the learned joint probability measure of the components of $\boldsymbol{H}$, using PLoM with the RODB. Additionally, and most importantly, it introduces a new projection basis for PLoM, derived from a time-dependent anisotropic kernel. The objective is to determine if this new basis, named the reduced-order transient basis (ROTB), improves the construction of statistical surrogate models for stochastic manifolds involving conditional statistics.

It is important to note the following points regarding this objective. We construct a time-dependent family of learned probability measures for $\boldsymbol{H}$ provided by PLoM, where the projection basis is a time-dependent family of reduced-order transient basis named as ROTB(t), associated with a transient anisotropic kernel. This construction is performed so that this time-dependent family converges to the solution provided by PLoM with the RODB as t approaches the initial time. The given points at the initial time are the points of the training dataset of $\boldsymbol{H}$. The novel basis ROTB for PLoM is obtained by selecting the basis ROTB(t_{opt}) from the time-dependent family ROTB(t) at the time t_{opt} for which the quality of the learned probability measure of $\boldsymbol{H}$ is optimal, according to the chosen information theory criterion.

The ISDE chosen to generate the time-dependent solution family must have a drift vector (which is nonlinear) and a diffusion matrix that are independent of time because we want the asymptotic solution for large t to converge to the stationary solution, for which the marginal law is the invariant measure. This choice was made to ensure this property. Building an ISDE with time-dependent coefficients would require arbitrarily constructing the drift vector and the diffusion matrix as functions of time t. These quantities should become independent of time for sufficiently large t; otherwise, there would be no stationary solution and thus no invariant measure. In the framework of the pro-

posed PLoM method, the only available information is the training dataset of $\boldsymbol{H}$, which is not generated by a known structured family parameterized by time. Therefore, in this context, there is no information available in the training dataset to construct time-dependent drift vector and diffusion matrix. Thus, for the framework used, the proposed approach is entirely data-driven without any assumptions beyond an intrinsic structure extracted from the data themselves. The stochastic differential equation and the associated FKP equation are formulated and solved without any assumptions. The proposed theory is independent of any underlying dynamics specified by the character of the data. If the training dataset is obtained from a non-autonomous system, then this character will exist in the training data set and will be encoded in the joint probability density function of the random observations. For constructing the novel projection basis for PLoM with the transient anisotropic kernel, we have chosen a first-order ISDE, while PLoM is based on a second-order ISDE (dissipative Hamiltonian stochastic system as mentioned above) for generating the learned realizations. The choice of a first-order ISDE is justified by our interest in transient stochastic responses. The choice of a second-order dissipative Hamiltonian stochastic equation for PLoM, made in 2016, was to quickly reach the stationary stochastic solution of the ISDE.

Finally, the methodological objective for dealing with real-word applications using the proposed approach can be summarized as follows. The PLoM method has been applied to a large number of complex applications, as explained above and in Section 1.2, demonstrating its usefulness for real-world applications. The random vector $\boldsymbol{H}$ is a quasi-equivalent reduced representation of the random vector $\boldsymbol{X}$ that can represent complex real-world systems. The proposed improvement of this PLoM method, which is connected to it as previously explained, concerns the quality of the learned joint probability measure of the components of $\boldsymbol{H}$. This quality is quantified by a criterion from information theory, allowing for intrinsic validation of the improvement obtained. This means that for each application, for which we only have a small training dataset, the quality of the learned joint probability measure constructed by PLoM with RODB can be quantified and potentially improved using PLoM with ROTB, thus quantifying this improvement if it exists. Furthermore, additional information is extracted from the training dataset that is not used in the original PLoM. Specifically, the path from each datapoint to the collective invariant measure carries a signature of heterogeneity, which is leveraged in the present work. The three proposed applications, for which the support of the probability measure is complex, show that in these cases, an improvement is obtained.

1.2. Framework of the considered problem

Machine learning tools and artificial intelligence [7, 8, 9, 10], such as probabilistic and statistical learning [11, 12, 13, 14, 15], are used in UQ for problems that would require computer resources not available with the most usual approaches. Thus, methods have emerged in the field of engineering sciences, such as learning on manifolds [16, 1, 17, 18, 19] and physics-informed probabilistic learning [20, 21, 22].

Probabilistic learning is a very active domain of research for constructing surrogate models (see for instance, [23, 16, 24, 25, 26, 18, 20]). Probabilistic Learning on Manifolds (PLoM) is a tool in computational statistics, introduced in 2016 [1], which can be viewed as a tool for scientific machine learning. The PLoM approach has specifically been developed for small dataset cases [1, 27, 28, 29, 30]. The method avoids the scattering of learned realizations associated with the probability distribution to preserve its concentration in the neighborhood of the random manifold defined by the parameterized computational model. This method allows for solving unsupervised and supervised problems under uncertainty when the training datasets are small. This situation is encountered in many problems in physics and engineering science with expensive function evaluations. The exploration of the admissible solution space in these situations is thus hampered by available computational resources.

Several extensions have been proposed to account for implicit constraints induced by physics, computational models, and measurements [21, 22, 31], to reduce the stochastic dimension using a statistical partition approach [32], and to update the prior probability distribution with a target dataset, whose points are, for instance, experimental realizations of the system observations [33]. Consequently, PLoM, constrained by a stochastic computational model and statistical moments or samples/realizations, allows for performing probabilistic learning inference and constructing predictive statistical surrogate models for large parameterized stochastic computational models.

This last capability of PLoM can also be viewed as an alternative method to Bayesian inference for high dimensions [34, 35, 36, 37, 38, 39, 40, 41, 42, 43] and is a complementary approach to existing methods in machine learning for sampling distributions on manifolds under constraints. Although a Bayesian inference methodology has also been developed using probabilistic learning on manifolds for high dimensions [29].

PLoM has successfully been adapted to tackle these challenges for several related problems, including nonconvex optimization under uncertainty [44, 45, 46, 47, 48, 49, 50, 51], fracture paths in random composites [52], concurrent multiscale simulations in random media [53], stochastic homogenization in random elastic media [54], ultrasonic transmission techniques in cortical bone microstructures [29], updating digital twins under uncertainties [55], updating under-observed dynamical systems [56, 57], calculation of the Sobol indices [58], dynamic monitoring [59], surrogate modeling of structural seismic response [60], probabilistic-learning-based stochastic surrogate models from small incomplete datasets [61], and polynomial-chaos-based conditional statistics for probabilistic learning of atomic collisions [62], as well as for aeroacoustic liner impedance metamodels from simulation and experimental data [63].

1.3. Methodology proposed and organization of the paper

Starting with a training dataset of n_d realizations of ν random variables, we consider the probability flow from each of these n_d realizations towards the ν-dimensional sampling probability distribution of the training dataset. We construct the associated coupled Fokker-Planck (FKP) equations with each of the realizations as initial condition. These describe the evolution of the transition probability measures over the graph described by the training dataset, that transport each independent realization, viewed as a concentrated measure at the initial time, into the joint probability measure, consisting of the common stationary probability measure (also called the steady-state solution) of the FKP equations. This evolution describes a trajectory along which transition probabilities are consistent with both the training dataset and its postulated joint probability density function (that we approximate using a Gaussian Kernel Density Estimate (KDE)). Each of these probabilities provides a distinct geometric characterization of the training dataset, with its own plausible model of statistical dependence, resulting in a transient anisotropic kernel from which we construct a time evolving PLoM. We then set criteria for selecting among these PLoM models, which is tantamount to identifying the most appropriate statistical dependence structure for the learned dataset, along the flow characterized by the FKP equations. The following description of the paper's organization provides a coherent summary of the proposed methodology.

In Section 2, we define the training dataset constituted of given realizations of a non-Gaussian normalized vector-valued random variable $\boldsymbol{H}$ (centered and with an identity covariance matrix) and define the associated probability measure $P_{\boldsymbol{H}}$, whose density is estimated using the Gaussian Kernel Density Estimation (GKDE) method.

Section 3 deals with a short summary of formal results, introducing an Itô stochastic differential equation (ISDE) and the derived Fokker-Planck (FPK) equation for which $P_{\boldsymbol{H}}$ is the invariant measure, which is the steady-state solution. We then introduce a formal formulation of the eigenvalue problem of the FPK operator and the nonstationary solution of the Fokker-Planck equation with a deterministic initial condition.

Section 4 is devoted to the time-dependent kernel k_t derived from the time-dependent solution of the Fokker-Planck equation with a deterministic initial condition. We then define the time-dependent operator K_t associated with the transient kernel k_t, which is a Hilbert-Schmidt operator. We propose a construction of its finite approximation, represented by a matrix $[\widehat{K}(t)]$, using a sampling of $P_{\boldsymbol{H}}$ of the bilinear form associated with K_t, and introduce the corresponding finite approximation of the eigenvalue problem.

In Section 5, we present the direct construction of matrix $[\widehat{K}(t)]$. We introduce the transition probability density function as the solution of the ISDE with a deterministic initial condition and study its existence, uniqueness, and properties. We then rewrite the ISDE in matrix form and propose a time-discrete approximation of this matrix-valued ISDE based on an Euler scheme. We introduce convergence criteria to verify convergence. Finally, we construct an explicit representation of the matrix $[\widehat{K}(t)]$ based on the nonstationary solution of the matrix-valued ISDE with a deterministic initial condition.

A numerical illustration of the proposed formulation is given in Section 6, for which an explicit solution is known.

Section 7 deals with the construction and study of the vector basis for PLoM derived from the transient anisotropic kernel, connected to the DMAPS basis constructed with the isotropic kernel. We construct a time-dependent matrix $[\tilde{K}(n\Delta t)]$ of the transient anisotropic kernel derived from the matrix $[\hat{K}(n\Delta t)]$, whose fundamental property is its convergence to the matrix $[K_{\text{DM}}]$ of the DMAPS isotropic kernel as time approaches zero. The time-dependent reduced-order transient basis is then the eigenvectors of $[\tilde{K}(n\Delta t)]$ associated with its dominant positive eigenvalues. In order to qualify and quantify the gain of the constructed reduced-order transient basis, ROTB($n\Delta t$) at a sampling time $n\Delta t$, with respect to the reduced-order DMAPS basis, RODB, we introduce the angle between the subspaces spanned by ROTB($n\Delta t$) and RODB. We propose a methodology for identifying the optimal value of the sampling time $n\Delta t$, which

maximizes a selection criterion. This criterion is based on the estimation of mutual information from Information Theory, normalized using entropy estimation to account for the effects of the number of realizations used in the statistical estimators. Such a criterion allows for selecting the best joint learned probability measure with respect to statistical dependencies.

In Section 8, we present three applications, each with a specific level of complexity and data heterogeneity in the training dataset. The first application is created such that the probability measure of $\boldsymbol{H}$ in $\mathbb{R}^9$, defined by the points of the training dataset, is concentrated in a multiconnected domain of $\mathbb{R}^9$. The constituent connected parts are manifolds of dimensions much lower than 9, each with different dimensions. These parts may or may not be connected to each other. The training dataset of the second application consists of realizations of the random vector $\boldsymbol{H}$ with values in $\mathbb{R}^8$, generated using a polynomial chaos expansion of degree 6 of a real-valued random variable, whose random germ has a dimension of 2, with each of the two random germs being a uniform random variable with different support. There are therefore 28 terms in this expansion, and the 8 components of $\boldsymbol{H}$ are defined as the random terms of given rank, defining a relatively complex random manifold in $\mathbb{R}^8$. The third application results from a statistical treatment of an experimental database containing photon measurements in the ATLAS detector at CERN. The PCA step of PLoM has been performed, and 45 components have been extracted to obtain the training dataset for the $\mathbb{R}^{45}$-valued random variable $\boldsymbol{H}$. This application is in higher dimension than the first two but has less statistical complexity.

The paper is completed by two appendices. Appendix A presents an overview of the probabilistic learning on manifolds (PLoM) algorithm and its parameterization, using either the DMAPS basis (RODB) or the transient basis (ROTB($n\Delta t$)) as the projection basis. Appendix B provides the formulas for estimating the Kullback-Leibler divergence, mutual information, and entropy from a set of realizations.

It should be noted that there are some very brief repetitions, which have been deliberately made to facilitate reading.

1.4. Convention for the variables, vectors, and matrices

x, η: lower-case Latin or Greek letters are deterministic real variables.
$\boldsymbol{x}, \boldsymbol{\eta}$: boldface lower-case Latin or Greek letters are deterministic vectors.
X: upper-case Latin letters are real-valued random variables.
$\boldsymbol{X}$: boldface upper-case Latin letters are vector-valued random variables.
$[x]$: lower-case Latin letters between brackets are deterministic matrices.
$[\boldsymbol{X}]$: boldface upper-case letters between brackets are matrix-valued random variables.

1.5. Algebraic notations

$\mathbb{N}, \mathbb{N}^*$: set of natural numbers including 0, excluding 0.
$\mathbb{R}, \mathbb{R}^+, \mathbb{R}^{+*}$: set of real numbers, subset $[0, +\infty[$, subset $]0, +\infty[$.
$\mathbb{R}^\nu$: Euclidean vector space of dimension ν.
$\mathbb{M}_{n,m}, \mathbb{M}_n$: set of the $(n \times m)$, $(n \times n)$, real matrices.
$\mathbb{M}_n^+, \mathbb{M}_n^{+0}$: set of the positive-definite, positive, $(n \times n)$ real matrices.
$[I_n]$: identity matrix in $\mathbb{M}_n$.
$\| [x] \|_F$: Frobenius norm of matrix $[x]$.
∇, div: gradient and divergence operators in $\mathbb{R}^\nu$.
δ_0: Dirac measure on $\mathbb{R}^\nu$ at point $\mathbf{0}_\nu$.

1.6. Convention used for random variables

In this paper, for any finite integer $m \geq 1$, the Euclidean space $\mathbb{R}^m$ is equipped with the σ-algebra $\mathcal{B}_{\mathbb{R}^m}$. If $\boldsymbol{Y}$ is a $\mathbb{R}^m$-valued random variable defined on the probability space $(\Theta, \mathcal{T}, \mathcal{P})$, $\boldsymbol{Y}$ is a mapping $\theta \mapsto \boldsymbol{Y}(\theta)$ from Θ into $\mathbb{R}^m$, measurable from $(\Theta, \mathcal{T})$ into $(\mathbb{R}^m, \mathcal{B}_{\mathbb{R}^m})$, and $\boldsymbol{Y}(\theta)$ is a realization (sample) of $\boldsymbol{Y}$ for $\theta \in \Theta$. The probability distribution of $\boldsymbol{Y}$ is the probability measure $P_{\boldsymbol{Y}}(d\boldsymbol{y})$ on the measurable set $(\mathbb{R}^m, \mathcal{B}_{\mathbb{R}^m})$ (we will simply say on $\mathbb{R}^m$). The Lebesgue measure on $\mathbb{R}^m$ is noted $d\boldsymbol{y}$ and when $P_{\boldsymbol{Y}}(d\boldsymbol{y})$ is written as $p_{\boldsymbol{Y}}(\boldsymbol{y})\, d\boldsymbol{y}$, $p_{\boldsymbol{Y}}$ is the probability density function (pdf) on $\mathbb{R}^m$ of $P_{\boldsymbol{Y}}(d\boldsymbol{y})$ with respect to $d\boldsymbol{y}$. Finally, E denotes the mathematical expectation operator that is such that $E\{\boldsymbol{Y}\} = \int_{\mathbb{R}^m} \boldsymbol{y}\, P_{\boldsymbol{Y}}(d\boldsymbol{y})$.

2. Defining the probability measure $P_{\boldsymbol{H}}$ of random vector $\boldsymbol{H}$

Let $\mathcal{D}_{\text{train}}(\boldsymbol{\eta}) = \{\boldsymbol{\eta}^j, j = 1, \ldots, n_d\}$ be the set of $n_d > 1$ independent realizations $\boldsymbol{\eta}^j \in \mathbb{R}^\nu$, with $\nu \geq 1$, of a second-order $\mathbb{R}^\nu$-valued random variable defined on a probability space $(\Theta, \mathcal{T}, \mathcal{P})$. Let $\underline{\boldsymbol{\eta}}_d \in \mathbb{R}^\nu$ and $[C_d] \in \mathbb{M}_{n_d}$ be the associated empirical estimates of the mean value and the covariance matrix constructed with the points of $\mathcal{D}_{\text{train}}(\boldsymbol{\eta})$,

$$\underline{\boldsymbol{\eta}}_d = \frac{1}{n_d} \sum_{j=1}^{n_d} \boldsymbol{\eta}^j \quad , \quad [C_d] = \frac{1}{n_d - 1} \sum_{j=1}^{n_d} (\boldsymbol{\eta}^j - \underline{\boldsymbol{\eta}}_d) \otimes (\boldsymbol{\eta}^j - \underline{\boldsymbol{\eta}}_d) \, . \tag{2.1}$$

It is assumed that $\mathcal{D}_{\text{train}}(\boldsymbol{\eta})$ is such that

$$\underline{\boldsymbol{\eta}}_d = \mathbf{0}_\nu \quad , \quad [C_d] = [I_\nu] \, . \tag{2.2}$$

Let $\boldsymbol{\eta} \mapsto p_{\boldsymbol{H}}(\boldsymbol{\eta})$ be the probability density function on $\mathbb{R}^\nu$, with respect to the Lebesgue measure $d\boldsymbol{\eta}$, defined by

$$p_{\boldsymbol{H}}(\boldsymbol{\eta}) = \frac{1}{n_d} \sum_{j=1}^{n_d} \frac{1}{(\sqrt{2\pi}\, \hat{s})^\nu} \exp\left(-\frac{1}{2\hat{s}^2} \| \boldsymbol{\eta} - \frac{\hat{s}}{s} \boldsymbol{\eta}^j \|^2 \right) \quad , \quad \forall \boldsymbol{\eta} \in \mathbb{R}^\nu \, , \tag{2.3}$$

where $\hat{s}$ and s are defined by

$$s = \left(\frac{4}{n_d(2+\nu)} \right)^{1/(\nu+4)} \quad , \quad \hat{s} = \frac{s}{\sqrt{s^2 + (n_d - 1)/n_d}} \, . \tag{2.4}$$

Eqs. (2.3) and (2.4) correspond to the Gaussian kernel-density estimation (KDE) constructed using the n_d independent realizations of $\mathcal{D}_{\text{train}}(\boldsymbol{\eta})$ involving the modification [64] of the usual formulation [65, 36, 66], in which s is the Silverman bandwidth. Let $\boldsymbol{H}$ be the second-order $\mathbb{R}^\nu$-valued random variable, defined on a probability space $(\Theta, \mathcal{T}, \mathcal{P})$, whose probability measure $P_{\boldsymbol{H}}(d\boldsymbol{\eta}) = p_{\boldsymbol{H}}(\boldsymbol{\eta})\, d\boldsymbol{\eta}$ on $\mathbb{R}^\nu$ is defined by the probability density function $p_{\boldsymbol{H}}$ given by Eq. (2.3). It can be seen that, for any fixed $n_d > 1$, we have

$$E\{\boldsymbol{H}\} = \int_{\mathbb{R}^\nu} \boldsymbol{\eta}\, P_{\boldsymbol{H}}(d\boldsymbol{\eta}) = \frac{1}{2\hat{s}^2}\, \underline{\boldsymbol{\eta}}_d = \mathbf{0}_\nu \, , \tag{2.5}$$

$$E\{\boldsymbol{H} \otimes \boldsymbol{H}\} = \int_{\mathbb{R}^\nu} \boldsymbol{\eta} \otimes \boldsymbol{\eta}\, P_{\boldsymbol{H}}(d\boldsymbol{\eta}) = \hat{s}^2\, [I_\nu] + \frac{\hat{s}^2}{s^2} \frac{(n_d - 1)}{n_d} [C_d] = [I_\nu] \, . \tag{2.6}$$

Eqs. (2.5) and (2.6) show that $\boldsymbol{H}$ is a normalized $\mathbb{R}^\nu$-valued random variable. The probability density function $p_{\boldsymbol{H}}$ defined by Eq. (2.3) is rewritten, for all $\boldsymbol{\eta}$ in $\mathbb{R}^\nu$, as

$$p_{\boldsymbol{H}}(\boldsymbol{\eta}) = c_\nu\, \xi(\boldsymbol{\eta}) \quad , \quad \xi(\boldsymbol{\eta}) = e^{-\Phi(\boldsymbol{\eta})} \, , \tag{2.7}$$

in which $c_\nu = (\sqrt{2\pi}\, \hat{s})^{-\nu}$ and where $\Phi(\boldsymbol{\eta}) = -\log(\xi(\boldsymbol{\eta}))$ is such that

$$\Phi(\boldsymbol{\eta}) = -\log\left(\frac{1}{n_d} \sum_{j=1}^{n_d} \exp\left(-\frac{1}{2\hat{s}^2} \| \boldsymbol{\eta} - \frac{\hat{s}}{s} \boldsymbol{\eta}^j \|^2 \right) \right) \quad , \quad \forall \boldsymbol{\eta} \in \mathbb{R}^\nu \, . \tag{2.8}$$

3. Short summary of formal results

This section is limited to a summary of essential results, which will be used in Section 4 and which are formally presented.

3.1. Itô stochastic differential equation related to $P_{\boldsymbol{H}}$

We introduce an Itô stochastic differential equation (ISDE) on $\mathbb{R}^\nu$, with initial condition, for which $P_{\boldsymbol{H}}(d\boldsymbol{\eta})$ is the invariant measure. A classical candidate to such an ISDE is written as

$$d\boldsymbol{Y}(t) = \boldsymbol{b}(\boldsymbol{Y}(t))\, dt + d\boldsymbol{W}(t) \quad , \quad t > 0\,, \tag{3.1}$$

$$\boldsymbol{Y}(0) = \boldsymbol{x} \in \mathbb{R}^\nu\,,\ a.s.\,, \tag{3.2}$$

where the drift vector is the function $\boldsymbol{y} \mapsto \boldsymbol{b}(\boldsymbol{y})$ from $\mathbb{R}^\nu$ into $\mathbb{R}^\nu$ defined by

$$\boldsymbol{b}(\boldsymbol{y}) = -\frac{1}{2}\,\nabla\Phi(\boldsymbol{y}) \quad , \quad \forall \boldsymbol{y} \in \mathbb{R}^\nu\,. \tag{3.3}$$

In Eq. (3.1), $\boldsymbol{W}(t) = (W_1(t), \ldots, W_\nu(t))$ is the normalized Wiener stochastic process [67] on $\mathbb{R}^+$, with values in $\mathbb{R}^\nu$, which is a stochastic process with independent increments, such that $\boldsymbol{W}(0) = \boldsymbol{0}_\nu$ $a.s$, and for $0 \le \tau < t < +\infty$, the increment $\Delta\boldsymbol{W}_{\tau t} = \boldsymbol{W}(t) - \boldsymbol{W}(\tau)$ is a Gaussian $\mathbb{R}^\nu$-valued second-order random variable, centered and with a covariance matrix that is written as

$$[C_{\Delta\boldsymbol{W}_{\tau t}}] = E\{\Delta\boldsymbol{W}_{\tau t} \otimes \Delta\boldsymbol{W}_{\tau t}\} = (t-\tau)\,[I_\nu]\,. \tag{3.4}$$

It should be noted that Eqs. (3.1) and (3.2) is equivalent to

$$\boldsymbol{Y}(t) = \boldsymbol{x} + \int_0^t \boldsymbol{b}(\boldsymbol{Y}(\tau))\, d\tau + \int_0^t d\boldsymbol{W}(\tau) \quad , \quad t \ge 0\,. \tag{3.5}$$

In Section 5, we will see that $\{\boldsymbol{Y}(t), t \in \mathbb{R}^+\}$ is a homogeneous diffusion stochastic process, which is asymptotically stationary for $t \to +\infty$. Assuming that the transition probability measure of $\boldsymbol{Y}(t)$ given $\boldsymbol{Y}(0) = \boldsymbol{x}$ admits a density with respect to d$\boldsymbol{y}$, such that, for all $t > 0$, for all $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbb{R}^\nu$, and for any Borelian $\mathcal{B}$ in $\mathbb{R}^\nu$, we have

$$\mathcal{P}\{\,\boldsymbol{Y}(t) \in \mathcal{B} \,|\, \boldsymbol{Y}(0) = \boldsymbol{x}\,\} = \int_{\mathcal{B}} \rho(\boldsymbol{y}, t \,|\, \boldsymbol{x}, 0)\, d\boldsymbol{y}\,, \tag{3.6}$$

$$\lim_{t \to 0_+} \rho(\boldsymbol{y}, t \,|\, \boldsymbol{x}, 0)\, d\boldsymbol{y} = \delta_0(\boldsymbol{y} - \boldsymbol{x})\,, \tag{3.7}$$

$$\int_{\mathbb{R}^\nu} \rho(\boldsymbol{y}, t \,|\, \boldsymbol{x}, 0)\, d\boldsymbol{y} = 1\,. \tag{3.8}$$

3.2. FKP equation associated with the ISDE

For all $\boldsymbol{x}$ in $\mathbb{R}^\nu$, the transition probability density function $(\boldsymbol{y}, t) \mapsto \rho(\boldsymbol{y}, t \,|\, \boldsymbol{x}, 0)$ from $\mathbb{R}^\nu \times \mathbb{R}^+$ into $\mathbb{R}^+$ verifies the following Fokker-Planck (FKP) equation (see for instance [68, 69, 70]),

$$\frac{\partial \rho}{\partial t} + L_{\text{FKP}}(\rho) = 0 \quad , \quad t > 0\,, \tag{3.9}$$

with the initial condition for $t = 0$ defined by Eq. (3.7). The Fokker-Planck operator L_{FKP} can be written, after a small algebraic manipulation and for any sufficiently differentiable function $\boldsymbol{y} \mapsto \mathrm{v}(\boldsymbol{y})$ from $\mathbb{R}^\nu$ into $\mathbb{R}$, as

$$\{L_{\text{FKP}}(\mathrm{v})\}(\boldsymbol{y}) = -\frac{1}{2}\mathrm{div}\left\{ p_{\boldsymbol{H}}(\boldsymbol{y}) \nabla\left(\frac{\mathrm{v}(\boldsymbol{y})}{p_{\boldsymbol{H}}(\boldsymbol{y})}\right)\right\}. \tag{3.10}$$

The detailed balance (the probability current vanishes) is satisfied and the steady state solution of Eq. (3.9) is the pdf $p_{\boldsymbol{H}}$ defined by Eq. (2.3) [70, 71, 72]. We then have, for $\mathrm{v} = p_{\boldsymbol{H}}$,

$$L_{\text{FKP}}(p_{\boldsymbol{H}}) = 0\,. \tag{3.11}$$

3.3. Return to the invariant measure $P_{\boldsymbol{H}}$

The invariant measure (see Proposition 5) $P_{\boldsymbol{H}}(d\boldsymbol{y}) = p_{\boldsymbol{H}}(\boldsymbol{y})\, d\boldsymbol{y}$ is such that, for all $\boldsymbol{y}$ in $\mathbb{R}^\nu$ and for all $t \ge 0$,

$$p_{\boldsymbol{H}}(\boldsymbol{y}) = \int_{\mathbb{R}^\nu} \rho(\boldsymbol{y}, t \,|\, \boldsymbol{x}, 0)\, p_{\boldsymbol{H}}(\boldsymbol{x})\, d\boldsymbol{x}\,. \tag{3.12}$$

The ISDE defined by Eqs. (3.1) and (3.2) admits an asymptotic ($t \to +\infty$) stationary solution whose marginal probability density function of order one is $p_{\boldsymbol{H}}$. Consequently, for all $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbb{R}^\nu$, we have

$$\lim_{t \to +\infty} \rho(\boldsymbol{y}, t \,|\, \boldsymbol{x}, 0) = p_{\boldsymbol{H}}(\boldsymbol{y})\,. \tag{3.13}$$

3.4. Formal formulation of the eigenvalue problem of the FKP operator

The eigenvalue problem, posed in an adapted functional space, is written as

$$L_{\mathrm{FKP}}(\mathrm{v}) = \lambda\,\mathrm{v}\,, \tag{3.14}$$

for which the current must vanish at infinity, yielding the condition,

$$\lim_{\|\boldsymbol{y}\|\to+\infty} p_{\boldsymbol{H}}(\boldsymbol{y})\,\|\,\nabla(p_{\boldsymbol{H}}(\boldsymbol{y})^{-1}\mathrm{v}(\boldsymbol{y}))\,\| = 0\,. \tag{3.15}$$

Continuing the development within a formal framework, such as that used in [72], we introduce the change of function,

$$\mathrm{v}(\boldsymbol{y}) = p_{\boldsymbol{H}}(\boldsymbol{y})^{1/2}\,q(\boldsymbol{y}) \quad , \quad \boldsymbol{y}\in\mathbb{R}^{\nu} \quad , \quad q:\mathbb{R}^{\nu}\to\mathbb{R}\,. \tag{3.16}$$

Let $\hat{L}_{\mathrm{FKP}}$ be the linear operator defined, for q belonging to an admissible set of functions,

$$\{\hat{L}_{\mathrm{FKP}}(q)\}(\boldsymbol{y}) = p_{\boldsymbol{H}}(\boldsymbol{y})^{-1/2}\,L_{\mathrm{FKP}}(p_{\boldsymbol{H}}(\boldsymbol{y})^{1/2}\,q(y)) \quad , \quad \boldsymbol{y}\in\mathbb{R}^{\nu}\,. \tag{3.17}$$

Therefore, the eigenvalue problem defined by Eqs. (3.14) and (3.15) can be rewritten in q as

$$\hat{L}_{\mathrm{FKP}}(q) = \lambda\,q\,, \tag{3.18}$$

with the condition at infinity,

$$\lim_{\|\boldsymbol{y}\|\to+\infty} p_{\boldsymbol{H}}(\boldsymbol{y})^{1/2}\,\|\,\nabla(p_{\boldsymbol{H}}(\boldsymbol{y})^{-1/2}q(\boldsymbol{y}))\,\| = 0\,. \tag{3.19}$$

Remark 1 (Another algebraic representation of operator $\hat{L}_{\mathrm{FKP}}$). Using Eq. (2.7), which shows that $p_{\boldsymbol{H}}(\boldsymbol{y})^{-1}\nabla p_{\boldsymbol{H}}(\boldsymbol{y}) = -\nabla\Phi$, and using Eqs. (3.10) and (3.17), it can be seen that

$$\{\hat{L}_{\mathrm{FKP}}(q)\}(\boldsymbol{y}) = \mathcal{V}(\boldsymbol{y})\,q(\boldsymbol{y}) - \frac{1}{2}\nabla^2 q(\boldsymbol{y}) \quad , \quad \boldsymbol{y}\in\mathbb{R}^{\nu}\,, \tag{3.20}$$

in which ∇^2 is the Laplacian operator in $\mathbb{R}^{\nu}$ and where $\boldsymbol{y}\mapsto\mathcal{V}(\boldsymbol{y})$ is the function from $\mathbb{R}^{\nu}$ into $\mathbb{R}$, which is defined, for all $\boldsymbol{y}$ in $\mathbb{R}^{\nu}$, as

$$\mathcal{V}(\boldsymbol{y}) = \frac{1}{8}\|\,\nabla\Phi(\boldsymbol{y})\,\|^2 - \frac{1}{4}\nabla^2\Phi(\boldsymbol{y})\,. \tag{3.21}$$

3.5. Properties of operator $\hat{L}_{\mathrm{FKP}}$

Let δq be a function from $\mathbb{R}^{\nu}$ into $\mathbb{R}$, belonging to the admissible set that allows the evaluation of the bracket

$$\langle\hat{L}_{\mathrm{FKP}}(q)\,,\delta q\rangle = \int_{\mathbb{R}^{\nu}}\{\hat{L}_{\mathrm{FKP}}(q)\}(\boldsymbol{y})\;\delta q(\boldsymbol{y})\,d\boldsymbol{y}\,.$$

Removing $\boldsymbol{y}$ and using Eq. (3.17) with Eq. (3.10) yields

$$\langle\hat{L}_{\mathrm{FKP}}(q)\,,\delta q\rangle = -\frac{1}{2}\int_{\mathbb{R}^{\nu}}(p_{\boldsymbol{H}}^{-1/2}\,\delta q)\,\mathrm{div}\{p_{\boldsymbol{H}}\,\nabla(p_{\boldsymbol{H}}^{-1/2}\,q)\}\,d\boldsymbol{y}\,. \tag{3.22}$$

Using the condition at infinity, defined by Eq. (3.19), Eq. (3.22) can be rewritten as,

$$\langle\hat{L}_{\mathrm{FKP}}(q)\,,\delta q\rangle = \frac{1}{2}\int_{\mathbb{R}^{\nu}} p_{\boldsymbol{H}}\,\langle\nabla(p_{\boldsymbol{H}}^{-1/2}\,q)\,,\nabla(p_{\boldsymbol{H}}^{-1/2}\,\delta q)\rangle_{\mathbb{R}^{\nu}}\,d\boldsymbol{y}\,. \tag{3.23}$$

(a) Eq. (3.23) shows that $\hat{L}_{\mathrm{FKP}}$ is a symmetric and positive operator.
(b) Eqs. (3.11) and (3.17) show that

$$\hat{L}_{\mathrm{FKP}}(q_0) = 0 \quad \text{for} \quad q_0 = p_{\boldsymbol{H}}^{1/2}\,. \tag{3.24}$$

In Proposition 5, it will be proven that the ISDE defined by Eq. (3.1), with the initial condition defined by (3.2), has a unique solution and a unique invariant measure $p_{\boldsymbol{H}}(\boldsymbol{y})\,d\boldsymbol{y}$. Consequently the dimension of the null space of operator L_{FKP} is 1. Since $p_{\boldsymbol{H}}(\boldsymbol{y})\,d\boldsymbol{y}$ is a bounded positive measure (probability measure), the right-hand side of Eq. (3.23) shows that the null space of $\hat{L}_{\mathrm{FKP}}$, which is also of dimension 1, is constituted of the function $q_0 = p_{\boldsymbol{H}}^{1/2}$. For $q = \delta q \neq q_0$, and $\|q_0\|\neq 0$, we have $\langle\hat{L}_{\mathrm{FKP}}(q)\,,q\rangle > 0$. Therefore, $\hat{L}_{\mathrm{FKP}}$ is a positive operator (in the quotient space by the null space).

Hypothesis 1 (On the spectrum of operator $\hat{L}_{\text{FKP}}$). It is assumed that $p_{\boldsymbol{H}}$ defined by Eq. (2.3), which is constructed with the n_d points $\{\boldsymbol{\eta}^j, j = 1, \ldots, n_d\}$ of the training dataset, is such that the spectrum of $\hat{L}_{\text{FKP}}$ is countable. Due to (a) and (b), we then deduce that the eigenvalues of $\hat{L}_{\text{FKP}}$ (defined by Eqs. (3.18) and (3.19)) are positive except one that is zero. We will also assume that the multiplicity of each eigenvalue is finite.

3.6. Eigenvalue problem for operator $\hat{L}_{\text{FKP}}$

Under Hypothesis 1, the eigenvalue problem $\hat{L}_{\text{FKP}}(q_\alpha) = \lambda_\alpha q_\alpha$ for operator $\hat{L}_{\text{FKP}}$, with the condition defined by Eq. (3.19), is such that

$$0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \ldots , \tag{3.25}$$

the multiplicity of each eigenvalue being finite. We will assume that the family $\{q_\alpha, \alpha \in \mathbb{N}\}$ of the eigenfunctions is a Hilbert basis of $L^2(\mathbb{R}^\nu)$. We then have

$$\langle q_\alpha, q_\beta \rangle_{L^2} = \int_{\mathbb{R}^\nu} q_\alpha(\boldsymbol{y})\, q_\beta(\boldsymbol{y})\, d\boldsymbol{y} = \delta_{\alpha\beta} . \tag{3.26}$$

The eigenfunction q_0 associated with $\lambda_0 = 0$, is such that (see Eq. (3.24)),

$$q_0 = p_{\boldsymbol{H}}^{1/2} \quad , \quad \| q_0 \|_{L^2} = 1 , \tag{3.27}$$

and we have

$$\sum_{\alpha \in \mathbb{N}} q_\alpha(\boldsymbol{y})\, q_\alpha(\boldsymbol{x})\, d\boldsymbol{y} = \delta_0(\boldsymbol{y} - \boldsymbol{x}) . \tag{3.28}$$

From Eqs. (3.26) and (3.27), it can be deduced that

$$\forall \alpha \geq 1 \quad , \quad \int_{\mathbb{R}^\nu} p_{\boldsymbol{H}}(\boldsymbol{y})^{1/2}\, q_\alpha(\boldsymbol{y})\, d\boldsymbol{y} = 0 . \tag{3.29}$$

3.7. Nonstationary solution of the Fokker-Planck equation with initial condition

The transition probability density function $\rho(\boldsymbol{y}, t \,|\, \boldsymbol{x}, 0)$ introduced in Section 3.1 and satisfying Eq. (3.9) with the initial condition defined by Eq. (3.7), can be written, using the Hilbert basis $\{q_\alpha, \alpha \in \mathbb{N}\}$ defined in Section 3.6, as

$$\rho(\boldsymbol{y}, t \,|\, \boldsymbol{x}, 0) = p_{\boldsymbol{H}}(\boldsymbol{y})^{1/2}\, p_{\boldsymbol{H}}(\boldsymbol{x})^{-1/2} \sum_{\alpha \in \mathbb{N}} e^{-\lambda_\alpha t} q_\alpha(\boldsymbol{y})\, q_\alpha(\boldsymbol{x}) \quad , \quad t > 0 . \tag{3.30}$$

This representation of $\rho(\boldsymbol{y}, t \,|\, \boldsymbol{x}, 0)$, defined by Eq. (3.30), actually satisfied all the required properties:
Eqs. (3.25), (3.27), and (3.29) yield $\int_{\mathbb{R}^\nu} \rho(\boldsymbol{y}, t \,|\, \boldsymbol{x}, 0)\, d\boldsymbol{y} = 1$.
Eqs. (3.25) and (3.27) yield $\lim_{t \to +\infty} \rho(\boldsymbol{y}, t \,|\, \boldsymbol{x}, 0)\, d\boldsymbol{y} = p_{\boldsymbol{H}}(\boldsymbol{y})$.
Eq. (3.28) yields $\lim_{t \to 0_+} \rho(\boldsymbol{y}, t \,|\, \boldsymbol{x}, 0)\, d\boldsymbol{y} = \delta_0(\boldsymbol{y} - \boldsymbol{x})$.
Eqs. (3.27) and (3.29) yield $\int_{\mathbb{R}^\nu} \rho(\boldsymbol{y}, t \,|\, \boldsymbol{x}, 0)\, p_{\boldsymbol{H}}(\boldsymbol{x})\, d\boldsymbol{x} = p_{\boldsymbol{H}}(\boldsymbol{y})$ that is Eq. (3.12).

4. Time-dependent kernel, its associated operator, and finite approximation

In this section, we define the kernel $k_t(\boldsymbol{y}, \boldsymbol{x})$ and give its basic properties directly deduced from the properties of $\rho(\boldsymbol{y}, t \,|\, \boldsymbol{x}, 0)$ and $p_{\boldsymbol{H}}(\boldsymbol{y})$, without using the spectral representation defined by Eq. (3.30). From the spectral representation of $k_t(\boldsymbol{y}, \boldsymbol{x})$, we deduce its spectral representation using the spectral representation of $\rho(\boldsymbol{y}, t \,|\, \boldsymbol{x}, 0)$, defined by Eq. (3.30). Finally, we define the linear operator K_t associated with kernel k_t and we give the spectral representation of operator K_t.

4.1. Definition of the kernel k_t and its basic probabilistic properties

The kernel $k_t(\boldsymbol{y}, \boldsymbol{x})$ associated with the transition probability density function $\rho(\boldsymbol{y}, t \,|\, \boldsymbol{x}, 0)$ is defined as follows.

Definition 1 (Kernel k_t on $\mathbb{R}^\nu \times \mathbb{R}^\nu$). For every fixed $t > 0$, the kernel function $(\boldsymbol{y}, \boldsymbol{x}) \mapsto k_t(\boldsymbol{y}, \boldsymbol{x})$, from $\mathbb{R}^\nu \times \mathbb{R}^\nu$ into $\mathbb{R}^+$, is defined by

$$k_t(\boldsymbol{y}, \boldsymbol{x}) = \frac{\rho(\boldsymbol{y}, t \,|\, \boldsymbol{x}, 0)}{p_{\boldsymbol{H}}(\boldsymbol{y})} \,. \tag{4.1}$$

The following Lemma gives basic properties of kernel k_t.

Lemma 1 (Properties of kernel k_t). *For every fixed $t > 0$, and for all $\boldsymbol{y}$ and $\boldsymbol{x}$ in $\mathbb{R}^\nu$, we have the following properties:*

$$(a) \quad \int_{\mathbb{R}^\nu} k_t(\boldsymbol{y}, \boldsymbol{x})\, p_{\boldsymbol{H}}(\boldsymbol{y})\, d\boldsymbol{y} = 1 \quad , \quad \int_{\mathbb{R}^\nu} k_t(\boldsymbol{y}, \boldsymbol{x})\, p_{\boldsymbol{H}}(\boldsymbol{x})\, d\boldsymbol{x} = 1 \,, \tag{4.2}$$

$$(b) \quad \int_{\mathbb{R}^\nu} \int_{\mathbb{R}^\nu} k_t(\boldsymbol{y}, \boldsymbol{x})\, p_{\boldsymbol{H}}(\boldsymbol{y})\, p_{\boldsymbol{H}}(\boldsymbol{x})\, d\boldsymbol{y}\, d\boldsymbol{x} = 1 \,, \tag{4.3}$$

$$(c) \quad \lim_{t \to 0_+} k_t(\boldsymbol{y}, \boldsymbol{x})\, p_{\boldsymbol{H}}(\boldsymbol{y})\, d\boldsymbol{y} = \delta_0(\boldsymbol{y} - \boldsymbol{x}) \quad , \quad \lim_{t \to +\infty} k_t(\boldsymbol{y}, \boldsymbol{x}) = 1 \,, \tag{4.4}$$

$$(d) \quad (\boldsymbol{y}, \boldsymbol{x}) \mapsto \rho(\boldsymbol{y}, t \,|\, \boldsymbol{x}, 0) \in C^0(\mathbb{R}^\nu \times \mathbb{R}^\nu) \Rightarrow (\boldsymbol{y}, \boldsymbol{x}) \mapsto k_t(\boldsymbol{y}, \boldsymbol{x}) \in C^0(\mathbb{R}^\nu \times \mathbb{R}^\nu) \,, \tag{4.5}$$

$$(e) \quad k_t(\boldsymbol{y}, \boldsymbol{x}) = k_t(\boldsymbol{x}, \boldsymbol{y}) \,. \tag{4.6}$$

PROOF. (Lemma 1). Definition 1 is used.
(a) The first equation in Eq. (4.2) is due to Eq. (3.8) and the second one is due to Eq. (3.12).
(b) Eq. (4.3) is directly deduced from Eq. (4.2).
(c) The first equation in Eq. (4.4) is due to Eq. (3.7) and the second one is due to Eq. (3.13).
(d) For all $\boldsymbol{y}$ in $\mathbb{R}^\nu$, $p_{\boldsymbol{H}}(y) > 0$ and $p_{\boldsymbol{H}} \in C^0(\mathbb{R}^\nu)$. The hypothesis $\rho(\cdot, t \,|\, \cdot, 0) \in C^0(\mathbb{R}^\nu \times \mathbb{R}^\nu)$ yields Eq. (4.5).
(e) let $p_{\boldsymbol{Y}(t),\boldsymbol{H}}(\boldsymbol{y}, t; \boldsymbol{x}, 0)$ be the joint pdf of $\boldsymbol{Y}(t)$ with $\boldsymbol{H}$ in which $\boldsymbol{Y}(t)$ is the solution of Eq. (3.1) for fixed $t > 0$, with the random initial condition $\boldsymbol{Y}(0) = \boldsymbol{H}$. We have the classical property related to the definition of the invariant measure, $p_{\boldsymbol{Y}(t)}(\boldsymbol{y}, t) = \int_{\mathbb{R}^\nu} p_{\boldsymbol{Y}(t),\boldsymbol{H}}(\boldsymbol{y}, t; \boldsymbol{x}, 0)\, d\boldsymbol{x} = \int_{\mathbb{R}^\nu} \rho(\boldsymbol{y}, t \,|\, \boldsymbol{x}, 0)\, p_{\boldsymbol{H}}(\boldsymbol{x})\, d\boldsymbol{x} = p_{\boldsymbol{H}}(\boldsymbol{y})$. For all $\boldsymbol{y}$ and $\boldsymbol{x}$ in $\mathbb{R}^\nu$, and for $t > 0$, we have, $\rho(\boldsymbol{y}, t \,|\, \boldsymbol{x}, 0)\, p_{\boldsymbol{H}}(\boldsymbol{x}) = p_{\boldsymbol{Y}(t),\boldsymbol{H}}(\boldsymbol{y}, t; \boldsymbol{x}, 0) = p_{\boldsymbol{H},\boldsymbol{Y}(t)}(\boldsymbol{x}, 0; \boldsymbol{y}, t) = \rho(\boldsymbol{x}, 0 \,|\, \boldsymbol{y}, t)\, p_{\boldsymbol{Y}(t)}(\boldsymbol{y}, t) = \rho(\boldsymbol{x}, 0 \,|\, \boldsymbol{y}, t)\, p_{\boldsymbol{H}}(\boldsymbol{y})$. Consequently, $k_t(\boldsymbol{y}, \boldsymbol{x}) = \rho(\boldsymbol{y}, t \,|\, \boldsymbol{x}, 0)\, p_{\boldsymbol{H}}(\boldsymbol{x}) / (p_{\boldsymbol{H}}(\boldsymbol{y})\, p_{\boldsymbol{H}}(\boldsymbol{x})) = \rho(\boldsymbol{x}, 0 \,|\, \boldsymbol{y}, t)\, p_{\boldsymbol{H}}(\boldsymbol{y}) / (p_{\boldsymbol{H}}(\boldsymbol{y})\, p_{\boldsymbol{H}}(\boldsymbol{x})) = k_t(\boldsymbol{x}, \boldsymbol{y})$.

4.2. Hypothesis and properties of kernel k_t from its representation

In Section 3, we introduced an hypothesis of existence of a discrete (countable) spectrum $\{\lambda_\alpha, \alpha \in \mathbb{N}\}$ of the FKP operator $\hat{L}_{\text{FKP}}$. In this section, we study the spectral representation of kernel k_t, for every fixed $t > 0$, deduced from the time-dependent spectral representation of $\rho(\boldsymbol{y}, t \,|\, \boldsymbol{x}, 0)$, defined by Eq. (3.30).

Definition 2 (Hilbert space $\mathbb{H} = L^2(\mathbb{R}^\nu; p_{\boldsymbol{H}})$). Let $\mathbb{H} = L^2(\mathbb{R}^\nu; p_{\boldsymbol{H}})$ be the Hilbert space of the square-integrable real-valued functions on $\mathbb{R}^\nu$, with respect to the probability measure $p_{\boldsymbol{H}}(\boldsymbol{y})\, d\boldsymbol{y}$ on $\mathbb{R}^\nu$, equipped with the inner product and the associated norm,

$$\langle u \,, \mathrm{v} \rangle_{\mathbb{H}} = \int_{\mathbb{R}^\nu} u(\boldsymbol{y})\, \mathrm{v}(\boldsymbol{y})\, p_{\boldsymbol{H}}(\boldsymbol{y})\, d\boldsymbol{y} \quad , \quad \|u\|_{\mathbb{H}} = \langle u \,, u \rangle_{\mathbb{H}}^{1/2} \,. \tag{4.7}$$

Lemma 2 (Hilbert basis in $\mathbb{H}$). *Let $\{q_\alpha, \alpha \in \mathbb{N}\}$ be the Hilbert basis of $L^2(\mathbb{R}^\nu)$ introduced in Section 3.6. For all $\alpha \in \mathbb{N}$, we defined the real-valued function ψ_α on $\mathbb{R}^\nu$ such that*

$$\psi_\alpha(\boldsymbol{y}) = q_\alpha(\boldsymbol{y})\, p_{\boldsymbol{H}}(\boldsymbol{y})^{-1/2} \quad , \quad \forall \boldsymbol{y} \in \mathbb{R}^\nu \,. \tag{4.8}$$

Then, $\{\psi_\alpha, \alpha \in \mathbb{N}\}$ is a Hilbert basis of $\mathbb{H}$ and we have,

$$(a) \quad \psi_\alpha \in \mathbb{H} \quad , \quad \|\psi_\alpha\|_{\mathbb{H}} = \|q_\alpha\|_{L^2} = 1 \quad , \quad \forall \alpha \in \mathbb{N}\,, \tag{4.9}$$

$$(b) \quad \langle \psi_\alpha \,, \psi_\beta \rangle_{\mathbb{H}} = \int_{\mathbb{R}^\nu} \psi_\alpha(\boldsymbol{y})\, \psi_\beta(\boldsymbol{y})\, p_{\boldsymbol{H}}(\boldsymbol{y})\, d\boldsymbol{y} = \delta_{\alpha\beta} \quad , \quad \forall (\alpha, \beta) \in \mathbb{N} \times \mathbb{N}\,, \tag{4.10}$$

$$(c) \quad \psi_0(\boldsymbol{y}) = 1\,,\ \forall \boldsymbol{y} \in \mathbb{R}^\nu \quad , \quad \|\psi_0\|_{\mathbb{H}} = 1\,, \tag{4.11}$$

$$(d) \quad \int_{\mathbb{R}^\nu} \psi_\alpha(\boldsymbol{y})\, p_{\boldsymbol{H}}(\boldsymbol{y})\, d\boldsymbol{y} = 0 \quad , \quad \forall \alpha \in \mathbb{N}^*\,, \tag{4.12}$$

$$(e) \quad \sum_{\alpha \in \mathbb{N}} \psi_\alpha(\boldsymbol{y})\, \psi_\beta(\boldsymbol{x})\, p_{\boldsymbol{H}}(\boldsymbol{y})\, d\boldsymbol{y} = \delta_0(\boldsymbol{y} - \boldsymbol{x}) \quad , \quad \forall (\boldsymbol{y}, \boldsymbol{x}) \in \mathbb{R}^\nu \times \mathbb{R}^\nu\,. \tag{4.13}$$

PROOF. (Lemma 2).
(a) Since $p_{\boldsymbol{H}}^{-1/2} \in C^0(\mathbb{R}^\nu)$, Eqs. (4.8) yields Eq. (4.9); combined with Eq. (3.29), this yields $\|\psi_\alpha\|_{\mathbb{H}} = \int_{\mathbb{R}^\nu} q_\alpha(\boldsymbol{y})^2\, d\boldsymbol{y} = \|q_\alpha\|_{L^2} = 1$.
(b) Using Eq. (3.26) yields $\langle \phi_\alpha \,, \psi_\beta \rangle_{\mathbb{H}} = \int_{\mathbb{R}^\nu} q_\alpha(\boldsymbol{y})\, q_\beta(\boldsymbol{y})\, d\boldsymbol{y} = \langle q_\alpha \,, q_\beta \rangle_{L^2} = \delta_{\alpha\beta}$. Thus $\{\psi_\alpha, \alpha \in \mathbb{N}\}$ is an orthonormal family in $\mathbb{H}$. For all u in $L^2(\mathbb{R}^\nu)$, the linear mapping $u \mapsto \mathrm{v} = u p_{\boldsymbol{H}}^{-1/2}$ is a continuous injection from $L^2(\mathbb{R}^\nu)$ into $\mathbb{H}$ with $\|\mathrm{v}\|_{\mathbb{H}} = \|u\|_{L^2}$. Therefore, $\{\psi_\alpha, \alpha \in \mathbb{N}\}$ is a Hilbert basis of $\mathbb{H}$.
(c) The two equations in Eq. (4.11) are directly deduced from Eqs. (3.27) and (4.8).
(d) Since $\langle \psi_0 \,, \psi_\alpha \rangle_{\mathbb{H}} = 0$ for all $\alpha \in \mathbb{N}^*$, we obtain Eq. (4.12).
(e) Since $\{\psi_\alpha, \alpha \in \mathbb{N}\}$ is a Hilbert basis of $\mathbb{H}$, Eq. (4.13) holds.

Proposition 1 (Spectral representation of kernel k_t). *Let $\{\psi_\alpha, \alpha \in \mathbb{N}\}$ be the Hilbert basis of $\mathbb{H}$ defined in Lemma 2.*
(a) For every fixed $t > 0$, the symmetric kernel k_t can be written, for all $\boldsymbol{y}$ and $\boldsymbol{x}$ in $\mathbb{R}^\nu$, as

$$k_t(\boldsymbol{y}, \boldsymbol{x}) = \sum_{\alpha \in \mathbb{N}} b_\alpha(t)\, \psi_\alpha(\boldsymbol{y})\, \psi_\alpha(\boldsymbol{x})\,, \tag{4.14}$$

in which the family of positive real numbers $\{b_\alpha(t) = exp(-\lambda_\alpha\, t)\,, \alpha \in \mathbb{N}\}$, is such that

$$1 = b_0(t) > b_1(t) \geq b_2(t) \geq \ldots\,. \tag{4.15}$$

(b) If for every fixed $t > 0$, kernel k_t satisfies

$$\int_{\mathbb{R}^\nu} \int_{\mathbb{R}^\nu} k_t(\boldsymbol{y}, \boldsymbol{x})^2\, p_{\boldsymbol{H}}(\boldsymbol{y})\, p_{\boldsymbol{H}}(\boldsymbol{x})\, d\boldsymbol{y}\, d\boldsymbol{x} = c_t^2 < +\infty\,, \tag{4.16}$$

where $c_t > 1$ is a positive constant depending on t, then,

$$\sum_{\alpha \in \mathbb{N}} b_\alpha(t)^2 = c_t^2 < +\infty\,. \tag{4.17}$$

PROOF. (Proposition 1).
(a) Substituting Eq. (3.30) with $q_\alpha(\boldsymbol{y}) = \psi_\alpha(\boldsymbol{y})\, p_{\boldsymbol{H}}(\boldsymbol{y})^{1/2}$ (see Eq. (4.8)) into Eq. (4.1) yields Eq. (4.14). From Eq. (3.25) and since $b_\alpha(t) = exp(-\lambda_\alpha\, t)$, we obtain Eq. (4.15).
(b) Assuming Eq. (4.16), substituting Eq. (4.14) into Eq. (4.16), and using Eq. (4.10) yields Eq. (4.17). Since $b_0(t) = 1$, it can be deduced that $c_t^2 > 1$ and thus, $c_t > 1$.

4.3. Hilbert-Schmidt operator K_t associated with kernel k_t

We now introduce the linear operator in $\mathbb{H}$, defined by kernel k_t, and we study its properties and spectrum.

Definition 3 (Operator K_t associated with kernel k_t). For every fixed $t > 0$, we defined the linear operator K_t from $\mathbb{H}$ into $\mathbb{H}$ such that, for all u and v in $\mathbb{H}$,

$$\langle K_t\, u \,, \mathrm{v} \rangle_{\mathbb{H}} = \int_{\mathbb{R}^\nu} \int_{\mathbb{R}^\nu} k_t(\boldsymbol{y}, \boldsymbol{x})\, u(\boldsymbol{x})\, \mathrm{v}(\boldsymbol{y})\, p_{\boldsymbol{H}}(\boldsymbol{y})\, p_{\boldsymbol{H}}(\boldsymbol{x})\, d\boldsymbol{y}\, d\boldsymbol{x}\,, \tag{4.18}$$

where the symmetric kernel k_t verifies the condition defined by Eq. (4.16).

Proposition 2 (K_t as a Hilbert-Schmidt operator in $\mathbb{H}$). *For every fixed $t > 0$, let K_t be the continuous linear operator defined by Eq.* (4.18)*, in which kernel k_t is symmetric on $\mathbb{R}^\nu \times \mathbb{R}^\nu$, and verifies Eq.* (4.16)*.*
(a) For all u and v in $\mathbb{H}$, operator K_t is such that

$$\langle K_t\, u\,,\mathrm{v}\rangle_{\mathbb{H}} = \sum_{\alpha\in\mathbb{N}} b_\alpha(t)\,\langle u\,,\psi_\alpha\rangle_{\mathbb{H}}\,\langle \mathrm{v}\,,\psi_\alpha\rangle_{\mathbb{H}}\,, \tag{4.19}$$

and is a positive symmetric operator in $\mathbb{H}$. For all u in $\mathbb{H}$,

$$K_t\, u = \sum_{\alpha\in\mathbb{N}} b_\alpha(t)\,\langle u\,,\psi_\alpha\rangle_{\mathbb{H}}\,\psi_\alpha\,. \tag{4.20}$$

(b) For all α in $\mathbb{N}$, $\psi_\alpha \in \mathbb{H}$ is the eigenfunction independent of t, associated with the positive eigenvalue $b_\alpha(t)$, satisfying Eq. (4.15)*, of operator K_t,*

$$K_t\,\psi_\alpha = b_\alpha(t)\,\psi_\alpha \quad , \quad b_\alpha(t) = \exp(-\lambda_\alpha t) \quad , \quad \alpha\in\mathbb{N}\,, \tag{4.21}$$

which shows that, for all α and β in $\mathbb{N}$,

$$\langle K_t\,\psi_\alpha\,,\psi_\beta\rangle_{\mathbb{H}} = b_\alpha(t)\,\delta_{\alpha\beta}\,. \tag{4.22}$$

(c) For all u in $\mathbb{H}$, we have

$$\|K_t\,u\|^2_{\mathbb{H}} = \sum_{\alpha\in\mathbb{N}} b_\alpha(t)^2\,\langle u\,,\psi_\alpha\rangle^2_{\mathbb{H}}\,, \tag{4.23}$$

and for Hilbert basis $\{\psi_\alpha\,,\alpha\in\mathbb{N}\}$ of $\mathbb{H}$,

$$\sum_{\alpha\in\mathbb{N}} \|K_t\,\psi_\alpha\|^2_{\mathbb{H}} = c_t^2 < +\infty\,, \tag{4.24}$$

where c_t^2, defined by Eq. (4.16)*, is such that $c_t^2 = \sum_{\alpha\in\mathbb{N}} b_\alpha(t)^2$, and therefore, is a Hilbert-Schmidt operator in $\mathbb{H}$.*

PROOF. (Proposition 2). Under the condition defined by Eq. (4.16), it is well known that operator K_t is continuous from $\mathbb{H}$ into $\mathbb{H}$.
(a) Substituting Eq. (4.14) into Eq. (4.18) and using Eq. (4.7) yield Eq. (4.19). This equation shows that K_t is a symmetric and positive operator because $\langle K_t\,u\,,u\rangle_{\mathbb{H}} > 0$ for all u in $\mathbb{H}$ with $\|u\|_{\mathbb{H}} \neq 0$. Note that Eq. (4.20) is directly deduced from Eq. (4.19).
(b) From Eqs. (4.8) and (4.9), we have $\psi_\alpha \in \mathbb{H}$. Taking $u = \psi_\beta$ in Eq. (4.20) and using Eq. (4.10) yield $K_t\,\psi_\beta = \sum_{\alpha\in\mathbb{N}} b_\alpha(t)\,\langle\psi_\beta\,,\psi_\alpha\rangle_{\mathbb{H}}\,\psi_\alpha = b_\beta(t)\,\psi_\beta$, Eq. (4.22) is obtained from Eq. (4.19) by taking $u = \psi_\alpha$ and $\mathrm{v} = \psi_\beta$. The relationship between $b_\alpha(t)$ and λ_α comes from Proposition 1.
(c) Using Eqs. (4.20) and (4.10) yields Eq. (4.23). Taking $u = \psi_\alpha$ in Eq. (4.23) yields $\|K_t\,\psi_\alpha\|^2_{\mathbb{H}} = b_\alpha(t)^2$. From Eq. (4.17) and $\sum_{\alpha\in\mathbb{N}}\|K_t\,\psi_\alpha\|^2_{\mathbb{H}} = \sum_{\alpha\in\mathbb{N}} b_\alpha(t)^2$, we obtain Eq. (4.24). It can be deduced (see for instance [73]) that K_t is a Hilbert-Schmidt operator in $\mathbb{H}$.

4.4. Finite approximation of operator K_t and of its eigenvalue problem

The Hilbert-Schmidt operator K_t defined by Eq. (4.18), operates in infinite dimension. The Hilbert basis $\{\psi_\alpha\,,\alpha\in\mathbb{N}\}$ (which relates to the Hilbert basis $\{q_\alpha\,,\alpha\in\mathbb{N}\}$, see Lemma 2 and Eq. (3.18)) is not explicitly known, thereby preventing the use of the representation defined by Eq. (4.20). We must thus construct a finite approximation of K_t. Since $\mathbb{R}^\nu$ is an unbounded set and ν can be very large, classical discretization such as finite-difference or finite-element methods (see [74, 75, 76, 77] for Fokker-Planck equation and [78] for fractional Fokker-Planck equation) or such methods based on shape-morphing modes for solving the Fokker-Planck equation as proposed in [79], are not directly adapted for solving the eigenvalue problem of operator $\hat{L}_{\mathrm{FKP}}$. Another classical method consists in introducing a finite family of functions in $\mathbb{H}$, generating a finite dimension subspace of $\mathbb{H}$, and in performing the projection of K_t on this finite subspace. Such an approach is not really adapted to operator $\hat{L}_{\mathrm{FKP}}$ for which a large number of eigenvalues and associated eigenfunctions have to be computed. It should be noted that a related problem, but distinct from the one addressed, is that of the numerical method for Schrödinger operator and the associated equation (see for instance [80] for solving the Schrödinger Equation, [81] for the solution of the Schrödinger equation by spectral methods, [82] for numerically solving the time-dependent Schrödinger equation, and [83] for the numerical solution of the Schrödinger

equation using finite-difference method). Nevertheless, such approaches are not well adapted to the objective of the actual developments, which has been detailed in Section 1. We then propose to use a statistical sampling of $\mathbb{R}^\nu$ equipped with the probability measure $p_{\boldsymbol{H}}(\boldsymbol{\eta})\, d\boldsymbol{\eta}$, which will be well adapted to our objective of performing a construction connected to the DMAPS approach.

Proposition 3 (Probabilistic interpretation of the bilinear form $\langle K_t u\,,\mathrm{v}\rangle_{\mathbb{H}}$). *Let us assume that, for every fixed $t > 0$, we have $(\boldsymbol{y},\boldsymbol{x}) \mapsto \rho(\boldsymbol{y},t\,|\,\boldsymbol{x},0) \in C^0(\mathbb{R}^\nu \times \mathbb{R}^\nu)$. From Eq. (4.5), it can be deduced that $(\boldsymbol{y},\boldsymbol{x}) \mapsto k_t(\boldsymbol{y},\boldsymbol{x}) \in C^0(\mathbb{R}^\nu \times \mathbb{R}^\nu)$. For every fixed $t > 0$, for all u and v in $\mathbb{H} \cap C^0(\mathbb{R}^\nu)$, the restriction to $(\mathbb{H} \cap C^0(\mathbb{R}^\nu)) \times (\mathbb{H} \cap C^0(\mathbb{R}^\nu))$ of the bilinear form $(u,\mathrm{v}) \mapsto \langle K_t u\,,\mathrm{v}\rangle_{\mathbb{H}}$, defined on $\mathbb{H}\times\mathbb{H}$ by Eq. (4.18) with the continuous symmetric function k_t verifying Eq. (4.16), can be written as*

$$\langle K_t u\,,\mathrm{v}\rangle_{\mathbb{H}} = E\{k_t(\boldsymbol{H},\underset{\sim}{\boldsymbol{H}})\, u(\underset{\sim}{\boldsymbol{H}})\, \mathrm{v}(\boldsymbol{H})\}\,, \tag{4.25}$$

in which $\underset{\sim}{\boldsymbol{H}}$ is an independent copy of $\boldsymbol{H}$. The joint probability measure $P_{\boldsymbol{H},\underset{\sim}{\boldsymbol{H}}}(d\boldsymbol{y},d\boldsymbol{x})$ of $\boldsymbol{H}$ with $\underset{\sim}{\boldsymbol{H}}$ is $p_{\boldsymbol{H}}(\boldsymbol{y}) \times p_{\boldsymbol{H}}(\boldsymbol{x})\, d\boldsymbol{y}\, d\boldsymbol{x}$. The real-valued random variables $u(\underset{\sim}{\boldsymbol{H}})$, $\mathrm{v}(\boldsymbol{H})$, and the positive-valued random variable $k_t(\boldsymbol{H},\underset{\sim}{\boldsymbol{H}})$, defined on $(\Theta,\mathcal{T},\mathcal{P})$, are second-order random variables,

$$E\{u(\underset{\sim}{\boldsymbol{H}})^2\} < +\infty \quad , \quad E\{\mathrm{v}(\boldsymbol{H})^2\} < +\infty \quad , \quad E\{k_t(\boldsymbol{H},\underset{\sim}{\boldsymbol{H}})^2\} < +\infty\,. \tag{4.26}$$

Let $Z_{t,u,\mathrm{v}}$ be the real-valued random variable, defined on $(\Theta,\mathcal{T},\mathcal{P})$, such that

$$Z_{t,u,\mathrm{v}} = k_t(\boldsymbol{H},\underset{\sim}{\boldsymbol{H}})\, u(\underset{\sim}{\boldsymbol{H}})\, \mathrm{v}(\boldsymbol{H})\,. \tag{4.27}$$

Then, $Z_{t,u,\mathrm{v}}$ is such that,

$$E\{Z_{t,u,\mathrm{v}}\} < +\infty\,. \tag{4.28}$$

PROOF. (Proposition 3).
(a) Since u and v are continuous functions in $\mathbb{H}$, $u(\boldsymbol{H})$ and $\mathrm{v}(\boldsymbol{H})$ are real-valued random variables defined on $(\Theta,\mathcal{T},\mathcal{P})$ and are second-order because,

$$E\{u(\underset{\sim}{\boldsymbol{H}})^2\} = \int_{\mathbb{R}^\nu} u(\boldsymbol{x})^2 p_{\boldsymbol{H}}(\boldsymbol{x})\, d\boldsymbol{x} = \|u\|^2_{\mathbb{H}} < +\infty \quad , \quad E\{\mathrm{v}(\boldsymbol{H})^2\} = \int_{\mathbb{R}^\nu} \mathrm{v}(\boldsymbol{y})^2 p_{\boldsymbol{H}}(\boldsymbol{y})\, d\boldsymbol{y} = \|\mathrm{v}\|^2_{\mathbb{H}} < +\infty\,.$$

Since $k_t \in C^0(\mathbb{R}^\nu \times \mathbb{R}^\nu)$ (see Eq. (4.5)) and due to Eq. (4.16), $k_t(\boldsymbol{H},\underset{\sim}{\boldsymbol{H}})$ is a second-order positive-valued random variable defined on $(\Theta,\mathcal{T},\mathcal{P})$,

$$E\{k_t(\boldsymbol{H},\underset{\sim}{\boldsymbol{H}})^2\} = \int_{\mathbb{R}^\nu}\int_{\mathbb{R}^\nu} k_t(\boldsymbol{y},\boldsymbol{x})^2\; p_{\boldsymbol{H}}(\boldsymbol{y})\, p_{\boldsymbol{H}}(\boldsymbol{x})\, d\boldsymbol{y}\, d\boldsymbol{x} = c_t^2 < +\infty\,.$$

(b) For all u in $\mathbb{H}$ and for $t > 0$, from Eq. (4.18), it can be seen that $(K_t u)(\boldsymbol{y}) = \int_{\mathbb{R}^\nu} k_t(\boldsymbol{y},\boldsymbol{x})\, p_{\boldsymbol{H}}(\boldsymbol{x})^{1/2}\, u(\boldsymbol{x})\, p_{\boldsymbol{H}}(\boldsymbol{x})^{1/2}\, d\boldsymbol{x}$. Using the Cauchy-Schwarz inequality and Eq. (4.16) yield

$$\|K_t u\|^2_{\mathbb{H}} \le \|u\|^2_{\mathbb{H}} \int_{\mathbb{R}^\nu}\int_{\mathbb{R}^\nu} k_t(\boldsymbol{y},\boldsymbol{x})^2\; p_{\boldsymbol{H}}(\boldsymbol{y})\, p_{\boldsymbol{H}}(\boldsymbol{x})\, d\boldsymbol{y}\, d\boldsymbol{x} = c_t^2\, \|u\|^2_{\mathbb{H}}\,,$$

(which, in passing, shows the continuity of the operator K_t in $\mathbb{H}$ as stated at the beginning of the proof of Proposition 2). Consequently, we have $\langle K_t u\,,\mathrm{v}\rangle^2_{\mathbb{H}} \le \|K_t u\|^2_{\mathbb{H}}\, \|\mathrm{v}\|^2_{\mathbb{H}}$, which shows that

$$\langle K_t u\,,\mathrm{v}\rangle^2_{\mathbb{H}} \le c_t^2\, \|u\|^2_{\mathbb{H}}\, \|\mathrm{v}\|^2_{\mathbb{H}} < +\infty\,, \tag{4.29}$$

and therefore, Eq. (4.28) holds.

Definition 4 (Estimator constructed with a statistical sampling and associated estimation). Let $\{\boldsymbol{H}^i, i = 1,\ldots,n_d\}$ and $\{\underset{\sim}{\boldsymbol{H}}^j, j = 1,\ldots,n_d\}$ be n_d independent copies of $\boldsymbol{H}$ and $\underset{\sim}{\boldsymbol{H}}$, respectively. For every fixed $t > 0$, and for all u and v in $\mathbb{H} \cap C^0(\mathbb{R}^\nu)$, let $Z^{ij}_{t,u,\mathrm{v}}$ be the real valued random variable on $(\Theta,\mathcal{T},\mathcal{P})$, such that, for all i and j in $\{1,\ldots,n_d\}$,

$$Z^{ij}_{t,u,\mathrm{v}} = k_t(\boldsymbol{H}^i,\underset{\sim}{\boldsymbol{H}}^j)\, u(\underset{\sim}{\boldsymbol{H}}^j)\, \mathrm{v}(\boldsymbol{H}^i)\,. \tag{4.30}$$

Let $\widehat{Z}^{(n_d)}_{t,u,\mathrm{v}}$ be the real-valued random variable on $(\Theta, \mathcal{T}, \mathcal{P})$, defined by

$$\widehat{Z}^{(n_d)}_{t,u,\mathrm{v}} = \frac{1}{n_d^2} \sum_{i=1}^{n_d} \sum_{j=1}^{n_d} Z^{ij}_{t,u,\mathrm{v}} \,. \tag{4.31}$$

Then, $\widehat{Z}^{(n_d)}_{t,u,\mathrm{v}}$ is an estimator of

$$\underline{z}_{t,u,\mathrm{v}} = E\{Z_{t,u,\mathrm{v}}\} = \langle K_t u \,, \mathrm{v} \rangle_{\mathbb{H}} \,, \tag{4.32}$$

where $Z_{t,u,\mathrm{v}}$ is given by Eq. (4.27). Let $\{\boldsymbol{\eta}^j, j = 1, \ldots, n_d\}$ be the n_d independent realizations of $\boldsymbol{H}$ introduced in Section 2. Since $\boldsymbol{H}^i$ and $\underline{\boldsymbol{H}}^j$ are independent copies of $\boldsymbol{H}$ (because $\underline{\boldsymbol{H}}$ is an independent copy of $\boldsymbol{H}$), an estimation of $\underline{z}_{t,u,\mathrm{v}}$ is a realization $\underline{z}^{(n_d)}_{t,u,\mathrm{v}}$ of the estimator $\widehat{Z}^{(n_d)}_{t,u,\mathrm{v}}$, which is written as

$$\underline{z}^{(n_d)}_{t,u,\mathrm{v}} = \frac{1}{n_d^2} \sum_{i=1}^{n_d} \sum_{j=1}^{n_d} k_t(\boldsymbol{\eta}^i, \boldsymbol{\eta}^j)\, u(\boldsymbol{\eta}^j)\, \mathrm{v}(\boldsymbol{\eta}^i) \,. \tag{4.33}$$

Lemma 3 (Convergence of the sequence of estimators $\{\widehat{Z}^{(n_d)}_{t,u,\mathrm{v}}\}_{n_d}$). *Under the hypotheses and notations of Definition 4, the sequence of real-valued random variables $\{\widehat{Z}^{(n_d)}_{t,u,\mathrm{v}}\}_{n_d}$ on $(\Theta, \mathcal{T}, \mathcal{P})$ is convergent in probability to $\underline{z}_{t,u,\mathrm{v}}$,*

$$\forall \epsilon > 0 \quad , \quad \lim_{n_d \to +\infty} \mathcal{P}\{ \,|\, \widehat{Z}^{(n_d)}_{t,u,\mathrm{v}} - \underline{z}_{t,u,\mathrm{v}} \,|\, \geq \epsilon \} = 0 \,. \tag{4.34}$$

We have also the almost sure convergence, thanks to the strong law of large numbers,

$$\mathcal{P}\{ \lim_{n_d \to +\infty} \widehat{Z}^{(n_d)}_{t,u,\mathrm{v}} = \underline{z}_{t,u,\mathrm{v}} \} = 1 \,. \tag{4.35}$$

PROOF. (Lemma 3). The proof uses the usual results from mathematical statistics (see for instance [84]).

Remark 2. (a) As is well known, the speed of convergence is proportional to $1/\sqrt{n_d^2} = 1/n_d$ and is independent of dimension ν. The quantification of the approximation error could traditionally be estimated using the central limit theorem, which involves the variance of the estimator (see, for instance, [84, 66, 85]). In the numerical illustration provided in Section 6, we will show the numerical calculation of the first eigenvalues of the Fokker-Planck operator for which a reference is known.
(b) Lemma 3 with Definition 4 and Proposition 3 allow a finite approximation of $\langle K_t u \,, \mathrm{v} \rangle_{\mathbb{H}}$ to be constructed and consequently, to deduce the corresponding finite approximation of the eigenvalue problem defined by Eq. (4.21).

Proposition 4 (Finite approximation of the eigenvalue problem). *For every fixed $t > 0$, for all u and v in $\mathbb{H} \cap C^0(\mathbb{R}^\nu)$, and for n_d sufficiently large, we have (in the sense of the convergence described in Lemma 3),*

$$\langle K_t u \,, \mathrm{v} \rangle_{\mathbb{H}} \simeq \langle [\widehat{K}(t)]\, \hat{\boldsymbol{u}} \,, \hat{\mathbf{v}} \rangle_{\mathbb{R}^{n_d}} \,, \tag{4.36}$$

in which $\langle \cdot, \cdot \rangle_{\mathbb{R}^{n_d}}$ is the usual Euclidean inner product in $\mathbb{R}^{n_d}$ and where $[\widehat{K}(t)] \in \mathbb{M}^{+0}_{n_d}$ is such that, for all i and j in $\{1, \ldots, n_d\}$,

$$[\widehat{K}(t)]_{ij} = \frac{1}{n_d}\, k_t(\boldsymbol{\eta}^i, \boldsymbol{\eta}^j) \,, \tag{4.37}$$

and where the vectors $\hat{\boldsymbol{u}}$ and $\hat{\mathbf{v}}$ in $\mathbb{R}^\nu$ are such that

$$\hat{\boldsymbol{u}} = \left(\frac{u(\boldsymbol{\eta}^1)}{\sqrt{n_d}} \,, \ldots \,, \frac{u(\boldsymbol{\eta}^{n_d})}{\sqrt{n_d}} \right) \quad , \quad \hat{\mathbf{v}} = \left(\frac{\mathrm{v}(\boldsymbol{\eta}^1)}{\sqrt{n_d}} \,, \ldots \,, \frac{\mathrm{v}(\boldsymbol{\eta}^{n_d})}{\sqrt{n_d}} \right) . \tag{4.38}$$

The corresponding finite approximation of the eigenvalue problem defined by Eq. (4.21) *is written as*

$$[\widehat{K}(t)]\, \hat{\boldsymbol{\psi}}_\alpha(t) = \hat{b}_\alpha(t)\, \hat{\boldsymbol{\psi}}_\alpha(t) \quad , \quad \hat{\boldsymbol{\psi}}_\alpha(t) \in \mathbb{R}^{n_d} \,, \tag{4.39}$$

in which

$$\hat{b}_0(t) \,\geq\, \hat{b}_1(t) \,\geq\, \ldots \,\geq\, \hat{b}_{n_d - 1}(t) \,\geq 0 \,, \tag{4.40}$$

and where the normalization of the eigenvectors is chosen so that for α and β in $\{0, 1, \ldots, n_d - 1\}$,

$$\langle \hat{\boldsymbol{\psi}}_\alpha(t), \hat{\boldsymbol{\psi}}_\beta(t) \rangle_{\mathbb{R}^{n_d}} = \delta_{\alpha\beta} \,. \tag{4.41}$$

PROOF. (Proposition 4). From Eqs. (4.32), (4.33), and (4.37), it can be deduced that, for n_d sufficiently large, $\langle K_t u\,,\mathrm{v}\rangle_{\mathbb{H}} \simeq n_d^{-2}\sum_{i=1}^{n_d}\sum_{j=1}^{n_d} k_t(\boldsymbol{\eta}^i,\boldsymbol{\eta}^j)\,u(\boldsymbol{\eta}^j)\,\mathrm{v}(\boldsymbol{\eta}^i)$. Since $k_t(\boldsymbol{\eta}^i,\boldsymbol{\eta}^j) = k_t(\boldsymbol{\eta}^j,\boldsymbol{\eta}^i)$ by symmetry of k_t (see Eq. (4.6)), the right-hand side member can be written as $\langle[\widehat{K}(t)]\,\hat{\boldsymbol{u}}\,,\hat{\mathbf{v}}\rangle_{\mathbb{R}^{n_d}}$ where $[\widehat{K}(t)]$, $\hat{\boldsymbol{u}}$, and $\hat{\mathbf{v}}$ are defined by Eqs. (4.37) and (4.38). From Proposition 2, K_t is a positive operator in $\mathbb{H}$, and as Hilbert-Schmidt operator, $b_\alpha(t) \to 0_+$ as $\alpha \to +\infty$. For the finite approximation, matrix $[\widehat{K}(t)]$ is then symmetric and positive. Using Eq. (4.36), the finite approximation of Eq. (4.21) is then written as Eq. (4.39). Since $[\widehat{K}(t)]$ is real, symmetric, and positive, we have Eqs. (4.40) and we choose the normalization of $\{\hat{\boldsymbol{\psi}}_\alpha(t)\}_\alpha$ so that (4.41) holds.

Remark 3 (About the finite approximation of the eigenvalue problem). (a) It should be noted that we have chosen the construction of $\{\hat{\boldsymbol{\psi}}_\alpha(t)\}_\alpha$ as an orthonormal basis in $\mathbb{R}^{n_d}$. Consequently, $\hat{\boldsymbol{\psi}}_\alpha(t)$ is not related to ψ_α by a simple sampling, similar to the one described in Eq. (4.38).
(b) In addition, $\hat{\boldsymbol{\psi}}_\alpha(t)$ depends, *a priori*, on t, while ψ_α is independent of t. Only for $n_d \to +\infty$, $\hat{\boldsymbol{\psi}}_\alpha(t)$ goes to a vector independent of t. Similarly, although $b_0(t) = \exp(-\lambda_0 t) = 1$ because $\lambda_0 = 0$ (see Eq. (4.15)), we do not have, *a priori*, $\hat{b}_0(t) = 1$, but this equality holds for $n_d \to +\infty$. Nevertheless, for n_d finite, we will exploit this existing dependence on t for the construction of the reduced transient basis at time t, which will be connected to the reduced DMAPS basis for $t \to 0$.
(c) Note that as $n_d \to +\infty$, $\hat{b}_\alpha(t)$ tends to $b_\alpha(t)$. Since $b_\alpha(t) = \exp(-\lambda_\alpha t)$, that is to say $\lambda_\alpha = -\frac{1}{t}\,\log\, b_\alpha(t)$, we choose to define $\hat{\lambda}_\alpha(t)$ by a similar formula, such that for $t > 0$ and α for which $\hat{b}_\alpha(t) > 0$,

$$\hat{\lambda}_\alpha(t) = -\frac{1}{t}\,\log\,\hat{b}_\alpha(t)\,. \tag{4.42}$$

(d) For every fixed $t > 0$, to solve the eigenvalue problem defined by Eq. (4.39), we have to construct matrix $[\widehat{K}(t)]$ with an adapted methodology. This will be the object of Section 5. In particular, the matrix defined by Eq. (4.37) will be evaluated thanks to Proposition 6.

5. Construction of the matrix of the finite approximation

In this section, we present the methodology to construct matrix $[\widehat{K}(t)]$ as defined in Proposition 4. This construction requires the numerical evaluation of kernel k_t, because the entry $[\widehat{K}(t)]_{ij}$ of $[\widehat{K}(t)]$ is given by $k_t(\boldsymbol{\eta}^i,\boldsymbol{\eta}^j)/n_d$ (see Eq. (4.37)). According to Definition 1 (see Eq. (4.1)), for i and j in $\{1,\ldots,n_d\}$,

$$[\widehat{K}(t)]_{ij} = \frac{1}{n_d}\,\frac{\rho(\boldsymbol{\eta}^i,t\,|\,\boldsymbol{\eta}^j,0)}{p_{\boldsymbol{H}}(\boldsymbol{\eta}^i)}\,. \tag{5.1}$$

In Eq. (5.1), $p_{\boldsymbol{H}}$ is explicitly defined by Eq. (2.3), and $\rho(\boldsymbol{y},t\,|\,\boldsymbol{x},0)$ is the transient probability density function of the stochastic process $\{\boldsymbol{Y}(t), t \geq 0\}$, starting from $\boldsymbol{Y}(0) = \boldsymbol{x}$ in $\mathbb{R}^\nu$. This function is the solution of the ISDE defined by Eq. (3.1) for $t > 0$ and with the initial condition $\boldsymbol{Y}(0) = \boldsymbol{x}$ (see Eq. (3.2)). We will begin by using a a classical mathematical result concerning the solution of the ISDE, which must be validated for the specific case where the invariant measure is defined in Section 2. Additionally, we will obtain a proof of the properties introduced in Section 3.2. Then, we will present the numerical method for constructing $[\widehat{K}(t)]$ by numerically solving the ISDE and using nonparametric statistics to estimate $\rho(\boldsymbol{\eta}^i,t\,|\,\boldsymbol{\eta}^j,0)$. Subsequently, we will derive an explicit algebraic formula for $[\widehat{K}(t)]_{ij}$.

5.1. Existence and uniqueness of the solution of ISDE and properties of the transition probability

Let $\boldsymbol{Y} = \{\boldsymbol{Y}(t), t \geq 0\}$ be the $\mathbb{R}^\nu$-valued stochastic process satisfying (see Eqs. (3.1) and (3.2)),

$$d\boldsymbol{Y}(t) = \boldsymbol{b}(\boldsymbol{Y}(t))\,dt + d\boldsymbol{W}(t) \quad , \quad t > 0\,, \tag{5.2}$$

$$\boldsymbol{Y}(0) = \boldsymbol{x} \in \mathbb{R}^\nu\,,\ a.s.\,, \tag{5.3}$$

with $\boldsymbol{x} \in \mathbb{R}^\nu$, where the drift $\boldsymbol{y} \mapsto \boldsymbol{b}(\boldsymbol{y}) : \mathbb{R}^\nu \mapsto \mathbb{R}^\nu$ is defined by Eq. (3.3) with Eq. (2.7), and where $\{\boldsymbol{W}(t), t \geq 0\}$ is the normalized Wiener process defined on $(\Theta,\mathcal{T},\mathcal{P})$.

Proposition 5 (Existence and uniqueness). *Eqs.* (5.2) *and* (5.3) *define a unique homogeneous diffusion* $\mathbb{R}^\nu$*-valued stochastic process* $\boldsymbol{Y}$ *defined on* $(\Theta, \mathcal{T}, \mathcal{P})$*, whose transition probability measure is homogeneous (that is to say, it depends only on* $t - 0 = t$*),*

$$P_{\boldsymbol{Y}(t)|\boldsymbol{Y}(0)}(\mathcal{B}, t\,|\,\boldsymbol{x}, 0) = \mathcal{P}\{\boldsymbol{Y}(t) \in \mathcal{B}\,|\,\boldsymbol{Y}(0) = \boldsymbol{x}\} \quad , \quad t > 0\,, \tag{5.4}$$

where $\mathcal{B}$ *is any Borel set in* $\mathbb{R}^\nu$*. For all* $t > 0$ *and for all* $\boldsymbol{x}$ *in* $\mathbb{R}^\nu$*,* $P_{\boldsymbol{Y}(t)|\boldsymbol{Y}(0)}(d\boldsymbol{y}, t\,|\,\boldsymbol{x}, 0)$ *admits a density function* $\boldsymbol{y} \mapsto \rho(\boldsymbol{y}, t\,|\,\boldsymbol{x}, 0) : \mathbb{R}^\nu \mapsto]0\,, +\infty[$ *with respect to the Lebesgue measure* $d\boldsymbol{y}$ *on* $\mathbb{R}^\nu$*, such that*

$$P_{\boldsymbol{Y}(t)\,|\,\boldsymbol{Y}(0)}(d\boldsymbol{y}, t\,|\,\boldsymbol{x}, 0) = \rho(\boldsymbol{y}, t\,|\,\boldsymbol{x}, 0)\, d\boldsymbol{y} \quad , \quad t > 0\,, \tag{5.5}$$

$$\lim_{t \to 0_+} \rho(\boldsymbol{y}, t\,|\,\boldsymbol{x}, 0)\, d\boldsymbol{y} = \delta_0(\boldsymbol{y} - \boldsymbol{x})\,. \tag{5.6}$$

Stochastic process $\boldsymbol{Y}$ *has almost-surely continuous trajectories and for all* $t > 0$*,* $\boldsymbol{Y}(t)$ *is a second-order random variable,*

$$\forall t \geq 0 \quad , \quad E\{\|\boldsymbol{Y}(t)\|^2\} < +\infty\,. \tag{5.7}$$

For $t \to +\infty$*,* $\boldsymbol{Y}$ *is asymptotic to a stationary stochastic process whose first-order marginal probability measure is the invariant measure* $p_{\boldsymbol{H}}(\boldsymbol{y})\, d\boldsymbol{y}$*,*

$$\lim_{t \to +\infty} \rho(\boldsymbol{y}, t\,|\,\boldsymbol{x}, 0)\, d\boldsymbol{y} = p_{\boldsymbol{H}}(\boldsymbol{y})\, d\boldsymbol{y}\,. \tag{5.8}$$

The function $(\boldsymbol{y}, \boldsymbol{x}) \mapsto \rho(\boldsymbol{y}, t\,|\,\boldsymbol{x}, 0)$ *is continuous from* $\mathbb{R}^\nu \times \mathbb{R}^\nu$ *into* $\mathbb{R}^{+*}$*,*

$$\rho(\cdot, t\,|\,\cdot, 0) \in C^0(\mathbb{R}^\nu \times \mathbb{R}^\nu, \mathbb{R}^{+*})\,. \tag{5.9}$$

PROOF. (Proposition 5). The proof is presented in five steps.
(a) It is easy to prove that drift function b is continuous on $\mathbb{R}^\nu$.

(b) Since the diffusion matrix is the identity matrix and b belongs to $C^0(\mathbb{R}^\nu, \mathbb{R}^\nu)$, we can establish the existence of a unique diffusion stochastic process (see [68] Ch. VIII, Sec. 2; [86] Ch. IV, Secs. 2, 3, and 5; or [69] Ch. V), provided that for all $\boldsymbol{y}$ and $\boldsymbol{y}'$ in $\mathbb{R}^\nu$, we have

$$\|\boldsymbol{b}(\boldsymbol{y}) - \boldsymbol{b}(\boldsymbol{y}')\| \leq\ c\,\|\boldsymbol{y} - \boldsymbol{y}'\| \quad , \quad \|\boldsymbol{b}(\boldsymbol{y})\| \leq\ C\,(1 + \|\boldsymbol{y}\|)\,. \tag{5.10}$$

Using Eqs. (2.3), (2.7), and (3.3), $\boldsymbol{b}(\boldsymbol{y})$ can be rewritten, for all $\boldsymbol{y}$ in $\mathbb{R}^\nu$, as

$$\boldsymbol{b}(\boldsymbol{y}) = \frac{1}{2}\,\xi(\boldsymbol{y})^{-1} \nabla \xi(\boldsymbol{y}) \quad , \quad \xi(\boldsymbol{y}) = \frac{1}{n_d} \sum_{j=1}^{n_d} \exp\left\{ -\frac{1}{2\hat{s}^2} \|\frac{\hat{s}}{s}\boldsymbol{\eta}^j - \boldsymbol{y}\|^2 \right\} > 0\,. \tag{5.11}$$

Calculating $\nabla \xi(\boldsymbol{y})$, it can be seen that

$$2\,\|\boldsymbol{b}(\boldsymbol{y})\| \leq\ \frac{1}{\xi(\boldsymbol{y})\, n_d\, \hat{s}^2} \sum_{j=1}^{n_d} \frac{\hat{s}}{s}\, \|\boldsymbol{\eta}^j\| \exp\left\{ -\frac{1}{2\hat{s}^2} \|\frac{\hat{s}}{s}\boldsymbol{\eta}^j - \boldsymbol{y}\|^2 \right\} + \frac{1}{\hat{s}^2}\, \|\boldsymbol{y}\|\,.$$

Since n_d is finite and since $E\{\|\boldsymbol{H}\|^2\} = \frac{1}{n_d} \sum_{j=1}^{n_d} \|\boldsymbol{\eta}^j\|^2 = \nu$ is also finite, we have $\sup_{j=1,\ldots,n_d} \|\boldsymbol{\eta}^j\| = c_\eta < +\infty$ and therefore,

$$\|\boldsymbol{b}(\boldsymbol{y})\| \leq\ \frac{1}{2}\left(\frac{1}{\hat{s}^2}\,\frac{\hat{s}}{s}\, c_\eta + \frac{1}{\hat{s}^2}\, \|\boldsymbol{y}\| \right) \leq\ C\,(1 + \|\boldsymbol{y}\|) \quad , \quad 0 < C < +\infty\,.$$

The second inequality in Eq. (5.10) is then proven. It can now be verified that

$$\|\boldsymbol{b}(\boldsymbol{y}) - \boldsymbol{b}(\boldsymbol{y}')\| \leq\ \frac{1}{2\hat{s}^2}(\,\|\boldsymbol{f}(\boldsymbol{y}) - \boldsymbol{f}(\boldsymbol{y}')\| + \|\boldsymbol{y} - \boldsymbol{y}'\|\,)\,,$$

in which $\boldsymbol{f}(\boldsymbol{y}) = \boldsymbol{a}(\boldsymbol{y})/\beta(\boldsymbol{y})$ where $\boldsymbol{a}(\boldsymbol{y})$ and $\beta(\boldsymbol{y})$ are written as

$$\boldsymbol{a}(\boldsymbol{y}) = \frac{\hat{s}}{s} \sum_{j=1}^{n_d} \boldsymbol{\eta}^j\, \exp\left\{ -\frac{1}{2\hat{s}^2} \|\frac{\hat{s}}{s}\boldsymbol{\eta}^j - \boldsymbol{y}\|^2 \right\} \quad , \quad \beta(\boldsymbol{y}) = \sum_{j=1}^{n_d} \exp\left\{ -\frac{1}{2\hat{s}^2} \|\frac{\hat{s}}{s}\boldsymbol{\eta}^j - \boldsymbol{y}\|^2 \right\}\,.$$

We must have $\|\boldsymbol{f}(\boldsymbol{y}) - \boldsymbol{f}(\boldsymbol{y}')\| \leq c\,\|\boldsymbol{y} - \boldsymbol{y}'\|$, that is true if $\|\,[\nabla \boldsymbol{f}(\boldsymbol{y})]\,\|_F \leq\ \tilde{c} < +\infty$, in which the $(\nu \times \nu)$-real matrix $[\nabla \boldsymbol{f}(\boldsymbol{y})]$ is written as $[\nabla \boldsymbol{f}(\boldsymbol{y})] = \beta(\boldsymbol{y})^{-1}\,[\nabla \boldsymbol{a}(\boldsymbol{y})] - \beta(\boldsymbol{y})^{-2}\,\boldsymbol{a}(\boldsymbol{y}) \otimes \nabla\beta(\boldsymbol{y})$, and where $\|\cdot\|_F$ is the Frobenius norm. Introducing, temporary, the notation $e_j = \exp\left\{-\frac{1}{2\hat{s}^2}\|\frac{\hat{s}}{s}\boldsymbol{\eta}^j - \boldsymbol{y}\|^2\right\}$, we have

$$[\nabla \boldsymbol{f}(\boldsymbol{y})] = \frac{1}{\hat{s}^2}\,\frac{\hat{s}^2}{s^2}\,\frac{\sum_{j=1}^{n_d}\sum_{j'=1}^{n_d} e_j e_{j'}(\boldsymbol{\eta}^j \otimes \boldsymbol{\eta}^j - \boldsymbol{\eta}^j \otimes \boldsymbol{\eta}^{j'})}{\sum_{j=1}^{n_d}\sum_{j'=1}^{n_d} e_j e_{j'}}\,.$$

It can then be deduced that,

$$\|\,[\nabla \boldsymbol{f}(\boldsymbol{y})]\,\|_F \leq\ \frac{1}{\hat{s}^2}\,\frac{\hat{s}^2}{s^2}\,\sup_{j,j'}\|\boldsymbol{\eta}^j \otimes \boldsymbol{\eta}^j - \boldsymbol{\eta}^j \otimes \boldsymbol{\eta}^{j'}\|_F\,.$$

As previously, since $\sup_j \|\boldsymbol{\eta}^j\| = c_\eta < +\infty$, we have $\|\,[\nabla \boldsymbol{f}(\boldsymbol{y})]\,\|_F \leq\ \tilde{c}_F$ with $\tilde{c}_F$ a finite positive constant independent of $\boldsymbol{y}$, that yields the proof for the Lipchitz continuity of $\boldsymbol{b}$.

(c) We will admit that, for all $t > 0$ and for all $\boldsymbol{x}$ in $\mathbb{R}^\nu$, the transition probability measure has a density $\boldsymbol{y} \mapsto \rho(\boldsymbol{y}, t\,|\,\boldsymbol{x}, 0)$ on $\mathbb{R}^\nu$ with respect to $d\boldsymbol{y}$, and that $(\boldsymbol{y}, \boldsymbol{x}) \mapsto \rho(\boldsymbol{y}, t\,|\,\boldsymbol{x}, 0)$ is continuous on $\mathbb{R}^\nu \times \mathbb{R}^\nu$ (for all $t > 0$) (see for instance [87] Ch III, Secs. 6 to 9, or [88] Ch. 10).

(d) As a diffusion stochastic process, the trajectories are almost surely continuous functions. Since $\boldsymbol{b}$ is continuous and $\|\boldsymbol{b}(\boldsymbol{y})\| \leq\ c\,(1 + \|\boldsymbol{y}\|)$ implies $\|\boldsymbol{b}(\boldsymbol{y})\|^2 \leq\ 2c^2\,(1 + \|\boldsymbol{y}\|^2)$, and since $E\{\|\boldsymbol{Y}(0)\|^2\} = \|\boldsymbol{x}\|^2 < +\infty$ for all fixed $\boldsymbol{x}$ in $\mathbb{R}^\nu$, we have Eq. (5.7) (see [86] Ch. IV, Sec. 2).

(e) The existence of the asymptotic stationary solution with Eq. (5.8) can be found in [87] Ch. III, or [70] Ch. VI, Secs. 5 and 6.

5.2. Rewriting the Itô equation in a matrix form

Let $\{\boldsymbol{Y}^j(t), t \geq 0\}$ be the solution of Eqs. (5.2) and (5.3) with the initial condition $\boldsymbol{Y}^j(0) = \boldsymbol{\eta}^j \in \mathbb{R}^\nu$, in which $\boldsymbol{\eta}^j$ is defined in Section 2. We then have

$$\rho(\boldsymbol{y}, t\,|\,\boldsymbol{\eta}^j, 0)\,d\boldsymbol{y} = P_{\boldsymbol{Y}^j(t)\,|\,\boldsymbol{Y}^j(0)}(d\boldsymbol{y}, t\,|\,\boldsymbol{\eta}^j, 0)\,. \tag{5.12}$$

Let $\{[\boldsymbol{Y}(t)]\,, t \geq 0\}$ be the $\mathbb{M}_{\nu,n_d}$-valued stochastic process and $[\eta_d]$ be the matrix in $\mathbb{M}_{\nu,n_d}$ such that

$$[\boldsymbol{Y}(t)] = [\boldsymbol{Y}^1(t) \ldots \boldsymbol{Y}^{n_d}(t)] \quad , \quad [\eta_d] = [\boldsymbol{\eta}^1 \ldots \boldsymbol{\eta}^{n_d}]\,. \tag{5.13}$$

Therefore, $\{[\boldsymbol{Y}(t)]\,, t \geq 0\}$ is solution of the matrix-valued ISDE,

$$d[\boldsymbol{Y}(t)] = \frac{1}{2}\,[L([\boldsymbol{Y}(t)])]\,dt + d[\boldsymbol{W}(t)] \quad , \quad t > 0\,, \tag{5.14}$$

$$[\boldsymbol{Y}(0)] = [\eta_d]\,,\ a.s.\,, \tag{5.15}$$

where $[y] \mapsto [L([y])]$ is the function from $\mathbb{M}_{\nu,n_d}$ into $\mathbb{M}_{\nu,n_d}$, defined, for $k \in \{1, \ldots, \nu\}$ and $j \in \{1, \ldots, n_d\}$, by

$$[L([y])]_{kj} = \frac{1}{\xi(\boldsymbol{y}^j)}\,\frac{\partial \xi(\boldsymbol{y}^j)}{\partial y_k^j}\,. \tag{5.16}$$

In Eq. (5.16), $\boldsymbol{y}^j$ is the j-th column of $[y] = [\boldsymbol{y}^1 \ldots \boldsymbol{y}^{n_d}]$, where $\boldsymbol{y}^j = (y_1^j, \ldots, y_\nu^j)$. Additionally, $\xi(\boldsymbol{y})$ is defined by Eq. (5.11), and $\{[\boldsymbol{W}(t)]\,, t \geq 0\}$ is the normalized $\mathbb{M}_{\nu,n_d}$-valued Wiener stochastic process.

5.3. Time-discrete approximation of the ISDE and convergence analysis

To estimate $[\widehat{K}(n\Delta t)]_{ij}$, defined by Eq. (5.1), using nonparametric statistics, we need to generate realizations of Eqs. (5.14) and (5.15). Consequently, a first stage involves introducing a time-discrete approximation of Eq. (5.14) and analyzing the convergence.

(i) Time sampling Δt and δt. We first define a time sampling $n\Delta t, n \geq 1$ to be used for estimating $[\widehat{K}(n\Delta t)]$. The time step Δt will be defined in Section 7. However, Δt may not be sufficiently small to achieve a satisfactory rate of convergence. We then introduce $\delta t \leq \Delta t$ such that $\Delta t = n_s \times \delta t$, with $n_s \geq 1$. To discretize Eq. (5.14), we use the time sampling $\{\mu\, \delta t, \mu \geq 1\}$.

(ii) ISDE discretization. Assuming that δt is sufficiently small relative to 1, employing the Euler scheme (as seen, for example, in [89]) to discretize the solution $\{[\boldsymbol{Y}(t)]\,, t \geq 0\}$ of Eqs. (5.14) and (5.15) yields

$$[\boldsymbol{Y}_\mu] = [\boldsymbol{Y}_{\mu-1}] + \frac{\delta t}{2}\,[L([\boldsymbol{Y}_{\mu-1}])] + \sqrt{\delta t}\,[\boldsymbol{\Gamma}_\mu] \quad , \quad \mu \geq 1\,, \tag{5.17}$$

$$[\boldsymbol{Y}_0] = [\eta_d]\,,\ a.s.\,, \tag{5.18}$$

where $[\boldsymbol{Y}_\mu]$ is the approximation of $[\boldsymbol{Y}(t)]$ at $t = \mu\,\delta t$, and where $\{[\boldsymbol{\Gamma}_\mu]_{kj}\,; k = 1, \ldots, \nu\,; j = 1, \ldots, n_d\,; \mu \geq 1\}$ is an infinite family of independent normalized Gaussian real-valued random variables.

(iii) Convergence of the time-discrete approximation. For each j in $1, \ldots, n_d$, Proposition 5 can be applied to the stochastic process $\{\boldsymbol{Y}^j(t), t \geq 0\}$, where $\boldsymbol{Y}^j(t)$ represents the j-th column of $[\boldsymbol{Y}(t)]$. Let $\{[\boldsymbol{Y}^\delta(t)], t \geq 0\}$ be the time-discrete approximation of $\{[\boldsymbol{Y}(t)], t \geq 0\}$. Take any fixed positive real number T, let $\mu_T = \mathrm{int}(T/\delta t)$ denotes the nearest integer to $T/\delta t$. It can be observed that as δt approaches 0, μ_T tends to $+\infty$. The following classical result holds (see, for instance, [89], Page 323).

Lemma 4 (Strong convergence). *Under Proposition 5, the time-discrete approximation $\{[\boldsymbol{Y}^\delta(t)], t \geq 0\}$ of $\{[\boldsymbol{Y}(t)], t \geq 0\}$ converges strongly at time T if*

$$\lim_{\delta t \to 0} E\{\,\|\,[\boldsymbol{Y}_{\mu_T}] - [\boldsymbol{Y}(T)]\,\|_F^2\} = 0\,, \tag{5.19}$$

where $[\boldsymbol{Y}_{\mu_T}]$ is determined by Eqs. (5.17) *and* (5.18) *up to $\mu = \mu_T$.*

(iv) Criterion to determine if δt is small enough. We can employ a criterion based on weak convergence, which is related to the covariance matrix $[C_{\boldsymbol{Y}^j(t)}] \in \mathbb{M}_\nu^+$ of the $\mathbb{R}^\nu$-valued random variable $\boldsymbol{Y}^j(t)$, for t fixed in the interval $]0, T]$. Let $\sigma^2_{\boldsymbol{Y}}(t)$ be defined by

$$\sigma^2_{\boldsymbol{Y}}(t) = E\{\,\|\,[\boldsymbol{Y}(t)] - E\{[\boldsymbol{Y}(t)]\}\,\|_F^2\} = \sum_{j=1}^{n_d} \sum_{k=1}^{\nu} E\{(Y_k^j(t) - E\{Y_k^j(t)\})^2\}\,. \tag{5.20}$$

It can easily be seen that

$$\sigma^2_{\boldsymbol{Y}}(t) = \sum_{j=1}^{n_d} \mathrm{tr}[C_{\boldsymbol{Y}^j(t)}]\,. \tag{5.21}$$

Let $\sigma^2_{\boldsymbol{Y}_{\mu_T}}$ be the corresponding quantity for random matrix $[\boldsymbol{Y}_{\mu_T}]$,

$$\sigma^2_{\boldsymbol{Y}_{\mu_T}} = E\{\,\|\,[\boldsymbol{Y}_{\mu_T}] - E\{[\boldsymbol{Y}_{\mu_T}]\}\,\|_F^2\} = \sum_{j=1}^{n_d} \mathrm{tr}[C_{\boldsymbol{Y}^j_{\mu_T}}]\,. \tag{5.22}$$

The criterion can then be based on the following properties,

$$\lim_{\delta t \to 0} |\,\sigma_{\boldsymbol{Y}_{\mu_T}} - \sigma_{\boldsymbol{Y}}(T)\,| = 0\,. \tag{5.23}$$

We will detailed this criterion in paragraph (vi).

(v) Generation of independent realizations of the time-discrete approximation. Let $n_s \geq 1$ and $N > 1$ be two fixed integers. Let Δt be fixed and let $\delta t = \Delta t / n_s$. Note that μ_T introduced in Section 5.3-(iv) is such that $\mu_T = n_s \times N$. We define T, $\mathcal{N}$, and $\mathcal{N}_\mu$ by,

$$T = N \times \Delta t = n_s \times N \times \delta t \quad , \quad \mathcal{N} = \{1, \ldots, N\} \quad , \quad \mathcal{N}_\mu = \{1, \ldots, n_s \times N\} . \tag{5.24}$$

For all μ in $\mathcal{N}_\mu$, let $\{[\gamma_\mu^\ell] \in \mathbb{M}_{\nu,n_d}, \ell = 1, \ldots, n_{\text{MC}}\}$ be n_{MC} independent realizations of the random matrix $[\mathbf{\Gamma}_\mu]$. As discussed in Section 5.3-(ii), it can be seen that the family $\{[\gamma_\mu^\ell], \ell = 1, \ldots, n_{\text{MC}}; \mu \in \mathcal{N}_\mu\}$ consists of $n_{\text{MC}} \times n_s \times N$ independent realizations. For each ℓ in $\{1, \ldots, n_{\text{MC}}\}$, the realization $\{[\tilde{y}_\mu^\ell], \mu \in \mathcal{N}_\mu\}$ of the time-discrete approximation $\{[\boldsymbol{Y}_\mu], \mu \in \mathcal{N}_\mu\}$ is computed using the following recurrence (refer to Eqs. (5.17) and (5.18)),

$$[\tilde{y}_\mu^\ell] = [\tilde{y}_{\mu-1}^\ell] + \frac{\delta t}{2} [L([\tilde{y}_{\mu-1}^\ell])] + \sqrt{\delta t} [\gamma_\mu^\ell] \quad , \quad \mu \in \mathcal{N}_\mu , \tag{5.25}$$

$$[\tilde{y}_0^\ell] = [\eta_d] , \tag{5.26}$$

In the following, for performing the statistical estimations, we will use the subsequence $\{[y_n^\ell], n \in \mathcal{N}\}$ of $\{[\tilde{y}_\mu^\ell], \mu \in \mathcal{N}_\mu\}$, such that

$$\forall n \in \mathcal{N} \quad , \quad [y_n^\ell] = [\tilde{y}_\mu^\ell] \quad , \quad \mu = n_s \times n . \tag{5.27}$$

(vi) Practical criteria for controlling the convergence parameters. Let Δt be fixed, as well as n_s, meaning δt is fixed, to ensure satisfaction of the convergence criteria introduced in Section 5.3-(iv). An additional practical criterion can be applied to verify the adequacy of all the convergence parameters, as provided in Lemma 5.

Lemma 5 (Practical criteria for controlling the convergence). *For $k \in \{1, \ldots, \nu\}$, $j \in \{1, \ldots, n_d\}$, and $n \in \mathcal{N}$, let $[\underline{y}_n]$ and $[\sigma_n]$ be the matrices in $\mathbb{M}_{\nu,n_d}$ defined by,*

$$[\underline{y}_n]_{kj} = \frac{1}{n_{\text{MC}}} \sum_{\ell=1}^{n_{\text{MC}}} [y_n^\ell]_{kj} \quad , \quad ([\sigma_n]_{kj})^2 = \frac{1}{n_{\text{MC}}} \sum_{\ell=1}^{n_{\text{MC}}} ([y_n^\ell]_{kj} - [\underline{y}_n]_{kj})^2 . \tag{5.28}$$

Let $n \mapsto \underline{y}(n)$ and $n \mapsto \underline{\sigma}(n)$ be the positive-valued functions defined, for $n \in \mathcal{N}$, by

$$\underline{y}(n) = \frac{1}{\sqrt{\nu \times n_d}} \| [\underline{y}_n] \|_F \quad , \quad \underline{\sigma}(n) = \frac{1}{\sqrt{\nu \times n_d}} \| [\sigma_n] \|_F . \tag{5.29}$$

Then, for $\delta t \to 0$ and $n_{\text{MC}} \to +\infty$, we have

$$\lim_{N \to +\infty} \underline{y}(n) = 0 \quad , \quad \lim_{N \to +\infty} \underline{\sigma}(n) = 1 . \tag{5.30}$$

PROOF. (Lemma 5). It can be seen that $[\underline{y}_n]_{kj}$ and $([\sigma_n]_{kj})^2$ are the empirical estimates of the mean value and the variance of the time-discrete approximation $[\boldsymbol{Y}_\mu]_{kj}$ for $\mu = n_s \times n$ of $Y_k^j(n\Delta t)$. From Proposition 5, we know that, for $t \to +\infty$, $\{\boldsymbol{Y}(t), t \geq 0\}$, and consequently, $\{\boldsymbol{Y}^j(t), t \geq 0\}$, is asymptotically stationary and that Eq. (5.8) holds. Since $E\{\boldsymbol{H}\} = \mathbf{0}_\nu$ and $E\{\boldsymbol{H} \otimes \boldsymbol{H}\} = [I_\nu]$ (see Eqs. (2.5) and (2.6)), if $\delta t \to 0$ and $n_{\text{MC}} \to +\infty$, we have for $N \to +\infty$, $[\underline{y}_n]_{kj} \to E\{H_k\} = 0$ and $([\sigma_n]_{kj})^2 \to [C_d]_{kk} = 1$. Since $\| [\underline{y}_n] \|_F^2 = \sum_{k=1}^{\nu} \sum_{j=1}^{n_d} ([\underline{y}_n]_{kj})^2$ and $\| [\sigma_n] \|_F^2 = \sum_{k=1}^{\nu} \sum_{j=1}^{n_d} ([\sigma_n]_{kj})^2$, using the same normalization $1/\sqrt{\nu \times n_d}$ for $\underline{y}(n)$ and $\underline{\sigma}(n)$, we obtain Eq. (5.30).

5.4. Estimation of the matrix of the finite approximation

We now can estimate $[\widehat{K}(n\Delta t)]_{ij}$ defined by Eq. (5.1) using the realizations $\{[y_n^\ell], \ell = 1, \ldots, n_{\text{MC}}\}$ for $n \in \mathcal{N}$, defined by Eq. (5.27).

Proposition 6 (Estimation of matrix $[\widehat{K}(n\Delta t)]$). *Under Propositions 4 and 5, results and notations of Section 5.3-(v), for $n \in \mathcal{N}$, an estimate of matrix $[\widehat{K}(n\Delta t)]$, whose entries are defined by Eq.* (4.37)*, is written as*

$$[\widehat{K}(n\Delta t)] = [\widehat{B}]^{-1} [\widehat{\mathcal{K}}(n\Delta t)] , \tag{5.31}$$

in which $[\widehat{B}]$ *is the diagonal matrix in* $\mathbb{M}_{n_d}$ *for which its entries are*

$$[\widehat{B}]_{ij} = \delta_{ij} \sum_{j'=1}^{n_d} \exp\left\{-\frac{1}{2\hat{s}^2}\|\boldsymbol{\eta}^i - \frac{\hat{s}}{s}\boldsymbol{\eta}^{j'}\|^2\right\} . \tag{5.32}$$

The entries of matrix $[\widehat{\mathcal{K}}(n\Delta t)]$ *in* $\mathbb{M}_{n_d}$ *are written as,*

$$[\widehat{\mathcal{K}}(n\Delta t)]_{ij} = \frac{1}{n_{\text{MC}}} \sum_{\ell=1}^{n_{\text{MC}}} \left(\frac{\hat{s}^\nu}{s_{\text{SB}}^\nu \prod_{k=1}^\nu [\sigma_n]_{kj}}\right) \exp\left\{-\frac{1}{2}\sum_{k=1}^\nu \left(\frac{\eta_k^i - [y_n^\ell]_{kj}}{s_{\text{SB}}[\sigma_n]_{kj}}\right)^2\right\} , \tag{5.33}$$

in which s *and* $\hat{s}$ *are defined by Eq.* (2.4)*, where* s_{SB} *is written as*

$$s_{\text{SB}} = \left\{\frac{4}{n_{\text{MC}}(2+\nu)}\right\}^{1/(\nu+4)} , \tag{5.34}$$

where $[y_n^\ell] \in \mathbb{M}_{\nu,n_d}$ *is defined by Eq.* (5.27)*, and where* $[\sigma_n] \in \mathbb{M}_{\nu,n_d}$ *is defined by Eq.* (5.28)*.*

PROOF. (Proposition 6). Eq. (5.31) is deduced from Eq. (5.1) with $t = n\Delta t$ and with $p_{\boldsymbol{H}}(\boldsymbol{\eta}^i)$ given by Eq. (2.3) in which $\boldsymbol{\eta} = \boldsymbol{\eta}^i$, and with the estimate of $\rho(\boldsymbol{\eta}^i, n\Delta t \,|\, \boldsymbol{\eta}^j, 0)$ given by the Gaussian kernel-density estimation method (see [65, 90]) with the Silverman bandwidth s_{SB} defined by Eq. (5.34), which is written as

$$\rho(\boldsymbol{\eta}^i, n\Delta t \,|\, \boldsymbol{\eta}^j, 0) = \frac{1}{n_{\text{MC}}} \sum_{\ell=1}^{n_{\text{MC}}} \frac{1}{(\sqrt{2\pi}s_{\text{SB}})^\nu \prod_{k=1}^\nu [\sigma_n]_{kj}} \exp\left\{-\frac{1}{2}\sum_{k=1}^\nu \left(\frac{\eta_k^i - [y_n^\ell]_{kj}}{s_{\text{SB}}[\sigma_n]_{kj}}\right)^2\right\} .$$

Remark 4. (a) In Section 6, we will provide an illustration by numerically solving the eigenvalue problem defined by Eq. (4.39) for $t = n\,\Delta t$, using Eq. (5.31). Specifically, we consider the Gaussian case with dimension $\nu = 1$. In this scenario, $\boldsymbol{H}$ represents a normalized Gaussian real-valued random variable. For this case, a reference solution is available.
(b) Nevertheless, it would be challenging to employ such a formulation with reasonable convergence in very high dimensions, where ν equals several tens or even hundreds. This would necessitate a large value of n_d.
(c) In fact, Proposition 6 will be employed in Section 7 to derive an expression that connects to the kernel $[\mathcal{K}_{\text{DM}}]_{ij} = \exp\left\{-\frac{1}{4\varepsilon_{\text{DM}}}\|\boldsymbol{\eta}^i - \boldsymbol{\eta}^j\|^2\right\}$, which is used to calculate the DMAPS basis and proves efficient for large values of ν.

6. Numerical illustration of the proposed formulation

As explained in Remark 4-(a), this section presents a numerical illustration of the formulation introduced in Section 5 to solve the approximated eigenvalue problem defined by Eq.(4.39), which is derived from the eigenvalue problem in Eq.(4.21). To validate the formulation, we select a reference case where the eigenvalue problem defined by Eq. (4.21) can be exactly solved. This is feasible when the dimension ν of $\boldsymbol{H}$ is 1 and $\boldsymbol{H}$ is a normalized Gaussian real-valued random variable.

6.1. Reference case definition and explicit solution

The quantities related to the reference case will be indexed by letter r. The probability density function of H_r on $\mathbb{R}$ is $p_{H_r}(y) = (2\pi)^{-1/2}\exp(-y^2/2)$. The potential function Φ_r (see Eq. (2.7)) is $\Phi_r(y) = y^2/2$ and the drift b_r defined by Eq. (3.3) is $b_r(y) = -y/2$. The ISDE defined by Eqs. (5.2) and (5.3) are rewritten as

$$dY_r(t) = -\frac{1}{2}Y_r(t)\,dt + dW_r(t) \quad , \quad t > 0\,, \tag{6.1}$$

$$Y_r(0) = x \in \mathbb{R}\,. \tag{6.2}$$

Consequently, $\{Y_r(t), t \geq 0\}$ is a second-order Gaussian stochastic process, which is explicitly defined by

$$Y_r(t) = x\, e^{-t/2} + \int_0^t e^{-(t-\tau)/2}\, dW_r(\tau)\,. \tag{6.3}$$

A simple calculation shows that the mean value $m_r(t) = E\{Y_r(t)\}$ and the standard deviation $\sigma_r(t) = (E\{(Y_r(t) - m_r(t))^2\})^{1/2}$ of the random variable $Y_r(t)$ for fixed $t > 0$, are written as

$$m_r(t) = x\, e^{-t/2} \quad , \quad \sigma_r(t) = \sqrt{1 - e^{-t}}\,. \tag{6.4}$$

For all $t > 0$, as $\{Y_r(t) \,|\, Y_r(0) = x\}$ is a Gaussian random variable, the transition probability density function is

$$\rho_r(y, t \,|\, x, 0) = \frac{1}{\sqrt{2\pi}\,\sigma_r(t)} \exp\left\{-\frac{1}{2\sigma_r(t)^2}(y - m_r(t))^2\right\}\,. \tag{6.5}$$

Note that Eqs. (6.4) and (6.5) show that, we effectively have $\lim_{t\to 0_+} \rho_r(y, t \,|\, x, 0)\, dy = \delta_0(y - x)$ (see Eq. (5.6)) and $\lim_{t\to+\infty} \rho_r(y, t \,|\, x, 0) = p_{H_r}(y)$ (see Eq. (5.8)). It can be deduced that the kernel $k_{r,t}(y, x)$ on $\mathbb{R}\times\mathbb{R}$, defined by Eq. (4.1) is written, for $t > 0$, as

$$k_{r,t}(y, x) = \frac{1}{\sqrt{1 - e^{-t}}} \exp\left\{-\frac{1}{2}\frac{e^{-t}}{(1 - e^{-t})}(y^2 + x^2 - 2e^{t/2} y\, x)^2\right\}\,. \tag{6.6}$$

Let $\{h_\alpha(y)\,, \alpha \in \mathbb{N}\}$ be the Hermite polynomials and $\{h_\alpha(y)/\sqrt{\alpha!}\,, \alpha \in \mathbb{N}\}$ the Hilbert basis in $L^2(\mathbb{R}; p_{H_r})$,

$$\int_{\mathbb{R}} \frac{h_\alpha(y)}{\sqrt{\alpha!}}\, \frac{h_\beta(y)}{\sqrt{\beta!}}\, p_{H_r}(y)\, dy = \delta_{\alpha\beta} \quad , \quad h_0(y) = 1 \quad , \quad h_1(y) = y \quad , \quad h_{\alpha+1}(y) = y\, h_\alpha(y) - \frac{d}{dy} h_\alpha(y)\,.$$

Let $\{\mathbb{h}_\alpha(y)\,, \alpha \in \mathbb{N}\}$ be the polynomials defined by

$$h_\alpha(y) = \frac{1}{\sqrt{2^\alpha}}\, \mathbb{h}_\alpha(\frac{y}{\sqrt{2}}) \quad , \quad \mathbb{h}_\alpha(y) = \sqrt{2^\alpha}\, h_\alpha(\sqrt{2}\, y)\,.$$

We have the formula [91],

$$\frac{1}{\sqrt{1 - 4a^2}} \exp\left\{\frac{4a}{1 - 4a^2}(\hat{x}\,\hat{y} - a\hat{x}^2 - a\hat{y}^2)\right\} = \sum_{\alpha\in\mathbb{N}} \frac{a^\alpha}{\alpha!}\, \mathbb{h}_\alpha(\hat{x})\, \mathbb{h}_\alpha(\hat{y})\,.$$

Taking $x = \sqrt{2}\,\hat{x}$, $y = \sqrt{2}\,\hat{y}$, and $a = \frac{1}{2}e^{-t/2} < 1/2$ for $t > 0$, Eq. (6.6) can be rewritten as

$$k_{r,t}(y, x) = \sum_{\alpha\in\mathbb{N}} e^{-\lambda_{r,\alpha} t} \psi_{r,\alpha}(y)\, \psi_{r,\alpha}(x)\,, \tag{6.7}$$

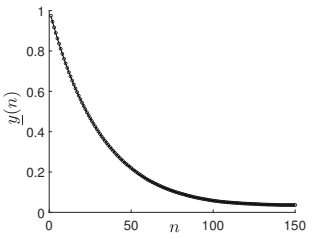
$$\lambda_{r,\alpha} = \frac{\alpha}{2} \quad , \quad \psi_{r,\alpha}(y) = \frac{1}{\sqrt{2^\alpha\, \alpha!}}\, \mathbb{h}_\alpha(\frac{y}{\sqrt{2}}) = \frac{1}{\sqrt{\alpha!}}\, h_\alpha(y)\,, \tag{6.8}$$

and thus, for α and β in $\mathbb{N}$,

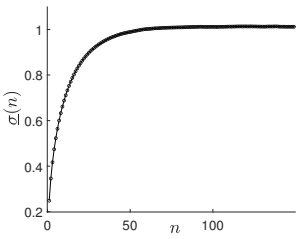
$$\int_{\mathbb{R}} \psi_{r,\alpha}(y)\, \psi_{r,\beta}(y)\, p_{H_r}(y)\, dy = \delta_{\alpha,\beta}\,. \tag{6.9}$$

Comparing Eqs. (6.7) and (6.9) with Eqs. (4.14) and (4.10) shows that, for this reference case, the eigenvalues of the Fokker-Planck operator L_{FKP} are

$$\lambda_{r,\alpha} = \frac{\alpha}{2} \quad , \quad \alpha \in \mathbb{N}\,. \tag{6.10}$$



(a) Graph of function $n \mapsto \underline{y}(n)$.

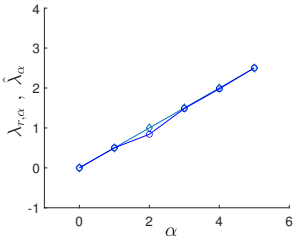


(b) Graph of function $n \mapsto \underline{\sigma}(n)$.

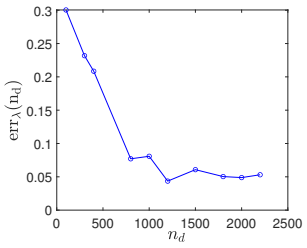
Figure 1: Criteria defined in Lemma 5 for controlling the convergence of the Gaussian-case reference with $\nu = 1$, $n_s = 1$, $n_d = 1200$, and $n_{\text{MC}} = 1200$.

6.2. Estimating the eigenvalues with the proposed numerical formulation

(i) For the convergence analysis, we consider 10 values of n_d constituting the set $\mathcal{N}_d = \{100, 300, 400, 800, 1000, 1200, 1500, 1800, 2000, 2200\}$. For each $n_d \in \mathcal{N}_d$, the matrix $[\eta_d] = [\eta^1 \dots \eta^{n_d}] \in \mathbb{M}_{1,n_d}$ is generated with an adapted generator (instruction randn$(1, n_d)$ for Matlab Gaussian generator).
(ii) For each value of n_d in $\mathcal{N}_d$, the generation of $n_{\text{MC}} = n_d$ independent realizations $\{[\tilde{y}_\mu^\ell], \mu \in \mathcal{N}_\mu\}$ with $\mathcal{N}_\mu = \{1, \dots, n_s \times N\}$ is performed using Eqs. (5.25) and (5.26) with $n_s = 1$, $\delta t = \Delta t = 0.061796$, $N = 150$. For $n_d = 1200$ and $n_{\text{MC}} = 1200$, Figs. 1a and 1b display the graphs of functions $n \mapsto \underline{y}(n)$ and $n \mapsto \underline{\sigma}(n)$ computed using Eq. (5.29) with Eq. (5.28). It can be seen that the criterion defined by Eq. (5.30) is satisfied.
(iii) For every $n_d \in \mathcal{N}_d$, matrix $[\widehat{K}(2\,\Delta t)] \in \mathbb{M}_{n_d}^{+0}$ is calculated using Eq. (5.31) with $\hat{s} = 0.2486$, $\hat{s}/s = 0.9690$, and $s_{\text{SB}} = 0.2565$. The first 6 largest eigenvalues, $\hat{b}_\alpha(2\,\Delta t)$ of $[\widehat{K}(2\,\Delta t)]$ and $\hat{\lambda}_\alpha(2\,\Delta t)$, simply denoted by $\hat{\lambda}_\alpha$, are obtained from Eq. (4.42). Fig. 2b shows the graph of function $n_d \mapsto \text{err}_\lambda(n_d)$ that quantifies the relative error between $\alpha \mapsto \lambda_{r,\alpha}$ and $\alpha \mapsto \hat{\lambda}_\alpha$, written as $\text{err}_\lambda(n_d) = \sum_{\alpha=0}^{5}(\lambda_{r,\alpha} - \hat{\lambda}_\alpha)^2 / \sum_{\alpha=0}^{5} \lambda_{r,\alpha}^2$. It can be seen that the error decreases as n_d increases. For $n_d = n_{\text{MC}} = 1200$, Fig. 2a compares the reference eigenvalues $\lambda_{r,\alpha} = \alpha/2$ (see Eq. (6.10)) with the computed eigenvalues $\hat{\lambda}_\alpha$ for $\alpha = 0, 1, \dots, 5$. The comparison is good enough.



(a) graphs $\alpha \mapsto \lambda_{r,\alpha}$ (diamond) and $\alpha \mapsto \hat{\lambda}_\alpha$ (circle).



(b) Graph $n_d \mapsto \text{err}_\lambda(n_d)$.

Figure 2: (a) Comparison of the reference eigenvalues $\lambda_{r,\alpha}$ (diamond) with the computed eigenvalues $\hat{\lambda}_\alpha$ (circle) for $\alpha = 0, 1, \dots, 5$. (b) Graph of $n_d \mapsto \text{err}_\lambda(n_d)$ quantifying the relative error between $\alpha \mapsto \lambda_{r,\alpha}$ and $\alpha \mapsto \hat{\lambda}_\alpha$.

7. Vector basis for PLoM derived from the transient anisotropic kernel, connected to the DMAPS basis

We revisit the objective presented in Section 1. The diffusion-maps (DMAPS) basis, used by PLoM [1, 30, 32], is associated with the isotropic kernel detailed in [2, 3]. This section introduces the construction of a transient vector basis, based on the transient anisotropic kernel described in Section 5 (see Proposition 6). This approach incorporates the requirement that as $\Delta t \to 0$, the transient kernel at the first time Δt coincides with the DMAPS isotropic kernel. Consequently, the two vector bases will be linked asymptotically as $t \to 0$, and we will refer to this property as "connected". To develop such a "connected" transient anisotropic kernel to the DMAPS isotropic kernel, we undertake a

reparameterization of Eqs.(5.31) to (5.33), which defines the matrix $[\widehat{K}(n\Delta t)] \in \mathbb{M}_{n_d}^{+0}$ for $n \in \mathcal{N}$. While the formulation defined by Eqs.(5.31) to (5.33) of Proposition 6 suffers from the curse of dimensionality, the reparameterization defined by Eqs.(7.1) to (7.3) does not suffer from the curse of dimensionality. This reformulation enables the construction of the vector basis for high dimensions (large values of ν) and relatively small n_d, as implemented in PLoM, designed for probabilistic learning with small training datasets.

7.1. Transient anisotropic kernel connected to the DMAPS isotropic kernel

The following definition provides the construction of a transient anisotropic kernel that is linked to the DMAPS isotropic kernel. This construction draws inspiration from the expression of $[\widehat{K}(\Delta t)]$ defined by Eq. (5.31).

Definition 5 (Transient anisotropic kernel connected to DMAPS). For each n in $\mathcal{N}$, for Δt defined in Section 5.3-(i), for n_{MC} and $[y_n^\ell]$ defined in Section 5.3-(v), for $[\sigma_n]$ defined by Eq. (5.28), and for a given real $\varepsilon_{\text{DM}} > 0$, we define the matrix $\tilde{K}(n\Delta t)] \in \mathbb{M}_{n_d}^{+0}$, such that

$$[\tilde{K}(n\Delta t)] = [B]^{-1}[\mathcal{K}(n\Delta t)]\,, \tag{7.1}$$

such that, for all i and j in $\{1,\ldots,n_d\}$, the matrix $[\mathcal{K}(n\Delta t)] \in \mathbb{M}_{n_d}$ has entries,

$$[\mathcal{K}(n\Delta t)]_{ij} = \frac{1}{n_{\text{MC}}} \sum_{\ell=1}^{n_{\text{MC}}} \Big(\Pi_{k=1}^{\nu}\{[\sigma_n]_{kj}/\Delta t\}\Big)^{-1} \exp\left\{-\frac{1}{4\varepsilon_{\text{DM}}} \sum_{k=1}^{\nu} \left(\frac{\eta_k^i - [y_n^\ell]_{kj}}{[\sigma_n]_{kj}/\sqrt{\Delta t}}\right)^2\right\}, \tag{7.2}$$

and $[B] \in \mathbb{M}_{n_d}$ is a diagonal matrix whose entries are,

$$[B]_{ij} = \delta_{ij} \sum_{j'=1}^{n_d} \exp\left\{-\frac{1}{4\varepsilon_{\text{DM}}}\|\boldsymbol{\eta}^i - \boldsymbol{\eta}^{j'}\|^2\right\}. \tag{7.3}$$

Remark 5 (About the choice of the smoothing parameter ε_{DM}). An optimal value, ε_{opt}, for the smoothing parameter ε_{DM}, is proposed for the PLoM algorithm in [32] and is detailed in Appendix A.2-(i). This optimal value enables the analysis of high-dimensional problems (large value of ν). It should be noted that Proposition 7 will explain the definition of $[\tilde{K}(n\Delta t)]$ as defined by Eqs. (7.1) to (7.3), showing the connection with DMAPS.

Proposition 7 (Limit of $[\tilde{K}(\Delta t)]$ for $\Delta t \to 0$). *We use the notation introduced in Definition 5. Let $[\mathcal{K}_{\text{DM}}] \in \mathbb{M}_{n_d}^{+0}$ be the matrix of the DMAPS isotropic kernel, defined for all $i, j \in \{1,\ldots,n_d\}$, by*

$$[\mathcal{K}_{\text{DM}}]_{ij} = \exp\left\{-\frac{1}{4\,\varepsilon_{\text{DM}}}\|\boldsymbol{\eta}^i - \boldsymbol{\eta}^j\|^2\right\}. \tag{7.4}$$

Let $[K_{\text{DM}}] \in \mathbb{M}_{n_d}$ be the matrix defined by

$$[K_{\text{DM}}] = [B]^{-1}[\mathcal{K}_{\text{DM}}]\,, \tag{7.5}$$

where $[B]$ is the diagonal matrix defined by Eq. (7.3). Let Δt be defined by

$$\Delta t = \hat{s}^2/\kappa \quad , \quad \kappa \geq 1\,, \tag{7.6}$$

where $\hat{s}$ is defined by Eq. (2.4). Then, as $\kappa \to +\infty$, and consequently $\Delta t \to 0$, we have

$$[\tilde{K}(\Delta t)] \to [K_{\text{DM}}]\,, \tag{7.7}$$

which means that, for $n = 1$, the matrix $[\tilde{K}(n\Delta t)]$ of the transient anisotropic kernel, defined by Eq. (7.1), converges to the matrix $[K_{\text{DM}}]$ of the DMAPS isotropic kernel as $\Delta t \to 0$.

PROOF. (Proposition 7). For $n = 1$, for $\kappa \to +\infty$, that is to say, for $\Delta t \to 0$, Eqs. (5.2) and (5.3) show that, for all $\ell \in \{1,\ldots,n_{\text{MC}}\}$, $k \in \{1,\ldots,\nu\}$, and $j \in \{1,\ldots,n_d\}$, we have $[y_1^\ell]_{kj} \to \eta_k^j$ and $[\sigma_n]_{kj} \to \sqrt{\Delta t}$, and consequently, $[\mathcal{K}(\Delta t)] \to [\mathcal{K}_{\text{DM}}]$. From Eqs. (7.1) and (7.5), it can be deduced Eq. (7.7).

7.2. Construction of the reduced-order transient basis and its counterpart for the DMAPS basis

In this section, we: (i) review the construction of the reduced-order diffusion-maps (DMAPS) basis (RODB) represented by a matrix $[g_{\rm DM}] \in \mathbb{M}_{n_d,m_{\rm opt}}$ with $m_{\rm opt} < n_d$, and (ii) construct, for each $n \in \mathcal{N}$, the reduced-order transient basis (ROTB($n\Delta t$)) represented by a matrix $[g(n\Delta t)] \in \mathbb{M}_{n_d,m_{\rm opt}}$.

(i) Reminder of the construction of the RODB. This construction, due to [2], is the one used by the PLoM algorithm [1, 30]. The matrix $[K_{\rm DM}]$, defined by Eq. (7.5), has positive entries and represents the transition matrix of a Markov chain. The eigenvalues $b_{{\rm DM},1},\ldots,b_{{\rm DM},n_d}$ and the associated eigenvectors $\boldsymbol{g}^1_{\rm DM},\ldots,\boldsymbol{g}^{n_d}_{\rm DM}$ are such that

$$[K_{\rm DM}]\,\boldsymbol{g}^\beta_{\rm DM} = b_{{\rm DM},\beta}\,\boldsymbol{g}^\beta_{\rm DM} \quad , \quad 1 = b_{{\rm DM},1} > b_{{\rm DM},2} \geq \ldots \geq b_{{\rm DM},n_d}\,. \tag{7.8}$$

Using Eq. (7.5), this eigenvalue problem is rewritten as the symmetric eigenvalue problem,

$$\mathbb{P}^S_{\rm DM}\,\boldsymbol{\phi}^\beta = b_{{\rm DM},\beta}\,\boldsymbol{\phi}^\beta \quad , \quad \langle \boldsymbol{\phi}^\beta, \boldsymbol{\phi}^{\beta'}\rangle = \delta_{\beta\beta'} \quad , \quad \boldsymbol{g}^\beta_{\rm DM} = [B]^{-1/2}\boldsymbol{\phi}^\beta\,, \tag{7.9}$$

in which $\mathbb{P}^S_{\rm DM} = [B]^{-1/2}[\mathcal{K}_{\rm DM}]\,[B]^{-1/2}$ is a symmetric matrix. The diffusion-maps basis $\{\boldsymbol{g}^1_{\rm DM},\ldots,\boldsymbol{g}^{n_d}_{\rm DM}\}$ forms a vector basis of $\mathbb{R}^{n_d}$. As explained in [1, 30], PLoM uses the RODB of order m, which is defined by $\{\boldsymbol{g}^1_{\rm DM}\ldots\boldsymbol{g}^m_{\rm DM}\}$. This basis depends on $\varepsilon_{\rm DM}$ and m. The optimal value $m_{\rm opt}$ of m is defined (see [32]) by

$$m_{\rm opt} = \nu + 1\,. \tag{7.10}$$

The optimal value $\varepsilon_{\rm opt}$ of $\varepsilon_{\rm DM}$ is estimated to obtain

$$1 = b_{{\rm DM},1} > b_{{\rm DM},2} \simeq \ldots \simeq b_{{\rm DM},m_{\rm opt}} \gg b_{{\rm DM},m_{\rm opt}+1} \geq \ldots \geq b_{{\rm DM},n_d} > 0\,, \tag{7.11}$$

in which the jump amplitude

$$J_{\rm DM} = b_{{\rm DM},m_{\rm opt}+1}/b_{{\rm DM},m_{\rm opt}}\,, \tag{7.12}$$

must be equal to $J_{\rm DM} = 0.1$ (following [30]), but which can also be chosen in the interval $[0.1\,,0.5]$ when ν is large. Therefore, the RODB is defined for $m = m_{\rm opt}$ and is represented by the matrix,

$$[g_{\rm DM}] = [\boldsymbol{g}^1_{\rm DM}\ldots\boldsymbol{g}^{m_{\rm opt}}_{\rm DM}] \in \mathbb{M}_{n_d,m_{\rm opt}}\,. \tag{7.13}$$

(ii) Construction of the ROTB($n\Delta t$). For each n in $\mathcal{N}$, the ROTB($n\Delta t$) represented by $[g(n\Delta t)]$ is constructed as the eigenvectors of matrix $[\tilde{K}(n\Delta t)]$ defined by Eq. (7.1). Taking into account Eq. (7.7), we want that for $n = 1$, $[g(\Delta t)] \in \mathbb{M}_{n_d,m_{\rm opt}}$ goes to $[g_{\rm DM}] \in \mathbb{M}_{n_d,m_{\rm opt}}$ as $\Delta t \to 0$. For $n > 1$, matrix $[\tilde{K}(n\Delta t)]$ is *a priori* symmetric, but matrix $[\mathcal{K}(n\Delta t)]$ is not symmetric. For applying a similar approach to the one defined by Eq. (7.9), we then symmetrize the matrix $[\mathbb{P}(n\Delta t)] = [B]^{-1/2}[\mathcal{K}(n\Delta t)]\,[B]^{-1/2}$, introducing $[\mathbb{P}^S(n\Delta t)] = ([\mathbb{P}(n\Delta t)] + [\mathbb{P}(n\Delta t)]^T)/2$. The transient basis $\{\boldsymbol{g}^1(n\Delta t),\ldots,\boldsymbol{g}^{n_d}(n\Delta t)\}$ associated with the eigenvalues $\tilde{b}_1(n\Delta t) \geq \tilde{b}_2(n\Delta t) \geq \ldots \geq \tilde{b}_{n_d}(n\Delta t)$ are then computed by solving the symmetric eigenvalue problem

$$[\mathbb{P}^S(n\Delta t)]\,\boldsymbol{\varphi}^\beta(n\Delta t) = \tilde{b}_\beta(n\Delta t)\,\boldsymbol{\varphi}^\beta(n\Delta t) \quad , \quad \langle \boldsymbol{\varphi}^\beta(n\Delta t)\,,\boldsymbol{\varphi}^{\beta'}(n\Delta t)\rangle = \delta_{\beta\beta'}\,. \tag{7.14}$$

$$\boldsymbol{g}^\beta(n\Delta t) = [B]^{-1/2}\boldsymbol{\varphi}^\beta(n\Delta t)\,. \tag{7.15}$$

The ROTB($n\Delta t$) is then defined by the matrix

$$[g(n\Delta t)] = [\boldsymbol{g}^1(n\Delta t)\ldots\boldsymbol{g}^{m_{\rm opt}}(n\Delta t)] \in \mathbb{M}_{n_d,m_{\rm opt}}\,, \tag{7.16}$$

in which $m_{\rm opt}$ is defined as explained in Section 7.2-(i).

7.3. Criteria for comparing the reduced-order transient basis with the reduced-order DMAPS basis

The PLoM algorithm is based on the use of the RODB (see Appendix A). With such an RODB based on the isotropic kernel, PLoM has proven generally efficient, even in extremely difficult cases, as demonstrated in numerous publications since 2016. This efficacy will also be evident through the applications presented in Section 8. However, for cases involving very heterogeneous data in the training dataset, the learned statistical dependence between the components of the random vector $\boldsymbol{H}$ (through the learned probability measure for $\boldsymbol{H}$) can *a priori* be improved using ROTB($n\Delta t$) for given n. To assess a possible improvement with this reduced-order transient basis compared to the reduced-order DMAPS basis, quantitative criteria are necessary. In this section, we introduce criteria for comparing the two vector bases, and then, in Section 7.4, we will present a methodology for identifying the instance, $n\Delta t$, that maximizes the selection criteria of the ROTB($n\Delta t$).

(i) The first criterion will be the angle between the two vector subspaces generated by the two vector bases, ROTB($n\Delta t$) and RODB. If this angle is close to zero then the two bases coincide. This must be the case when we choose the instant $n = 1$ for the ROTB($n\Delta t$) (See Proposition 7).

(ii) The second criterion is the concentration of the learned probability measure in relation to the concentration of the probability measure of the training dataset. It has been proven that the PLoM algorithm, which uses the RODB, was designed to preserve the concentration of the learned probability measure from a small training dataset. For this, in [30, 32], we introduced the indicator d^2, linked to the mean-square convergence, which we will recall and use. We also introduce a second concentration criterion, KL, based on the Kullback-Leibler divergence [92, 93] between the learned probability measure and the probability measure of the training dataset.

(iii) The third criterion is mutual information [94, 95], which is defined as the relative entropy introduced by Kullback and Leibler [93]. This will be used to quantify the level of statistical dependencies among the components of the centered random vector $\boldsymbol{H}$, whose covariance matrix is the identity matrix. Such mutual information will be estimated for the learned probability measure generated by PLoM, comparing the RODB and the ROTB($n\Delta t$) as a function of n.

(iv) The last criterion is the Entropy from Information Theory [96, 95, 97], introduced by Shannon [98]. The estimation of this entropy is a function of the number of realizations used for estimating the probability measure. This property will be used in Section 7.4 to normalize the estimation of the mutual information.

7.3.1. Angle between the subspaces spanned by ROTB($n\Delta t$) and RODB

Let n be fixed in $\mathcal{N}$. It is assumed that $\text{rank}([g_{\text{DM}}]) = \text{rank}([g(n\Delta t)]) = m_{\text{opt}}$ in which the two matrices are defined by Eqs. (7.13) and (7.16). In addition, it is assumed that the null space of the matrix $[g(n\Delta t)]^T\,[g_{\text{DM}}]$ is $\{0\}$. Let $V(n\Delta t) = \text{span}\{\boldsymbol{g}^1(n\Delta t), \ldots, \boldsymbol{g}^{m_{\text{opt}}}(n\Delta t)\}$ be the m_{opt}-dimension subspace of $\mathbb{R}^{n_d}$ (spanned by the ROTB($n\Delta t$)). Let $V_{\text{DM}} = \text{span}\{\boldsymbol{g}^1_{\text{DM}}, \ldots, \boldsymbol{g}^{m_{\text{opt}}}_{\text{DM}}\}$ be the m_{opt}-dimension subspace of $\mathbb{R}^{n_d}$ (spanned by the RODB). For $\beta \in \{1, \ldots m_{\text{opt}}\}$, let $\hat{g}^\beta_{\text{DM}} = g^\beta_{\text{DM}}/\|g^\beta_{\text{DM}}\|$ be the normalized vector g^β_{DM} and let $\hat{g}^\beta(n\Delta t) = g^\beta(n\Delta t)/\|g^\beta(n\Delta t)\|$ be the normalized vector $g^\beta(n\Delta t)$. Let $[\hat{g}_{\text{DM}}] = [\hat{g}^1_{\text{DM}} \ldots \hat{g}^{m_{\text{opt}}}_{\text{DM}}]$ and $[\hat{g}(n\Delta t)] = [\hat{g}^1(n\Delta t) \ldots \hat{g}(n\Delta t)^{m_{\text{opt}}}]$ be matrices in $\mathbb{M}_{n_d, m_{\text{opt}}}$. The angle $\gamma(n\Delta t)$ between the subspaces $V(n\Delta t)$ and V_{DM} is defined by

$$\gamma(n\Delta t) = \arccos(\sigma_{\min}([\hat{g}(n\Delta t)]^T\,[\hat{g}_{\text{DM}}]))\,, \tag{7.17}$$

in which $\sigma_{\min}$ denotes the smallest singular value. If the angle is close to 0, the two subspaces are nearly linearly dependent.

7.3.2. Indicators related to the learned probability measure

In this section, we detail the indicators used for comparison: concentration of the learned probability measure, mutual information, and entropy. These will be employed to compare the probability measures associated with the training dataset and the learned dataset generated by the PLoM algorithm, as summarized in Appendix A. Specifically, we use the reduced-order DMAPS basis (RODB) and, alternatively, the reduced-order transient basis (ROTB($n\Delta t$)). To facilitate the comparisons presented in the numerical illustrations, we introduce the necessary notations for clarifying the diverse quantities and their numerical calculations.

- *Training dataset*. In Section 2, the independent realizations of the $\mathbb{R}^\nu$-valued random variable $\boldsymbol{H}$ are $\boldsymbol{\eta}^1, \ldots, \boldsymbol{\eta}^{n_d}$, and the matrix $[\eta_d] = [\boldsymbol{\eta}^1 \ldots \boldsymbol{\eta}^{n_d}] \in \mathbb{M}_{\nu, n_d}$ represents one realization of the random matrix $[\boldsymbol{H}]$, used by PLoM, and defined in Appendix A.1. The probability density function of $\boldsymbol{H}$ is $p_{\boldsymbol{H}}$ defined by Eq. (2.3).

- *Learned dataset*. In Appendix A, the learned realizations generated by PLoM are those of the $\mathbb{M}_{\nu,n_d}$-valued random variable $[\boldsymbol{H}_{\rm ar}]$ and are written as $[\eta^\ell_{\rm ar}]$, $\ell = 1,\ldots,n_{\rm MCH}$ (see Appendix A.3). The corresponding realizations $\boldsymbol{\eta}^{\ell'}_{\rm ar}$, $\ell = 1,\ldots,n_{\rm ar}$, with $n_{\rm ar} = n_d \times n_{\rm MCH}$, of the $\mathbb{R}^\nu$-valued random variable $\boldsymbol{H}_{\rm ar}$ are obtained by reshaping. The learned probability density function of $\boldsymbol{H}_{\rm ar}$ is denoted by $p_{\boldsymbol{H}_{\rm ar}}$. When PLoM is used with RODB, $[\boldsymbol{H}_{\rm ar}]$, $[\eta^\ell_{\rm ar}]$, $\boldsymbol{\eta}^{\ell'}_{\rm ar}$, and $p_{\boldsymbol{H}_{\rm ar}}$ will be rewritten as $[\boldsymbol{H}_{\rm DB}]$, $[\eta^\ell_{\rm DB}]$, $\boldsymbol{\eta}^{\ell'}_{\rm DB}$, and $p_{\rm DB}$, respectively. When PLoM is used with ROTB($n\Delta t$), $[\boldsymbol{H}_{\rm ar}]$, $[\eta^\ell_{\rm ar}]$, $\boldsymbol{\eta}^{\ell'}_{\rm ar}$, and $p_{\boldsymbol{H}_{\rm ar}}$ will be rewritten as $[\boldsymbol{H}_{\rm TB}(n\Delta t)]$, $[\eta^\ell_{\rm TB}(n\Delta t)]$, $\boldsymbol{\eta}^{\ell'}_{\rm TB}(n\Delta t)$, and $p_{\rm TB}(.\,;n\Delta t)$, respectively.

(i) Concentration of the learned probability measure for RODB and ROTB($n\Delta t$). Based on the mean-square norm, for PLoM with RODB and ROTB($n\Delta t$), the learned probability measure concentration is written (see Eq. (A.6)) as

$$d^2_{\rm DB}(m_{\rm opt}) = E\{\,\|\,[\boldsymbol{H}_{\rm DB}] - [\eta_d]\,\|^2\,\}\,/\,\|\,[\eta_d]\,\|^2\,, \tag{7.18}$$

$$d^2_{\rm TB}(m_{\rm opt};n\Delta t) = E\{\,\|\,[\boldsymbol{H}_{\rm TB}(n\Delta t)] - [\eta_d]\,\|^2\,\}\,/\,\|\,[\eta_d]\,\|^2\,. \tag{7.19}$$

Using the realizations, these quantities are estimated by

$$\hat{d}^2_{\rm DB}(m_{\rm opt}) = (1/n_{\rm MCH})\sum_{\ell=1}^{n_{\rm MCH}}\{\,\|\,[\boldsymbol{\eta}^\ell_{\rm DB}] - [\eta_d]\,\|^2\,\}\,/\,\|\,[\eta_d]\,\|^2\,. \tag{7.20}$$

$$\hat{d}^2_{\rm DB}(m_{\rm opt};n\Delta t) = (1/n_{\rm MCH})\sum_{\ell=1}^{n_{\rm MCH}}\{\,\|\,[\boldsymbol{\eta}^\ell_{\rm TB}(n\Delta t)] - [\eta_d]\,\|^2\,\}\,/\,\|\,[\eta_d]\,\|^2\,. \tag{7.21}$$

The concentration of the learned probability measure can also be estimated using the Kullback-Leiber divergence and its estimation from a set of realizations as presented in Appendix B.1. For the PLoM formulated with RODB and ROTB($n\Delta t$), we then have,

$$D(p_{\rm DB}\,\|\,p_{\boldsymbol{H}}) = \int_{\mathbb{R}^\nu} p_{\rm DB}(\boldsymbol{\eta})\,\log\left(\frac{p_{\rm DB}(\boldsymbol{\eta})}{p_{\boldsymbol{H}}(\boldsymbol{\eta})}\right)d\boldsymbol{\eta}\,, \tag{7.22}$$

$$D(p_{\rm TB}(.\,;n\Delta t)\,\|\,p_{\boldsymbol{H}}) = \int_{\mathbb{R}^\nu} p_{\rm TB}(\boldsymbol{\eta}\,;n\Delta t)\,\log\left(\frac{p_{\rm TB}(\boldsymbol{\eta};n\Delta t)}{p_{\boldsymbol{H}}(\boldsymbol{\eta})}\right)d\boldsymbol{\eta}\,. \tag{7.23}$$

(ii) Mutual-information-based components statistical dependencies of the learned probability measure for RODB and ROTB($n\Delta t$). The mutual information and its estimation from a set of realizations are presented in Appendix B.2. For the training dataset and for PLoM formulated with RODB and ROTB($n\Delta t$), we have , for $\mathbb{R}^\nu$-valued random variables $\boldsymbol{H}$, $\boldsymbol{H}_{\rm DB}$, and $\boldsymbol{H}_{\rm TB}(n\Delta t)$,

$$I(\boldsymbol{H}) = D(p_{\boldsymbol{H}}\,\|\,\otimes_{k=1}^\nu p_{H_k}) = \int_{\mathbb{R}^\nu} p_{\boldsymbol{H}}(\boldsymbol{\eta})\,\log\left(\frac{p_{\boldsymbol{H}}(\boldsymbol{\eta})}{\Pi_{k=1}^\nu p_{H_k}(\eta_k)}\right)d\boldsymbol{\eta}\,, \tag{7.24}$$

$$I(\boldsymbol{H}_{\rm DB}) = D(p_{\rm DB}\,\|\,\otimes_{k=1}^\nu p_{{\rm DB},k}) = \int_{\mathbb{R}^\nu} p_{\rm DB}(\boldsymbol{\eta})\,\log\left(\frac{p_{\rm DB}(\boldsymbol{\eta})}{\Pi_{k=1}^\nu p_{{\rm DB},k}(\eta_k)}\right)d\boldsymbol{\eta}\,, \tag{7.25}$$

$$I(\boldsymbol{H}_{\rm TB};n\Delta t) = D(p_{\rm TB}(\cdot\,;n\Delta t)\,\|\,\otimes_{k=1}^\nu p_{{\rm TB},k}(\cdot\,;n\Delta t)) = \int_{\mathbb{R}^\nu} p_{\rm TB}(\boldsymbol{\eta}\,;n\Delta t)\,\log\left(\frac{p_{\rm TB}(\boldsymbol{\eta}\,;n\Delta t)}{\Pi_{k=1}^\nu p_{{\rm TB},k}(\eta_k\,;n\Delta t)}\right)d\boldsymbol{\eta}\,, \tag{7.26}$$

in which $p_{{\rm DB},k}$ is the pdf of component $H_{{\rm DB},k}$ of $\boldsymbol{H}_{\rm DB}$ and where $p_{{\rm TB},k}(\cdot\,;n\Delta t)$ is the pdf of component $H_{{\rm TB},k}(n\Delta t)$ of $\boldsymbol{H}_{\rm TB}(n\Delta t)$.

(iii) Entropy for RODB and ROTB($n\Delta t$). The entropy and its estimation from a set of realizations are presented in Appendix B.3. For the training dataset and for PLoM formulated with RODB and ROTB($n\Delta t$), we have for $\mathbb{R}^\nu$-valued random variables $\boldsymbol{H}$, $\boldsymbol{H}_{\rm DB}$, and $\boldsymbol{H}_{\rm TB}(n\Delta t)$,

$$S_{\boldsymbol{H}} = -\int_{\mathbb{R}^\nu} p_{\boldsymbol{H}}(\boldsymbol{\eta})\,\log p_{\boldsymbol{H}}(\boldsymbol{\eta})\,d\boldsymbol{\eta} \quad , \quad S_{\rm DB} = -\int_{\mathbb{R}^\nu} p_{\rm DB}(\boldsymbol{\eta})\,\log p_{\rm DB}(\boldsymbol{\eta})\,d\boldsymbol{\eta}\,, \tag{7.27}$$

$$S_{\rm TB}(n\Delta t) = -\int_{\mathbb{R}^\nu} p_{\rm TB}(\boldsymbol{\eta}\,;n\Delta t)\,\log p_{\rm TB}(\boldsymbol{\eta}\,;n\Delta t)\,d\boldsymbol{\eta}\,. \tag{7.28}$$

7.4. Identification methodology for the instant maximizing the selection criterion of the ROTB($n\Delta t$)

In this section, all quantities denoted with a hat (such as $\hat{I}(\boldsymbol{H})$) represent the estimated values of the corresponding quantities without a hat (such as $I(\boldsymbol{H})$), using the realizations as explained in Appendix B. Let n_d be the number of points in the training dataset and $n_{\text{ar}} = n_d \times n_{\text{MCH}} \gg n_d$ be the number of learned realizations with PLoM (see Appendix A.3), either with the RODB or with the ROTB($n\Delta t$).

(i) Defining the subset $\mathscr{C}_N$ of $\mathcal{N}$ containing the admissible values of n. Let N be the largest value of n for which the ROTB($n\Delta t$), represented by matrix $[g(n\Delta t)]$, is computed. This value N being fixed, the set $\mathcal{N} = \{1, \ldots, N\}$ defined by Eq. (5.24) is fixed. The subset $\mathscr{C}_N$ of the admissible values of n is then defined by

$$\mathscr{C}_N = \{n \in \mathcal{N}\,,\ \frac{1}{\nu}\,\hat{d}^2_{\text{TB}}(m_{\text{opt}}; n\Delta t) \leq \tau_c \ll 1\}\,, \tag{7.29}$$

where $\hat{d}^2_{\text{TB}}(m_{\text{opt}}; n\Delta t)$ is defined by Eq. (7.21) and where $0 < \tau_c \ll 1$ is fixed sufficiently small with respect to 1 in order to preserve the concentration of the learned probability measure (see Section 7.3-(i) and Appendix A.4). It should be noted that the maximum value of $d^2_{\text{DB}}(n_d)$ is 2 (corresponding to a measure concentration completely lost, which can be obtained with classical MCMC algorithms). The PLoM algorithm allows for preserving the concentration of the learned probability measure [30] yielding values of the order 0.01 to 0.1 through all the performed applications. For defining the concentration criterion in Eq. (7.29), we have normalized with respect to dimension ν.

(ii) Characterizing a better ROTB($n\Delta t$) compared to RODB. From Proposition 7, in particular from Eq. (7.7), it can be deduced that, for $n = 1$, we have $\hat{I}(\boldsymbol{H}_{\text{TB}}\,; 1 \times \Delta t) \simeq \hat{I}(\boldsymbol{H}_{\text{DB}})$ for Δt sufficiently small (which is the considered case). Let us assume that $\hat{I}(\boldsymbol{H}) < \hat{I}(\boldsymbol{H}_{\text{DM}})$. For n fixed in $\mathscr{C}_N$, we will say that the ROTB($n\Delta t$), represented by matrix $[g(n\Delta t)]$, is better than the RODB, represented by matrix $[g_{\text{DM}}]$, if

$$\hat{I}(\boldsymbol{H}) \leq \hat{I}(\boldsymbol{H}_{\text{TB}}\,; n\Delta t) < \hat{I}(\boldsymbol{H}_{\text{DM}})\,. \tag{7.30}$$

(iii) Determining the optimal value n_{opt} of n in $\mathscr{C}_N$. The underlying idea is to select a reduced-order transient basis that gives a learned probability measure whose mutual information is as close as possible to the mutual information of the probability measure $p_{\boldsymbol{H}}(\boldsymbol{\eta}), d\boldsymbol{\eta}$. Based on paragraph (ii) above, the optimal transient basis $[g(n_{\text{opt}})]$ among the set of possible transient bases $\{[g(n\Delta t)], n \in \mathcal{N}\}$ is obtained for $n = n_{\text{opt}}$ such that

$$n_{\text{opt}} = \arg\min_{n \in \mathscr{C}_N} \hat{I}(\boldsymbol{H}_{\text{TB}}\,; n\Delta t)\,. \tag{7.31}$$

(iv) Defining the normalized estimate of the mutual information. As explained in Appendix B.3, if n_{samp} is the number of realizations (either n_d or n_{ar}), the entropy estimation asymptotically decreases as $-log(n_{\text{samp}})$ when n_{samp} increases. For comparing the estimate of the mutual information for $\boldsymbol{H}$, which uses n_d realizations, with the one of $\boldsymbol{H}_{\text{DB}}$ or $\boldsymbol{H}_{\text{TB}}(n\Delta t)$, which uses $n_{\text{ar}} \gg n_d$ learned realizations, we must normalize the estimated mutual information with respect to the number of realizations. Taking into account the relationship between the mutual information and the entropy estimates (see Appendix B.3), we chose to normalize the estimated mutual information by dividing by the function $\chi + \log(n_{\text{samp}})$ (with $n_{\text{samp}} = n_d$ or $n_{\text{samp}} = n_{\text{ar}}$), where χ is a real number that must be identified and which must be such that $\chi + log(n_d) > 0$. Since $n_{\text{ar}} > n_d$ this condition implies that $\chi + log(n_{\text{ar}}) > 0$ is automatically verified. For the $\mathbb{R}^\nu$-valued random variables $\boldsymbol{H}$, $\boldsymbol{H}_{\text{DB}}$, and $\boldsymbol{H}_{\text{TB}}(n\Delta t)$, the normalized estimated mutual information are then defined by

$$\hat{I}_{\text{norm}}(\boldsymbol{H}) = \frac{\hat{I}(\boldsymbol{H})}{\chi + log(n_d)} \quad , \quad \hat{I}_{\text{norm}}(\boldsymbol{H}_{\text{DB}}) = \frac{\hat{I}(\boldsymbol{H}_{\text{DB}})}{\chi + log(n_{\text{ar}})} \quad , \quad \hat{I}_{\text{norm}}(\boldsymbol{H}_{\text{TB}}\,; n\Delta t) = \frac{\hat{I}(\boldsymbol{H}_{\text{TB}}\,; n\Delta t)}{\chi + log(n_{\text{ar}})}\,. \tag{7.32}$$

in which $\hat{I}(\boldsymbol{H})$, $\hat{I}(\boldsymbol{H}_{\text{DB}})$, and $\hat{I}(\boldsymbol{H}_{\text{TB}}\,; n\Delta t)$ are the estimates (computed with Eq. (B.6)) of the mutual information $I(\boldsymbol{H})$, $I(\boldsymbol{H}_{\text{DB}})$, and $I(\boldsymbol{H}_{\text{TB}}\,; n\Delta t)$ defined by Eqs. (7.24) to (7.26). Based on paragraph (iii) above, constant χ is defined as follows. Let χ_{opt} be the solution in χ of the equation,

$$\frac{\hat{I}(\boldsymbol{H})}{\chi + log(n_d)} = \frac{\hat{I}(\boldsymbol{H}_{\text{TB}}\,; n_{\text{opt}}\,\Delta t)}{\chi + log(n_{\text{ar}})}\,. \tag{7.33}$$

If $\chi_{\text{opt}} + \log(n_d) > 0$, then χ_{opt} is the desired value of χ. It should be noted that the optimal value n_{opt} is calculated with the estimation of the non-normalized mutual information (see Eq. (7.31). Normalization is only introduced for the purpose of comparing the value of this criterion for $\boldsymbol{H}$ and $\boldsymbol{H}_{\text{TB}}$.

(v) Identification of the instant maximizing the selection criterion of the ROTB($n\Delta t$). Let n_{opt} be defined by Eq. (7.31), and let χ_{opt} be identified by solving Eq. (7.33). Then, we have

$$\hat{I}_{\text{norm}}(\boldsymbol{H}) = \hat{I}_{\text{norm}}(\boldsymbol{H}_{\text{TB}}\,; n_{\text{opt}}\Delta t) \leq \hat{I}_{\text{norm}}(\boldsymbol{H}_{\text{TB}}\,; n\Delta t) \quad , \quad \forall n \in \mathscr{C}_N\,. \tag{7.34}$$

We can then conclude that the ROTB($n_{\text{opt}}\Delta t$), represented by matrix $[g(n_{\text{opt}}\Delta t)]$, is better than the RODB, represented by matrix $[g_{\text{DM}}]$. The proof is straightforward.

8. Numerical applications

8.1. Preamble

We will present three applications, each with its own specificities. However, as the generation of the training dataset related to the random vector $\boldsymbol{X}$ is relatively complex to describe, we will start with the training dataset related to the normalized random vector $\boldsymbol{H}$ with values in $\mathbb{R}^{\nu}$. The number of realizations in the training dataset is n_d, and the training dataset is represented by the matrix $[\eta_d] \in \mathbb{M}_{\nu,n_d}$. This means that the PCA step of PLoM, which transforms $\boldsymbol{X}$ into $\boldsymbol{H}$ (see Appendix A.1), is not detailed here. Readers interested in the training dataset for $\boldsymbol{H}$ can request, from the "corresponding author" of the article, the transfer of the matrix $[\eta_d]$ for each application presented. We will still briefly give the main specificities of these 3 applications.

Application 1. This application (*Appli 1*) was created so that the probability measure of $\boldsymbol{H}$ in $\mathbb{R}^9$ ($\nu = 9$), which is defined by the points of the training dataset, is concentrated in a multiconnected domain of $\mathbb{R}^9$, with the constituent connected parts being manifolds of dimensions much lower than 9, each having different dimensions. These parts may or may not be connected to each other.

Application 2. For the second application (*Appli 2*), the n_d realizations of the random vector $\boldsymbol{H}$ with values in $\mathbb{R}^8$ are generated using a polynomial chaos expansion of degree 6 of a real random variable, whose random germ is of dimension 2, with each of the two random germs being a uniform random variable of different support. There are therefore 28 terms in this expansion, and the 8 components of $\boldsymbol{H}$ are defined as the random terms of rank 2, 3, 6, 8, 12, 13, 17, and 19. We thus define a relatively complex random manifold in $\mathbb{R}^8$.

Application 3. The third application (*Appli 3*) results from a statistical treatment of an experimental database containing photon measurements in the ATLAS detector at CERN. This dataset was obtained by loading the file 'pid22˙E262144˙eta˙20˙25˙voxalisation.csv' from the free access CERN Open Data Portal. The PCA step of PLoM has been performed, and an extraction of 45 components has been done to obtain the training dataset for the $\mathbb{R}^{45}$-valued random variable $\boldsymbol{H}$. This application is in higher dimension than the first two, but the statistical complexity is less.

The following text gives explanations for how the figures must be read and interpreted. The probability density functions (pdfs) of components of $\boldsymbol{H}$, which correspond to the training dataset, and the learned dataset constructed with a classical MCMC algorithm, are plotted. Since the training dataset is small, there is a well-known smoothing effect; however, there are training data points in the multiconnected domain that should induce a multimodality in each first-order marginal pdf (this aspect can be visualized by examining the clouds of points). On the other hand, the use of a classical MCMC method results in a scattering of the learned realizations, and therefore the pdfs are smooth, causing the multimodal character to be lost. In other words, the required concentration of the learned probability measure is lost. Conversely, the figures show that PLoM with RODB preserves the concentration of the probability measure, allowing for a multimodal first-order marginal probability density function. These good properties of PLoM with RODB are also obtained by PLoM with ROTB, and for the first-order marginal pdfs, there are no significant differences, which are the expected results. As explained in Section 1.1, the objective of PLoM with ROTB instead of RODM is to see if it is possible to improve the learned joint probability measure of the components of $\boldsymbol{H}$, because the quality of this joint probability measure directly impacts the quality of the statistical surrogate model of $\boldsymbol{Q}$ given $\boldsymbol{W}$, built using conditional statistics. Such an improvement cannot be qualified by a simple graphical comparison of

the joint probability density function of the components of $\boldsymbol{H}$, which is of dimension $\nu \gg 1$ (in the applications ν is 8, 9 and 46). This is why a quantification has been introduced by the objective criterion of mutual information from the Information Theory. This improvement or not of the quality of the joint probability measure is demonstrated by comparing the graphs of mutual information $n \mapsto \hat{I}_{\text{norm}}(\boldsymbol{H})$ and $n \mapsto \hat{I}_{\text{norm}}(\boldsymbol{H}_{\text{TB}}; n\Delta t)$, and which allow the optimal time to be identified.

8.2. Additional convergence analysis conducted for the three applications

For each of the three applications, we will provide detailed analyses of the calculation of the optimal value of the instant $n_{\text{opt}}\Delta t$, which allows for the selection of the best reduced-order transient basis (ROTB). We will also present the convergence results of the PLoM algorithms under the normalization constraints (see Appendix A.6). However, we cannot show all the convergence results for the other parameters. Below are the different analyses that were carried out with respect to the parameters that control the construction of the reduced-order transient basis:

(i) Convergence with respect to the value of κ. This point is important because Δt must be sufficiently small to apply Proposition 7. For the three applications, we found that $\kappa = 30$ is an appropriate value.

(ii) Convergence with respect to the size of δt, that is to say, the value of n_s. For the three applications, we found that once Δt is fixed by the value of κ, the criterion associated with Eq. (5.30) was satisfied for $n_s = 1$, as well as the evolution of the angle $\gamma(n_{\text{opt}}\Delta t)$ at the optimal instant was very insensitive to the values of n_s greater than 1. This is due to the fact that Δt is already small enough to achieve good accuracy with the Euler scheme used to integrate the ISDE. We found that $n_s = 1$, thus $\delta t = \Delta t$ is a good value.

(iii) Convergence with respect to n_{MC}. This analysis was performed by examining the convergence of the angle $\gamma(n\Delta t)$ for $n \in \mathcal{N}$. We observed that a large value of n_{MC} was necessary to achieve convergence. This point is particularly important and must be carefully checked for the applications.

8.3. Parameters defining the training dataset and controlling the construction of the RODB

For each application, Table 1 provides the values of the parameters that define the training dataset and the probability measure $P_{\boldsymbol{H}}(d\boldsymbol{\eta}) = p_{\boldsymbol{H}}(\boldsymbol{\eta})\, d\boldsymbol{\eta}$ on $\mathbb{R}^{\nu}$, as well as the parameters defined in Section 7.2-(i), which control the construction of the reduced-order DMAPS basis (RODB).

Table 1: For each application, values of the parameters defining the training dataset and parameters controlling the construction of the RODB

	Training					RODB		
	ν	n_d	s	$\hat{s}$	$\hat{s}/s$	J_{DM}	m_{opt}	ε_{DM}
Appli 1	9	400	0.5835	0.5044	0.8645	0.2	10	56
Appli 2	8	400	0.5623	0.4946	0.8725	0.1	10	53
Appli 3	46	560	0.8357	0.6416	0.7677	0.5	46	74

8.4. Parameters controlling the construction of the ROTB($n\Delta t$)

For each application, Table 2 provides the values of the parameters that control the construction of the reduced-order transient basis ROTB($n\Delta t$). In particular, the optimal time $n_{\text{opt}}\Delta t$ is estimated using the indicators defined in Section 7.3.2. The evolution of these indicators as a function of n will be presented in detail in Sections 8.6 to 8.8 for the three applications.

Table 2: For each application, values of the parameters controlling the construction of the ROTB($n\Delta t$)

	ROTB($n\Delta t$)							
	N	τ_c	κ	Δt	n_s	δt	n_{MC}	n_{opt}
Appli 1	9	0.002	30	0.00848	1	0.00848	400000	9
Appli 2	10	0.002	30	0.00802	1	0.00802	400000	5
Appli 3	9	0.002	30	0.01370	1	0.01370	448000	9

8.5. Parameters of PLoM with RODB and ROTB

For each of the three applications, Table 3 provides the values of the parameters defined in Appendix A.5 and Appendix A.6 used by the PLoM algorithm with the reduced-order DMAPS basis (RODM) or the reduced-order transient basis ROTB($n\Delta t$). Taking into account the convergence analysis carried out on n_{MC} and the expertise on the statistical convergence of the quantities considered, we choose $n_{\text{MCH}} = n_{\text{MC}}$, a sufficiently large value (see Appendix A.5). For the PLoM algorithm, the constraints related to normalization, defined by Eq. (A.12), are always applied in the computation (see Appendix A.6)

Table 3: Values of the parameters of PLoM with RODB and ROTB($n\Delta t$) for each application

	PLoM with RODB and with ROTB($n_{\text{opt}}\Delta t$)									
	f_0	Δt_{SV}	M_0	n_{MCH}	n_{ar}	β_1	β_2	i_2	i_{last}	err(i_{last})
Appli 1 with RODB	4	0.1585	30	1000	400000	0.001	0.05	20	2563	0.000998
Appli 1 with ROTB	4	0.1585	30	1000	400000	0.001	0.05	20	3438	0.000994
Appli 2 with RODB	4	0.1541	30	1000	400000	0.001	0.05	20	1882	0.000999
Appli 2 with ROTB	4	0.1541	30	1000	400000	0.001	0.05	20	2589	0.000998
Appli 3 with RODB	4	0.2016	30	800	448000	0.001	0.05	20	6000	0.00261
Appli 3 with ROTB	4	0.2016	30	800	448000	0.001	0.05	20	6000	0.00383

8.6. Results for Application 1

(i) Figure 3 displays the graphs of the probability density function (pdf) of components 1, 2, 4, and 5 for $\boldsymbol{H}$ estimated with the n_d realizations of the training dataset, and the pdf estimated with n_{ar} learned realizations, for $\boldsymbol{H}_{\text{ar}}$ using MCMC without PLoM (a,d,g,j), for $\boldsymbol{H}_{\text{DB}}$ using PLoM with RODB (b, e, h, k), and for $\boldsymbol{H}_{\text{TB}}$ using PLoM with ROTB($n_{\text{opt}}\Delta t$) (c, f, i, l).
(ii) Figure 4 shows the joint probability density function of components 4 and 5 of $\boldsymbol{H}$ estimated with the n_d realizations of the training dataset (a) and estimated with n_{ar} learned realizations for $\boldsymbol{H}_{\text{ar}}$ using MCMC without PLoM (b), for $\boldsymbol{H}_{\text{DB}}$ using PLoM with RODB (c), and for $\boldsymbol{H}_{\text{TB}}$ using PLoM with ROTB($n_{\text{opt}}\Delta t$) (d).
(iii) In Fig. 5, the clouds of n_{ar} points corresponding to n_{ar} learned realizations can be seen for components 1, 2, 3 (a,b,c) and components 3, 4, 5 (d,e,f). These are shown for $\boldsymbol{H}_{\text{ar}}$ using MCMC without PLoM (a,d), for $\boldsymbol{H}_{\text{DB}}$ using PLoM with RODB (b,e), and for $\boldsymbol{H}_{\text{TB}}$ using PLoM with ROTB($n_{\text{opt}}\Delta t$) (c,f).
(iv) Figure 6 plots the functions that characterize the reduced-order transient basis ROTB($n\Delta t$) as a function of time $n\Delta t$:

- The eigenvalues of matrix $[K_{\text{DM}}]$ and those of the of symmetrized matrix $[\tilde{K}(n\Delta t)]$ are shown in Fig. 6a.
- The probability-measure concentration using the d^2/ν-criterion is shown in Fig. 6b. For the learning without PLoM, the d^2-concentration is 0.6465, which shows that the concentration is lost, and for the PLoM with the RODM, the concentration is 0.0116, which shows that the concentration is preserved.

- The other criterion of the probability-measure concentration is given by Kullback-Leibler divergence measure, shown in Fig. 6c. For the learning without PLoM, Kullback-Leibler divergence is 0.3486, and for the PLoM with the RODM, Kullback-Leibler divergence is 3.9435. Comparing Figs. 6b and 6c shows that the two criteria are consistent and give the same analysis of the concentration.

- The angle between the subspaces spanned by RODB and ROTB($n\Delta t$) is displayed in Fig. 6d. It can be seen that, for the optimal time $9\,\Delta t$, the angle is 53.03°, which is a significant angle showing that the two bases are different while the d^2-concentration remains small at 0.0174.

- The entropy estimation of pdf $p_{\text{TB}}(\cdot\,; n\Delta t)$ is given in Fig. 6e.

- The normalized mutual information (MI) of the pdfs $p_{\boldsymbol{H}}$ and $p_{\text{TB}}(\cdot\,; n\Delta t)$ is shown in Fig. 6f. This figure shows that the optimal value of n is $n_{\text{opt}} = 9$. For the non-normalized estimation of the mutual information, we have $\hat{I}(\boldsymbol{H}) = 4.4668$, $\hat{I}(\boldsymbol{H}_{\text{TB}}\,; n_{\text{opt}}\Delta t) = 6.9996$, and $\hat{I}(\boldsymbol{H}_{\text{DM}}) = 7.0945$. For the normalized one, we have $\hat{I}_{\text{norm}}(\boldsymbol{H}) = \hat{I}_{\text{norm}}(\boldsymbol{H}_{\text{TB}}\,; n_{\text{opt}}\Delta t) = 0.3666$ and $\hat{I}_{\text{norm}}(\boldsymbol{H}_{\text{DM}}) = 0.3716$.

(v) Finally, examination of these figures shows that traditional learning without PLoM gives poor results compared to PLoM, which allows the concentration to be preserved and properly learns the geometry of the probability measure support. We also see that PLoM with the optimal ROTB provides an improvement in learning compared to PLoM with RODM and, therefore, should improve the estimates of conditional statistics thanks to better learning of the joint probability measure.

8.7. Results for Application 2

(i) Figure 7 displays the graphs of the probability density function (pdf) of components 3, 5, 6, and 8 for $\boldsymbol{H}$ estimated with the n_d realizations of the training dataset, and the pdf estimated with n_{ar} learned realizations, for $\boldsymbol{H}_{\text{ar}}$ using MCMC without PLoM (a,d,g,j), for $\boldsymbol{H}_{\text{DB}}$ using PLoM with RODB (b, e, h, k), and for $\boldsymbol{H}_{\text{TB}}$ using PLoM with ROTB($n_{\text{opt}}\Delta t$) (c, f, i, l).
(ii) Figure 8 shows the joint probability density function of components 6 and 8 of $\boldsymbol{H}$ estimated with the n_d realizations of the training dataset (a) and estimated with n_{ar} learned realizations for $\boldsymbol{H}_{\text{ar}}$ using MCMC without PLoM (b), for $\boldsymbol{H}_{\text{DB}}$ using PLoM with RODB (c), and for $\boldsymbol{H}_{\text{TB}}$ using PLoM with ROTB($n_{\text{opt}}\Delta t$) (d).
(iii) In Fig. 9, the clouds of n_{ar} points corresponding to n_{ar} learned realizations can be seen for components 2, 5, 6 (a,b,c) and components 3, 6, 8 (d,e,f). These are shown for $\boldsymbol{H}_{\text{ar}}$ using MCMC without PLoM (a,d), for $\boldsymbol{H}_{\text{DB}}$ using PLoM with RODB (b,e), and for $\boldsymbol{H}_{\text{TB}}$ using PLoM with ROTB($n_{\text{opt}}\Delta t$) (c,f).
(iv) Figure 10 plots the functions that characterize the reduced-order transient basis ROTB($n\Delta t$) as a function of time $n\Delta t$:

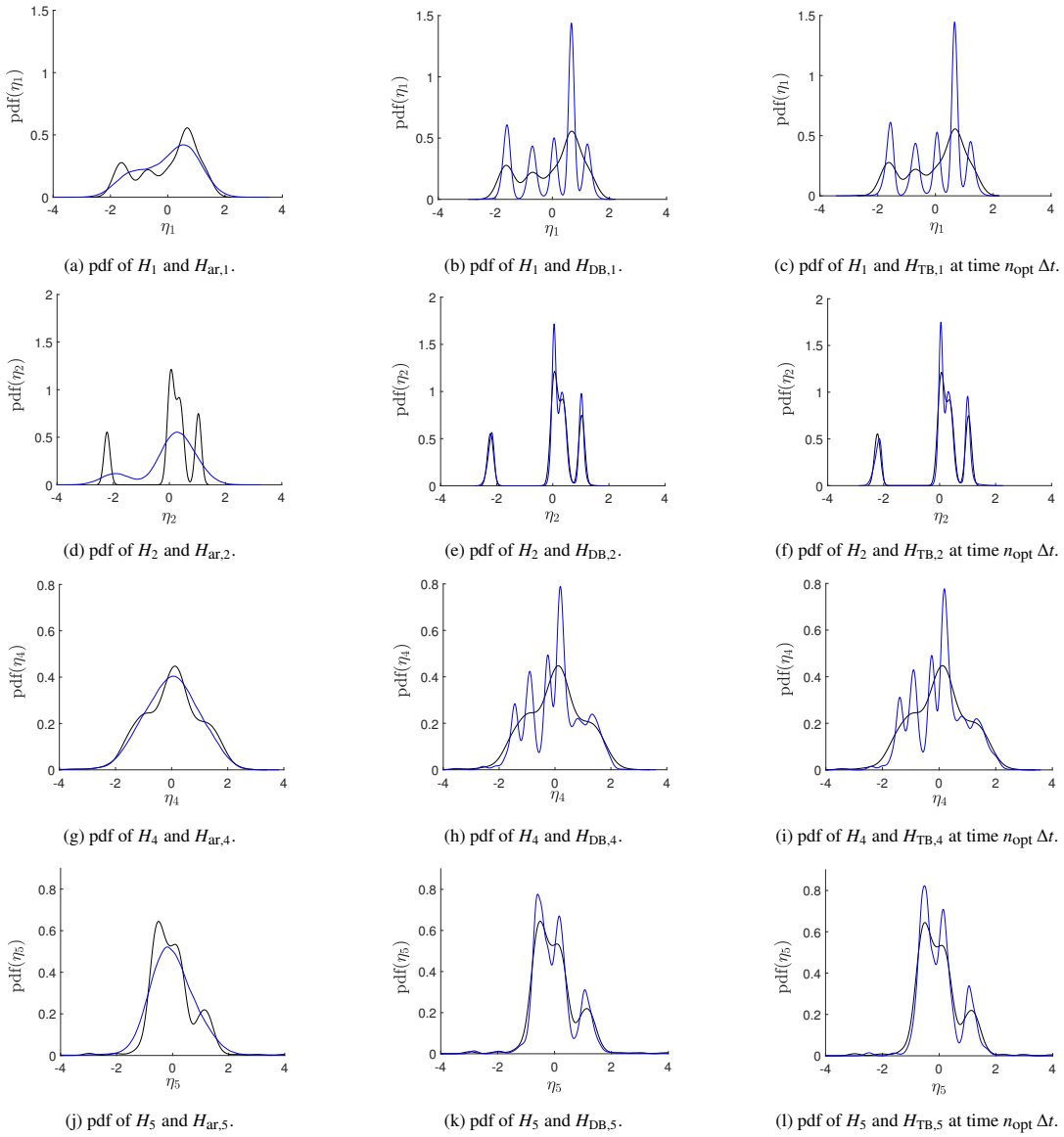
- The eigenvalues of matrix $[K_{\text{DM}}]$ and those of the of symmetrized matrix $[\tilde{K}(n\Delta t)]$ are shown in Fig. 10a.

- The probability-measure concentration using the d^2/ν-criterion is shown in Fig. 10b. For the learning without PLoM, the d^2-concentration is 0.951, which shows that the concentration is lost, and for the PLoM with the RODM, the concentration is 0.0091, which shows that the concentration is preserved.

- The other criterion of the probability-measure concentration is given by Kullback-Leibler divergence measure, shown in Fig. 10c. For the learning without PLoM, Kullback-Leibler divergence is 0.4594, and for the PLoM with the RODM, Kullback-Leibler divergence is 3.4473. Comparing Figs. 10b and 10c shows, similarly to Application 1 that the two criteria are consistent and give the same analysis of the concentration.

- The angle between the subspaces spanned by RODB and ROTB($n\Delta t$) is displayed in Fig. 10d. It can be seen that, for the optimal time $5\,\Delta t$, the angle is 17.1°, which is significant, although less than the optimal angle of Application 1. This shows that the two bases are different while the d^2-concentration remains small at 0.0153.

- The entropy of pdf $p_{\text{TB}}(\cdot\,; n\Delta t)$ is given in Fig. 6e.



(a) pdf of H_1 and $H_{\text{ar},1}$. (b) pdf of H_1 and $H_{\text{DB},1}$. (c) pdf of H_1 and $H_{\text{TB},1}$ at time $n_{\text{opt}}\,\Delta t$.

(d) pdf of H_2 and $H_{\text{ar},2}$. (e) pdf of H_2 and $H_{\text{DB},2}$. (f) pdf of H_2 and $H_{\text{TB},2}$ at time $n_{\text{opt}}\,\Delta t$.

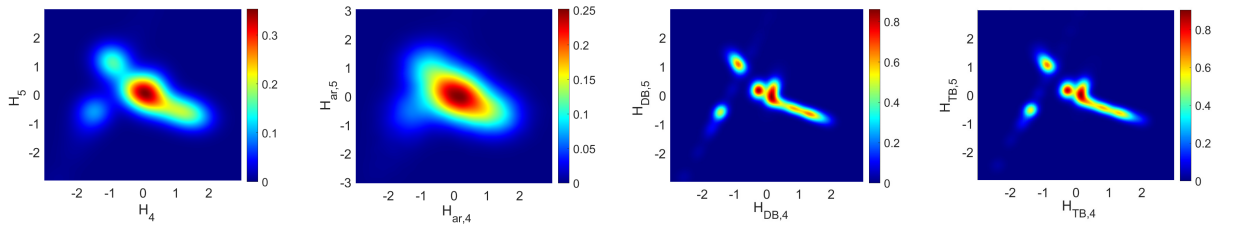
(g) pdf of H_4 and $H_{\text{ar},4}$. (h) pdf of H_4 and $H_{\text{DB},4}$. (i) pdf of H_4 and $H_{\text{TB},4}$ at time $n_{\text{opt}}\,\Delta t$.

(j) pdf of H_5 and $H_{\text{ar},5}$. (k) pdf of H_5 and $H_{\text{DB},5}$. (l) pdf of H_5 and $H_{\text{TB},5}$ at time $n_{\text{opt}}\,\Delta t$.

Figure 3: Application 1. Probability density function (pdf) of components 1, 2, 4, and 5 for $\boldsymbol{H}$ estimated with the n_d realizations of the training dataset (thin black line) and pdf estimated with n_{ar} learned realizations (thick blue line), for $\boldsymbol{H}_{\text{ar}}$ using MCMC without PLoM (a,d,g,j), for $\boldsymbol{H}_{\text{DB}}$ using PLoM with RODB (b,e,h,k), and for $\boldsymbol{H}_{\text{TB}}$ using PLoM with ROTB($n_{\text{opt}}\Delta t$) (c, f, i, l).

- The normalized mutual information (MI) of the pdfs $p_{\boldsymbol{H}}$ and $p_{\text{TB}}(\cdot\,; n\Delta t)$ is shown in Fig. 10f. This figure shows that the optimal value of n is $n_{\text{opt}} = 5$. Unlike Application 1, the normalized mutual information presents a local minimum, which is also a global minimum over the admissible set $\mathscr{C}_N$. For the non-normalized estimation of the mutual information, we have $\hat{I}(\boldsymbol{H}) = 3.4994$, $\hat{I}(\boldsymbol{H}_{\text{TB}}\,; n_{\text{opt}}\Delta t) = 5.8763$, and $\hat{I}(\boldsymbol{H}_{\text{DM}}) = 5.9382$. For the normalized one, we have $\hat{I}_{\text{norm}}(\boldsymbol{H}) = \hat{I}_{\text{norm}}(\boldsymbol{H}_{\text{TB}}\,; n_{\text{opt}}\Delta t) = 0.3441$, and $\hat{I}_{\text{norm}}(\boldsymbol{H}_{\text{DM}}) = 0.3477$.

(v) As for Application 1, examination of these figures shows that traditional learning without PLoM gives poor results compared to PLoM, which allows the concentration to be preserved and properly learns the geometry of the probability measure support. We also see that PLoM with the optimal ROTB provides an improvement in learning compared to PLoM with RODM. However, this improvement is less than in the case of Application 1 for which the data are much more heterogeneous (in correlation with the geometric complexity of the probability-measure support). Nevertheless,



(a) joint pdf of H_4 with H_5. (b) joint pdf of $H_{\text{ar},4}$ with $H_{\text{ar},5}$. (c) joint pdf of $H_{\text{DB},4}$ with $H_{\text{DB},5}$. (d) joint pdf of $H_{\text{TB},4}$ with $H_{\text{TB},5}$ at time $n_{\text{opt}}\,\Delta t$.

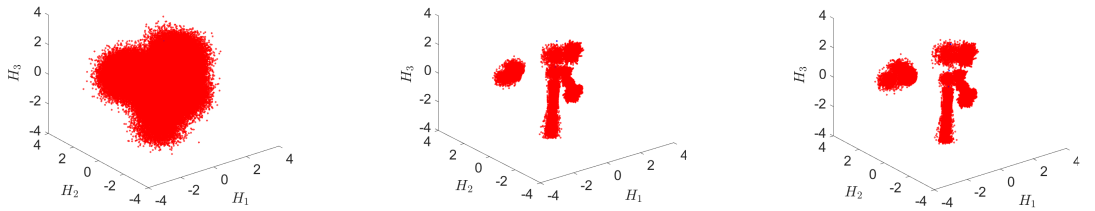
Figure 4: Application 1. Joint probability density function of components 4 with 5 of $\boldsymbol{H}$ estimated with the n_d realizations of the training dataset (a) and estimated with n_{ar} learned realizations, for $\boldsymbol{H}_{\text{ar}}$ using MCMC without PLoM (b), for $\boldsymbol{H}_{\text{DB}}$ using PLoM with RODB (c), and for $\boldsymbol{H}_{\text{TB}}$ using PLoM with ROTB($n_{\text{opt}}\Delta t$) (d).

PLoM with the optimal ROTB is an improvement over PLoM with RODB and consequently, should improve the estimates of conditional statistics thanks to better learning of the joint probability measure.

8.8. Results for Application 3

(i) Figure 11 displays the graphs of the probability density function (pdf) of components 1, 6, 25, and 40 for $\boldsymbol{H}$ estimated with the n_d realizations of the training dataset, and the pdf estimated with n_{ar} learned realizations, for $\boldsymbol{H}_{\text{ar}}$ using MCMC without PLoM (a,d,g,j), for $\boldsymbol{H}_{\text{DB}}$ using PLoM with RODB (b, e, h, k), and for $\boldsymbol{H}_{\text{TB}}$ using PLoM with ROTB($n_{\text{opt}}\Delta t$) (c, f, i, l).
(ii) Figure 12 shows the joint probability density function of components 25 and 40 of $\boldsymbol{H}$ estimated with the n_d realizations of the training dataset (a) and estimated with n_{ar} learned realizations for $\boldsymbol{H}_{\text{ar}}$ using MCMC without PLoM (b), for $\boldsymbol{H}_{\text{DB}}$ using PLoM with RODB (c), and for $\boldsymbol{H}_{\text{TB}}$ using PLoM with ROTB($n_{\text{opt}}\Delta t$) (d).
(iii) In Fig. 13, the clouds of n_{ar} points corresponding to n_{ar} learned realizations can be seen for components 1, 6, 12 (a,b,c) and components 12, 25, 40 (d,e,f). These are shown for $\boldsymbol{H}_{\text{ar}}$ using MCMC without PLoM (a,d), for $\boldsymbol{H}_{\text{DB}}$ using PLoM with RODB (b,e), and for $\boldsymbol{H}_{\text{TB}}$ using PLoM with ROTB($n_{\text{opt}}\Delta t$) (c,f).
(iv) Figure 14 plots the functions that characterize the reduced-order transient basis ROTB($n\Delta t$) as a function of time $n\Delta t$:

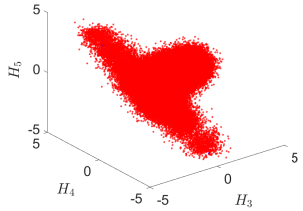
- The eigenvalues of matrix $[K_{\text{DM}}]$ and those of the of symmetrized matrix $[\tilde{K}(n\Delta t)]$ are shown in Fig. 14a.
- The probability-measure concentration using the d^2/ν-criterion is shown in Fig. 14b. For the learning without PLoM, the d^2-concentration is 0.574, which shows that the concentration is lost, and for the PLoM with the RODM, the concentration is 0.067, which shows that the concentration is preserved.
- The other criterion of the probability-measure concentration is given by Kullback-Leibler divergence measure, shown in Fig. 14c. For the learning without PLoM, Kullback-Leibler divergence is 14.35, and for the PLoM with the RODM, Kullback-Leibler divergence is 5.72. Comparing Figs. 14b and 14c shows, similarly to Applications 1 and 2 that the two criteria are consistent and give the same analysis of the concentration.
- The angle between the subspaces spanned by RODB and ROTB($n\Delta t$) is displayed in Fig. 14d. It can be seen that, for the optimal time $9\,\Delta t$, the angle is 9.7°, which is significant, although less than the optimal angle of Applications 1 and 2. This shows that the two bases are different while the d 2-concentration remains small at 0.078.
- The entropy of pdf $p_{\text{TB}}(\cdot\,; n\Delta t)$ is given in Fig. 14e.
- The normalized mutual information (MI) of the pdfs $p_{\boldsymbol{H}}$ and $p_{\text{TB}}(\cdot\,; n\Delta t)$ is shown in Fig. 14f. This figure shows that the optimal value of n is $n_{\text{opt}} = 9$. The behavior of the normalized mutual information is similar to that of Application 1 and does not present a local minimum as in Application 2. For the non-normalized estimation of the mutual information, we have $\hat{I}(\boldsymbol{H}) = 24.167$, $\hat{I}(\boldsymbol{H}_{\text{TB}}\,; n_{\text{opt}}\Delta t) = 25.115$, and $\hat{I}(\boldsymbol{H}_{\text{DM}}) = 25.673$. For the normalized one, we have $\hat{I}_{\text{norm}}(\boldsymbol{H}) = \hat{I}_{\text{norm}}(\boldsymbol{H}_{\text{TB}}\,; n_{\text{opt}}\Delta t) = 0.1418$ and $\hat{I}_{\text{norm}}(\boldsymbol{H}_{\text{DM}}) = 0.1450$.



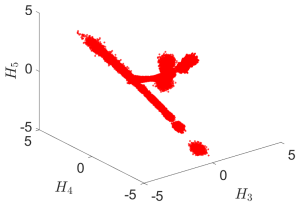
(a) clouds for $(H_{\mathrm{ar},1}, H_{\mathrm{ar},2}, H_{\mathrm{ar},3})$.

(b) clouds for $(H_{\mathrm{DB},1}, H_{\mathrm{DB},2}, H_{\mathrm{DB},3})$.

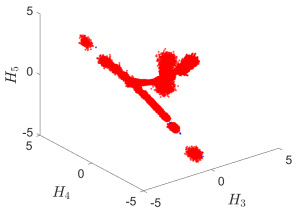
(c) clouds for $(H_{\mathrm{TB},1}, H_{\mathrm{TB},2}, H_{\mathrm{TB},3})$ at time $n_{\mathrm{opt}}\,\Delta t$.



(d) clouds for $(H_{\mathrm{ar},3}, H_{\mathrm{ar},4}, H_{\mathrm{ar},5})$.



(e) clouds for $(H_{\mathrm{DB},3}, H_{\mathrm{DB},4}, H_{\mathrm{DB},5})$.



(f) clouds for $(H_{\mathrm{TB},3}, H_{\mathrm{TB},4}, H_{\mathrm{TB},5})$ at time $n_{\mathrm{opt}}\,\Delta t$.

Figure 5: Application 1. Clouds of n_{ar} points corresponding to n_{ar} learned realizations, for components 1, 2, 3 (a,b,c) and components 3, 4, 5 (d,e,f), for $\boldsymbol{H}_{\mathrm{ar}}$ using MCMC without PLoM (a,d), for $\boldsymbol{H}_{\mathrm{DB}}$ using PLoM with RODB (b,e), and for $\boldsymbol{H}_{\mathrm{TB}}$ using PLoM with ROTB($n_{\mathrm{opt}}\Delta t$) (c,f).

(v) As for Applications 1 and 2, examination of these figures shows that traditional learning without PLoM gives poor results compared to PLoM, which allows the concentration to be preserved and properly learns the geometry of the probability measure support. We also see that PLoM with the optimal ROTB provides an improvement in learning compared to PLoM with RODM. However, this improvement is less than in the case of Applications 1 and 2. For this application, relative to a relatively high dimension of $\nu = 45$, the data are more homogeneous than for the other applications (in correlation with the geometric complexity of the probability-measure support). Nevertheless, PLoM with the optimal ROTB is an improvement over PLoM with RODB and, consequently, should improve the estimates of conditional statistics thanks to better learning of the joint probability measure.

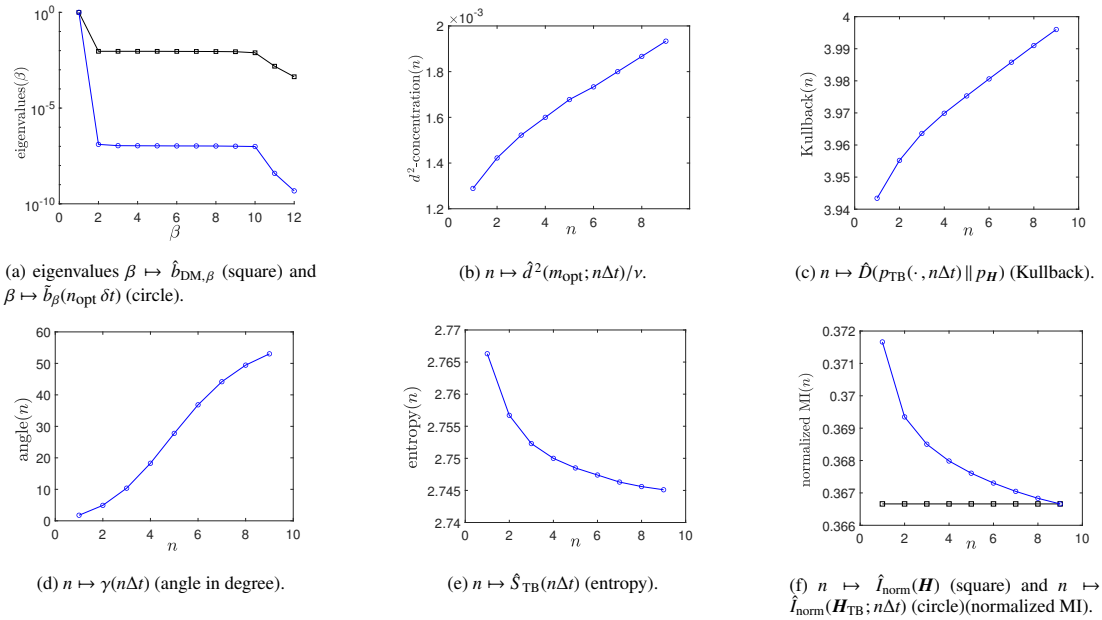
9. Conclusion

In this paper, we have presented the theoretical elements of constructing a time-dependent anisotropic kernel, which allows us to create a data projection basis for PLoM. This basis serves as an alternative to the DMAPS basis built using a time-independent isotropic kernel used by PLoM. We have demonstrated that an optimal time can be determined to obtain an optimal transient basis, best respecting the statistical dependence between the components for the learned joint probability measure.

The proposed theory has been developed to improve PLoM in cases of highly heterogeneous data. The improvement of the learned joint probability measure is quantified by estimating an objective criterion from information theory, namely the mutual information, which we have normalized relative to the number of realizations using entropy.

This theory is consistent in the sense that, for a time close to the initial time, the DMAPS basis constructed with the time-independent isotropic kernel coincides with the transient basis constructed with the time-dependent anisotropic kernel. Thus, we can characterize the difference between the two bases by the angle of the vector subspaces they generate.

The theory is illustrated through three applications with decreasing levels of data heterogeneity. The three applications confirm that PLoM with the DMAPS basis (time-independent isotropic kernel) always results in learning that preserves the concentration of the measure, unlike the classic MCMC approach. The applications show that it is possible to improve the learned joint probability measure with the transient anisotropic kernel, which a priori allows for better estimates of conditional statistics.



(a) eigenvalues $\beta \mapsto \hat{b}_{\mathrm{DM},\beta}$ (square) and $\beta \mapsto \tilde{b}_\beta(n_{\mathrm{opt}}\,\delta t)$ (circle).

(b) $n \mapsto \hat{d}^2(m_{\mathrm{opt}}; n\Delta t)/\nu$.

(c) $n \mapsto \hat{D}(p_{\mathrm{TB}}(\cdot\,, n\Delta t)\,\|\,p_{\boldsymbol{H}})$ (Kullback).

(d) $n \mapsto \gamma(n\Delta t)$ (angle in degree).

(e) $n \mapsto \hat{S}_{\mathrm{TB}}(n\Delta t)$ (entropy).

(f) $n \mapsto \hat{I}_{\mathrm{norm}}(\boldsymbol{H})$ (square) and $n \mapsto \hat{I}_{\mathrm{norm}}(\boldsymbol{H}_{\mathrm{TB}}; n\Delta t)$ (circle)(normalized MI).

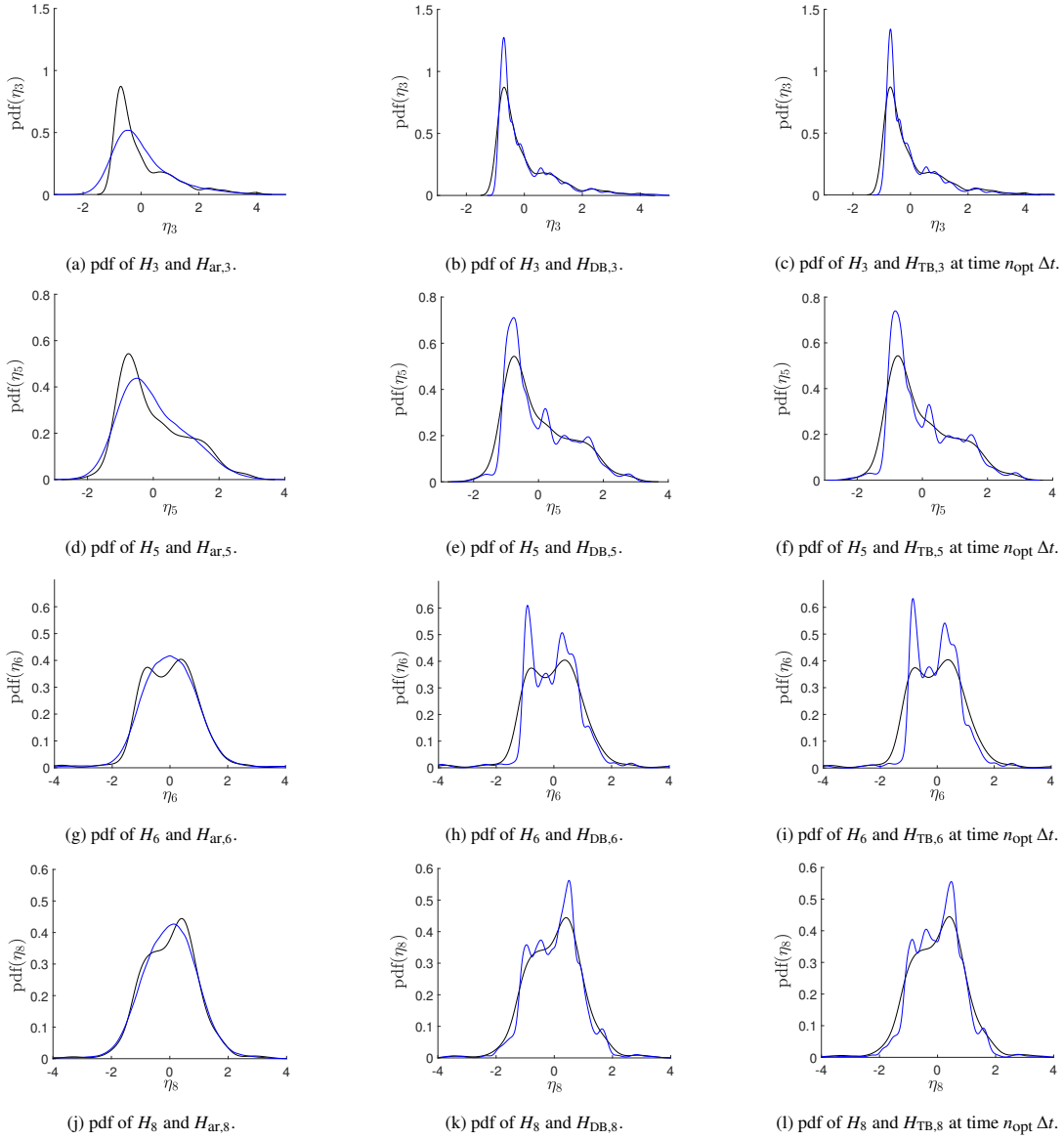
Figure 6: Application 1. Functions characterizing the reduced-order transient basis ROTB($n\Delta t$) as a function of time $n\Delta t$: eigenvalues of $[K_{\mathrm{DM}}]$ and of symmetrized $[\tilde{K}(n\Delta t)]$ (a); probability-measure concentration with d^2/ν-criterion (b) and with Kullback-Leibler divergence criterion (c); angle between the subspaces spanned by RODB and ROTB($n\Delta t$) (d); entropy of pdf $p_{\mathrm{TB}}(\cdot\,; n\Delta t)$ (e); normalized mutual information (MI) of pdf $p_{\boldsymbol{H}}$ and $p_{\mathrm{TB}}(\cdot\,; n\Delta t)$ (f).

Acknowledgments

The authors acknowledge partial funding from DOE SciDAC FASTMath Institute, and an ONR MURI on Modeling Turbulence and Chemistry in High Speed Reactive Flows.

Appendix A. Overview of the probabilistic learning on manifolds (PLoM) algorithm and its parameterization

The PLoM approach [1, 30, 32], which has specifically been developed for small data (as opposed to big data) starts from a training dataset $\mathcal{D}_d$ made up of a relatively small number n_d of points. It is assumed that $\mathcal{D}_d$ is generated with an underlying stochastic manifold related to a $\mathbb{R}^{n_x}$-valued random variable $\boldsymbol{X} = (\boldsymbol{Q}, \boldsymbol{W})$, defined on a probability space $(\Theta, \mathcal{T}, \mathcal{P})$, in which $\boldsymbol{Q}$ is the quantity of interest that is a $\mathbb{R}^{n_q}$-random variable, where $\boldsymbol{W}$ is the control parameter that is a $\mathbb{R}^{n_w}$-random variable, and where $n_x = n_q + n_w$. Another $\mathbb{R}^{n_u}$- valued random variable $\boldsymbol{U}$ defined on $(\Theta, \mathcal{T}, \mathcal{P})$ is also be considered, which is an uncontrolled parameter and/or a noise. Random variable $\boldsymbol{Q}$ is assumed to be written as $\boldsymbol{Q} = \boldsymbol{f}(\boldsymbol{U}, \boldsymbol{W})$ in which the measurable mapping $\boldsymbol{f}$ is not explicitly known. The joint probability distribution $P_{\boldsymbol{W},\boldsymbol{U}}(d\boldsymbol{w}, d\boldsymbol{u})$ of $\boldsymbol{W}$ and $\boldsymbol{U}$ is assumed to be given. The non-Gaussian probability measure $P_{\boldsymbol{X}}(\boldsymbol{x}) = P_{\boldsymbol{Q},\boldsymbol{W}}(d\boldsymbol{q}, d\boldsymbol{w})$ of $\boldsymbol{X} = (\boldsymbol{Q}, \boldsymbol{W})$ is concentrated in a region of $\mathbb{R}^{n_x}$ for which the only available information is the cloud of the points of training dataset $\mathcal{D}_d$. The PLoM method makes it possible to generate the learned dataset $\mathcal{D}_{\mathrm{ar}}$ for $\boldsymbol{X}$ whose $n_{\mathrm{ar}} \gg n_d$ points (learned realizations) are generated by the non-Gaussian probability measure that is estimated using the training dataset. The concentration of the probability measure is preserved thanks to the use of a diffusion-maps basis that allows to enrich the available information from the training dataset. The training dataset $\mathcal{D}_d$ is made up of the n_d independent realizations $\boldsymbol{x}_d^j = (\boldsymbol{q}_d^j, \boldsymbol{w}_d^j)$ in $\mathbb{R}^{n_x} = \mathbb{R}^{n_q} \times \mathbb{R}^{n_w}$ for $j \in \{1, \ldots, n_d\}$ of random variable $\boldsymbol{X} = (\boldsymbol{Q}, \boldsymbol{W})$. The PLoM method allows for generating the learned dataset $\mathcal{D}_{\mathrm{ar}}$ made up of $n_{\mathrm{ar}} \gg n_d$ learned realizations $\{\boldsymbol{x}_{\mathrm{ar}}^\ell, \ell = 1, \ldots, n_{\mathrm{ar}}\}$ of random vector $\boldsymbol{X}$. As soon as the learned dataset has been constructed, the learned realizations for $\boldsymbol{Q}$ and $\boldsymbol{W}$ can be extracted as $(\boldsymbol{q}_{\mathrm{ar}}^\ell, \boldsymbol{w}_{\mathrm{ar}}^\ell) = \boldsymbol{x}_{\mathrm{ar}}^\ell$ for $\ell = 1, \ldots, n_{\mathrm{ar}}$. Using the learned dataset $\mathcal{D}_{\mathrm{ar}}$, PLoM allows for carrying out any conditional statistics such as $\boldsymbol{w} \mapsto E\{\boldsymbol{\xi}(\boldsymbol{Q})|\boldsymbol{W} = \boldsymbol{w}\}$ from $\mathscr{C}_w$ in $\mathbb{R}^{n_\xi}$, in which $\boldsymbol{\xi}$ is a given measurable mapping from $\mathbb{R}^{n_q}$ into $\mathbb{R}^{n_\xi}$, that is to say to construct statistical surrogate models (metamodels) in a probabilistic framework.



(a) pdf of H_3 and $H_{\mathrm{ar},3}$. (b) pdf of H_3 and $H_{\mathrm{DB},3}$. (c) pdf of H_3 and $H_{\mathrm{TB},3}$ at time $n_{\mathrm{opt}}\,\Delta t$.

(d) pdf of H_5 and $H_{\mathrm{ar},5}$. (e) pdf of H_5 and $H_{\mathrm{DB},5}$. (f) pdf of H_5 and $H_{\mathrm{TB},5}$ at time $n_{\mathrm{opt}}\,\Delta t$.

(g) pdf of H_6 and $H_{\mathrm{ar},6}$. (h) pdf of H_6 and $H_{\mathrm{DB},6}$. (i) pdf of H_6 and $H_{\mathrm{TB},6}$ at time $n_{\mathrm{opt}}\,\Delta t$.

(j) pdf of H_8 and $H_{\mathrm{ar},8}$. (k) pdf of H_8 and $H_{\mathrm{DB},8}$. (l) pdf of H_8 and $H_{\mathrm{TB},8}$ at time $n_{\mathrm{opt}}\,\Delta t$.

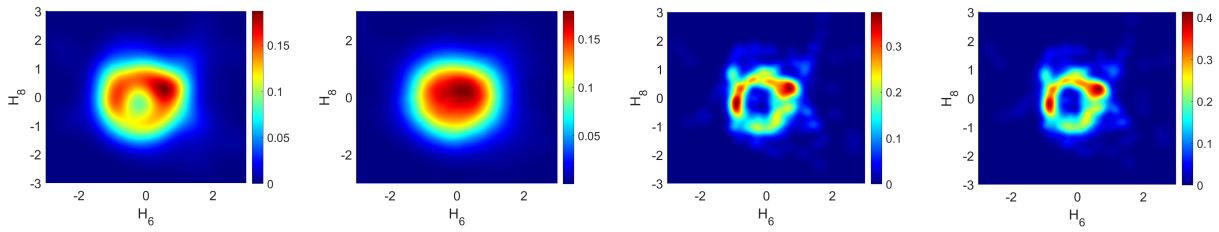
Figure 7: Application 2. Probability density function (pdf) of components 3, 5, 6, and 8 for $\boldsymbol{H}$ estimated with the n_d realizations of the training dataset (thin black line) and pdf estimated with n_{ar} learned realizations (thick blue line), for $\boldsymbol{H}_{\mathrm{ar}}$ using MCMC without PLoM (a,d,g,j), for $\boldsymbol{H}_{\mathrm{DB}}$ using PLoM with RODB (b,e,h,k), and for $\boldsymbol{H}_{\mathrm{TB}}$ using PLoM with ROTB($n_{\mathrm{opt}}\Delta t$) (c, f, i, l).

Appendix A.1. Reduced representation

The n_d independent realizations $\{\boldsymbol{x}_d^j,\, j = 1,\ldots,n_d\}$ are represented by the matrix $[x_d] = [\boldsymbol{x}_d^1 \ldots \boldsymbol{x}_d^{n_d}]$ in $\mathbb{M}_{n_x,n_d}$. Let $[\boldsymbol{X}] = [\boldsymbol{X}^1,\ldots,\boldsymbol{X}^{n_d}]$ be the random matrix with values in $\mathbb{M}_{n_x,n_d}$, whose columns are n_d independent copies of random vector $\boldsymbol{X}$. Using the PCA of $\boldsymbol{X}$, random matrix $[\boldsymbol{X}]$ is written as,

$$[\boldsymbol{X}] = [\underline{x}] + [\varphi]\,[\zeta]^{1/2}\,[\boldsymbol{H}]\,, \tag{A.1}$$

in which $[\boldsymbol{H}] = [\boldsymbol{H}^1,\ldots,\boldsymbol{H}^{n_d}]$ is a $\mathbb{M}_{\nu,n_d}$-valued random matrix, where $\nu \le n_x$, and where $[\zeta]$ is the $(\nu \times \nu)$ diagonal matrix of the ν positive eigenvalues of the empirical estimate of the covariance matrix of $\boldsymbol{X}$. The $(n_x \times \nu)$ matrix $[\varphi]$ is made up of the associated eigenvectors such $[\varphi]^T\,[\varphi] = [I_\nu]$. The matrix $[\underline{x}]$ in $\mathbb{M}_{n_x,n_d}$ has identical columns, each one being equal to the empirical estimate $\underline{\boldsymbol{x}} \in \mathbb{R}^{n_x}$ of the mean value of random vector $\boldsymbol{X}$. The columns of $[\boldsymbol{H}]$ are



(a) joint pdf of H_6 with H_8. (b) joint pdf of $H_{\text{ar},6}$ with $H_{\text{ar},8}$. (c) joint pdf of $H_{\text{DB},6}$ with $H_{\text{DB},8}$. (d) joint pdf of $H_{\text{TB},6}$ with $H_{\text{TB},8}$ at time $n_{\text{opt}}\,\Delta t$.

Figure 8: Application 2. Joint probability density function of components 6 with 8 of $\boldsymbol{H}$ estimated with the n_d realizations of the training dataset (a) and estimated with n_{ar} learned realizations, for $\boldsymbol{H}_{\text{ar}}$ using MCMC without PLoM (b), for $\boldsymbol{H}_{\text{DB}}$ using PLoM with RODB (c), and for $\boldsymbol{H}_{\text{TB}}$ using PLoM with ROTB($n_{\text{opt}}\Delta t$) (d).

n_d independent copies of a random vector $\boldsymbol{H}$ with values in $\mathbb{R}^\nu$. The realization $[\eta_d] = [\boldsymbol{\eta}^1 \dots \boldsymbol{\eta}^{n_d}] \in \mathbb{M}_{\nu,n_d}$ of $[\boldsymbol{H}]$ is computed by $[\eta_d] = [\zeta]^{-1/2}[\varphi]^T\,([x_d] - [\underline{x}])$. The value ν is classically calculated in order that the L^2- error function $\nu \mapsto \text{err}_{\boldsymbol{X}}(\nu)$ defined by

$$\text{err}_{\boldsymbol{X}}(\nu) = 1 - \frac{\sum_{\alpha=1}^{\nu} \zeta_\alpha}{E\{\|\boldsymbol{X}\|^2\}}\,, \tag{A.2}$$

be smaller than ε_{PCA}. If $\nu < n_x$, then there is a statistical reduction.

Appendix A.2. Construction of a reduced-order diffusion-maps basis (RODB) and reduced-order transient basis (ROTB($n\Delta t$))

In this section, we begin with the construction of the RODB that is the basis initially used in the PLoM algorithm (see [1]). Concerning the construction of the ROTB($n\Delta t$), we refer the read to 7.2-(ii).

(i) Construction of RODB. This construction corresponds to the one initailly proposed in the PLoM algorithm. For preserving the concentration of the learned realizations in the region in which the points of the training dataset are concentrated, the PLoM relies on the diffusion-maps method [3, 99]. This is an algebraic basis of vector space $\mathbb{R}^{n_d}$, which is constructed using the diffusion maps. Let $[\mathcal{K}_{\text{DM}}]$ and $[B]$ be the matrices such that, for all i and j in $\{1,\dots,n_d\}$, $[\mathcal{K}_{\text{DM}}]_{ij} = \exp\{-(4\,\varepsilon_{\text{DM}})^{-1}\|\boldsymbol{\eta}^i - \boldsymbol{\eta}^j\|^2\}$ and $[B]_{ij} = \delta_{ij}\exp\{-(4\,\varepsilon_{\text{DM}})^{-1}\|\boldsymbol{\eta}^i - \boldsymbol{\eta}^j\|^2\}$, in which $\varepsilon_{\text{DM}} > 0$ is a smoothing parameter. The eigenvalues $b_{\text{DM},1},\dots,b_{\text{DM},n_d}$ and the associated eigenvectors $\boldsymbol{g}^1_{\text{DM}},\dots,\boldsymbol{g}^{n_d}_{\text{DM}}$ of the right-eigenvalue problem $[B]^{-1}[\mathcal{K}_{\text{DM}}]\,\boldsymbol{g}^\beta_{\text{DM}} = b_{\text{DM},\beta}\,\boldsymbol{g}^\beta_{\text{DM}}$ are such that $1 = b_{\text{DM},1} > b_{\text{DM},2} \geq \dots \geq b_{\text{DM},n_d}$ and are computed by solving the eigenvalue problem $[B]^{-1/2}[\mathcal{K}_{\text{DM}}]\,[B]^{-1/2}\,\boldsymbol{\phi}^\beta = b_{\text{DM},\beta}\,\boldsymbol{\phi}^\beta$ with the normalization $\langle\boldsymbol{\phi}^\beta,\boldsymbol{\phi}^{\beta'}\rangle = \delta_{\beta\beta'}$, and $\boldsymbol{g}^\beta_{\text{DM}} = [B]^{-1/2}\boldsymbol{\phi}^\beta$. The eigenvector $\boldsymbol{g}^1_{\text{DM}}$ associated with $b_{\text{DM},1} = 1$ is a constant vector. The diffusion-maps basis $\{\boldsymbol{g}^1_{\text{DM}},\dots,\boldsymbol{g}^\alpha_{\text{DM}},\dots,\boldsymbol{g}^{n_d}_{\text{DM}}\}$ is a vector basis of $\mathbb{R}^{n_d}$. For a given integer $m < n_d$, the reduced-order diffusion-maps basis of order m is defined as the family $\{\boldsymbol{g}^1_{\text{DM}},\dots,\boldsymbol{g}^m_{\text{DM}}\}$. This basis depends on two parameters, ε_{DM} and m, which have to be identified. As explained in [32], the optimal value m_{opt} of m_{DM} is chosen as $m_{\text{opt}} = \nu + 1$, and the optimal value ε_{opt} of ε_{DM} is such that

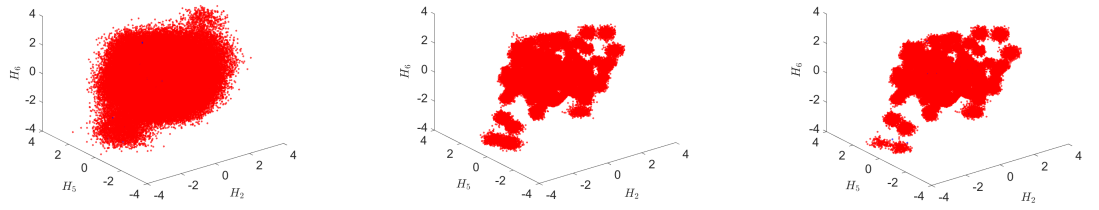
$$1 = b_{\text{DM},1} > b_{\text{DM},2} \simeq \dots \simeq b_{\text{DM},m_{\text{opt}}} \gg b_{\text{DM},m_{\text{opt}}+1} \geq \dots \geq b_{\text{DM},n_d} > 0\,, \tag{A.3}$$

with the jump amplitude $J_{\text{DM}} = b_{\text{DM},m_{\text{opt}}+1}/b_{\text{DM},m_{\text{opt}}}$, which is $J_{\text{DM}} = 0.1$ (following [30]), but which can also be chosen in the interval $[0.1\,,0.5]$. Consequently, the RODB is defined for $m = m_{\text{opt}}$ and is represented by the matrix

$$[g_{\text{DM}}] = [\boldsymbol{g}^1_{\text{DM}} \dots \boldsymbol{g}^{m_{\text{opt}}}_{\text{DM}}] \in \mathbb{M}_{n_d,m_{\text{opt}}}\,. \tag{A.4}$$

(ii) Construction of ROTB($n\Delta t$). Because PLoM will also use the reduced-order transient basis to quantify its efficiency relative to the reduced-order DMAPS basis, we introduce this basis in this Appendix. For n fixed in $\{1,\dots,N\}$, the reduced-order transient basis, ROTB($n\Delta t$), is represented by the matrix

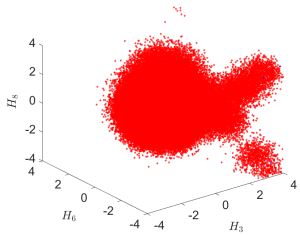
$$[g(n\Delta t)] = [\boldsymbol{g}^1(n\Delta t) \dots \boldsymbol{g}^{m_{\text{opt}}}(n\Delta t)] \in \mathbb{M}_{n_d,m_{\text{opt}}}\,, \tag{A.5}$$



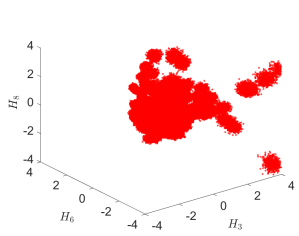
(a) clouds for $(H_{\mathrm{ar},2}, H_{\mathrm{ar},5}, H_{\mathrm{ar},6})$.

(b) clouds for $(H_{\mathrm{DB},2}, H_{\mathrm{DB},5}, H_{\mathrm{DB},6})$.

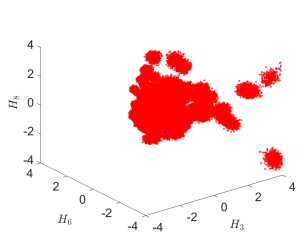
(c) clouds for $(H_{\mathrm{TB},2}, H_{\mathrm{TB},5}, H_{\mathrm{TB},6})$ at time $n_{\mathrm{opt}}\,\Delta t$.



(d) clouds for $(H_{\mathrm{ar},3}, H_{\mathrm{ar},6}, H_{\mathrm{ar},8})$.



(e) clouds for $(H_{\mathrm{DB},3}, H_{\mathrm{DB},6}, H_{\mathrm{DB},8})$.



(f) clouds for $(H_{\mathrm{TB},3}, H_{\mathrm{TB},6}, H_{\mathrm{TB},8})$ at time $n_{\mathrm{opt}}\,\Delta t$.

Figure 9: Application 2. Clouds of n_{ar} points corresponding to n_{ar} learned realizations, for components 2, 5, 6 (a,b,c) and components 3, 6, 8 (d,e,f), for $\boldsymbol{H}_{\mathrm{ar}}$ using MCMC without PLoM (a,d), for $\boldsymbol{H}_{\mathrm{DB}}$ using PLoM with RODB (b,e), and for $\boldsymbol{H}_{\mathrm{TB}}$ using PLoM with ROTB($n_{\mathrm{opt}}\Delta t$) (c,f).

in which m_{opt} is the optimal value identified in Appendix A.2-(i), and where $[g(n\Delta t)]$ is constructed in Section 7.2-(ii) (see Eq. (7.16).

(iii) Reduced-order basis for PLoM. In this Appendix, the PLoM reduced-order basis will be represented by the matrix $[g_{\mathrm{opt}}] \in \mathbb{M}_{n_d,m_{\mathrm{opt}}}$. Depending on the context of its use, this matrix will either be $[g_{\mathrm{DM}}]$, representing the reduced-order DMAPS basis (RODB) as used in the initial construction of the PLoM, or $[g(n\Delta t)]$ for a fixed n, representing the reduced-order transient basis (ROTB($n\Delta t$)) as proposed in this paper. The latter is introduced with the goal of comparing the efficiency of the two reduced-order vector bases.

Appendix A.3. Reduced-order representation of the random matrices

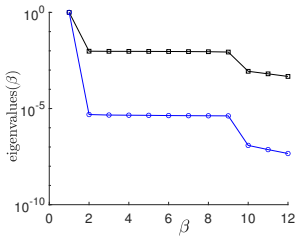
The reduced-order basis represented by matrix $[g_{\mathrm{opt}}] \in \mathbb{M}_{n_d,m_{\mathrm{opt}}}$ spans a subspace of $\mathbb{R}^{n_d}$ that characterizes, for the optimal values m_{opt} and $\varepsilon_{\mathrm{opt}}$, the local geometry structure of dataset $\{\boldsymbol{\eta}^j, j = 1, \ldots, n_d\}$. So the PLoM method introduces the $\mathbb{M}_{\nu,n_d}$-valued random matrix $[\boldsymbol{H}_{\mathrm{ar}}] = [\boldsymbol{Z}]\,[g_{\mathrm{opt}}]^T$ with $m_{\mathrm{opt}} < n_d$, corresponding to a data-reduction representation of random matrix $[\boldsymbol{H}]$, in which $[\boldsymbol{Z}]$ is a $\mathbb{M}_{\nu,m_{\mathrm{opt}}}$-valued random matrix. The MCMC generator of random matrix $[\boldsymbol{Z}]$ belongs to the class of Hamiltonian Monte Carlo methods, is explicitly described in [1], and is mathematically detailed in Theorem 6.3 of [30]. This generator allows for computing n_{MCH} realizations $\{[\boldsymbol{z}^\ell_{\mathrm{ar}}], \ell = 1, \ldots, n_{\mathrm{MCH}}\}$ of $[\boldsymbol{Z}]$ and therefore, for deducing the n_{MCH} realizations $\{[\boldsymbol{\eta}^\ell_{\mathrm{ar}}], \ell = 1, \ldots, n_{\mathrm{MCH}}\}$ of $[\boldsymbol{H}_{\mathrm{ar}}]$. The reshaping of matrix $[\boldsymbol{\eta}^\ell_{\mathrm{ar}}] \in \mathbb{M}_{\nu,n_d}$ allows for obtaining $n_{\mathrm{ar}} = n_{\mathrm{MCH}} \times n_d$ learned realizations $\{\boldsymbol{\eta}^{\ell'}_{\mathrm{ar}}, \ell' = 1, \ldots, n_{\mathrm{ar}}\}$ of $\boldsymbol{H}_{\mathrm{ar}}$. These learned realizations allow for estimating converged statistics on $\boldsymbol{H}_{\mathrm{ar}}$ and then on $\boldsymbol{X}_{\mathrm{ar}} = \underline{x} + [\varphi]\,[\zeta]^{1/2}\,\boldsymbol{H}_{\mathrm{ar}}$, such as pdf, moments, or conditional expectation of the type $E\{\boldsymbol{\xi}(\boldsymbol{Q}) \,|\, \boldsymbol{W} = \boldsymbol{w}\}$ for $\boldsymbol{w}$ given in $\mathbb{R}^{n_w}$ and for any given vector-valued function $\boldsymbol{\xi}$ defined on $\mathbb{R}^{n_q}$.

Appendix A.4. Criterion for quantifying the concentration of the probability measure of random matrix $[\boldsymbol{H}_{\mathrm{ar}}]$

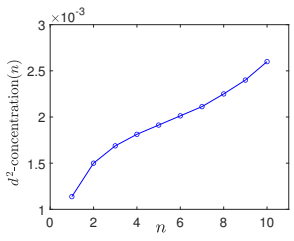
The concentration of the probability measure of random matrix $[\boldsymbol{H}_{\mathrm{ar}}]$ is defined (see [30]) by

$$d^2(m_{\mathrm{opt}}) = E\{\|[\boldsymbol{H}_{\mathrm{ar}}] - [\eta_d]\|^2\}/\|[\eta_d]\|^2\,. \tag{A.6}$$

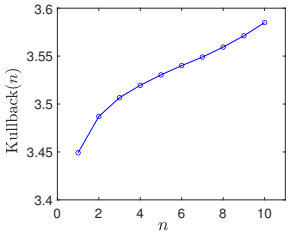
Let $\mathcal{M} = \{m_{\mathrm{opt}}, m_{\mathrm{opt}}+1, \ldots, n_d\}$ in which m_{opt} is the optimal value of m. Theorem 7.8 of [30] shows that $\min_{m\in\mathcal{M}} d^2(m) \leq 1 + m_{\mathrm{opt}}/(n_d - 1) < d^2(n_d)$, which means that the PLoM method, for $m = m_{\mathrm{opt}}$ and $[g_{\mathrm{opt}}]$ is a better method than the



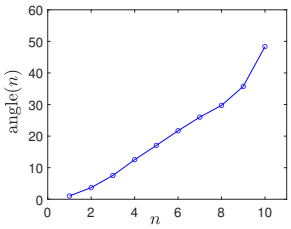
(a) eigenvalues $\beta \mapsto \hat{b}_{\mathrm{DM},\beta}$ (square) and $\beta \mapsto \tilde{b}_{\beta}(n_{\mathrm{opt}}\,\delta t)$ (circle).



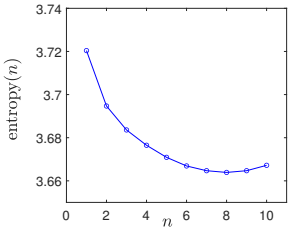
(b) $n \mapsto \hat{d}^2(m_{\mathrm{opt}}; n\Delta t)/\nu$.



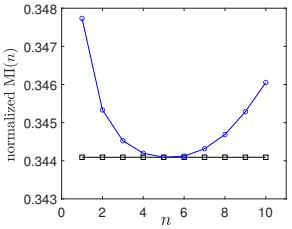
(c) $n \mapsto \hat{D}(p_{\mathrm{TB}}(\cdot\,, n\Delta t)\,\|\,p_{\boldsymbol{H}})$ (Kullback).



(d) $n \mapsto \gamma(n\Delta t)$ (angle in degree).



(e) $n \mapsto \hat{S}_{\mathrm{TB}}(n\Delta t)$ (entropy).



(f) $n \mapsto \hat{I}_{\mathrm{norm}}(\boldsymbol{H})$ (square) and $n \mapsto \hat{I}_{\mathrm{norm}}(\boldsymbol{H}_{\mathrm{TB}}; n\Delta t)$ (circle)(normalized MI).

Figure 10: Application 2. Functions characterizing the reduced-order transient basis ROTB($n\Delta t$) as a function of time $n\Delta t$: eigenvalues of $[K_{\mathrm{DM}}]$ and of symmetrized $[\tilde{K}(n\Delta t)]$ (a); measure concentration with d^2/ν-criterion (b) and with Kullback-Leibler divergence criteria (c); angle between the subspaces spanned by RODB and ROTB($n\Delta t$) (d); entropy of pdf $p_{\mathrm{TB}}(\cdot\,; n\Delta t)$ (e); normalized mutual information (MI) of pdf $p_{\boldsymbol{H}}$ and$p_{\mathrm{TB}}(\cdot\,; n\Delta t)$ (f).

usual one corresponding to $d^2(n_d) = 1 + n_d/(n_d - 1) \simeq 2$. Using the n_{MCH} realizations $\{[\boldsymbol{\eta}_{\mathrm{ar}}^{\ell}], \ell = 1, \ldots, n_{\mathrm{MCH}}\}$ of $[\boldsymbol{H}_{\mathrm{ar}}]$, we have the estimate,

$$\hat{d}^2(m_{\mathrm{opt}}) = (1/n_{\mathrm{MCH}}) \sum_{\ell=1}^{n_{\mathrm{MCH}}} \{\|[\boldsymbol{\eta}_{\mathrm{ar}}^{\ell}] - [\eta_d]\|^2\}/\|[\eta_d]\|^2 \,. \tag{A.7}$$

Appendix A.5. Generation of learned realizations $\{\boldsymbol{\eta}_{\mathrm{ar}}^{\ell'}, \ell' = 1, \ldots, n_{\mathrm{ar}}\}$ of random vector $\boldsymbol{H}_{\mathrm{ar}}$

Let $\{([\mathcal{Z}(t)], [\boldsymbol{\mathcal{Y}}(t)]),\ t \in \mathbb{R}^+\}$ be the unique asymptotic (for $t \to +\infty$) stationary diffusion stochastic process with values in $\mathbb{M}_{\nu,m_{\mathrm{opt}}} \times \mathbb{M}_{\nu,m_{\mathrm{opt}}}$, of the following reduced-order ISDE (stochastic nonlinear second-order dissipative Hamiltonian dynamic system), for $t > 0$,

$$d[\mathcal{Z}(t)] = [\boldsymbol{\mathcal{Y}}(t)]\, dt\,,$$
$$d[\boldsymbol{\mathcal{Y}}(t)] = [\mathcal{L}([\mathcal{Z}(t)])]\, dt - \frac{1}{2} f_0\, [\boldsymbol{\mathcal{Y}}(t)]\, dt + \sqrt{f_0}\, [d\boldsymbol{\mathcal{W}}^{\mathrm{wien}}(t)]\,,$$

with $[\mathcal{Z}(0)] = [\eta_d]\,[a]$ and $[\boldsymbol{\mathcal{Y}}(0)] = [\boldsymbol{\mathcal{N}}]\,[a]$, in which

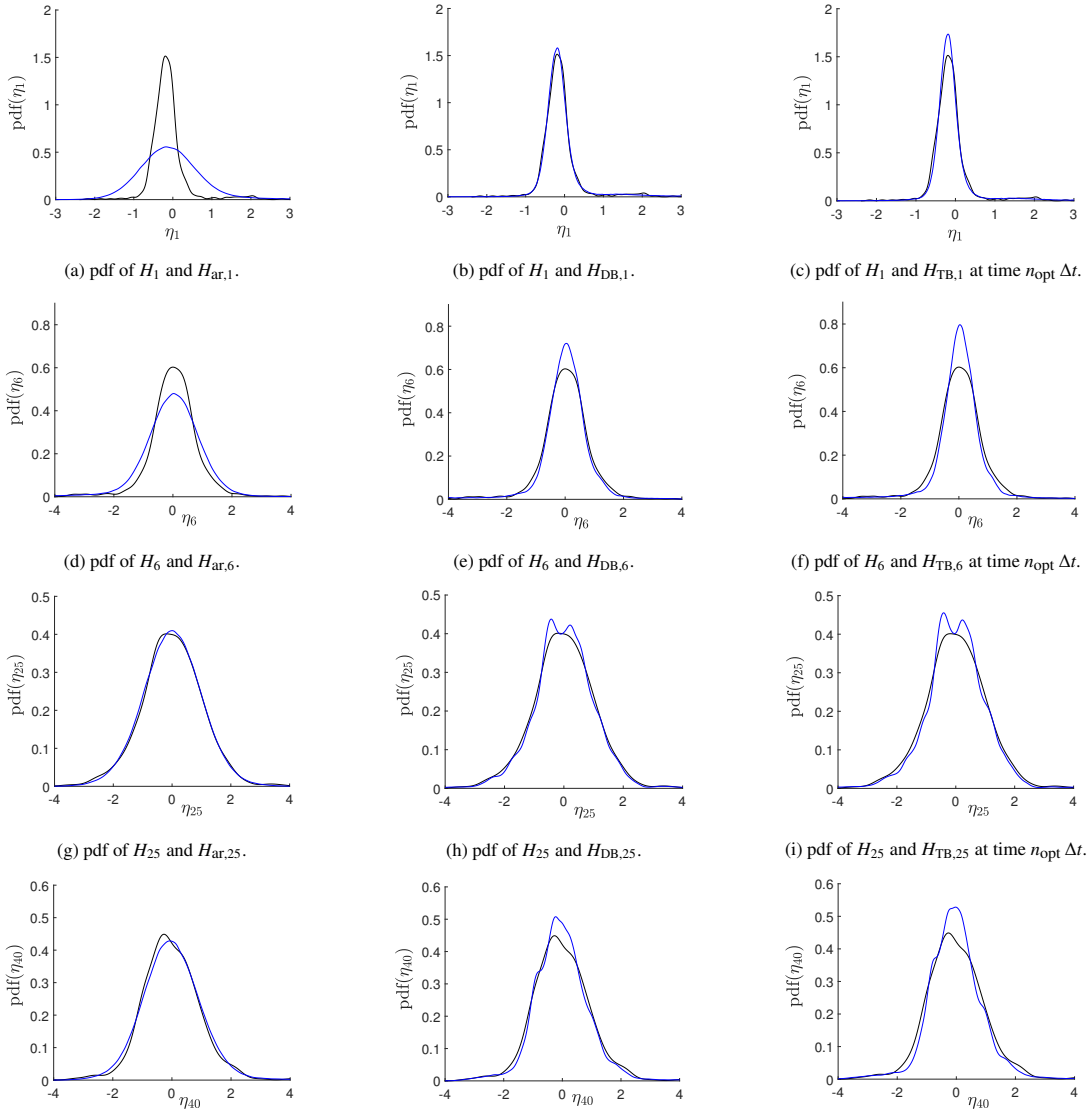
$$[a] = [g_{\mathrm{opt}}]\,([g_{\mathrm{opt}}]^T\,[g_{\mathrm{opt}}])^{-1} \in \mathbb{M}_{n_d,m_{\mathrm{opt}}}\,.$$

(1) $[\mathcal{L}([\mathcal{Z}(t)])] = [L([\mathcal{Z}(t)]\,[g_{\mathrm{opt}}]^T)]\,[a]$ is a random matrix with values in $\mathbb{M}_{\nu,m_{\mathrm{opt}}}$. For all $[u] = [\boldsymbol{u}^1 \ldots \boldsymbol{u}^{n_d}]$ in $\mathbb{M}_{\nu,n_d}$ with $\boldsymbol{u}^j = (u_1^j, \ldots, u_\nu^j)$ in $\mathbb{R}^\nu$, the matrix $[L([u])]$ in $\mathbb{M}_{\nu,n_d}$ is defined, for all $k = 1, \ldots, \nu$ and for all $j = 1, \ldots, n_d$, by

$$[L([u])]_{kj} = \frac{1}{p(\boldsymbol{u}^j)}\, \{\boldsymbol{\nabla}_{\boldsymbol{u}^j}\, p(\boldsymbol{u}^j)\}_k\,, \tag{A.8}$$

$$p(\boldsymbol{u}^j) = \frac{1}{n_d} \sum_{j'=1}^{n_d} \exp\{-\frac{1}{2\hat{s}^2} \|\frac{\hat{s}}{s}\boldsymbol{\eta}^{j'} - \boldsymbol{u}^j\|^2\}\,,$$

$$\boldsymbol{\nabla}_{\boldsymbol{u}^j}\, p(\boldsymbol{u}^j) = \frac{1}{\hat{s}^2\, n_d} \sum_{j'=1}^{n_d} (\frac{\hat{s}}{s}\boldsymbol{\eta}^{j'} - \boldsymbol{u}^j)\, \exp\{-\frac{1}{2\hat{s}^2} \|\frac{\hat{s}}{s}\boldsymbol{\eta}^{j'} - \boldsymbol{u}^j\|^2\}\,,$$



(a) pdf of H_1 and $H_{\mathrm{ar},1}$. (b) pdf of H_1 and $H_{\mathrm{DB},1}$. (c) pdf of H_1 and $H_{\mathrm{TB},1}$ at time $n_{\mathrm{opt}}\,\Delta t$.

(d) pdf of H_6 and $H_{\mathrm{ar},6}$. (e) pdf of H_6 and $H_{\mathrm{DB},6}$. (f) pdf of H_6 and $H_{\mathrm{TB},6}$ at time $n_{\mathrm{opt}}\,\Delta t$.

(g) pdf of H_{25} and $H_{\mathrm{ar},25}$. (h) pdf of H_{25} and $H_{\mathrm{DB},25}$. (i) pdf of H_{25} and $H_{\mathrm{TB},25}$ at time $n_{\mathrm{opt}}\,\Delta t$.

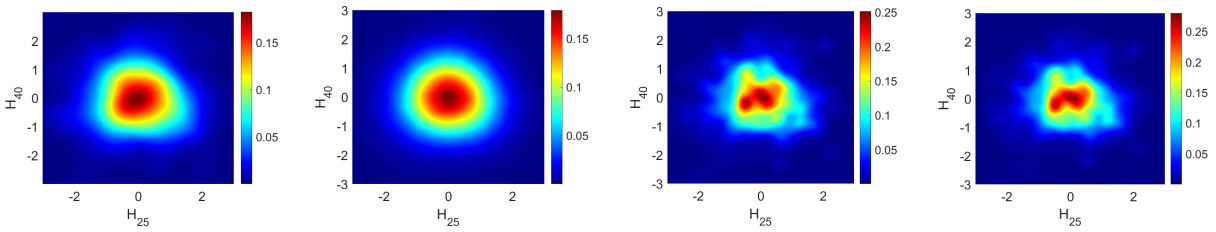
(j) pdf of H_{40} and $H_{\mathrm{ar},40}$. (k) pdf of H_{40} and $H_{\mathrm{DB},40}$. (l) pdf of H_{40} and $H_{\mathrm{TB},40}$ at time $n_{\mathrm{opt}}\,\Delta t$.

Figure 11: Application 3. Probability density function (pdf) of components 1, 6, 25, and 40 for $\boldsymbol{H}$ estimated with the n_d realizations of the training dataset (thin black line) and pdf estimated with n_{ar} learned realizations (thick blue line), for $\boldsymbol{H}_{\mathrm{ar}}$ using MCMC without PLoM (a,d,g,j), for $\boldsymbol{H}_{\mathrm{DB}}$ using PLoM with RODB (b,e,h,k), and for $\boldsymbol{H}_{\mathrm{TB}}$ using PLoM with ROTB($n_{\mathrm{opt}}\Delta t$) (c, f, i, l).

in which $\hat{s}$ is the modified Silverman bandwidth s, which has been introduced in [64],

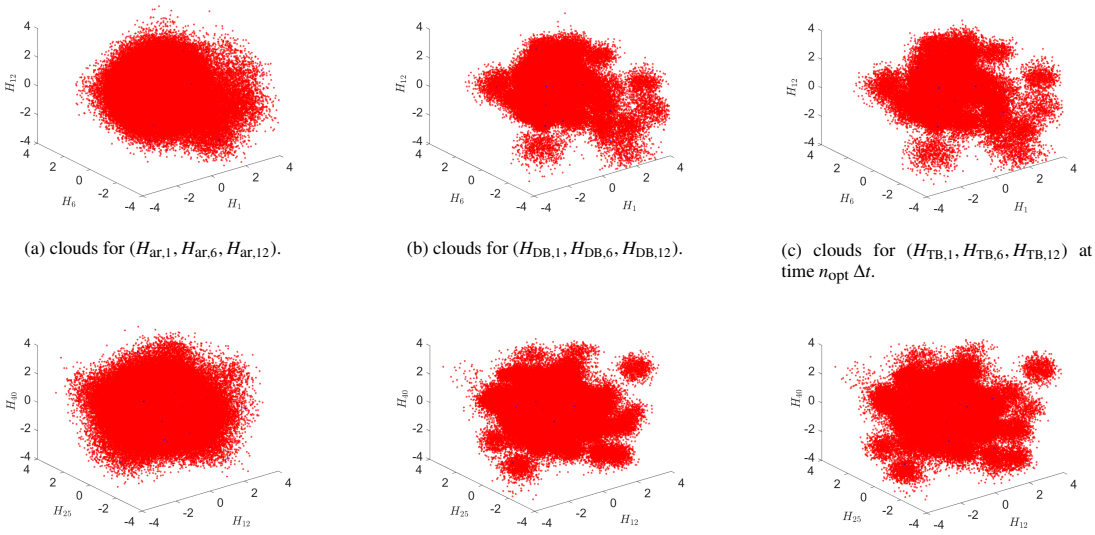
$$\hat{s} = \frac{s}{\sqrt{s^2 + \dfrac{n_d - 1}{n_d}}} \quad , \quad s = \left\{ \frac{4}{n_d(2+\nu)} \right\}^{1/(\nu+4)} \quad .$$

(2) $[\boldsymbol{\mathcal{W}}^{\mathrm{wien}}(t)] = [\mathbb{W}^{\mathrm{wien}}(t)]\,[a]$ where $\{[\mathbb{W}^{\mathrm{wien}}(t)],\, t \in \mathbb{R}^+\}$ is the $\mathbb{M}_{\nu,n_d}$-valued normalized Wiener process.
(3) $[\boldsymbol{\mathcal{N}}]$ is the $\mathbb{M}_{\nu,n_d}$-valued normalized Gaussian random matrix that is independent of process $[\mathbb{W}^{\mathrm{wien}}]$.
(4) The free parameter f_0, such that $0 < f_0 < 4/\hat{s}$, allows the dissipation term of the nonlinear second-order dynamic system (dissipative Hamiltonian system) to be controlled in order to kill the transient part induced by the initial conditions. A common value is $f_0 = 4$ (note that $\hat{s} < 1$).



(a) joint pdf of H_{25} with H_{40}. (b) joint pdf of $H_{\mathrm{ar},25}$ with $H_{\mathrm{ar},40}$. (c) joint pdf of $H_{\mathrm{DB},25}$ with $H_{\mathrm{DB},40}$. (d) joint pdf of $H_{\mathrm{TB},25}$ with $H_{\mathrm{TB},40}$ at time $n_{\mathrm{opt}}\,\Delta t$.

Figure 12: Application 3. Joint probability density function of components 24 with 40 of $\boldsymbol{H}$ estimated with the n_d realizations of the training dataset (a) and estimated with n_{ar} learned realizations, for $\boldsymbol{H}_{\mathrm{ar}}$ using MCMC without PLoM (b), for $\boldsymbol{H}_{\mathrm{DB}}$ using PLoM with RODB (c), and for $\boldsymbol{H}_{\mathrm{TB}}$ using PLoM with ROTB($n_{\mathrm{opt}}\Delta t$) (d).



(a) clouds for $(H_{\mathrm{ar},1}, H_{\mathrm{ar},6}, H_{\mathrm{ar},12})$. (b) clouds for $(H_{\mathrm{DB},1}, H_{\mathrm{DB},6}, H_{\mathrm{DB},12})$. (c) clouds for $(H_{\mathrm{TB},1}, H_{\mathrm{TB},6}, H_{\mathrm{TB},12})$ at time $n_{\mathrm{opt}}\,\Delta t$.

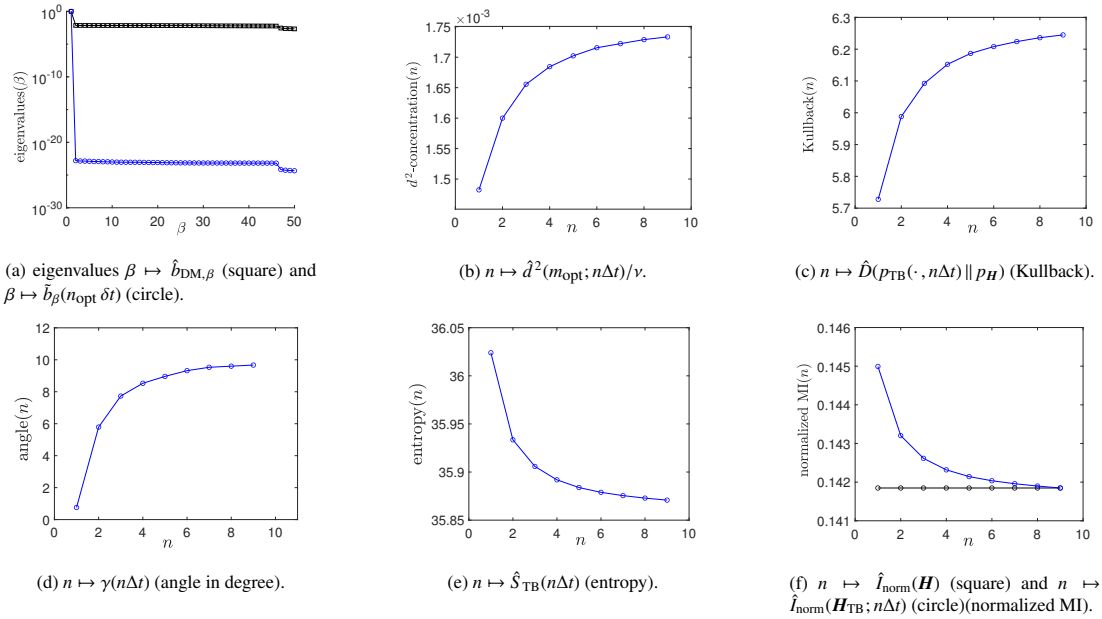
(d) clouds for $(H_{\mathrm{ar},12}, H_{\mathrm{ar},25}, H_{\mathrm{ar},40})$. (e) clouds for $(H_{\mathrm{DB},12}, H_{\mathrm{DB},25}, H_{\mathrm{DB},40})$. (f) clouds for $(H_{\mathrm{TB},12}, H_{\mathrm{TB},25}, H_{\mathrm{TB},40})$ at time $n_{\mathrm{opt}}\,\Delta t$.

Figure 13: Application 3. Clouds of n_{ar} points corresponding to n_{ar} learned realizations, for components 1, 6, 12 (a,b,c) and components 12, 25, 40 (d,e,f), for $\boldsymbol{H}_{\mathrm{ar}}$ using MCMC without PLoM (a,d), for $\boldsymbol{H}_{\mathrm{DB}}$ using PLoM with RODB (b,e), and for $\boldsymbol{H}_{\mathrm{TB}}$ using PLoM with ROTB($n_{\mathrm{opt}}\Delta t$) (c,f).

(5) We then have $[\boldsymbol{Z}] = \lim_{t\to+\infty}[\mathcal{Z}(t)]$ in probability distribution. The Störmer-Verlet scheme is used [1] for solving the reduced-order ISDE, which allows for generating the learned realizations, $[z^1_{\mathrm{ar}}], \ldots, [z^{n_{\mathrm{MCH}}}_{\mathrm{ar}}]$, and then generating the learned realizations $[\eta^1_{\mathrm{ar}}], \ldots, [\eta^{n_{\mathrm{MCH}}}_{\mathrm{ar}}]$ such that $[\eta^\ell_{\mathrm{ar}}] = [z^\ell_{\mathrm{ar}}]\,[g_{\mathrm{opt}}]^T$. The implementation of the Störmer-Verlet scheme is detailed, for instance, in the Appendix of [22] for parallel computation, introducing the following parameters: the integration time step Δt_{SV}, the initial time $t_i = 0$, and the final integration time $t_f = M_0 \times \Delta t_{\mathrm{SV}}$, at which the stationary solution is reached.
(6) The learned realizations $\{\boldsymbol{x}^{\ell'}_{\mathrm{ar}}, \ell' = 1, \ldots, n_{\mathrm{ar}}\}$ of random vector $\boldsymbol{X}$ are then calculated (see Eq. (A.1)) by $\boldsymbol{x}^{\ell'}_{\mathrm{ar}} = \underline{\boldsymbol{x}} + [\varphi]\,[\mu]^{1/2}\,\boldsymbol{\eta}^{\ell'}_{\mathrm{ar}}$.

Appendix A.6. Constraints on the second-order moments of the components of $\boldsymbol{H}_{\mathrm{ar}}$

In general, the mean value of $\boldsymbol{H}_{\mathrm{ar}}$ estimated using the n_{ar} learned realizations $\{\boldsymbol{\eta}^{\ell'}_{\mathrm{ar}}, \ell' = 1, \ldots, n_{\mathrm{ar}}\}$, is sufficiently close to zero. Likewise, the estimate of the covariance matrix of $\boldsymbol{H}_{\mathrm{ar}}$, which must be the identity matrix, is sufficiently close to a diagonal matrix. However, sometimes the diagonal entries of the estimated covariance matrix can be lower



(a) eigenvalues $\beta \mapsto \hat{b}_{\mathrm{DM},\beta}$ (square) and $\beta \mapsto \tilde{b}_\beta(n_{\mathrm{opt}}\,\delta t)$ (circle).

(b) $n \mapsto \hat{d}^2(m_{\mathrm{opt}}; n\Delta t)/\nu$.

(c) $n \mapsto \hat{D}(p_{\mathrm{TB}}(\cdot\,, n\Delta t)\,\|\,p_{\boldsymbol{H}})$ (Kullback).

(d) $n \mapsto \gamma(n\Delta t)$ (angle in degree).

(e) $n \mapsto \hat{S}_{\mathrm{TB}}(n\Delta t)$ (entropy).

(f) $n \mapsto \hat{I}_{\mathrm{norm}}(\boldsymbol{H})$ (square) and $n \mapsto \hat{I}_{\mathrm{norm}}(\boldsymbol{H}_{\mathrm{TB}}; n\Delta t)$ (circle)(normalized MI).

Figure 14: Application 3. Functions characterizing the reduced-order transient basis ROTB($n\Delta t$) as a function of time $n\Delta t$: eigenvalues of $[K_{\mathrm{DM}}]$ and of symmetrized $[\tilde{K}(n\Delta t)]$ (a); measure concentration with d^2/ν-criterion (b) and with Kullback-Leibler divergence criteria (c); angle between the subspaces spanned by RODB and ROTB($n\Delta t$) (d); entropy of pdf $p_{\mathrm{TB}}(\cdot\,; n\Delta t)$ (e); normalized mutual information (MI) of pdf $p_{\boldsymbol{H}}$ and $p_{\mathrm{TB}}(\cdot\,; n\Delta t)$ (f).

than 1. Normalization can be recovered by imposing constraints

$$E\{(H_{\mathrm{ar},k})^2\} = 1\ ,\ k = 1, \ldots, \nu\,, \tag{A.9}$$

in the algorithm presented in Appendix A.5. For that, we use the method and the iterative algorithm presented in [32] (that is based on Sections 5.5 and 5.6 of [21]). The constraints are imposed by using the Kullback-Leibler minimum cross-entropy principle. The resulting optimization problem is formulated using a Lagrange multiplier $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_\nu)$ associated with the constraints. The optimal solution of the Lagrange multiplier is computed using an efficient iterative algorithm. At each iteration, the MCMC generator detailed in Appendix A.5 is used. The constraints are rewritten as

$$E\{\boldsymbol{h}(\boldsymbol{H}_{\mathrm{ar}})\} = \boldsymbol{b}\,, \tag{A.10}$$

in which the function $\boldsymbol{h} = (h_1, \ldots, h_\nu)$ and the vector $\boldsymbol{b} = (b_1, \ldots, b_\nu)$ are such that $h_k(\boldsymbol{H}_{\mathrm{ar}}) = (H_{\mathrm{ar},k})^2$ and $b_k = 1$ for k in $\{1, \ldots, \nu\}$. To take into account the constraints in the algorithm presented in Appendix A.5, Eq. (A.8) is replaced by the following one,

$$[L_\lambda([u])]_{kj} = \frac{1}{p(\boldsymbol{u}^j)}\,\{\boldsymbol{\nabla}_{\boldsymbol{u}^j}\,p(\boldsymbol{u}^j)\}_k - 2\,\lambda_k u_k^j\,. \tag{A.11}$$

It should be noted that Eqs. (A.9) to (A.11) can be straightforwardly extended to the case in which the constraint defined by Eq. (A.9) is replaced by the full second-order moment constraints $E\{\boldsymbol{H}_{\mathrm{ar}}\} = \boldsymbol{0}_\nu$ and $E\{\boldsymbol{H}_{\mathrm{ar}} \otimes \boldsymbol{H}_{\mathrm{ar}}\} = [I_\nu]$, that is to say,

$$E\{H_{\mathrm{ar},k}\} = 0\,,\ E\{(H_{\mathrm{ar},k})^2\} = 1\ ,\ k = 1, \ldots, \nu \quad , \quad E\{H_{\mathrm{ar},k} H_{\mathrm{ar},k'}\} = 0\ ,\ 1 \le k < k' \le \nu\,. \tag{A.12}$$

The iteration algorithm for computing $\boldsymbol{\lambda}^{i+1}$ as a function of $\boldsymbol{\lambda}^i$ is the following,

$$\begin{aligned} &\boldsymbol{\lambda}^{i+1} = \boldsymbol{\lambda}^i - \alpha_i [\Gamma''(\boldsymbol{\lambda}^i)]^{-1}\,\boldsymbol{\Gamma}'(\boldsymbol{\lambda}^i) \quad , \quad i \ge 0\,, \\ &\boldsymbol{\lambda}^0 = \boldsymbol{0}_\nu\,, \end{aligned}$$

in which $\mathbf{\Gamma}'(\boldsymbol{\lambda}^i) = \boldsymbol{b} - E\{\boldsymbol{h}(\boldsymbol{H}_{\boldsymbol{\lambda}^i})\}$ and $[\Gamma''(\boldsymbol{\lambda}^i)] = [\mathrm{cov}\{\boldsymbol{h}(\boldsymbol{H}_{\boldsymbol{\lambda}^i})\}]$ (the covariance matrix), and where α_i is a relaxation function (less than 1) that is introduced for controlling the convergence as a function of iteration number i. For given $i_2 \geq 2$, for given β_1 and β_2 such that $0 < \beta_1 < \beta_2 \leq 1$, α_i can be defined by:
- for $i \leq i_2$, $\alpha_i = \beta_1 + (\beta_2 - \beta_1)(i-1)/(i_2 - 1)$;
- for $i > i_2$, $\alpha_i = \beta_2$.
The convergence of the iteration algorithm is controlled by the error function $i \mapsto \mathrm{err}(i)$ defined by

$$\mathrm{err}(i) = \|\boldsymbol{b} - E\{\boldsymbol{h}(\boldsymbol{H}_{\boldsymbol{\lambda}^i})\}\|/\|\boldsymbol{b}\| \,. \tag{A.13}$$

At each iteration i, $E\{\boldsymbol{h}(\boldsymbol{H}_{\boldsymbol{\lambda}^i})\}$ and $[\mathrm{cov}\{\boldsymbol{h}(\boldsymbol{H}_{\boldsymbol{\lambda}^i})\}]$ are estimated by using the n_{ar} learned realizations of $\boldsymbol{H}_{\boldsymbol{\lambda}^i}$ obtained by reshaping the learned realizations. If i_{last} is the last iteration corresponding to convergence, we have $\boldsymbol{H}_{\mathrm{ar}} = \boldsymbol{H}_{\boldsymbol{\lambda}^i}$ with $i = i_{\mathrm{last}}$.

Appendix B. Estimation of the Kullback-Leibler divergence, the mutual information, and the entropy from a set of realizations

The definition Kullback-Leibler divergence, the mutual information, and the entropy can be found in [95, 96, 98]. The estimation of these quantities from a set of independent realizations is carried out using the Gaussian kernel density estimation (GKDE) method [65, 36, 66].

For $\nu > 1$, let $\boldsymbol{X} = (X_1, \ldots, X_\nu)$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_\nu)$ be $\mathbb{R}^\nu$-valued random variables defined on the probability space $(\Theta, \mathcal{T}, \mathcal{P})$, whose probability measures are $P_{\boldsymbol{X}}(d\boldsymbol{x}) = p_{\boldsymbol{X}}(\boldsymbol{x})\, d\boldsymbol{x}$ and $P_{\boldsymbol{Y}}(d\boldsymbol{y}) = p_{\boldsymbol{Y}}(\boldsymbol{y})\, d\boldsymbol{y}$, in which the probability density functions $p_{\boldsymbol{X}}$ and $p_{\boldsymbol{Y}}$ are assumed to be strictly positive. Let $\{\boldsymbol{x}^\ell, \ell = 1, \ldots, N_x\}$ be N_x independent realizations of $\boldsymbol{X}$ and let $\{\boldsymbol{y}^j, j = 1, \ldots, N_y\}$ be N_y independent realizations of $\boldsymbol{Y}$. For $k = 1, \ldots, \nu$, let σ_{X_k} and σ_{Y_k} be the standard deviation of X_k and Y_k that are estimated (empirical estimator) with the independent realizations. Finally, we introduce the Silverman bandwidth for the Gaussian KDE estimation of $p_{\boldsymbol{X}}$ and $p_{\boldsymbol{Y}}$,

$$s_x = \left\{\frac{4}{N_x(2+\nu)}\right\}^{1/(\nu+4)} \quad , \quad s_y = \left\{\frac{4}{N_y(2+\nu)}\right\}^{1/(\nu+4)} \,. \tag{B.1}$$

Appendix B.1. Estimation of the Kullback-Leibler divergence from a set of realizations

The Kullback-Leibler divergence (or the relative entropy) between $p_{\boldsymbol{X}}$ and $p_{\boldsymbol{Y}}$ is defined by

$$D(p_{\boldsymbol{X}} \| p_{\boldsymbol{Y}}) = \int_{\mathbb{R}^\nu} p_{\boldsymbol{X}}(\boldsymbol{x}) \, \log\left(\frac{p_{\boldsymbol{X}}(\boldsymbol{x})}{p_{\boldsymbol{Y}}(\boldsymbol{x})}\right) d\boldsymbol{x} = E\left\{\log\left(\frac{p_{\boldsymbol{X}}(\boldsymbol{X})}{p_{\boldsymbol{Y}}(\boldsymbol{X})}\right)\right\} \,. \tag{B.2}$$

The GKDE, $\hat{D}(p_{\boldsymbol{X}} \| p_{\boldsymbol{Y}})$, of $D(p_{\boldsymbol{X}} \| p_{\boldsymbol{Y}})$ yields the formula,

$$\begin{aligned} \hat{D}(p_{\boldsymbol{X}} \,\|\, p_{\boldsymbol{Y}}) =& \nu \log\left(\frac{s_y}{s_x}\right) + \log\left(\frac{N_y}{N_x}\right) + \log\left(\frac{\sigma_{Y_1} \times \ldots \times \sigma_{Y_\nu}}{\sigma_{X_1} \times \ldots \times \sigma_{X_\nu}}\right) \\ &+ \frac{1}{N_x} \sum_{\ell'=1}^{N_x} \log\left\{\frac{\sum_{\ell=1}^{N_x} \exp\left(-\frac{1}{2s_x^2} \sum_{k=1}^{\nu} (\frac{x_k^{\ell'} - x_k^{\ell}}{\sigma_{X_k}})^2\right)}{\sum_{j=1}^{N_y} \exp\left(-\frac{1}{2s_y^2} \sum_{k=1}^{\nu} (\frac{x_k^{\ell'} - y_k^{j}}{\sigma_{Y_k}})^2\right)}\right\} \,. \end{aligned} \tag{B.3}$$

Appendix B.2. Estimation of the mutual information from a set of realizations

The mutual information $I(\boldsymbol{X})$ of $\boldsymbol{X}$ allows to quantify the level of statistical dependencies of the components $X_1, \ldots, X_\nu$ of $\boldsymbol{X} = (X_1, \ldots, X_\nu)$. Let p_{X_k} be the pdf of real-valued random variable X_k,

$$p_{X_k}(x_k) = \int_{\mathbb{R}^{\nu-1}} p_{\boldsymbol{X}}(x_1, \ldots, x_{k-1}, x_k, x_{k+1}, \ldots, x_\nu)\, dx_1 \ldots dx_{k-1}\, dx_{k+1} \ldots\, dx_\nu \,. \tag{B.4}$$

The mutual information $I(\boldsymbol{X})$ is defined by

$$I(\boldsymbol{X}) = D(p_{\boldsymbol{X}} \,\|\, \otimes_{k=1}^{\nu} p_{X_k}) = E\left\{\log\left(\frac{p_{\boldsymbol{X}}(\boldsymbol{X})}{p_{X_1}(X_1) \times \ldots \times p_{X_\nu}(X_\nu)}\right)\right\} \,. \tag{B.5}$$

Eq. (B.5) shows that, if the components $X_1, \ldots, X_\nu$ are statistically independent, then $I(\boldsymbol{X}) = 0$. The GKDE, $\hat{I}(\boldsymbol{X})$ of $I(\boldsymbol{X})$ yields the formula,

$$\hat{I}(\boldsymbol{X}) = \frac{1}{N_x} \sum_{\ell'=1}^{N_x} \log \left\{ \frac{\frac{1}{N_x} \sum_{\ell=1}^{N_x} \exp\left(-\frac{1}{2s_x^2} \sum_{k=1}^{\nu} (\frac{x_k^{\ell'} - x_k^{\ell}}{\sigma_{X_k}})^2 \right)}{\prod_{k=1}^{\nu} \left(\frac{1}{N_x} \sum_{\ell''=1}^{N_x} \exp\left(-\frac{1}{2s_x^2} (\frac{x_k^{\ell'} - x_k^{\ell''}}{\sigma_{X_k}})^2 \right) \right)} \right\} . \tag{B.6}$$

Appendix B.3. Estimation of the entropy from a set of realizations

The entropy related to $p_{\boldsymbol{X}}$ is defined by

$$S_{\boldsymbol{X}} = -\int_{\mathbb{R}^\nu} p_{\boldsymbol{X}}(\boldsymbol{x}) \, \log p_{\boldsymbol{X}}(\boldsymbol{x}) \, d\boldsymbol{x} = -E\left\{ \log p_{\boldsymbol{X}}(\boldsymbol{X}) \right\} . \tag{B.7}$$

The GKDE, $\hat{S}_{\boldsymbol{X}}$, of $S_{\boldsymbol{X}}$ yields the formula,

$$\begin{aligned} \hat{S}_{\boldsymbol{X}} =& \nu \log(s_x \sqrt{2\pi}) + \log(\sigma_{X_1} \times \ldots \times \sigma_{X_\nu}) \\ & - \frac{1}{N_x} \sum_{\ell'=1}^{N_x} \log \left\{ \frac{1}{N_x} \sum_{\ell=1}^{N_x} \exp\left(-\frac{1}{2s_x^2} \sum_{k=1}^{\nu} (\frac{x_k^{\ell'} - x_k^{\ell}}{\sigma_{X_k}})^2 \right) \right\} . \end{aligned} \tag{B.8}$$

Since $J_S(N_x) = \nu \log(s_x \sqrt{2\pi})$ is asymptotically for $N_x \to +\infty$ in $-\log(N_x)$, the entropy decreases when N_x increases.

Conflict of interest

The author declares that he has no conflict of interest.

References

[1] C. Soize, R. Ghanem, Data-driven probability concentration and sampling on manifold, Journal of Computational Physics 321 (2016) 242–258. doi:10.1016/j.jcp.2016.05.044.

[2] R. Coifman, S. Lafon, A. Lee, M. Maggioni, B. Nadler, F. Warner, S. Zucker, Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps, PNAS 102 (21) (2005) 7426–7431. doi:10.1073/pnas.0500334102.

[3] R. Coifman, S. Lafon, Diffusion maps, Applied and Computational Harmonic Analysis 21 (1) (2006) 5–30. doi:10.1016/j.acha.2006.04.006.

[4] B. Nadler, S. Lafon, R. R. Coifman, I. G. Kevrekidis, Diffusion maps, spectral clustering and reaction coordinates of dynamical systems, Applied and Computational Harmonic Analysis 21 (1) (2006) 113–127. doi:10.1016/j.acha.2005.07.004.

[5] R. R. Coifman, I. G. Kevrekidis, S. Lafon, M. Maggioni, B. Nadler, Diffusion maps, reduction coordinates, and low dimensional representation of stochastic systems, Multiscale Modeling & Simulation 7 (2) (2008) 842–864. doi:10.1137/070696325.

[6] A. Singer, R. Erban, I. G. Kevrekidis, R. R. Coifman, Detecting intrinsic slow variables in stochastic dynamical systems by anisotropic diffusion maps, Proceedings of the National Academy of Sciences 106 (38) (2009) 16090–16095. doi:10.1073/pnas.0905547106.

[7] K. B. Korb, A. E. Nicholson, Bayesian artificial intelligence, CRC press, Boca Raton, 2010.

[8] K. P. Murphy, Machine Learning: A Probabilistic Perspective, MIT press, 2012.

[9] Z. Ghahramani, Probabilistic machine learning and artificial intelligence, Nature 521 (7553) (2015) 452–459. doi:10.1038/nature14541.

[10] S. Russel, P. Norvig, Artifical Intelligence, A Modern Approach, Third Edition, Pearson, Harlow, 2016.

[11] V. Vapnik, The Nature of Statistical Learning Theory, Springer, New York, 2000. doi:10.1007/978-1-4757-3264-1.

[12] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, Second Edition, Springer, 2009. doi:10.1007/b94608.

[13] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning, Vol. 112, Springer, 2013.

[14] J. Taylor, R. J. Tibshirani, Statistical learning and selective inference, Proceedings of the National Academy of Sciences 112 (25) (2015) 7629–7634. doi:10.1073/pnas.1507583112.

[15] R. Swischuk, L. Mainini, B. Peherstorfer, K. Willcox, Projection-based model reduction: Formulations for physics-based machine learning, Computers & Fluids 179 (2019) 704–717. doi:10.1016/j.compfluid.2018.07.021.

[16] A. C. Öztireli, M. Alexa, M. Gross, Spectral sampling of manifolds, ACM Transactions on Graphics (TOG) 29 (6) (2010) 1–8. doi:10.1145/1882261.1866190.

[17] G. Perrin, C. Soize, S. Marque-Pucheu, J. Garnier, Nested polynomial trends for the improvement of Gaussian process-based predictors, Journal of Computational Physics 346 (2017) 389–402. doi:10.1016/j.jcp.2017.05.051.

[18] Y. Kevrekidis, Manifold learning for parameter reduction, Bulletin of the American Physical Society 65 (2020).

[19] K. Kontolati, D. Loukrezis, K. R. dos Santos, D. G. Giovanis, M. D. Shields, Manifold learning-based polynomial chaos expansions for high-dimensional surrogate models, International Journal for Uncertainty Quantification 12 (4) (2022). doi:10.1615/Int.J.UncertaintyQuantification.2022039936.
[20] S. Pan, K. Duraisamy, Physics-informed probabilistic learning of linear embeddings of nonlinear dynamics with guaranteed stability, SIAM Journal on Applied Dynamical Systems 19 (1) (2020) 480–509. doi:10.1137/19M1267246.
[21] C. Soize, R. Ghanem, Physics-constrained non-Gaussian probabilistic learning on manifolds, International Journal for Numerical Methods in Engineering 121 (1) (2020) 110–145. doi:10.1002/nme.6202.
[22] C. Soize, R. Ghanem, Probabilistic learning on manifolds constrained by nonlinear partial differential equations for small datasets, Computer Methods in Applied Mechanics and Engineering 380 (2021) 113777. doi:10.1016/j.cma.2021.113777.
[23] A. Talwalkar, S. Kumar, H. Rowley, Large-scale manifold learning, in: 2008 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2008, pp. 1–8. doi:10.1109/CVPR.2008.4587670.
[24] Y. Marzouk, T. Moselhy, M. Parno, A. Spantini, Sampling via measure transport: An introduction, Handbook of uncertainty quantification (2016) 1–41doi:10.1007/978-3-319-11259-6_23-1.
[25] M. D. Parno, Y. M. Marzouk, Transport map accelerated markov chain Monte Carlo, SIAM/ASA Journal on Uncertainty Quantification 6 (2) (2018) 645–682. doi:10.1137/17M1134640.
[26] G. Perrin, C. Soize, N. Ouhbi, Data-driven kernel representations for sampling with an unknown block dependence structure under correlation constraints, Computational Statistics & Data Analysis 119 (2018) 139–154. doi:10.1016/j.csda.2017.10.005.
[27] C. Soize, R. Ghanem, Polynomial chaos representation of databases on manifolds, Journal of Computational Physics 335 (2017) 201–221. doi:10.1016/j.jcp.2017.01.031.
[28] C. Soize, R. Ghanem, C. Safta, X. Huan, Z. P. Vane, J. C. Oefelein, G. Lacaze, H. N. Najm, Q. Tang, X. Chen, Entropy-based closure for probabilistic learning on manifolds, Journal of Computational Physics 388 (2019) 528–533. doi:10.1016/j.jcp.2018.12.029.
[29] C. Soize, R. Ghanem, C. Desceliers, Sampling of Bayesian posteriors with a non-Gaussian probabilistic learning on manifolds from a small dataset, Statistics and Computing 30 (5) (2020) 1433–1457. doi:10.1007/s11222-020-09954-6.
[30] C. Soize, R. Ghanem, Probabilistic learning on manifolds, Foundations of Data Science 2 (3) (2020) 279–307. doi:10.3934/fods.2020013.
[31] C. Soize, Probabilistic learning inference of boundary value problem with uncertainties based on Kullback-Leibler divergence under implicit constraints, Computer Methods in Applied Mechanics and Engineering 395 (2022) 115078. doi:10.1016/j.cma.2022.115078.
[32] C. Soize, R. Ghanem, Probabilistic learning on manifolds (PLoM) with partition, International Journal for Numerical Methods in Engineering 123 (1) (2022) 268–290. doi:10.1002/nme.6856.
[33] C. Soize, Probabilistic learning constrained by realizations using a weak formulation of fourier transform of probability measures, Computational Statistics 38 (4) (2023) 1879–1925. doi:10.1007/s00180-022-01300-w.
[34] M. C. Kennedy, A. O'Hagan, Bayesian calibration of computer models, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63 (3) (2001) 425–464. doi:10.1111/1467-9868.00294.
[35] Y. M. Marzouk, H. N. Najm, L. A. Rahn, Stochastic spectral methods for efficient Bayesian solution of inverse problems, Journal of Computational Physics 224 (2) (2007) 560–586. doi:10.1016/j.jcp.2006.10.010.
[36] J. E. Gentle, Computational statistics, Springer, New York, 2009. doi:10.1007/978-0-387-98144-4.
[37] A. M. Stuart, Inverse problems: a Bayesian perspective, Acta Numerica 19 (2010) 451–559. doi:10.1017/S0962492910000061.
[38] H. Owhadi, C. Scovel, T. Sullivan, On the brittleness of Bayesian inference, SIAM Review 57 (4) (2015) 566–582. doi:10.1137/130938633.
[39] H. G. Matthies, E. Zander, B. V. Rosić, A. Litvinenko, O. Pajonk, Inverse problems in a Bayesian setting, in: Computational Methods for Solids and Fluids, Vol. 41, Springer, 2016, pp. 245–286. doi:10.1007/978-3-319-27996-1_10.
[40] M. Dashti, A. M. Stuart, The Bayesian approach to inverse problems, in: R. Ghanem, D. Higdon, O. Houman (Eds.), Handbook of Uncertainty Quantification, Springer, Cham, Switzerland, 2017, Ch. 10, pp. 311–428. doi:10.1007/978-3-319-12385-1_7.
[41] R. Ghanem, D. Higdon, H. Owhadi, Handbook of Uncertainty Quantification, Vol. 1 to 3, Springer, Cham, Switzerland, 2017. doi:10.1007/978-3-319-12385-1.
[42] A. Spantini, T. Cui, K. Willcox, L. Tenorio, Y. Marzouk, Goal-oriented optimal approximations of Bayesian linear inverse problems, SIAM Journal on Scientific Computing 39 (5) (2017) S167–S196. doi:10.1137/16M1082123.
[43] G. Perrin, C. Soize, Adaptive method for indirect identification of the statistical properties of random fields in a Bayesian framework, Computational Statistics 35 (1) (2020) 111–133. doi:10.1007/s00180-019-00936-5.
[44] R. Ghanem, C. Soize, Probabilistic nonconvex constrained optimization with fixed number of function evaluations, International Journal for Numerical Methods in Engineering 113 (4) (2018) 719–741. doi:10.1002/nme.5632.
[45] C. Soize, Design optimization under uncertainties of a mesoscale implant in biological tissues using a probabilistic learning algorithm, Computational Mechanics 62 (3) (2018) 477–497. doi:10.1007/s00466-017-1509-x.
[46] R. Ghanem, C. Soize, C. Thimmisetty, Optimal well-placement using probabilistic learning, Data-Enabled Discovery and Applications 2 (1) (2018) 1–16. doi:10.1007/s41688-017-0014-x.
[47] C. Farhat, R. Tezaur, T. Chapman, P. Avery, C. Soize, Feasible probabilistic learning method for model-form uncertainty quantification in vibration analysis, AIAA Journal 57 (11) (2019) 4978–4991. doi:10.2514/1.J057797.
[48] R. Ghanem, C. Soize, C. Safta, X. Huan, G. Lacaze, J. C. Oefelein, H. N. Najm, Design optimization of a scramjet under uncertainty using probabilistic learning on manifolds, Journal of Computational Physics 399 (2019) 108930. doi:10.1016/j.jcp.2019.108930.
[49] J. O. Almeida, F. A. Rochinha, A probabilistic learning approach applied to the optimization of wake steering in wind farms, Journal of Computing and Information Science in Engineering 23 (1) (2022) 011003. doi:10.1115/1.4054501.
[50] E. Capiez-Lernout, C. Soize, Nonlinear stochastic dynamics of detuned bladed disks with uncertain mistuning and detuning optimization using a probabilistic machine learning tool, International Journal of Non-Linear Mechanics 143 (2022) 104023. doi:10.1016/j.ijnonlinmec.2022.104023.
[51] J. O. Almeida, F. A. Rochinha, A probabilistic learning approach applied to the optimization of wake steering in wind farms, Journal of Computing and Information Science in Engineering 23 (1) (2023) 011003. doi:10.1115/1.4054501.
[52] J. Guilleminot, J. E. Dolbow, Data-driven enhancement of fracture paths in random composites, Mechanics Research Communications 103

(2020) 103443. doi:10.1016/j.mechrescom.2019.103443.
[53] P. Chen, J. Guilleminot, C. Soize, Concurrent multiscale simulations of nonlinear random materials using probabilistic learning, Computer Methods in Applied Mechanics and Engineering 422 (2024) 116837. doi:10.1016/j.cma.2024.116837.
[54] C. Soize, An overview on uncertainty quantification and probabilistic learning on manifolds in multiscale mechanics of materials, Mathematics and Mechanics of Complex Systems 11 (1) (2023) 87–174. doi:10.2140/memocs.2023.11.87.
[55] R. Ghanem, C. Soize, L. Mehrez, V. Aitharaju, Probabilistic learning and updating of a digital twin for composite material systems, International Journal for Numerical Methods in Engineering 123 (13) (2022) 3004–3020. doi:10.1002/nme.6430.
[56] O. Ezvan, C. Soize, C. Desceliers, R. Ghanem, Updating an uncertain and expensive computational model in structural dynamics based on one single target frf using a probabilistic learning tool, Computational Mechanics 71 (2023) 1161–1177. doi:10.1007/s00466-023-02301-2.
[57] E. Capiez-Lernout, C. Ezvan, Olivier Soize, Updating nonlinear stochastic dynamics of an uncertain nozzle model using probabilistic learning with partial observability and incomplete dataset, ASME Journal of Computing and Information Science in Engineering 24 (6) (2024) 061006,1–17. doi:10.1115/1.4065312.
[58] M. Arnst, C. Soize, K. Bulthies, Computation of sobol indices in global sensitivity analysis from small data sets by probabilistic learning on manifolds, International Journal for Uncertainty Quantification 11 (2) (2021) 1–23. doi:10.1615/Int.J.UncertaintyQuantification.2020032674.
[59] C. Soize, A. Orcesi, Machine learning for detecting structural changes from dynamic monitoring using the probabilistic learning on manifolds, Structure and Infrastructure Engineering Journal 17 (10) (2021) 1418–1430. doi:10.1080/15732479.2020.1811991.
[60] K. Zhong, J. G. Navarro, S. Govindjee, G. G. Deierlein, Surrogate modeling of structural seismic response using Probabilistic Learning on Manifolds, Earthquake Engineering and Structural Dynamics 52 (8) (2023) 2407–2428. doi:10.1002/eqe.3839.
[61] C. Soize, R. Ghanem, Probabilistic-learning-based stochastic surrogate model from small incomplete datasets for nonlinear dynamical systems, Computer Methods in Applied Mechanics and Engineering 418 (2024) 116498. doi:10.1016/j.cma.2023.116498.
[62] C. Soize, Q.-D. To, Polynomial-chaos-based conditional statistics for probabilistic learning with heterogeneous data applied to atomic collisions of helium on graphite substrate, Journal of Computational Physics 496 (2024) 112582,1–20. doi:10.1016/j.jcp.2023.112582.
[63] A. Sinha, C. Soize, C. Desceliers, G. Cunha, Aeroacoustic liner impedance metamodel from simulation and experimental data using probabilistic learning, AIAA Journal 61 (11) (2023) 4926–4934. doi:10.2514/1.J062991.
[64] C. Soize, Polynomial chaos expansion of a multimodal random vector, SIAM-ASA Journal on Uncertainty Quantification 3 (1) (2015) 34–60. doi:10.1137/140968495.
[65] A. Bowman, A. Azzalini, Applied Smoothing Techniques for Data Analysis: The Kernel Approach With S-Plus Illustrations, Vol. 18, Oxford University Press, Oxford: Clarendon Press, New York, 1997. doi:10.1007/s001800000033.
[66] G. Givens, J. Hoeting, Computational Statistics, 2nd Edition, John Wiley and Sons, Hoboken, New Jersey, 2013.
[67] J. L. Doob, Stochastic processes, John Wiley & Sons, New York, 1953.
[68] I. I. Guikhman, A. Skorokhod, Introduction à la Théorie des Processus Aléatoires, Edition Mir, 1980.
[69] A. Friedman, Stochastic Differential Equations and Applications, Dover Publications, Inc., Mineola, New York, 2006.
[70] C. Soize, The Fokker-Planck Equation for Stochastic Dynamical Systems and its Explicit Steady State Solutions, Vol. Series on Advances in Mathematics for Applied Sciences: Vol 17, World Scientific, Singapore, 1994. doi:10.1142/2347.
[71] C. W. Gardiner, Handbook of Stochastic Methods, Second Edition, Springer Verlag, Berlin, Heidelberg, 1985.
[72] H. Risken, The Fokker-Planck Equation, Second Edition, Springer Verlag, Berlin, Heidelberg, 1989.
[73] I. M. Gelfand, N. I. Vilenkin, Les Distributions. Tome 4. Application de l'Analyse Harmonique, Dunod, 1967. doi:10.1016/0375-9474(67)90547-7.
[74] B. Spencer, L. Bergman, On the numerical solution of the Fokker-Planck equation for nonlinear stochastic systems, Nonlinear Dynamics 4 (1993) 357–372. doi:10.1007/BF00120671.
[75] A. Masud, L. A. Bergman, Application of multi-scale finite element methods to the solution of the Fokker-Planck equation, Computer methods in applied mechanics and engineering 194 (12-16) (2005) 1513–1526. doi:10.1016/j.cma.2004.06.041.
[76] P. Kumar, S. Narayanan, Solution of Fokker-Planck equation by finite element and finite difference methods for nonlinear systems, Sadhana 31 (2006) 445–461. doi:10.1007/BF02716786.
[77] L. Pichler, A. Masud, L. A. Bergman, Numerical solution of the Fokker-Planck equation by finite difference and finite element methods - a comparative study, Computational Methods in Stochastic Dynamics: Volume 2 (2013) 69–85doi:10.1007/978-94-007-5134-7˙5.
[78] W. Deng, Finite element method for the space and time fractional Fokker-Planck equation, SIAM journal on numerical analysis 47 (1) (2009) 204–226. doi:10.1137/080714130.
[79] W. Anderson, M. Farazmand, Fisher information and shape-morphing modes for solving the Fokker-Planck equation in higher dimensions, Applied Mathematics and Computation 467 (2024) 128489. doi:10.1016/j.amc.2023.128489.
[80] P. L. Popelier, Solving the Schrödinger Equation: Has Everything Been Tried?, Imperial College Press, 2011. doi:10.1142/p780.
[81] M. Feit, J. Fleck Jr, A. Steiger, Solution of the Schrödinger equation by a spectral method, Journal of Computational Physics 47 (3) (1982) 412–433. doi:10.1016/0021-9991(82)90091-2.
[82] T. Iitaka, Solving the time-dependent Schrödinger equation numerically, Physical Review E 49 (5) (1994) 4684. doi:10.1103/PhysRevE.49.4684.
[83] T. Simos, P. Williams, A finite-difference method for the numerical solution of the Schrödinger equation, Journal of Computational and Applied Mathematics 79 (2) (1997) 189–205. doi:10.1016/S0377-0427(96)00156-2.
[84] R. J. Serfling, Approximation theorems of mathematical statistics, Vol. 162, John Wiley & Sons, 1980.
[85] C. Soize, Uncertainty Quantification. An Accelerated Course with Advanced Applications in Computational Engineering, Springer, New York, 2017. doi:10.1007/978-3-319-54339-0.
[86] N. Ikeda, S. Watanabe, Stochastic Differential Equations and Diffusion Processes, North-Holland, Amsterdam, 1981.
[87] R. Has'minski, Stochastic Stability of Differential Equations, Sijthoff & Noordhoff, Alphen aan den Rijn, The Netherlands, 1980, first English edition Khasminskii,2021, Spinger.
[88] D. W. Stroock, S. S. Varadhan, Multidimensional Diffusion Processes, Vol. 233, Springer-Verlag, Berlin,Heidelberg, 1997.
[89] P. Kloeden, E. Platen, Numerical Solution of Stochastic Differentials Equations, Springer-Verlag, Heidelberg, 1992.

[90] T. Duong, A. Cowling, I. Koch, M. Wand, Feature significance for multivariate kernel density estimation, Computational Statistics & Data Analysis 52 (9) (2008) 4225–4242. doi:10.1016/j.csda.2008.02.035.
[91] E. R. Hansen, A Table of Series and Products, Prentice-Hall, New York, 1975.
[92] A. Bhattacharyya, On the measures of divergence between two statistical populations defined by their probability distributions, Bulletin of the Calcultta Mathematical Society 35 (1943) 99–109.
[93] S. Kullback, R. A. Leibler, On information and sufficiency, The Annals of Mathematical Statistics 22 (1) (1951) 79–86. doi:10.1214/aoms/1177729694.
[94] A. Kolmogorov, On the shannon theory of information transmission in the case of continuous signals, IRE Transactions on Information Theory 2 (4) (1956) 102–108.
[95] T. M. Cover, J. A. Thomas, Elements of Information Theory, Second Edition, John Wiley & Sons, Hoboken, 2006.
[96] J. N. Kapur, H. K. Kesavan, Entropy Optimization Principles with Applications, Academic Press, San Diego, 1992.
[97] R. M. Gray, Entropy and Information Theory, 2nd Edition, Springer, New York, 2011. doi:10.1007/978-1-4419-7970-4.
[98] C. E. Shannon, A mathematical theory of communication, Bell system technical journal 27 (3) (1948) 379–423 & 623–659. doi:10.1002/j.1538-7305.1948.tb01338.x.
[99] S. Lafon, A. B. Lee, Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization, IEEE transactions on pattern analysis and machine intelligence 28 (9) (2006) 1393–1403. doi:10.1109/TPAMI.2006.184.