

Calculation of Lagrange multipliers in the construction of maximum entropy distributions in high stochastic dimension

Anas Batou, Christian Soize

► **To cite this version:**

Anas Batou, Christian Soize. Calculation of Lagrange multipliers in the construction of maximum entropy distributions in high stochastic dimension. SIAM/ASA Journal on Uncertainty Quantification, ASA, American Statistical Association, 2013, 1 (1), pp.431-451. 10.1137/120901386 . hal-00851201

HAL Id: hal-00851201

<https://hal-upec-upem.archives-ouvertes.fr/hal-00851201>

Submitted on 13 Aug 2013

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.

Calculation of Lagrange multipliers in the construction of maximum entropy distributions in high stochastic dimension *

A. Batou[†] and C. Soize[†]

Abstract. The research addressed here concerns the construction of the probability distribution of a random vector in high dimension using the maximum entropy (MaxEnt) principle under constraints defined by the available information. In this paper, a new algorithm, adapted to the high stochastic dimension, is proposed to identify the Lagrange multipliers introduced to take into account the constraints in the MaxEnt principle. This new algorithm is based on (1) the minimization of an appropriate convex functional and (2) the construction of the probability distribution defined as the invariant measure of an Itô Stochastic Differential Equation. The methodology is validated through an application devoted to the generation of accelerograms which are physically consistent and spectrum compatible.

Key words. Maximum Entropy principle; High stochastic dimension; Lagrange multipliers; Accelerograms.

AMS subject classifications.

1. Introduction. The construction of the probability distribution of random vectors under some constraints defined by the available information requires the use of an adapted method, such as the maximum entropy (MaxEnt) principle, in order to obtain a well-posed problem. The available information can be obtained from experimental realizations and/or physical considerations. In the case of an engineering design, this information can also be provided by specifications. In general, due to the limited number of experimental realizations, the direct nonparametric statistical estimation of the probability distribution is inaccurate in high stochastic dimension and a parametric representation is a way to circumvent this difficulty. The parameters of this representation can be estimated using parametric statistics. The MaxEnt principle [23, 14, 31, 24, 18, 15] is a method which allows such a parametric representation to be constructed and if a solution exists, then to uniquely identify the parameters of the probability distribution (Lagrange multipliers associated with the constraints) using only the available information. This method has been applied successfully for a large class of applications, for instance, in image reconstruction and image processing [25], in acoustic processing [20], for modeling species geographic distributions with presence-only data [21], for econometric applications [11, 33], for natural language processing [3], for many applications in the area of physics [16, 17, 32, 8] and mechanics [26, 28, 5], and in many other fields.

Even if the MaxEnt principle allows the construction of the probability distribution to be carried out, two difficulties remain. The first one concerns the construction of a generator of independent realizations of the random vector in high dimension. The second one is related to the estimation of the high-dimension integrals which have to be computed in order to calculate the optimal value of the Lagrange multipliers. These two difficulties increase with

*This research was supported by the "Agence Nationale de la Recherche", Contract TYCHE, ANR-2010-BLAN-0904.

[†]Université Paris-Est, Laboratoire Modélisation et Simulation Multi Echelle, MSME UMR 8208 CNRS, 5 bd Descartes, 77454 Marne-la-Vallée, France

the stochastic dimension.

Recently, a new methodology [29], which consists in constructing the MaxEnt probability distribution as the invariant measure of an Itô Stochastic Differential Equation (ISDE), has been introduced for the low and high stochastic dimension. This method allows independent realizations of a high-dimension random vector constructed using the MaxEnt principle to be generated and then the high dimensional integrals to be estimated using the Monte Carlo simulation method or using the ergodic method.

The objective of this paper is to improve the methodology introduced in [29] by introducing a more efficient algorithm for the identification of the Lagrange multipliers. The proposed algorithm is based (1) on the minimization of the convex cost function introduced by [1], (2) at each iteration, on the use of an ISDE to construct independent realizations allowing the gradient and the Hessian to be estimated, (3) on the use of efficient initial conditions for the ISDE. This algorithm allows the Lagrange multipliers to be identified with a few number of Newton iterations.

In Section 2, the construction of the MaxEnt probability distribution is briefly summarized. Section 3 is devoted to the identification of the Lagrange multipliers introduced in this construction. In Section 4, the methodology is validated in high dimension (random vector with 1,600 components) through an application which deals with the generation of accelerograms which are physically consistent and spectrum compatible .

2. Construction of the MaxEnt probability distribution. Let $\mathbf{A} = (A_1, \dots, A_N)$ be a \mathbb{R}^N -valued second-order random variable. The objective of this section is to construct the probability density function $\mathbf{a} \mapsto p_{\mathbf{A}}(\mathbf{a})$ of the random vector \mathbf{A} using the MaxEnt principle under the constraints defined by the available information relative to random vector \mathbf{A} . The support of the probability density function (pdf) is assumed to be all the set \mathbb{R}^N .

2.1. Available information. The available information for the random vector \mathbf{A} can be deduced from physical considerations and/or experimental measurements. In some cases, such as the design of a structure, the available information can also be deduced from engineering specifications. Let $E\{\cdot\}$ be the mathematical expectation. We suppose that the available information is written as

$$E\{\mathbf{g}(\mathbf{A})\} = \mathbf{f}, \quad (2.1)$$

in which $\mathbf{a} \mapsto \mathbf{g}(\mathbf{a})$ is a given function from \mathbb{R}^N into \mathbb{R}^μ and where \mathbf{f} is a given vector in \mathbb{R}^μ . Equation (2.1) can be rewritten as

$$\int_{\mathbb{R}^N} \mathbf{g}(\mathbf{a}) p_{\mathbf{A}}(\mathbf{a}) d\mathbf{a} = \mathbf{f}. \quad (2.2)$$

An additional constraint relative to the normalization of the distribution $p_{\mathbf{A}}(\mathbf{a})$ is introduced such that

$$\int_{\mathbb{R}^N} p_{\mathbf{A}}(\mathbf{a}) d\mathbf{a} = 1. \quad (2.3)$$

It should be noted that the available information type defined by Eq. (2.1) is not a limitation. Indeed, most of the information available in physical phenomena are written or can be

rewritten in the form defined by Eq. (2.1). As it will be seen in the applications of Section 4, we can also control each realization of random vector \mathbf{A} by imposing appropriate constraints.

2.2. MaxEnt probability distribution. The entropy of the probability density function $\mathbf{a} \mapsto p_{\mathbf{A}}(\mathbf{a})$ is defined by

$$S(p) = - \int_{\mathbb{R}^N} p_{\mathbf{A}}(\mathbf{a}) \log(p_{\mathbf{A}}(\mathbf{a})) d\mathbf{a}, \quad (2.4)$$

which is defined as a measure of uncertainty for $p_{\mathbf{A}}(\mathbf{a})$ and where \log is the natural logarithm. Let \mathcal{C} be the set of all the probability density functions defined on \mathbb{R}^N with values in \mathbb{R}^+ , verifying the constraints defined by Eqs. (2.2) and (2.3). Then the MaxEnt principle consists in constructing the probability density function $\mathbf{a} \mapsto p_{\mathbf{A}}(\mathbf{a})$ as the unique pdf in \mathcal{C} which maximizes the entropy defined by Eqs. (2.4). Then by introducing a Lagrange multiplier λ_0 in \mathbb{R}^+ associated with Eq. (2.3) and a Lagrange multiplier $\boldsymbol{\lambda}$ associated with Eq. (2.2) and belonging to an admissible open subset \mathcal{L}_μ of \mathbb{R}^μ , it can be shown (see [14, 15]) that the MaxEnt solution, if it exists, is defined by

$$p_{\mathbf{A}}(\mathbf{a}) = c_0^{\text{sol}} \exp(-\langle \boldsymbol{\lambda}^{\text{sol}}, \mathbf{g}(\mathbf{a}) \rangle), \quad (2.5)$$

in which $c_0^{\text{sol}} = \exp(-\lambda_0^{\text{sol}})$, $\langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1 + \dots + x_\mu y_\mu$ and where λ_0^{sol} and $\boldsymbol{\lambda}^{\text{sol}}$ are respectively the values of λ_0 and $\boldsymbol{\lambda}$ for which Eqs. (2.2) and (2.3) are satisfied. Using Eqs. (2.3) and (2.5), the parameter c_0^{sol} is given by

$$c_0^{\text{sol}} = \left\{ \int_{\mathbb{R}^N} \exp(-\langle \boldsymbol{\lambda}^{\text{sol}}, \mathbf{g}(\mathbf{a}) \rangle) d\mathbf{a} \right\}^{-1}. \quad (2.6)$$

Therefore, Eq. (2.5) can be rewritten as

$$p_{\mathbf{A}}(\mathbf{a}) = c_0(\boldsymbol{\lambda}^{\text{sol}}) \exp(-\langle \boldsymbol{\lambda}^{\text{sol}}, \mathbf{g}(\mathbf{a}) \rangle), \quad (2.7)$$

in which $\boldsymbol{\lambda}^{\text{sol}}$ is such that Eq. (2.2) is satisfied and where $c_0(\boldsymbol{\lambda})$ is defined by

$$c_0(\boldsymbol{\lambda}) = \left\{ \int_{\mathbb{R}^N} \exp(-\langle \boldsymbol{\lambda}, \mathbf{g}(\mathbf{a}) \rangle) d\mathbf{a} \right\}^{-1}. \quad (2.8)$$

The identification of the vector $\boldsymbol{\lambda}^{\text{sol}}$ of the Lagrange multipliers is the objective of this paper and is addressed in the next Section.

3. Calculation of the Lagrange multipliers. In this section, we propose a methodology for the calculation of $\boldsymbol{\lambda}^{\text{sol}}$. We first present the general methodology for the calculation of this vector. Then we present the Gaussian particular case for which the gradient and the Hessian of the objective function can explicitly be calculated. The Gaussian case is useful (i) either if the available information consists only of linear or affine transformations of statistical second-order moments (ii) or for initializing the value of the Lagrange multipliers in the iterative algorithm devoted to the non-Gaussian case. Finally, we present the case for a general MaxEnt distribution for which integrals on \mathbb{R}^N have to be estimated using an appropriate methodology.

3.1. Objective function and methodology. Using Eqs. (2.7) and (2.2), vector $\boldsymbol{\lambda}^{\text{sol}}$ is the solution in $\boldsymbol{\lambda}$ of the following set of μ nonlinear algebraic equations

$$\int_{\mathbb{R}^N} \mathbf{g}(\mathbf{a}) c_0(\boldsymbol{\lambda}) \exp(-\langle \boldsymbol{\lambda}, \mathbf{g}(\mathbf{a}) \rangle) d\mathbf{a} = \mathbf{f}. \quad (3.1)$$

A more convenient way to calculate vector $\boldsymbol{\lambda}^{\text{sol}}$ consists in solving the following optimization problem (see [1, 11, 19, 6]),

$$\boldsymbol{\lambda}^{\text{sol}} = \arg \min_{\boldsymbol{\lambda} \in \mathcal{L}_\mu \subset \mathbb{R}^\mu} \Gamma(\boldsymbol{\lambda}), \quad (3.2)$$

in which the objective function Γ is written as

$$\Gamma(\boldsymbol{\lambda}) = \langle \boldsymbol{\lambda}, \mathbf{f} \rangle - \log(c_0(\boldsymbol{\lambda})). \quad (3.3)$$

It should be noted that the great interest of such a formulation is that the introduced objective function Γ does not depend on the Lagrange multiplier λ_0 associated with the constant of normalization. We recall that, under the same hypotheses, the usual strictly convex objective function $(\lambda_0, \boldsymbol{\lambda}) \mapsto \tilde{\Gamma}(\lambda_0, \boldsymbol{\lambda})$ is written (see for instance [15]) as $\tilde{\Gamma}(\lambda_0, \boldsymbol{\lambda}) = \lambda_0 + \langle \boldsymbol{\lambda}, \mathbf{f} \rangle + e^{-\lambda_0} \{c_0(\boldsymbol{\lambda})\}^{-1}$, which depends on the constant of normalization. Let $\{\mathbf{A}_\lambda, \boldsymbol{\lambda} \in \mathcal{L}_\mu\}$ be a family of random variables for which the pdf is defined, for all $\boldsymbol{\lambda}$ in \mathcal{L}_μ , by

$$p_{\mathbf{A}_\lambda}(\mathbf{a}) = c_0(\boldsymbol{\lambda}) \exp(-\langle \boldsymbol{\lambda}, \mathbf{g}(\mathbf{a}) \rangle). \quad (3.4)$$

Then the gradient vector $\nabla \Gamma(\boldsymbol{\lambda})$ and the Hessian matrix $[H(\boldsymbol{\lambda})]$ of function $\boldsymbol{\lambda} \mapsto \Gamma(\boldsymbol{\lambda})$ are written as

$$\nabla \Gamma(\boldsymbol{\lambda}) = \mathbf{f} - E\{\mathbf{g}(\mathbf{A}_\lambda)\}. \quad (3.5)$$

$$[H(\boldsymbol{\lambda})] = E\{\mathbf{g}(\mathbf{A}_\lambda)\mathbf{g}(\mathbf{A}_\lambda)^T\} - E\{\mathbf{g}(\mathbf{A}_\lambda)\} E\{\mathbf{g}(\mathbf{A}_\lambda)\}^T, \quad (3.6)$$

in which \mathbf{u}^T is the transpose of \mathbf{u} . It can be noted that the Hessian matrix $[H(\boldsymbol{\lambda})]$ is the covariance matrix of the random vector $\mathbf{g}(\mathbf{A}_\lambda)$. It is assumed that the constraints defined by Eq. (2.2) are algebraically independent. Consequently, the Hessian matrix is positive definite and therefore, function $\boldsymbol{\lambda} \mapsto \Gamma(\boldsymbol{\lambda})$ is strictly convex and reaches its minimum for $\boldsymbol{\lambda}^{\text{sol}}$ which is such that $\nabla \Gamma(\boldsymbol{\lambda}) = \mathbf{0}$ for $\boldsymbol{\lambda} = \boldsymbol{\lambda}^{\text{sol}}$. It can then be deduced that the minimum of function $\boldsymbol{\lambda} \mapsto \Gamma(\boldsymbol{\lambda})$ corresponds to the solution of Eq. (3.1). The optimization problem defined by Eqs. (3.2) and (3.3) can be solved using any minimization algorithm. Since function Γ is strictly convex, the Newton iterative method can be applied to the increasing function $\boldsymbol{\lambda} \mapsto \nabla \Gamma(\boldsymbol{\lambda})$ for searching $\boldsymbol{\lambda}^{\text{sol}}$ such that $\nabla \Gamma(\boldsymbol{\lambda}^{\text{sol}}) = \mathbf{0}$. This iterative method is not unconditionally convergent. Consequently, an under-relaxation is introduced and the iterative algorithm is written as

$$\boldsymbol{\lambda}^{i+1} = \boldsymbol{\lambda}^i - \alpha [H(\boldsymbol{\lambda}^i)]^{-1} \nabla \Gamma(\boldsymbol{\lambda}^i), \quad (3.7)$$

in which α belongs to $]0, 1]$ in order to ensure the convergence towards the solution $\boldsymbol{\lambda}^{\text{sol}}$. For $\alpha = 1$ there is no under-relaxation and it is under-relaxed for $0 < \alpha < 1$. The Newton iterative

method which is proposed will not require the calculation of the normalization constant $c_0(\boldsymbol{\lambda})$ of the pdf in the $\Gamma(\boldsymbol{\lambda})$ function defined by Eq. (3.3) (such a calculation would constitute a difficult problem in high dimension). In fact, as we will present in Section 3.3, an algorithm belonging to the Markov Chain Monte Carlo (MCMC) class will be used to calculate the mathematical expectations in Eqs. (3.5) and (3.6) which will not require the calculation of the normalization constant $c_0(\boldsymbol{\lambda})$ of the pdf. At each iteration i , the error is defined by

$$\text{error}(i) = \frac{\|\mathbf{f} - E\{\mathbf{g}(\mathbf{A}_{\boldsymbol{\lambda}^i})\}\|}{\|\mathbf{f}\|} = \frac{\|\nabla\Gamma(\boldsymbol{\lambda}^i)\|}{\|\mathbf{f}\|}, \quad (3.8)$$

and is calculated in order to control the convergence of the algorithm. The performance of this algorithm depends on the choice of the initial condition for which details are given in Section 3.3.2.

3.2. Gaussian case. In this section, we consider the particular case for which the available information leads us to a second-order centered random vector \mathbf{A} which is Gaussian. Note that if \mathbf{A} was not centered, with a given mean value \mathbf{m} , we could only consider the associated centered random vector $\mathbf{A} - \mathbf{m}$. This case is interesting to calculate the covariance matrix of the Gaussian centered random vector \mathbf{A} when the μ constraints are defined by μ quadratic forms on \mathbb{R}^N . Since these constraints must be algebraically independent, it is assumed that $1 < \mu \leq N(N+1)/2$. It should be noted that the case $\mu = 1$ is trivial. The available information is then defined by the following constraints,

$$E\{\mathbf{A}\} = \mathbf{0}, \quad (3.9)$$

$$E\left\{\frac{1}{2}\langle\mathbf{A}, [K_j]\mathbf{A}\rangle\right\} = l_j \quad , \quad j = 1, \dots, \mu, \quad (3.10)$$

in which $\{[K_j]\}_{j=1, \dots, \mu}$ are μ symmetric ($N \times N$) real matrices which are assumed to be algebraically independent. Let $\boldsymbol{\lambda}_m$ be the Lagrange parameter relative to the constraint defined by Eq. (3.9), $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_\mu)$ be the Lagrange parameter relative to the constraint defined by Eq. (3.10). In this case, the MaxEnt pdf is written as

$$p_{\mathbf{A}(\boldsymbol{\lambda}_m, \boldsymbol{\lambda})}(\mathbf{a}) = \frac{\exp(-\langle\boldsymbol{\lambda}_m, \mathbf{a}\rangle - \sum_{j=1}^{\mu} \frac{1}{2} \lambda_j \langle\mathbf{a}, [K_j]\mathbf{a}\rangle)}{\int_{\mathbb{R}^N} \exp(-\langle\boldsymbol{\lambda}_m, \mathbf{a}\rangle - \sum_{j=1}^{\mu} \frac{1}{2} \lambda_j \langle\mathbf{a}, [K_j]\mathbf{a}\rangle) d\mathbf{a}}, \quad (3.11)$$

Using Eq. (3.9), it can be shown that $\boldsymbol{\lambda}_m = \mathbf{0}$. Below, $\mathbf{A}_{(0, \boldsymbol{\lambda})}$ is rewritten as $\mathbf{A}_{\boldsymbol{\lambda}}$. We then have

$$p_{\mathbf{A}_{\boldsymbol{\lambda}}}(\mathbf{a}) = \frac{\exp(-\sum_{j=1}^{\mu} \frac{1}{2} \lambda_j \langle\mathbf{a}, [K_j]\mathbf{a}\rangle)}{\int_{\mathbb{R}^N} \exp(-\sum_{j=1}^{\mu} \frac{1}{2} \lambda_j \langle\mathbf{a}, [K_j]\mathbf{a}\rangle) d\mathbf{a}}. \quad (3.12)$$

Let $[\mathcal{K}_{\boldsymbol{\lambda}}] = \sum_{j=1}^{\mu} \lambda_j [K_j]$ which is assumed to be positive definite for all $\boldsymbol{\lambda}$ in \mathcal{L}_{μ} . In this case, Eq. (3.12) can be rewritten as

$$p_{\mathbf{A}_{\boldsymbol{\lambda}}}(\mathbf{a}; \boldsymbol{\lambda}) = c_0(\boldsymbol{\lambda}) \exp\left(-\frac{1}{2} \langle\mathbf{a}, [\mathcal{K}_{\boldsymbol{\lambda}}]\mathbf{a}\rangle\right), \quad (3.13)$$

in which $c_0(\boldsymbol{\lambda})$ is the normalization constant such that

$$c_0(\boldsymbol{\lambda}) = \left\{ \int_{\mathbb{R}^N} \exp\left(-\frac{1}{2}\langle \mathbf{a}, [\mathcal{K}_\lambda] \mathbf{a} \rangle\right) d\mathbf{a} \right\}^{-1} = (2\pi)^{-\frac{N}{2}} \sqrt{\det[\mathcal{K}_\lambda]}. \quad (3.14)$$

Equations (3.13) and (3.14) show that \mathbf{A}_λ is effectively a real centered Gaussian random vector for which the covariance matrix is $[\mathcal{K}(\boldsymbol{\lambda})]^{-1}$. For this Gaussian case, function $\Gamma(\boldsymbol{\lambda})$, gradient vector $\nabla\Gamma(\boldsymbol{\lambda})$ and Hessian matrix $[H(\boldsymbol{\lambda})]$ can explicitly be derived. Using Eqs. (3.3) and (3.14), we then have

$$\Gamma(\boldsymbol{\lambda}) = \langle \boldsymbol{\lambda}, \mathbf{l} \rangle + \frac{N}{2} \log(2\pi) - \frac{1}{2} \log(\det([\mathcal{K}_\lambda])). \quad (3.15)$$

in which $\mathbf{l} = (l_1, \dots, l_\mu)$. The gradient vector is then written as

$$(\nabla\Gamma(\boldsymbol{\lambda}))_i = l_i - \frac{1}{2} \text{tr}([\mathcal{K}_\lambda]^{-1} [K_i]), \quad (3.16)$$

in which tr is the trace for matrices. The Hessian matrix is written as

$$[H(\boldsymbol{\lambda})]_{ij} = \frac{1}{2} \text{tr}([\mathcal{K}_\lambda]^{-1} [K_i] [\mathcal{K}_\lambda]^{-1} [K_j]). \quad (3.17)$$

Finally, Eq. (3.7) allows $\boldsymbol{\lambda}^{\text{sol}}$ to be calculated iteratively.

3.3. General non-Gaussian case. For the Gaussian case presented in the previous section, the integrals in high dimension involved in Eqs. (3.3), (3.5) and (3.6) have explicitly been calculated which facilitates the identification of vector $\boldsymbol{\lambda}^{\text{sol}}$ of the Lagrange multipliers. In general, for the non-Gaussian case, these integrals cannot explicitly be calculated. Due to the high dimension, these integrals can not be discretized in \mathbb{R}^N such as proposed in [34] in the context of the MaxEnt methodology. In the same context, in [6, 9], the authors construct a set of linear equation using an integrating by parts of Eq. (2.2) and, in [2], the moment problem is solved in approximating the MaxEnt distribution on an adapted family of basis functions (these two methods are well adapted for $N = 1$).

In this paper, the integrals involved in Eqs. (3.3), (3.5) and (3.6) are estimated using the Monte Carlo simulation method [22] for which independent realizations of the random vector \mathbf{A}_λ are generated using an algorithm belonging to the MCMC class which is adapted to the high dimension, as proposed in [29]. In this Section, we first introduce a generator of independent realizations of a random vector in high dimension for which the pdf is constructed using the MaxEnt principle. Then an algorithm for the calculation of vector $\boldsymbol{\lambda}^{\text{sol}}$ of the Lagrange multipliers is presented.

3.3.1. Generator of independent realizations. The objective of this section is to provide a generator of independent realizations of the random variable \mathbf{A}_λ for all $\boldsymbol{\lambda}$ fixed in \mathcal{L}_μ . A generator of independent realizations for MaxEnt distribution has been proposed in [29] in the class of the MCMC algorithms. The methodology introduced consists in constructing the pdf of random vector \mathbf{A}_λ as the density of the invariant measure, $p_{\mathbf{A}_\lambda}(\mathbf{a})d\mathbf{a}$, associated with the stationary solution of a second-order nonlinear ISDE. The advantages of this generator

compared to the other MCMC generators such as the Metropolis-Hastings algorithm (see [12]) are: (1) The mathematical results concerning the existence and the uniqueness of an invariant measure can be used, (2) a damping matrix can be introduced in order to rapidly reach the invariant measure and (3) there is no need to introduce a proposal distribution which can induce difficulties in high dimension. In this paper, two modifications are introduced compared to the methodology introduced in [29]. The first one concerns the introduction of a damping matrix instead of a homogeneous damping parameter and therefore, which allows the convergence rate to be better controlled for all the components of the random vector. The second one concerns an optimal choice of the probability distribution of the initial condition of the ISDE which has to be solved for each given value of the Lagrange multiplier during the iterative algorithm.

i - Construction of the ISDE

Let $\mathbf{u} \mapsto \Phi(\mathbf{u}, \boldsymbol{\lambda})$ be a potential function defined by

$$\Phi(\mathbf{u}, \boldsymbol{\lambda}) = \langle \boldsymbol{\lambda}, \mathbf{g}(\mathbf{u}) \rangle. \quad (3.18)$$

We then introduce the stochastic process $\{(\mathbf{U}(r), \mathbf{V}(r)), r \geq 0\}$ with values in $\mathbb{R}^N \times \mathbb{R}^N$ satisfying, for all $r \geq 0$, the following ISDE (see [13], [27])

$$\begin{aligned} d\mathbf{U}(r) &= \mathbf{V}(r) dr \\ d\mathbf{V}(r) &= -\nabla_{\mathbf{u}}\Phi(\mathbf{U}(r), \boldsymbol{\lambda}) dr - \frac{1}{2}[D_{\boldsymbol{\lambda}}]\mathbf{V}(r) dr + [S_{\boldsymbol{\lambda}}] d\mathbf{W}(r), \end{aligned} \quad (3.19)$$

with the initial conditions

$$\mathbf{U}(0) = \mathbf{U}_0 \quad , \quad \mathbf{V}(0) = \mathbf{V}_0 \quad a.s. \quad (3.20)$$

In Eq. (3.19), $\nabla_{\mathbf{u}}\Phi(\mathbf{u}, \boldsymbol{\lambda})$ is the gradient of function $\Phi(\mathbf{u}, \boldsymbol{\lambda})$ with respect to \mathbf{u} , such that

$$\nabla_{\mathbf{u}}\Phi(\mathbf{u}, \boldsymbol{\lambda}) = [\nabla_{\mathbf{u}}\mathbf{g}(\mathbf{u})]\boldsymbol{\lambda}, \quad (3.21)$$

in which the $(N \times \mu)$ matrix $[\nabla_{\mathbf{u}}\mathbf{g}(\mathbf{u})]$ is the gradient of function $\mathbf{g}(\mathbf{u})$ with respect to \mathbf{u} . The matrix $[D_{\boldsymbol{\lambda}}]$ is a symmetric positive-definite damping matrix, the lower triangular matrix $[S_{\boldsymbol{\lambda}}]$ is such that $[D_{\boldsymbol{\lambda}}] = [S_{\boldsymbol{\lambda}}][S_{\boldsymbol{\lambda}}]^T$ and $\mathbf{W} = (W_1, \dots, W_N)$ is the normalized Wiener stochastic process indexed by \mathbb{R}^+ . The random initial condition $(\mathbf{U}_0, \mathbf{V}_0)$ is a second-order random variable independent of the Wiener stochastic process $\{\mathbf{W}(r), r \geq 0\}$. The probability distribution of the initial condition will be discussed in the next section. Then it can be proven (see [27, 29]) that, if $\mathbf{u} \mapsto \Phi(\mathbf{u}, \boldsymbol{\lambda})$ is continuous on \mathbb{R}^N , if $\mathbf{u} \mapsto \|\nabla_{\mathbf{u}}\Phi(\mathbf{u}, \boldsymbol{\lambda})\|$ is locally bounded on \mathbb{R}^N (i.e. is bounded on all compact set in \mathbb{R}^N), and if

$$\inf_{\|\mathbf{u}\| > R} \Phi(\mathbf{u}, \boldsymbol{\lambda}) \rightarrow +\infty \quad \text{if } R \rightarrow +\infty, \quad (3.22)$$

$$\inf_{\mathbf{u} \in \mathbb{R}^N} \Phi(\mathbf{u}, \boldsymbol{\lambda}) = \Phi_{\min} \quad \text{with } \Phi_{\min} \in \mathbb{R}, \quad (3.23)$$

$$\int_{\mathbb{R}^N} \|\nabla_{\mathbf{u}} \Phi(\mathbf{u}, \boldsymbol{\lambda})\| p_{\mathbf{A}_{\boldsymbol{\lambda}}}(\mathbf{u}) d\mathbf{u} < +\infty, \quad (3.24)$$

then, the ISDE defined by Eqs. (3.19) and (3.20) admits an invariant measure defined by the pdf $\rho_{\boldsymbol{\lambda}}(\mathbf{u}, \mathbf{v})$ with respect to $d\mathbf{u} d\mathbf{v}$, which is the unique solution of a steady-state Fokker-Planck equation and which is written as

$$\rho_{\boldsymbol{\lambda}}(\mathbf{u}, \mathbf{v}) = c_0(\boldsymbol{\lambda}) \exp(-\Phi(\mathbf{u}, \boldsymbol{\lambda})) \times (2\pi)^{-N/2} \exp\left(-\frac{1}{2}\|\mathbf{v}\|^2\right), \quad (3.25)$$

which shows that

$$p_{\mathbf{A}_{\boldsymbol{\lambda}}}(\mathbf{a}) = \int_{\mathbb{R}^N} \rho_{\boldsymbol{\lambda}}(\mathbf{a}, \mathbf{v}) d\mathbf{v}. \quad (3.26)$$

It can then be deduced that, for $r \rightarrow +\infty$, the stochastic process $\{(\mathbf{U}(r), \mathbf{V}(r)), r \geq 0\}$ tends to a stationary stochastic process in probability distribution, for which the one-order marginal probability distribution is $\rho_{\boldsymbol{\lambda}}(\mathbf{u}, \mathbf{v}) d\mathbf{u} d\mathbf{v}$. It can then be concluded that, for $r \rightarrow +\infty$, the stochastic process $\{\mathbf{U}(r), r \geq 0\}$ tends to a stationary stochastic process in probability distribution, for which the one-order marginal probability distribution is $p_{\mathbf{A}_{\boldsymbol{\lambda}}}(\mathbf{a}) d\mathbf{a}$ and we can briefly write,

$$\lim_{r \rightarrow +\infty} \mathbf{U}(r) = \mathbf{A}_{\boldsymbol{\lambda}} \quad \text{in probability distribution.} \quad (3.27)$$

In addition, for $r \rightarrow +\infty$, (1) $\mathbf{U}(r)$ and $\mathbf{V}(r)$ tend to independent random variables, (2) $\mathbf{U}(r)$ and $\mathbf{V}(r)$ tend to be independent of the random initial conditions, (3) $\mathbf{V}(r)$ tends to a centered Gaussian random vector with covariance matrix equal to the identity matrix. Therefore, using an independent realization of the Wiener stochastic process \mathbf{W} and an independent realization of the initial condition $(\mathbf{U}_0, \mathbf{V}_0)$, an independent realization of the random vector $\mathbf{A}_{\boldsymbol{\lambda}}$ can be constructed as the solution of the ISDE defined by Eqs. (3.19) and (3.20), for r sufficiently large. The value r_0 of r for which the invariant measure is approximatively reached depends on the choice of the damping matrix $[D_{\boldsymbol{\lambda}}]$ and on the probability distribution of the random initial conditions. The damping induced by the matrix $[D_{\boldsymbol{\lambda}}]$ has to be sufficiently large in order to rapidly kill the transient response but a too large damping introduces increasing errors in the numerical integration of the ISDE. Concerning the initial conditions, the more the probability distribution of the initial conditions is close to the invariant measure, the shorter is the transient response. Particularly, if the probability distribution $p_{\mathbf{U}_0, \mathbf{V}_0}(\mathbf{u}, \mathbf{v}) d\mathbf{u} d\mathbf{v}$ is equal to the invariant measure $\rho_{\boldsymbol{\lambda}}(\mathbf{u}, \mathbf{v}) d\mathbf{u} d\mathbf{v}$, then for all fixed $r > 0$, the probability distribution of the random vector $(\mathbf{U}(r), \mathbf{V}(r))$ is this invariant measure. The choice of the pdf $p_{\mathbf{U}_0, \mathbf{V}_0}$, which allows the invariant measure to be rapidly reached, is discussed in Section 3.3.2.

ii - Discretization of the ISDE

Several numerical integration schemes of an ISDE have been proposed in the literature. They are classically divided into three types: Explicit schemes, implicit schemes and semi-implicit schemes. A review and comparisons of the main integration schemes can be found in [4].

For the second-order ISDE under consideration, the choice of the integration scheme depends on the regularity of the potential function $\Phi(\mathbf{u}, \boldsymbol{\lambda})$ involved in the ISDE and the required precision. The advantage of implicit schemes is their unconditional stability with respect to the integration step size. Nevertheless, such an implicit scheme is time consuming because it requires to solve a nonlinear equation at each integration step. The explicit schemes only require the evaluation of the potential $\Phi(\mathbf{u}, \boldsymbol{\lambda})$ at each integration step but the stability of such schemes depends on the integration step size. If function $\mathbf{u} \mapsto \nabla_{\mathbf{u}}\Phi(\mathbf{u}, \boldsymbol{\lambda})$ is very "stiff" (presence of very large eigenvalues for the linearized second-order differential equation), then a small integration step size is required to guarantee the stability of the integration scheme. In [29], a semi-implicit integration scheme is proposed in order to avoid the resolution of an algebraic nonlinear equation at each step size while allowing a significant increase in the time step compared to a purely explicit scheme. The semi-implicit integration scheme, used for the application presented in Section 4.1, is summarized below.

A semi-implicit scheme can easily be used when the potential can be written as the sum of a positive-definite quadratic form $\mathcal{Q}(\mathbf{u})$ in \mathbf{u} with a nonlinear function of \mathbf{u} . In such a case, the gradient with respect to \mathbf{u} of the potential exhibits a linear part $[\mathcal{Q}']\mathbf{u}$ in which the Hessian $[\mathcal{Q}']$ is a positive-definite symmetric real matrix independent of \mathbf{u} . An implicit scheme can then be used for the linear part while the nonlinear part is considered with an explicit scheme. As a consequence, the equation which has to be solved at each integration step is an algebraic linear equation and therefore, its resolution is not time consuming.

Let us then assume that, for all $\boldsymbol{\lambda} \in \mathcal{L}_\mu$, the potential function can be written as

$$\Phi(\mathbf{u}, \boldsymbol{\lambda}) = \frac{1}{2}\langle \mathbf{u}, [\mathcal{K}_{\lambda_L}]\mathbf{u} \rangle + \Phi_{\text{NL}}(\mathbf{u}, \boldsymbol{\lambda}_{\text{NL}}), \quad (3.28)$$

for which the following decomposition $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_L, \boldsymbol{\lambda}_{\text{NL}}) \in \mathcal{L}_\mu$ of the Lagrange multipliers has been introduced and where $[\mathcal{K}_{\lambda_L}]$ is a positive-definite symmetric $(N \times N)$ real matrix depending on $\boldsymbol{\lambda}_L$. Therefore, the gradient is written as

$$\nabla_{\mathbf{u}}\Phi(\mathbf{u}, \boldsymbol{\lambda}) = [\mathcal{K}_{\lambda_L}]\mathbf{u} + \nabla_{\mathbf{u}}\Phi_{\text{NL}}(\mathbf{u}, \boldsymbol{\lambda}_{\text{NL}}). \quad (3.29)$$

Let Δr_λ be the integration step size and let $\{r_k = (k-1)\Delta r_\lambda, k = 1, \dots, M\}$ be the sampling points of the interval $[0, (M-1)\Delta r_\lambda]$, M being a positive integer. The semi-implicit integration scheme of the ISDE defined by Eq. (3.19) is then written, for $k = 1, \dots, M-1$, as

$$\begin{aligned} [\mathcal{E}_\lambda]\mathbf{V}^{k+1} &= [\mathcal{B}_\lambda]\mathbf{V}^k - \Delta r_\lambda[\mathcal{K}_{\lambda_L}]\mathbf{U}^k + \Delta r_\lambda\mathbf{L}_{\text{NL}}^k + [\mathcal{S}_\lambda]\Delta\mathbf{W}^{k+1} \\ \mathbf{U}^{k+1} &= \mathbf{U}^k + \frac{1}{2}\Delta r_\lambda(\mathbf{V}^{k+1} + \mathbf{V}^k), \end{aligned} \quad (3.30)$$

in which $[\mathcal{E}_\lambda] = [I_N] + \frac{1}{4}\Delta r_\lambda[D_\lambda] + \frac{1}{4}\Delta r_\lambda^2[\mathcal{K}_{\lambda_L}]$ and $[\mathcal{B}_\lambda] = [I_N] - \frac{1}{4}\Delta r_\lambda[D_\lambda] - \frac{1}{4}\Delta r_\lambda^2[\mathcal{K}_{\lambda_L}]$ where $[I_N]$ is the $(N \times N)$ identity matrix. The vector $\Delta\mathbf{W}^{k+1}$ is a second-order Gaussian centered random vector with covariance matrix equal to $\Delta r_\lambda[I_N]$ and the random vectors $\Delta\mathbf{W}^1, \dots, \Delta\mathbf{W}^M$ are mutually independent. The vector \mathbf{L}_{NL}^k is defined by $\mathbf{L}_{\text{NL}}^k = -\{\nabla_{\mathbf{u}}\Phi_{\text{NL}}(\mathbf{u}, \boldsymbol{\lambda}_{\text{NL}})\}_{\mathbf{u}=\mathbf{U}^k}$. If the gradient function $\nabla_{\mathbf{u}}\Phi_{\text{NL}}(\mathbf{u}, \boldsymbol{\lambda}_{\text{NL}})$ cannot explicitly be calculated, then \mathbf{L}_{NL}^k is calculated using a finite difference approximation of $\nabla_{\mathbf{u}}\Phi_{\text{NL}}(\mathbf{u}, \boldsymbol{\lambda}_{\text{NL}})$.

The initial conditions are $\mathbf{U}^1 = \mathbf{U}_0$ and $\mathbf{V}^1 = \mathbf{V}_0$.

iii - Estimation of the mathematical expectations

The integrals in high dimension involved in Eqs. (3.4), (3.5) and (3.6) are estimated using the Monte Carlo method and the generator of independent realizations previously defined. For $\ell = 1, \dots, n_s$, using n_s independent realizations $\{\Delta \mathbf{W}^{k+1, \ell}, k = 1, \dots, M-1\}$ of the family of random vectors $\{\Delta \mathbf{W}^{k+1}, k = 1, \dots, M-1\}$ and n_s independent realizations $(\mathbf{U}_0^\ell, \mathbf{V}_0^\ell)$ of the random initial conditions $(\mathbf{U}_0, \mathbf{V}_0)$ (which are also independent of $\{\Delta \mathbf{W}^{k+1}, k = 1, \dots, M-1\}$), then the n_s independent realizations $\mathbf{U}^{M, \ell}$ of the vector random \mathbf{U}^M are generated by solving n_s times, for $k = 1, \dots, M-1$, the following equations

$$\begin{aligned} [\mathcal{E}_\lambda] \mathbf{V}^{k+1, \ell} &= [\mathcal{B}_\lambda] \mathbf{V}^{k, \ell} - \Delta r_\lambda [\mathcal{K}_{\lambda_L}] \mathbf{U}^{k, \ell} + \Delta r_\lambda \mathbf{L}_{\text{NL}}^{k, \ell} + [S_\lambda] \Delta \mathbf{W}^{k+1, \ell}, \\ \mathbf{U}^{k+1, \ell} &= \mathbf{U}^{k, \ell} + \frac{1}{2} \Delta r_\lambda (\mathbf{V}^{k+1, \ell} + \mathbf{V}^{k, \ell}), \\ \mathbf{U}^{1, \ell} &= \mathbf{U}_0^\ell, \quad \mathbf{V}^{1, \ell} = \mathbf{V}_0^\ell. \end{aligned} \quad (3.31)$$

Then, if M is sufficiently large, using Eq. (3.27), the n_s independent realizations of the random vector \mathbf{A}_λ are constructed such that $\mathbf{A}_\lambda^\ell \simeq \mathbf{U}^{M, \ell}$ for $\ell = 1, \dots, n_s$. Therefore the mean value $E\{\mathbf{g}(\mathbf{A}_\lambda)\}$ and the correlation matrix $E\{\mathbf{g}(\mathbf{A}_\lambda) \mathbf{g}(\mathbf{A}_\lambda)^T\}$ are estimated by

$$E\{\mathbf{g}(\mathbf{A}_\lambda)\} \simeq \frac{1}{n_s} \sum_{\ell=1}^{n_s} \mathbf{g}(\mathbf{A}_\lambda^\ell), \quad (3.32)$$

$$E\{\mathbf{g}(\mathbf{A}_\lambda) \mathbf{g}(\mathbf{A}_\lambda)^T\} \simeq \frac{1}{n_s} \sum_{\ell=1}^{n_s} \mathbf{g}(\mathbf{A}_\lambda^\ell) \mathbf{g}(\mathbf{A}_\lambda^\ell)^T. \quad (3.33)$$

As we have previously explained, we need not to calculate the normalization constant $c_0(\lambda)$ for the proposed iterative algorithm. It should be noted that quantities M and n_s can depend on the current value of λ but such a dependence has been removed to simplify the notation.

3.3.2. Implementation of the iterative algorithm for the calculation of the Lagrange multipliers. In this section, we present the implementation of the iterative algorithm presented in Section 3.1 for the calculation of vector λ^{sol} and also, the construction of n_s independent realizations of random vector \mathbf{A} .

(i) Concerning the initialization of the algorithm, an initial value λ_{init} of λ has to be provided in \mathcal{L}_μ . A pdf of the random vector $(\mathbf{U}_{0, \text{init}}, \mathbf{V}_{0, \text{init}})$ corresponding to the random initial condition $(\mathbf{U}_0, \mathbf{V}_0)$ for $\lambda = \lambda_{\text{init}}$ has to be provided too. As explained in Section 3.3.1-i, the pdf $p_{\mathbf{U}_0, \mathbf{V}_0}(\mathbf{u}_0, \mathbf{v}_0)$ of the random initial condition has to be as close as possible to the invariant pdf $(\mathbf{u}, \mathbf{v}) \mapsto \rho_\lambda(\mathbf{u}, \mathbf{v})$. Concerning \mathbf{V}_0 , since for all λ in \mathcal{L}_μ and for r sufficiently large, the random vector $\mathbf{V}(r)$ is Gaussian, centered, with covariance matrix equal to the identity matrix, the pdf of $\mathbf{V}_{0, \text{init}}$ is chosen as the normalized Gaussian distribution, $\mathbf{V}_{0, \text{init}} \sim \mathcal{N}(\mathbf{0}, [I_n])$. Concerning $\mathbf{U}_{0, \text{init}}$, if the potential function can be written as in Eq. (3.28) and if we set

$\boldsymbol{\lambda}_{\text{NL,init}} = \mathbf{0}$, then $\mathbf{A}_{\boldsymbol{\lambda}_{\text{init}}}$ is a Gaussian centered random vector for which the covariance matrix is $[\mathcal{K}_{\boldsymbol{\lambda}_{\text{L,init}}}]^{-1}$. Therefore the pdf of $\mathbf{U}_{0,\text{init}}$ is chosen such that $\mathbf{U}_{0,\text{init}} \sim \mathcal{N}(\mathbf{0}, [\mathcal{K}_{\boldsymbol{\lambda}_{\text{L,init}}}]^{-1})$, and the initial value $\boldsymbol{\lambda}_{\text{L,init}}$ can be identified using the methodology developed in Section 3.2 devoted to the Gaussian case. If the potential function cannot be written as in Eq. (3.28), then $\mathbf{U}_{0,\text{init}} = \mathbf{0}$ almost surely can be taken. At this stage, $\boldsymbol{\lambda}_{\text{L,init}}$ has been identified and $\boldsymbol{\lambda}_{\text{NL,init}}$ must be chosen such that $(\boldsymbol{\lambda}_{\text{L,init}}, \boldsymbol{\lambda}_{\text{NL,init}})$ is in \mathcal{L}_μ .

(ii) At each iteration of the algorithm, the calculation of $\boldsymbol{\lambda}^{i+1}$ given $\boldsymbol{\lambda}^i$ required the calculation of the gradient and the Hessian for $\boldsymbol{\lambda}^i$. These quantities are estimated solving the ISDE with the pdf of the random initial condition which is chosen as the pdf of the invariant measure constructed for $\boldsymbol{\lambda}^{i-1}$ for which independent realizations are known. The algorithm for the identification of the Lagrange multipliers is summarized in Algorithm 1.

Algorithm 1: Identification of the Lagrange multipliers

INITIALIZATION:

$i = 1$;
 $\boldsymbol{\lambda}^i = \boldsymbol{\lambda}_{\text{init}}$;
 $\text{error}(i) = +\infty$;
for $\ell = 1, \dots, n_s$ **do**
 $\mathbf{U}_0^\ell = \mathbf{U}_{0,\text{init}}^\ell$;
 Generate $\mathbf{V}_0^\ell \sim \mathcal{N}(\mathbf{0}, [I_n])$;

LOOP:

while $\text{error}(i) > \epsilon$ **do**
 for $\ell = 1, \dots, n_s$ **do**
 $\mathbf{U}^{1,\ell} \leftarrow \mathbf{U}_0^\ell$;
 $\mathbf{V}^{1,\ell} \leftarrow \mathbf{V}_0^\ell$;
 Solve the ISDE $\rightarrow (\mathbf{U}^{M,\ell}, \mathbf{V}^{M,\ell})$;
 $\mathbf{A}_{\boldsymbol{\lambda}^i}^\ell \leftarrow \mathbf{U}^{M,\ell}$;
 $\mathbf{U}_0^\ell \leftarrow \mathbf{U}^{M,\ell}$;
 $\mathbf{V}_0^\ell \leftarrow \mathbf{V}^{M,\ell}$;
 Estimate $E\{\mathbf{g}(\mathbf{A}_{\boldsymbol{\lambda}^i})\} \simeq (1/n_s) \sum_{\ell=1}^{n_s} \mathbf{g}(\mathbf{A}_{\boldsymbol{\lambda}^i}^\ell)$;
 Estimate $E\{\mathbf{g}(\mathbf{A}_{\boldsymbol{\lambda}^i})\mathbf{g}(\mathbf{A}_{\boldsymbol{\lambda}^i})^T\} \simeq (1/n_s) \sum_{\ell=1}^{n_s} \mathbf{g}(\mathbf{A}_{\boldsymbol{\lambda}^i}^\ell)\mathbf{g}(\mathbf{A}_{\boldsymbol{\lambda}^i}^\ell)^T$;
 Estimate $\nabla\Gamma(\boldsymbol{\lambda}^i) = \mathbf{f} - E\{\mathbf{g}(\mathbf{A}_{\boldsymbol{\lambda}^i})\}$;
 Estimate $[H(\boldsymbol{\lambda}^i)] = E\{\mathbf{g}(\mathbf{A}_{\boldsymbol{\lambda}^i})\mathbf{g}(\mathbf{A}_{\boldsymbol{\lambda}^i})^T\} - E\{\mathbf{g}(\mathbf{A}_{\boldsymbol{\lambda}^i})\}E\{\mathbf{g}(\mathbf{A}_{\boldsymbol{\lambda}^i})\}^T$;
 Estimate $\text{error}(i+1) = \|\nabla\Gamma(\boldsymbol{\lambda}^i)\| / \|\mathbf{f}\|$;
 $\boldsymbol{\lambda}^{i+1} = \boldsymbol{\lambda}^i - [H(\boldsymbol{\lambda}^i)]^{-1} \nabla\Gamma(\boldsymbol{\lambda}^i)$;
 $i \leftarrow i + 1$;

4. Application.

4.1. Gaussian case: Generation of physically consistent accelerograms. The application concerns the generation of realizations of a non-stationary Gaussian centered stochastic process in the framework of the generation of physically consistent accelerograms. The

available information is constituted of second-moments of the stochastic process and other constraints which guarantees that the generated trajectories are physically consistent.

4.1.1. Definition of the available information. A time sampling of the acceleration stochastic process is introduced yielding a time series $\{A_1, \dots, A_N\}$ for which the \mathbb{R}^N -valued random vector $\mathbf{A} = (A_1, \dots, A_N)$ is associated with. We then have to construct the probability distribution of random vector \mathbf{A} . As explained in Section 3.2, since random vector \mathbf{A} is centered, the lagrange multiplier relative to the mean vector is zero and therefore, the available information relative to the mean vector is useless. The available information for the random vector \mathbf{A} is then defined as follows,

$$E\{A_j^2\} = \sigma_j^2 < +\infty \quad , \quad \forall j = \{1, \dots, N\}, \quad (4.1)$$

$$E\left\{\left(\sum_{k=1}^N A_k\right)^2\right\} = 0, \quad (4.2)$$

$$E\left\{\left(\sum_{k=1}^N (N - k + 1) A_k\right)^2\right\} = 0, \quad (4.3)$$

$$E\left\{\left(\sum_{k=1}^N (N - k + 1)^2 A_k\right)^2\right\} = 0. \quad (4.4)$$

The N constraints defined by Eq. (4.1) are relative to the variance of each component A_j of random vector \mathbf{A} . The constraints defined by Eqs. (4.2) and (4.3) allow us to impose, for the generated trajectories and in the mean-square sense, a zero end-velocity $\mathbb{V}_N = \sum_{j=1}^N A_j$ and a zero end-displacement $\mathbb{D}_N = \sum_{j=1}^N (N - j + 1)A_j$. The constraint defined by Eq. (4.4) implies that the integral over the time observation of the displacement process associated with the acceleration process, is zero in the mean-square sense (in terms of earthquake engineering, this constraint allows the rock ground motions to be eliminated in the generated trajectories). Modeling this constraint and using Eqs. (4.2) and (4.3) yield the constraint defined by Eq. (4.4). For $j = \{1, \dots, N\}$, the matrices $[K_j]$ introduced in Eq. (3.10), relative to Eq. (4.1), are such that $[K_j]_{k\ell} = 2 \delta_{kj} \delta_{\ell j}$ in which δ_{kj} is the Kronecker symbol, and $l_j = \sigma_j^2$. The matrix $[K_{N+1}]$, relative to Eq. (4.2), is defined by $[K_{N+1}]_{k\ell} = 2$ and $l_{N+1} = 0$. The matrix $[K_{N+2}]$, relative to Eq. (4.3), is defined by $[K_{N+2}]_{k\ell} = 2(N - k + 1)(N - \ell + 1)$ and $l_{N+2} = 0$. The matrix $[K_{N+3}]$, relative to Eq. (4.4), is defined by $[K_{N+3}]_{k\ell} = 2(N - k + 1)^2(N - \ell + 1)^2$ and $l_{N+3} = 0$. Then, $\mu = N + 3$ and the matrix $[\mathcal{K}_\lambda]$ is defined by

$$[\mathcal{K}_\lambda] = \sum_{j=1}^N \{\lambda_j [K_j]\} + \lambda_{N+1} [K_{N+1}] + \lambda_{N+2} [K_{N+2}] + \lambda_{N+3} [K_{N+3}], \quad (4.5)$$

The admissible space \mathcal{L}_μ for the vector $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_\mu)$ is defined as $\mathcal{L}_\mu =]0, +\infty[^\mu$. For $p = 1, 2, 3$, the matrices $[K_{N+p}]$ can be rewritten as $[K_{N+p}] = 2\mathbf{z}_p \mathbf{z}_p^T$ in which the vectors

\mathbf{z}_p are such that $\{\mathbf{z}_p\}_k = (N - k + 1)^{p-1}$. It can be seen that $\langle \mathbf{a}, [\mathcal{K}_\lambda] \mathbf{a} \rangle = 2 \sum_{j=1}^N \lambda_j a_j^2 + 2 \sum_{p=1}^3 \lambda_{N+p} \langle \mathbf{a}, \mathbf{z}_p \rangle^2$. For all $\boldsymbol{\lambda}$ in \mathcal{L}_μ (thus $\|\boldsymbol{\lambda}\| \neq 0$) and for all vector \mathbf{a} in \mathbb{R}^N such that $\|\mathbf{a}\| \neq 0$, we have $\sum_{j=1}^N \lambda_j a_j^2 > 0$ and $\sum_{p=1}^3 \lambda_{N+p} \langle \mathbf{a}, \mathbf{z}_p \rangle^2 \geq 0$, which show that $\langle \mathbf{a}, [\mathcal{K}_\lambda] \mathbf{a} \rangle > 0$ and therefore, $[\mathcal{K}_\lambda]$ is positive definite.

4.1.2. Application. The accelerogram is sampled into $N = 1,600$ time steps. The final time is 20 s and the time step is $\Delta t = 0.0125$ s. The target standard deviation $\{\sigma_j, j = 1, \dots, N\}$ is defined by

$$\begin{aligned} \sigma_j &= 1.3985 (j\Delta t)^2 / 16 + 0.14 \quad , \quad j = 1, \dots, 320 & (4.6) \\ \sigma_j &= 1.5383 \quad , \quad j = 321, \dots, 1280 \\ \sigma_j &= 1.3985 \exp(-1.15 (j\Delta t) - 16) + 0.14 \quad , \quad j = 1281, \dots, 1600 \end{aligned}$$

This target standard deviation is plotted in Fig. 4.1. The methodology developed in Sec-

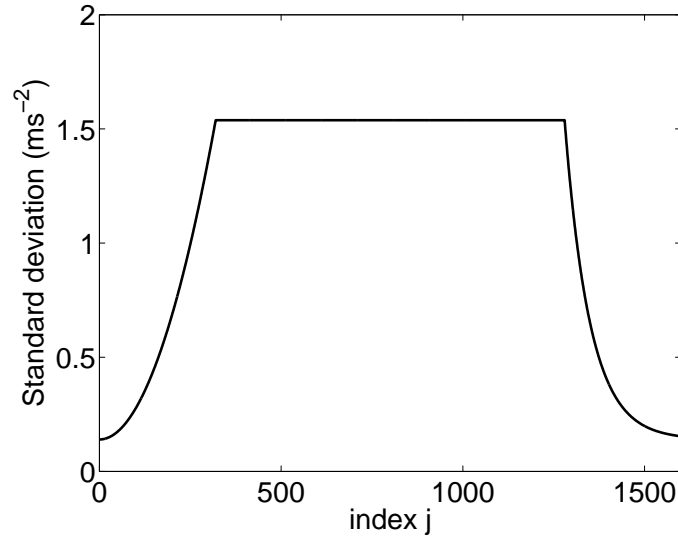


Figure 4.1. Graph of function $j \mapsto \sigma_j$.

tions 3.1 and 3.2 is applied here using 30 iterations. The error function $i \mapsto \text{error}(i)$ is plotted in Fig. 4.2. Figure 4.3 shows two independent realizations of the random vector $\mathbf{A}_{\boldsymbol{\lambda}_{\text{sol}}}$ which is generated using a classical generator for Gaussian random variable and which are representative of two independent realizations of the random accelerogram. The corresponding trajectories of the velocity times series \mathbb{V}_N and of the displacement times series \mathbb{D}_N result from two successive numerical integrations of each realization of the random accelerogram and are plotted in Figs. 4.4 and 4.5. As expected, it can be seen that the end velocity and the end displacements are equal to zero.

4.2. General case: Generation of physically consistent and spectrum-compatible accelerograms. With respect to the constraints defined in Section 4.1, an additional constraint is added in order to specify the mean Velocity Response Spectrum (VRS), see [7, 30]. The target spectrum is constructed from the European Code 8 (see [10]).

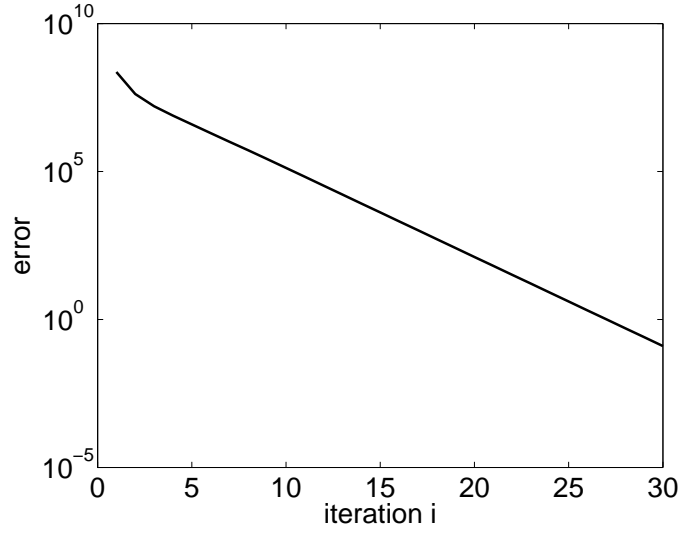


Figure 4.2. Graph $i \mapsto \text{error}(i)$ of the error function in log scale.

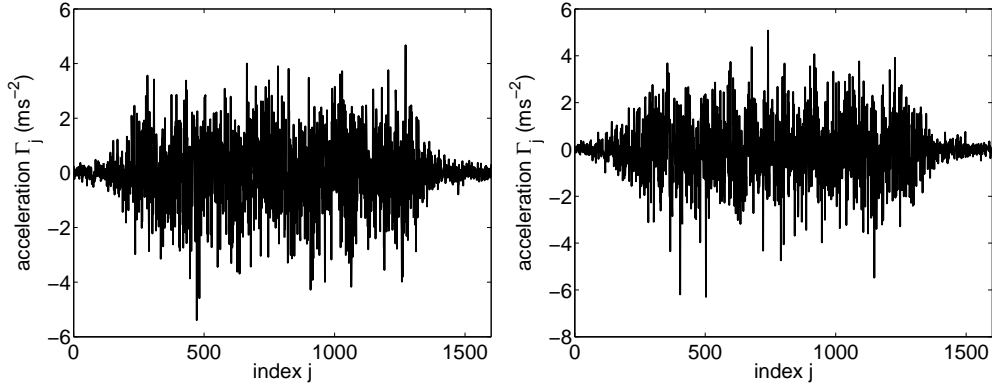


Figure 4.3. Two independent realizations of the random accelerogram.

4.2.1. Available information. The random VRS is represented by a random vector $\mathbf{S} = \mathbf{s}(\mathbf{A})$ with values in \mathbb{R}^κ , in which $\mathbf{a} \mapsto \mathbf{s}(\mathbf{a}) = (s_1(\mathbf{a}), \dots, s_\kappa(\mathbf{a}))$ is a nonlinear mapping from \mathbb{R}^N to \mathbb{R}^κ defined, for all k in $\{1, \dots, \kappa\}$, by

$$s_k(\mathbf{a}) = \omega_k \max\{|y_1^k(\mathbf{a})|, \dots, |y_N^k(\mathbf{a})|\} \quad , \quad y_j^k(\mathbf{a}) = \{[B^k]\mathbf{a}\}_j \quad , \quad (4.7)$$

in which ω_k is a given frequency and $[B^k]$ is a family of $(N \times N)$ matrices defined by

$$[B^k]_{ij} = -\frac{\Delta t}{\omega_k \sqrt{1 - \xi^2}} \exp(-(i - j) \xi \omega_k \Delta t) \sin((i - j) \omega_k \Delta t) \quad , \quad (4.8)$$

in which $0 < \xi < 1$ is the given damping ratio and where Δt is the time step relative to time series \mathbf{A} . Then the available information relative to random vector \mathbf{A} is defined by

$$E\{A_j^2\} = \sigma_j^2 < +\infty \quad , \quad \forall j \in \{1, \dots, N\} \quad , \quad (4.9)$$

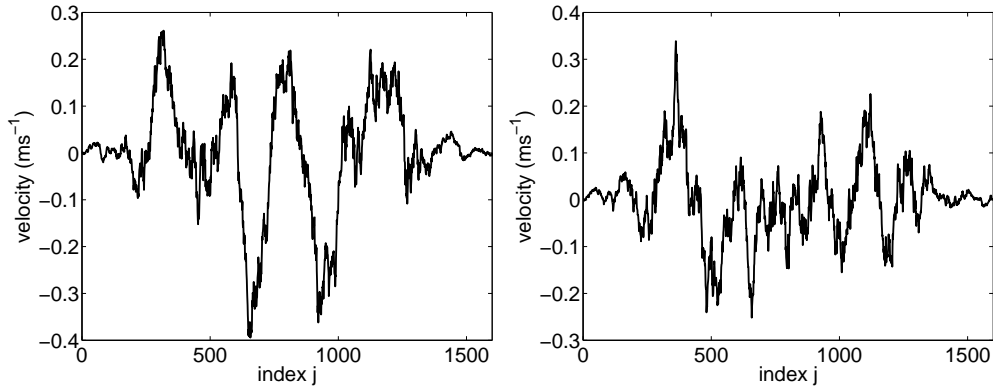


Figure 4.4. Two independent realizations of the random velocity.

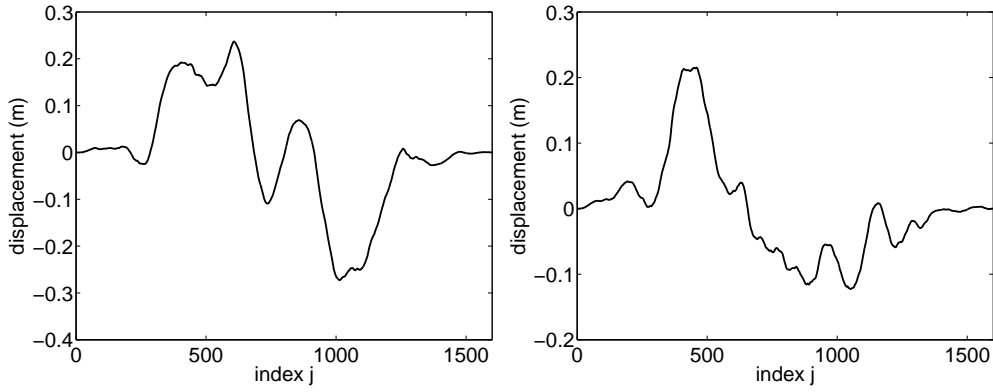


Figure 4.5. Two independent realizations of the random displacement.

$$E\left\{\left(\sum_{k=1}^N A_k\right)^2\right\} = 0, \quad (4.10)$$

$$E\left\{\left(\sum_{k=1}^N k A_k\right)^2\right\} = 0, \quad (4.11)$$

$$E\left\{\left(\sum_{k=1}^N k^2 A_k\right)^2\right\} = 0. \quad (4.12)$$

$$E\{s_k(\mathbf{A})\} = \underline{s}_k, \quad \forall k \in \{1, \dots, \kappa\}, \quad (4.13)$$

where $\underline{\mathbf{s}} = (\underline{s}_1, \dots, \underline{s}_\kappa)$ is the mean VRS which is chosen as the target. It should be noted that a constraint relative to the zero mean value should be added. Nevertheless, using Eq. (2.5) and remarking that the constraints defined by Eqs. (4.9) to (4.13) are even functions in \mathbf{A} , the

Lagrange multipliers relative to the mean value are zero and therefore this constraint is not taken into account. Therefore, $\mu = N + 3 + \kappa$ and the function $\mathbf{g}(\mathbf{u})$ with values in \mathbb{R}^μ is defined by $\mathbf{g}(\mathbf{u}) = (u_1^2, \dots, u_N^2, (\sum_{k=1}^N u_k)^2, (\sum_{k=1}^N k u_k)^2, (\sum_{k=1}^N k^2 u_k)^2, s_1(\mathbf{u}), \dots, s_\kappa(\mathbf{u}))$. For this application, the gradient vector $\nabla_{\mathbf{u}}\Phi(\mathbf{u}, \boldsymbol{\lambda})$ can be written as in Eq. (3.29), for which $[\mathcal{K}_{\boldsymbol{\lambda}_L}]$ is given by Eq. (4.5) in which $\boldsymbol{\lambda}$ is replaced by $\boldsymbol{\lambda}_L$ and for which the nonlinear part (induced by the constraint defined by Eq. (4.13)) is written as $\nabla_{\mathbf{u}}\Phi_{\text{NL}}(\mathbf{u}, \boldsymbol{\lambda}_{\text{NL}}) = [\nabla_{\mathbf{u}}\mathbf{g}_{\text{NL}}(\mathbf{u})]\boldsymbol{\lambda}_{\text{NL}}$ in which

$$[\nabla_{\mathbf{u}}\mathbf{g}_{\text{NL}}(\mathbf{u})]_{jk} = \omega_k [B^k]_{\text{imax}(k), j} \text{sgn}(y_{\text{imax}(k)}^k(\mathbf{u})), \quad (4.14)$$

in which $\text{sgn}(\alpha)$ is the sign of α , $\mathbf{y}^k(\mathbf{u})$ is defined in Eq. (4.7) and where imax is a function from $\{1, \dots, N\}$ to $\{1, \dots, N\}$ such that

$$\text{imax}(k) = \arg \max_{j \in \{1, \dots, N\}} |y_j^k(\mathbf{u})|. \quad (4.15)$$

For $\|\mathbf{u}\| \neq 0$, it should be noted that the derivative given by Eq. (4.14) holds if $\text{imax}(k)$ constructed using Eq. (4.15) is unique. This derivative is used for \mathbf{u} equal to $\mathbf{U}^{k, \ell}$ in which $\mathbf{U}^{k, \ell}$ is a given realization $\{\mathbf{U}^{k, \ell}, k = 1, \dots, M-1\}$ of the time series $\{\mathbf{U}^k, k = 1, \dots, M-1\}$. For a given realization ℓ , the probability that there exist two solutions for Eq. (4.15) is zero. For $\|\mathbf{u}\| = 0$, this derivative does not exist but, since the initial condition is such that $\|\mathbf{u}\| \neq 0$, the probability that $\|\mathbf{U}^{k, \ell}\|$ be equal to zero is zero.

4.2.2. Application. The sampling parameters and the standard deviation $\{\sigma_j, j = 1, \dots, N\}$ are the same as in Section 4.1.2. Concerning the VRS, $\kappa = 20$, $\xi = 0.05$ and the frequencies ω_k (in rad/s) are 1.04, 1.34, 1.73, 2.23, 2.86, 3.69, 4.74, 6.11, 7.86, 10.11, 13.01, 16.74, 21.53, 27.70, 35.64, 45.86, 59.00, 75.91, 97.67 and 125.66. The target mean VRS $\{\underline{s}_k, k = 1 \dots, \kappa\}$, which is constructed following the Eurocode 8, is plotted in Fig. 4.6. For the identification

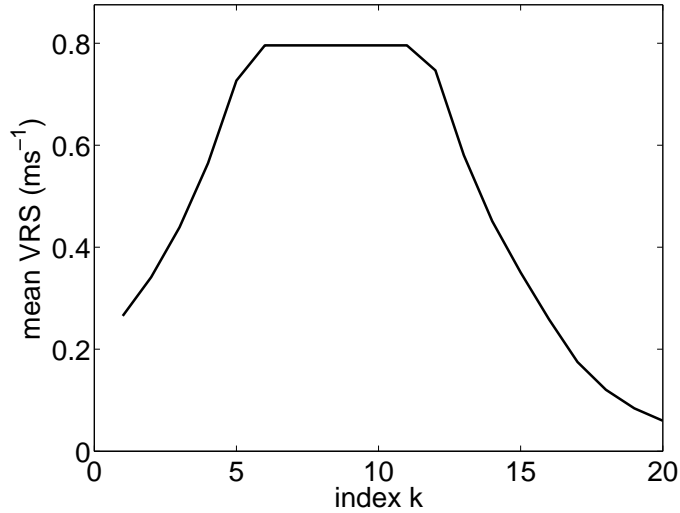


Figure 4.6. Mean VRS $k \mapsto \underline{s}_k$.

of the Lagrange multipliers, the algorithm developed in Section 3.3.2 and the semi-implicit

integration scheme are used. For the initial value $\boldsymbol{\lambda}_{\text{init}}$ of the Lagrange multipliers, the non-linear part $\boldsymbol{\lambda}_{NL,\text{init}}$ is zero while the linear part $\boldsymbol{\lambda}_{L,\text{init}}$ is the result of the identification carried out in Section 4.1.2. As explained in Section 3.3.2, the initial value $\boldsymbol{U}_{0,\text{init}}$ of the random initial condition is generated such that $\boldsymbol{U}_{0,\text{init}} \sim \mathcal{N}(\mathbf{0}, [\mathcal{K}_{\boldsymbol{\lambda}_{L,\text{init}}}]^{-1})$. At each iteration, the time step is $\Delta r_{\boldsymbol{\lambda}} = 2\pi/(\beta\sqrt{2\lambda_{\text{max}}})$, in which $\lambda_{\text{max}} = \max\{\{\boldsymbol{\lambda}_L\}_i, i = 1, \dots, N\}$ and $\beta = 80$. At each iteration, the damping matrix $[D_{\boldsymbol{\lambda}}]$ is a diagonal matrix such that $[D_{\boldsymbol{\lambda}}]_{ii} = 2\xi_{\text{ito}}\sqrt{2\{\boldsymbol{\lambda}_L\}_i}$, in which $\xi_{\text{ito}} = 0.7$. For the ISDE, the number of integration steps is $M = 600$. At each iteration, $n_s = 900$ Monte Carlo simulations are carried out. The under-relaxation parameter is $\alpha = 0.3$. The error function $i \mapsto \text{error}(i)$ for 30 iterations is plotted in Fig. 4.7. Figure 4.8

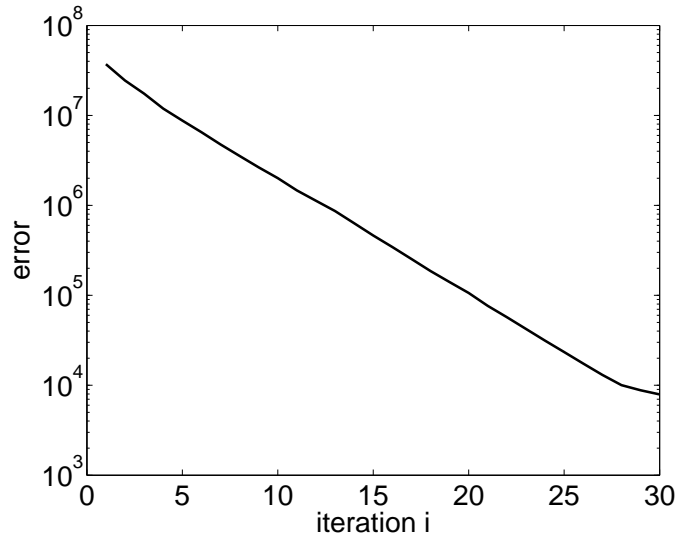


Figure 4.7. Graph $i \mapsto \text{error}(i)$ of the error function in log scale.

shows two independent realizations of the random vector $\mathbf{A}_{\boldsymbol{\lambda}_{\text{sol}}}$ which are representative of two independent realizations of the random accelerogram. The corresponding trajectories of the velocity times series \mathbb{V}_N and of the displacement times series \mathbb{D}_N result from two successive numerical integrations of each realization of the random accelerogram and are plotted in Figs. 4.9 and 4.10. As expected, it can be seen that the end-velocity and the end-displacements are equal to zero. Figure 4.11 displays a comparison of the mean variance with the target variance. Figure 4.12 shows a comparison of the mean VRS with the target mean VRS. It can be seen in Figures 4.8 to 4.12 that the results are very good. Figure 4.12 also shows the mean VRS estimated using realizations of the Gaussian accelerograms constructed in the Section 4.1. This figure shows the influence of the probability distribution on the random accelerogram and shows that the Gaussian modeling yields a completely different mean VRS compared with the non-Gaussian modeling for which the constraints relative to the mean VRS is taken into account. The confidence region of the random VRS with a probability level $P_c = 0.95$ is plotted in Fig. 4.13.

5. Conclusions. We have presented a new algorithm for the identification in high stochastic dimension of the Lagrange multipliers of a probability density function constructed using

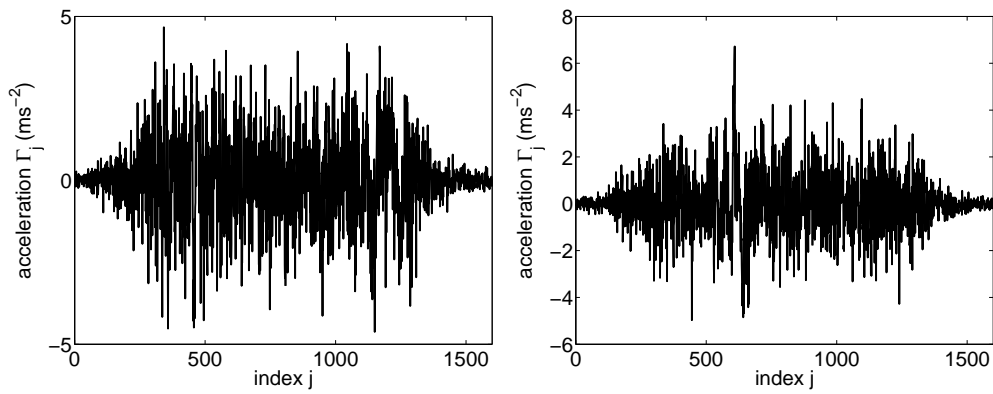


Figure 4.8. *Two independent realizations of the random accelerogram.*

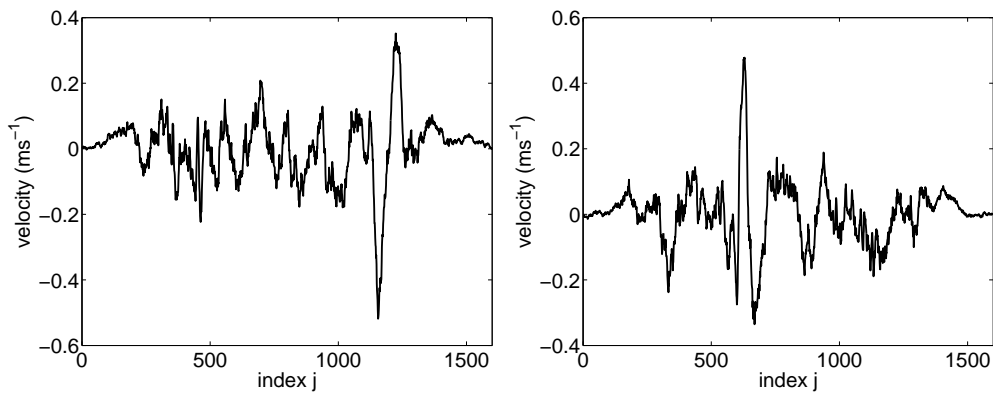


Figure 4.9. *Two independent realizations of the random velocity.*

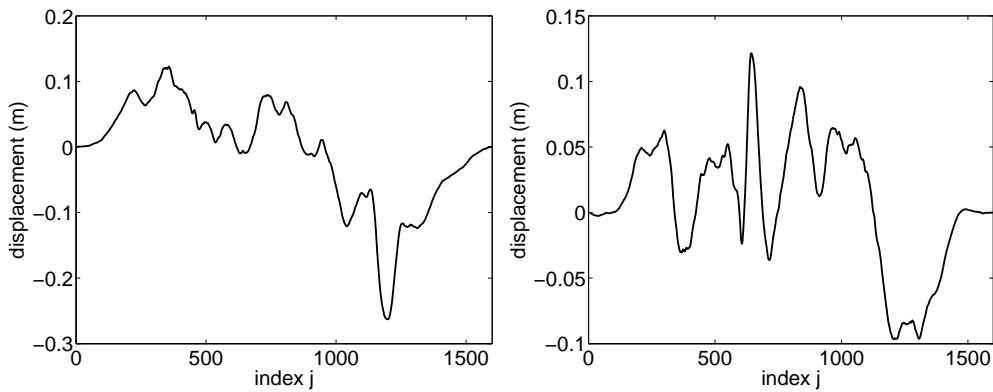


Figure 4.10. *Two independent realizations of the random displacement.*

the MaxEnt principle under constraints defined by the available information. The high dimensional integrals are evaluated by simulating independent realizations of an ISDE. The two applications show the efficiency of the proposed algorithm even in presence of strongly

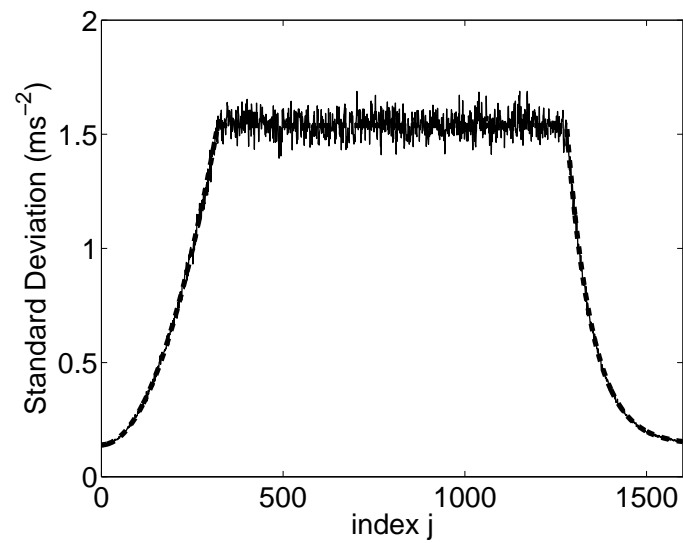


Figure 4.11. Variance: Target (thick dashed line) and estimation (thin solid line).

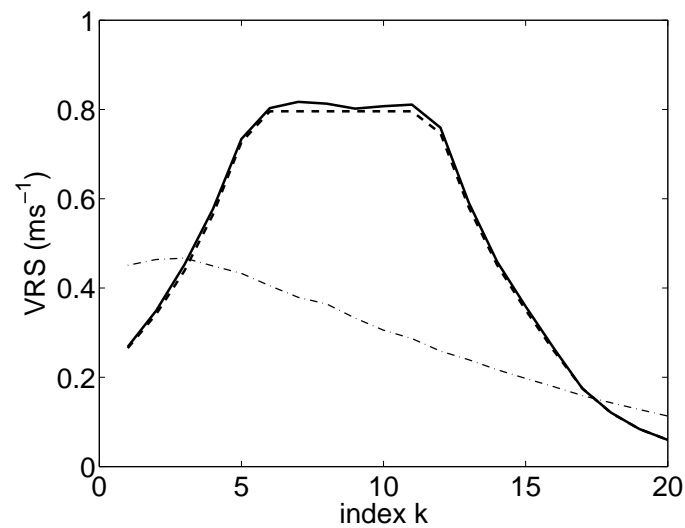


Figure 4.12. Mean VRS: Target (dashed line), estimation (solid line), estimation with a Gaussian distribution (mixed line).

nonlinear constraints in the available information.

REFERENCES

- [1] N. AGMON, Y. ALHASSID, R. D. LEVINE, An algorithm for finding the distribution of maximal entropy, *Journal of Computational Physics*, 30, 250-258, 1979.
- [2] A. BALESTRINO, A. CAITI, E. CRISOSTOMI, Efficient numerical approximations of maximum entropy estimates, *International Journal of Control*, 79(9), 1145-1155, 2006.

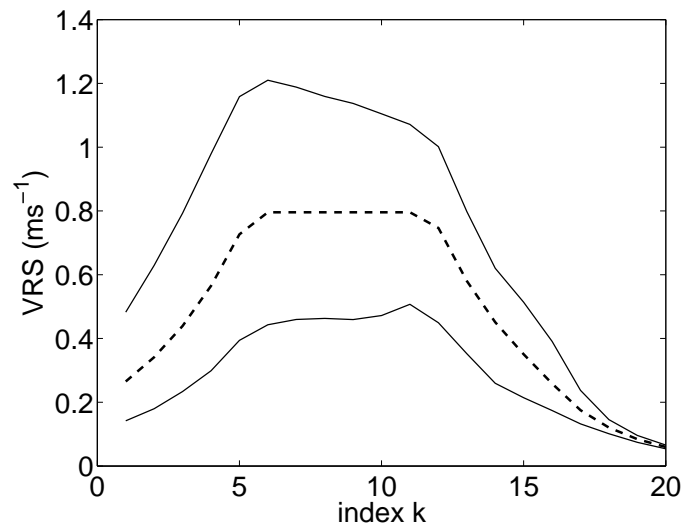


Figure 4.13. Random VRS: Confidence region of the random VRS (delimited by the upper and the lower solid lines) and mean value of the random VRS (dashed line).

- [3] A. L. BERGER, S. A. DELLAPIETRA, V. J. DELLAPIETRA, A maximum entropy approach to natural language processing, *Computational Linguistics*, 22(1), 39-71, 1996.
- [4] K. BURRAGE, Numerical methods for second-order stochastic differential equation, *SIAM Journal of Scientific Computing*, 29(1), 245-264, 2010.
- [5] K. D. CHAN, G. SUBBARAYAN, L. NGUYEN, Maximum-entropy principle for modeling damage and fracture in solder joints, *Journal of Electronic Materials*, 41(2), 398-411, 2012.
- [6] B. CHEN, J. HU, Y. ZHU, Computing maximum entropy densities: A hybrid approach, *Signal Processing: An International Journal*, 4(2), 114-122, 2010.
- [7] R. W. CLOUGH, J. PENZIEN, *Dynamics of Structures*, McGraw-Hill, New York, 1975.
- [8] R. DEWAR, Information theory explanation of the fluctuation theorem, maximum entropy production and self-organized criticality in non-equilibrium stationary states, *Journal of Physics A-Mathematical and General*, 36(3), 631-641, 2003.
- [9] D. ERDOGMUS, K. E. HILD II, Y. N. RAO, J. C. PRINCIPE, Minimax mutual information approach for independent components analysis, *Neural Computation*, 16(6), 1235-1252, 2004.
- [10] CEN, *Eurocode 8: Design of Structures for Earthquake Resistance-Part 1: General Rules, Seismic Actions and Rules for Buildings. EN 1998-1: 2003 European Committee for Standardization*, Brussels, 2003.
- [11] A. GOLAN, G. JUDGE, D. MILLER, *Maximum entropy econometrics: robust estimation with limited data*, Wiley, New York, 1996.
- [12] W. K. HASTINGS, Monte Carlo sampling methods using Markov Chains and their applications, *Biometrika*, 109, 57-97, 1970.
- [13] K. ITÔ, On stochastic differential equations, *Memoirs American Mathematical Society*, 4, 1951.
- [14] E. T. JAYNES, Information theory and statistical mechanics, *Physical Review*, 106(4), 620-630 and 108(2), 171-190, 1957.
- [15] J. N. KAPUR, H. K. KEVASAN, *Entropy Optimization Principles with Applications*, Academic Press, San Diego, 1992.
- [16] A. K. LIVESEY, J. SKILLING, Maximum-entropy theory, *Acta Crystallographica Section A*, 41, 113-122, 1985.
- [17] A. LYGRE, H. E. KROGSTAD, Maximum-entropy of the directional distribution of ocean wave *Journal of Physical Oceanography*, 16(12), 2052-2060, 1986.
- [18] R. MEAD, N. PAPANICOLAOU, Maximum entropy in the problem of moments, *Journal of Mathematical Physics*, 24:2404-2417., 1984.

-
- [19] D. ORMONEIT, H. WHITE, An efficient algorithm to compute maximum entropy densities, *Econometrics reviews*, 18(2), 127-140, 1999.
- [20] A. PAPOULIS, Maximum-entropy and spectral estimation - A Review, *IEEE Transaction on Acoustics Speech and Signal Processing*, 29(6), 1176-1186, 1981.
- [21] S. J. PHILLIPS, R. P. ANDERSON, R. E. SCHAPIRE, Maximum entropy modeling of species geographic distributions, *Ecological Modelling*, 190(3-4), 231-259, 2006.
- [22] R. Y. RUBINSTEIN, D. P. KROESE, *Simulation and the Monte Carlo Method*, Second Edition, John Wiley and Sons, New York, 2008.
- [23] C. E. SHANNON, A mathematical theory of communication, *Bell System Technology Journal*, 27, 379-423 and 623-659, 1948.
- [24] J. E. SHORE, R. W. JOHNSON, Axiomatic derivation of the principle of maximum-entropy and the principle of minimum cross-entropy, *IEEE Transactions on Information Theory*, 26(1), 26-37, 1980.
- [25] J. SKILLING, R. K. BRYAN, Maximum-entropy image-reconstruction - general algorithm. *Monthly Notices of the Royal Astronomical Society*, 211(1), 111-124, 1984.
- [26] K. SOBCZYK, J. TRABICKI, Maximum-entropy principle and nonlinear stochastic oscillators, *Physica A*, 193(3-4), 448-468, 1993.
- [27] C. SOIZE, *The Fokker-Planck Equation for Stochastic Dynamical Systems and its Explicit Steady State Solution*, World Scientific: Singapore, 1994.
- [28] C. SOIZE, Maximum entropy approach for modeling random uncertainties in transient elastodynamics, *Journal of the Acoustical Society of America*, 109(5), 1979-1996, 2001.
- [29] C. SOIZE, Construction of probability distributions in high dimension using the maximum entropy principle. Applications to stochastic processes, random fields and random matrices, *International Journal for Numerical Methods in Engineering*, 76(10), 1583-1611, 2008.
- [30] C. SOIZE, Information theory for generation of accelerograms associated with shock response spectra, *Computer-Aided Civil and Infrastructure Engineering*, 25, 334-347, 2010.
- [31] T. J. ULRYCH, T. N. BISHOP, Maximum entropy spectral analysis and autoregressive decomposition, *Review of Geophysics*, 13(1), 1975.
- [32] W. VONDERLINDEN, Maximum-entropy data-analysis, *Applied Physics A-Materials Science & Processing*, 60(2), 155-165, 1995.
- [33] X. WU, Calculation of maximum entropy densities with application to income distribution, *Journal of Econometrics*, 115, 347-354, 2003.
- [34] A. ZELLNER, R. A. HIGHFIELD, Calculation of maximum entropy distributions and approximation of marginal posterior distributions, *Journal of Econometrics*, 37, 195-209, 1988.