



HAL
open science

Probabilistic learning on manifolds

Christian Soize, Roger Ghanem

► **To cite this version:**

Christian Soize, Roger Ghanem. Probabilistic learning on manifolds. Foundations of Data Science, 2020, 2 (3), pp.279-307. 10.3934/fods.2020013 . hal-02919127

HAL Id: hal-02919127

<https://hal.science/hal-02919127>

Submitted on 21 Aug 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Probabilistic Learning on Manifolds

Christian SOIZE^{a,*}, Roger GHANEM^b

^aUniversité Gustave Eiffel, MSME UMR 8208 CNRS, 5 bd Descartes, 77454 Marne-la-Vallée, France

^bUniversity of Southern California, Viterbi School of Engineering, 210 KAP Hall, Los Angeles, CA 90089, United States

Abstract

This paper presents novel mathematical results in support of the probabilistic learning on manifolds (PLoM) recently introduced by the authors. An initial dataset, constituted of a small number of points given in an Euclidean space, is given. The points are independent realizations of a vector-valued random variable for which its non-Gaussian probability measure is unknown but is, *a priori*, concentrated in an unknown subset of the Euclidean space. A learned dataset, constituted of additional realizations, is constructed. A transport of the probability measure estimated with the initial dataset is done through a linear transformation constructed using a reduced-order diffusion-maps basis. It is proven that this transported measure is a marginal distribution of the invariant measure of a reduced-order Itô stochastic differential equation. The concentration of the probability measure is preserved. This property is shown by analyzing a distance between the random matrix constructed with the PLoM and the matrix representing the initial dataset, as a function of the dimension of the basis. It is further proven that this distance has a minimum for a dimension of the reduced-order diffusion-maps basis that is strictly smaller than the number of points in the initial dataset.

Notations

The following notations are used:

x : lower-case Latin of Greek letters are deterministic real variables.

\mathbf{x} : boldface lower-case Latin of Greek letters are deterministic vectors.

X : upper-case Latin or Greek letters are real-valued random variables.

\mathbf{X} : boldface upper-case Latin or Greek letters are vector-valued random variables.

$[x]$: lower-case Latin of Greek letters between brackets are deterministic matrices.

$[\mathbf{X}]$: boldface upper-case letters between brackets are matrix-valued random variables.

\mathbb{N}, \mathbb{R} : set of all the integers $\{0, 1, 2, \dots\}$, set of all the real numbers.

\mathbb{R}^n : Euclidean vector space on \mathbb{R} of dimension n .

$\mathbf{x} = (x_1, \dots, x_n)$: point in \mathbb{R}^n .

$\langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1 + \dots + x_n y_n$: inner product in \mathbb{R}^n .

$\|\mathbf{x}\|$: norm in \mathbb{R}^n such that $\|\mathbf{x}\|^2 = \langle \mathbf{x}, \mathbf{x} \rangle$.

$\mathbb{M}_{n,m}$: set of all the $(n \times m)$ real matrices.

\mathbb{M}_n : set of all the square $(n \times n)$ real matrices.

\mathbb{M}_n^{+0} : set of all the positive symmetric $(n \times n)$ real matrices.

\mathbb{M}_n^+ : set of all the positive-definite symmetric $(n \times n)$ real matrices.

$\delta_{kk'}$: Kronecker's symbol.

δ_{0_v} and $\delta_{0_{\mathbb{M}_{v,N}}}$: Dirac measure at the origin of \mathbb{R}^v and of $\mathbb{M}_{v,N}$.

$[I_n]$: identity matrix in \mathbb{M}_n .

$[x]^T$: transpose of matrix $[x]$.

$\text{Tr}\{[x]\}$: trace of the square matrix $[x]$.

*Corresponding author: C. Soize, christian.soize@univ-eiffel.fr

Email addresses: christian.soize@univ-eiffel.fr (Christian SOIZE), ghanem@usc.edu (Roger GHANEM)

$\langle [x], [y] \rangle_F = \text{Tr} \{ [x]^T [y] \}$, inner product of matrices $[x]$ and $[y]$ in $\mathbb{M}_{n,m}$.
 $\|x\|$ or $\|[x]\|$: Frobenius norm of matrix $[x]$ such that $\|x\|^2 = \langle [x], [x] \rangle_F$.
 E : mathematical expectation.
 $\mathbf{1}$: vector $(1, \dots, 1) \in \mathbb{R}^N$.

1. Introduction

In this paper, novel mathematical results are presented for justifying and clarifying the methodology of *probabilistic learning on manifolds* (PLoM), initially introduced in [1]. The proposed PLoM, which can be viewed either as a supervised or an unsupervised machine learning method, considers a given initial dataset constituted of N given points $\boldsymbol{\eta}_d^1, \dots, \boldsymbol{\eta}_d^N$ in \mathbb{R}^v , which are interpreted as independent realizations of a \mathbb{R}^v -valued random variable \mathbf{H} for which its non-Gaussian probability measure $p_{\mathbf{H}}(\boldsymbol{\eta}) d\boldsymbol{\eta}$ on \mathbb{R}^v is unknown but is, *a priori*, concentrated in an unknown subset of \mathbb{R}^v . In this paper, a quantity indexed by d means that this quantity is relative to the initial dataset and consequently, is a given quantity. Denoting by $p_{\mathbf{H}}^{(N)}$ the nonparametric statistical estimation of $p_{\mathbf{H}}$, the sequence of probability measures $\{p_{\mathbf{H}}^{(N)}(\boldsymbol{\eta}) d\boldsymbol{\eta}\}_N$ on \mathbb{R}^v is convergent to $p_{\mathbf{H}}(\boldsymbol{\eta}) d\boldsymbol{\eta}$ for $N \rightarrow +\infty$. In the PLoM, the number N of data points is fixed and is presumed to be relatively small (case for which only small data are available in opposite to the big-data case). Nevertheless, it is assumed that N is larger than some lower bound N_0 needed for to be a sufficiently accurate estimate of $p_{\mathbf{H}}(\boldsymbol{\eta}) d\boldsymbol{\eta}$. Let us now define the random vector $\mathbf{H}^{(N)}$ such that its probability measure is $p_{\mathbf{H}}^{(N)}(\boldsymbol{\eta}) d\boldsymbol{\eta}$. We define the random matrix $[\mathbf{H}^N] = [\mathbf{H}^1 \dots \mathbf{H}^N]$ with values in $\mathbb{M}_{v,N}$, whose columns $\mathbf{H}^1, \dots, \mathbf{H}^N$ are N independent copies of $\mathbf{H}^{(N)}$. The matrix $[\eta_d] = [\boldsymbol{\eta}_d^1 \dots \boldsymbol{\eta}_d^N] \in \mathbb{M}_{v,N}$ is then interpreted as one realization of random matrix $[\mathbf{H}^N]$. A reduced-order diffusion-maps basis $[g_m] \in \mathbb{M}_{N,m}$ of order $m < N$ is introduced by the authors for constructing a $\mathbb{M}_{v,N}$ -valued reduced-order representation $[\mathbf{H}_m^N] = [\mathbf{Z}_m][g_m]$ of random matrix $[\mathbf{H}^N]$. A MCMC generator of the random matrix $[\mathbf{Z}_m]$ with values $\mathbb{M}_{v,m}$ is explicitly constructed as a reduced-order Itô stochastic differential equation (ISDE) associated with a dissipative Hamiltonian dynamical system. We then consider the family $\{p_{[\mathbf{H}_m^N]}([\boldsymbol{\eta}]) d[\boldsymbol{\eta}]\}_{1 \leq m \leq N}$ of probability measures on $\mathbb{M}_{v,N}$ for which the reduced-order ISDE is the MCMC generator. We prove that there exists an optimal value $m_{\text{opt}} < N$ such that the probability measure $p_{[\mathbf{H}_{m_{\text{opt}}}^N]}([\boldsymbol{\eta}]) d[\boldsymbol{\eta}]$ allows for generating an arbitrary number $n_{\text{MC}} \gg N$ of independent realizations of $[\mathbf{H}_{m_{\text{opt}}}^N]$ (the learned dataset) in preserving the concentration of the measure. This property is shown by analyzing the function $m \mapsto d_N^2(m) = E\{\|[\mathbf{H}_m^N] - [\eta_d]\|^2\} / E\{\|\eta_d\|^2\}$, which is minimized for $m = m_{\text{opt}}$ and such that $d_N^2(m_{\text{opt}}) \ll d_N^2(N)$. It should be noted that, for $m = N$, the value $d_N^2(N)$ represents the distance of the random matrix $[\mathbf{H}_N^N]$ to the initial dataset $[\eta_d]$, for which the learned dataset would be generated without using the PLoM and consequently, would involve a scattering of the generated realizations corresponding to a loss of concentration.

In the formulation proposed and analyzed, N is fixed. There is *a priori* no sense in studying the convergence of the distance for N going towards infinity, on the one hand because N has a limited value that is supposed to be rather small and on the other hand because N also represents the number of columns of the random matrix $[\mathbf{H}^N]$. However, for a fixed small value of N , the convergence of the probabilistic learning with respect to N can be considered by introducing an ordered subset of integers $N_1 < N_2 < \dots < N_{i_{\text{max}}}$ with $N_1 > 1$ and $N_{i_{\text{max}}} = N$, and by studying the convergence as a function of N_i for $i = 1, \dots, i_{\text{max}}$. If the convergence is reached for $i \leq i_{\text{max}}$, then the learning process is successful; if not, this means that the value of N is too small and has to be increased, that is to say, by increasing the number of points in the initial dataset. This question is outside the scope of this paper and we refer the reader to the references given in the first paragraph of this introduction, references in which this question is dealt with.

1.1. Framework and objective of the PLoM

In the framework of supervised machine learning, a typical problem for the use of the PLoM is the following. Let $(\mathbf{w}, \mathbf{u}) \mapsto \mathbf{f}(\mathbf{w}, \mathbf{u})$ be any measurable mapping on $\mathbb{R}^{n_w} \times \mathbb{R}^{n_u}$ with values in \mathbb{R}^{n_q} representing a computational model coming, for instance, from the discretization of a boundary value problem, in which n_w , n_u , and n_q are any finite integers. Let \mathbf{W} and \mathbf{U} be two independent (non-Gaussian) random variables defined on a probability space $(\Theta, \mathcal{T}, \mathcal{P})$ with values in \mathbb{R}^{n_w} and \mathbb{R}^{n_u} , for which the probability measures $P_{\mathbf{W}}(d\mathbf{w}) = p_{\mathbf{W}}(\mathbf{w}) d\mathbf{w}$ and $P_{\mathbf{U}}(d\mathbf{u}) = p_{\mathbf{U}}(\mathbf{u}) d\mathbf{u}$ are defined by the probability density functions $p_{\mathbf{W}}$ and $p_{\mathbf{U}}$ with respect to the Lebesgue measures $d\mathbf{w}$ and $d\mathbf{u}$ on \mathbb{R}^{n_w} and \mathbb{R}^{n_u} . Random vector \mathbf{W} is made up of a part of the random parameters of the computational model, which are used for controlling the system, while random vector \mathbf{U} is made up of the other part of these random parameters, which

are not used for controlling the system. Let \mathbf{Q} be the quantities of interest (QoI) that is a random variable defined on $(\Theta, \mathcal{T}, \mathcal{P})$ with values in \mathbb{R}^{n_q} such that

$$\mathbf{Q} = \mathbf{f}(\mathbf{W}, \mathbf{U}). \quad (1)$$

Let us assume that N calculations have been performed with the computational model (the training) whose solution is represented by Eq. (1), allowing N independent realizations $\{\mathbf{q}^j, j = 1, \dots, N\}$ of \mathbf{Q} to be computed such that $\mathbf{q}^j = \mathbf{f}(\mathbf{w}^j, \mathbf{u}^j)$, in which $\{\mathbf{w}^j, j = 1, \dots, N\}$ and $\{\mathbf{u}^j, j = 1, \dots, N\}$ are N independent realizations of (\mathbf{W}, \mathbf{U}) , which have been generated using an adapted generator for $p_{\mathbf{W}}$ and $p_{\mathbf{U}}$. We then consider the random variable \mathbf{X} with values in \mathbb{R}^n , such that

$$\mathbf{X} = (\mathbf{Q}, \mathbf{W}) \quad , \quad n = n_q + n_w. \quad (2)$$

The probabilistic learning is performed for \mathbf{X} . The unnormalized initial dataset \mathcal{D}_N related to random vector \mathbf{X} is then made up of the N independent realizations $\{\mathbf{x}^j, j = 1, \dots, N\}$ in which $\mathbf{x}^j = (\mathbf{q}^j, \mathbf{w}^j) \in \mathbb{R}^n$. In this paper, it is assumed that the measurable mapping \mathbf{f} is such that the non-Gaussian probability measure $P_{\mathbf{X}}(d\mathbf{x})$ of $\mathbf{X} = (\mathbf{Q}, \mathbf{W})$ admits a density $p_{\mathbf{X}}(\mathbf{x})$ with respect to the Lebesgue measure $d\mathbf{x}$ on \mathbb{R}^n . The probability measure of \mathbf{X} is unknown and is assumed to be concentrated in a subset of \mathbb{R}^n that is also unknown (this concentration property is due to Eqs. (1) and (2)). The objective of the PLoM proposed in [1] is to construct a probabilistic model of non-Gaussian random vector \mathbf{X} using only the unnormalized initial dataset \mathcal{D}_N , which allows for generating $\nu_{\text{sim}} \gg N$ additional independent realizations $\{\mathbf{x}_{\text{ar}}^1, \dots, \mathbf{x}_{\text{ar}}^{\nu_{\text{sim}}}\}$ in \mathbb{R}^n of random vector \mathbf{X} , preserving the concentration of its probability measure and without using the computational model. It can then be deduced ν_{sim} additional realizations $\{(\mathbf{q}_{\text{ar}}^\ell, \mathbf{w}_{\text{ar}}^\ell), \ell = 1, \dots, \nu_{\text{sim}}\}$ that are such that $(\mathbf{q}_{\text{ar}}^\ell, \mathbf{w}_{\text{ar}}^\ell) = \mathbf{x}_{\text{ar}}^\ell$. These additional realizations allow, for instance, a cost function $\mathbb{J}(\mathbf{w}) = E\{J(\mathbf{Q}, \mathbf{W}) | \mathbf{W} = \mathbf{w}\}$ to be evaluated, in which $(\mathbf{q}, \mathbf{w}) \mapsto J(\mathbf{q}, \mathbf{w})$ is a given measurable real-valued mapping on $\mathbb{R}^{n_q} \times \mathbb{R}^{n_w}$ as well as constraints related to a nonconvex optimization problem [2] and this, without calling the computational model.

1.2. Organization of the paper and its what are the main results

In Section 2, we introduce the \mathbb{R}^ν -valued random variable \mathbf{H} resulting from the principal component analysis (PCA) of the \mathbb{R}^n -valued random variable \mathbf{X} with $\nu \leq n$. Section 3 is devoted to the nonparametric statistical estimate $p_{\mathbf{H}}^{(N)}$ of the pdf $p_{\mathbf{H}}$ of \mathbf{H} and we give Theorem 1 concerning the consistency of the sequence of estimators of $p_{\mathbf{H}}(\boldsymbol{\eta})$ for all $\boldsymbol{\eta}$ fixed in \mathbb{R}^ν . Section 4 deals with the definition of random matrix $[\mathbf{H}^N]$ and Proposition 1 gives an explicit expression of the pdf $p_{[\mathbf{H}^N]}$ of random matrix $[\mathbf{H}^N]$. In Section 5, we present the construction of the reduced-order diffusion-maps basis $[g_m]$ that is used by the PLoM method and we introduce the estimation of the optimal values ε_{opt} and m_{opt} of the hyperparameter ε_{DM} and of the reduced order m . In particular, we compare the hyperparameter ε_{DM} and the modified Silverman bandwidth \hat{s} ; we conclude that the invariant probability measure $p^{\text{eDM}}(i)$ of the Markov chain allowing the diffusion-maps basis to be constructed is different from the probability measure $p_{\mathbf{H}}^{(N)}(\boldsymbol{\eta}) d\boldsymbol{\eta}$ of random vector $\mathbf{H}^{(N)}$ that is considered by the PLoM. Section 6 is devoted to the construction of the probability measure and its generator related to the probabilistic learning on manifolds. We introduce the reduced-order representation $[\mathbf{H}_m^N] = [\mathbf{Z}_m][g_m]^T$ of random matrix $[\mathbf{H}^N]$. In a first central Theorem 3, we prove that the transported probability measure $p_{[\mathbf{Z}_m]}([z]) d[z]$ of random matrix $[\mathbf{Z}_m]$ is the marginal distribution of the invariant measure of the reduced-order ISDE that is used as the MCMC generator of random matrix $[\mathbf{Z}_m]$. We also prove in Proposition 2 that $p_{[\mathbf{Z}_m]}$ has a "Gaussian representation", which is a linear combination of N^N products of ν Gaussian pdf on \mathbb{R}^N . Consequently, the use of Theorem 3 effectively allows realizations of random matrix $[\mathbf{Z}_m]$ to be generated, while a Gaussian generator that would be based on the Gaussian representation is unthinkable for $N > 10$, for instance. Section 7 deals with the square of the relative distance $d_N^2(m)$ of random matrix $[\mathbf{H}_m^N]$ to matrix $[\eta_d]$ of the initial dataset. This distance allows for quantifying, as a function of m , the concentration of the measure $p_{[\mathbf{H}_m^N]}([\eta]) d[\eta]$ in the subset of $\mathbb{M}_{\nu, N}$ where the initial dataset (represented by $[\eta_d]$) is located. We show that the usual MCMC generator of random matrix $[\mathbf{H}^N]$ corresponding to $m = N$, yields $d_N^2(N) \simeq 2$ (see Lemma 2), and induces a loss of concentration of the probability measure. Under a "reasonable hypothesis", the second central Theorem 4 proves that $d_N^2(m_{\text{opt}}) \ll d_N^2(N)$ in which $m_{\text{opt}} < N$ is the optimal value of m . This result demonstrates that the PLoM method is a better method than the usual one because it keeps the concentration of the measure. In Section 8, we present a justification of the hypothesis introduced in Theorem 4, based on the use of the maximum entropy principle from Information Theory. Section 9 is devoted to a brief numerical application that illustrates the mathematical results. The conclusions follow in Section 10.

The novelty of the mathematical results presented are mainly those of Section 6.3 related to the explicit expression of pdf $p_{[\mathbf{z}_m]}$ (Proposition 2) and its connection to the invariant measure of the reduced-order ISDE (Theorem 3), and those of Section 7 related to the characterization of the concentration of the probability measure of random matrix $[\mathbf{H}_m^N]$ by introducing distance $d_N(m)$ (Propositions 3 and 4), and to the existence of a minimum of this distance for an optimal value $m_{\text{opt}} < N$ of m (Theorem 4), which demonstrates that the PLoM method allows the concentration of the measure to be kept.

2. De-correlation and normalization of random vector \mathbf{X} by PCA

The first step of the PLoM consists in performing a principal component analysis (PCA) of \mathbf{X} in order to obtain a scaled random vector \mathbf{H} through de-correlation and normalization. Let $\hat{\mathbf{x}} \in \mathbb{R}^n$ and $[\hat{\mathbf{C}}] \in \mathbb{M}_n^{+0}$ be the classical empirical estimates of the mean vector and the covariance matrix of \mathbf{X} , constructed using \mathcal{D}_N . Let $[\hat{\boldsymbol{\mu}}] \in \mathbb{M}_v$ be the diagonal matrix of the first v eigenvalues $\hat{\mu}_1 \geq \hat{\mu}_2 \geq \dots \geq \hat{\mu}_v > 0$ of $[\hat{\mathbf{C}}]$ and let $[\hat{\Phi}] \in \mathbb{M}_{n,v}$ be the matrix of the associated orthonormal eigenvectors. For any $\varepsilon > 0$ fixed, $v \leq n$ is chosen such that $\text{err}_{\text{PCA}}(v) = 1 - (\hat{\mu}_1 + \dots + \hat{\mu}_v) / (\text{Tr}[\hat{\mathbf{C}}]) \leq \varepsilon$. This PCA allows for representing \mathbf{X} by

$$\mathbf{X}^v = \hat{\mathbf{x}} + [\hat{\Phi}] [\hat{\boldsymbol{\mu}}]^{1/2} \mathbf{H} \quad , \quad E\{\|\mathbf{X} - \mathbf{X}^v\|^2\} \leq \varepsilon E\{\|\mathbf{X}\|^2\}. \quad (3)$$

Throughout this paper, it will be assumed that $v < N$. The N independent realizations $\{\boldsymbol{\eta}_d^j, j = 1, \dots, N\}$ of the second-order \mathbb{R}^v -valued random variable \mathbf{H} defined on probability space $(\Theta, \mathcal{T}, \mathcal{P})$ are such that

$$\boldsymbol{\eta}_d^j = [\hat{\boldsymbol{\mu}}]^{-1/2} [\hat{\Phi}]^T (\mathbf{x}^j - \hat{\mathbf{x}}) \in \mathbb{R}^v \quad , \quad j = 1, \dots, N. \quad (4)$$

The initial dataset D_N related to random vector \mathbf{H} is then defined as $D_N = \{\boldsymbol{\eta}_d^1, \dots, \boldsymbol{\eta}_d^N\}$.

Let $[\eta_d] = [\boldsymbol{\eta}_d^1 \dots \boldsymbol{\eta}_d^N] \in \mathbb{M}_{v,N}$ be the matrix of the N realizations of \mathbf{H} . The empirical estimates $\mathbf{m}_N \in \mathbb{R}^v$ and $[C_N] \in \mathbb{M}_v^+$ of the mean vector and the covariance matrix of \mathbf{H} , are such that

$$\mathbf{m}_N = \frac{1}{N} \sum_{j=1}^N \boldsymbol{\eta}_d^j = \mathbf{0}_v \quad , \quad [C_N] = \frac{1}{N-1} [\eta_d] [\eta_d]^T = [I_v]. \quad (5)$$

It can be seen that the Frobenius norm $\|\eta_d\|$ of matrix $[\eta_d] \in \mathbb{M}_{v,N}$ is such that

$$\|\eta_d\|^2 = \text{Tr}\{[\eta_d]^T [\eta_d]\} = \sum_{j=1}^N \|\boldsymbol{\eta}_d^j\|^2 = v(N-1). \quad (6)$$

3. Nonparametric estimate of the pdf of \mathbf{H}

As proposed in [1], the modification of the multidimensional Gaussian kernel-density estimation method [3, 4, 5, 6] is used for constructing the estimation $p_{\mathbf{H}}^{(N)}$ on \mathbb{R}^v of the pdf $p_{\mathbf{H}}$ of random vector \mathbf{H} , which is written, $\forall \boldsymbol{\eta} \in \mathbb{R}^v$, as

$$p_{\mathbf{H}}^{(N)}(\boldsymbol{\eta}) = \frac{1}{N} \sum_{j=1}^N \pi_{v,N}(\frac{\hat{s}}{s} \boldsymbol{\eta}_d^j - \boldsymbol{\eta}) \quad , \quad \pi_{v,N}(\boldsymbol{\eta}) = \frac{1}{(\sqrt{2\pi} \hat{s})^v} \exp\{-\frac{1}{2\hat{s}^2} \|\boldsymbol{\eta}\|^2\}, \quad (7)$$

$$s = \left(\frac{4}{N(v+2)} \right)^{1/(v+4)} \quad , \quad \hat{s} = \frac{s}{\sqrt{s^2 + \frac{N-1}{N}}}, \quad (8)$$

in which s is the usual Silverman bandwidth (since $[C_N] = [I_v]$) (see for instance, [7]) and where \hat{s} has been introduced in order that $\int_{\mathbb{R}^v} \boldsymbol{\eta} p_{\mathbf{H}}^{(N)}(\boldsymbol{\eta}) d\boldsymbol{\eta} = \mathbf{0}_v$ and that $\int_{\mathbb{R}^v} \boldsymbol{\eta} \otimes \boldsymbol{\eta} p_{\mathbf{H}}^{(N)}(\boldsymbol{\eta}) d\boldsymbol{\eta} = [I_v]$, because, in the framework of the PLoM, we need to preserve the centering and the orthogonality property. Finally, for fixed v ,

$$\text{if } N \rightarrow +\infty, \quad \text{then } s \rightarrow 0, \quad \hat{s} \rightarrow 0, \quad \frac{\hat{s}}{s} \rightarrow 1, \quad \frac{s}{\hat{s}} \rightarrow 1. \quad (9)$$

Theorem 1 (Sequence of estimators of $p_{\mathbf{H}}(\boldsymbol{\eta})$ [8]). Let us assume that $p_{\mathbf{H}}$ is continuous on \mathbb{R}^{ν} . For ν fixed and for $\boldsymbol{\eta}$ given in \mathbb{R}^{ν} , let $\{P^{(N)}(\boldsymbol{\eta})\}_N$ be the sequence of estimators of $p_{\mathbf{H}}(\boldsymbol{\eta})$ for which $P^{(N)}(\boldsymbol{\eta}) = \frac{1}{N} \sum_{j=1}^N \pi_{\nu,N}(\frac{\widehat{\mathbf{H}}^j - \boldsymbol{\eta}}{s})$ is a positive-valued random variable where $\widehat{\mathbf{H}}^1, \dots, \widehat{\mathbf{H}}^N$ are N independent copies of \mathbf{H} . Thus, $\forall \boldsymbol{\eta} \in \mathbb{R}^{\nu}$, the mean value $\underline{P}^{(N)}(\boldsymbol{\eta}) = E\{P^{(N)}(\boldsymbol{\eta})\}$ of $P^{(N)}(\boldsymbol{\eta})$ is such that $\lim_{N \rightarrow +\infty} \underline{P}^{(N)}(\boldsymbol{\eta}) = p_{\mathbf{H}}(\boldsymbol{\eta})$ and the variance is such that $\text{Var}\{P^{(N)}(\boldsymbol{\eta})\} = E\{(P^{(N)}(\boldsymbol{\eta}) - \underline{P}^{(N)}(\boldsymbol{\eta}))^2\} \leq N^{-4/(\nu+4)} \beta_{\nu,N} \underline{P}^{(N)}(\boldsymbol{\eta})$, in which, for ν fixed and for $N \rightarrow +\infty$, the positive constant $\beta_{\nu,N}$ is such that $\beta_{\nu,N} \sim (2\pi)^{-\nu/2} ((2 + \nu)/4)^{\nu/(\nu+4)}$.

PROOF. The proof, inspired from [8], is adapted to the modification $\widehat{s} \neq s$ used for defining the estimator. Since $p_{\mathbf{H}}$ is assumed to be a continuous function, $\forall \boldsymbol{\eta} \in \mathbb{R}^{\nu}$, $p_{\mathbf{H}}(\boldsymbol{\eta}) = E\{\delta_{0,\nu}(\mathbf{H} - \boldsymbol{\eta})\}$. Using the second Eq. (7), for all $\widetilde{\boldsymbol{\eta}}$ and $\boldsymbol{\eta}$ in \mathbb{R}^{ν} , we have $\pi_{\nu,N}(\frac{\widehat{s}}{s} \widetilde{\boldsymbol{\eta}} - \boldsymbol{\eta}) d\widetilde{\boldsymbol{\eta}} = (s/\widehat{s})^{\nu} (\sqrt{2\pi}s)^{-\nu} \exp\{-\frac{1}{2s^2} \|\widetilde{\boldsymbol{\eta}} - \frac{s}{\widehat{s}} \boldsymbol{\eta}\|^2\} d\widetilde{\boldsymbol{\eta}}$. Using Eq. (9) yields the following equality in the space of the bounded measures on \mathbb{R}^{ν} , $\lim_{N \rightarrow +\infty} \pi_{\nu,N}(\frac{\widehat{s}}{s} \widetilde{\boldsymbol{\eta}} - \boldsymbol{\eta}) d\widetilde{\boldsymbol{\eta}} = \delta_{0,\nu}(\widetilde{\boldsymbol{\eta}} - \boldsymbol{\eta})$. Since $\widehat{\mathbf{H}}^1, \dots, \widehat{\mathbf{H}}^N$ are independent copies of \mathbf{H} , we have $\underline{P}^{(N)}(\boldsymbol{\eta}) = E\{\pi_{\nu,N}(\frac{\widehat{s}}{s} \mathbf{H} - \boldsymbol{\eta})\} = \int_{\mathbb{R}^{\nu}} \pi_{\nu,N}(\frac{\widehat{s}}{s} \widetilde{\boldsymbol{\eta}} - \boldsymbol{\eta}) p_{\mathbf{H}}(\widetilde{\boldsymbol{\eta}}) d\widetilde{\boldsymbol{\eta}}$. Using the two last above equations yields the expression for the mean. Similarly, $E\{(P^{(N)}(\boldsymbol{\eta}))^2\} = \frac{1}{N} E\{(\pi_{\nu,N}(\frac{\widehat{s}}{s} \mathbf{H} - \boldsymbol{\eta}))^2\} + (1 - \frac{1}{N}) (\underline{P}^{(N)}(\boldsymbol{\eta}))^2$. Consequently, $\text{Var}\{P^{(N)}(\boldsymbol{\eta})\} = \frac{1}{N} E\{(\pi_{\nu,N}(\frac{\widehat{s}}{s} \mathbf{H} - \boldsymbol{\eta}))^2\} - \frac{1}{N} (\underline{P}^{(N)}(\boldsymbol{\eta}))^2 \leq \frac{1}{N} E\{(\pi_{\nu,N}(\frac{\widehat{s}}{s} \mathbf{H} - \boldsymbol{\eta}))^2\} \leq \frac{1}{N} (\sup_{\widetilde{\boldsymbol{\eta}}} \pi_{\nu,N}(\frac{\widehat{s}}{s} \widetilde{\boldsymbol{\eta}} - \boldsymbol{\eta})) E\{\pi_{\nu,N}(\frac{\widehat{s}}{s} \mathbf{H} - \boldsymbol{\eta})\} = \frac{1}{N} (\sqrt{2\pi}\widehat{s})^{-\nu} \underline{P}^{(N)}(\boldsymbol{\eta})$. From Eq. (8) and the last inequality yield the expression for the variance in which $\beta_{\nu,N} = (2\pi)^{-\nu/2} \{(2 + \nu)/4\}^{\nu/(\nu+4)} \{1 - 1/N\}^{\nu/2} \{1 + (4/(2 + \nu))^{2/(\nu+4)} N^{-2/(\nu+4)} (1 - 1/N)^{-1}\}^{\nu/2}$ that is the expression given in the theorem for N sufficiently large.

Remark 1 (Properties of the sequence of estimators of $p_{\mathbf{H}}(\boldsymbol{\eta})$). Theorem 1 shows that estimator $P^{(N)}(\boldsymbol{\eta})$ is asymptotically unbiased. Since $\forall \boldsymbol{\eta} \in \mathbb{R}^{\nu}$, $E\{(P^{(N)}(\boldsymbol{\eta}) - p_{\mathbf{H}}(\boldsymbol{\eta}))^2\} = \text{Var}\{P^{(N)}(\boldsymbol{\eta})\} + (\underline{P}^{(N)}(\boldsymbol{\eta}) - p_{\mathbf{H}}(\boldsymbol{\eta}))^2$, we have $\lim_{N \rightarrow +\infty} E\{(P^{(N)}(\boldsymbol{\eta}) - p_{\mathbf{H}}(\boldsymbol{\eta}))^2\} = 0$, which shows that estimator $P^{(N)}(\boldsymbol{\eta})$ is consistent. This mean-square convergence implies the convergence in probability.

4. Definition of the random matrix $[\mathbf{H}^N]$ and its pdf

The introduction of a $(\nu \times N)$ random matrix $[\mathbf{H}^N]$ will allow the initial dataset D_N to be represented using the diffusion-maps basis.

Definition 1 (Matrices $[\boldsymbol{\eta}]$, $[\eta_d]$, $[\eta_d(\mathbf{j})]$, and set \mathcal{J}). Let $[\boldsymbol{\eta}]$ be any matrix in $\mathbb{M}_{\nu,N}$ that is written as

$$[\boldsymbol{\eta}] = [\boldsymbol{\eta}^1 \dots \boldsymbol{\eta}^N] \in \mathbb{M}_{\nu,N} \quad , \quad \boldsymbol{\eta}^{\ell} = (\eta_{1,\ell}^{\ell}, \dots, \eta_{\nu,\ell}^{\ell}) \in \mathbb{R}^{\nu} \quad , \quad \ell = 1, \dots, N, \quad (10)$$

and let $d[\boldsymbol{\eta}] = \otimes_{\ell=1}^N d\boldsymbol{\eta}^{\ell}$ be the measure on $\mathbb{M}_{\nu,N}$ induced by the Lebesgue measures $d\boldsymbol{\eta}^1, \dots, d\boldsymbol{\eta}^N$ on \mathbb{R}^{ν} . Let $[\eta_d] \in \mathbb{M}_{\nu,N}$ be the matrix constructed using the N points $\boldsymbol{\eta}^j \in \mathbb{R}^{\nu}$ defined by Eq. (4),

$$[\eta_d] = [\boldsymbol{\eta}_d^1 \dots \boldsymbol{\eta}_d^N] \in \mathbb{M}_{\nu,N} \quad , \quad \boldsymbol{\eta}_d^j = (\eta_{d,1}^j, \dots, \eta_{d,\nu}^j) \in \mathbb{R}^{\nu} \quad , \quad j = 1, \dots, N. \quad (11)$$

Let $\mathbf{j} = (j_1, \dots, j_N) \in \mathcal{J}$ be the multi-index of dimension N with $\mathcal{J} = \{1, 2, \dots, N\}^N \subset \mathbb{N}^N$. For all \mathbf{j} in \mathcal{J} , the matrix $[\eta_d(\mathbf{j})] \in \mathbb{M}_{\nu,N}$ is defined by

$$[\eta_d(\mathbf{j})]_{k\ell} = \eta_{d,k}^{j_{\ell}} \quad , \quad k = 1, \dots, \nu \quad , \quad \ell = 1, \dots, N. \quad (12)$$

Finally, we will use the following notation, $\sum_{\mathbf{j} \in \mathcal{J}} = \sum_{j_1=1}^N \dots \sum_{j_N=1}^N$.

Note that matrix $[\eta_d]$ defined by Eq. (11) has to carefully be distinguished from matrix $[\eta_d(\mathbf{j})]$ defined by Eq. (12). Nevertheless, it can be seen that for $\mathbf{j}_0 = (1, 2, \dots, N) \in \mathcal{J}$, we have $[\eta_d(\mathbf{j}_0)] = [\eta_d]$.

Definition 2 (Random matrix $[\mathbf{H}^N]$). Let $\mathbf{H}^{(N)}$ be the \mathbb{R}^{ν} -valued random variable defined on $(\Theta, \mathcal{T}, \mathcal{P})$ for which the pdf is $p_{\mathbf{H}}^{(N)}$ defined by Eqs. (7) and (8). We then define the random matrix $[\mathbf{H}^N]$ with values in $\mathbb{M}_{\nu,N}$ such that $[\mathbf{H}^N] = [\mathbf{H}^1 \dots \mathbf{H}^N]$ in which $\mathbf{H}^1, \dots, \mathbf{H}^N$ are N independent copies of $\mathbf{H}^{(N)}$. From Section 3, it can be seen that $E\{\mathbf{H}^{(N)}\} = 0_{\nu}$ and that $E\{\mathbf{H}^{(N)} \otimes \mathbf{H}^{(N)}\} = [I_{\nu}]$.

Note that in Definition 2, $\mathbf{H}^1, \dots, \mathbf{H}^N$ are not taken as N independent copies of \mathbf{H} whose pdf $p_{\mathbf{H}}$ is unknown, but are taken as N independent copies of $\mathbf{H}^{(N)}$ whose pdf $p_{\mathbf{H}}^{(N)}$ is known.

Proposition 1 (Probability density function of random matrix $[\mathbf{H}^N]$). *The probability measure of random matrix $[\mathbf{H}^N]$ with values in $\mathbb{M}_{v,N}$ admits the following density $[\eta] \mapsto p_{[\mathbf{H}^N]}([\eta])$ on $\mathbb{M}_{v,N}$ with respect to $d[\eta]$,*

$$p_{[\mathbf{H}^N]}([\eta]) = \prod_{\ell=1}^N \left\{ \frac{1}{N} \sum_{j=1}^N \frac{1}{(\sqrt{2\pi} \hat{s})^v} \exp\left\{-\frac{1}{2\hat{s}^2} \left\| \frac{\hat{s}}{s} \boldsymbol{\eta}_d^j - \boldsymbol{\eta}^\ell \right\|^2\right\} \right\}. \quad (13)$$

PROOF. Using Definition 2 yields, for all $[\eta]$ in $\mathbb{M}_{v,N}$, $p_{[\mathbf{H}^N]}([\eta]) = \prod_{\ell=1}^N p_{\mathbf{H}}^{(N)}(\boldsymbol{\eta}^\ell)$ and using Eq. (7) yields Eq. (13).

5. Construction of a reduced-order diffusion-maps basis

To identify the subset around which the initial data are concentrated, the PLoM method relies on the diffusion-maps method [9, 10, 11, 12]. We use the Gaussian kernel such that, for all $\boldsymbol{\eta}$ and $\boldsymbol{\eta}'$ in \mathbb{R}^v , $k_{\varepsilon_{\text{DM}}}(\boldsymbol{\eta}, \boldsymbol{\eta}') = \exp\{-(4\varepsilon_{\text{DM}})^{-1} \|\boldsymbol{\eta} - \boldsymbol{\eta}'\|^2\}$ in which $\varepsilon_{\text{DM}} > 0$. The matrices $[K]$ and $[b]$ are defined, for all i and j in $\{1, \dots, N\}$, by $[K]_{ij} = \exp\{-(4\varepsilon_{\text{DM}})^{-1} \|\boldsymbol{\eta}_d^i - \boldsymbol{\eta}_d^j\|^2\}$ and $[b]_{ij} = \delta_{ij} b_i$ with $b_i = \sum_{j=1}^N [K]_{ij}$. It is assumed that $[\eta_d]$ is such that $[K] \in \mathbb{M}_N^+$. Hence, the diagonal matrix $[b]$ belongs to \mathbb{M}_N^+ . Let $[\mathbb{P}] = [b]^{-1}[K] \in \mathbb{M}_N$ be the non symmetric matrix with positive entries such that $\sum_j [\mathbb{P}]_{ij} = 1$ for all i . Matrix $[\mathbb{P}]$ is the transition matrix of a Markov chain that yields the probability of transition in one step.

5.1. Diffusion-maps basis as a non orthogonal vector basis in \mathbb{R}^N

The eigenvalues $\lambda_1, \dots, \lambda_N$ and the associated eigenvectors $\boldsymbol{\psi}^1, \dots, \boldsymbol{\psi}^N$ of the right-eigenvalue problem $[\mathbb{P}] \boldsymbol{\psi}^\alpha = \lambda_\alpha \boldsymbol{\psi}^\alpha$ are such that $1 = \lambda_1 > \lambda_2 \geq \dots \geq \lambda_N$ and can be computed by solving the generalized eigenvalue problem $[K] \boldsymbol{\psi}^\alpha = \lambda_\alpha [b] \boldsymbol{\psi}^\alpha$ with the normalization $\langle [b] \boldsymbol{\psi}^\alpha, \boldsymbol{\psi}^\beta \rangle = \delta_{\alpha\beta}$. The eigenvector $\boldsymbol{\psi}^1$ associated with $\lambda_1 = 1$ is a constant vector that can be written as $\boldsymbol{\psi}^1 = N^{-1/2} \|\boldsymbol{\psi}^1\| \mathbf{1}$ with $\mathbf{1} = (1, \dots, 1) \in \mathbb{R}^N$.

Definition 3 (Reduced-order diffusion-maps basis $[g_m]$ of order m). *For a given integer $\kappa \geq 0$, the diffusion-maps basis $\{\mathbf{g}^1, \dots, \mathbf{g}^\alpha, \dots, \mathbf{g}^N\}$ is a vector basis of \mathbb{R}^N defined by $\mathbf{g}^\alpha = \lambda_\alpha^\kappa \boldsymbol{\psi}^\alpha$ such that $\langle [b] \mathbf{g}^\alpha, \mathbf{g}^\beta \rangle = \lambda_\alpha^{2\kappa} \delta_{\alpha\beta}$. For a given integer m with $2 < m \leq N$, we define the reduced-order diffusion-maps basis of order m as the family $\{\mathbf{g}^1, \dots, \mathbf{g}^m\}$ that we represent by the matrix $[g_m] = [\mathbf{g}^1 \dots \mathbf{g}^m] \in \mathbb{M}_{N,m}$ with $\mathbf{g}^\alpha = (g_\alpha^1, \dots, g_\alpha^m)$ and $[g_m]_{\ell\alpha} = g_\alpha^\ell$.*

Note that $\{\mathbf{g}^\alpha\}_\alpha$ is not orthogonal for the inner product $\langle \cdot, \cdot \rangle$, but is orthogonal for the one defined by $(\mathbf{u}, \mathbf{v}) \mapsto \langle [b] \mathbf{u}, \mathbf{v} \rangle$ on $\mathbb{R}^N \times \mathbb{R}^N$. It can also be seen that the construction of the reduced-order diffusion-maps basis $[g_m]$ depends, *a priori*, on three parameters: the smoothing parameter ε_{DM} , the order m , and the integer κ . Nevertheless, we will see in Section 5.4 that κ has not a role from a theoretical point of view in the proposed method, in contrary to the one used in [9]. In the PLoM, its role is the one of an additional scaling; its value can therefore be fixed arbitrarily (for instance, it can be set to 1 or even to 0; in the latter case, we have $\mathbf{g}^\alpha = \boldsymbol{\psi}_\alpha$). As a result, the only two parameters that will be considered will be ε_{DM} and m .

5.2. Estimation of the optimal values ε_{opt} and m_{opt} of ε_{DM} and m

Hypothesis 1 (On the initial data represented by matrix $[\eta_d]$). *For a given matrix $[\eta_d]$, the eigenvalues λ_α depend on ε_{DM} . It is assumed that there exist a value ε_{opt} of ε_{DM} and a value $m_{\text{opt}} > 2$ of integer m such that $1 = \lambda_1 > \lambda_2(\varepsilon_{\text{opt}}) \geq \dots \geq \lambda_{m_{\text{opt}}}(\varepsilon_{\text{opt}}) \gg \lambda_{m_{\text{opt}}+1}(\varepsilon_{\text{opt}}) \geq \dots \geq \lambda_N(\varepsilon_{\text{opt}}) > 0$.*

Under Hypothesis 1, the following algorithm associated with the given initial dataset is proposed for estimating the optimal value ε_{opt} of ε_{DM} and an optimal value m_{opt} of order m . Let $\varepsilon_{\text{DM}} \mapsto \widehat{m}(\varepsilon_{\text{DM}})$ be the function from $]0, +\infty[$ into \mathbb{N} such that

$$\widehat{m}(\varepsilon_{\text{DM}}) = \arg \min_{\alpha \mid \alpha \geq 3} \left\{ \frac{\lambda_\alpha(\varepsilon_{\text{DM}})}{\lambda_2(\varepsilon_{\text{DM}})} < 0.1 \right\}. \quad (14)$$

If function \widehat{m} is a decreasing (that is to say, nonincreasing) function of ε_{DM} , then the optimal value ε_{opt} of ε_{DM} can be chosen as the smallest value of the integer $\widehat{m}(\varepsilon_{\text{opt}})$ such that

$$\{\widehat{m}(\varepsilon_{\text{opt}}) < \widehat{m}(\varepsilon_{\text{DM}}), \forall \varepsilon_{\text{DM}} \in]0, \varepsilon_{\text{opt}}[\} \cap \{\widehat{m}(\varepsilon_{\text{opt}}) = \widehat{m}(\varepsilon_{\text{DM}}), \forall \varepsilon_{\text{DM}} \in]\varepsilon_{\text{opt}}, \varepsilon_{\text{max}}[\}, \quad (15)$$

in which ε_{\max} is an arbitrary value chosen sufficiently large in order that $\varepsilon_{\text{DM}} \mapsto \hat{m}(\varepsilon_{\text{DM}})$ be a flat function on the interval $]\varepsilon_{\text{opt}}, \varepsilon_{\max}[$. The corresponding optimal value m_{opt} of m is then given by $m_{\text{opt}} = \hat{m}(\varepsilon_{\text{opt}})$.

In fact, we are looking for the couple $(\varepsilon_{\text{opt}}, m_{\text{opt}})$ so that we have $1 = \lambda_1 > \lambda_2(\varepsilon_{\text{opt}}) \simeq \dots \simeq \lambda_{m_{\text{opt}}}(\varepsilon_{\text{opt}}) \gg \lambda_{m_{\text{opt}}+1}(\varepsilon_{\text{opt}}) \geq \dots \geq \lambda_N(\varepsilon_{\text{opt}}) > 0$ (see the illustration shown in Fig. 1 (left)). Consequently, Hypothesis 1 is satisfied. We have seen through all the performed numerical experiments that ε_{\max} is generally less than $1.5 \varepsilon_{\text{opt}}$.

In practice, Hypothesis 1 is easy to verify. A value of ε_{\max} is fixed *a priori*. The interval $]0, \varepsilon_{\max}[$ of the values of ε_{DM} is explored defining an ordered partition of it. For each value ε_{DM} in this partition, the eigenvalues are computed as explained in Section 5.1 and the criterion defined by Eqs. (14) and (15) are examined. Note that this estimation has been used with success for many applications (see for instance, [13, 2, 14, 15, 16, 17, 18, 19, 20]).

If function \hat{m} is not a decreasing function of ε_{DM} , then a general method based on Information Theory is proposed in [21]. Alternatively, if data come from a mixture model, the method proposed in [22] could also be used.

5.3. On the relationship between hyperparameter ε_{DM} and the modified Silverman bandwidth \hat{s}

The invariant measure associated with transition matrix $[\mathbb{P}]$ of the one-step Markov chain is $p^{\varepsilon_{\text{DM}}}(i) = b_i (\sum_{j=1}^N b_j)^{-1}$, which is such that $\sum_{i=1}^N p(j|i) p^{\varepsilon_{\text{DM}}}(i) = p^{\varepsilon_{\text{DM}}}(j)$ in which $p(j|i) = [\mathbb{P}]_{ij}$. Let us compare the measure $p^{\varepsilon_{\text{DM}}}(i) = (\sum_{j=1}^N b_j)^{-1} \sum_{j=1}^N \exp\{-(4 \varepsilon_{\text{DM}})^{-1} \|\boldsymbol{\eta}_d^i - \boldsymbol{\eta}_d^j\|^2\}$ with $p_{\mathbf{H}}^{(N)}(\boldsymbol{\eta}) d\boldsymbol{\eta}$ in which $p_{\mathbf{H}}^{(N)}(\boldsymbol{\eta})$ is defined by Eqs. (7) and (8), which is written, for N sufficiently large (that is to say for $\hat{s}/s \sim 1$ and $\hat{s} \sim s$), as $p_{\mathbf{H}}^{(N)}(\boldsymbol{\eta}) \simeq N^{-1} (\sqrt{2\pi}s)^{-\nu} \sum_{j=1}^N \exp\{-(2s^2)^{-1} \|\boldsymbol{\eta} - \boldsymbol{\eta}_d^j\|^2\}$. In general, for ν sufficiently large (for instance, $\nu \sim 10$), the optimal value ε_{opt} defined by Eqs. (14) and (15) is such that $\varepsilon_{\text{opt}} \gg 1$ while, since $\nu \leq N$, Eq. (8) shows that $s^2/2 < 1$. Therefore, $p^{\varepsilon_{\text{DM}}}(i)$ is very different from the probability measure $p_{\mathbf{H}}^{(N)}(\boldsymbol{\eta}) d\boldsymbol{\eta}$ that corresponds to an observation of the initial dataset from inside it, that is to say, for an observation at the smallest scale. In contrast, the probability measure $p^{\varepsilon_{\text{DM}}}(i)$ is the one for which the initial dataset is observed from outside it, that is to say, for an observation at a larger scale.

5.4. Properties of the reduced-order diffusion-maps basis

Definition 4 (Matrices $[a_m]$ and $[G_m]$). For all fixed m , let $[g_m] \in \mathbb{M}_{N,m}$ be the matrix defined in Definition 3. Since matrix $[g_m]^T [g_m] \in \mathbb{M}_m$ is invertible, we define the matrix $[a_m] = [g_m] ([g_m]^T [g_m])^{-1} \in \mathbb{M}_{N,m}$ and the matrix $[G_m] = [a_m] [g_m]^T = [g_m] ([g_m]^T [g_m])^{-1} [g_m]^T \in \mathbb{M}_N$.

It should be noted that, as announced at the end of Section 5.1, matrix $[G_m]$ is independent of $\lambda_1^k, \dots, \lambda_m^k$ and thus, is independent of κ .

Lemma 1 (Properties of $[G_m]$). For all m such that $1 \leq m \leq N - 1$:

- (i) $\text{rank}\{[G_m]\} = m$, $\text{Tr}\{[G_m]\} = m$, $[G_m]^T = [G_m]$, and $[G_m] \in \mathbb{M}_N^{+0}$.
- (ii) for $m = N$, we have $[G_N] = [I_N]$.
- (iii) $[G_m]^2 = [G_m]$, thus $[G_m]$ is idempotent and is a projection operator.
- (iv) the eigenvalue problem $[G_m] \boldsymbol{\varphi}^\alpha = \mu_\alpha \boldsymbol{\varphi}^\alpha$ is such that $\mu_1 = \dots = \mu_m = 1$ and $\mu_{m+1} = \dots = \mu_N = 0$. Matrix $[G_m]$ can be written as $[G_m] = \sum_{\alpha=1}^N \mu_\alpha \boldsymbol{\varphi}^\alpha \otimes \boldsymbol{\varphi}^\alpha = \sum_{\alpha=1}^m \boldsymbol{\varphi}^\alpha \otimes \boldsymbol{\varphi}^\alpha$ in which the eigenvectors are such that $\langle \boldsymbol{\varphi}^\alpha, \boldsymbol{\varphi}^\beta \rangle = \delta_{\alpha\beta}$.
- (v) $[I_N] - [G_m] \in \mathbb{M}_N^{+0}$.

PROOF. The proof is left to the reader.

6. Probabilistic learning on manifolds (PLoM): construction of the probability measure and its generator

The three main steps of the PLoM introduced in [1] are the following. 1) Construction of a MCMC generator for random matrix $[\mathbf{H}^N]$ defined in Definition 2, based on a nonlinear Itô stochastic differential equation (ISDE) that will be introduced in Section 6.1, for which the probability measure $p_{[\mathbf{H}^N]}([\boldsymbol{\eta}]) d[\boldsymbol{\eta}]$ is a marginal probability distribution of the unique invariant measure of this ISDE. 2) Definition of a reduced representation $[\mathbf{H}_m^N] = [\mathbf{Z}_m] [g_m]^T$ of order $m < N$ for random matrix $[\mathbf{H}^N]$ using the reduced-order diffusion-maps basis $[g_m]$ and where $[\mathbf{Z}_m]$ is a random matrix with values in $\mathbb{M}_{\nu,m}$ for which its probability measure is $p_{[\mathbf{Z}_m]}([z]) d[z]$. 3) Construction of a reduced-order ISDE for which $p_{[\mathbf{Z}_m]}([z]) d[z]$ is a marginal probability distribution of its unique invariant measure. We will then

obtain a MCMC generator of random matrix $[\mathbf{Z}_m]$ and then of random matrix $[\mathbf{H}_m^N]$, which allows a learned dataset $\{[\eta_{ar}^\ell], \ell = 1, \dots, n_{MC}\}$ to be generated with an arbitrary number n_{MC} of realizations of $[\mathbf{H}_{m_{opt}}^N]$.

As already explained, the PLoM methodology has been developed for small values of N (small data) for which the probability measure $p_{\mathbf{H}}^{(N)}(\boldsymbol{\eta}) d\boldsymbol{\eta}$ is not necessarily converged. Therefore additional realizations that would be generated with this measure would not provide good realizations preserving the concentration. This is the reason why, the measure $p_{[\mathbf{H}^N]}([\boldsymbol{\eta}]) d[\boldsymbol{\eta}]$ is improved by introducing the transported probability measure $p_{[\mathbf{Z}_{m_{opt}}]}([z]) d[z]$ of random matrix $[\mathbf{Z}_{m_{opt}}]$. It should be noted that the additional realizations of $[\mathbf{H}_{m_{opt}}^N]$ are not constructed using the projection of realizations of $[\mathbf{H}^N]$ on the subspace spanned by the reduced-order diffusion-maps basis $[g_{m_{opt}}]$ (that would not be correct for a small value of N), but are constructed using the reduced-order ISDE associated with the transported probability measure $p_{[\mathbf{Z}_{m_{opt}}]}([z]) d[z]$ allowing additional realizations of $[\mathbf{Z}_{m_{opt}}]$ to be generated and then deducing the additional realizations of $[\mathbf{H}_{m_{opt}}^N] = [\mathbf{Z}_{m_{opt}}] [g_{m_{opt}}]^T$.

6.1. MCMC generator for random matrix $[\mathbf{H}^N]$

The PLoM method begins with the construction of a MCMC generator for random matrix $[\mathbf{H}^N]$ whose pdf $p_{[\mathbf{H}^N]}$ is given by Eq. (13). It is based on a nonlinear ISDE, formulated for a dissipative Hamiltonian dynamical system [23] for a diffusion stochastic process $\{([\mathbf{U}(r)], [\mathbf{V}(r)]), r \geq 0\}$ with values in $\mathbb{M}_{v,N} \times \mathbb{M}_{v,N}$, which admits a unique invariant measure for which the marginal probability distribution with respect to $[\mathbf{U}]$ is the probability measure $p_{[\mathbf{H}^N]}([\boldsymbol{\eta}]) d[\boldsymbol{\eta}]$. This MCMC generator is adapted to perform its projection on the subspace spanned by the reduced-order diffusion-maps basis and in addition, a dissipative term allows the transient part of the response to be rapidly killed. This MCMC generator belongs to the class of Hamiltonian Monte Carlo methods [24, 25], which is an MCMC algorithm [26, 27, 28].

Notation 1 (Matrix-valued Wiener process $[\mathbf{W}]$ and parameter f_0). Let us introduce the stochastic process $\{[\mathbf{W}(r)] = [\mathbf{W}^1(r) \dots \mathbf{W}^N(r)], r \geq 0\}$ defined on $(\Theta, \mathcal{T}, \mathcal{P})$, with values in $\mathbb{M}_{v,N}$, independent of random matrix $[\mathbf{H}^N]$, in which the columns $\mathbf{W}^1, \dots, \mathbf{W}^N$ are N independent copies of the normalized Wiener stochastic process $\mathbf{W} = (W_1, \dots, W_v)$, defined on $(\Theta, \mathcal{T}, \mathcal{P})$, indexed by \mathbb{R}^+ , with values in \mathbb{R}^v , such that $\mathbf{W}(0) = 0_v$ a.s., $E\{\mathbf{W}(r)\} = 0_v$, and $E\{\mathbf{W}(r) \otimes \mathbf{W}(r')\} = \min(r, r') [I_v]$. Let $f_0 > 0$ be a free parameter that will allow the dissipation term of the nonlinear ISDE (dissipative Hamiltonian system) to be controlled.

Theorem 2 (ISDE as the MCMC generator of matrix $[\mathbf{H}^N]$). Using Notation 1, we consider the stochastic process $\{([\mathbf{U}(r)], [\mathbf{V}(r)]), r \geq 0\}$ with values in $\mathbb{M}_{v,N} \times \mathbb{M}_{v,N}$, which verifies the following ISDE for $r > 0$, with the initial conditions for $r = 0$,

$$d[\mathbf{U}(r)] = [\mathbf{V}(r)] dr, \quad (16a)$$

$$d[\mathbf{V}(r)] = [L([\mathbf{U}(r)])] dr - \frac{1}{2} f_0 [\mathbf{V}(r)] dr + \sqrt{f_0} d[\mathbf{W}(r)], \quad (16b)$$

$$[\mathbf{U}(0)] = [\eta_d] \text{ a.s.}, \quad [\mathbf{V}(0)] = [v_0] \text{ a.s.}, \quad (16c)$$

in which $[\eta_d]$ is defined by Eq. (11) and where $[v_0]$ a given matrix in $\mathbb{M}_{v,N}$. For $k = 1, \dots, v$ and $\ell = 1, \dots, N$, and for $\mathbf{u}^\ell = (u_1^\ell, \dots, u_v^\ell)$ with $u_k^\ell = [u]_{k\ell}$, the matrix $[L([u])] \in \mathbb{M}_{v,N}$ is defined, as a function of a potential \mathcal{V} , by $[L([u])]_{k\ell} = -\partial \mathcal{V}(\mathbf{u}^\ell) / \partial u_k^\ell$ in which $\mathcal{V}(\mathbf{u}^\ell) = -\log\{\frac{1}{N} \sum_{j=1}^N \exp\{-\frac{1}{2s^2} \|\frac{\hat{s}}{s} \boldsymbol{\eta}_d^j - \mathbf{u}^\ell\|^2\}\}$. The ISDE defined by Eqs. (16a) and (16b) admits the unique invariant measure, $p_{[\mathbf{H}^N], [\mathbf{V}^N]}([\boldsymbol{\eta}], [v]) d[\boldsymbol{\eta}] \otimes d[v] = (p_{[\mathbf{H}^N]}([\boldsymbol{\eta}]) d[\boldsymbol{\eta}]) \otimes (p_{[\mathbf{V}^N]}([v]) d[v])$ on $\mathbb{M}_{v,N} \times \mathbb{M}_{v,N}$, in which $p_{[\mathbf{V}^N]}$ is the Gaussian density $[v] \mapsto (2\pi)^{-vN/2} \exp\{-\|v\|^2/2\}$ on $\mathbb{M}_{v,N}$ and where the pdf $p_{[\mathbf{H}^N]}([\boldsymbol{\eta}])$ is defined by Eq. (13). Matrix $[v_0]$ is any realization of the Gaussian pdf $p_{[\mathbf{V}^N]}$, independent of $\{[\mathbf{W}(r)], r \geq 0\}$.

PROOF. Since the columns $\mathbf{H}^1, \dots, \mathbf{H}^N$ of random matrix $[\mathbf{H}^N]$ are independent copies of random vector $\mathbf{H}^{(N)}$ (see Definition 2), and since the pdf of random matrix $[\mathbf{H}^N]$ is $p_{[\mathbf{H}^N]}$ defined by Eq. (13), Theorems 4, 6, and 7 in Pages 211 to 214 of [29] and the expression of the invariant measure given in Page 211 of the same reference, for which the Hamiltonian is $\mathcal{H}(\mathbf{u}, \mathbf{v}) = \|\mathbf{v}\|^2/2 + \mathcal{V}(\mathbf{u})$, prove that the invariant measure is the one given in Theorem 2 and is unique.

6.2. Reduced representation $[\mathbf{H}_m^N]$ of random matrix $[\mathbf{H}^N]$

Definition 5 (Random matrix $[\mathbf{H}_m^N]$). For given ε_{DM} , m , and κ , the random matrix $[\mathbf{H}_m^N]$ on $(\Theta, \mathcal{T}, \mathcal{P})$, with values in $\mathbb{M}_{v,N}$, is defined by $[\mathbf{H}_m^N] = [\mathbf{Z}_m][g_m]^T$ with $[g_m] \in \mathbb{M}_{N,m}$ defined in Definition 3 and where $[\mathbf{Z}_m]$ is a random matrix with values in $\mathbb{M}_{v,m}$ for which its probability measure admits a pdf $p_{[\mathbf{Z}_m]}([z])$ with respect to $d[z]$.

Notation 2 (Random vectors $\hat{\mathbf{H}}^k$ and $\hat{\mathbf{Z}}^k$). For $k \in \{1, \dots, v\}$, let $\hat{\mathbf{H}}^k = (\hat{H}_1^k, \dots, \hat{H}_N^k)$ be the random vector in \mathbb{R}^N such that $\hat{H}_j^k = [\mathbf{H}_m^N]_{kj}$ for $j \in \{1, \dots, N\}$ and let $\hat{\mathbf{Z}}^k = (\hat{Z}_1^k, \dots, \hat{Z}_m^k)$ be the random vector in \mathbb{R}^m such that $\hat{Z}_\alpha^k = [\mathbf{H}_m^N]_{k\alpha}$ for $\alpha \in \{1, \dots, m\}$. Consequently, $\hat{\mathbf{H}}^k = \sum_{\alpha=1}^m \hat{Z}_\alpha^k \mathbf{g}^\alpha$ in which \mathbf{g}^α is defined in Definition 3.

Let $\mathcal{L}^0(\Theta, \mathbb{R}^N)$ be the vector space of all the random variables, defined on $(\Theta, \mathcal{T}, \mathcal{P})$, with values in \mathbb{R}^N . It can be seen that each \mathbb{R}^N -valued random variable $\hat{\mathbf{H}}^k$ belongs to the subspace $\mathcal{L}^0(\Theta, \mathcal{E}_m) \subset \mathcal{L}^0(\Theta, \mathbb{R}^N)$ in which $\mathcal{E}_m \subset \mathbb{R}^N$ is the subspace of \mathbb{R}^N spanned by $\{\mathbf{g}^1, \dots, \mathbf{g}^m\}$. Note that, contrarily to the PCA that is a reduction following the physical coordinates axis, the representation constructed with the reduced-order diffusion-maps basis is a reduction following the data axis.

Remark 2 (Relationship between $[\mathbf{H}_N^N]$ and $[\mathbf{H}^N]$). Since $[g_N]$ is a vector basis of \mathbb{R}^N (see Definition 3), for $m = N$, the random matrix $[\mathbf{H}_N^N]$ is an independent copy of random matrix $[\mathbf{H}^N]$ introduced in Definition 2, in which $[\mathbf{H}_N^N] = [\mathbf{Z}_N][g_N]^T$ is a representation of $[\mathbf{H}^N]$ with $[\mathbf{Z}_N] = [\mathbf{H}^N][a_N]$, where $[a_N]$ is given by Definition 4 for $m = N$.

6.3. Explicit expression of pdf $p_{[\mathbf{Z}_m]}$ and reduced-order ISDE

Theorem 3 (Reduced-order ISDE and pdf $p_{[\mathbf{Z}_m]}$). The notations introduced in Definition 4 and in Theorem 2 are used. For given ε_{DM} , m , and κ , let $\{([\mathbf{Z}(r)], [\mathbf{Y}(r)]), r \geq 0\}$ be the stochastic process defined on $(\Theta, \mathcal{T}, \mathcal{P})$, with values in $\mathbb{M}_{v,m} \times \mathbb{M}_{v,m}$, which verifies the following reduced-order ISDE for all $r > 0$, with the initial conditions for $r = 0$,

$$d[\mathbf{Z}(r)] = [\mathbf{Y}(r)] dr, \quad (17a)$$

$$d[\mathbf{Y}(r)] = [\mathcal{L}([\mathbf{Z}(r)])] dr - \frac{1}{2} f_0 [\mathbf{Y}(r)] dr + \sqrt{f_0} d[\mathbf{W}(r)] [a_m], \quad (17b)$$

$$[\mathbf{Z}(0)] = [\eta_d] [a_m] \text{ a.s.}, \quad [\mathbf{Y}(0)] = [v_0] [a_m] \text{ a.s.}, \quad (17c)$$

in which, $\forall [z] \in \mathbb{M}_{v,m}$, $[\mathcal{L}([z])] = [L([z][g_m]^T)] [a_m] \in \mathbb{M}_{v,m}$. Eqs. (17a) and (17b) admit the unique invariant measure on $\mathbb{M}_{v,m} \times \mathbb{M}_{v,m}$,

$$p_{[\mathbf{Z}_m], [\mathbf{Y}_m]}([z], [y]) d[z] \otimes d[y] = (p_{[\mathbf{Z}_m]}([z]) d[z]) \otimes (p_{[\mathbf{Y}_m]}([y]) d[y]), \quad (18)$$

in which $p_{[\mathbf{Y}_m]}$ is the Gaussian density $[y] \mapsto (2\pi)^{-vm/2} \exp\{-\|y\|^2/2\}$ on $\mathbb{M}_{v,m}$ and where the pdf $[z] \mapsto p_{[\mathbf{Z}_m]}([z])$ on $\mathbb{M}_{v,m}$ is written as

$$p_{[\mathbf{Z}_m]}([z]) = c_{vm} \prod_{\ell=1}^N \left\{ \sum_{j=1}^N \exp\left\{-\frac{1}{2\hat{s}^2} \left\| \frac{\hat{s}}{s} \boldsymbol{\eta}_d^j - \sum_{\alpha=1}^m \mathbf{z}^\alpha g_\ell^\alpha \right\|^2\right\} \right\}. \quad (19)$$

The positive parameter c_{vm} is the constant of normalization, $[z] = [\mathbf{z}^1 \dots \mathbf{z}^m] \in \mathbb{M}_{v,m}$ with $\mathbf{z}^\alpha = (z_1^\alpha, \dots, z_v^\alpha) \in \mathbb{R}^v$ and with $z_k^\alpha = [z]_{k\alpha}$, and g_ℓ^α is given by Definition 3. The reduced-order ISDE with initial conditions, defined by Eqs. (17a) to (17c), has a unique stochastic solution $\{([\mathbf{Z}(r)], [\mathbf{Y}(r)]), r \geq 0\}$ that is a second-order diffusion stochastic process, which is asymptotic, for $r \rightarrow +\infty$, to a stationary and ergodic stochastic process $\{([\mathbf{Z}_{\text{st}}(r_{\text{st}})], [\mathbf{Y}_{\text{st}}(r_{\text{st}})]), r_{\text{st}} \geq 0\}$ for the right-shift semi-group on $\mathbb{R}^+ = [0, +\infty[$. For all r_{st} fixed in \mathbb{R}^+ , the joint probability measure of the random matrices $[\mathbf{Z}_{\text{st}}(r_{\text{st}})]$ and $[\mathbf{Y}_{\text{st}}(r_{\text{st}})]$ is the invariant measure defined by Eq. (18) and the pdf of random matrix $[\mathbf{Z}_{\text{st}}(r_{\text{st}})]$ is defined by Eq. (19). Consequently, Eqs. (17a) to (17c) yield a MCMC generator of random matrix $[\mathbf{Z}_m]$ and parameter f_0 allows for killing the transient regime induced by initial conditions, in order to reach the stationary solution more quickly.

PROOF. We introduce the stochastic process $\{([\mathbf{Z}(r)], [\mathbf{Y}(r)]), r \geq 0\}$ with values in $\mathbb{M}_{v,m} \times \mathbb{M}_{v,m}$, such that, for all $r \geq 0$, $[\mathbf{U}(r)] = [\mathbf{Z}(r)][g_m]^T$ and $[\mathbf{V}(r)] = [\mathbf{Y}(r)][g_m]^T$ in which $[g_m] \in \mathbb{M}_{N,m}$ is given by Definition 3 and where $\{([\mathbf{U}(r)], [\mathbf{V}(r)]), r \geq 0\}$ is the stochastic process with values in $\mathbb{M}_{v,N} \times \mathbb{M}_{v,N}$, introduced in Theorem 2. Considering this

change of stochastic processes, substituting them in Eqs. (16a) and (16b), and right multiplying these two equations by matrix $[a_m]$, yield Eqs. (17a) and (17b). The initial conditions defined by Eq. (17c) are similarly obtained.

(i) Proof of Eq. (19). For m fixed, since the reduced representation of random matrix $[\mathbf{H}^N]$ (for which its pdf $p_{[\mathbf{H}^N]}$ is given by Eq. (13)) is defined as the random matrix $[\mathbf{H}_m^N] = [\mathbf{Z}_m] [g_m]^T$ (see Definition 5), the theorem of the image of a measure by a measurable mapping allows for deducing Eq. (19) of the pdf $p_{[\mathbf{Z}_m]}$ of random matrix $[\mathbf{Z}_m]$ with values in $\mathbb{M}_{\nu, m}$.

(ii) Proof that $p_{[\mathbf{Z}_m], [\mathbf{Y}_m]}([z], [y]) d[z] \otimes d[y]$ defined by Eq. (18), with $p_{[\mathbf{Z}_m]}([z])$ given by Eq. (19), is the invariant measure of Eqs. (17a) and (17b). For proving that, there are several possibilities. We chose to use an algebraic-based demonstration, which allows for introducing notations that will be reused in Proposition 2. For simplifying the writing, the Itô equation (17a)-(17b) is rewritten as the following second-order stochastic differential equation that has to be read as an equality of generalized stochastic processes (see for instance, Chapter XI of [30]),

$$D_r^2[\mathbf{Z}] + \frac{1}{2} f_0 D_r[\mathbf{Z}] - [\mathcal{L}([\mathbf{Z})]] = \sqrt{f_0} D_r[\mathbf{W}] [a_m], \quad (20)$$

in which $D_r[\mathbf{W}]$ is the generalized normalized Gaussian white process resulting from the generalized derivative with respect to r of the $\mathbb{M}_{\nu, N}$ -valued Wiener stochastic process defined in Notation 1. For $k = 1, \dots, \nu$, for $\alpha = 1, \dots, m$, and for $\ell = 1, \dots, N$, we define $\mathbf{z}^\alpha = (z_1^\alpha, \dots, z_\nu^\alpha) \in \mathbb{R}^\nu$ and $\hat{\mathbf{z}}^k = (\hat{z}_1^k, \dots, \hat{z}_m^k) \in \mathbb{R}^m$ with $z_k^\alpha = \hat{z}_\alpha^k = [z]_{k\alpha}$. Similarly, we define the real functions $(\mathbf{z}^1, \dots, \mathbf{z}^m) \mapsto \Phi(\mathbf{z}^1, \dots, \mathbf{z}^m)$ on $\mathbb{R}^\nu \times \dots \times \mathbb{R}^\nu$ and $(\hat{\mathbf{z}}^1, \dots, \hat{\mathbf{z}}^\nu) \mapsto \hat{\Phi}(\hat{\mathbf{z}}^1, \dots, \hat{\mathbf{z}}^\nu)$ on $\mathbb{R}^m \times \dots \times \mathbb{R}^m$, such that

$$\Phi(\mathbf{z}^1, \dots, \mathbf{z}^m) = \sum_{\ell=1}^N \mathcal{V}(\mathbf{u}^\ell), \quad \mathbf{u}^\ell = \sum_{\alpha=1}^m \mathbf{z}^\alpha g_\ell^\alpha, \quad \hat{\Phi}(\hat{\mathbf{z}}^1, \dots, \hat{\mathbf{z}}^\nu) = \Phi(\mathbf{z}^1, \dots, \mathbf{z}^m). \quad (21)$$

Eq. (20) can be rewritten as ν coupled generalized stochastic equations on \mathbb{R}^m ,

$$D_r^2 \hat{\mathbf{Z}}^k + \frac{1}{2} f_0 D_r \hat{\mathbf{Z}}^k + ([g_m]^T [g_m])^{-1} \nabla_{\hat{\mathbf{z}}^k} \hat{\Phi}(\hat{\mathbf{Z}}^1, \dots, \hat{\mathbf{Z}}^\nu) = \sqrt{f_0} [a_m]^T D_r \hat{\mathbf{W}}^k,$$

with $k \in \{1, \dots, \nu\}$ and where $\{\hat{\mathbf{W}}^k\}_\alpha = [\mathbf{W}]_{k\alpha}$. Left multiplying this last equation by the invertible matrix $[g_m]^T [g_m] \in \mathbb{M}_m$ yields, for $k \in \{1, \dots, \nu\}$, the following coupled equations,

$$[g_m]^T [g_m] D_r^2 \hat{\mathbf{Z}}^k + \frac{1}{2} f_0 [g_m]^T [g_m] D_r \hat{\mathbf{Z}}^k + \nabla_{\hat{\mathbf{z}}^k} \hat{\Phi}(\hat{\mathbf{Z}}^1, \dots, \hat{\mathbf{Z}}^\nu) = \sqrt{f_0} [g_m]^T D_r \hat{\mathbf{W}}^k.$$

Using the mathematical results given in Chapter X of [29], it can be deduced that the ISDE corresponding to the previous ν coupled generalized stochastic equations admits a unique invariant measure on $(\Pi_{k=1}^\nu \mathbb{R}^m) \times (\Pi_{k=1}^\nu \mathbb{R}^m)$, defined by the following density with respect to $(\otimes_{k=1}^\nu d\hat{\mathbf{z}}^k) \otimes (\otimes_{k=1}^\nu d\hat{\mathbf{y}}^k)$, which is

$$p(\hat{\mathbf{z}}^1, \dots, \hat{\mathbf{z}}^\nu; \hat{\mathbf{y}}^1, \dots, \hat{\mathbf{y}}^\nu) = \hat{c}_{2\nu m} \exp\left\{-\frac{1}{2} \sum_{k=1}^\nu \langle [g_m]^T [g_m] \hat{\mathbf{y}}^k, \hat{\mathbf{y}}^k \rangle - \hat{\Phi}(\hat{\mathbf{z}}^1, \dots, \hat{\mathbf{z}}^\nu)\right\}.$$

Consequently, the joint probability density function of the \mathbb{R}^ν -valued random variables $\mathbf{Z}^1, \dots, \mathbf{Z}^m$ with respect to $\otimes_{\alpha=1}^m d\mathbf{z}^\alpha$ is given, using the third Eq. (21), by $p_{\mathbf{Z}^1, \dots, \mathbf{Z}^m}(\mathbf{z}^1, \dots, \mathbf{z}^m) = \hat{c}_{\nu m} \exp\{-\Phi(\mathbf{z}^1, \dots, \mathbf{z}^m)\}$ and thus, using Eq. (21), the pdf of random matrix $[\mathbf{Z}_m]$ with respect to $d[z]$ is

$$p_{[\mathbf{Z}_m]}([z]) = \hat{c}_{\nu m} \exp\left\{-\sum_{\ell=1}^N \mathcal{V}\left(\sum_{\alpha=1}^m \mathbf{z}^\alpha g_\ell^\alpha\right)\right\}.$$

Using the expression of $\mathcal{V}(\mathbf{u}^\ell)$ defined in Theorem 2 and introducing $c_{\nu m} = \hat{c}_{\nu m}/N^N$, this pdf can be rewritten as Eq. (19).

(iii) Proof of uniqueness of an asymptotic stationary and ergodic solution of Eqs. (17a) to (17c). The use of Theorem 9 in Page 216 of [29] yields the proof that Eqs. (17a) to (17c) has a unique solution $\{([\mathbf{Z}(r)], [\mathbf{Y}(r)]), r \geq 0\}$ that is a second-order diffusion stochastic process, which is asymptotic, for $r \rightarrow +\infty$, to a unique stationary stochastic process

($[\mathbf{Z}_{st}]$, $[\mathbf{Y}_{st}]$) having the properties given in Theorem 3. The ergodicity of the stationary solution is directly deduced from [31] or from [32].

Proposition 2 (Explicit expression of the pdf $p_{[\mathbf{Z}_m]}$ of $[\mathbf{Z}_m]$). (i) The pdf $p_{[\mathbf{Z}_m]}$ of random matrix $[\mathbf{Z}_m]$ defined by Eq. (19) can be rewritten, for all $[z]$ in $\mathbb{M}_{v,m}$, as

$$p_{[\mathbf{Z}_m]}([z]) = \sum_{\mathbf{j} \in \mathcal{J}} p_{\mathbf{j}}(m) \prod_{k=1}^v p_{\widehat{\mathbf{z}}^k}(\widehat{\mathbf{z}}^k; \mathbf{j}), \quad \widehat{\mathbf{z}}^k = (\widehat{z}_1^k, \dots, \widehat{z}_m^k) \in \mathbb{R}^m, \quad \widehat{z}_\alpha^k = [z]_{k\alpha}, \quad (22)$$

in which for all \mathbf{j} in \mathcal{J} (see Definition 1),

$$p_{\mathbf{j}}(m) = \gamma_{\mathbf{j}}(m) \left(\sum_{\mathbf{j}' \in \mathcal{J}} \gamma_{\mathbf{j}'}(m) \right)^{-1}, \quad \sum_{\mathbf{j} \in \mathcal{J}} p_{\mathbf{j}}(m) = 1, \quad (23a)$$

$$\gamma_{\mathbf{j}}(m) = \exp\left\{-\frac{1}{2s^2} \langle [I_N] - [G_m], [M_d(\mathbf{j})] \rangle_F\right\}. \quad (23b)$$

Matrix $[G_m] \in \mathbb{M}_N$ (see Definition 4) and the matrix $[M_d(\mathbf{j})] \in \mathbb{M}_N^{+0}$ is defined by

$$[M_d(\mathbf{j})] = [\eta_d(\mathbf{j})]^T [\eta_d(\mathbf{j})] \quad , \quad [M_d(\mathbf{j})]_{\ell\ell'} = \langle \eta_d^{\ell}(\mathbf{j}), \eta_d^{\ell'}(\mathbf{j}) \rangle, \quad (24)$$

in which $[\eta_d(\mathbf{j})] \in \mathbb{M}_{v,N}$ is defined by Eq. (12). For all k in $\{1, \dots, v\}$, $p_{\widehat{\mathbf{z}}^k}(\cdot; \mathbf{j})$ is the Gaussian pdf, such that, for all $\widehat{\mathbf{z}}^k$ in \mathbb{R}^m ,

$$p_{\widehat{\mathbf{z}}^k}(\widehat{\mathbf{z}}^k; \mathbf{j}) = (2\pi)^m \det[C_m]^{-1/2} \exp\left\{-\frac{1}{2} \langle [C_m]^{-1} (\widehat{\mathbf{z}}^k - \widehat{\mathbf{z}}^k(\mathbf{j})), \widehat{\mathbf{z}}^k - \widehat{\mathbf{z}}^k(\mathbf{j}) \rangle\right\}, \quad (25)$$

in which $[C_m] = \widehat{s}^2 ([g_m]^T [g_m])^{-1} \in \mathbb{M}_m^+$, where $\widehat{\mathbf{z}}^k(\mathbf{j}) = (\widehat{s}/s) [a_m]^T \widehat{\eta}_d^k(\mathbf{j}) \in \mathbb{R}^m$ with $[a_m] \in \mathbb{M}_{N,m}$ given by Definition 4, and where $\widehat{\eta}_d^k(\mathbf{j}) = (\widehat{\eta}_{d,1}^k(\mathbf{j}), \dots, \widehat{\eta}_{d,N}^k(\mathbf{j})) \in \mathbb{R}^N$ with $\widehat{\eta}_{d,\ell}^k(\mathbf{j}) = \eta_{d,k}^{\ell}(\mathbf{j}) = [\eta_d(\mathbf{j})]_{k\ell}$ (see Eq. (12)).

(ii) For all \mathbf{j} in \mathcal{J} and for all $1 \leq m \leq N-1$, we have $a_{\mathbf{j}}(m) \stackrel{\text{def}}{=} \langle [I_N] - [G_m], [M_d(\mathbf{j})] \rangle_F \geq 0$, $0 < \gamma_{\mathbf{j}}(m) < 1$, $0 < p_{\mathbf{j}}(m) < 1$, and for $m = N$, $a_{\mathbf{j}}(N) = 0$, $\gamma_{\mathbf{j}}(N) = 1$, and $p_{\mathbf{j}}(N) = 1/N^N$.

PROOF. (i) Eq. (19) can be written as

$$p_{[\mathbf{Z}_m]}([z]) = c_{vm} \sum_{\mathbf{j} \in \mathcal{J}} \prod_{k=1}^v \exp\left\{-\frac{1}{2\widehat{s}^2} \left\| \frac{\widehat{s}}{s} \widehat{\eta}_d^k(\mathbf{j}) - [g_m] \widehat{\mathbf{z}}^k \right\|^2\right\}.$$

On the other hand, $\forall k \in \{1, \dots, v\}$, we have $-(2\widehat{s}^2)^{-1} \left\| \frac{\widehat{s}}{s} \widehat{\eta}_d^k(\mathbf{j}) - [g_m] \widehat{\mathbf{z}}^k \right\|^2 = -(1/2) \langle [C_m]^{-1} (\widehat{\mathbf{z}}^k - \widehat{\mathbf{z}}^k(\mathbf{j})), \widehat{\mathbf{z}}^k - \widehat{\mathbf{z}}^k(\mathbf{j}) \rangle + (2s^2)^{-1} \langle [G_m] \widehat{\eta}_d^k(\mathbf{j}), \widehat{\eta}_d^k(\mathbf{j}) \rangle - \|\widehat{\eta}_d^k(\mathbf{j})\|^2$. Combining the previous equations allows $p_{[\mathbf{Z}_m]}([z])$ to be rewritten as

$$p_{[\mathbf{Z}_m]}([z]) = c_{vm} \sum_{\mathbf{j} \in \mathcal{J}} \gamma_{\mathbf{j}}(m) \prod_{k=1}^v \exp\left\{-\frac{1}{2} \langle [C_m]^{-1} (\widehat{\mathbf{z}}^k - \widehat{\mathbf{z}}^k(\mathbf{j})), \widehat{\mathbf{z}}^k - \widehat{\mathbf{z}}^k(\mathbf{j}) \rangle\right\},$$

in which $\gamma_{\mathbf{j}}(m) = \prod_{k=1}^v \exp\left\{-\frac{1}{2s^2} (\|\widehat{\eta}_d^k(\mathbf{j})\|^2 - \langle [G_m] \widehat{\eta}_d^k(\mathbf{j}), \widehat{\eta}_d^k(\mathbf{j}) \rangle)\right\}$, which can, finally, be rewritten as Eq. (23b) with Eq. (24). The above expression of $p_{[\mathbf{Z}_m]}([z])$ is rewritten as

$$p_{[\mathbf{Z}_m]}([z]) = c_{vm} (2\pi)^{vm/2} (\det[C_m])^{v/2} \sum_{\mathbf{j} \in \mathcal{J}} \gamma_{\mathbf{j}}(m) \prod_{k=1}^v p_{\widehat{\mathbf{z}}^k}(\widehat{\mathbf{z}}^k; \mathbf{j}).$$

The constant c_{vm} of normalization is calculated by $\int_{\mathbb{M}_{v,m}} p_{[\mathbf{Z}_m]}([z]) d[z] = 1$. Since $\int_{\mathbb{R}^m} p_{\widehat{\mathbf{z}}^k}(\widehat{\mathbf{z}}^k; \mathbf{j}) d\widehat{\mathbf{z}}^k = 1$, it is deduced that $c_{vm} = \{(2\pi)^{vm/2} (\det[C_m])^{v/2} \sum_{\mathbf{j} \in \mathcal{J}} \gamma_{\mathbf{j}}(m)\}^{-1}$. Eq. (22) can then be deduced in which $p_{\mathbf{j}}(m)$ is given by Eq. (23a). Finally, $[M_d(\mathbf{j})] = \sum_{k=1}^v \widehat{\eta}_d^k(\mathbf{j}) \otimes \widehat{\eta}_d^k(\mathbf{j}) = [\eta_d(\mathbf{j})]^T [\eta_d(\mathbf{j})]$, which shows that $[M_d(\mathbf{j})] \in \mathbb{M}_N^{+0}$ because $v < N$ (see Section 2).

(ii) From Lemma 1-(v), and using Eq. (24), it can be seen that $a_j(m) \geq 0$. The end of the proof is easy to do.

Remark 3 (About the algebraic representation of pdf $p_{[\mathbf{Z}_m]}$ and its generator).

(i) Eq. (22) shows that pdf $p_{[\mathbf{Z}_m]}$ on $\mathbb{M}_{\nu,m}$ is a linear combination of N^N products of ν Gaussian pdf on \mathbb{R}^N . Consequently, the use of the reduced-order ISDE given by Theorem 3 effectively allows realizations of random matrix $[\mathbf{Z}_m]$ to be generated, while a Gaussian generator that would be based on the representation given by Eq. (22) is unthinkable.
(ii) The generation of $n_{\text{MC}} \gg 1$ independent realizations $\{[z^\ell], \ell = 1, \dots, n_{\text{MC}}\}$ of random matrix $[\mathbf{Z}_m]$ is performed by using the MCMC generator defined by Theorem 3 in which the reduced-order stochastic Eqs. (17a) to (17c) are solved using the Störmer-Verlet scheme [33, 34], which is well adapted to stochastic Hamiltonian dynamical systems and which is detailed in [1]. We can then deduce the learned dataset $\{[\eta_{\text{ar}}^\ell], \ell = 1, \dots, n_{\text{MC}}\}$ of random matrix $[\mathbf{H}_m^N]$ such that $[\eta_{\text{ar}}^\ell] = [z^\ell][g_m]^T$, with an arbitrary value of realizations.

7. L^2 -distance of random matrix $[\mathbf{H}_m^N]$ to matrix $[\eta_d]$ of the initial dataset and its analysis

In this section, N, ν, κ , and $\varepsilon_{\text{DM}} = \varepsilon_{\text{opt}}$ are fixed. The optimal value of m associated with ε_{opt} is m_{opt} as defined in Section 5.2. Integer m varies in $\{1, \dots, N\}$. The measure of the concentration of the probability measure $p_{[\mathbf{H}_m^N]}([\eta]) d[\eta]$, which is informed by the initial dataset represented by matrix $[\eta_d]$, will be analyzed as a function of m by using the square $d_N^2(m)$ of the $L^2(\Theta, \mathbb{M}_{\nu,N})$ -distance between random matrix $[\mathbf{H}_m^N]$ and matrix $[\eta_d]$.

Definition 6 (Square of the relative distance $d_N^2(m)$ of $[\mathbf{H}_m^N]$ to $[\eta_d]$). For m fixed, the square of the relative distance of random matrix $[\mathbf{H}_m^N]$ with values in $\mathbb{M}_{\nu,N}$ to matrix $[\eta_d] \in \mathbb{M}_{\nu,N}$ is defined as $d_N^2(m) = E\{\|[\mathbf{H}_m^N] - [\eta_d]\|^2\} / E\{\|[\eta_d]\|^2\}$.

The following Lemma gives the value of $d_N^2(N)$, which corresponds to the value of the distance if the PLoM method is not used ($m = N$). In this case, the MCMC generator of random matrix $[\mathbf{H}_N^N]$ is given by Theorem 2.

Lemma 2 (Value of $d_N^2(m)$ for $m = N$). For $m = N$, the random matrix $[\mathbf{H}_N^N]$, which is an independent copy of random matrix $[\mathbf{H}^N]$ (see Definition 2 and Remark 2) is such that $E\{[\mathbf{H}_N^N]\} = [0_{\nu,N}]$, $E\{\|[\mathbf{H}_N^N]\|^2\} = \nu N$, and the value of $d_N^2(m)$ for $m = N$ is $d_N^2(N) = 1 + N/(N - 1)$.

PROOF. Note that $\mathbf{H}^1, \dots, \mathbf{H}^N$ are independent copies of $\mathbf{H}^{(N)}$ (see Definition 2). (i) $E\{[\mathbf{H}_N^N]\} = E\{[\mathbf{H}^N]\} = [E\{\mathbf{H}^{(N)}\} \dots E\{\mathbf{H}^{(N)}\}]$, and since $E\{\mathbf{H}^{(N)}\} = 0_\nu$, we have $E\{[\mathbf{H}_N^N]\} = [0_{\nu,N}]$. (ii) $E\{\|[\mathbf{H}_N^N]\|^2\} = E\{\|[\mathbf{H}^N]\|^2\} = \sum_{j=1}^N E\{\|\mathbf{H}^j\|^2\} = N E\{\|\mathbf{H}^{(N)}\|^2\}$, therefore, $E\{\|[\mathbf{H}^{(N)}]\|^2\} = \text{Tr}\{[I_\nu]\} = \nu$, and consequently, we have $E\{\|[\mathbf{H}_N^N]\|^2\} = \nu N$. (iii) Using Definition 6 and Eq. (6) yields $d_N^2(N) = (\nu(N - 1))^{-1} (E\{\|[\mathbf{H}_N^N]\|^2\} - 2 \langle E\{[\mathbf{H}_N^N]\}, [\eta_d] \rangle_F + \|\eta_d\|^2)$. The result is obtained using (i), (ii), and Eq. (6).

It should be noted that the concentration is measured by the mean-square distance between the random matrix $[\mathbf{H}_m^N]$ and the deterministic matrix $[\eta_d]$. For $m = N$, this distance would be zero only if the probability measure of $[\mathbf{H}_N^N]$ were the Dirac measure in the space $\mathbb{M}_{\nu,N}$ at point $[\eta_d] \in \mathbb{M}_{\nu,N}$, which is obviously not the case.

Proposition 3 (Expression of $d_N^2(m)$). Let m be fixed. We have

$$E\{[\mathbf{H}_m^N]\} = \sum_{\mathbf{j} \in \mathcal{J}} p_{\mathbf{j}}(m) \frac{\widehat{s}}{s} [\eta_d(\mathbf{j})] [G_m] \in \mathbb{M}_{\nu,N}, \quad (26a)$$

$$E\{\|[\mathbf{H}_m^N]\|^2\} = \sum_{\mathbf{j} \in \mathcal{J}} p_{\mathbf{j}}(m) \left(\nu \widehat{s}^2 m + \frac{\widehat{s}^2}{s^2} \langle [G_m], [M_d(\mathbf{j})] \rangle_F \right), \quad (26b)$$

in which $p_{\mathbf{j}}(m)$ is given by Eq. (23a), $[\eta_d(\mathbf{j})]$ is defined by Eq. (12), $[G_m]$ by Definition 4, and $[M_d(\mathbf{j})]$ by Eq. (24), and we have

$$d_N^2(m) = 1 + \frac{m \widehat{s}^2}{N - 1} + \frac{1}{\|\eta_d\|^2} \sum_{\mathbf{j} \in \mathcal{J}} p_{\mathbf{j}}(m) \langle [G_m], [B_d(\mathbf{j})] \rangle_F, \quad (27a)$$

$$[B_d(\mathbf{j})] = \frac{\widehat{s}^2}{s^2} [M_d(\mathbf{j})] - 2 \frac{\widehat{s}}{s} [\eta_d(\mathbf{j})]^T [\eta_d] \in \mathbb{M}_N. \quad (27b)$$

The entries of $[B_d(\mathbf{j})]$ are $[B_d(\mathbf{j})]_{\ell\ell'} = (\hat{s}^2/s^2) \langle \boldsymbol{\eta}_d^{j_\ell}, \boldsymbol{\eta}_d^{j_{\ell'}} \rangle - (2\hat{s}/s) \langle \boldsymbol{\eta}_d^{j_\ell}, \boldsymbol{\eta}_d^{j_{\ell'}} \rangle$.

PROOF. Since $[\mathbf{H}_m^N] = [\mathbf{Z}_m] [g_m]^T$ (see Definition 5), it can be seen that $E\{[\mathbf{H}_m^N]\} = [\mathcal{M}_1^N(m)] [g_m]^T$ and $E\{\|[\mathbf{H}_m^N]\|^2\} = \langle [\mathcal{M}_2^N(m)], [g_m]^T [g_m] \rangle_F$, in which, using Eq. (22) for $p_{[\mathbf{z}_m]}$,

$$\begin{aligned} [\mathcal{M}_1^N(m)] &= \sum_{\mathbf{j} \in \mathcal{J}} p_{\mathbf{j}}(m) \int_{\mathbb{R}^m} \dots \int_{\mathbb{R}^m} [\hat{\mathbf{z}}^1 \dots \hat{\mathbf{z}}^v]^T \otimes_{k=1}^v \{p_{\hat{\mathbf{z}}^k}(\hat{\mathbf{z}}^k; \mathbf{j}) d\hat{\mathbf{z}}^k\} \in \mathbb{M}_{v,m}, \\ [\mathcal{M}_2^N(m)] &= \sum_{k'=1}^v \sum_{\mathbf{j} \in \mathcal{J}} p_{\mathbf{j}}(m) \prod_{k=1}^v \left\{ \int_{\mathbb{R}^m} \hat{\mathbf{z}}^{k'} \otimes \hat{\mathbf{z}}^k p_{\hat{\mathbf{z}}^k}(\hat{\mathbf{z}}^k; \mathbf{j}) d\hat{\mathbf{z}}^k \right\} \in \mathbb{M}_{v,m}. \end{aligned}$$

(i) Calculation of $E\{[\mathbf{H}_m^N]\}$. Using Eq. (25) and the expression of $\hat{\mathbf{z}}^k(\mathbf{j})$ defined in Proposition 2 yield $[\mathcal{M}_1^N(m)] = \sum_{\mathbf{j} \in \mathcal{J}} p_{\mathbf{j}}(m) [\hat{\mathbf{z}}(\mathbf{j})]^T$ with $[\hat{\mathbf{z}}(\mathbf{j})] = [\hat{\mathbf{z}}^1(\mathbf{j}) \dots \hat{\mathbf{z}}^v(\mathbf{j})] = (\hat{s}/s) [a_m]^T [\eta_d(\mathbf{j})]^T$. Since $[a_m] [g_m]^T = [G_m]$ (see Definition 4), we obtain Eq. (26a).

(ii) Calculation of $E\{\|[\mathbf{H}_m^N]\|^2\}$. Eq. (25) shows that $\int_{\mathbb{R}^m} \hat{\mathbf{z}}^k \otimes \hat{\mathbf{z}}^k p_{\hat{\mathbf{z}}^k}(\hat{\mathbf{z}}^k; \mathbf{j}) d\hat{\mathbf{z}}^k = [C_m] + \hat{\mathbf{z}}^k(\mathbf{j}) \otimes \hat{\mathbf{z}}^k(\mathbf{j})$. Since $\int_{\mathbb{R}^m} p_{\hat{\mathbf{z}}^k}(\hat{\mathbf{z}}^k; \mathbf{j}) d\hat{\mathbf{z}}^k = 1$ and that $\sum_{k'=1}^v \hat{\mathbf{z}}^{k'}(\mathbf{j}) \otimes \hat{\mathbf{z}}^{k'}(\mathbf{j}) = [\hat{\mathbf{z}}(\mathbf{j})] [\hat{\mathbf{z}}(\mathbf{j})]^T$, we have $[\mathcal{M}_2^N(m)] = \sum_{\mathbf{j} \in \mathcal{J}} p_{\mathbf{j}}(m) (v[C_m] + [\hat{\mathbf{z}}(\mathbf{j})] [\hat{\mathbf{z}}(\mathbf{j})]^T)$. It can then be deduced that $E\{\|[\mathbf{H}_m^N]\|^2\} = \sum_{\mathbf{j} \in \mathcal{J}} p_{\mathbf{j}}(m) (v \mathcal{I}_1(m) + \mathcal{I}_2(m, \mathbf{j}))$. Using $[C_m] = \hat{s}^2 ([g_m]^T [g_m])^{-1}$ (see Proposition 2) and the expression of $[G_m]$ given in Definition 4 yield $\mathcal{I}_1(m) = \langle [C_m], [g_m]^T [g_m] \rangle_F = \hat{s}^2 \text{Tr}\{[G_m]\} = \hat{s}^2 m$. On the other hand, $\mathcal{I}_2(m, \mathbf{j}) = \|[\hat{\mathbf{z}}(\mathbf{j})]^T [g_m]^T\|^2$ and since $[\hat{\mathbf{z}}(\mathbf{j})]^T = (\hat{s}/s) [\eta_d(\mathbf{j})] [a_m]$, we have $[\hat{\mathbf{z}}(\mathbf{j})]^T [g_m]^T = (\hat{s}/s) [\eta_d(\mathbf{j})] [G_m]$. Therefore, $\mathcal{I}_2(m, \mathbf{j}) = (\hat{s}^2/s^2) \|[\eta_d(\mathbf{j})] [G_m]\|^2$. Since $[G_m]^2 = [G_m]$ (see Lemma 1-(iii)) and that $[M_d(\mathbf{j})] = [\eta_d(\mathbf{j})]^T [\eta_d(\mathbf{j})]$ (see Eq. (24)), we obtain $\mathcal{I}_2(m, \mathbf{j}) = (\hat{s}^2/s^2) \langle [G_m], [M_d(\mathbf{j})] \rangle_F$. By substitution, we obtain Eq. (26b).

(iii) Calculation of $d_N^2(m)$. From Definition 6 of $d_N^2(m)$, we have $d_N^2(m) = \|\eta_d\|^{-2} (E\{\|[\mathbf{H}_m^N]\|^2\} - 2 \langle E\{[\mathbf{H}_m^N]\}, [\eta_d] \rangle_F + \|\eta_d\|^2)$. Using Eqs. (26a) and (26b) allows for proving Eq. (27a) with Eq. (27b).

In order to apply Proposition 3 for the case $m = N$, we need the results given in the following lemma.

Lemma 3. Using Definition 1 and Eqs. (24) and (27b), we have

$$\frac{1}{N^N} \sum_{\mathbf{j} \in \mathcal{J}} [\eta_d(\mathbf{j})] = [0_{v,N}], \quad (28a)$$

$$\frac{1}{N^N} \sum_{\mathbf{j} \in \mathcal{J}} [M_d(\mathbf{j})] = \frac{1}{N} \|\eta_d\|^2 [I_N] \quad (28b)$$

$$\frac{1}{N^N} \sum_{\mathbf{j} \in \mathcal{J}} [B_d(\mathbf{j})] = \frac{\hat{s}^2}{s^2} \frac{1}{N} \|\eta_d\|^2 [I_N]. \quad (28c)$$

PROOF. (i) Proof of Eq. (28a): for all fixed ℓ , we have

$$N^{-N} \sum_{\mathbf{j} \in \mathcal{J}} \boldsymbol{\eta}_d^{j_\ell} = N^{-N} \sum_{j_1=1}^N \dots \sum_{j_N=1}^N \boldsymbol{\eta}_d^{j_\ell} = N^{-1} \sum_{j_\ell=1}^N \boldsymbol{\eta}_d^{j_\ell} = \mathbf{0}_v,$$

taking into account Eq. (5). (ii) Proof of Eq. (28b): for $\ell = \ell'$, it can be seen that

$$N^{-N} \sum_{\mathbf{j} \in \mathcal{J}} [M_d(\mathbf{j})] = N^{-N} \sum_{j_1=1}^N \dots \sum_{j_N=1}^N \|\boldsymbol{\eta}_d^{j_\ell}\|^2 = N^{-1} \sum_{j_\ell=1}^N \|\boldsymbol{\eta}_d^{j_\ell}\|^2.$$

For $\ell \neq \ell'$, due to Eq. (28a), it can be seen that

$$N^{-N} \sum_{\mathbf{j} \in \mathcal{J}} [M_d(\mathbf{j})] = N^{-N+2} \sum_{\mathbf{j} \neq \{j_\ell \cap j_{\ell'}\}} \langle N^{-1} \sum_{j_\ell=1}^N \boldsymbol{\eta}_d^{j_\ell}, N^{-1} \sum_{j_{\ell'}=1}^N \boldsymbol{\eta}_d^{j_{\ell'}} \rangle = 0.$$

Grouping the two cases yields Eq. (28b). (iii) Proof of Eq. (28c): this result can easily be deduced from Eqs. (27b), (28a), and (28b).

Corollary 1 (Value of $d_N^2(N)$ as a corollary of Proposition 3). Taking $m = N$ in Eq. (27a) yields the value $d_N^2(N) = 1 + N/(N - 1)$ given in Lemma 2.

PROOF. The proof is easy to obtain using Proposition 2-(ii), Lemma 1-(ii), Eqs. (27a) and (27b) with $m = N$, and Eq. (8).

Definition 7 (Matrix $[\eta_d^m]$ and function ε_d). Let $[\eta_d] \in \mathbb{M}_{v,N}$ be the matrix defined by Eq. (11) and let $[G_m] \in \mathbb{M}_N$ be the matrix defined in Definition 4. For all m such that $1 \leq m \leq N$, the matrix $[\eta_d^m] \in \mathbb{M}_N$ is defined by

$$[\eta_d^m] = [\eta_d] [G_m]. \quad (29)$$

The function $m \mapsto \varepsilon_d(m)$ with values in \mathbb{R}^+ is defined by

$$\|[\eta_d] - [\eta_d^m]\| = \varepsilon_d(m) \|\eta_d\|. \quad (30)$$

Lemma 4 (Properties of $m \mapsto \varepsilon_d(m)$ and expression of $\|\eta_d^m\|^2$). (i) $m \mapsto \varepsilon_d(m)$ is a decreasing function from $\{1, \dots, N\}$ into $[0, 1]$ and we have $\varepsilon_d(1) = 1$ and $\varepsilon_d(N) = 0$. (ii) For all m such that $1 \leq m \leq N$, the square of the Frobenius norm of matrix $[\eta_d^m]$ that is defined by Eq. (29), is written as

$$\|\eta_d^m\|^2 = (1 - \varepsilon_d(m)^2) \|\eta_d\|^2. \quad (31)$$

PROOF. (i) From Eq. (29) and Lemma 1-(ii), it can be seen that $[\eta_d^N] = [\eta_d]$. Therefore, Eq. (30) yields $\varepsilon_d(N) = 0$. From Section 5.1, it can be seen that $\mathbf{g}^1 = N^{-1/2} \|\boldsymbol{\psi}^1\| \mathbf{1}$ and using Definition 4 yield $[G_1] = N^{-1} \mathbf{1} \otimes \mathbf{1}$. We have $\|[\eta_d] - [\eta_d^1]\|^2 = \|[\eta_d]([I_N] - [G_1])\|^2 = \langle [\eta_d]([I_N] - [G_1]), [\eta_d]([I_N] - [G_1]) \rangle_F = \langle [\eta_d]^T [\eta_d], [I_N] - [G_1] \rangle_F$ because $([I_N] - [G_1])^2 = [I_N] - 2[G_1] + [G_1]^2 = [I_N] - [G_1]$. Hence, $\|[\eta_d] - [\eta_d^1]\|^2 = \|\eta_d\|^2 - N^{-1} \langle [\eta_d]^T [\eta_d], \mathbf{1} \otimes \mathbf{1} \rangle_F$ and $\langle [\eta_d]^T [\eta_d], \mathbf{1} \otimes \mathbf{1} \rangle_F = \|\sum_{j=1}^N \boldsymbol{\eta}_d^j\|^2 = 0$ (due to the first Eq. (5)), we deduce that $\|[\eta_d] - [\eta_d^1]\|^2 = \|\eta_d\|^2$, which proves that $\varepsilon_d(1) = 1$. Since $[G_m] = \sum_{\alpha=1}^m \boldsymbol{\varphi}^\alpha \otimes \boldsymbol{\varphi}^\alpha$ (see Lemma 1-(iv)), we have $\|[\eta_d] - [\eta_d^m]\|^2 = \langle [\eta_d]^T [\eta_d], [I_N] - [G_m] \rangle_F = \|\eta_d\|^2 - \sum_{\alpha=1}^m \|[\eta_d] \boldsymbol{\varphi}^\alpha\|^2$, which proves that $m \mapsto \varepsilon_d(m)$ is a decreasing function. (ii) Developing the left-hand side of Eq. (30) yields $\|\eta_d\|^2 - 2 \langle [\eta_d], [\eta_d^m] \rangle_F + \|\eta_d^m\|^2 = \varepsilon_d(m)^2 \|\eta_d\|^2$. On the other hand, $\langle [\eta_d], [\eta_d^m] \rangle_F = \langle [\eta_d], [\eta_d] [G_m] \rangle_F$ and since $[G_m] = [G_m]^2$ (see Lemma 1-(iii)), it can be deduced that, $\langle [\eta_d], [\eta_d^m] \rangle_F = \|\eta_d^m\|^2$, and then Eq. (31) holds.

In Hypothesis 1, based on the properties of $\{\lambda_\alpha(\varepsilon_{\text{DM}})\}_\alpha$, we have introduced the existence of optimal values ε_{opt} and m_{opt} of ε_{DM} and m , respectively. In the following hypothesis, we introduce the connection between $m \mapsto \varepsilon_d(m)$ and $\{\lambda_\alpha(\varepsilon_{\text{DM}})\}_\alpha$.

Hypothesis 2 (Relative to $m \mapsto \varepsilon_d(m)$). Under Hypothesis 1, it is assumed that m_{opt} is such that $1 = \varepsilon_d(1) > \dots > \varepsilon_d(m_{\text{opt}} - 1) \gg \varepsilon_d(m_{\text{opt}}) > \varepsilon_d(m_{\text{opt}} + 1) > \dots > \varepsilon_d(N) = 0$.

Remark 4 (Comments about hypotheses 1 and 2 related to m_{opt}). Regarding Hypothesis 1 devoted to the existence of the optimal values ε_{opt} of ε_{DM} and m_{opt} of m , Fig. 1 (left) displays the graph of function $\alpha \mapsto \log(\lambda_\alpha(\varepsilon_{\text{opt}}))$. The graph of function $m \mapsto \varepsilon_d(m)$ corresponding to Hypothesis 2 is shown in Fig. 1 (right). For $m > m_{\text{opt}}$, we have $\varepsilon_d(m) \ll 1$ while $\varepsilon_d(1) = 1$ and $\varepsilon_d(N) = 0$.

Proposition 4 (Adapted expression of $d_N^2(m)$). For all m such that $1 \leq m \leq N$, $d_N^2(m)$ given by Eq. (27a) can be written as

$$d_N^2(m) = f_d(m) + h_d(m), \quad (32a)$$

$$f_d(m) = \frac{m \widehat{s}^2}{N - 1} + \varepsilon_d(m)^2, \quad (32b)$$

$$h_d(m) = \sum_{\mathbf{j} \in \mathcal{J}} p_{\mathbf{j}}(m) \frac{1}{\|\eta_d\|^2} \|[\eta_d^m] - \frac{\widehat{s}}{s} [\eta_d^m(\mathbf{j})]\|^2, \quad (32c)$$

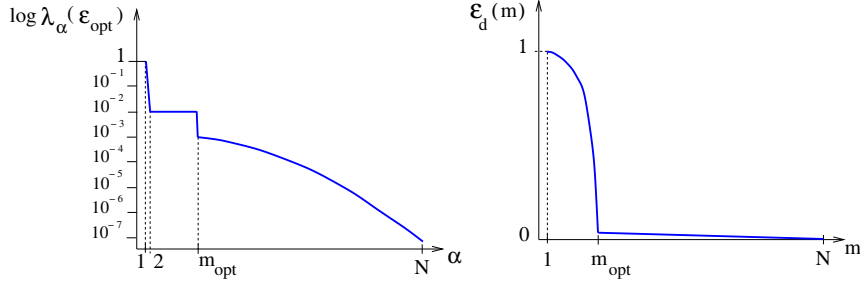


Figure 1: Left figure: for $\varepsilon_{\text{DM}} = \varepsilon_{\text{opt}}$, distribution of the eigenvalues $\lambda_\alpha(\varepsilon_{\text{opt}})$ in log scale as a function of rank α . Right figure: graph of function $m \mapsto \varepsilon_d(m)$.

in which, $\forall \mathbf{j} \in \mathcal{J}$, $p_{\mathbf{j}}(m)$ is defined by Eq. (23a) as a function of $\gamma_{\mathbf{j}}(m)$ (defined by Eq. (23b)) that can be rewritten as

$$\gamma_{\mathbf{j}}(m) = \exp \left\{ -\frac{1}{2s^2} \|\eta_d(\mathbf{j}) - [\eta_d^m(\mathbf{j})]\|^2 \right\}, \quad (33)$$

where $[\eta_d^m(\mathbf{j})] \in \mathbb{M}_{v,N}$ is defined by

$$[\eta_d^m(\mathbf{j})] = [\eta_d(\mathbf{j})] [G_m], \quad (34)$$

with $[\eta_d(\mathbf{j})] \in \mathbb{M}_{v,N}$ defined by Eq. (12).

PROOF. (i) From Eqs. (27b) and (24), it can be seen that $\langle [G_m], [B_d(\mathbf{j})] \rangle_F = (\widehat{s}/s)^2 \langle [G_m], [\eta_d(\mathbf{j})]^T [\eta_d(\mathbf{j})] \rangle_F - 2(\widehat{s}/s) \langle [G_m], [\eta_d(\mathbf{j})]^T [\eta_d] \rangle_F$. Since $[G_m] = [G_m]^2$ (see Lemma 1-(iii)), we obtain $\langle [G_m], [B_d(\mathbf{j})] \rangle_F = (\widehat{s}/s)^2 \|\eta_d(\mathbf{j})\|^2 - 2(\widehat{s}/s) \langle [\eta_d(\mathbf{j})] [G_m], [\eta_d] [G_m] \rangle_F$, which can be rewritten, using Eqs. (29) and (34), as $\langle [G_m], [B_d(\mathbf{j})] \rangle_F = \|\eta_d^m(\mathbf{j}) - (\widehat{s}/s) [\eta_d^m(\mathbf{j})]\|^2 - \|\eta_d^m(\mathbf{j})\|^2$. By substitution into Eq. (27a), since $\sum_{\mathbf{j} \in \mathcal{J}} p_{\mathbf{j}}(m) = 1$, and using Eq. (31) yield Eqs. (32a) to (32b). (ii) We have $([I_N] - [G_m])^2 = [I_N] - 2[G_m] + [G_m]^2 = [I_N] - [G_m]$. Consequently, $\langle [I_N] - [G_m], [M_d(\mathbf{j})] \rangle_F = \langle ([I_N] - [G_m])^2, [\eta_d(\mathbf{j})]^T [\eta_d(\mathbf{j})] \rangle_F = \|\eta_d(\mathbf{j}) ([I_N] - [G_m])\|^2$. Using Eq. (34) allows $\langle [I_N] - [G_m], [M_d(\mathbf{j})] \rangle_F = \|\eta_d(\mathbf{j}) - [\eta_d^m(\mathbf{j})]\|^2$ to be written. By substitution into Eq. (23b) yields Eq. (33).

Lemma 5 (Rewriting function h_d). For all m such that $1 \leq m \leq N$ and for all \mathbf{j} in \mathcal{J} , let $g_{\mathbf{j}}(m)$, $\underline{g}(m)$, $\underline{\gamma}(m)$, and $r(m)$ be defined by

$$g_{\mathbf{j}}(m) = \frac{\|\eta_d^m(\mathbf{j}) - \widehat{s} [\eta_d^m(\mathbf{j})]\|^2}{\|\eta_d\|^2}, \quad \underline{g}(m) = \frac{1}{N^N} \sum_{\mathbf{j} \in \mathcal{J}} g_{\mathbf{j}}(m), \quad \underline{\gamma}(m) = \frac{1}{N^N} \sum_{\mathbf{j} \in \mathcal{J}} \gamma_{\mathbf{j}}(m), \quad (35)$$

in which $\gamma_{\mathbf{j}}(m)$ is given by Eq. (33), and

$$r(m) \stackrel{\text{def}}{=} \sum_{\mathbf{j} \in \mathcal{J}} p_{\mathbf{j}}(m) \frac{g_{\mathbf{j}}(m)}{\underline{g}(m)} = \frac{\frac{1}{N^N} \sum_{\mathbf{j} \in \mathcal{J}} \gamma_{\mathbf{j}}(m) g_{\mathbf{j}}(m)}{\underline{\gamma}(m) \underline{g}(m)}, \quad (36)$$

in which $p_{\mathbf{j}}(m)$ is defined by Eq. (23a). Using these definitions, we have

$$\underline{g}(m) = 1 + \frac{\widehat{s}^2}{s^2} \frac{m}{N} - \varepsilon_d(m)^2, \quad \underline{g}(m) > 0, \quad \underline{g}(N) = 1 + \frac{\widehat{s}^2}{s^2}, \quad (37)$$

in which $\varepsilon_d(m)$ is defined by Eq. (30), and

$$\underline{\gamma}(m) > 0, \quad \underline{\gamma}(N) = 1, \quad r(m) > 0, \quad r(N) = 1. \quad (38)$$

Function $m \mapsto h_d(m)$ defined by Eq. (32c) can be rewritten as

$$h_d(m) = r(m) \underline{g}(m). \quad (39)$$

PROOF. Using Eq. (35) yields

$$\underline{g}(m) = \|\eta_d\|^{-2} \left\{ N^{-N} \sum_{\mathbf{j} \in \mathcal{J}} \|\eta_d^m(\mathbf{j})\|^2 + (\widehat{s}^2/s^2) N^{-N} \sum_{\mathbf{j} \in \mathcal{J}} \|\eta_d^m(\mathbf{j})\|^2 - 2(\widehat{s}/s) \langle [\eta_d^m], N^{-N} \sum_{\mathbf{j} \in \mathcal{J}} [\eta_d^m(\mathbf{j})] \rangle_F \right\}.$$

It can be seen that $\|\eta_d^m(\mathbf{j})\|^2 = \langle [\eta_d(\mathbf{j})] [G_m], [\eta_d(\mathbf{j})] [G_m] \rangle_F$. Since $[G_m]^2 = [G_m]$ (Lemma 1-(iii)) and $[\eta_d(\mathbf{j})]^T [\eta_d(\mathbf{j})] = [M_d(\mathbf{j})]$ (Eq. (24)), we have $N^{-N} \sum_{\mathbf{j} \in \mathcal{J}} \|\eta_d^m(\mathbf{j})\|^2 = \langle N^{-N} \sum_{\mathbf{j} \in \mathcal{J}} [M_d(\mathbf{j})], [G_m] \rangle_F$ that can be rewritten, using Eq. (28b) and Lemma 1-(i), as $N^{-N} \sum_{\mathbf{j} \in \mathcal{J}} \|\eta_d^m(\mathbf{j})\|^2 = N^{-1} \|\eta_d\|^2 \text{Tr}\{[G_m]\} = (m/N) \|\eta_d\|^2$. Substituting $\|\eta_d^m\|^2$ given by Eq. (31) in the above expression of $\underline{g}(m)$ and using Eq. (28a) yield Eq. (37). Due to Eq. (35), $\underline{g}(m) > 0$ and since $\varepsilon_d(N) = 0$ (see Lemma 4-(i)), the first Eq. (37) with $m = N$ yields the third Eq. (37). Since $[G_N] = [I_N]$ (see Lemma 1-(ii)) and using the definition of $\gamma_j(m)$ given by Eq. (23b) yield $\gamma_j(N) = 1$. The other results of the lemma are easy to prove.

Lemma 6 (Property of function f_d). For N and ν fixed, let \widehat{s} be defined by Eq. (8). Function $m \mapsto f_d(m)$ from $\{1, \dots, N\}$ into \mathbb{R}^+ , defined by Eq. (32b), is such that $f_d(1) = 1 + \widehat{s}^2/(N-1)$ and $f_d(N) = N \widehat{s}^2/(N-1)$. Let m_{opt} be the value of m defined in Hypothesis 1. If

$$\varepsilon_d(m_{\text{opt}})^2 < \frac{\widehat{s}^2}{N-1} < \varepsilon_d(m_{\text{opt}} - 1)^2, \quad (40)$$

then function $m \mapsto f_d(m)$ has a unique local minimum that is a global minimum, which is reached for $m = m_{\text{opt}}$.

$$m_{\text{opt}} = \arg \min_{1 \leq m \leq N} f_d(m). \quad (41)$$

PROOF. The value of $f_d(1)$ and $f_d(N)$ are directly deduced from Eq. (32b) and also from the values $\varepsilon_d(1) = 1$ and $\varepsilon_d(N) = 0$ (see Lemma 4). (i) Let m be such that $m_{\text{opt}} \leq m \leq N-1$, and let $\Delta_m^+ = f_d(m+1) - f_d(m) = \widehat{s}^2/(N-1) + \varepsilon_d(m+1)^2 - \varepsilon_d(m)^2$. Since ε_d is a decreasing function (see Lemma 4), $\varepsilon_d(m)^2 \leq \varepsilon_d(m_{\text{opt}})^2$, and consequently, $\Delta_m^+ \geq \widehat{s}^2/(N-1) + \varepsilon_d(m+1)^2 - \varepsilon_d(m_{\text{opt}})^2$. Since $\varepsilon_d(m+1)^2 > 0$ and since $\widehat{s}^2/(N-1) - \varepsilon_d(m_{\text{opt}})^2 > 0$ (due to Eq. (40)), we have $\Delta_m^+ > 0$ and therefore, f_d is an increasing function on $\{m_{\text{opt}}, \dots, N\}$. (ii) Let m be such that $1 \leq m \leq m_{\text{opt}}$, and let $\Delta_m^- = f_d(m) - f_d(m-1) = \widehat{s}^2/(N-1) + \varepsilon_d(m)^2 - \varepsilon_d(m-1)^2$. From Hypothesis 2, it can be deduced that $\varepsilon_d(m)^2 - \varepsilon_d(m-1)^2 < 0$. For all $2 \leq m \leq m_{\text{opt}}$, we have $\varepsilon_d(m-1)^2 \geq \varepsilon_d(m_{\text{opt}} - 1)^2 \gg \varepsilon_d(m_{\text{opt}})^2$ and Eq. (40) shows that $\Delta_m^- < 0$. (iii) Since $\Delta_m^- < 0$ for $1 \leq m \leq m_{\text{opt}}$ and $\Delta_m^+ > 0$ for $m_{\text{opt}} \leq m \leq N$ yield Eq. (41).

Theorem 4 (Existence of a minimum of $d_N^2(m)$ for $m < N$). Let $\mathcal{M}_{\text{opt}} = \{m_{\text{opt}}, m_{\text{opt}} + 1, \dots, N\}$ in which m_{opt} is defined in Hypothesis 1.

$$\text{If } \forall m \in \mathcal{M}_{\text{opt}} \quad , \quad r(m) \leq 1, \quad (42a)$$

$$\text{then } \min_{m \in \mathcal{M}_{\text{opt}}} d_N^2(m) \leq \min_{m \in \mathcal{M}_{\text{opt}}} d_N^{2,\text{sup}}(m) < d_N^2(N), \quad (42b)$$

in which $d_N^{2,\text{sup}}(m)$ is written as

$$d_N^{2,\text{sup}}(m) = 1 + \frac{m}{N-1}. \quad (43)$$

Eqs. (42a) and (42b) shows that

$$\min_{m \in \mathcal{M}_{\text{opt}}} d_N^2(m) \leq 1 + \frac{m_{\text{opt}}}{N-1} < d_N^2(N), \quad (44)$$

which means that, if hypothesis defined by Eq. (42a) holds, then the PLoM method is a better method than the usual one corresponding to $d_N^2(N)$.

PROOF. Eqs. (32a) and (39) yield $d_N^2(m) = f_d(m) + r(m) \underline{g}(m)$. If $r(m) \leq 1$ for all m in \mathcal{M}_{opt} , then

$$d_N^2(m) \leq d_N^{2,\text{sup}}(m) \quad , \quad \forall m \in \mathcal{M}_{\text{opt}}, \quad (45)$$

in which $d_N^{2,\text{sup}}(m) = f_d(m) + \underline{g}(m)$. From Eqs. (32b) and (37), it can be seen that $d_N^{2,\text{sup}}(m) = 1 + m \widehat{s}^2/(N-1) +$

$(\widehat{s}^2/s^2)(m/N)$ that can be rewritten, using Eq. (8), as Eq. (43). For $m = N$, Eq. (43) and Lemma 2 yield $d_N^{2,\text{sup}}(N) = d_N^2(N) = 1 + N/(N-1)$. From Eqs. (43) and (45), it can then be deduced that $\min_{m \in \mathcal{M}_{\text{opt}}} d_N^2(m) \leq \min_{m \in \mathcal{M}_{\text{opt}}} d_N^{2,\text{sup}}(m) = 1 + m_{\text{opt}}/(N-1)$, and since $d_N^2(N) = 1 + N/(N-1)$, we obtain Eq. (44).

Remark 5 (Concerning the hypothesis $r(m) \leq 1, \forall m \in \mathcal{M}_{\text{opt}}$). For $m \in \{1, \dots, N\}$, $r(m)$ defined by Eq. (36) does not seem to be calculable either explicitly or numerically (since there are N^N elements in set \mathcal{J}). This is the reason why we have introduced the hypothesis defined by Eq. (42a) in order to formulate the theorem. Obviously, this hypothesis has numerically been verified by a direct Monte Carlo simulation of $d_N^2(m)$ given in Definition 6 using Section 6.3 and Remark 3. In Section 8, we give additional developments and comments about this hypothesis.

8. Justification of the hypothesis introduced in Theorem 4

This theorem holds if the hypothesis defined by Eq. (42a) is verified. As explained in Remark 5, $r(m)$ defined by Eq. (36) cannot explicitly be calculated for $1 \leq m \leq N-1$ (for $m = N$, we have $r(N) = 1$). The authors have not been able to perform the exact calculation and therefore this question can be considered as an open problem. Consequently, Section 8 has been developed to support the validation of this hypothesis. We then propose an estimation of $r(m)$ using the maximum entropy principle from Information Theory, and finally, we propose a rough approximation of $r(m)$. In Section 9, devoted to a numerical illustration, we will compare the two last estimations of $d_N^2(m)$ with the "true" function $d_N^2(m)$ estimated as explained in Remark 5.

Remark 6 (Preliminary remark). For all \mathbf{j} in \mathcal{J} and for all m in $\{1, \dots, N\}$, let $a_j(m) \geq 0$ be defined in Proposition 2-(ii). Therefore, $\gamma_j(m)$, which is defined by Eq. (23b), can be rewritten as $\gamma_j(m) = \exp\{-\frac{1}{2s^2} a_j(m)\}$ and consequently, $p_j(m)$ defined by Eq. (23a), can also be rewritten as

$$p_j(m) = \frac{\exp\{-\frac{1}{2s^2} a_j(m)\}}{\sum_{j \in \mathcal{J}} \exp\{-\frac{1}{2s^2} a_j(m)\}} \quad , \quad \sum_{j \in \mathcal{J}} p_j(m) = 1. \quad (46)$$

The discrete random variable $A(m)$ with values in $\{a_j(m), \mathbf{j} \in \mathcal{J}\}$, whose probability distribution $\{p_j(m), \mathbf{j} \in \mathcal{J}\}$ is defined by Eq. (46), is a Maxwell-Boltzmann distribution. Its mean value is $\underline{a}(m) = E\{A(m)\} = \sum_{j \in \mathcal{J}} a_j(m) p_j(m)$, and its entropy is written as $S(\{p_j(m)\}_j) = -\sum_{j \in \mathcal{J}} p_j(m) \log p_j(m) = \frac{1}{2s^2} \underline{a}(m) + \log\left(\sum_{j \in \mathcal{J}} \exp\{-\frac{1}{2s^2} a_j(m)\}\right)$. From Eq. (36), we then have to calculate, for all $m \in \{1, \dots, N-1\}$, $r(m) = \sum_{j \in \mathcal{J}} (g_j(m) / \underline{g}(m)) p_j(m)$, in which $g_j(m)$ is defined by Eq. (35). As we have explained, such a calculation cannot be performed neither explicitly nor numerically (there are N^N elements in \mathcal{J}).

In the following, we construct an estimation of $r(m)$ using the maximum entropy principle.

Definition 8 (Discrete random matrix $[\mathbf{A}]$). Let $[\mathbf{A}]$ be the discrete random variable with values in $\{[\eta_d(\mathbf{j})], \mathbf{j} \in \mathcal{J}\}$ with $[\eta_d(\mathbf{j})] \in \mathbb{M}_{v,N}$ defined by Eq. (12), and for which the probability distribution is $\{\widehat{p}_j, \mathbf{j} \in \mathcal{J}\}$ with $\widehat{p}_j = 1/N^N$,

$$P_{[\mathbf{A}]}(d[a]) = \sum_{j \in \mathcal{J}} \widehat{p}_j \delta_{0_{\mathbb{M}_{v,N}}}([a] - [\eta_d(\mathbf{j})]). \quad (47)$$

Lemma 7 (Second-order moments of random matrix $[\mathbf{A}]$).

$$E\{[\mathbf{A}]\} = [0_{v,N}] \quad , \quad E\{[\mathbf{A}]^T [\mathbf{A}]\} = \frac{1}{N} \|\eta_d\|^2 [I_N]. \quad (48)$$

PROOF. It can be seen that $E\{[\mathbf{A}]\} = \int_{\mathbb{M}_{v,N}} [a] \sum_{j \in \mathcal{J}} \widehat{p}_j \delta_{0_{\mathbb{M}_{v,N}}}([a] - [\eta_d(\mathbf{j})])$ and $E\{[\mathbf{A}]^T [\mathbf{A}]\} = \int_{\mathbb{M}_{v,N}} [a]^T [a] \sum_{j \in \mathcal{J}} \widehat{p}_j \delta_{0_{\mathbb{M}_{v,N}}}([a] - [\eta_d(\mathbf{j})])$ yielding $E\{[\mathbf{A}]\} = N^{-N} \sum_{j \in \mathcal{J}} [\eta_d(\mathbf{j})]$ and $E\{[\mathbf{A}]^T [\mathbf{A}]\} = N^{-N} \sum_{j \in \mathcal{J}} [\eta_d(\mathbf{j})]^T [\eta_d(\mathbf{j})]$. Using Eqs. (28a) and (24) with Eq. (28b) yield Eq. (48).

Lemma 8 (Expression of $r(m)$ as a function of random matrix $[\mathbf{A}]$). Let m be fixed in $\{1, \dots, N\}$. Function $h_d(m) = r(m) \underline{g}(m)$ defined by Eq. (39) in which $r(m)$ is defined by Eq. (36) can be rewritten as

$$r(m) \underline{g}(m) = 1 - \varepsilon_d(m)^2 + \frac{1}{\underline{\gamma}(m) \|\eta_d\|^2} < \frac{\widehat{s}^2}{s^2} [T_2(m)] - 2 \frac{\widehat{s}}{s} [T_1(m)] [\eta_d], [G_m] >_F, \quad (49a)$$

$$\underline{\gamma}(m) = E \left\{ \exp \left(- \frac{1}{2s^2} < [I_N] - [G_m], [\mathbf{A}]^T [\mathbf{A}] >_F \right) \right\}, \quad (49b)$$

$$[T_1(m)] = E \left\{ [\mathbf{A}]^T \exp \left(- \frac{1}{2s^2} < [I_N] - [G_m], [\mathbf{A}]^T [\mathbf{A}] > \right) \right\}, \quad (49c)$$

$$[T_2(m)] = E \left\{ [\mathbf{A}]^T [\mathbf{A}] \exp \left(- \frac{1}{2s^2} < [I_N] - [G_m], [\mathbf{A}]^T [\mathbf{A}] >_F \right) \right\}. \quad (49d)$$

PROOF. For all \mathbf{j} in \mathcal{J} , $g_j(m)$ (defined by Eq. (35)) can be rewritten, using the proof of Proposition 4 and Eq. (27b), as follows,

$$g_j(m) = \|\eta_d\|^{-2} \{ \|\eta_d^m\|^2 + < (\widehat{s}^2/s^2) [M_d(\mathbf{j})] - 2(\widehat{s}/s) [\eta_d(\mathbf{j})]^T [\eta_d], [G_m] >_F \},$$

or using Eq. (31),

$$g_j(m) = 1 - \varepsilon_d(m)^2 + \|\eta_d\|^{-2} \{ < (\widehat{s}^2/s^2) [\eta_d(\mathbf{j})]^T [\eta_d(\mathbf{j})] - 2(\widehat{s}/s) [\eta_d(\mathbf{j})]^T [\eta_d], [G_m] >_F \}.$$

Substituting this expression of $g_j(m)$ into Eq. (36) and using Eq. (35) yield Eq. (49a) in which $\underline{\gamma}(m) = \sum_{\mathbf{j} \in \mathcal{J}} \widehat{p}_{\mathbf{j}} \exp(-\frac{1}{2s^2} < [I_N] - [G_m], [\eta_d(\mathbf{j})]^T [\eta_d(\mathbf{j})] >_F)$, where $[T_1(m)] = \sum_{\mathbf{j} \in \mathcal{J}} \widehat{p}_{\mathbf{j}} [\eta_d(\mathbf{j})]^T \exp(-\frac{1}{2s^2} < [I_N] - [G_m], [\eta_d(\mathbf{j})]^T [\eta_d(\mathbf{j})] >_F)$ and $[T_2(m)] = \sum_{\mathbf{j} \in \mathcal{J}} \widehat{p}_{\mathbf{j}} [\eta_d(\mathbf{j})]^T [\eta_d(\mathbf{j})] \exp(-\frac{1}{2s^2} < [I_N] - [G_m], [\eta_d(\mathbf{j})]^T [\eta_d(\mathbf{j})] >_F)$. Using Eq. (47) allows $\underline{\gamma}(m)$, $[T_1(m)]$, and $[T_2(m)]$ to be rewritten as Eqs. (49b), (49c), and (49d).

Below, an approximation $[\mathbf{A}^c]$ of random matrix $[\mathbf{A}]$ is constructed using the maximum entropy principle [35, 36, 37, 38] under the available information defined by Eq. (48).

Definition 9 (Random matrix $[\mathbf{A}^c]$). Let $[\mathbf{A}^c]$ be the random matrix with values in $\mathbb{M}_{v,N}$ whose probability measure $P_{[\mathbf{A}^c]}(d[a])$ is defined by a pdf $[a] \mapsto p_{[\mathbf{A}^c]}([a])$ on $\mathbb{M}_{v,N}$ with respect to $d[a]$. This pdf is constructed as the unique solution of the following maximum entropy (MaxEnt) problem,

$$p_{[\mathbf{A}^c]} = \max_{p \in \mathcal{C}_{\text{ad}}} S(p), \quad (50)$$

in which the entropy is written as $S(p) = - \int_{\mathbb{M}_{v,N}} p([a]) \log(p([a])) d[a]$ and where the admissible set is defined by $\mathcal{C}_{\text{ad}} = \{[a] \mapsto p([a]) : \mathbb{M}_{v,N} \rightarrow \mathbb{R}^+, \int_{\mathbb{M}_{v,N}} p([a]) d[a] = 1, \int_{\mathbb{M}_{v,N}} [a] p([a]) d[a] = [0_{v,N}], \int_{\mathbb{M}_{v,N}} [a]^T [a] p([a]) d[a] = (1/N) \|\eta_d\|^2 [I_N]\}$.

Proposition 5 (Explicit expression of pdf $p_{[\mathbf{A}^c]}$). The optimization problem defined by Eq. (50) has a unique solution written, for all $[a]$ in $\mathbb{M}_{v,N}$, as

$$p_{[\mathbf{A}^c]}([a]) = \frac{1}{(2\pi)^{vN/2}} \frac{1}{\sigma^{vN}} \exp \left\{ - \frac{1}{2\sigma^2} \|a\|^2 \right\}, \quad \sigma^2 = \frac{1}{vN} \|\eta_d\|^2 = 1 - \frac{1}{N}. \quad (51)$$

PROOF. The proof is left to the reader.

Remark 7 (Independence of the entries of random matrix $[\mathbf{A}^c]$). Eq. (51) shows that the real-valued random variables $\{[\mathbf{A}^c]_{k\ell} ; k = 1, \dots, v ; \ell = 1, \dots, N\}$ are independent and, each one $\mathbf{A}_{k\ell}^c = [\mathbf{A}^c]_{k\ell}$, is a second-order, centered, Gaussian random variable for which its variance is σ^2 ,

$$p_{\mathbf{A}_{k\ell}^c}(a_{k\ell}) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ - \frac{1}{2\sigma^2} a_{k\ell}^2 \right\}, \quad E\{\mathbf{A}_{k\ell}^c\} = 0, \quad E\{(\mathbf{A}_{k\ell}^c)^2\} = \sigma^2. \quad (52)$$

A simple calculation shows that $[\mathbf{A}^c]$ satisfies the constraints defined by the available information, that is, $E\{[\mathbf{A}^c]\} = [0_{v,N}]$ and $E\{[\mathbf{A}^c]^T [\mathbf{A}^c]\} = (1/N) \|\eta_d\|^2 [I_N]$.

Remark 8 (Comparison of the entropy of measures $P_{[\mathbf{A}]}$ and $P_{[\mathbf{A}^c]}$). Let us compare the entropy of $P_{[\mathbf{A}]}(d[a])$ defined by Eq. (47) with the entropy of $P_{[\mathbf{A}^c]}(d[a]) = p_{[\mathbf{A}^c]}([a]) d[a]$ whose pdf is defined by Eq. (51). Since $\hat{p}_j = 1/N^N$, we have $(P_{[\mathbf{A}]}) = -\sum_{j \in \mathcal{J}} \hat{p}_j \log \hat{p}_j = N \log N$. On the other hand, we have $S(P_{[\mathbf{A}^c]}) = -\int_{\mathbb{M}_{v,N}} p_{[\mathbf{A}^c]}([a]) \log p_{[\mathbf{A}^c]}([a]) d[a] = (\nu N/2)(\log(2\pi e) + \log(1 - 1/N))$. Consequently, $S(P_{[\mathbf{A}^c]})/S(P_{[\mathbf{A}]}) = \nu(2 \log N)^{-1} (\log(2\pi e) + \log(1 - 1/N))$. Clearly, the approximation will be optimal if $S(P_{[\mathbf{A}^c]}) \sim S(P_{[\mathbf{A}]})$, which, for $\nu \geq 2$, is reached if $N \sim (2\pi e)^{\nu/2}$. In general, $N < (2\pi e)^{\nu/2}$ and consequently, the level of uncertainties associated with probability measure $P_{[\mathbf{A}^c]}(d[a])$ is larger than the one for the probability measure $P_{[\mathbf{A}]}$. For instance, in Section 9 devoted to the numerical illustration, we have $\nu = 9$ and $N = 200$, which yields $S(P_{[\mathbf{A}^c]})/S(P_{[\mathbf{A}]}) = 2.4$.

Definition 10 (Approximation $r^c(m)$ of $r(m)$). For all m in $\{1, \dots, N\}$, the approximations $r^c(m)$, $\gamma^c(m)$, $[T_1^c(m)]$, and $[T_2^c(m)]$ of $r(m)$, $\underline{\gamma}(m)$, $[T_1(m)]$, and $[T_2(m)]$ is obtained by replacing $[\mathbf{A}]$ by $[\mathbf{A}^c]$ in Eqs. (49a) to (49d),

$$r^c(m) \underline{g}(m) = 1 - \varepsilon_d(m)^2 + \frac{1}{\gamma^c(m) \|\eta_d\|^2} < \frac{\hat{s}^2}{s^2} [T_2^c(m)] - 2 \frac{\hat{s}}{s} [T_1^c(m)] [\eta_d], [G_m] >_F, \quad (53a)$$

$$\underline{\gamma}^c(m) = E \left\{ \exp \left(-\frac{1}{2s^2} < [I_N] - [G_m], [\mathbf{A}^c]^T [\mathbf{A}^c] >_F \right) \right\}, \quad (53b)$$

$$[T_1^c(m)] = E \left\{ [\mathbf{A}^c]^T \exp \left(-\frac{1}{2s^2} < [I_N] - [G_m], [\mathbf{A}^c]^T [\mathbf{A}^c] >_F \right) \right\}, \quad (53c)$$

$$[T_2^c(m)] = E \left\{ [\mathbf{A}^c]^T [\mathbf{A}^c] \exp \left(-\frac{1}{2s^2} < [I_N] - [G_m], [\mathbf{A}^c]^T [\mathbf{A}^c] >_F \right) \right\}. \quad (53d)$$

Lemma 9 (Explicit calculation of $\underline{\gamma}^c(m)$, $[T_1^c(m)]$, and $[T_2^c(m)]$). For all integer m in $\{1, \dots, N\}$, we have

$$\underline{\gamma}^c(m) = \left(1 + \frac{\sigma^2}{s^2}\right)^{-\nu(N-m)/2}, \quad [T_1^c(m)] = [0_{N,\nu}], \quad [T_2^c(m)] = \underline{\gamma}^c(m) \nu [b_m]^{-1} \quad (54)$$

in which σ^2 is defined by the second Eq. (51) and where the matrix $[b_m]$ is defined by $[b_m] = s^{-2}([I_N] - [G_m]) + \sigma^{-2}[I_N]$ and belongs to \mathbb{M}_N^+ (and thus is invertible).

PROOF. Let f be a mapping on $\mathbb{M}_{v,N}$ such that the following quantity be defined, $\mathcal{L}_f(m) = E \left\{ f([\mathbf{A}^c]) \exp \left(-\frac{1}{2s^2} < [I_N] - [G_m], [\mathbf{A}^c]^T [\mathbf{A}^c] >_F \right) \right\}$. Using Eq. (51), it can be seen that $\mathcal{L}_f(m) = (2\pi\sigma^2)^{-\frac{\nu N}{2}} \int_{\mathbb{M}_{v,N}} f([a]) \exp \left(-\frac{1}{2} < [a]^T [a], [b_m] >_F \right) d[a]$. From Lemma 1-(iv), $[G_m] = \sum_{\alpha=1}^N \mu_\alpha \varphi^\alpha \otimes \varphi^\alpha$ with $< \varphi^\alpha, \varphi^\beta > = \delta_{\alpha\beta}$, $\mu_1 = \dots = \mu_m = 1$, and $\mu_{m+1} = \dots = \mu_N = 0$. Since $[I_N] = \sum_{\alpha=1}^N \varphi^\alpha \otimes \varphi^\alpha$, matrix $[b_m]$ can be rewritten as $[b_m] = \sum_{\alpha=1}^N \zeta_\alpha(m) \varphi^\alpha \otimes \varphi^\alpha$ in which $\zeta_\alpha(m) = 1/\sigma^2$ if $\alpha \leq m$, and $\zeta_\alpha(m) = 1/s^2 + 1/\sigma^2$ if $\alpha > m$. Since $\zeta_\alpha(m) > 0, \forall \alpha$, it can be seen that $[b_m] \in \mathbb{M}_N^+$. Hence, $[b_m]^{-1} = \sum_{\alpha=1}^N (1/\zeta_\alpha(m)) \varphi^\alpha \otimes \varphi^\alpha$ and consequently, $\det\{[b_m]^{-1}\} = (\zeta_1(m) \times \dots \times \zeta_N(m))^{-1} = \sigma^{2N} (1 + \frac{\sigma^2}{s^2})^{-(N-m)}$. Hence, $\mathcal{L}_f(m)$ can be rewritten as

$$\mathcal{L}_f(m) = (1 + \sigma^2/s^2)^{-\nu(N-m)/2} \int_{\mathbb{R}^N} \dots \int_{\mathbb{R}^N} f([\hat{a}]^T) p(\hat{\mathbf{a}}^1) \times \dots \times p(\hat{\mathbf{a}}^\nu) d\hat{\mathbf{a}}^1 \dots d\hat{\mathbf{a}}^\nu,$$

where $\forall k \in \{1, \dots, \nu\}$, $\hat{\mathbf{a}}^k \in \mathbb{R}^N$ is such that $\forall \ell \in \{1, \dots, N\}$, $\hat{a}_\ell^k = \{[\hat{a}]^T\}_{\ell k} = [a]_{k\ell}$, and where

$$p(\hat{\mathbf{a}}^k) = ((2\pi)^{N/2} \sqrt{\det\{[b_m]^{-1}\}})^{-1} \exp\left(-\frac{1}{2} < [b_m] \hat{\mathbf{a}}^k, \hat{\mathbf{a}}^k >\right)$$

is the pdf of a Gaussian centered second-order \mathbb{R}^N -valued random variable $\hat{\mathbf{A}}^k$ whose covariance matrix is $[b_m]^{-1}$. (i) Taking $f([a]) = 1$, Eq. (53b) is written as $\underline{\gamma}^c(m) = (1 + \sigma^2/s^2)^{-\nu(N-m)/2} \prod_{k=1}^\nu \left\{ \int_{\mathbb{R}^N} p(\hat{\mathbf{a}}^k) d\hat{\mathbf{a}}^k \right\}$ that gives the first Eq. (54).

(ii) Taking $f([a]) = [a]^T$, Eq. (53c) is written as

$$[T_1^c(m)] = (1 + \sigma^2/s^2)^{-\nu(N-m)/2} \int_{\mathbb{R}^N} \dots \int_{\mathbb{R}^N} [\hat{a}] p(\hat{\mathbf{a}}^1) \times \dots \times p(\hat{\mathbf{a}}^\nu) d\hat{\mathbf{a}}^1 \dots d\hat{\mathbf{a}}^\nu,$$

which is equal to $[0_{N,\nu}]$ because $\hat{\mathbf{A}}^k$ is centered, and therefore, the second Eq. (54) is proven. (iii) Finally, taking

$f([a]) = [a]^T [a]$, Eq. (53d) is written as

$$[T_2^c(m)] = (1 + \sigma^2/s^2)^{-\nu(N-m)/2} \int_{\mathbb{R}^N} \dots \int_{\mathbb{R}^N} [\widehat{a}][\widehat{a}]^T p(\widehat{\mathbf{a}}^1) \times \dots \times p(\widehat{\mathbf{a}}^\nu) d\widehat{\mathbf{a}}^1 \dots d\widehat{\mathbf{a}}^\nu$$

whose entries are

$$[T_2^c(m)]_{\ell\ell'} = (1 + \sigma^2/s^2)^{-\nu(N-m)/2} \sum_{k=1}^{\nu} \int_{\mathbb{R}^N} \dots \int_{\mathbb{R}^N} \widehat{a}_\ell^k \widehat{a}_{\ell'}^k p(\widehat{\mathbf{a}}^1) \times \dots \times p(\widehat{\mathbf{a}}^\nu) d\widehat{\mathbf{a}}^1 \dots d\widehat{\mathbf{a}}^\nu,$$

which shows that

$$[T_2^c(m)] = (1 + \sigma^2/s^2)^{-\nu(N-m)/2} \sum_{k=1}^{\nu} E\{\widehat{\mathbf{A}}^k \otimes \widehat{\mathbf{A}}^k\}.$$

The third Eq. (54) is then directly deduced.

Proposition 6 (Expression of $r^c(m)$). For all m in $\{1, \dots, N\}$, we have

$$r^c(m) = 1. \quad (55)$$

PROOF. Substituting the second and the third Eq. (54) into Eq. (53a) yields

$$r^c(m) \underline{g}(m) = 1 - \varepsilon_d(m)^2 + (\underline{\gamma}^c(m) \|\eta_d\|^2)^{-1} (\widehat{s}^2/s^2) \underline{\gamma}^c(m) \nu < [b_m]^{-1}, [G_m] >_F.$$

Since $[G_m] = \sum_{\alpha=1}^N \mu_\alpha \boldsymbol{\varphi}^\alpha \otimes \boldsymbol{\varphi}^\alpha$ and using the proof of Lemma 9, we have $< [b_m]^{-1}, [G_m] >_F = \sum_{\alpha=1}^N \sum_{\beta=1}^N (\mu_\beta / \zeta_\alpha(m)) < \boldsymbol{\varphi}^\alpha \otimes \boldsymbol{\varphi}^\alpha, \boldsymbol{\varphi}^\beta \otimes \boldsymbol{\varphi}^\beta >_F = \sum_{\alpha=1}^m 1/\zeta_\alpha(m) = m \sigma^2 = m \|\eta_d\|^2 / (\nu N)$. Therefore, $r^c(m) \underline{g}(m) = 1 - \varepsilon_d(m)^2 + (\widehat{s}^2/s^2)(m/N)$. It can be seen that the right-hand side of this equation is $\underline{g}(m)$ defined by Eq. (37). Consequently, $r^c(m) = 1$.

Remark 9 (MaxEnt approximation $d_N^{2,c}(m)$ of $d_N^2(m)$ for all $m \geq m_{\text{opt}}$). Using Theorem 4 and Eq. (32a), the Max-Ent approximation of $d_N^2(m)$ is defined, for all $m \geq m_{\text{opt}}$, as $d_N^{2,c}(m) = f_d(m) + r^c(m) \underline{g}(m)$ in which $f_d(m)$ is defined by Eq. (32b) and $\underline{g}(m)$ by the first Eq. (37). From Eqs. (55), Theorem 4 and its proof, it can be deduced that for all $m \geq m_{\text{opt}}$, $d_N^{2,c}(m) = f_d(m) + \underline{g}(m) = d_N^{2,\text{sup}}(m)$, and consequently, using Eq. (43),

$$\forall m \geq m_{\text{opt}}, d_N^{2,c}(m) = 1 + \frac{m}{N-1}; \quad d_N^{2,c}(N) = d_N^2(N) = 1 + \frac{N}{N-1}. \quad (56)$$

Remark 10 (Rough approximation $d_N^{2,\text{app}}(m)$ of $d_N^2(m)$ for all $m \geq m_{\text{opt}}$). In this remark, for $m \geq m_{\text{opt}}$, we define a "rough approximation" $r_{\text{app}}(m)$ of $r(m)$ defined by Eq. (36). Let $\mathbf{j}_o = (1, 2, \dots, N) \in \mathcal{J} \subset \mathbb{N}^N$. Eq. (12) shows that $[\eta_d(\mathbf{j}_o)] = [\eta_d]$, and consequently, Eqs. (29) and (34) yield $[\eta_d^m(\mathbf{j}_o)] = [\eta_d^m]$. Hence, Eq. (33) yields $\gamma_{\mathbf{j}_o}(m) = \exp(-(2s^2)^{-1} \|\eta_d - \eta_d^m\|^2)$, which can be rewritten, using Eq. (30), as $\gamma_{\mathbf{j}_o}(m) = \exp(-(2s^2)^{-1} \varepsilon_d(m)^2 \|\eta_d\|^2)$. Let us assume that, for $m \geq m_{\text{opt}}$, $\underline{\gamma}(m) \simeq \underline{\gamma}(N) = 1$ (due to Eq. (38)). Starting from Eq. (36), we define

$$r_{\text{app}}(m) = N^{-N} \left(\sum_{\mathbf{j} \in \mathcal{J}} \gamma_{\mathbf{j}_o}(m) g_{\mathbf{j}}(m) \right) (\underline{\gamma}(N) \underline{g}(m))^{-1} = \gamma_{\mathbf{j}_o}(m).$$

Therefore, for all $m \geq m_{\text{opt}}$, $r_{\text{app}}(m) = \exp(-(2s^2)^{-1} \varepsilon_d(m)^2 \|\eta_d\|^2)$. Since $\varepsilon_d(N) = 0$ (see Lemma 4) and since $\gamma_{\mathbf{j}_o}(N) = 1$, it can be seen that $r_{\text{app}}(N) = r(N) = 1$. Finally, using Eqs. (32a) and (39), the corresponding approximation $d_N^{2,\text{app}}(m)$ of $d_N^2(m)$ is written as

$$\forall m \geq m_{\text{opt}}, d_N^{2,\text{app}}(m) = f_d(m) + \underline{g}(m) \exp\left(-\frac{1}{2s^2} \varepsilon_d(m)^2 \|\eta_d\|^2\right), \quad (57)$$

in which $f_d(m)$ is defined by Eq. (32b), $\underline{g}(m)$ by the first Eq. (37), and $\varepsilon_d(m)$ by Eq. (30). Since $r_{\text{app}}(N) = 1$, using Lemma 6, the third Eq. (37), and Eq. (8) yield $d_N^{2,\text{app}}(N) = d_N^2(N) = 1 + N/(N-1)$.

9. Numerical illustration

The numerical illustration proposed is the application (AP1) of reference [36] (see the "Electronic supplementary material" of the online version of this reference). For reasons of limitation of the paper length, we cannot reproduce the description of this application and we refer the reader to this reference. The reader can find further illustrations concerning the estimation of m_{opt} and the preservation of the concentration of the probability measure in [39, 13, 2, 40, 14, 15, 17, 18, 19, 20].

With respect to the notation introduced in Section 1.1, we have $n_w = 20$, $n_q = 200$, $n = 220$, and $N = 200$. For the PCA (see Section 2) and for $\varepsilon = 10^{-6}$ in the second Eq. (3), we have $\nu = 9$. Consequently, $\text{err}_{\text{PCA}}(\nu) \leq 10^{-6}$. Concerning the nonparametric estimate (see Section 3), the values of the parameters defined by Eq. (8) are $s = 0.615$, $\hat{s} = 0.525$, and $\hat{s}/s = 0.853$. The use of Eqs. (14) and (15) yields $\varepsilon_{\text{opt}} = 60$ and $m_{\text{opt}} = 10$. Parameter κ has been fixed to 1. The graph of function $\alpha \mapsto \log(\lambda_\alpha(\varepsilon_{\text{opt}}))$ (see Section 5.1) is displayed in Fig. 2 (left) and the graph of function $m \mapsto \varepsilon_d(m)$ defined by Eq. (30) is shown in Fig. 2 (right). In order to better visualize these graphs, a zoom has been done for the abscissa ($\alpha \leq 50$ and $m \leq 50$ instead of the upper bound $N = 200$). It can be seen that these graphs are similar to the ones shown in Fig. 1 and that Hypotheses 1 and 2 are well verified. The graph of function

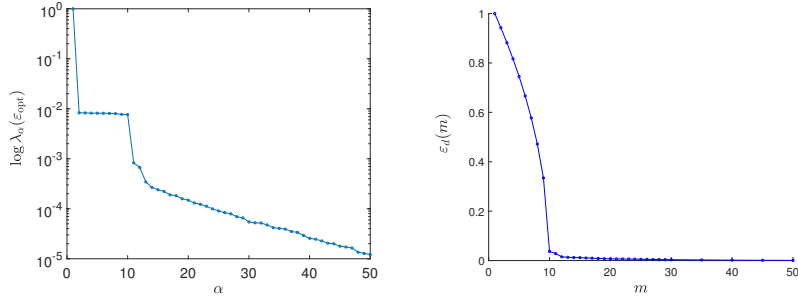


Figure 2: Left figure: distribution of the eigenvalues $\lambda_\alpha(\varepsilon_{\text{opt}})$ in log scale as a function of rank $\alpha \leq 50$ for $\varepsilon_{\text{DM}} = \varepsilon_{\text{opt}} = 60$. Right figure: graph of function $m \mapsto \varepsilon_d(m)$ for $m \leq 50$.

$m \mapsto f_d(m)$ defined by Eq. (32b) is displayed in Fig. 3 (left) and the graph of function $m \mapsto g(m)$ defined by Eq. (37) is shown in Fig. 2 (right). It can be seen that f_d has a minimum for $m = m_{\text{opt}} = 10$. For all m such that $1 \leq m \leq N$,

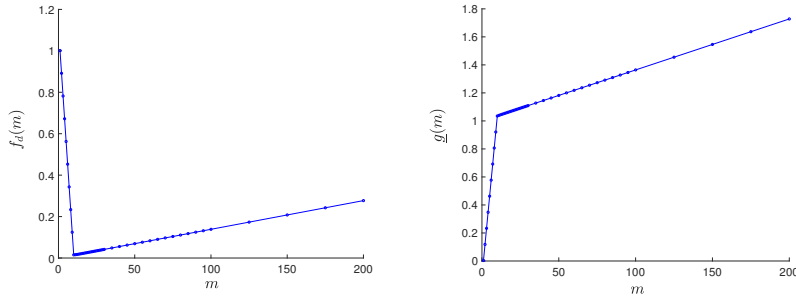


Figure 3: Left figure: graph of function $m \mapsto f_d(m)$. Right figure: graph of function $m \mapsto g(m)$.

the estimation $d_N^{2,\text{sim}}(m) = \|\eta_d\|^{-2} \frac{1}{n_{\text{MC}}} \sum_{\ell=1}^{n_{\text{MC}}} \|\eta_{\text{ar}}^\ell - [\eta_d]\|^2$ of $d_N^2(m)$ defined in Definition 6 has been carried out using the learned dataset $\{\eta_{\text{ar}}^\ell, \ell = 1, \dots, n_{\text{MC}}\}$ with $n_{\text{MC}} = 320\,000$ realizations of random matrix $[\mathbf{H}_m^N]$, which have been computed with the PLoM method presented in Section 6. It has been verified that the L^2 -convergence is obtained for this value of n_{MC} . Left Fig. 4 shows the graph of function $m \mapsto d_N^{2,\text{sim}}(m)$. It can be seen that the local minimum is a global minimum obtained for $m = m_{\text{opt}}$ as expected and that $d_N^{2,\text{sim}}(N) \simeq 2$ (in agreement with Lemma 2). Right Fig. 4 shows three curves: again the graph of $m \mapsto d_N^{2,\text{sim}}(m)$ in order to have a reference, and for $m \geq m_{\text{opt}}$, the graph of function $m \mapsto d_N^{2,c}(m)$ computed with Eq. (56) and the graph of function $m \mapsto d_N^{2,\text{app}}(m)$ computed with Eq. (57). It can

be seen that the graph of $m \mapsto d_N^{2,c}(m)$ is in coherence with Theorem 4 and that the graph of function $m \mapsto d_N^{2,\text{app}}(m)$ has a minimum in $m = m_{\text{opt}} = 10$ on \mathcal{M}_{opt} , as expected.

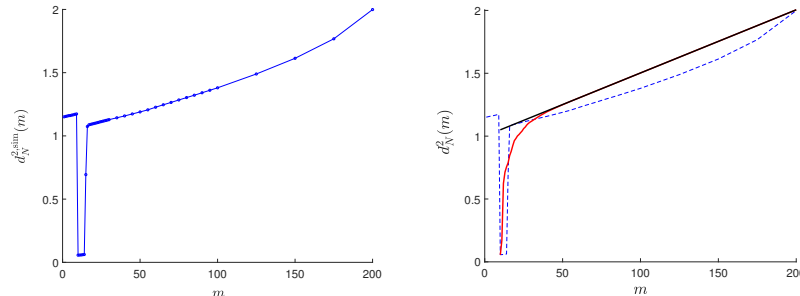


Figure 4: Left figure: graph of function $m \mapsto d_N^{2,\text{sim}}(m)$. Right figure: graph of function $m \mapsto d_N^{2,\text{sim}}(m)$ (blue dashed line), and for $m \geq m_{\text{opt}}$, graphs of $m \mapsto d_N^{2,c}(m)$ (dark thick straight line) and $m \mapsto d_N^{2,\text{app}}(m)$ (red thick curve line).

10. Conclusions

In this paper, we have presented novel mathematical results that justify, highlight, and better explain the probabilistic learning on manifolds proposed in [1]. We have formulated and proven several results, which show that the PLoM methodology is efficient for probabilistic learning as it has been demonstrated in the framework of applications performed for complex engineering systems. The distance introduced for the mathematical analysis of the concentration properties of the probability measure could be used to estimate the optimal dimension of the reduced-order diffusion maps basis and thus to replace the algorithm previously introduced, which uses only the initial dataset. However, the criterion based on this distance would require to generate a large number of replicates of the learned dataset and therefore would induce a larger numerical cost.

Acknowledgments

This research was partially supported by the PIRATE project funded under DARPA's AIRA program and by FASTMATH SciDac Institute supported under DOE's ASCR program.

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. 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.
- [3] T. Duong, M. L. Hazelton, Cross-validation bandwidth matrices for multivariate kernel density estimation, *Scandinavian Journal of Statistics* 32 (3) (2005) 485–506. doi:10.1111/j.1467-9469.2005.00445.x.
- [4] 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.
- [5] M. Filippone, G. Sanguinetti, Approximate inference of the bandwidth in multivariate kernel density estimation, *Computational Statistics & Data Analysis* 55 (12) (2011) 3104–3122. doi:10.1016/j.csda.2011.05.023.
- [6] N. Zougab, S. Adjabi, C. C. Kokonendji, Bayesian estimation of adaptive bandwidth matrices in multivariate kernel density estimation, *Computational Statistics & Data Analysis* 75 (2014) 28–38. doi:10.1016/j.csda.2014.02.002.
- [7] 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.
- [8] E. Parzen, On estimation of a probability density function and mode, *Annals of Mathematical Statistics* 33 (3) (1962) 1065–1076. doi:10.1214/aoms/1177704472.
- [9] 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.
- [10] R. Coifman, S. Lafon, Diffusion maps, *Applied and Computational Harmonic Analysis* 21 (1) (2006) 5–30. doi:10.1016/j.acha.2006.04.006.

- [11] R. Coifman, S. Lafon, Geometric harmonics: A novel tool for multiscale out-of-sample extension of empirical functions, *Applied and Computational Harmonic Analysis* 21 (2006) 31–52. doi:10.1016/j.acha.2005.07.005.
- [12] 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.
- [13] 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.
- [14] R. Ghanem, C. Soize, C. Thimmisetty, Optimal well-placement using probabilistic learning, *Data-Enabled Discovery and Applications* 2 (1) (2018) 4,1–16. doi:10.1007/s41688-017-0014-x.
- [15] R. G. 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.
- [16] J. Guillemot, J. Dolbow, Data-driven enhancement of fracture paths in random composites, *Mechanics Research Communications* 103 (103443) (2019) 1–12. doi:10.1016/j.mechrescom.2019.103443.
- [17] 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.
- [18] C. Soize, C. Farhat, Probabilistic learning for modeling and quantifying model-form uncertainties in nonlinear computational mechanics, *International Journal for Numerical Methods in Engineering* 117 (2019) 819–843. doi:10.1002/nme.5980.
- [19] C. Soize, R. G. 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.
- [20] C. Soize, R. G. 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.
- [21] C. Soize, R. G. 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.
- [22] N. G. Trillos, F. Hoffmann, B. Hosseini, Geometric structure of graph laplacian embeddings, arXiv:1901.10651, (2019).
- [23] C. Soize, Construction of probability distributions in high dimension using the maximum entropy principle. applications to stochastic processes, random fields and random matrices, *International Journal for Numerical Methods in Engineering* 76 (10) (2008) 1583–1611. doi:10.1002/nme.2385.
- [24] R. Neal, MCMC using hamiltonian dynamics, in: S. Brooks, A. Gelman, G. Jones, X.-L. Meng (Eds.), *Handbook of Markov Chain Monte Carlo*, Chapman and Hall-CRC Press, Boca Raton, 2011, Ch. 5. doi:10.1201/b10905-6.
- [25] M. Girolami, B. Calderhead, Riemann manifold Langevin and Hamiltonian Monte Carlo methods, *Journal of the Royal Statistics Society* 73 (2) (2011) 123–214. doi:10.1111/j.1467-9868.2010.00765.x.
- [26] J. Kaipio, E. Somersalo, *Statistical and Computational Inverse Problems*, Vol. 160, Springer Science & Business Media, 2005. doi:10.1007/b138659.
- [27] C. Robert, G. Casella, *Monte Carlo Statistical Methods*, Springer Science & Business Media, 2005. doi:10.1007/978-1-4757-4145-2.
- [28] J. C. Spall, *Introduction to Stochastic Search and Optimization*, Wiley-Interscience, 2003. doi:10.1002/0471722138.
- [29] 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.
- [30] P. Krée, C. Soize, *Mathematics of Random Phenomena*, Reidel Pub. Co, 1986, (first published by Bordas in 1983 and also published by Springer Science & Business Media in 2012).
- [31] J. L. Doob, *Stochastic processes*, John Wiley & Sons, New York, 1953.
- [32] R. Khasminskii, *Stochastic Stability of Differential Equations*, Vol. 66, Springer-Verlag, Berlin, Heidelberg, 2012, originally published in Russian, by Nauka, Moskow, 1969. First English edition published in 1980 under R.Z. Has'minski in the series *Mechanics: Analysis by Sijthoff & Noordhoff*. doi:10.1007/978-3-642-23280-0.
- [33] E. Hairer, C. Lubich, G. Wanner, *Geometric Numerical Integration. Structure-Preserving Algorithms for Ordinary Differential Equations*, Vol. 31, Springer Science & Business Media, 2006. doi:10.1007/3-540-30666-8.
- [34] K. Burrage, I. Lenane, G. Lythe, Numerical methods for second-order stochastic differential equations, *SIAM Journal on Scientific Computing* 29 (1) (2007) 245–264. doi:10.1137/050646032.
- [35] 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.
- [36] E. T. Jaynes, Information theory and statistical mechanics, *Physical Review* 106 (4) (1957) 620–630. doi:10.1103/PhysRev.106.620.
- [37] E. T. Jaynes, Information theory and statistical mechanics. ii, *Physical Review* 108 (2) (1957) 171–190. doi:10.1103/PhysRev.108.171.
- [38] T. M. Cover, J. A. Thomas, *Elements of Information Theory*, Second Edition, John Wiley & Sons, Hoboken, 2006.
- [39] 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 online*, 18 August 2020 (2020) 1–34. doi:10.1615/Int.J.UncertaintyQuantification.2020032674.
- [40] 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* (2020) 1–21doi:10.1002/nme.6430.