



HAL
open science

Probabilistic nonconvex constrained optimization with fixed number of function evaluations

Roger Ghanem, Christian Soize

► **To cite this version:**

Roger Ghanem, Christian Soize. Probabilistic nonconvex constrained optimization with fixed number of function evaluations. *International Journal for Numerical Methods in Engineering*, 2018, 113 (4), pp.719-741. 10.1002/nme.5632 . hal-01576263

HAL Id: hal-01576263

<https://hal.science/hal-01576263>

Submitted on 22 Aug 2017

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 Nonconvex Constrained Optimization with Fixed Number of Function Evaluations *

Roger Ghanem^{1*} and Christian Soize²

¹University of Southern California, 210 KAP Hall, Los Angeles, CA 90089, United States. ²Université Paris-Est, Laboratoire Modélisation et Simulation Multi-Echelle, MSME UMR 8208 CNRS, 5 bd Descartes, 77454 Marne-La-Vallée, France.

SUMMARY

A methodology is proposed for the efficient solution of probabilistic nonconvex constrained optimization problems with uncertain. Statistical properties of the underlying stochastic generator are characterized from an initial statistical sample of function evaluations. A diffusion manifold over the initial set of data points is first identified and an associated basis computed. The joint probability density function of this initial set is estimated using a kernel density model and an Itô stochastic differential equation constructed with this model as its invariant measure. This ISDE is adapted to fluctuate around the manifold yielding additional joint realizations of the uncertain parameters, design variables, and function values are obtained as solutions of the ISDE. The expectations in the objective function and constraints are then accurately evaluated without performing additional function evaluations. The methodology brings together novel ideas from manifold learning and stochastic Hamiltonian dynamics to tackle an outstanding challenge in stochastic optimization. Three examples are presented to highlight different aspects of the proposed methodology.

Received . . .

KEY WORDS: Optimization under uncertainty, Probabilistic optimization, Nonconvex constrained optimization, Probability distribution on manifolds, MCMC generator, Diffusion maps

1. INTRODUCTION

The efficient exploration of the set of design parameters is crucial to the optimization of expensive functions. The development of mathematical and algorithmic constructs that promote learning with successive optimization steps continues to be a key challenge in that regard. These methods have progressed along many directions, including gradient-based learning, adapted to convex problems [1, 2] and global search algorithms including stochastic, genetic, and evolutionary algorithms [3, 4]. Statistical learning methods, whereby a deterministic problem is construed as the representative from a class of stochastic problems have also been developed with the benefit of enabling statistical learning [5]. The learning process is typically manifested in the form of a surrogate model from which approximations of the expensive function can be readily evaluated [6, 7]. The resulting error and its repercussions on the attained optimal solution distinguish the various algorithms [8]. The global character of the surrogate is typically achieved either through a deterministic interpolation process, or a stochastic model whereby biases induced by complex dependencies between model outputs and design parameters are captured through statistical correlations over parameter space. Although Gaussian process models are most commonly used in this context [9, 10], more robust alternatives based on Bayesian optimization [5, 11, 12] have also proven useful.

*Correspondence to: ¹University of Southern California, 210 KAP Hall, Los Angeles, CA 90089, United States. Email: ghanem@usc.edu

Recent research in the field of uncertainty quantification [13, 14, 15, 16, 17] has underscored the need for optimization algorithms with underlying stochastic operators and constraints. In these situations, referred to as optimization under uncertainty (OUU), the challenge is magnified since for each design point along the optimization path, a sufficiently large statistical sample of function outputs must be computed to evaluate the required expectations [8]. In essence, the function output must be characterized as a stochastic process over the set of design variables in order to facilitate such evaluations. For expensive function evaluations exhibiting uncertainty, computational challenges remain currently significant enough to require simplifying assumptions in the form of surrogate models for the stochastic function itself or approximations to relevant probabilities [18, 19, 20, 7].

The present paper addresses this challenge and proposes an algorithm that maintains the number of function evaluations required for OUU at a level essentially equal to that of the deterministic problem. This is achieved by first recognizing that the expensive function evaluator generates samples that fluctuate around a manifold. An algorithm is then introduced to sample the neighborhood of this manifold from the joint distribution of random parameters and design variables. The underlying manifold is learned from a diffusion process on a data set [21, 22, 23] synthesized by evaluating one sample of the expensive function for each of a handful of design variables. A target multivariate probability density function is then constructed from this data set using nonparametric kernel density estimation and smoothing. An Itô stochastic differential equation, constructed with this distribution as its invariant measure, is then projected on the manifold, ensuring that the ensuing solution remains in the neighborhood of that manifold. The construction of this Itô equation, associated with a stochastic nonlinear dissipative Hamiltonian dynamical system, follows recent developments that accelerate the convergence of MCMC algorithms [24] to their steady-state distribution (invariant measure). Such a generator belongs to the class of Hamiltonian Monte Carlo methods [24, 25, 26], which is an MCMC algorithm [27, 28, 2]. This paper extends recent work by the authors [29], where the above sampling on manifolds was first introduced, to the case where the joint distribution of multiple vectors, is constructed and used to evaluate the conditional expectations that define objective functions and constraints in an OUU problem. The paper is organized as follows. In Section 2, the probabilistic nonconvex constrained optimization problem is defined. Section 3 deals with a methodology based on the use of a nonparametric statistical estimation of the conditional mathematical expectation and the algorithm for evaluating the objective function and the constraints function at any point in the admissible set by using only the given dataset. The method for generating additional samples without performing additional function evaluations and the algorithm are presented in Section 5. Finally three applications are presented for validating the method proposed. Some details concerning the algorithms are given in three Appendices.

Notations

A lower case letter such as x , η , or u , is a real deterministic variable.

A boldface lower case letter such as \mathbf{x} , $\boldsymbol{\eta}$, or \mathbf{u} is a real deterministic vector.

An upper case letter such as X , H , or U , is a real random variable.

A boldface upper case letter, \mathbf{X} , \mathbf{H} , or \mathbf{U} , is a real random vector.

A lower case letter between brackets such as $[x]$, $[\eta]$, or $[u]$, is a real deterministic matrix.

A boldface upper case letter between brackets such as $[\mathbf{X}]$, $[\mathbf{H}]$, or $[\mathbf{U}]$, is a real random matrix.

$(\Theta, \mathcal{T}, \mathcal{P})$: Probability triple.

$\mathbb{N} = \{0, 1, 2, \dots\}$: set of all the null and positive integers.

\mathbb{R} : set of all the real numbers.

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

$\|\mathbf{x}\|$: usual Euclidean norm in \mathbb{R}^n .

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

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

$[x]_{kj}$: entry of matrix $[x]$.

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

$\|[x]\|_F$: Frobenius norm of matrix $[x]$ such that $\|x\|_F^2 = \text{tr}\{[x]^T [x]\}$.

$[I_\nu]$: identity matrix in \mathbb{M}_ν .

$\delta_{kk'}$: Kronecker's symbol such that $\delta_{kk'} = 0$ if $k \neq k'$ and $= 1$ if $k = k'$.

$\mathbb{1}_{\mathcal{A}}(a)$ is the indicator function of set \mathcal{A} : $\mathbb{1}_{\mathcal{A}}(a) = 1$ if $a \in \mathcal{A}$ and $= 0$ if $a \notin \mathcal{A}$.

E : Mathematical expectation.

pdf: probability density function.

ISDE: Itô Stochastic Differential Equation.

MCMC: Markov Chain Monte Carlo.

2. PROBLEM SET-UP

2.1. Definition of the optimization problem to be solved and objective of the paper

Let $m_w \geq 1$ and $m_c \geq 1$ be two given integers. Let $\mathbf{w} = (w_1, \dots, w_{m_w})$ be a vector of design parameters that belongs to an admissible set \mathcal{C}_w , which is a subset of \mathbb{R}^{m_w} . We consider the following probabilistic nonconvex constrained optimization problem with nonlinear constraints,

$$\mathbf{w}^{\text{opt}} = \arg \min_{\substack{\mathbf{w} \in \mathcal{C}_w \\ \mathbf{c}(\mathbf{w}) < 0}} f(\mathbf{w}), \quad (1)$$

in which $\mathbf{w} \mapsto f(\mathbf{w})$ is the objective function defined on \mathcal{C}_w , with values in \mathbb{R} , written as

$$f(\mathbf{w}) = E\{\mathcal{Q}(\mathbf{w})\}, \quad (2)$$

where $\mathbf{w} \mapsto \mathbf{c}(\mathbf{w}) = (c_1(\mathbf{w}), \dots, c_{m_c}(\mathbf{w}))$ is the constraints function defined on \mathcal{C}_w , with values in \mathbb{R}^{m_c} , such that

$$\mathbf{c}(\mathbf{w}) = E\{\mathcal{B}(\mathbf{w})\}, \quad (3)$$

and where E is the mathematical expectation. In equations (2) and (3), $\{\mathcal{Q}(\mathbf{w}), \mathbf{w} \in \mathcal{C}_w\}$ and $\{\mathcal{B}(\mathbf{w}) = (\mathcal{B}_1(\mathbf{w}), \dots, \mathcal{B}_{m_c}(\mathbf{w})), \mathbf{w} \in \mathcal{C}_w\}$ are dependent second-order stochastic processes defined on a probability space $(\Theta, \mathcal{T}, \mathcal{P})$, indexed by \mathcal{C}_w , with values in \mathbb{R} and \mathbb{R}^{m_c} respectively. Consequently, for all \mathbf{w} fixed in \mathcal{C}_w , the random variables $\mathcal{Q}(\mathbf{w})$ and $\mathcal{B}(\mathbf{w})$ are the mappings $\theta \mapsto \mathcal{Q}(\mathbf{w}; \theta)$ and $\theta \mapsto \{\mathcal{B}(\mathbf{w}; \theta)\}$, from Θ into \mathbb{R} and \mathbb{R}^{m_c} respectively, which are such that

$$E\{\mathcal{Q}(\mathbf{w})^2\} = \int_{\Theta} \mathcal{Q}(\mathbf{w}; \theta)^2 d\mathcal{P}(\theta) < +\infty,$$

$$E\{\|\mathcal{B}(\mathbf{w})\|^2\} = \int_{\Theta} \|\mathcal{B}(\mathbf{w}; \theta)\|^2 d\mathcal{P}(\theta) < +\infty.$$

It is assumed that, for \mathbf{w} given in \mathcal{C}_w , the values $f(\mathbf{w})$ and $\mathbf{c}(\mathbf{w})$ of the cost and constraints functions are calculated using a computational model with uncertainties (stochastic computational model), and that the probabilistic optimization problem defined by equation (1) has a unique solution \mathbf{w}^{opt} in \mathcal{C}_w .

This paper proposes a probabilistic formulation that permits to solve the above probabilistic nonconvex constrained optimization problem using a small number of evaluations of the objective and constraints functions, thus limiting the calls to the expensive stochastic computational model.

Remarks.

(i) It should be noted that the constraints function \mathbf{c} defined by equation (3) is quite general. For instance, let us consider the k -th constraint is specified in the form,

$$\text{Proba}\{\mathcal{G}_k(\mathbf{w}) \geq \kappa\} > P_c,$$

in which $\{\mathcal{G}_k(\mathbf{w}), \mathbf{w} \in \mathcal{C}_{\mathbf{w}}\}$ is a second-order real-valued stochastic process of interest, where κ is a given real number, and where P_c is a given level of probability. Consequently, $c_k(\mathbf{w}) = E\{\mathcal{B}_k(\mathbf{w})\}$, with

$$\mathcal{B}_k(\mathbf{w}) = P_c - \mathbb{1}_{[\kappa, +\infty[}(\mathcal{G}_k(\mathbf{w})).$$

(ii) A typical procedure for calculating the mathematical expectations that appear on the right-hand sides of equations (2) and (3) is as follows. For every \mathbf{w} given in $\mathcal{C}_{\mathbf{w}}$, the stochastic computational model evaluates N'_s independent samples (or realizations, or draws), $\mathcal{Q}(\mathbf{w}; \theta_{\ell'})$ and $\mathcal{B}(\mathbf{w}; \theta_{\ell'})$ of random variables $\mathcal{Q}(\mathbf{w})$ and $\mathcal{B}(\mathbf{w})$, for $\theta_{\ell'}$ in Θ with $\ell' = 1, \dots, N'_s$. For N'_s sufficiently large, an accurate estimation of $f(\mathbf{w})$ and $\mathbf{c}(\mathbf{w})$ can be computed according to,

$$E\{\mathcal{Q}(\mathbf{w})\} \simeq \frac{1}{N'_s} \sum_{\ell'=1}^{N'_s} \mathcal{Q}(\mathbf{w}; \theta_{\ell'}) \quad , \quad E\{\mathcal{B}(\mathbf{w})\} \simeq \frac{1}{N'_s} \sum_{\ell'=1}^{N'_s} \mathcal{B}(\mathbf{w}; \theta_{\ell'}). \quad (4)$$

If the optimization algorithm requires N_s evaluations $f(\mathbf{w}_{\ell})$ and $\mathbf{c}(\mathbf{w}_{\ell})$ of f and \mathbf{c} for $\ell = 1, \dots, N_s$, then the stochastic computational model must be called $N'_s \times N_s$ times, which could be prohibitive for expensive function calls. The probabilistic approach proposed in this paper will drastically reduce the number of required calls to the stochastic computational model to a value N of similar magnitude to N_s .

2.2. Definition of the dataset generated by the optimization algorithm with a fixed number of function evaluations and definition of the associated random variables

In this section, we first define the dataset of a fixed number N of data points denoted by $\mathbf{x}^{\ell} = (\mathbf{w}^{\ell}, q^{\ell}, \mathbf{b}^{\ell})$ for $\ell = 1, \dots, N$ that are generated by the optimization algorithm and which require N calls of the stochastic computational model. We then construct the random variables \mathbf{W} , Q , \mathbf{B} , and $\mathbf{X} = (\mathbf{W}, Q, \mathbf{B})$ that admits the N data points $\mathbf{x}^{\ell} = (\mathbf{w}^{\ell}, q^{\ell}, \mathbf{b}^{\ell})$ for $\ell = 1, \dots, N$ as N independent samples. The probabilistic properties of these random variables depend on N , but this dependence is dropped, with any loss of generality, for notational simplicity.

Definition of the dataset. For any \mathbf{w} fixed in $\mathcal{C}_{\mathbf{w}}$ and for any θ_{ℓ} fixed in Θ , let $\mathcal{Q}(\mathbf{w}, \theta_{\ell})$ and $\mathcal{B}(\mathbf{w}, \theta_{\ell})$ be samples of the dependent random variables $\mathcal{Q}(\mathbf{w})$ and $\mathcal{B}(\mathbf{w})$ computed using the stochastic computational model. Let us now consider a fixed number N of values $\mathbf{w}^1, \dots, \mathbf{w}^N$ in $\mathcal{C}_{\mathbf{w}}$ of vector \mathbf{w} . These values can correspond either to a training procedure applied to \mathbf{w} or are some values of \mathbf{w} generated by an optimization algorithm as it explores the feasible domain. Let q^1, \dots, q^N be real numbers in \mathbb{R} and let $\mathbf{b}^1, \dots, \mathbf{b}^N$ be real vectors in \mathbb{R}^{m_c} such that

$$q^{\ell} = \mathcal{Q}(\mathbf{w}^{\ell}, \theta_{\ell}) \in \mathbb{R} \quad , \quad \mathbf{b}^{\ell} = \mathcal{B}(\mathbf{w}^{\ell}, \theta_{\ell}) \in \mathbb{R}^{m_c} \quad , \quad \ell = 1, \dots, N. \quad (5)$$

Let us now introduce the N data points $\mathbf{x}^1, \dots, \mathbf{x}^N$ in \mathbb{R}^n such that

$$\mathbf{x}^{\ell} = (\mathbf{w}^{\ell}, q^{\ell}, \mathbf{b}^{\ell}) \in \mathbb{R}^n = \mathbb{R}^{m_w} \times \mathbb{R} \times \mathbb{R}^{m_c} \quad , \quad \ell = 1, \dots, N, \quad (6)$$

in which

$$n = m_w + 1 + m_c.$$

Definition of the random variables associated with the dataset. We now define the random variables \mathbf{W} , Q , \mathbf{B} , and \mathbf{X} . Let $\mathbf{W} = (W_1, \dots, W_{m_w})$, Q , and $\mathbf{B} = (B_1, \dots, B_{m_c})$ be the second-order random variables defined on $(\Theta, \mathcal{T}, \mathcal{P})$ with values in \mathbb{R}^{m_w} , \mathbb{R} , and \mathbb{R}^{m_c} , statistically dependent, for which the joint probability distribution on $\mathcal{C}_{\mathbf{w}} \times \mathbb{R} \times \mathbb{R}^{m_c}$ is unknown but for which a set of N independent samples is given and is defined as follows. For $\ell = 1, \dots, N$, $\mathbf{w}^{\ell} = (w_1^{\ell}, \dots, w_{m_w}^{\ell})$, q^{ℓ} , and $\mathbf{b}^{\ell} = (b_1^{\ell}, \dots, b_{m_c}^{\ell})$ are N given independent samples of \mathbf{W} , Q , and \mathbf{B} in which \mathbf{w}^{ℓ} are the given vectors in \mathbb{R}^{m_w} introduced above, and where q^{ℓ} and \mathbf{b}^{ℓ} are defined by equation (5). Let $\mathbf{X} = (X_1, \dots, X_n)$ be the second-order random variable defined on $(\Theta, \mathcal{T}, \mathcal{P})$ with values in \mathbb{R}^n

with $n = m_w + 1 + m_c$, that is written as

$$\mathbf{X} = (\mathbf{W}, Q, \mathbf{B}). \quad (7)$$

The probability distribution $P_{\mathbf{X}}(d\mathbf{x})$ will be assumed to be represented by a pdf $p_{\mathbf{X}}(\mathbf{x})$ with respect to the Lebesgue measure $d\mathbf{x}$ on \mathbb{R}^n . While $p_{\mathbf{X}}(\mathbf{x})$ is unknown, $\mathbf{x}^1, \dots, \mathbf{x}^N$ in \mathbb{R}^n defined by equation (6) are N statistically independent samples of \mathbf{X} , which are represented by the matrix,

$$[x_d] = [\mathbf{x}^1 \dots \mathbf{x}^N] \in \mathbb{M}_{n,N}.$$

Remarks.

(i) The stochastic computational model defines a mapping between parameter \mathbf{w} and the random variables $\mathcal{Q}(\mathbf{w})$ and $\mathbf{B}(\mathbf{w})$. The unknown pdf $p_{\mathbf{X}}$ is thus concentrated in a neighborhood of a subset \mathcal{S}_n of \mathbb{R}^n associated with this mapping. This neighborhood of subset \mathcal{S}_n will be discovered by using the method developed in [29] and summarized in Section 4.

(ii) For a fixed number N of function evaluations, the available dataset is thus made up of the deterministic vectors $\mathbf{x}^1, \dots, \mathbf{x}^N$ in \mathbb{R}^n . Consequently, for \mathbf{w}^ℓ given in \mathbb{R}^{m_w} , only one realization $(q^\ell, \mathbf{b}^\ell) = (\mathcal{Q}(\mathbf{w}^\ell, \theta_\ell), \mathbf{B}(\mathbf{w}^\ell, \theta_\ell))$ of random variable $(\mathcal{Q}(\mathbf{w}^\ell), \mathbf{B}(\mathbf{w}^\ell))$ with values in $\mathbb{R} \times \mathbb{R}^{m_c}$ is calculated by using the stochastic computational model. In accordance to Remark (ii) in Section 2.1, for a given value \mathbf{w}^ℓ of \mathbf{w} , $N'_s \gg 1$ samples $\{(\mathcal{Q}(\mathbf{w}^\ell, \theta^{\ell'}), \mathbf{B}(\mathbf{w}^\ell, \theta^{\ell'})), \ell' = 1, \dots, N'_s\}$ are not computed by calling N'_s times the stochastic computational model.

(iii) For any \mathbf{w} fixed in \mathcal{C}_w , the probability distribution of the random vector $(\mathcal{Q}(\mathbf{w}), \mathbf{B}(\mathbf{w}))$ with values in $\mathbb{R} \times \mathbb{R}^{m_w}$ is not explicitly known.

3. METHODOLOGY FOR EVALUATING THE OBJECTIVE FUNCTION AND THE CONSTRAINTS FUNCTION AT ANY POINT \mathbf{w}^0 IN THE ADMISSIBLE SET USING THE DATASET

The available information is only constituted of the fixed number N of data points $\mathbf{x}^\ell = (\mathbf{w}^\ell, q^\ell, \mathbf{b}^\ell)$ for $\ell = 1, \dots, N$, which are N independent samples of the constructed random variable $\mathbf{X} = (\mathbf{W}, Q, \mathbf{B})$ defined by equation (7) (see Section 2.2). The problem consists in calculating, for any point \mathbf{w}^0 given in \mathcal{C}_w , an estimate of $f(\mathbf{w}^0)$ defined by equation (2) and an estimate of $\mathbf{c}(\mathbf{w}^0)$ defined by equation (3). It can easily be seen that, if N was sufficiently large, then these estimates would be written as

$$f(\mathbf{w}^0) \simeq E\{Q \mid \mathbf{W} = \mathbf{w}^0\}, \quad (8)$$

$$c_k(\mathbf{w}^0) \simeq E\{B_k \mid \mathbf{W} = \mathbf{w}^0\} \quad , \quad k = 1, \dots, m_c, \quad (9)$$

in which $E\{Q \mid \mathbf{W} = \mathbf{w}^0\}$ and $E\{B_k \mid \mathbf{W} = \mathbf{w}^0\}$ are the conditional mathematical expectations of random variables Q and B_k given $\mathbf{W} = \mathbf{w}^0$ in \mathcal{C}_w . We then have to estimate these conditional mathematical expectations by using a data smoothing technique based on the available information defined by the dataset. For that, we begin by introducing a generic problem that we solve using nonparametric statistical methods.

3.1. Definition of a generic problem related to the estimation of a conditional expectation from a given dataset

Let (\mathbf{W}, R) be the second-order random variable defined on $(\Theta, \mathcal{T}, \mathcal{P})$ with values in $\mathbb{R}^{m_w} \times \mathbb{R}$ in which \mathbf{W} is the second-order \mathbb{R}^{m_w} -valued random variable defined in Section 2.2 and where R is a second-order \mathbb{R} -valued random variable that depends on \mathbf{W} (R will refer to either Q or B_k). The dataset is made up of $\nu_{\text{sim}} > 1$ given independent samples of (\mathbf{W}, R) ,

$$(\mathbf{w}^\ell, r^\ell) \in \mathbb{R}^{m_w} \times \mathbb{R} \quad , \quad \ell = 1, \dots, \nu_{\text{sim}}. \quad (10)$$

The generic problem consists in estimating $E\{R | \mathbf{W} = \mathbf{w}^0\}$ that is the conditional mathematical expectation of the real-valued random variable R given $\mathbf{W} = \mathbf{w}^0$ in \mathcal{C}_w , using only the ν_{sim} samples $\{(\mathbf{w}^\ell, r^\ell), \ell = 1, \dots, \nu_{\text{sim}}\}$ defined by equation (10).

Remarks.

(i) Taking $\nu_{\text{sim}} = N$, $R = Q$, and $r^\ell = q^\ell$ for $\ell = 1, \dots, N$, and if N is sufficiently large, an estimate of the objective function at point \mathbf{w}^0 is given by equation (8). Similarly, for $k = 1, \dots, m_c$, by taking $\nu_{\text{sim}} = N$, $R = B_k$, and $r^\ell = b_k^\ell$ for $\ell = 1, \dots, N$, and if N is sufficiently large, an estimate of the component $c_k(\mathbf{w}^0)$ of constraints function \mathbf{c} at point \mathbf{w}^0 is given by equation (9).

(ii) As we have explained in Remark (ii) of Section 2.2, the available information is only made up of the samples $(\mathbf{w}^\ell, r^\ell)$ for $\ell = 1, \dots, \nu_{\text{sim}}$. Consequently, for any \mathbf{w}^0 fixed in \mathcal{C}_w , the conditional mathematical expectation $E\{R | \mathbf{W} = \mathbf{w}^0\}$ cannot be estimated using classical statistical method because, for given \mathbf{w}^ℓ , only one sample r^ℓ is given, and these latter require a large number of samples $r^{\ell,1}, \dots, r^{\ell,N'}$ of R . In order to overcome this difficulty, a data smoothing technique based on the use of the Gaussian kernel-density estimation method is used for estimating the conditional mathematical expectation $E\{R | \mathbf{W} = \mathbf{w}^0\}$.

3.2. Nonparametric statistical estimation of the conditional mathematical expectation

In order to apply nonparametric statistics for estimating the conditional mathematical expectation $E\{R | \mathbf{W} = \mathbf{w}^0\}$ on the basis of the dataset that is made up of the ν_{sim} independent realizations $\{(\mathbf{w}^\ell, r^\ell), \ell = 1, \dots, \nu_{\text{sim}}\}$, it is necessary to first normalize the dataset in order to obtain well-conditioned numerical calculations.

Normalizing the dataset. For $j = 1, \dots, m_w$, let \underline{w}_j and σ_j be the empirical estimates of the mean value and of the standard deviation of random variable W_j constructed using the ν_{sim} independent samples $\{w_j^\ell, \ell = 1, \dots, \nu_{\text{sim}}\}$. Similarly, let \underline{r} and σ be the empirical estimates of the mean value and of the standard deviation of random variable R constructed with the ν_{sim} independent samples $\{r^\ell, \ell = 1, \dots, \nu_{\text{sim}}\}$. We then introduce the normalized random variables \widetilde{W}_j for $j = 1, \dots, m_w$ and \widetilde{R} defined by

$$\widetilde{W}_j = (W_j - \underline{w}_j)/\sigma_j \quad , \quad \widetilde{R} = (R - \underline{r})/\sigma \quad , \quad (11)$$

for which the ν_{sim} independent samples are given by

$$\widetilde{w}_j^\ell = (w_j^\ell - \underline{w}_j)/\sigma_j \quad , \quad \widetilde{r}^\ell = (r^\ell - \underline{r})/\sigma \quad , \quad \ell = 1, \dots, \nu_{\text{sim}} \quad . \quad (12)$$

For any $\mathbf{w}^0 = (w_1^0, \dots, w_{m_w}^0)$ fixed in \mathcal{C}_w , we then have,

$$E\{R | \mathbf{W} = \mathbf{w}^0\} = \underline{r} + \sigma E\{\widetilde{R} | \widetilde{\mathbf{W}} = \widetilde{\mathbf{w}}^0\} \quad , \quad (13)$$

in which $\widetilde{\mathbf{W}} = (\widetilde{W}_1, \dots, \widetilde{W}_{m_w})$ and where $\widetilde{\mathbf{w}}^0 = (\widetilde{w}_1^0, \dots, \widetilde{w}_{m_w}^0)$ with

$$\widetilde{w}_j^0 = (w_j^0 - \underline{w}_j)/\sigma_j \quad , \quad j = 1, \dots, m_w \quad . \quad (14)$$

Introducing the joint pdf $p_{\widetilde{\mathbf{W}}, \widetilde{R}}(\widetilde{\mathbf{w}}, \widetilde{r})$ with respect to $d\widetilde{\mathbf{w}} d\widetilde{r}$ of random variables $\widetilde{\mathbf{W}}$ and \widetilde{R} and the pdf $p_{\widetilde{\mathbf{W}}}(\widetilde{\mathbf{w}}) = \int_{\mathbb{R}} p_{\widetilde{\mathbf{W}}, \widetilde{R}}(\widetilde{\mathbf{w}}, \widetilde{r}) d\widetilde{r}$ with respect to $d\widetilde{\mathbf{w}}$ of random variable $\widetilde{\mathbf{W}}$, the conditional mathematical expectation $E\{\widetilde{R} | \widetilde{\mathbf{W}} = \widetilde{\mathbf{w}}^0\}$ can be written as

$$E\{\widetilde{R} | \widetilde{\mathbf{W}} = \widetilde{\mathbf{w}}^0\} = \frac{1}{p_{\widetilde{\mathbf{W}}}(\widetilde{\mathbf{w}}^0)} \int_{\mathbb{R}} \widetilde{r} p_{\widetilde{\mathbf{W}}, \widetilde{R}}(\widetilde{\mathbf{w}}^0, \widetilde{r}) d\widetilde{r} \quad . \quad (15)$$

Nonparametric statistical estimation of the conditional mathematical expectation $E\{\widetilde{R} | \widetilde{\mathbf{W}} = \widetilde{\mathbf{w}}^0\}$. Each one of the dependent random variables $\widetilde{W}_1, \dots, \widetilde{W}_{m_w}$ and \widetilde{R} , has a zero empirical mean value and a unit empirical standard deviation (calculated using the ν_{sim} samples $(\widetilde{\mathbf{w}}^\ell, \widetilde{r}^\ell)$). The

nonparametric statistical estimation of the joint pdf $p_{\tilde{\mathbf{w}}, \tilde{r}}(\tilde{\mathbf{w}}^0, \tilde{r})$ at point $(\tilde{\mathbf{w}}^0, \tilde{r})$, constructed using the Gaussian kernel-density estimation method [30, 31] and using the ν_{sim} independent realizations $\{(\tilde{\mathbf{w}}^\ell, \tilde{r}^\ell), \ell = 1, \dots, \nu_{\text{sim}}\}$, is written, for ν_{sim} sufficiently large, as

$$p_{\tilde{\mathbf{w}}, \tilde{r}}(\tilde{\mathbf{w}}^0, \tilde{r}) \simeq \frac{1}{\nu_{\text{sim}}} \sum_{\ell=1}^{\nu_{\text{sim}}} \frac{1}{(\sqrt{2\pi} s)^{m_w+1}} \exp \left\{ -\frac{1}{2s^2} \{ \|\tilde{\mathbf{w}}^\ell - \tilde{\mathbf{w}}^0\|^2 + (\tilde{r}^\ell - \tilde{r})^2 \} \right\}, \quad (16)$$

in which s is the bandwidth parameter that can be chosen as the usual multidimensional optimal Silverman bandwidth (in taking into account that the empirical estimation of the standard deviation of each component is unity),

$$s = \left\{ \frac{4}{\nu_{\text{sim}}(2 + m_w + 1)} \right\}^{1/(4+m_w+1)}. \quad (17)$$

From equation (16), it can be deduced that $p_{\tilde{\mathbf{w}}}(\tilde{\mathbf{w}}^0) = \int_{\mathbb{R}} p_{\tilde{\mathbf{w}}, \tilde{r}}(\tilde{\mathbf{w}}^0, \tilde{r}) d\tilde{r}$ can be estimated, for ν_{sim} sufficiently large, by

$$p_{\tilde{\mathbf{w}}}(\tilde{\mathbf{w}}^0) \simeq \frac{1}{\nu_{\text{sim}}} \sum_{\ell=1}^{\nu_{\text{sim}}} \frac{1}{(\sqrt{2\pi} s)^{m_w}} \exp \left\{ -\frac{1}{2s^2} \|\tilde{\mathbf{w}}^\ell - \tilde{\mathbf{w}}^0\|^2 \right\}. \quad (18)$$

Using equations (15), (16), and (18), it can be deduced that, for ν_{sim} sufficiently large, an estimate of $E\{\tilde{R} | \tilde{\mathbf{W}} = \tilde{\mathbf{w}}^0\}$ is given by

$$E\{\tilde{R} | \tilde{\mathbf{W}} = \tilde{\mathbf{w}}^0\} \simeq \frac{\sum_{\ell=1}^{\nu_{\text{sim}}} \tilde{r}^\ell \exp \left\{ -\frac{1}{2s^2} \|\tilde{\mathbf{w}}^\ell - \tilde{\mathbf{w}}^0\|^2 \right\}}{\sum_{\ell=1}^{\nu_{\text{sim}}} \exp \left\{ -\frac{1}{2s^2} \|\tilde{\mathbf{w}}^\ell - \tilde{\mathbf{w}}^0\|^2 \right\}}. \quad (19)$$

3.3. Algorithm for estimating the objective function and the constraints function using only the given N data points

The proposed algorithm can be used for estimating the value $f(\mathbf{w}_g^i)$ of the objective function and the value $\mathbf{c}(\mathbf{w}_g^i)$ of the constraints function, in ν_g given points \mathbf{w}_g^i in \mathcal{C}_w belonging to the subset $\mathcal{C}_{w_g} = \{\mathbf{w}_g^i = (w_{g,1}^i, w_{g,2}^i), i = 1, \dots, \nu_g\} \subset \mathcal{C}_w$ using only the N data points represented by matrix $[x_d] \in \mathbb{M}_{n,N}$.

Data allowing the algorithm to be initialized.

1. Given m_w, m_c, N .
2. Deducing $n = m_w + 1 + m_c$.
3. Given $[x_d] = [\mathbf{x}^1 \dots \mathbf{x}^N] \in \mathbb{M}_{n,N}$ such that for $\ell = 1, \dots, N$:
 $\mathbf{x}^\ell = (\mathbf{w}^\ell, q^\ell, \mathbf{b}^\ell) \in \mathbb{R}^n$ with $\mathbf{w}^\ell \in \mathbb{R}^{m_w}, q^\ell \in \mathbb{R}, \mathbf{b}^\ell \in \mathbb{R}^{m_c}$.

Steps of the algorithm for estimating the objective function and the constraints function in \mathcal{C}_{w_g} using only the given N data points.

1. Computing
 - (a) $\underline{q} = \frac{1}{N} \sum_{\ell=1}^N q^\ell$ and $\sigma_q^2 = \frac{1}{N} \sum_{\ell=1}^N (q^\ell - \underline{q})^2$.
 - (b) $\underline{b}_k = \frac{1}{N} \sum_{\ell=1}^N b_k^\ell$ and $\sigma_{b_k}^2 = \frac{1}{N} \sum_{\ell=1}^N (b_k^\ell - \underline{b}_k)^2, k = 1, \dots, m_c$.
2. Applying equation (11) for normalizing the samples, which yields:
 - (a) N samples $(\tilde{\mathbf{w}}^\ell, \tilde{q}^\ell) \in \mathbb{R}^{m_w} \times \mathbb{R}$ for $\ell = 1, \dots, N$.
 - (b) N samples $(\tilde{\mathbf{w}}^\ell, \tilde{b}_k^\ell) \in \mathbb{R}^{m_w} \times \mathbb{R}$ for $\ell = 1, \dots, N$ and for $k = 1, \dots, m_c$.

3. Computing s using equation (17) with $\nu_{\text{sim}} = N$.
4. For each point \mathbf{w}_g^i in $\mathcal{C}_{\mathbf{w}_g} \subset \mathcal{C}_{\mathbf{w}}$, computing $f(\mathbf{w}_g^i)$ and $\mathbf{c}(\mathbf{w}_g^i)$ as follows.
 - (a) Computing the normalization $\tilde{\mathbf{w}}_g^i$ of \mathbf{w}_g^i using equation (14) with $\tilde{\mathbf{w}}^0 = \tilde{\mathbf{w}}_g^i$ and $\mathbf{w}^0 = \mathbf{w}_g^i$.
 - (b) Computing $e^\ell(\tilde{\mathbf{w}}_g^i) = \exp\{-\frac{1}{2s^2}\|\tilde{\mathbf{w}}^\ell - \tilde{\mathbf{w}}_g^i\|^2\}$ for $\ell = 1, \dots, N$.
 - (c) Computing $\gamma(\tilde{\mathbf{w}}_g^i) = \sum_{\ell=1}^N e^\ell(\tilde{\mathbf{w}}_g^i)$.
 - (d) Computing the estimate of $f(\mathbf{w}_g^i)$ using equations (8), (13), and (19):

$$f(\mathbf{w}_g^i) \simeq \underline{q} + \frac{\sigma_q}{\gamma(\tilde{\mathbf{w}}_g^i)} \sum_{\ell=1}^N \tilde{q}^\ell e^\ell(\tilde{\mathbf{w}}_g^i).$$

- (e) For $k = 1, \dots, m_c$, computing the estimate of $c_k(\mathbf{w}_g^i)$ by using equations (9), (13), and (19):

$$c_k(\mathbf{w}_g^i) \simeq b_k + \frac{\sigma_{b_k}}{\gamma(\tilde{\mathbf{w}}_g^i)} \sum_{\ell=1}^N \tilde{b}_k^\ell e^\ell(\tilde{\mathbf{w}}_g^i).$$

End of the algorithm.

4. METHOD FOR GENERATING ADDITIONAL SAMPLES WITHOUT PERFORMING ADDITIONAL FUNCTION EVALUATIONS

As explained in Remark (i) of Section 2.2, the unknown pdf $p_{\mathbf{X}}$ of random variable \mathbf{X} defined by equation (7) is concentrated in the neighborhood of an unknown subset \mathcal{S}_n of \mathbb{R}^n and the number N of its samples, which are represented by matrix $[x_d] \in \mathbb{M}_{n,N}$, is fixed and is relatively small so as to limit the number of calls to the stochastic computational model. With such a small value of N , if we choose $\nu_{\text{sim}} = N$ for estimating the conditional mathematical expectation $E\{\tilde{R} \mid \tilde{\mathbf{W}} = \tilde{\mathbf{w}}^0\}$ defined by equation (19), the estimate may not be sufficiently accurate. As we have explained, the idea is to generate additional samples for \mathbf{X} in order to use equation (19) with $\nu_{\text{sim}} \gg N$ for obtaining a good estimation of the conditional mathematical expectations, without performing any additional function evaluations with the stochastic computational model. For doing that, we need to construct the pdf $p_{\mathbf{X}}$ of \mathbf{X} , to discover the subset \mathcal{S}_n , and to construct the associated generator of additional samples, by using only the dataset $[x_d] \in \mathbb{M}_{n,N}$. A solution to this non-trivial problem is given in applying the recent methodology presented in [29], which is briefly summarized in this section. It should be noted that it was shown in [29] that a direct sampling of a random vector obtained by using a nonparametric statistical estimation of its pdf and a MCMC method for generating additional samples, yields samples that are not concentrated around the subset of interest, but are scattered throughout the ambient Euclidean space. This was indeed the motivation for the method proposed in [29], which is based on the use of the diffusion maps and which has been developed in order to preserve the concentration of the probability measure around the manifold.

4.1. Introducing the random matrix $[\mathbf{X}]$ associated with random vector \mathbf{X} and normalization of $[\mathbf{X}]$ in a random matrix $[\mathbf{H}]$

Let \mathbf{X} be the \mathbb{R}^n -valued second-order random vector defined by equation (7) for which the dataset $[x_d] \in \mathbb{M}_{n,N}$, defined in Section 2.2, is assumed to be scaled. If it is not the case, a scaling must be performed as explained in [29]. Let $[\mathbf{X}] = [\mathbf{X}^1, \dots, \mathbf{X}^N]$ be the random matrix with values in $\mathbb{M}_{n,N}$, whose columns are N independent copies of random vector \mathbf{X} . The normalization of random matrix $[\mathbf{X}]$ is attained with random matrix $[\mathbf{H}] = [\mathbf{H}^1, \dots, \mathbf{H}^N]$ with values in $\mathbb{M}_{\nu,N}$, whose columns are N independent copies of a random vector \mathbf{H} , with $\nu \leq n$, obtained by using principal component analysis resulting in,

$$[\mathbf{X}] = [\underline{x}] + [\varphi] [\lambda]^{1/2} [\mathbf{H}], \quad (20)$$

in which $[\lambda]$ is the $(\nu \times \nu)$ diagonal matrix of the ν positive eigenvalues of the empirical estimate $[\text{cov}] \in \mathbb{M}_n$ of the covariance matrix of \mathbf{X} (computed using equation (55)), where $[\varphi]$ is the $(n \times \nu)$ matrix of the associated eigenvectors such $[\varphi]^T [\varphi] = [I_\nu]$, and where $[\underline{x}]$ is the matrix in $\mathbb{M}_{n,N}$ with identical columns, each equal to the empirical estimate $\underline{\mathbf{x}} \in \mathbb{R}^n$ of the mean value of random vector \mathbf{X} (computed using equation (55)). The sample $[\eta_d] = [\boldsymbol{\eta}^1 \dots \boldsymbol{\eta}^N] \in \mathbb{M}_{\nu,N}$ of $[\mathbf{H}]$ (associated with the sample $[x_d]$ of $[\mathbf{X}]$) is computed by

$$[\eta_d] = [\lambda]^{-1/2} [\varphi]^T ([x_d] - [\underline{x}]). \quad (21)$$

Consequently, the empirical estimates of the mean value and of the covariance matrix of random vector \mathbf{H} are exactly $\mathbf{0}_\nu$ and $[I_\nu]$, respectively. Such a normalization is required for obtaining well-conditioned numerical calculations. In low dimension (n not too big), no statistical reduction is performed *a priori*, and if all the eigenvalues are positive, then $\nu = n$; if some eigenvalues are zero, they must be eliminated and then $\nu < n$. In high dimension (n is big), a statistical reduction can be done as usual and therefore $\nu < n$ in such a case.

4.2. Reduced-order representation of random matrix $[\mathbf{H}]$ by projection on a subspace spanned by a diffusion-maps basis

As previously explained, the introduction of a reduced-order representation of random matrix $[\mathbf{H}]$, which is constructed by projection on a diffusion-maps basis, allows for preserving the concentration of the probability measure of random matrix $[\mathbf{H}]$. Let $k_\varepsilon(\boldsymbol{\eta}, \boldsymbol{\eta}') = \exp(-\frac{1}{4\varepsilon} \|\boldsymbol{\eta} - \boldsymbol{\eta}'\|^2)$ be the kernel defined on $\mathbb{R}^\nu \times \mathbb{R}^\nu$, depending on a real smoothing parameter $\varepsilon > 0$. This kernel can be replaced by another one satisfying the symmetry, the positivity preserving, and the positive semi-definiteness properties. For $m \leq N$, let $[g] = [\mathbf{g}^1 \dots \mathbf{g}^m] \in \mathbb{M}_{N,m}$ be the "diffusion-maps basis" associated with kernel k_ε , which is defined and constructed in Appendix A (for $m = N$, $[g]$ is an algebraic basis of \mathbb{R}^N). For $\alpha = 1, \dots, m$, the diffusion-maps vector $\mathbf{g}^\alpha \in \mathbb{R}^N$ is defined by equation (50). The subspace of \mathbb{R}^N spanned by the vector basis $\{\mathbf{g}^\alpha\}_\alpha$ allows for characterizing the local geometry structure of the dataset concentrated in the neighborhood of a subset of \mathbb{R}^N . The reduced-order representation is obtained in projecting each column of the $\mathbb{M}_{N,\nu}$ -valued random matrix $[\mathbf{H}]^T$ on the subspace of \mathbb{R}^N , spanned by $\{\mathbf{g}^1 \dots \mathbf{g}^m\}$. Introducing the random matrix $[\mathbf{Z}]$ with values in $\mathbb{M}_{\nu,m}$, the following reduced-order representation of $[\mathbf{H}]$ is defined,

$$[\mathbf{H}] = [\mathbf{Z}] [g]^T. \quad (22)$$

As the matrix $[g]^T [g] \in \mathbb{M}_m$ is invertible, equation (22) yields the least squares approximation to \mathbf{Z} in the form,

$$[\mathbf{Z}] = [\mathbf{H}] [a] \quad , \quad [a] = [g] ([g]^T [g])^{-1} \in \mathbb{M}_{N,m}. \quad (23)$$

In particular, matrix $[\eta_d] \in \mathbb{M}_{\nu,N}$ can be written as $[\eta_d] = [z_d] [g]^T$ in which the matrix $[z_d] \in \mathbb{M}_{\nu,m}$ is given by

$$[z_d] = [\eta_d] [a] \in \mathbb{M}_{\nu,m}. \quad (24)$$

Consequently, the following representation of random matrix $[\mathbf{X}]$ as function of random matrix $[\mathbf{Z}]$ is deduced from equations (20) and (23),

$$[\mathbf{X}] = [\underline{x}] + [\varphi] [\lambda]^{1/2} [\mathbf{Z}] [g]^T. \quad (25)$$

The dimension m of the reduced-order representation is estimated by analyzing the convergence of the representation with respect to m . For a given value of integer ζ related to the analysis scale of the local geometric structure of the dataset (see equation (50) in Appendix A) and for a given value of the smoothing parameter $\varepsilon > 0$, the decay in the graph $\alpha \mapsto \Lambda_\alpha$ of the positive eigenvalues of the transition matrix $[\mathbb{P}]$ (see Appendix A) yields a criterion for choosing the value of m that allows the local geometric structure of the dataset represented by $[\eta_d]$ to be discovered. Nevertheless, this criterion may not be sufficient, and the L^2 -convergence may need to be enforced by increasing, as required, the value of m . However, if the value of m is chosen too large, the localization of the geometric structure of the dataset is lost. Consequently, a compromise must be reached between a

very small value of m identified by the decay of the eigenvalues and a larger value of m necessary for obtaining a reasonable mean-square convergence. A criterion for estimating an optimal value of m is given in Appendix B.

4.3. Generation of additional samples $[x_{\text{ar}}^1], \dots, [x_{\text{ar}}^{n_{\text{MC}}}]$ of random matrix $[\mathbf{X}]$

The generation of additional samples $[z_{\text{ar}}^1], \dots, [z_{\text{ar}}^{n_{\text{MC}}}]$ of random matrix $[\mathbf{Z}]$ is carried out by using an unusual MCMC method which is constructed as the projection on the diffusion-maps basis of an ISDE related to a dissipative Hamiltonian dynamical system for which the invariant measure is the pdf of random matrix $[\mathbf{H}]$ constructed with the Gaussian kernel-density estimation method. This method preserves the concentration of the probability measure and avoids the scatter phenomenon described above. For m , ε , and ζ fixed, we introduce the Markov stochastic process $\{([\mathbf{Z}(r)], [\mathbf{Y}(r)]), r \in \mathbb{R}^+\}$, defined on $(\Theta, \mathcal{T}, \mathcal{P})$, indexed by $\mathbb{R}^+ = [0, +\infty[$, with values in $\mathbb{M}_{\nu, m} \times \mathbb{M}_{\nu, m}$, which is the unique second-order stationary (for the shift semi-group on \mathbb{R}^+) and ergodic diffusion stochastic process, of the following reduced-order ISDE, for $r > 0$,

$$d[\mathbf{Z}(r)] = [\mathbf{Y}(r)] dr, \quad (26)$$

$$d[\mathbf{Y}(r)] = [\mathcal{L}([\mathbf{Z}(r)])] dr - \frac{1}{2} f_0 [\mathbf{Y}(r)] dr + \sqrt{f_0} [d\mathbf{W}(r)], \quad (27)$$

with the initial condition

$$[\mathbf{Z}(0)] = [\mathbf{H}_d][a], \quad [\mathbf{Y}(0)] = [\mathcal{N}][a] \quad a.s., \quad (28)$$

in which the random matrices $[\mathcal{L}([\mathbf{Z}(r)])]$ and $[d\mathbf{W}(r)]$ with values in $\mathbb{M}_{\nu, m}$ are such that

$$[\mathcal{L}([\mathbf{Z}(r)])] = [L([\mathbf{Z}(r)])[g^T]][a], \quad [d\mathbf{W}(r)] = [d\mathbb{W}(r)][a]. \quad (29)$$

(i) For all $[u] = [\mathbf{u}^1 \dots \mathbf{u}^N]$ in $\mathbb{M}_{\nu, N}$ with $\mathbf{u}^\ell = (u_1^\ell, \dots, u_\nu^\ell)$ in \mathbb{R}^ν , the matrix $[L([u])]$ in $\mathbb{M}_{\nu, N}$ is defined, for all $k = 1, \dots, \nu$ and for all $\ell = 1, \dots, N$, by

$$[L([u])]_{k\ell} = \frac{1}{p(\mathbf{u}^\ell)} \{ \nabla_{\mathbf{u}^\ell} p(\mathbf{u}^\ell) \}_k, \quad (30)$$

$$p(\mathbf{u}^\ell) = \frac{1}{N} \sum_{j=1}^N \exp\left\{-\frac{1}{2\hat{s}_\nu^2} \left\| \frac{\hat{s}_\nu}{s_\nu} \boldsymbol{\eta}^j - \mathbf{u}^\ell \right\|^2\right\}, \quad (31)$$

$$\nabla_{\mathbf{u}^\ell} p(\mathbf{u}^\ell) = \frac{1}{\hat{s}_\nu^2} \frac{1}{N} \sum_{j=1}^N \left(\frac{\hat{s}_\nu}{s_\nu} \boldsymbol{\eta}^j - \mathbf{u}^\ell \right) \exp\left\{-\frac{1}{2\hat{s}_\nu^2} \left\| \frac{\hat{s}_\nu}{s_\nu} \boldsymbol{\eta}^j - \mathbf{u}^\ell \right\|^2\right\}, \quad (32)$$

$$s_\nu = \left\{ \frac{4}{N(2 + \nu)} \right\}^{1/(\nu+4)}, \quad \hat{s}_\nu = \frac{s_\nu}{\sqrt{s_\nu^2 + \frac{N-1}{N}}}. \quad (33)$$

(ii) The stochastic process $\{[d\mathbb{W}(r)], r \geq 0\}$ with values in $\mathbb{M}_{\nu, N}$ is such that $[d\mathbb{W}(r)] = [d\mathbb{W}^1(r) \dots d\mathbb{W}^N(r)]$ in which the columns $\mathbb{W}^1, \dots, \mathbb{W}^N$ are N independent copies of the normalized Wiener process \mathbb{W} defined on $(\Theta, \mathcal{T}, \mathcal{P})$, indexed by \mathbb{R}^+ with values in \mathbb{R}^ν . The matrix-valued autocorrelation function $[R_{\mathbb{W}}(r, r')] = E\{\mathbb{W}(r)\mathbb{W}(r')^T\}$ of \mathbb{W} is then written as $[R_{\mathbb{W}}(r, r')] = \min(r, r') [I_\nu]$.

(iii) The probability distribution of the random matrix $[\mathbf{H}_d]$ with values in $\mathbb{M}_{\nu, N}$ is identical to the probability distribution of random matrix $[\mathbf{H}]$. A known sample of $[\mathbf{H}_d]$ is matrix $[\eta_d]$ defined by equation (21). The random matrix $[\mathcal{N}]$ with values in $\mathbb{M}_{\nu, N}$ is written as $[\mathcal{N}] = [\mathcal{N}^1 \dots \mathcal{N}^N]$ in which the columns $\mathcal{N}^1, \dots, \mathcal{N}^N$ are N independent copies of the normalized Gaussian vector \mathcal{N} with values in \mathbb{R}^ν (this means that $E\{\mathcal{N}\} = \mathbf{0}$ and $E\{\mathcal{N}\mathcal{N}^T\} = [I_\nu]$). The random matrices $[\mathbf{H}_d]$

and $[\mathcal{N}]$, and the normalized Wiener process $\{\mathbb{W}(r), r \geq 0\}$ are assumed to be independent.

(iv) The free parameter $f_0 > 0$ allows the dissipation term of the nonlinear second-order dynamical system (dissipative Hamiltonian system) to be controlled.

(v) The algorithm for solving equations (26) to (28) is detailed in [29] and is summarized in Appendix C.

The generation of additional samples $[x_{\text{ar}}^1], \dots, [x_{\text{ar}}^{n_{\text{MC}}}]$ of random matrix $[\mathbf{X}]$ is then performed as follows. For θ fixed in Θ , the deterministic quantities $\{[\mathbf{W}(r; \theta)], r \geq 0\}$, $[\mathcal{Z}(0; \theta)] = [\eta_d] [a]$, and $[\mathcal{Y}(0; \theta)] = [\mathcal{N}(\theta)] [a]$ are independent samples of the stochastic process $\{[\mathbf{W}(r)], r \geq 0\}$, of the random matrix $[\mathcal{Z}(0)]$, and of the random matrix $[\mathcal{Y}(0)]$. Let $\{([\mathcal{Z}(r; \theta)], [\mathcal{Y}(r; \theta)]), r \in \mathbb{R}^+\}$ be the corresponding sample of the unique stationary diffusion process $\{([\mathcal{Z}(r)], [\mathcal{Y}(r)]), r \in \mathbb{R}^+\}$ of the problem defined by equations (26) to (28)). The reduced-order ISDE is solved on the finite interval $\mathcal{R} = [0, M \Delta r]$, in which Δr is the sampling step of the continuous index parameter r used in the integration scheme (see Appendix C). The integer M is written as $M = M_0 \times n_{\text{MC}}$, in which M_0 and n_{MC} are positive integers greater or equal to 1. Introducing $\rho = M_0 \Delta r$, the n_{MC} additional samples $[z_{\text{ar}}^1], \dots, [z_{\text{ar}}^{n_{\text{MC}}}]$ of random matrix $[\mathbf{Z}]$ and the corresponding samples $[\eta_{\text{ar}}^1], \dots, [\eta_{\text{ar}}^{n_{\text{MC}}}]$ of random matrix $[\mathbf{H}]$ are given by

$$[z_{\text{ar}}^\ell] = [\mathcal{Z}(\ell \times \rho; \theta)] \quad , \quad [\eta_{\text{ar}}^\ell] = [z_{\text{ar}}^\ell] [g]^T \quad , \quad \ell = 1, \dots, n_{\text{MC}} . \quad (34)$$

- If $M_0 = 1$, then $\rho = \Delta r$ and the n_{MC} additional samples are dependent, but the ergodic property of $\{[\mathcal{Z}(r)], r \in \mathbb{R}^+\}$ can be invoked for ensuring the convergence of statistics constructed using $[z_{\text{ar}}^1], \dots, [z_{\text{ar}}^{n_{\text{MC}}}]$ for random matrix $[\mathbf{Z}]$.
- If integer M_0 is chosen sufficiently large (such that ρ is much larger than the relaxation time of the dissipative Hamiltonian dynamical system), then $[z_{\text{ar}}^1], \dots, [z_{\text{ar}}^{n_{\text{MC}}}]$ can approximately be considered as independent samples of random matrix $[\mathbf{Z}]$.

Using equation (25), the n_{MC} additional samples $[x_{\text{ar}}^1], \dots, [x_{\text{ar}}^{n_{\text{MC}}}]$ of random matrix $[\mathbf{X}]$ can be generated according to the following expression,

$$[x_{\text{ar}}^\ell] = [\underline{x}] + [\varphi] [\lambda]^{1/2} [z_{\text{ar}}^\ell] [g]^T \quad , \quad \ell = 1, \dots, n_{\text{MC}} . \quad (35)$$

5. PROPOSED ALGORITHM FOR SOLVING THE PROBABILISTIC NONCONVEX CONSTRAINED OPTIMIZATION PROBLEM WITH A FIXED NUMBER OF FUNCTION EVALUATIONS

The proposed algorithm used for solving the probabilistic nonconvex constrained optimization problem defined by equations (1) to (3) with a fixed number N of function evaluations is based on the following procedure. Instead of computing a large number of samples \mathbf{x}^ℓ , each one requiring to call the stochastic computational model a large number of times, the number N of computed samples \mathbf{x}^ℓ by calling the stochastic computational model is restricted to a relatively small value N . In this condition, N is not sufficiently large for obtaining a good estimate of the objective function and of the constraints function using the algorithm detailed in Section 3.3. The method proposed herein consists in enriching the N samples that are represented by the columns of matrix $[x_d] = [\mathbf{x}^1 \dots \mathbf{x}^N]$ in $\mathbb{M}_{n, N}$, with n_{MC} additional samples $[x_{\text{ar}}^1], \dots, [x_{\text{ar}}^{n_{\text{MC}}}]$ that are computed without calling the stochastic computational model but that are computed by using the generator of samples presented in Section 4. Using these $\nu_{\text{sim}} = n_{\text{MC}} \times N$ additional samples, a good estimate of $E\{R|\mathbf{W} = \mathbf{w}^0\}$ can be obtained from equations (13) and (19). The algorithm is then as follows:

Data allowing the algorithm to be initialized.

1. Given m_w, m_c, N .
2. Deducing $n = m_w + 1 + m_c$.

3. Given $[x_d] = [\mathbf{x}^1 \dots \mathbf{x}^N] \in \mathbb{M}_{n,N}$ such that for $\ell = 1, \dots, N$:
 $\mathbf{x}^\ell = (\mathbf{w}^\ell, q^\ell, \mathbf{b}^\ell) \in \mathbb{R}^n$ with $\mathbf{w}^\ell \in \mathbb{R}^{m_w}, q^\ell \in \mathbb{R}, \mathbf{b}^\ell \in \mathbb{R}^{m_c}$.
4. If necessary, scaling the data set represented by matrix $[x_d]$ in $\mathbb{M}_{n,N}$.

Steps of the algorithm.

1. Computing the mean $\underline{\mathbf{x}} \in \mathbb{R}^n$ and the covariance matrix $[\text{cov}] \in \mathbb{M}_n$ with equation (55).
2. Constructing matrix $[x] = [\underline{\mathbf{x}} \dots \underline{\mathbf{x}}] \in \mathbb{M}_{n,N}$ (the N columns are the same).
3. Solving the eigenvalue problem for the covariance matrix $[\text{cov}]$: computing $\nu \leq n$, matrices $[\lambda]$ and $[\varphi]$ (see equation (20)).
4. For ε fixed, computing the matrix $[g]$ in $\mathbb{M}_{N,m}$ constituted of the diffusion-maps basis by using the algorithm detailed in Appendix A, for which the value of $m < N$ is selected as explained in Appendix B.
5. Calculating the matrix $[a]$ in $\mathbb{M}_{N,m}$ defined by equation (23).
6. For given $f_0, \Delta r, M_0$, and n_{MC} , computing n_{MC} additional samples $[x_{\text{ar}}^\ell] \in \mathbb{M}_{n,N}$ for $\ell = 1, \dots, n_{\text{MC}}$ of random matrix $[\mathbf{X}]$ by using the algorithm detailed in Section 4.3.
7. Extracting from the $\nu_{\text{sim}} = n_{\text{MC}} \times N$ columns of the n_{MC} matrices $[x_{\text{ar}}^1], \dots, [x_{\text{ar}}^{n_{\text{MC}}}]$:
 - (a) ν_{sim} samples $(\mathbf{w}_{\text{ar}}^\ell, q_{\text{ar}}^\ell) \in \mathbb{R}^{m_w} \times \mathbb{R}$ for $\ell = 1, \dots, \nu_{\text{sim}}$.
 - (b) ν_{sim} samples $(\mathbf{w}_{\text{ar}}^\ell, \mathbf{b}_{\text{ar}}^\ell) \in \mathbb{R}^{m_w} \times \mathbb{R}^{m_c}$ for $\ell = 1, \dots, \nu_{\text{sim}}$.
8. Computing
 - (a) $\underline{q}_{\text{ar}} = \frac{1}{\nu_{\text{sim}}} \sum_{\ell=1}^{\nu_{\text{sim}}} q_{\text{ar}}^\ell$ and $\sigma_{q_{\text{ar}}}^2 = \frac{1}{\nu_{\text{sim}}} \sum_{\ell=1}^{\nu_{\text{sim}}} (q_{\text{ar}}^\ell - \underline{q}_{\text{ar}})^2$.
 - (b) $\underline{b}_{\text{ar},k} = \frac{1}{\nu_{\text{sim}}} \sum_{\ell=1}^{\nu_{\text{sim}}} b_{\text{ar},k}^\ell$ and $\sigma_{b_{\text{ar},k}}^2 = \frac{1}{\nu_{\text{sim}}} \sum_{\ell=1}^{\nu_{\text{sim}}} (b_{\text{ar},k}^\ell - \underline{b}_{\text{ar},k})^2, k = 1, \dots, m_c$.
9. Applying equation (11) for normalizing the samples, which yields:
 - (a) ν_{sim} samples $(\tilde{\mathbf{w}}_{\text{ar}}^\ell, \tilde{q}_{\text{ar}}^\ell) \in \mathbb{R}^{m_w} \times \mathbb{R}$ for $\ell = 1, \dots, \nu_{\text{sim}}$.
 - (b) ν_{sim} samples $(\tilde{\mathbf{w}}_{\text{ar}}^\ell, \tilde{b}_{\text{ar},k}^\ell) \in \mathbb{R}^{m_w} \times \mathbb{R}$ for $\ell = 1, \dots, \nu_{\text{sim}}$ and for $k = 1, \dots, m_c$.
10. Computing s by using equation (17).
11. For each point \mathbf{w}^0 in \mathcal{C}_w proposed by the optimization algorithm, computing $f(\mathbf{w}^0)$ and $\mathbf{c}(\mathbf{w}^0)$ by replacing the function evaluation with the stochastic computational model by the following procedure that does not use any function evaluation with the stochastic computational model:
 - (a) Computing the normalization $\tilde{\mathbf{w}}^0$ of \mathbf{w}^0 by using equation (14).
 - (b) Computing $e^\ell(\tilde{\mathbf{w}}^0) = \exp\{-\frac{1}{2s^2} \|\tilde{\mathbf{w}}_{\text{ar}}^\ell - \tilde{\mathbf{w}}^0\|^2\}$ for $\ell = 1, \dots, \nu_{\text{sim}}$.
 - (c) Computing $\gamma(\tilde{\mathbf{w}}^0) = \sum_{\ell=1}^{\nu_{\text{sim}}} e^\ell(\tilde{\mathbf{w}}^0)$.
 - (d) Computing the estimate of $f(\mathbf{w}^0)$ by using equations (8), (13), and (19):

$$f(\mathbf{w}^0) \simeq \underline{q}_{\text{ar}} + \frac{\sigma_{q_{\text{ar}}}}{\gamma(\tilde{\mathbf{w}}^0)} \sum_{\ell=1}^{\nu_{\text{sim}}} \tilde{q}_{\text{ar}}^\ell e^\ell(\tilde{\mathbf{w}}^0).$$

- (e) For $k = 1, \dots, m_c$, computing the estimate of $c_k(\mathbf{w}^0)$ by using equations (9), (13), and (19):

$$c_k(\mathbf{w}^0) \simeq \underline{b}_{\text{ar},k} + \frac{\sigma_{b_{\text{ar},k}}}{\gamma(\tilde{\mathbf{w}}^0)} \sum_{\ell=1}^{\nu_{\text{sim}}} \tilde{b}_{\text{ar},k}^\ell e^\ell(\tilde{\mathbf{w}}^0).$$

End of the algorithm.

6. APPLICATIONS

6.1. Application 1

Definition of the probabilistic nonconvex constrained optimization problem. The admissible set $\mathcal{C}_{\mathbf{w}} \subset \mathbb{R}^{m_w}$ with $m_w = 2$ and the stochastic process $\{\mathcal{Q}(\mathbf{w}), \mathbf{w} \in \mathcal{C}_{\mathbf{w}}\}$ with values in \mathbb{R} of the optimization problem (see equations (1) to (3)) are defined by

$$\mathcal{C}_{\mathbf{w}} = \{\mathbf{w} = (w_1, w_2) \in [0, 1.15]^2 \subset \mathbb{R}^2\},$$

and

$$\mathcal{Q}(\mathbf{w}) = 0.11 \beta_1 (\beta_2 + 0.05 w_1) (\beta_2 + 0.05 w_2) \left(\frac{\sin(\mathfrak{r})}{\mathfrak{r}^{0.10}} + 0.08 \beta_3 \right), \quad (36)$$

in which

$$w_1 = 20 w_1 - 10 \quad , \quad w_2 = 20 w_2 - 10 \quad , \quad \mathfrak{r} = \sqrt{w_1^2 + 0.9 w_2^2} + 0.01.$$

The components of the \mathbb{R}^{m_c} -valued stochastic process $\{\mathcal{B}(\mathbf{w}), \mathbf{w} \in \mathcal{C}_{\mathbf{w}}\}$ with $m_c = 4$ are defined as,

$$\mathcal{B}_1(\mathbf{w}) = w_1 + w_2 - 1.24 + \beta_1, \quad (37)$$

$$\mathcal{B}_2(\mathbf{w}) = w_2 - w_1 + \beta_2, \quad (38)$$

$$\mathcal{B}_3(\mathbf{w}) = w_1 - w_2 - 0.60 + \beta_3, \quad (39)$$

$$\mathcal{B}_4(\mathbf{w}) = -w_1 - w_2 + 0.4 + \beta_4. \quad (40)$$

In the previous equations, the \mathbb{R}^4 -valued random variable $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)$ on $(\Theta, \mathcal{T}, \mathcal{P})$ is defined by

$$\beta_1 = 1 + 2 \delta_1 \chi_2 (\chi_1 - 0.5),$$

$$\beta_2 = 1 + \delta_2 \chi_1,$$

$$\beta_3 = 1 + \delta_3 \chi_2,$$

$$\beta_4 = \delta_4 (\chi_1 - 0.5),$$

in which χ_1 and χ_2 are two independent uniform random variables on $[0, 1]$ defined on $(\Theta, \mathcal{T}, \mathcal{P})$, and where $\delta_1 \geq 0$, $\delta_2 \geq 0$, $\delta_3 \geq 0$, and $\delta_4 \geq 0$ are the hyperparameters that allow the level of uncertainties to be controlled, and that have been fixed for the numerical application to the following values $\delta_1 = \delta_2 = \delta_3 = 0.2$ and $\delta_4 = 0.05$.

Construction of the optimal solution of reference, $\mathbf{w}_r^{\text{opt}}$. The admissible set $\mathcal{C}_{\mathbf{w}}$ is meshed with a regular cartesian grid of 60×60 nodes that define $\nu_g = 3600$ points $\mathbf{w}_g^i = (w_{g,1}^i, w_{g,2}^i)$ with $i = 1, \dots, \nu_g$ in $\mathcal{C}_{\mathbf{w}}$. For every point $\mathbf{w} = \mathbf{w}_g^i$ of the grid, $f(\mathbf{w})$ and $\mathbf{c}(\mathbf{w})$ are estimated by using equation (4) with $N'_s = 10,000$ and equations (36) to (40). The graphs of the objective function and of the constraints function are displayed in figure 1. The optimal solution of reference, $\mathbf{w}_r^{\text{opt}} = (w_{r,1}^{\text{opt}}, w_{r,2}^{\text{opt}})$, of the probabilistic constrained optimization problem is estimated by

$$\mathbf{w}_r^{\text{opt}} \simeq \arg \min_{\substack{i=1, \dots, \nu_g \\ \mathbf{c}(\mathbf{w}_g^i) < 0}} f(\mathbf{w}_g^i)$$

and yields $w_{r,1}^{\text{opt}} = 0.74$, $w_{r,2}^{\text{opt}} = 0.49$, and $f(\mathbf{w}_r^{\text{opt}}) = -0.123$. It should be noted that, if the constraints were removed, the solution, noted $\mathbf{w}_{\text{rsc}}^{\text{opt}}$, would be $w_{\text{rsc},1}^{\text{opt}} = 1.11$, $w_{\text{rsc},2}^{\text{opt}} = 1.15$, and $f(\mathbf{w}_{\text{rsc}}^{\text{opt}}) = -0.218$, indicating that the constraints are indeed active.

Definition of the fixed number N of data points. For this numerical application, $n = m_w + 1 + m_c = 7 = 2 + 1 + 4$, and 3 values of N are considered in order to evaluate the efficiency of the proposed algorithm: $N = 100$, $N = 400$, and $N = 900$. For a fixed value of N , we define the data set $\{\mathbf{x}^\ell = (\mathbf{w}^\ell, q^\ell, \mathbf{b}^\ell) \in \mathbb{R}^n, \ell = 1, \dots, N\}$ introduced in Section 2.2 as follows. For $\ell = 1, \dots, N$, the

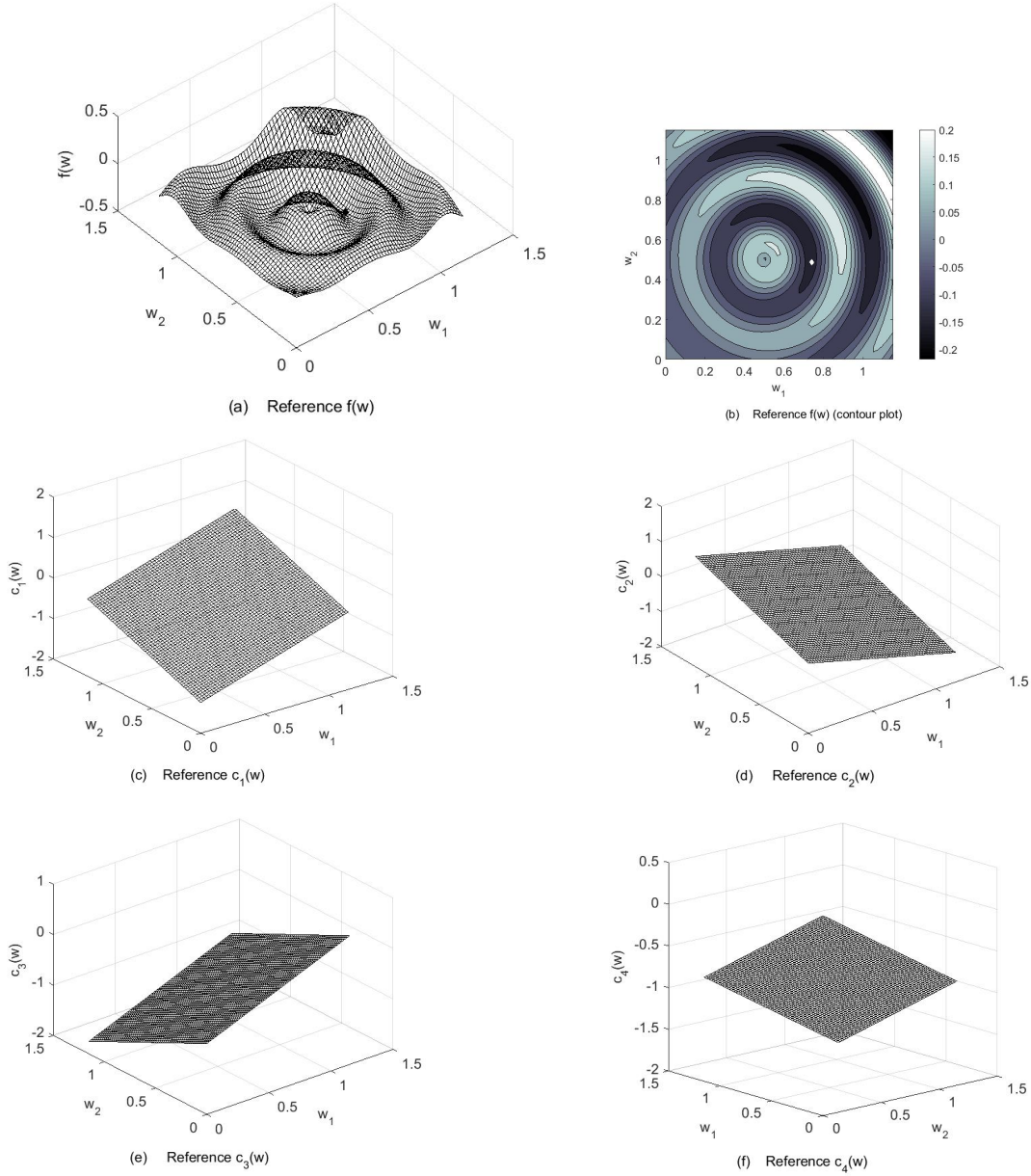


Figure 1. Solution of reference: Graph of the objective function $\mathbf{w} \mapsto f(\mathbf{w})$ in 3D (a) and in contour plot (b) evaluated in the ν_g points $\mathbf{w}_g^i = (w_{g,1}^i, w_{g,2}^i)$, $i = 1, \dots, \nu_g$ of the grid. The optimal solution of reference is the black diamond in (a) and the white diamond in (b). Graphs of the constraint functions $\mathbf{w} \mapsto c_1(\mathbf{w})$ (c), $c_2(\mathbf{w})$ (d), $c_3(\mathbf{w})$ (e), and $c_4(\mathbf{w})$ (f) evaluated in the ν_g points $\mathbf{w}_g^i = (w_{g,1}^i, w_{g,2}^i)$, $i = 1, \dots, \nu_g$ of the grid.

points $\mathbf{w}^\ell = (w_1^\ell, w_2^\ell)$ are generated by $w_1^\ell = 1.15 \chi_3^\ell$ and $w_2^\ell = 1.15 \chi_4^\ell$ in which $\{\chi_3^\ell\}_\ell$ and $\{\chi_4^\ell\}_\ell$ are N independent samples of the independent uniform random variables χ_3 and χ_4 on $[0, 1]$ that are independent of χ_1 and χ_2 . By using the stochastic model $\mathcal{Q}(\mathbf{w})$ and $\mathcal{B}(\mathbf{w})$ defined by equations (36) to (40), we thus deduce $q^\ell \in \mathbb{R}$ and $\mathbf{b}^\ell \in \mathbb{R}^{m_c}$ for $\ell = 1, \dots, N$ by using equation (5). It should be noted that the surface defined by the graph of the objective function $\mathbf{w} \mapsto f(\mathbf{w})$ of the reference (see figure 1) presents many oscillations and consequently, it cannot be reconstructed using a small number N of data points. To explore this point, results were obtained for N varying from 100 to 900.

Construction of the optimal solution, $\mathbf{w}_d^{\text{opt}}$, estimated with the N data points. The methodology presented in Section 3 is used for estimating the optimal solution $\mathbf{w}_d^{\text{opt}} = (w_{d,1}^{\text{opt}}, w_{d,2}^{\text{opt}})$ of the probabilistic constrained optimization problem, using only the N data points, and which is given by

$$\mathbf{w}_d^{\text{opt}} \simeq \arg \min_{\substack{i=1, \dots, \nu_g \\ \mathbf{c}(\mathbf{w}_g^i) < 0}} f(\mathbf{w}_g^i),$$

in which the objective function f and the constraints function \mathbf{c} are computed by using the algorithm presented in Section 3.3, and which are estimated by using only the N data points. In Table I, for the three values of N , the three lines (c) entitled "Data points" give the results for $\mathbf{w}_d^{\text{opt}}$. Figures 2-(a), 3-(a), and 4-(a) display the graph of the objective function $\mathbf{w} \mapsto f(\mathbf{w})$ in contour plot, evaluated in the ν_g points $\mathbf{w}_g^i = (w_{g,1}^i, w_{g,2}^i)$, $i = 1, \dots, \nu_g$ of the grid, and estimated with $N = 100$, $N = 400$, and $N = 500$. These results show that the optimal solution of reference, $\mathbf{w}_r^{\text{opt}}$, cannot be approximated by using only the N data points represented by $[\mathbf{x}_d]$, even for $N = 900$.

Construction of the optimal solution, $\mathbf{w}_{\text{ar}}^{\text{opt}}$, estimated with the fixed number N of evaluations and with additional samples without additional function evaluations. For each one of the values of N , the algorithm presented in Section 5 is used. For the three values of N : $\nu = 4$ (the null space of matrix $[\text{cov}]$ is found to be equal to 3), the smoothing parameter $\varepsilon = 5$, and m is estimated by using Appendix B yielding $m = 5$, independent of N . The relative error $e_{\text{red}}(m)$ computed with equation (54) yields 1.44×10^{-3} for $N = 100$, 8.25×10^{-4} for $N = 400$, and 9.89×10^{-4} for $N = 900$. The parameters for the generation of the additional samples performed by using Section 4.3 are $f_0 = 1.5$, $\Delta r = 0.11828$, and $M_0 = 100$. For each one of the considered values of N , the number of additional samples that are computed is $\nu_{\text{sim}} = n_{\text{MC}}N$ with $n_{\text{MC}} = 2$, $n_{\text{MC}} = 10$, and $n_{\text{MC}} = 100$. Steps 7 to 11 of the algorithm presented in Section 5 are carried out for evaluating the estimated objective function f and the constraints function \mathbf{c} at the ν_g points $\mathbf{w}^0 = \mathbf{w}_g^i = (w_{g,1}^i, w_{g,2}^i)$, $i = 1, \dots, \nu_g$ of the grid. The optimal solution $\mathbf{w}_{\text{ar}}^{\text{opt}} = (w_{\text{ar},1}^{\text{opt}}, w_{\text{ar},2}^{\text{opt}})$ of the probabilistic constrained optimization problem, using the ν_{sim} additional samples is given by

$$\mathbf{w}_{\text{ar}}^{\text{opt}} \simeq \arg \min_{\substack{i=1, \dots, \nu_g \\ \mathbf{c}(\mathbf{w}_g^i) < 0}} f(\mathbf{w}_g^i).$$

In Table I, for the three values of N , the three blocks of lines (c) entitled "Additional samples" give the results for $\mathbf{w}_{\text{ar}}^{\text{opt}}$ as a function of the number ν_{sim} of additional samples. For $N = 100$, $N = 400$, and $N = 500$, Figures 2, 3, and 4 display the graph of the objective function $\mathbf{w} \mapsto f(\mathbf{w})$ in contour plot, evaluated at the ν_g points $\mathbf{w}_g^i = (w_{g,1}^i, w_{g,2}^i)$, $i = 1, \dots, \nu_g$ of the grid, estimated with ν_{sim} additional samples. These figures clearly show that $N = 100$ is not sufficient for reconstructing the surface for any number of additional samples and consequently, $\mathbf{w}_r^{\text{opt}}$ cannot be approximated by $\mathbf{w}_{\text{ar}}^{\text{opt}}$. On the other hand, for $N = 400$ and *a fortiori* for $N = 900$, as soon as ν_{sim} is sufficiently large, then the surface is reconstructed with a reasonable accuracy and consequently, $\mathbf{w}_{\text{ar}}^{\text{opt}}$ yields a good approximation of $\mathbf{w}_r^{\text{opt}}$.

For illustration, for $N = 900$ data points, figure 5-(a) displays the eigenvalues in \log_{10} -scale of the transition matrix for random vector \mathbf{H} , which clearly shows that 5 is a good value for parameter m , and figure 5-(b) shows the graph of the objective function $\mathbf{w} \mapsto f(\mathbf{w})$ in 3D evaluated in the ν_g points $\mathbf{w}_g^i = (w_{g,1}^i, w_{g,2}^i)$, $i = 1, \dots, \nu_g$ of the grid, and estimated with $\nu_{\text{sim}} = 9000$ additional samples corresponding to $n_{\text{MC}} = 10$. It can be seen that this surface, which is reconstructed from the 900 data points by using 9000 additional samples, is close to the surface of reference shown in figure 1-(a).

Quantitative summary of the results obtained for the three values of N . In Table I, the first line (a) of results entitled "Reference no constraint" is the optimal solution of reference $\mathbf{w}^{\text{opt}} = \mathbf{w}_{\text{rsc}}^{\text{opt}}$ computed without the constraints while the second line (b), entitled "Reference" is the optimal solution of reference $\mathbf{w}^{\text{opt}} = \mathbf{w}_r^{\text{opt}}$ computed with the constraints. In Table I, line (c) corresponds to the optimal solution with constraints, $\mathbf{w}^{\text{opt}} = \mathbf{w}_d^{\text{opt}}$, estimated by using only the N data points and lines (d) correspond to the optimal solution with constraints, $\mathbf{w}^{\text{opt}} = \mathbf{w}_a^{\text{opt}}$, estimated by using ν_{sim}

additional samples. The results given in lines (c) and (d) must be compared to line (b) "Reference". These results provide a closer quantification of the illustrations in figures 2 to 4.

Table I. Table defining the optimal values of the optimization problem (a) corresponding to the reference solution without the constraints, (b) corresponding to the reference solution with the constraints, and for the optimization problem with the constraints, (c) estimated with the N data points, (d) estimated with ν_{sim} additional samples.

	N	n_{MC}	ν_{sim}	w_1^{opt}	w_2^{opt}	$f(\mathbf{w}^{\text{opt}})$
(a) Reference no constraint				1.11	1.15	-0.218
(b) Reference				0.74	0.49	-0.123
(c) Data points	100			0.74	0.00	-0.025
(d) Additional samples	100	2	200	0.64	0.00	-0.027
	100	10	1000	0.68	0.55	-0.049
	100	100	10 000	0.68	0.37	-0.063
(c) Data points	400			0.17	0.17	0.008
(d) Additional samples	400	2	800	0.21	0.00	-0.085
	400	10	4000	0.70	0.53	-0.067
	400	100	40 000	0.72	0.45	-0.107
(c) Data points	900			0.70	0.00	0.001
(d) Additional samples	900	2	1800	0.68	0.55	-0.013
	900	10	9000	0.70	0.49	-0.081
	900	100	90 000	0.72	0.49	-0.112

6.2. Application 2

The second application is devoted to the probabilistic nonconvex unconstrained optimization problem for which the objective function is the Rosenbrock function [32] in dimension 4 for which both a multiplicative noise and an additive noise have been added.

Definition of the probabilistic nonconvex unconstrained optimization problem. The admissible set $\mathcal{C}_{\mathbf{w}} \subset \mathbb{R}^{m_w}$ with $m_w = 4$ and the stochastic process $\{\mathcal{Q}(\mathbf{w}), \mathbf{w} \in \mathcal{C}_{\mathbf{w}}\}$ with values in \mathbb{R} of the optimization problem (see equations (1) to (2)) are defined by

$$\mathcal{C}_{\mathbf{w}} = \{(\mathbf{w} = (w_1, w_2, w_3, w_4) \in [0.85, 1.15]^4 \subset \mathbb{R}^4)\},$$

$$\mathcal{Q}(\mathbf{w}) = \beta_1 \mathcal{R}(\mathbf{w}) + \beta_2, \quad (41)$$

in which $\mathbf{w} \mapsto \mathcal{R}(\mathbf{w})$ is the Rosenbrock function that, for $m_w = 4$, is written as

$$\mathcal{R}(\mathbf{w}) = \sum_{i=1}^2 100 (w_{2i} - w_{2i-1}^2)^2 + (1 - w_{2i-1})^2, \quad (42)$$

and where the \mathbb{R}^2 -valued random variable $\beta = (\beta_1, \beta_2)$ on $(\Theta, \mathcal{T}, \mathcal{P})$ is defined by

$$\beta_1 = 1 + 2 \delta_1 \chi_2 (\chi_1 - 0.5), \quad (43)$$

$$\beta_2 = \delta_2 (\chi_1 - 0.5), \quad (44)$$

in which χ_1 and χ_2 are two independent uniform random variables on $[0, 1]$ defined on $(\Theta, \mathcal{T}, \mathcal{P})$, and where $\delta_1 \geq 0$ and $\delta_2 \geq 0$ are the hyperparameters that allow the level of uncertainties to be controlled, and that have been fixed for the numerical application as $\delta_1 = \delta_2 = 0.4$.

Construction of the optimal solution of reference, $\mathbf{w}_r^{\text{opt}}$. (i) As $E\{\beta_1\} = 1$ and $E\{\beta_2\} = 0$, the exact expression of the objective function defined by equation (2) is $f(\mathbf{w}) = \mathcal{R}(\mathbf{w})$. Consequently, the exact solution of the probabilistic optimization problem defined by equations (1) and (2) is $w_1^{\text{opt}} = w_2^{\text{opt}} = w_3^{\text{opt}} = w_4^{\text{opt}} = 1$ and $f(\mathbf{w}^{\text{opt}}) = 0$.

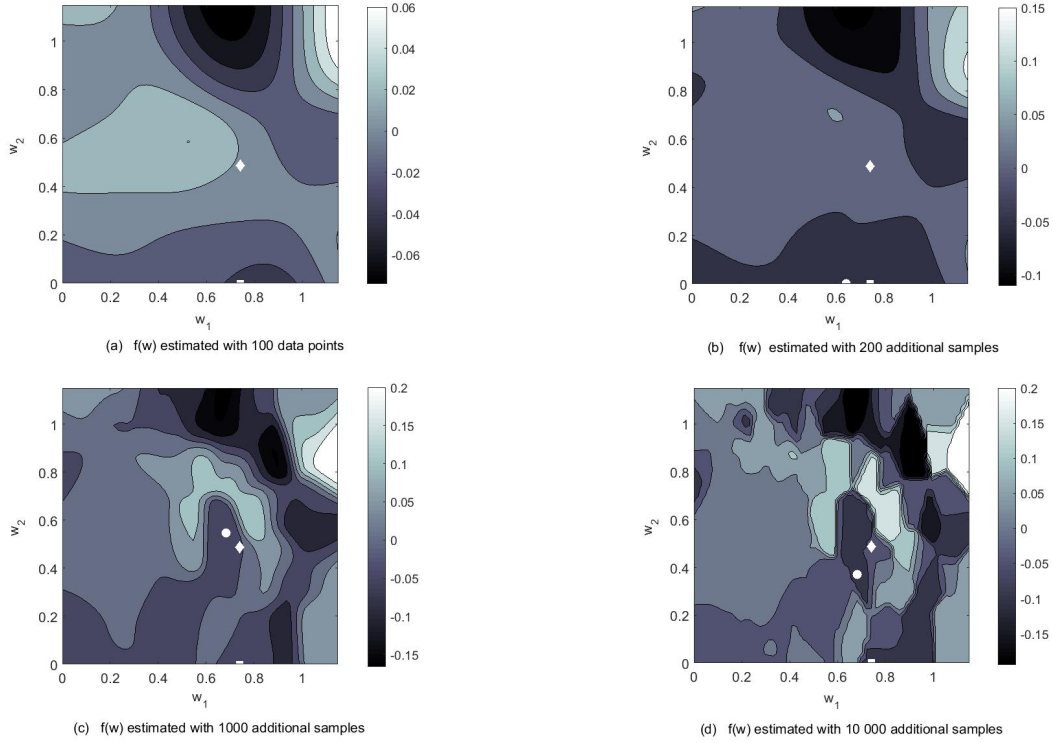


Figure 2. Graph of the objective function $\mathbf{w} \mapsto f(\mathbf{w})$ in contour plot, evaluated in the ν_g points $\mathbf{w}_g^i = (w_{g,1}^i, w_{g,2}^i)$, $i = 1, \dots, \nu_g$ of the grid, and estimated with $N = 100$ data points (a), with $\nu_{\text{sim}} = 200$ additional samples corresponding to $n_{\text{MC}} = 2$ (b), with $\nu_{\text{sim}} = 1000$ additional samples corresponding to $n_{\text{MC}} = 10$ (c), and with $\nu_{\text{sim}} = 10000$ additional samples corresponding to $n_{\text{MC}} = 100$ (d). The white diamond is the optimal solution of reference, the white disc is the optimal solution estimated with the additional samples, and the white square is the optimal solution estimated with the data points.

(ii) The optimal solution of reference is constructed numerically as follows. The admissible set $\mathcal{C}_{\mathbf{w}}$ is meshed with a regular cartesian grid of μ^4 nodes with $\mu \geq 2$, which define ν_g points $\mathbf{w}_g^i = (w_{g,1}^i, w_{g,2}^i, w_{g,3}^i, w_{g,4}^i)$ with $i = 1, \dots, \nu_g$ in $\mathcal{C}_{\mathbf{w}}$. For every point \mathbf{w}_g^i of the grid, $f(\mathbf{w}_g^i)$ is estimated by using equation (4) with $N_s^i = 100,000$ and equations (41) to (44). The optimal solution of reference, $\mathbf{w}_r^{\text{opt}} = (w_{r,1}^{\text{opt}}, w_{r,2}^{\text{opt}}, w_{r,3}^{\text{opt}}, w_{r,4}^{\text{opt}})$, of the probabilistic unconstrained optimization problem is estimated by

$$\mathbf{w}_r^{\text{opt}} \simeq \arg \min_{i=1, \dots, \nu_g} f(\mathbf{w}_g^i), \quad (45)$$

and the results are given in Table II as a function of the resolution μ of the grid. The extreme sensitivity of the reference solution to μ is clear. For the grids 13^4 , 15^4 , and 17^4 , the exact solution is reached, but it is by chance associated with the fact that the regular grid has a node that coincides with the exact solution and in addition, 100 000 samples are used to evaluate the mathematical expectations for each point $\mathbf{w} = \mathbf{w}_g^i$ of the grid. This aspect must be carefully kept in mind because, for the method proposed in this paper, we will use a dataset of $N = 256$ data points $\{\mathbf{w}^\ell, \ell = 1, \dots, N\}$ (very small number for 4 dimensions) that are randomly distributed in $\mathcal{C}_{\mathbf{w}}$ (and therefore, that are not nodes of a regular fineness grid) and for which, in each one of these data points, we have only one sample (and not 100 000 samples as for the reference solution) for estimating the objective function. Consequently, the statistical fluctuations in $\mathcal{Q}(\mathbf{w})$ play a fundamental role. The accuracy of the localization of the solution that is constructed with the proposed algorithm without additional evaluations of the objective function, must be interpreted statistically. A grid with 15^4 nodes is adopted in the following.

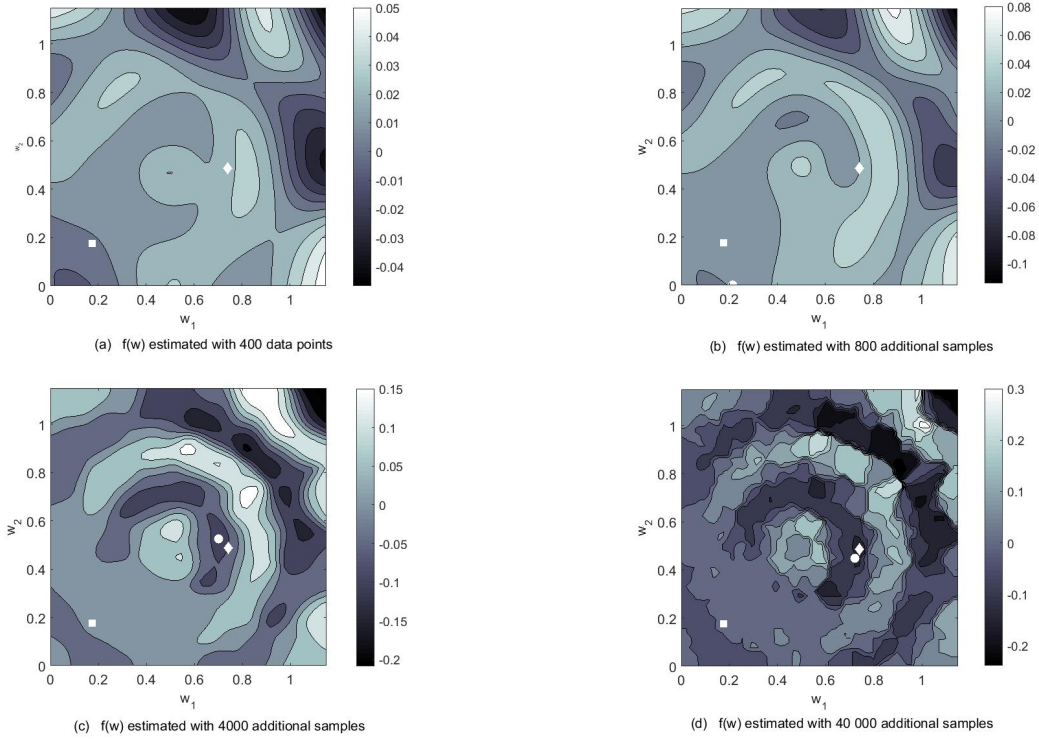


Figure 3. Graph of the objective function $\mathbf{w} \mapsto f(\mathbf{w})$ in contour plot, evaluated in the ν_g points $\mathbf{w}_g^i = (w_{g,1}^i, w_{g,2}^i)$, $i = 1, \dots, \nu_g$ of the grid, and estimated with $N = 400$ data points (a), with $\nu_{\text{sim}} = 800$ additional samples corresponding to $n_{\text{MC}} = 2$ (b), with $\nu_{\text{sim}} = 4000$ additional samples corresponding to $n_{\text{MC}} = 10$ (c), and with $\nu_{\text{sim}} = 40\,000$ additional samples corresponding to $n_{\text{MC}} = 100$ (d). The white diamond is the optimal solution of reference, the white disc is the optimal solution estimated with the additional samples, and the white square is the optimal solution estimated with the data points.

Table II. Table defining the optimal solution of reference $\mathbf{w}_r^{\text{opt}} = (w_{r,1}^{\text{opt}}, w_{r,2}^{\text{opt}}, w_{r,3}^{\text{opt}}, w_{r,4}^{\text{opt}})$ as a solution of equation (45) in function of the fineness of the grid.

Number of nodes in the grid	$w_{r,1}^{\text{opt}}$	$w_{r,2}^{\text{opt}}$	$w_{r,3}^{\text{opt}}$	$w_{r,4}^{\text{opt}}$	$f(\mathbf{w}_r^{\text{opt}})$
$13^4 = 28\,561$	1.0000	1.0000	1.0000	1.0000	-0.0001
$14^4 = 38\,416$	0.9192	0.8500	0.9192	0.8500	0.0180
$15^4 = 50\,625$	1.0000	1.0000	1.0000	1.0000	-0.0001
$16^4 = 65\,536$	0.9300	0.8700	0.9300	0.8700	0.0149
$17^4 = 83\,521$	1.0000	1.0000	1.0000	1.0000	-0.0001
$18^4 = 104\,976$	1.0618	1.1324	1.0618	1.1324	0.0126

Definition of the fixed number N of data points. For this numerical application, $n = m_w + 1 = 5 = 4 + 1$, and the value of N is 256. The data set $\mathbf{x}^\ell = (\mathbf{w}^\ell, q^\ell)$ for $\ell = 1, \dots, N$, which has been defined in Section 2.2 is constructed as follows. For $\ell = 1, \dots, N$, the points $\mathbf{w}^\ell = (w_1^\ell, w_2^\ell, w_3^\ell, w_4^\ell)$ are generated, for $k = 1, \dots, 4$ by $w_k^\ell = 0.85 + 0.3 \widehat{\chi}_k^\ell$ in which $\{\widehat{\chi}_k^\ell\}_{k,\ell}$ are $4 \times N$ independent samples of the independent uniform random variables $\{\widehat{\chi}_k, k = 1, \dots, 4\}$ on $[0, 1]$, which are independent of χ_1 and χ_2 . By using the stochastic model $\mathcal{Q}(\mathbf{w})$ defined by equations (41) to (44), we thus deduce q^ℓ for $\ell = 1, \dots, N$ by using equation (5).

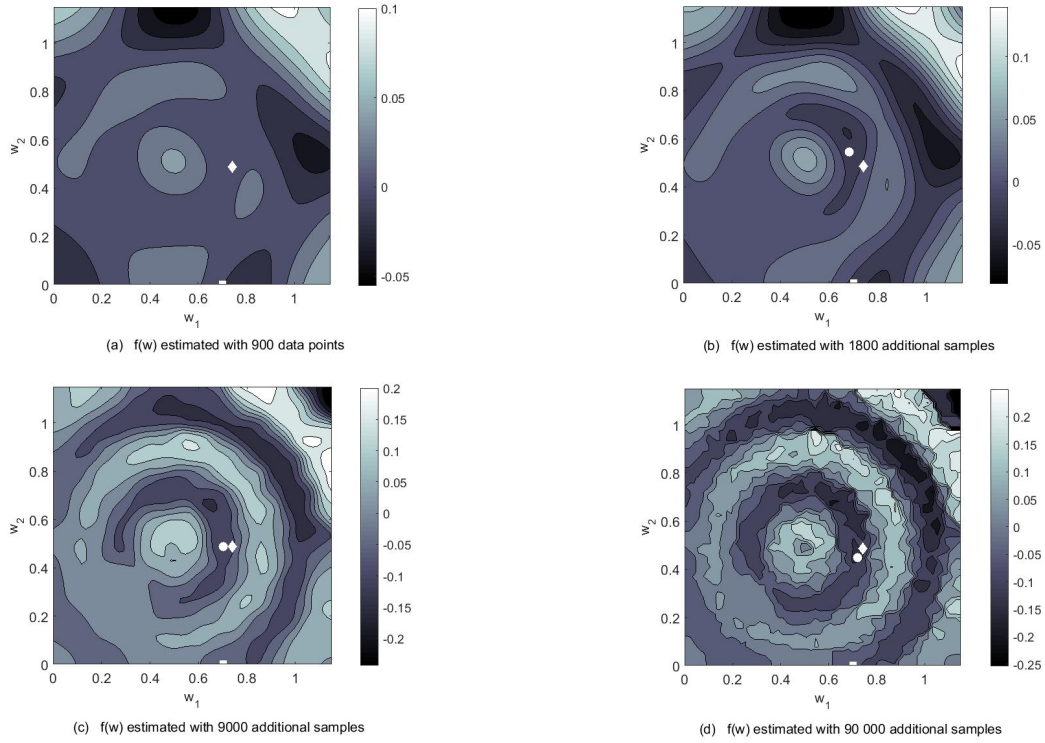


Figure 4. Graph of the objective function $\mathbf{w} \mapsto f(\mathbf{w})$ in contour plot, evaluated in the ν_g points $\mathbf{w}_g^i = (w_{g,1}^i, w_{g,2}^i), i = 1, \dots, \nu_g$ of the grid, and estimated with $N = 900$ data points (a), with $\nu_{\text{sim}} = 1800$ additional samples corresponding to $n_{\text{MC}} = 2$ (b), with $\nu_{\text{sim}} = 9000$ additional samples corresponding to $n_{\text{MC}} = 10$ (c), and with $\nu_{\text{sim}} = 90\,000$ additional samples corresponding to $n_{\text{MC}} = 100$ (d). The white diamond is the optimal reference solution, the white disc is the optimal solution estimated with the additional samples, and the white square is the optimal solution estimated with the data points.

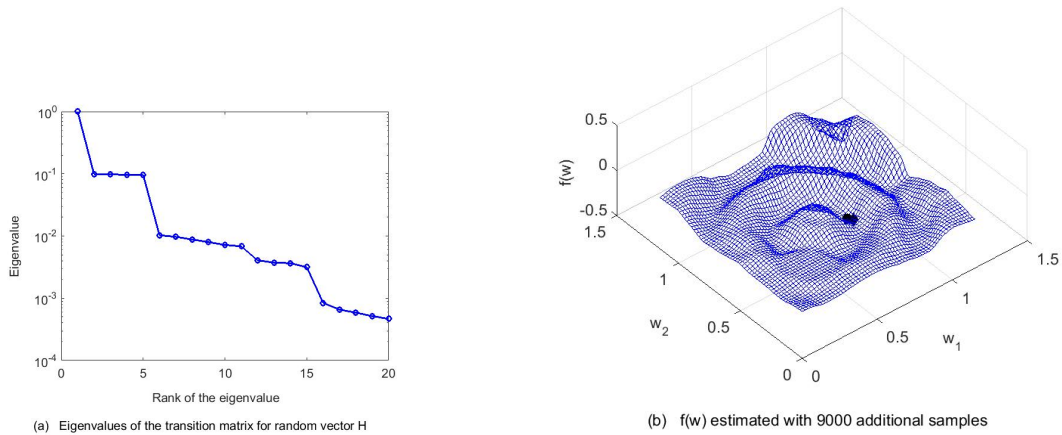


Figure 5. For $N = 900$ data points: (a) Eigenvalues in \log_{10} -scale of the transition matrix for random vector \mathbf{H} . (b) Graph of the objective function $\mathbf{w} \mapsto f(\mathbf{w})$ in 3D evaluated in the ν_g points $\mathbf{w}_g^i = (w_{g,1}^i, w_{g,2}^i), i = 1, \dots, \nu_g$ of the grid, and estimated with $\nu_{\text{sim}} = 9000$ additional samples corresponding to $n_{\text{MC}} = 10$. The black diamond is the optimal solution of reference, the black disc is the optimal solution estimated with the additional samples.

Construction of the optimal solution, $\mathbf{w}_d^{\text{opt}}$, estimated with the N data points. The methodology presented in Section 3 is used for estimating the optimal solution $\mathbf{w}_d^{\text{opt}} = (w_{d,1}^{\text{opt}}, w_{d,2}^{\text{opt}}, w_{d,3}^{\text{opt}}, w_{d,4}^{\text{opt}})$ of

the probabilistic unconstrained optimization problem, using only the N data points, and which is given by

$$\mathbf{w}_d^{\text{opt}} \simeq \arg \min_{i=1, \dots, \nu_g} f(\mathbf{w}_g^i),$$

in which the objective function f is computed by using the algorithm presented in Section 3.3, and which is estimated by using only the N data points. In Table III, the first line entitled "Data points" gives the results for $\mathbf{w}_d^{\text{opt}}$. These results show that the optimal reference solution, $\mathbf{w}_r^{\text{opt}}$, cannot be approximated by using only the N data points represented by $[\mathbf{x}_d]$ with $N = 256$ in dimension 4.

Table III. Table defining the optimal values of the probabilistic unconstrained optimization problem estimated with the N data points and estimated with $\nu_{\text{sim}} = N \times n_{\text{MC}}$ additional samples without evaluation of the objective functions.

N = 256	n_{MC}	$w_{\text{ar},1}^{\text{opt}}$	$w_{\text{ar},2}^{\text{opt}}$	$w_{\text{ar},3}^{\text{opt}}$	$w_{\text{ar},4}^{\text{opt}}$	$f(\mathbf{w}_{\text{ar}}^{\text{opt}})$
Data points		1.1071	1.1500	1.0000	1.1500	1.3140
Additional samples	2	1.0857	1.1286	1.0214	1.0429	1.2663
	10	1.0857	1.1500	1.0214	1.0429	0.9453
	50	0.9786	0.9357	1.0429	1.1071	0.7963
	100	0.9786	0.9357	1.0429	1.0857	0.7272
	300	0.9786	0.9357	1.0429	1.0857	0.6715
	700	0.9786	0.9357	1.0429	1.1071	0.6345
	800	1.0429	1.0643	1.0000	0.9357	0.3112
	900	1.0429	1.0429	0.9786	0.9357	0.3065
	1000	1.0429	1.0429	0.9786	0.9357	0.3024

Construction of the optimal solution, $\mathbf{w}_{\text{ar}}^{\text{opt}}$, estimated with the fixed number N of evaluations and with additional samples without additional function evaluations. The algorithm presented in Section 5 is used. The null space of matrix [cov] is found to be equal to 0 and therefore, $\nu = n$. The smoothing parameter $\varepsilon = 10$, and m is estimated by using Appendix B yielding $m = 6$. The relative error $e_{\text{red}}(m)$ computed with equation (54) yields 6.29×10^{-3} . The parameters for the generation of the additional samples performed by using Section 4.3 are $f_0 = 1.5$, $\Delta r = 0.14239$, and $M_0 = 100$. The number of additional samples that are computed is $\nu_{\text{sim}} = n_{\text{MC}}N$ for the values of n_{MC} given in Table III. Steps 7 to 11 of the algorithm presented in Section 5 are carried out for evaluating the estimated objective function f at the ν_g points $\mathbf{w}^0 = \mathbf{w}_g^i, i = 1, \dots, \nu_g$ of the grid. For each value of n_{MC} , the optimal solution $\mathbf{w}_{\text{ar}}^{\text{opt}}$ of the probabilistic unconstrained optimization problem is estimated using the ν_{sim} additional samples,

$$\mathbf{w}_{\text{ar}}^{\text{opt}} \simeq \arg \min_{i=1, \dots, \nu_g} f(\mathbf{w}_g^i).$$

In Table III, the block of 9 lines entitled "Additional samples" gives the results for $\mathbf{w}_{\text{ar}}^{\text{opt}}$ as a function of the number ν_{MC} such that $\nu_{\text{sim}} = N \times n_{\text{MC}}$ is the number of additional samples. These results show that $f(\mathbf{w}_{\text{ar}}^{\text{opt}})$ decreases when n_{MC} increases and that the fluctuations around 1 of each coordinate $w_{\text{ar},k}^{\text{opt}}$ also decreases when n_{MC} increases. For $n_{\text{MC}} = 900$, the level of these fluctuations around 1 is of the order of the fluctuations given in Table II that are generated by the variation of the fineness of the grid.

Remark. For the deterministic optimization problem for which the objective function $f(\mathbf{w}) = \mathcal{R}(\mathbf{w})$ is the Rosenbrock function defined by equation (42), the Dakota software is [33]. For this dimension $m_{\mathbf{w}} = 4$ and using an initial with coordinates point $w_1 = 0.86$, $w_2 = 0.9$, $w_3 = 1.1$, and $w_4 = 1.15$, for which $f(\mathbf{w}) = 2.96$, the following two observations are made,

(i) the gradient-based search algorithm finds the solution $\mathbf{w}_r^{\text{opt}} = (0.9998, 0.9996, 0.9997, 0.9994)$ with $f(\mathbf{w}_r^{\text{opt}}) = 1.27 \times 10^{-7}$ by using $N_e = 55$ evaluations of function f .

(ii) the global search algorithm finds the solutions:

$\mathbf{w}_r^{\text{opt}} = (0.9532, 0.9086, 1.0479, 1.0982)$ with $f(\mathbf{w}_r^{\text{opt}}) = 4.48 \times 10^{-3}$ after $N_e = 3,400$ evaluations of

function f .

$\mathbf{w}_r^{\text{opt}} = (0.9968, 0.9936, 1.0029, 1.0059)$ with $f(\mathbf{w}_r^{\text{opt}}) = 1.91 \times 10^{-5}$ by using $N_e = 34,710$ evaluations of the function f .

For cases (i) and (ii), if the expression of f was not explicitly known (which is generally the case of interest here), then $f(\mathbf{w})$ would have to be estimated using N'_s samples $\mathcal{Q}(\mathbf{w}; \theta'_s)$ of $\mathcal{Q}(\mathbf{w})$, and consequently, the total number of evaluations of \mathcal{Q} would be $N_e \times N'_s$. For instance, for $N'_s = 100$ (that is a minimum), such an algorithm would require 5,000 evaluations of \mathcal{Q} for case (i) and 340,000 or 3,471,000 for case (ii) (what has to be compared to $N = 256$ evaluations of \mathcal{Q} for the proposed method).

6.3. Application 3

The third application is introduced to demonstrate the applicability of the proposed method in higher dimensions. Specifically, we consider a value $m_w = 10$ in a problem that is somewhat similar to Application 2 above. We will thus refrain from detailing this problem and we limit ourselves to briefly summarizing the specifications of the objective functions, the associated parameters and the solutions sets.

Definition of the probabilistic nonconvex unconstrained optimization problem. The admissible set $\mathcal{C}_w \subset \mathbb{R}^{m_w}$ with $m_w = 10$ and the stochastic process $\{\mathcal{Q}(\mathbf{w}), \mathbf{w} \in \mathcal{C}_w\}$ with values in \mathbb{R} of the optimization problem (see equations (1) to (2)) are defined by $\mathcal{C}_w = \{\mathbf{w} = (w_1, \dots, w_{m_w}) \in [-0.1, 0.1]^{m_w} \subset \mathbb{R}^{m_w}\}$ and

$$\mathcal{Q}(\mathbf{w}) = \beta_1 \mathcal{H}(\mathbf{w}) + \beta_2, \quad (46)$$

in which $\mathbf{w} \mapsto \mathcal{H}(\mathbf{w})$ is the function that is written as

$$\mathcal{H}(\mathbf{w}) = -\sin(4 \|\mathbf{w}\|) / (4 \|\mathbf{w}\|). \quad (47)$$

In equation (47), the \mathbb{R}^2 -valued random variable $\beta = (\beta_1, \beta_2)$ is defined by equations (43) and (44) for which the hyperparameters that control the level of uncertainty are fixed at values of $\delta_1 = \delta_2 = 0.1$. It can be easily seen that that function $\mathbf{w} \mapsto f(\mathbf{w}) = E\{\mathcal{Q}(\mathbf{w})\}$ is not convex on \mathcal{C}_w . The exact solution of the probabilistic nonconvex unconstrained optimization problem defined by equations (1) and (2) is $w_j^{\text{opt}} = 0$ for all $j = 1, \dots, m_w$ and $f(\mathbf{w}^{\text{opt}}) = -1$.

Definition of the fixed number N of data points. For this numerical application, $n = m_w + 1 = 10 + 1 = 11$, and the value of N is 7,000. The data set $\mathbf{x}^\ell = (\mathbf{w}^\ell, q^\ell)$ for $\ell = 1, \dots, N$, which has been defined in Section 2.2 is constructed as follows. For $\ell = 1, \dots, N$, the points $\mathbf{w}^\ell = (w_1^\ell, \dots, w_{m_w}^\ell)$ are generated, for $k = 1, \dots, m_w$ by $w_k^\ell = -0.1 + 0.2 \widehat{\chi}_k^\ell$ in which $\{\widehat{\chi}_k^\ell\}_{k,\ell}$ are $m_w \times N$ independent samples of the independent uniform random variables $\{\widehat{\chi}_k, k = 1, \dots, m_w\}$ on $[0, 1]$, which are independent of χ_1 and χ_2 . By using the stochastic model $\mathcal{Q}(\mathbf{w})$ defined by equation (46), we thus deduce q^ℓ for $\ell = 1, \dots, N$ by using equation (5).

Construction of the optimal solution, $\mathbf{w}_{\text{ar}}^{\text{opt}}$, estimated with the fixed number N of evaluations and with additional samples without additional function evaluations. The algorithm presented in Section 5 is used. The null space of matrix $[\text{cov}]$ is found to be equal to 0 and therefore, $\nu = n$. The smoothing parameter is taken as $\varepsilon = 10$, and seven eigenvectors are retained for the diffusion maps ($m = 11$) which are associated with the eigenvalues of rank 2 to 12. The relative error $e_{\text{red}}(m)$ computed with equation (54) yields 5.44×10^{-4} . The parameters for the generation of the additional samples performed by using Section 4.3 are $f_0 = 150$, $\Delta r = 0.12582$, $M_0 = 1$, and $n_{\text{MC}} = 100$. The number of additional samples that are computed is $\nu_{\text{sim}} = n_{\text{MC}} N = 700,000$. Steps 7 to 11 of the algorithm presented in Section 5 are carried out for evaluating the estimated objective function f at the points \mathbf{w}_0 generated by the genetic algorithm that is used for solving the nonconvex unconstrained optimization problem. The initial population for the genetic algorithm is a sample of size 7,000 drawn from a uniform distribution on \mathcal{C}_w . The optimal solution obtained is $\mathbf{w}_r^{\text{opt}} = (-0.037, 0.033, 0.009, 0.008, 0.020, 0.018, 0.025, -0.027, 0.017, -0.006)$ with $f(\mathbf{w}_r^{\text{opt}}) = -1.001$, which is in close agreement with the exact solution.

It is clear from the sequence of applications presented in this paper that the proposed methodology

relies on the availability of a repertoire of deterministic optimization algorithms adapted to the problem at hand. The proposed methodology then accelerates statistical evaluations and averages required by these algorithms. From this perspective, it is apparent that scalability of the proposed method is inextricably tied to scalability of the underlying optimization algorithm. The added cost associated with sampling from the manifold-constrained distribution is a very small fraction of the cost of typical function evaluations. This involves solving an initial value problem for a system of coupled ordinary differential equations, requiring no additional calls to the expensive function.

It is also apparent from the three examples presented above that although the proposed method was not designed to accelerate the convergence rate of deterministic functions, it does serve that purpose to a significant extent. Specifically, accuracy in the estimated optimal solution is achieved that is consistent with orders of magnitude more function evaluations than is actually used. While this improvement has the most significant ramifications in OUU problems where the statistical fluctuations can be traced to specific random variables in the underlying models, it remains very valuable in deterministic problems. In these cases, the convergence acceleration to the deterministic problem must also be interpreted and analyzed statistically, a task that is outside the scope of this paper.

7. CONCLUSIONS

The paper presented a new perspective on optimization under uncertainty whereby the objective function is viewed as the average of data scattered around a manifold. By integrating methods from machine learning, statistical smoothing, and stochastic analysis, the distribution of this scattered data is characterized and an Itô equation projected onto an underlying manifold is constructed to generate associated statistical samples. The savings from the proposed procedure will be greater the more uncertainty is present in the problem, resulting in greater scatter in the data. In these situations, standard procedures typically require a very large number of function evaluations at each design point.

The purpose of the present paper was to introduce the method with its mathematical background and numerical demonstrations. The extent of comparison with other methods is currently limited to the impressive reduction in the number of function evaluations required for estimating the statistical averages present in the objective and constraint functions. Applications to high dimensional optimization problems will greatly benefit from the development of adapted optimization logic, a task beyond the scope of the present paper.

A. CONSTRUCTION OF THE DIFFUSION-MAPS BASIS

In this appendix, we summarize the construction of the diffusion map basis based on [21, 22] and detailed in [29]. Let $[\mathbf{b}]$ be the positive-definite diagonal real matrix in \mathbb{M}_N such that $[\mathbf{b}]_{ij} = \delta_{ij} \sum_{j'=1}^N [K]_{ij'}$ in which $[K]_{ij'} = k_\varepsilon(\boldsymbol{\eta}^i, \boldsymbol{\eta}^{j'})$. Let $[\mathbf{P}]$ be the transition matrix in \mathbb{M}_N such that $[\mathbf{P}] = [\mathbf{b}]^{-1} [K]$ and let $[\mathbf{P}_S]$ be the symmetric matrix in \mathbb{M}_N such that $[\mathbf{P}_S] = [\mathbf{b}]^{1/2} [\mathbf{P}] [\mathbf{b}]^{-1/2} = [\mathbf{b}]^{-1/2} [K] [\mathbf{b}]^{-1/2}$. Let m be an integer such that $1 < m \leq N$. The eigenvalues of $[\mathbf{P}_S]$ $\phi^\alpha = \Lambda_\alpha \phi^\alpha$ are positive and such that $1 = \Lambda_1 > \Lambda_2 \geq \dots \geq \Lambda_m$. Let $[\phi]$ be the matrix in $\mathbb{M}_{N,m}$ such that $[\phi]^T [\phi] = [I_m]$, whose columns are the m orthonormal eigenvectors ϕ^1, \dots, ϕ^m associated with $\Lambda_1, \dots, \Lambda_m$. The right eigenvectors ψ^1, \dots, ψ^m associated with $\Lambda_1, \dots, \Lambda_m$, which are such that $[\mathbf{P}] \psi^\alpha = \Lambda_\alpha \psi^\alpha$, are written as

$$\psi^\alpha = [\mathbf{b}]^{-1/2} \phi^\alpha \in \mathbb{R}^N, \quad \alpha = 1, \dots, m, \quad (48)$$

and consequently, the matrix $[\psi] = [\psi^1 \dots \psi^m] = [\mathbf{b}]^{-1/2} [\phi] \in \mathbb{M}_{N,m}$ is such that

$$[\psi]^T [\mathbf{b}] [\psi] = [I_m] \quad (49)$$

which defines the normalization of the right eigenvectors of $[\mathbb{P}]$. A "diffusion-maps basis" is defined by $[g] = [\mathbf{g}^1 \dots \mathbf{g}^m] \in \mathbb{M}_{N,m}$ (which is an algebraic basis of \mathbb{R}^N for $m = N$) such that

$$\mathbf{g}^\alpha = \Lambda_\alpha^\zeta \boldsymbol{\psi}^\alpha \in \mathbb{R}^N, \quad \alpha = 1, \dots, m, \quad (50)$$

in which ζ is an integer that is chosen for fixing the analysis scale of the local geometric structure of the dataset. It should be noted that the family $\{\Psi_\zeta\}$ of diffusion maps are defined [21, 22] by the vector $\Psi_\zeta = (\Lambda_1^\zeta \boldsymbol{\psi}^1, \dots, \Lambda_m^\zeta \boldsymbol{\psi}^m)$ in order to construct a diffusion distance, and integer ζ is thus such that the probability of transition is in ζ steps. However, as it has been explained in [29], we do not use such a diffusion distance.

B. CRITERION FOR ESTIMATING AN OPTIMAL VALUE OF m

In this Appendix, we recall the criterion introduced in [29] for estimating a value of dimension m . Let $[x_d] \in \mathbb{M}_{n,N}$ be the matrix of the dataset introduced in Section 4.1 and let $[\eta_d] \in \mathbb{M}_{\nu,N}$ be the matrix computed with equation (21). We then introduce the matrix $[x_{\text{red}}(m)] \in \mathbb{M}_{n,N}$ such that (see equations (23) and (25)),

$$[x_{\text{red}}(m)] = [\underline{x}] + [\varphi] [\lambda]^{1/2} [z_d] [g]^T, \quad [z_d] = [\eta_d] [a]. \quad (51)$$

Let $\mathbf{x}_{\text{red}}^1(m), \dots, \mathbf{x}_{\text{red}}^N(m)$ be the N vectors in \mathbb{R}^n , which constitute the N columns of matrix $[x_{\text{red}}(m)] \in \mathbb{M}_{n,N}$. We then introduce the empirical estimates $\underline{\mathbf{x}}_{\text{red}}(m) \in \mathbb{R}^n$ and $[\text{cov}_{\text{red}}(m)] \in \mathbb{M}_n$ of the mean value and of the covariance matrix calculated with the sample $[x_{\text{red}}(m)] = [\mathbf{x}_{\text{red}}^1(m) \dots \mathbf{x}_{\text{red}}^N(m)] \in \mathbb{M}_{n,N}$ such that

$$\underline{\mathbf{x}}_{\text{red}}(m) = \frac{1}{N} \sum_{\ell=1}^N \mathbf{x}_{\text{red}}^\ell(m), \quad (52)$$

$$[\text{cov}_{\text{red}}(m)] = \frac{1}{N-1} \sum_{\ell=1}^N (\mathbf{x}_{\text{red}}^\ell(m) - \underline{\mathbf{x}}_{\text{red}}(m)) (\mathbf{x}_{\text{red}}^\ell(m) - \underline{\mathbf{x}}_{\text{red}}(m))^T. \quad (53)$$

A criterion for the mean-square convergence is then specified as

$$e_{\text{red}}(m) = \frac{\|[\text{cov}_{\text{red}}(m)] - [\text{cov}]\|_F}{\|[\text{cov}]\|_F}. \quad (54)$$

in which $[\text{cov}]$ is the empirical estimate of the covariance matrix of random vector \mathbf{X} , which is such that

$$[\text{cov}] = \frac{1}{N-1} \sum_{\ell=1}^N (\mathbf{x}^\ell - \underline{\mathbf{x}}) (\mathbf{x}^\ell - \underline{\mathbf{x}})^T, \quad \underline{\mathbf{x}} = \frac{1}{N} \sum_{\ell=1}^N \mathbf{x}^\ell. \quad (55)$$

Since $[x_{\text{red}}(N)] = [x_d]$, it can be deduced that $e_{\text{red}}(m) \rightarrow 0$ when m goes to N . For a fixed reasonable value $\epsilon_0 > 0$ of the relative tolerance $e_{\text{red}}(m)$, an estimate of m will consist of looking for the smallest value of m such that $e_{\text{red}}(m) \leq \epsilon_0$.

C. ALGORITHM FOR SOLVING THE REDUCED-ORDER ISDE

The algorithm for solving the reduced-order ISDE defined by equations (26) to (28) is detailed in [29] and is summarized hereinafter. The Störmer-Verlet scheme is used. Let $M = n_{\text{MC}} \times M_0$ be the positive integer in which n_{MC} and M_0 have been introduced in Section 4.3. The reduced-order ISDE is solved on the finite interval $\mathcal{R} = [0, M \Delta r]$, in which Δr is the sampling step of the continuous index parameter r . The integration scheme is based on the use of the $M + 1$ sampling points $r_{\ell'}$ such that $r_{\ell'} = \ell' \Delta r$ for $\ell' = 0, \dots, M$. The following notations are introduced: $[\mathcal{Z}_{\ell'}] = [\mathcal{Z}(r_{\ell'})]$,

$[\mathcal{Y}^{\ell'}] = [\mathcal{Y}(r^{\ell'})]$, and $[\mathcal{W}^{\ell'}] = [\mathcal{W}(r^{\ell'})]$, for $\ell' = 0, \dots, M$, with

$$[\mathcal{Z}_0] = [\mathbf{H}_d][a] \quad , \quad [\mathcal{Y}_0] = [\mathcal{N}][a] \quad , \quad [\mathcal{W}_0] = [0_{\nu, m}] \quad a.s. \quad (56)$$

For $\ell' = 0, \dots, M - 1$, let

$$[\Delta \mathcal{W}^{\ell'+1}] = [\Delta \mathbb{W}^{\ell'+1}][a], \quad (57)$$

be the sequence of random matrices with values in $\mathbb{M}_{\nu, m}$, in which $[\Delta \mathbb{W}^{\ell'+1}] = [\mathbb{W}^{\ell'+1}] - [\mathbb{W}^{\ell'}]$. The increments $[\Delta \mathbb{W}_1], \dots, [\Delta \mathbb{W}_M]$ are M independent random matrices with values in $\mathbb{M}_{\nu, N}$. For all $k = 1, \dots, \nu$ and for all $j = 1, \dots, N$, the real-valued random variables $\{[\Delta \mathbb{W}^{\ell'+1}]_{kj}\}_{kj}$ are independent, Gaussian, second-order, and centered random variables such that

$$E\{[\Delta \mathbb{W}^{\ell'+1}]_{kj}[\Delta \mathbb{W}^{\ell'+1}]_{k'j'}\} = \Delta r \delta_{kk'} \delta_{jj'}. \quad (58)$$

For $\ell' = 0, \dots, M - 1$, the Störmer-Verlet scheme applied to equations. (26) and (27) yields

$$[\mathcal{Z}^{\ell'+\frac{1}{2}}] = [\mathcal{Z}^{\ell'}] + \frac{\Delta r}{2} [\mathcal{Y}^{\ell'}], \quad (59)$$

$$[\mathcal{Y}^{\ell'+1}] = \frac{1-b}{1+b} [\mathcal{Y}^{\ell'}] + \frac{\Delta r}{1+b} [\mathcal{L}^{\ell'+\frac{1}{2}}] + \frac{\sqrt{f_0}}{1+b} [\Delta \mathcal{W}^{\ell'+1}], \quad (60)$$

$$[\mathcal{Z}^{\ell'+1}] = [\mathcal{Z}^{\ell'+\frac{1}{2}}] + \frac{\Delta r}{2} [\mathcal{Y}^{\ell'+1}], \quad (61)$$

with the initial condition defined by equation (28), where $b = f_0 \Delta r / 4$, and where $[\mathcal{L}^{\ell'+\frac{1}{2}}]$ is the $\mathbb{M}_{\nu, m}$ -valued random variable such that

$$[\mathcal{L}^{\ell'+\frac{1}{2}}] = [\mathcal{L}([\mathcal{Z}^{\ell'+\frac{1}{2}}])] = [L([\mathcal{Z}^{\ell'+\frac{1}{2}}][g^T])[a], \quad (62)$$

in which, for all $[u] = [\mathbf{u}^1 \dots \mathbf{u}^N]$ in $\mathbb{M}_{\nu, N}$ with $\mathbf{u}^{\ell'} = (u_1^{\ell'}, \dots, u_\nu^{\ell'})$ in \mathbb{R}^ν , the entries of matrix $[L([u])]$ in $\mathbb{M}_{\nu, N}$ are defined by equations (30) to (33).

ACKNOWLEDGMENTS

This research was supported by the ScramJet-UQ project funded under DARPA's EQUIPS Program.

REFERENCES

1. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 1998; :2278–2324.
2. Spall J. *Introduction to Stochastic Search and Optimization*. John Wiley and Sons: Hoboken, New Jersey, 2003.
3. Konak A, Coit D, Smith A. Multi-objective optimization using genetic algorithms: A tutorial. *Reliability Engineering and System Safety* 2006; **91**(9):992–1007.
4. Coello C, Carlos A. Evolutionary multi-objective optimization: A historical view of the field. *IEEE Computational Intelligence Magazine* 2006; **1**(1):28–36.
5. Jones D, Schonlau M, Welch W. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization* 1998; **13**:455–492.
6. Queipo N, Haftka R, Shyy W, Goel T, Vaidyanathan R, Tucker K. Surrogate-based analysis and optimization. *Progress in Aerospace Science* 2005; **41**:1–28.
7. Byrd R, Chin G, Neveitt W, Nocedal J. On the use of stochastic hessian information in optimization methods for machine learning. *SIAM Journal of Optimization* 2011; **21**(3):977–995.
8. Homem-de Mello T, Bayraksan G. Monte Carlo sampling-based methods for stochastic optimization. *Surveys in Operations Research and Management Science* 2014; **19**:56–85.
9. Kleijnen J, van Beers W, van Nieuwenhuysse I. Constrained optimization in expensive simulation: Novel approach. *European Journal of Operational Research* 2010; **202**:164–174.
10. Sui Y, Gotovos A, Burdick J, Krause A. Safe exploration for optimization with gaussian processes. *Proceedings of the 32nd International Conference on Machine Learning*, vol. JMLR: W&CP Volume 37, Lille, France, 2015.
11. Wang Z, Zoghi M, Hutter F, Matheson D, de Freitas N. Bayesian optimization in a billion dimensions via random embeddings. *Journal of Artificial Intelligence Research* 2016; **55**:361–387, doi:10.1613/jair.4806. URL <http://jair.org/media/4806/live-4806-9131-jair.pdf>.

12. Xie J, Frazier P, Chick S. Bayesian optimization via simulation with pairwise sampling and correlated pair beliefs. *Operations Research* 2016; **64**(2):542–559.
13. Ghanem R, Spanos P. *Stochastic Finite Elements: A Spectral Approach*. Springer-Verlag, 1991.
14. Soize C. A nonparametric model of random uncertainties for reduced matrix models in structural dynamics. *Probabilistic Engineering Mechanics* 2000; **15**(3):277–294.
15. Soize C, Ghanem R. Physical systems with random uncertainties: Chaos representations with arbitrary probability measure. *SIAM Journal of Scientific Computing* 2004; **26**(2):395–410.
16. Thimmisetty C, Khodabakhshnejad A, Jabbari N, Aminzadeh F, Ghanem R, Rose K, Bauer J, Disenhof C. Multiscale stochastic representation in high-dimensional data using gaussian processes with implicit diffusion metrics. *Lecture Notes in Computer Science* 2015; **8964**.
17. Ghanem R, Higdon D, Owhadi HE. *Handbook of Uncertainty Quantification*. Springer-Verlag, 2017.
18. Du X, Chen W. Sequential optimization and reliability assessment method for efficient probabilistic design. *ASME Journal of Mechanical Design* 2004; **126**(2):225–233.
19. Yao W, Chen X, Luo W, vanTooren M, Guo J. Review of uncertainty-based multidisciplinary design optimization methods for aerospace vehicles. *Progress in Aerospace Sciences* 2011; **47**:450–479.
20. Eldred M. Design under uncertainty employing stochastic expansion methods. *International Journal for Uncertainty Quantification* 2011; **1**(2):119–146.
21. Coifman R, Lafon S, Lee A, Maggioni M, Nadler B, Warner F, Zucker S. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *PNAS* 2005; **102**(21):7426–7431.
22. Coifman R, Lafon S. Diffusion maps, applied and computational harmonic analysis. *Applied and Computational Harmonic Analysis* 2006; **21**(1):5–30.
23. Talmon R, Coifman R. Intrinsic modeling of stochastic dynamical systems using empirical geometry. *Applied and Computational Harmonic Analysis* 2015; **39**(1):138–160.
24. Soize C. 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* 2008; **76**(10):1583–1611.
25. Neal R. MCMC using hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, Brooks S, Gelman A, Jones G, Meng X (eds.), Chapman and Hall-CRC Press, Boca Raton, 2010.
26. Girolami M, Calderhead B. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society* 2011; **73**:123–214.
27. Kaipio J, Somersalo E. *Statistical and Computational Inverse Problems*. Springer-Verlag: New York, 2005.
28. Robert C, Casella G. *Monte Carlo Statistical Methods*. Springer-Verlag: New York, 2005.
29. Soize C, Ghanem R. Data-driven probability concentration and sampling on manifold. *Journal of Computational Physics* 2016; **321**:242–258, doi:10.1016/j.jcp.2016.05.044.
30. Bowman A, Azzalini A. *Applied Smoothing Techniques for Data Analysis*. Oxford University Press: Oxford, UK, 1997.
31. Scott D. *Multivariate Density Estimation: Theory, Practice, and Visualization*. 2nd edn., John Wiley and Sons: New York, 2015.
32. Rosenbrock H. An automatic method for finding the greatest or least value of a function. *The Computer Journal* 1960; **3**(3):175–184, doi:10.1093/comjnl/3.3.175.
33. Adams B, Bauman L, Bohnhoff W, Dalbey K, Ebeida M, Eddy J, Eldred M, Hough P, Hu K, Jakeman J, *et al.*. Dakota, a multilevel parallel object-oriented framework for design optimization, parameter estimation, uncertainty quantification, and sensitivity analysis: Version 6.0 users manual. *Sandia Technical Report SAND2014-4633*, Sandia National Laboratories 2015.