Approximation of probability density functions for SPDEs using truncated series expansions[0]

Giacomo Capodaglio[*]   Max Gunzburger[*]   Henry P. Wynn [†]

Abstract

The probability density function (PDF) of a random variable associated with the solution of a stochastic partial differential equation (SPDE) is approximated using a truncated series expansion. The SPDE is solved using two stochastic finite element (SFEM) methods, Monte Carlo sampling and the stochastic Galerkin method with global polynomials. The random variable is a functional of the solution of the SPDE, such as the average over the physical domain. The truncated series are obtained considering a finite number of terms in the Gram-Charlier or Edgeworth series expansions. These expansions approximate the PDF of a random variable in terms of another PDF, and involve coefficients that are functions of the known cumulants of the random variable. To the best of our knowledge, their use in the framework of SPDEs has not yet been explored.

1   Introduction

In the field of uncertainty quantification, the problem of describing the probability density function (PDF) of the output of a model, given a known distribution of the input is of major relevance [13]. For instance, for the solution of inverse problems involving partial differential equations (PDEs) with Bayesian inference [55, 20, 11], the PDF for the forward problem may be sought to get insights on the forward uncertainty propagation [38, 37]. The forward PDF is inferred from the histogram, which may be obtained according to two different strategies. The first is a direct approach, where Monte Carlo (MC) sampling is used to obtain realizations of the quantity of interest (QoI) by solving the forward problem for each sample. Alternatively, once an approximation of the stochastic solution of the forward problem is obtained, realizations may be obtained with MC sampling without solving the problem for each sample. This is the method used for instance in [38], where the stochastic solution of the forward problem is approximated with polynomial chaos, or in [37], where adaptive sparse grid collocation is employed. It is important to observe that despite providing a simple and easy approximation, histograms are not continuous, hence they cannot be differentiated if necessary. The differentiability of the estimator is an important feature in many applications [14, 27, 12, 52, 44]. In this work, we propose an alternative strategy to estimate the PDF of a quantity of interest, given a known distribution for the input data. The quantities of interest are associated with the solution of a stochastic partial differential equation (SPDE), which is approximated using a stochastic finite element method (SFEM). Once the QoI has been obtained from the SPDE solution, its PDF is estimated by means of a truncated Gram-Charlier (GC) or Edgeworth (ED) series expansion [35]. These series have been used in several fields such as chemistry [46, 22], finance [33, 50, 41], physics and astrophysics [23, 21, 45, 34, 8, 17], material science [51], oceanology [60], power systems engineering [24], and other branches of applied mathematics [47]. However, to the best of our knowledge, their use in the field of SPDEs has not yet been explored. The GC and ED expansions approximate a PDF adding successive corrections to a known PDF that is used as a first approximation. The known PDF is usually chosen to be Gaussian, and only a few efforts have been made to generalize the GC expansion to the case of a non-Gaussian kernel [9, 7]. The terms in the GC and ED series involve derivatives of the known PDF and are functions of the cumulants of the PDF to approximate. The cumulants of a given random variable can be obtained analytically from its moments, which involve integrals of powers of the random variable [35]. For this reason, a stochastic Galerkin method (SGM) with global polynomials is employed for the solution of the SPDE. With this choice, QoIs can be defined as polynomial functions in the stochastic variable, so exact moments can be

computed with appropriate quadrature rules. Hence, also the coefficients of the GC and ED expansions can be computed exactly. The SGM is validated against the Monte Carlo method (MCM) for small input variances of the random data. The method we propose still makes use of the histogram to determine the PDF of the QoI, although the histogram is used only as a crude approximation to determine the most appropriate truncation order of the series. The approximate PDF is given analytically as a truncated series and it does not depend directly on the histogram, so it is not affected by the dimension of the bins used to construct it and by the number of samples. Finally, if the kernel is a C^∞ function (i.e. a Gaussian distribution), the truncated GC and ED expansions are continuous and infinitely differentiable, therefore potentially more useful than just a histogram approximation.

The paper is structured as follows: in Section 2, the mathematical problem and the input data description are introduced; in Section 3, the two SFEMs employed for the solution of the SPDEs are described, namely the Monte Carlo method and the Stochastic Galerkin method; the theory on the Gram-Charlier and Edgeworth expansions is laid out in Section 4. Arguments on asymptotic expansions and convergence are given in Section 5 and Section 6, respectively. Numerical results are reported in Section 7, for different types of output distributions. The paper is concluded with Section 9, where our findings are discussed.

2 Formulation of the problem

For SPDEs, the stochasticity is taken into account assuming that the input data depends on a random variable, other than the physical variable as in standard partial differential equations [56, 16, 43, 59, 42, 58]. The input data consists of coefficient functions and forcing term. For simplicity, here it is assumed that the stochastic contribution is introduced only by the coefficients and not by the forcing term. The model SPDE considered is the Poisson problem defined on $D \times \Omega$, where $D \subset \mathbb{R}^d$ and Ω is the sample space

$$\begin{cases} -\nabla \cdot \left(a(\mathbf{x}, \omega) \, \nabla u(\mathbf{x}, \omega) \right) = f(\mathbf{x}) & \text{in} \quad D \times \Omega \\ u(\mathbf{x}, \omega) = 0 & \text{on} \quad \partial D \times \Omega. \end{cases} \tag{1}$$

As in [29, 3, 43, 42], the following assumption is made.

Assumption 1 *The coefficient function $a(\mathbf{x}, \omega)$ in system (1) has the following properties:*

1. *There exists two positive constants $a_{\min}$ and $a_{\max} < \infty$ such that*

$$a_{\min} \leq a(\mathbf{x}, \omega) \leq a_{\max},$$

   almost surely on Ω, for all $\mathbf{x} \in D$.

2. *$a(\mathbf{x}, \omega) = a(\mathbf{x}, \boldsymbol{y}(\omega))$ in $\overline{D} \times \Omega$, where $\boldsymbol{y}(\omega) = (y_1(\omega), y_2(\omega), \dots, y_N(\omega))$ is a vector of real-valued uncorrelated random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.*

3. *$a(\mathbf{x}, \boldsymbol{y}(\omega))$ is measurable with respect to $\boldsymbol{y}$.*

For any $n = 1, \dots, N$, let $\Gamma_n := y_n(\Omega) \subset \mathbb{R}$ and let us define the parameter space $\Gamma := \prod_{n=1}^N \Gamma_n$. The joint probability density function (PDF) for $\{y_n\}_{n=1}^N$ is denoted by $\rho(\boldsymbol{y}) : \Gamma \to \mathbb{R}^+$. Several choices of coefficient functions are possible to ensure that Assumption 1 is satisfied. Some examples are given in [29]. The choice made in this work is described in the next section.

2.1 Karhunen-Loève Expansion

The Karhunen-Loève (KL) series expansion has been widely used in the field of uncertainty quantification to represent random fields, such as the coefficient function in system (1), as an infinite sum of random variables [36, 53, 26]. For numerical simulations, due to the finite computational capability available, the series is truncated, resulting in the following approximation

$$a(\mathbf{x}, \boldsymbol{y}(\omega)) \approx \mathbb{E}[a(\mathbf{x}, \cdot)] + \sum_{n=1}^N \sqrt{\lambda_n} b_n(\mathbf{x}) y_n(\omega), \tag{2}$$

where, for $n = 1, \dots, N$, λ_n and b_n are, respectively, the eigenvalues and eigenfunctions of the covariance function of $a(\mathbf{x}, \boldsymbol{y}(\omega))$.

Remark 1 *We assume that the input random processes approximated with a truncated KL expansion are Gaussian, hence the random variables $\{y_n(\omega)\}_{n=1}^N$ are standard independent and identically distributed. This allows us to generate realizations of the truncated KL expansion by Monte Carlo sampling from Gaussian distributions.*

To bound the coefficient function away from zero, the KL expansion is used on the logarithm of $a(\mathbf{x}, \boldsymbol{y}(\omega)) - a_{\min}$ rather than on the function itself. Thus, let us define the function $\gamma(\mathbf{x}, \boldsymbol{y}(\omega))$ as follows

$$\gamma(\mathbf{x}, \boldsymbol{y}(\omega)) := \log(a(\mathbf{x}, \omega) - a_{\min}). \tag{3}$$

The above definition implies that

$$a(\mathbf{x}, \boldsymbol{y}(\omega)) = a_{\min} + \exp\Big(\gamma(\mathbf{x}, \boldsymbol{y}(\omega))\Big). \tag{4}$$

Using a truncated KL expansion to approximate $\gamma(\mathbf{x}, \boldsymbol{y}(\omega))$, we get

$$\gamma(\mathbf{x}, \boldsymbol{y}(\omega)) \approx \mu_\gamma + \sum_{n=1}^N \sqrt{\lambda_n} b_n(\mathbf{x}) y_n(\omega), \tag{5}$$

where $\mu_\gamma := \mathbb{E}[\gamma(\mathbf{x}, \cdot)]$, and N is the dimension of Γ. Note that, in Eq. (5), (λ_n, b_n) are the eigenpairs associated with the covariance function of $\gamma(\mathbf{x}, \boldsymbol{y}(\omega))$. Moreover, according to Remark 1, when the truncated KL is used to approximate $\gamma(\mathbf{x}, \boldsymbol{y}(\omega))$, the assumption of Gaussian distribution is made on $\gamma(\mathbf{x}, \boldsymbol{y}(\omega))$, and consequently $a(\mathbf{x}, \boldsymbol{y}(\omega))$ has a log-normal distribution, as shown in Eq. (4). It follows from Eq. (4) that $a(\mathbf{x}, \boldsymbol{y}(\omega))$ is approximated by

$$a(\mathbf{x}, \boldsymbol{y}(\omega)) \approx a_{\min} + \exp\Big(\mu_\gamma + \sum_{n=1}^N \sqrt{\lambda_n} b_n(\mathbf{x}) y_n(\omega)\Big). \tag{6}$$

The eigenvalues and eigenfunctions in Eq. (6) are obtained solving the generalized eigenvalue problem

$$\int_D C_\gamma(\mathbf{x}, \widehat{\mathbf{x}}) b_n(\mathbf{x}) d\mathbf{x} = \lambda_n b_n(\widehat{\mathbf{x}}), \tag{7}$$

where $C_\gamma(\mathbf{x}, \widehat{\mathbf{x}})$ is the covariance function of the field $\gamma(\mathbf{x}, \boldsymbol{y}(\omega))$. The covariance structure is generally unknown and usually a specific covariance function is assumed. In this work, the covariance function of $\gamma(\mathbf{x}, \boldsymbol{y}(\omega))$ is assumed to be

$$C_\gamma(\mathbf{x}, \widehat{\mathbf{x}}) = \sigma_\gamma^2 \exp\Big[-\frac{1}{L}\Big(\sum_{i=1}^d |x_i - \widehat{x}_i|\Big)\Big], \tag{8}$$

where σ_γ denotes the standard deviation of $\gamma(\mathbf{x}, \boldsymbol{y}(\omega))$, d is the dimension of the spatial variable and $L > 0$ is a correlation length satisfying $L \leq \mathrm{diam}(D)$.

The SPDE in Eq. (1) is solved using stochastic finite element (SFEM) methods, and the solution of Eq. (7) is obtained according to a Galerkin approach, following the procedure presented, for instance, in [53, 31]. Assume that the physical domain D is discretized with a regular finite element grid $\mathcal{T}_h$ of size h [15, 10, 2], and let J_h be the total number of degrees of freedom. If $\{\phi_j\}_{j=1}^{J_h}$ denotes the global nodal basis associated with $\mathcal{T}_h$, the eigenfunction b_n of the covariance function of $\gamma(\mathbf{x}, \boldsymbol{y}(\omega))$ in Eq. (6) is approximated by its nodal interpolant

$$b_n(\mathbf{x}) \approx \sum_{j=1}^{J_h} b_{j,n} \phi_j(\mathbf{x}), \tag{9}$$

where $b_{j,n} := b_n(\mathbf{x}_j)$, with $\mathbf{x}_j$ being the j-th degree of freedom of $\mathcal{T}_h$. A substitution of Eq. (9) in Eq. (7) transforms the continuous problem into a finite dimensional one

$$\sum_{j=1}^{J_h} b_{j,n}\Big(\int_D C_\gamma(\mathbf{x}, \widehat{\mathbf{x}}) \phi_j(\mathbf{x}) d\mathbf{x} - \lambda_n \phi_j(\widehat{\mathbf{x}})\Big) = 0. \tag{10}$$

Next, Eq. (10) is multiplied by ϕ_i, and integrated with respect to $\widehat{\mathbf{x}}$, to obtain

$$\sum_{j=1}^{J_h} b_{j,n}\Big(\int_D \int_D C_\gamma(\mathbf{x}, \widehat{\mathbf{x}}) \phi_j(\mathbf{x}) \phi_i(\widehat{\mathbf{x}}) d\mathbf{x} d\widehat{\mathbf{x}} - \lambda_n \int_D \phi_j(\widehat{\mathbf{x}}) \phi_i(\widehat{\mathbf{x}}) d\widehat{\mathbf{x}}\Big) = 0. \tag{11}$$

Let us define the two real symmetric $J_h \times J_h$ matrices $\boldsymbol{C}$ and $\boldsymbol{M}$ by

$$C_{ij} := \int_D \int_D C_\gamma(\mathbf{x}, \widehat{\mathbf{x}}) \phi_j(\mathbf{x}) \phi_i(\widehat{\mathbf{x}}) d\mathbf{x} d\widehat{\mathbf{x}}, \qquad M_{ij} := \int_D \phi_j(\widehat{\mathbf{x}}) \phi_i(\widehat{\mathbf{x}}) d\widehat{\mathbf{x}}. \qquad (12)$$

Note that $\boldsymbol{M}$ is the standard finite element mass matrix, and is positive definite. With these matrices, Eq. (11) can be rewritten as the vector equation

$$\boldsymbol{C}\boldsymbol{b}_n = \lambda_n \boldsymbol{M}\boldsymbol{b}_n, \qquad (13)$$

Because $\boldsymbol{C}$ is symmetric, its eigenvectors $\boldsymbol{b}_n$ associated with distinct eigenvalues are orthogonal with respect to the mass matrix $\boldsymbol{M}$, that is $\boldsymbol{b}_n^T \boldsymbol{M} \boldsymbol{b}_m = 0$ if $m \neq n$. This orthogonality property implies that the approximations of the functions $b_n(\mathbf{x})$ in Eq. (9) are orthogonal in $L^2(D)$. Orthonormality can be obtained dividing these functions by their $L^2(D)$ norm. Note that orthonormality in $L^2(D)$ is a requirement for the functions $b_n(\mathbf{x})$.

For the numerical tests, the generalized eigenvalue problem in Eq. (13) is implemented in the in-house finite element code FEMuS [1], and solved with the SLEPc library [30].

2.2 Quantities of Interest

The aim of the paper is to approximate probability density functions of functionals of the SPDE solution, which we refer to as quantities of interest. Let $u_{J_h}(\mathbf{x}, \boldsymbol{y})$ be the approximate solution of system (1), obtained with an SFEM. The methods used to solve such a system are discussed in the next section. Given $u_{J_h}(\mathbf{x}, \boldsymbol{y})$, examples of quantities of interest include the spatial average over the physical domain

$$\mathcal{Q}_u(\boldsymbol{y}) = \frac{1}{|D|} \int_D u_{J_h}(\mathbf{x}, \boldsymbol{y}) d\mathbf{x}, \qquad (14)$$

the integral of the square

$$\mathcal{Q}_u(\boldsymbol{y}) = \int_D \left(u_{J_h}(\mathbf{x}, \boldsymbol{y}) \right)^2 d\mathbf{x}, \qquad (15)$$

or the maximum over the physical domain

$$\mathcal{Q}_u(\boldsymbol{y}) = \max_{\mathbf{x} \in D} u_{J_h}(\mathbf{x}, \boldsymbol{y}). \qquad (16)$$

Note that the dependence on the physical variable $\mathbf{x}$ is eliminated in the quantities of interest, as $\mathcal{Q}_u(\boldsymbol{y})$ is a scalar function that only depends on the stochastic variable $\boldsymbol{y}$.

3 Solution of the SPDE

The SPDE (1) is solved numerically using stochastic finite element methods. A detailed description of SFEMs for SPDEs with random input data such as the one in system (1) can be found in the review article [29] and references therein. The SFEM used here are described in the next two sections, together with the procedures to compute stochastic quantities such as moments, once the approximate SFEM solution is obtained.

3.1 The Monte Carlo Method

The first SFEM considered is the classical Monte Carlo Method (MCM), which is a stochastic sampling method [29]. With MCM, M points $\{\boldsymbol{y}_m\}_{m=1}^M$ are chosen randomly in the parameter domain Γ, and system (1) is solved independently for each of these points. In such a way, M realization of the solution of the SPDE are obtained and M uncoupled finite element systems are solved. For more details on MCM and error estimates, see [25, 39, 29]. From now on, the number M will be referred to as the number of MCM samples. According to Remark 1, the MCM samples are randomly drawn from Gaussian distributions. The mean μ_γ and standard deviation σ_γ of $\gamma(\mathbf{x}, \boldsymbol{y}(\omega))$ in Eq. (3) are not in general zero and one, respectively. Hence, if one chooses samples for the KL expansion from a standard Gaussian distribution, because of the presence of μ_γ in the KL and σ_γ in the covariance function, the result is the same as sampling from a non-standard Gaussian distribution with mean μ_γ and standard deviation σ_γ. If $\boldsymbol{y}_m$ denotes one of the M samples obtained from the Monte Carlo Method, let $u_{J_h}(\mathbf{x}, \boldsymbol{y}_m)$ be the MCM solution of system (1) associated with the sample $\boldsymbol{y}_m$ and let $\mathcal{Q}_u(\boldsymbol{y}_m)$ be a value of a quantity of interest obtained from the realization $u_{J_h}(\mathbf{x}, \boldsymbol{y}_m)$. Then $\mathcal{Q}_u(\boldsymbol{y}_m)$ is not a function of a random variable but

rather just a scalar quantity. On the other hand, the quantity of interest $\mathcal{Q}_u(\boldsymbol{y})$ is a function of a random variable, and a random variable itself. Hence stochastic quantities such as moments and cumulants can be computed. In the framework of MCM, the mean of $\mathcal{Q}_u(\boldsymbol{y})$ is approximated with Monte Carlo integration as follows

$$\mathbb{E}\Big[\mathcal{Q}_u(\boldsymbol{y})\Big] \approx \frac{1}{M} \sum_{m=1}^{M} \mathcal{Q}_u(\boldsymbol{y}_m) := \mu_{\mathcal{Q}_u}. \tag{17}$$

For $l > 1$, the l-th moment is approximated in the same fashion by

$$\mathbb{E}\Big[\big(\mathcal{Q}_u(\boldsymbol{y})\big)^l\Big] \approx \frac{1}{M} \sum_{m=1}^{M} \big(\mathcal{Q}_u(\boldsymbol{y}_m)\big)^l. \tag{18}$$

An estimate of the variance is given by

$$\mathbb{E}\Big[\big(\mathcal{Q}_u(\boldsymbol{y}) - \mathbb{E}\big[\mathcal{Q}_u(\boldsymbol{y})\big]\big)^2\Big] \approx \frac{1}{M} \sum_{m=1}^{M} \big(\mathcal{Q}_u(\boldsymbol{y}_m) - \mu_{\mathcal{Q}_u}\big)^2.$$

The accuracy of Monte Carlo integration improves with the number of samples M, but the sampling error increases with the magnitude of the input variance [29]. For this reason, here the MCM is used only for small values of the input variance and to validate the stochastic Galerkin method, which is later employed for larger values of the input variance.

3.2 The Stochastic Galerkin Method

The second SFEM used to obtain an approximation of the solution of system (1) is a stochastic Galerkin method (SGM). With SGMs, the stochastic function space is approximated using a Galerkin procedure, as it is done for the physical function space in standard finite element methods [28, 4, 5]. Let us assume that the parameters are independent, so that the input joint probability function $\rho(\boldsymbol{y})$ can be expressed as $\rho(\boldsymbol{y}) = \prod_{n=1}^{N} \rho_n(y_n)$. According to a Galerkin methodology, the infinite dimensional space $L_\rho^2(\Gamma)$ is approximated by the finite dimensional space

$$\mathcal{P}_{\mathcal{J}(p)}(\Gamma) = \text{span}\Big\{ \prod_{n=1}^{N} y_n^{p_n} \Big| \boldsymbol{p} \in \mathcal{J}(p), y_n \in \Gamma_n \Big\},$$

$$\text{where } \mathcal{J}(p) = \Big\{ \boldsymbol{p} \in \mathbb{N}^N \Big| \sum_{n=1}^{N} p_n \leq p \Big\}, \tag{19}$$

which corresponds to the total degree (TD) multivariate polynomial space from [29]. Gaussian PDFs are assumed for the input variables, hence the multivariate probabilist Hermite polynomials are employed as an orthogonal basis for $\mathcal{P}_{\mathcal{J}(p)}(\Gamma)$. Moreover, it holds that $\Gamma_n = \mathbb{R}$ for all $n = 1, \ldots, N$, and so $\Gamma = \mathbb{R}^N$. The multivariate polynomials are obtained in a tensor product fashion from the univariate probabilist Hermite polynomials, which are appropriately scaled so that they form an orthonormal basis with respect to the PDF

$$\rho_n(y_n) = \frac{\exp\left(-y_n^2/2\right)}{\sqrt{2\,\pi}}. \tag{20}$$

The scaling is carried out by dividing the p_n-th univariate probabilist Hermite polynomial $H_{e_{p_n}}$ by $\sqrt{n!}$, to obtain

$$\int_{\mathbb{R}} \frac{H_{e_{p_n}}}{\sqrt{n!}} \frac{H_{e_{p_m}}}{\sqrt{m!}} \rho_n(y_n) dy_n = \delta_{nm}, \tag{21}$$

where $\rho_n(y_n)$ is as in Eq. (20) and δ_{nm} is Kronecker's delta. Multivariate $L_\rho^2(\Gamma)$-orthonormal Hermite polynomials are then defined as

$$H_{e_{\boldsymbol{p}}}(\boldsymbol{y}) = \prod_{n=1}^{N} H_{e_{p_n}}(y_n). \tag{22}$$

The SGM approximation of the solution of system (1) is defined as [29]

$$u_{J_h\,M_p}^{gSG}(\mathbf{x}, \boldsymbol{y}) = \sum_{\boldsymbol{p} \in \mathcal{J}(p)} u_{\boldsymbol{p}}(\mathbf{x}) H_{e_{\boldsymbol{p}}}(\boldsymbol{y}), \qquad \text{where } u_{\boldsymbol{p}}(\mathbf{x}) = \sum_{j=1}^{J_h} u_{\boldsymbol{p},j}\, \phi_j(\mathbf{x}).$$

We consider quantities of the form

$$\mathcal{Q}_u(\boldsymbol{y}) = \sum_{\boldsymbol{p} \in \mathcal{J}(p)} \beta_{\boldsymbol{p}} H_{e_{\boldsymbol{p}}}(\boldsymbol{y}). \tag{23}$$

When $\beta_{\boldsymbol{p}}$ is the average of $u_{\boldsymbol{p}}(\mathbf{x})$, we have

$$\mathcal{Q}_u(\boldsymbol{y}) = \sum_{\boldsymbol{p} \in \mathcal{J}(p)} \left(\frac{1}{|D|} \int_D u_{\boldsymbol{p}}(\mathbf{x}) d\mathbf{x} \right) H_{e_{\boldsymbol{p}}}(\boldsymbol{y}), \tag{24}$$

as in Eq. (14). When $\beta_{\boldsymbol{p}}$ is the integral of the square of $u_{\boldsymbol{p}}(\mathbf{x})$, we have

$$\mathcal{Q}_u(\boldsymbol{y}) = \sum_{\boldsymbol{p} \in \mathcal{J}(p)} \left(\int_D u_{\boldsymbol{p}}^2(\mathbf{x}) d\mathbf{x} \right) H_{e_{\boldsymbol{p}}}(\boldsymbol{y}), \tag{25}$$

As opposed to what it is obtained with MCM, the quantities of interest given by the SGM are now polynomial functions in the variable $\boldsymbol{y}$. Such a feature gives a considerable advantage in terms of accuracy, because moments can now be computed exactly with appropriate quadrature rules, rather than being approximated by Monte Carlo integration. The l-th moment of the $\mathcal{Q}_u(\boldsymbol{y})$ in Eq. (24) is defined as

$$\mathbb{E}\left[\left(\mathcal{Q}_u(\boldsymbol{y}) \right)^l \right] := \int_{\mathbb{R}^N} \left(\mathcal{Q}_u(\boldsymbol{y}) \right)^l \rho(\boldsymbol{y}) d\boldsymbol{y}. \tag{26}$$

The multivariate integral in the above equation can be computed exactly with a multidimensional Hermite quadrature rule. Exploiting Eq. (21), monodimensional quadrature points can be obtained as the zeros of the probabilist Hermite polynomials. The quadrature weights are given by the following formula

$$w_i = \frac{(n-1)!}{n \left(H_{e_{n-1}}(x_i) \right)^2}, \tag{27}$$

where x_i is the i-th quadrature point and $H_{e_{n-1}}$ is the $(n-1)$-th probabilist Hermite polynomial. By construction, the weights w_i have the property that $\sum_i^{N_q} w_i = 1$, where N_q is the total number of quadrature points employed in the numerical integration. Multidimensional quadrature points and weights can be obtained in a tensor product fashion. For the SGM, given $q \in \mathbb{N}$, the coefficient function $a(\mathbf{x}, \boldsymbol{y}(\omega))$ in system (1) is computed as [29]

$$a_{SGM}(\mathbf{x}, \boldsymbol{y}) = \sum_{\boldsymbol{q} \in \mathcal{J}(q)} a_{\boldsymbol{q}}(\mathbf{x}) H_{e_{\boldsymbol{q}}}(\boldsymbol{y}),$$
$$a_{\boldsymbol{q}}(\mathbf{x}) = \int_{\Gamma} a(\mathbf{x}, \boldsymbol{y}(\omega)) H_{e_{\boldsymbol{q}}}(\boldsymbol{y}) \rho(\boldsymbol{y}) d\boldsymbol{y}, \tag{28}$$

with $a(\mathbf{x}, \boldsymbol{y}(\omega))$ as in Eq. (4). In the numerical results, with the exception of the integral for the data projection in Eq. (28), all integrals are computed by an exact quadrature rule. Among the integrals that are computed exactly, there are also the moments in Eq. (26). It is worth mentioning that, if a non-standard Gaussian distribution with mean μ_γ and standard deviation σ_γ is considered for the input variables, the change of variable $(x - \mu_\gamma)/\sigma_\gamma$ would have to be made in all integrals over the parameter space. Although, one can avoid this change of variable by taking into account the non-standard distribution in the input data, as it has been discussed in Section 3.1.

4 The Gram-Charlier and Edgeworth expansions

The goal of this paper is to obtain the PDF $f_{\mathcal{Q}_u} : \mathbb{R} \to \mathbb{R}^+$ of a given quantity of interest $\mathcal{Q}_u(\boldsymbol{y})$ using the Gram-Charlier (GC) and the Edgeworth (ED) series expansions, which express the PDF in terms of a reference PDF. The coefficients of these expansions are functions of the cumulants of $\mathcal{Q}_u(\boldsymbol{y})$, which can be easily obtained once the moments of $\mathcal{Q}_u(\boldsymbol{y})$ have been computed. In general, if we denote by m_l and κ_l the l-th moment and the l-th cumulant respectively, the first six cumulants are expressed in terms of

the moments as follows [35, 7]:

$$
\begin{aligned}
\kappa_1 &= m_1, \\
\kappa_2 &= m_2 - m_1^2, \\
\kappa_3 &= m_3 + 2m_1^3 - 3m_1\, m_2, \\
\kappa_4 &= m_4 - 6m_1^4 + 12m_1^2 m_2 - 3m_2^2 - 4m_1 m_3, \\
\kappa_5 &= m_5 - 5m_4 m_1 - 10m_3 m_2 + 20m_3 m_1^2 + 30m_2^2 m_1 - 60m_2 m_1^3 + 24m_1^5, \\
\kappa_6 &= m_6 - 6m_5 m_1 - 15m_4 m_2 + 30m_4 m_1^2 - 10m_3^2 + 120m_3 m_2 m_1 \\
&\quad - 120m_3 m_1^3 + 30m_2^3 - 270m_2^2 m_1^2 + 360m_2 m_1^4 - 120m_1^6.
\end{aligned}
\tag{29}
$$

In practical applications, the GC and ED series are truncated and only a finite number of terms is retained. The truncation results in an approximation of the PDF $f_{\mathcal{Q}_u}$. The behavior of the truncated series as the number of terms is increased will be subject to a discussion in the next sections.

4.1 Derivation of the expansions

We choose the formulation given in [9] to formally describe the GC and ED series. Let X be any random variable with PDF $f_X(x)$. The characteristic function $\Psi_X(t)$ of X is the Fourier transform of $f_X(x)$ [35]. The cumulant generating function $K(t)$ is defined as $K(t) := \ln(\Psi_X(t))$, and it holds that

$$
\ln(\Psi_X(t)) = \sum_{l=1}^{\infty} \kappa_{X,l} \frac{(it)^l}{l!},
\tag{30}
$$

where i denotes the imaginary unit and $\kappa_{X,l}$ is the l-th cumulant of X. Let φ denote the standard Gaussian distribution defined in Eq. (20). With the same notation used for X, the following equation can be obtained

$$
\ln(\Psi_\varphi(t)) = \sum_{l=1}^{\infty} \kappa_{\varphi,l} \frac{(it)^l}{l!}.
\tag{31}
$$

Using the properties of exponentials and convergent series, Eq. (30) and Eq. (31) can be combined to get

$$
\Psi_X(t) = \exp\Big(\sum_{l=1}^{\infty} (\kappa_{X,l} - \kappa_{\varphi,l}) \frac{(it)^l}{l!} \Big) \Psi_\varphi(t).
\tag{32}
$$

The GC and ED expansions follow from different manipulations of Eq. (32).

The GC is obtained with the use of Bell polynomials $B_l(x_1, \ldots, x_l)$ [6, 40]. These polynomials have the property that

$$
\exp\Big(\sum_{l=1}^{\infty} x_l \frac{(it)^l}{l!} \Big) = \sum_{l=0}^{\infty} B_l(x_1, \ldots, x_l) \frac{(it)^l}{l!},
\tag{33}
$$

with $B_0 := 1$. Therefore, Eq. (32) becomes

$$
\Psi_X(t) = \Big(1 + \sum_{l=1}^{\infty} B_l(\kappa_{X,1} - \kappa_{\varphi,1}, \ldots, \kappa_{X,l} - \kappa_{\varphi,l}) \frac{(it)^l}{l!} \Big) \Psi_{\varphi(t)}.
\tag{34}
$$

The inverse Fourier transform applied to Eq. (34) gives the GC expansion

$$
f_X(x) = \Big(1 + \sum_{l=1}^{\infty} B_l(\kappa_{X,1} - \kappa_{\varphi,1}, \ldots, \kappa_{X,l} - \kappa_{\varphi,l}) \frac{(-1)^l D_x^{(l)}}{l!} \Big) \varphi(x),
\tag{35}
$$

where $D_x^{(l)}$ represents the l-th derivative operator with respect to x and

$$
D_x^{(l)}\big(\varphi(x) \big) = (-1)^l H_{e_l}(x) \varphi(x).
\tag{36}
$$

The first six Bell polynomials are given by [9]

$$
\begin{aligned}
B_1(x_1) &= x_1, \\
B_2(x_1, x_2) &= x_1^2 + x_2, \\
B_3(x_1, x_2, x_3) &= x_1^3 + 3x_1 x_2 + x_3, \\
B_4(x_1, x_2, x_3, x_4) &= x_1^4 + 6x_1^2 x_2 + 3x_2^2 + 4x_1 x_3 + x_4, \\
B_5(x_1, x_2, x_3, x_4, x_5) &= x_1^5 + 10x_1^3 x_2 + 15x_1 x_2^2 + 10x_1^2 x_3 + 10x_2 x_3 + 5x_1 x_4 + x_5, \\
B_6(x_1, x_2, x_3, x_4, x_5, x_6) &= x_1^6 + 15x_1^4 x_2 + 45x_1^2 x_2^2 + 15x_2^3 + 20x_1^3 x_3 + 60x_1 x_2 x_3 \\
&\quad + 10x_3^2 + 15x_1^2 x_4 + 15x_2 x_4 + 6x_1 x_5 + x_6.
\end{aligned}
\tag{37}
$$

Note that when X is a standardized variable, $\kappa_{X,0} = 0$ and $\kappa_{X,1} = 1$. Since $\kappa_{\varphi,0} = 0$, $\kappa_{\varphi,1} = 1$, and $\kappa_{\varphi,l} = 0$ for $l \geq 3$, Eq. (35) reduces to

$$
f_X(x) = \left(1 + \sum_{l=3}^{\infty} B_l(0, 0, \kappa_{X,3}, \ldots, \kappa_{X,l}) \frac{(-1)^l D_x^{(l)}}{l!} \right) \varphi(x),
\tag{38}
$$

To derive ED, one assumes that X can be written as the standardized sum [48, 49]

$$
X = \frac{1}{\sqrt{r}} \sum_{i=1}^{r} \frac{Z_i - \mu}{\sigma},
\tag{39}
$$

where the random variables $Z_1, Z_2, \ldots, Z_r$ are independent and identically distributed, each with mean μ, standard deviation σ, and l-th cumulant $\kappa_{Z,l}$. Note that $\kappa_{Z,1} = \mu$ and $\kappa_{Z,2} = \sigma^2$. Define $\nu_l = \kappa_{Z,l}/\sigma^l$, then [9]

$$
\kappa_{X,0} = 0, \qquad \kappa_{X,1} = 1, \qquad \kappa_{X,l} = \frac{\nu_l}{r^{\frac{l}{2}-1}}, \quad l \geq 3.
\tag{40}
$$

Substituting the above equation in Eq. (32), we get

$$
\Psi_X(t) = \exp\left(\sum_{l=3}^{\infty} \frac{\nu_l}{r^{\frac{l}{2}-1}} \frac{(it)^l}{l!} \right) \Psi_\varphi(t).
\tag{41}
$$

With a proper shift of the summation index and an inverse Fourier transform, the above series gives the ED expansion

$$
f_X(x) = \left(1 + \sum_{l=1}^{\infty} B_l(a_1, a_2, \ldots, a_l) \frac{1}{r^{l/2} \, l!} \right) \varphi(x),
\tag{42}
$$

The interested reader can consult [9] for more details on how the above equation is obtained. The coefficients a_l are given by

$$
a_l = \frac{\nu_{l+2}(-1)^{l+2} D_x^{(l+2)}}{(l+1)(l+2)}.
\tag{43}
$$

For ease of notation, if we write the series in Eq. (42) as

$$
f_X(x) = \sum_{l=0}^{\infty} (-1)^l \frac{\vartheta_l(x)}{r^{l/2}},
\tag{44}
$$

the first six coefficient functions $\vartheta_l(x)$ are given by [8]

$$
\begin{aligned}
\vartheta_0(x) &= \varphi(x), \\
\vartheta_1(x) &= \frac{1}{3!} \nu_3 D_x^{(3)}\big(\varphi(x) \big), \\
\vartheta_2(x) &= \frac{1}{4!} \nu_4 D_x^{(4)}\big(\varphi(x) \big) + \frac{1}{72} \nu_3^2 D_x^{(6)}\big(\varphi(x) \big), \\
\vartheta_3(x) &= \frac{1}{5!} \nu_5 D_x^{(5)}\big(\varphi(x) \big) + \frac{1}{144} \nu_3 \nu_4 D_x^{(7)}\big(\varphi(x) \big) + \frac{1}{1296} \nu_3^3 D_x^{(9)}\big(\varphi(x) \big), \\
\vartheta_4(x) &= \frac{1}{6!} \nu_6 D_x^{(6)}\big(\varphi(x) \big) + \left(\frac{1}{1152} \nu_4^2 + \frac{1}{720} \nu_3 \nu_5 \right) D_x^{(8)}\big(\varphi(x) \big) \\
&\quad + \frac{1}{1728} \nu_3^2 \nu_4 D_x^{(10)}\big(\varphi(x) \big) + \frac{1}{31104} \nu_3^4 D_x^{(12)}\big(\varphi(x) \big).
\end{aligned}
\tag{45}
$$

A schematic representation of the necessary steps to get from the coefficient function $a(\mathbf{x}, \boldsymbol{y}(\omega))$ to the approximation of the PDF $f_{\mathcal{Q}_u}$ is summarized in Figure 1. Referring to Figure 1, the physical variable $\mathbf{x}$ and the stochastic variable $\boldsymbol{y}$ are given as inputs to the coefficient function $a(\mathbf{x}, \boldsymbol{y}(\omega))$, which is used to assemble the SFEM system associated with the SPDE in (1). The system is then solved with an SFEM and an approximate solution is obtained. From this solution, quantities of interest are evaluated and their moments are computed, in the way associated with the SFEM chosen. Once the moments have been obtained, they are fed to the truncated series expansions described in this section and an approximation of the PDF $f_{\mathcal{Q}_u}$ is produced. In the next two sections, important considerations on the convergence of the GC and ED expansions are made. First, the concept of asymptotic expansion is discussed.
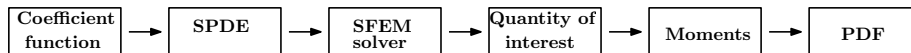

Figure 1: Summary of the steps to obtain an approximation of the PDF $f_{\mathcal{Q}_u}$.

5 Asymptotic expansions

Let $\{F_r(x)\}$ be a sequence of functions to be approximated by any partial sum of the series

$$\sum_{i=0}^{\infty} \frac{A_i(x)}{(\sqrt{r})^i}. \tag{46}$$

According to [54, 57], for a given r, the series (46) is an asymptotic expansion valid to n terms if the first $n+1$ partial sums have the property

$$\left| F_r(x) - \sum_{i=0}^{n} \frac{A_i(x)}{(\sqrt{r})^i} \right| \leq \frac{C_n(x)}{(\sqrt{r})^{n+1}}. \tag{47}$$

Moreover, if $C_n(x)$ does not depend on x, the expansion is said to be valid uniformly in x. As pointed out by Wallace [57], the asymptotic property is a property of finite partial sums and, for a given r, the series (46) may or may not be convergent. If the series (46) converges to $F_r(x)$, we know that for every $\varepsilon > 0$ there is an $R \in \mathbb{N}$ such that

$$\left| F_r(x) - \sum_{i=0}^{n} \frac{A_i(x)}{(\sqrt{r})^i} \right| < \varepsilon, \qquad \forall n \geq R. \tag{48}$$

While the above inequality states that for all partial sums with $n \geq R$, the error will be less than ε, it does not give specific information of what happens to the error before n reaches R. On the other hand, if r is sufficiently large, an asymptotic expansion valid to n terms has the property that the error will be uniformly reduced as more terms are added to the finite partial sum, from 1 up to $n+1$ terms, after which there is no longer the guarantee that adding successive terms will provide a uniform reduction of the error [54, 57]. With small r, the situation is a little different. Because the bounds $C_n(x)$ typically increase rapidly with n, a small value of r may be unable to make the denominator grow faster than the numerator of the right hand side in Ineq. (47), and only the first few terms would be improvements, as pointed out in [54, 57]. If a series is convergent, regardless of it being also asymptotic, the error will eventually go to zero as the number of terms in the partial sums is increased. If a series is asymptotic but not convergent, regardless of the value of r, there will be a minimum error that can be achieved, which limits the accuracy of the series. To summarize: convergent series give information on what happens to the error as the number of terms in the partial sum goes to infinity, whereas asymptotic series give information on the error as more terms are added starting from the first, and up to the $(n+1)$-th term.

We note that divergent asymptotic expansion have a long and useful history, especially with regards to applications. Excellent expositions on the subject are found in [54, 57].

6 Considerations on convergence

In [35] (page 152), the following sufficient condition for the convergence of the GC expansion is given

Proposition 1 *If the PDF $f(x)$ is of bounded variation on every finite interval and*

$$\int_{-\infty}^{\infty} |f(x)| \exp(-x^2/4)dx < \infty, \tag{49}$$

then the GC expansion of $f(x)$ is convergent.

While the integral in Proposition 1 is finite for PDFs with bounded support, the requirement of being of bounded variations on every finite interval may be in general harder to satisfy. We could not find in the literature any explicit sufficient condition addressing the convergence of the Edgeworth (ED) expansion.

Although, as pointed out by several authors [35, 19, 41, 24], convergence of the GC or ED series should not be a concern, because for practical applications the series are truncated. Of course, if the series are convergent, the error will eventually go to zero, but this is not useful for practical applications because the number of terms necessary to achieve convergence will likely be too big. The real question is how well a truncated GC or ED series can approximate the PDF of interest. This is directly connected to the definition of asymptotic expansion introduced in the previous section. With an asymptotic expansion, there is the guarantee that the initial error will be reduced by adding a certain number of successive terms. This property is relevant for our purpose regardless of the convergence of the series, because only a finite number of terms after the initial approximation are considered. Unfortunately, the GC series is not an asymptotic expansion [57, 8, 47, 34]. On the other hand, it has been shown in [18] that the ED expansion is asymptotic uniformly in x. Even though explicit bounds on errors were not given in [18], the asymptotic nature of ED expansion still makes it more valuable than the GC for the applications of our interest, where the series are truncated to a small number of terms.

In light of these observations, we believe that the best approach is composed of the following components. First we assume that we have in hand a stochastic Galerkin approximation of the quantity of interest. We do not include this step in the description of our algorithm because having in hand an SGM approximation of the QoI is something one might want for any approach for estimating PDFs of QoIs.

At the heart of our algorithms is the *construction of a truncated Edgeworth or truncated Gram-Charlier expansion approximation of the PDF for the QoI.*

1. Recursively for $l = 3, \ldots$, compute the l-th moment of the QoI. Note that we initialize by computing three moments.

   – Supposing we have in hand l moments so obtained, we use those moments to construct an l-term truncated ED or GC expansion, determining cumulants from the moments.

   – The moments are obtained via exact numerical integration of the SGM approximation of the QoI so that one need only integrate polynomials.

We need to *determine the "optimal" number of terms in the ED or GC expansions.*

2. This is done by comparing, each time we have incremented l, the l-term expansion with the $(l–1)$-term expansion previously determined.

   – The comparison can be done visually or by computing, e.g., using a sampling of the two expansions, the ℓ^2-norm of the difference between them.

At this point we have two possibilities.

3a. *If the ED or GC expansion is known to be convergent or*, using the comparisons done in **2** above, *the expansion "seems" to be convergent*, we stop the recursion when the difference computed in **2** is smaller that a prescribed tolerance.

   – In this case, the "optimal" number of terms in the truncated ED or GC expansion is determined by the prescribed tolerance.

3b. *If the ED or GC expansion is known to an be asymptotic but divergent or*, using the comparisons done in **2** above, *the expansion "seems" to be divergent*, increasing the number of terms in the expansion may not result in a better approximation so that we instead determine the "optimal" number of terms kept in the expansion as follows.

   – Run a Monte Carlo simulation to obtain samples of the QoI that are used to construct a histogram approximation of the PDF for the QoI, including bounds on the support of that PDF, in case it is bounded.
      – The Monte Carlo samples are determined from the SGM approximation of the QoI and not by doing expensive solves of the discretized SPDE.
      – As a result, the histogram approximation of the PDF for the QoI can be obtained at almost no cost. Even so, we do not need to take a huge number of samples because we only need to have in hand a crude histogram approximation.
   – By comparing the crude histogram so computed with the ED or GC approximate PDFs for all values of the number of terms used, one can determine the optimal number terms as that for which the expansion approximation is "closest" to the histogram.
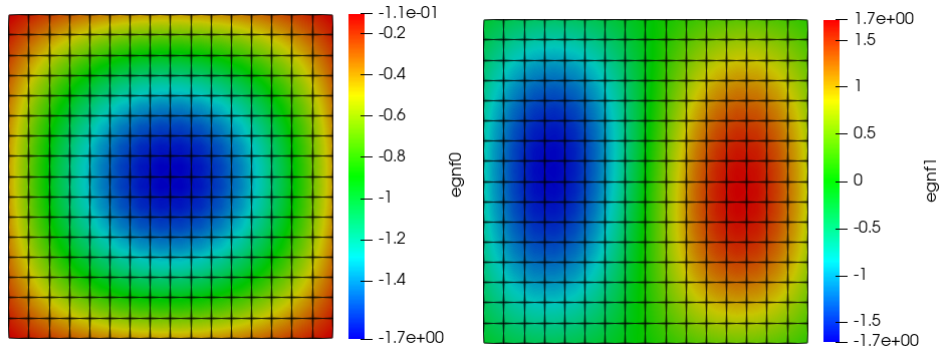
Figure 2: First eigenfunction (left) and second eigenfunction (right) in the KL expansion from Eq. (6) for the case of $\sigma_\gamma = 0.08$ and $N = 2$.

We also remark that the number of terms kept in ED or GC expansion approximations of PDFs is often considerable lower than, e.g., kernel density estimators (KDE) [32]. The number of terms in the latter is equal to the number of samples taken of the SGM approximation of the QoI. On the other hand, for ED or GC approximations of the PDF, the SGM approximation is sampled only at the quadrature points used to approximate moment integrals, so that the number of terms kept in the expansions is not otherwise related to the number of samples.

For the sake of simplicity, the numerical results reported in the next section do not follow every detail of the above algorithm, but those results are sufficient to illustrate the efficacy of our approach. A full implementation of the above algorithm would cast our approach in an even more favorable light as would some implementation steps that we have not discussed. An example of the latter is to take advantage of the use of SGM approximations of the QoI to avoid approximating integrals of polynomials that are known to vanish due to the use of orthogonal polynomials in the construction of the SGM approximation.

7   Numerical Results

We consider random variables associated with the solution of (1), obtained with the SFEMs discussed above. The methods are implemented in the in-house C++ FEM solver FEMuS [1], whereas the generalized eigenvalue problem in Eq. (13) is solved with the SLEPc library [30]. All numerical tests consider $f \equiv -1$ and dimension $d = 2$, for simplicity. The physical domain D is a unit square, with coarse grid composed of four bi-quadratic quadrilateral elements. The mesh for the simulations is obtained by refining three times the coarse grid according to a midpoint refinement procedure. The random field $a(\mathbf{x}, \boldsymbol{y}(\omega))$ is given by Eq. (6), and it is assumed that $\mu_\gamma = 0$. We chose $a_{\min} = 1/100$, and $L = 0.1$ for the covariance function in Eq. (8). The stochastic dimension is $N = 2$. The input standard deviation σ_γ is varied during the simulations. The quantities of interest used are given in (24), and (25). For MCM, the number of samples is $M = 10^5$. Unless otherwise stated, the same amount of samples is used to obtain the crude histogram approximations. We point out that a smaller value of M can be chosen, but a fairly large number of samples allows the crude histogram to also be used for comparison with the GC and ED expansions. Within the proposed method, the histogram is obtained sampling random values of $\boldsymbol{y}$ from a standard Gaussian distribution and plugging them in Eq. (23), once the SGM system has been solved. With MCM, the histogram is obtained in the standard way. For SGM, we choose the values of $p = 4$ in Eq. (19), and $q = 5$ in Eq. (28).

Remark 2 *When using GC, the quantity of interest is standardized before computing the moments, so the expansion adopted for the tests is Eq. (38). For ED, considering Eq. (42), it is assumed that $r = 1$ and that $Z_1 = \mathcal{Q}_u$. Hence, for both the GC and the ED expansions, the PDF refers to the standardized quantity of interest.*

7.1   Tests with nearly Gaussian output distribution

We begin with simulations where the output distribution of the quantity of interest is close to a standard Gaussian. For the QoI in Eq. (24), this is achieved with $\sigma_\gamma \in \{0.02, 0.04, 0.06, 0.08\}$. Because the truncated GC and ED expansions perform successive corrections of a standard Gaussian distribution, it is expected that the expansions will perform very well in this context. In Figure 2, the eigenfunctions $b_1(\mathbf{x})$ and $b_2(\mathbf{x})$ from Eq. (6) are reported for $\sigma_\gamma = 0.08$. The values of the moments and cumulants

SFEM	QoI in Eq. (24) ($N = 2$)					
	$\sigma_\gamma = 0.08$					
	m_1	m_2	m_3	m_4	m_5	m_6
MCM	-3.4798e-02	1.2110e-03	-4.2151e-05	1.4674e-06	-5.1082e-08	1.7786e-09
SGM	-3.4798e-02	1.2110e-03	-4.2152e-05	1.4673e-06	-5.1084e-08	1.7787e-09
	κ_1	κ_2	κ_3	κ_4	κ_5	κ_6
MCM	-3.4798e-02	1.4246e-07	0	0	0	0
SGM	-3.4798e-02	1.4356e-07	0	0	0	0
	$\sigma_\gamma = 0.06$					
	m_1	m_2	m_3	m_4	m_5	m_6
MCM	-3.4797e-02	1.2109e-03	-4.2141e-05	1.4667e-06	-5.1050e-08	1.7770e-09
SGM	-3.4797e-02	1.2109e-03	-4.2142e-05	1.4667e-06	-5.1051e-08	1.7770e-09
	κ_1	κ_2	κ_3	κ_4	κ_5	κ_6
MCM	-3.4797e-02	8.0123e-08	0	0	0	0
SGM	-3.4797e-02	8.0745e-08	0	0	0	0
	$\sigma_\gamma = 0.04$					
	m_1	m_2	m_3	m_4	m_5	m_6
MCM	-3.4796e-02	1.2108e-03	-4.2135e-05	1.4663e-06	-5.1027e-08	1.7758e-09
SGM	-3.4797e-02	1.2108e-03	-4.2135e-05	1.4663e-06	-5.1028e-08	1.7759e-09
	κ_1	κ_2	κ_3	κ_4	κ_5	κ_6
MCM	-3.4796e-02	3.5607e-08	0	0	0	0
SGM	-3.4796e-02	3.5884e-08	0	0	0	0
	$\sigma_\gamma = 0.02$					
	m_1	m_2	m_3	m_4	m_5	m_6
MCM	-3.4796e-02	1.2108e-03	-4.2131e-05	1.4660e-06	-5.1014e-08	1.7752e-09
SGM	-3.4796e-02	1.2108e-03	-4.2131e-05	1.4660e-06	-5.1014e-08	1.7752e-09
	κ_1	κ_2	κ_3	κ_4	κ_5	κ_6
MCM	-3.4796e-02	8.9013e-09	0	0	0	0
SGM	-3.4796e-02	8.9705e-09	0	0	0	0

Table 1: Behavior of moments and cumulants of the QoI in Eq. (24) for $\sigma_\gamma \in \{0.02, 0.04, 0.06, 0.08\}$.

for the quantity of interest in Eq. (24) are shown in Table 1. Note that such values are associated with the non-standardized quantity of interest. As discussed in [29], the sampling error of the MCM increases with the input standard deviation, hence we use this method only for small values of σ_γ, such as those considered in this section. The results obtained with MCM are relevant because they serve as a validation of the SGM, that is employed in the next section for more general output distributions. Observing Table 1, it is fair to say that convergence has been reached for the moments, given that there is a good agreement between the MCM and the SGM predictions. Therefore, the SGM can be considered validated by the MCM results. All cumulants κ_l with $l \geq 3$ are zero, confirming that the PDF is close to a Gaussian distribution, which is the one distribution that has $\kappa_l = 0$ for $l \geq 3$.

The histograms for $\sigma_\gamma = 0.08$ are in Figure 3. As expected, both suggest a nearly-Gaussian PDF: the major difference with an actual Gaussian distribution is that the in this case the PDFs have bounded support. This is due to the fact that the quantity of interest is bounded. Our tests have shown that larger values of σ_γ cause an increased skewness on the histograms, although this effect is very weak for the values of σ_γ considered here, and so histograms for $\sigma_\gamma = 0.02, 0.04$ and 0.06 are not shown due to their strong similarity to those in Figure 3. The increased skewness will be clearly visible when larger values of σ_γ are considered in the next section.

In Figure 3, the truncated GC expansion is displayed for $\sigma_\gamma = 0.08$. GC expansions for other values of σ_γ are not reported due to their strong resemblance to those in Figure 3. The notation GCj means that the GC expansion in Eq. (38) has been truncated at $l = j$. Truncations up to $l = 6$ have been computed. The GC expansion is a function defined on all $\mathbb{R}$, hence it could be graphed for ideally any x. Although, after the range of values of the quantity of interest is obtained from the crude approximation, the graph of the GC is plotted only for values of x in such a range. The same is done for the truncated Edgeworth expansions. Referring to Figure 3, we see that the GC expansions are in good agreement with the respective histogram.

In Figure 4 the truncated ED expansions are shown for MCM and SGM. Only the results for $\sigma_\gamma = 0.08$ are reported, because other values produced similar graphs. Note that, for GC, all the truncated series
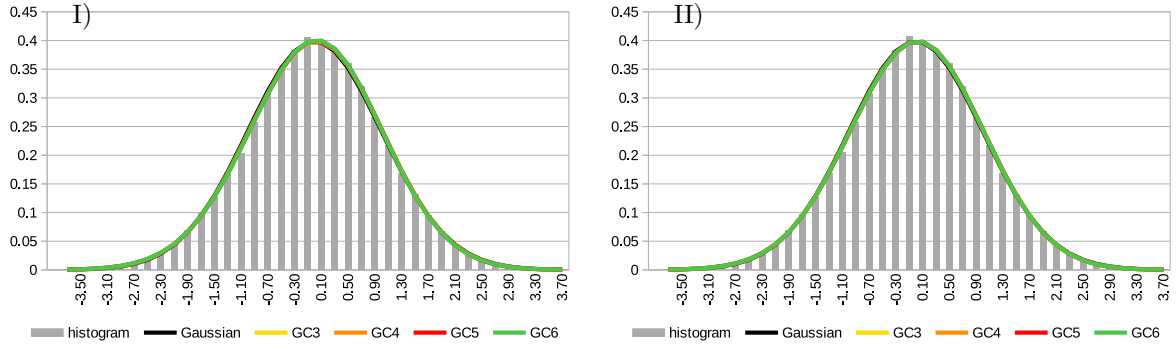
Figure 3: Gram-Charlier expansions of the standardized QoI in Eq. (24) for $\sigma_\gamma = 0.08$ and $N = 2$. I): GC - MCM, II): GC - SGM.
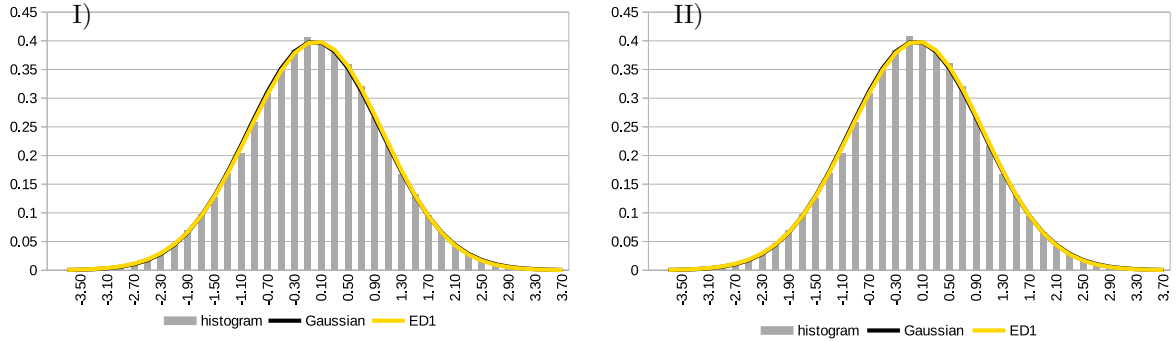

Figure 4: Edgeworth expansions of the standardized QoI in Eq. (24) for $N = 2$ and $\sigma_\gamma = 0.08$. I): ED - MCM, II): ED - SGM.

	Standardized QoI in Eq. (24) (SGM, $N = 2$)							
	Moments				Cumulants			
σ_γ	$m_{3_{STD}}$	$m_{4_{STD}}$	$m_{5_{STD}}$	$m_{6_{STD}}$	$\kappa_{3_{STD}}$	$\kappa_{4_{STD}}$	$\kappa_{5_{STD}}$	$\kappa_{6_{STD}}$
1.6	-0.793290	4.16802	-10.4848	46.2597	-0.79329	1.16802	-2.55194	7.44632
1.4	-0.68550	3.87021	-8.49255	36.8592	-0.68550	0.87021	-1.63752	4.10686
1.2	-0.58120	3.62412	-6.80469	29.8402	-0.58120	0.62412	-0.99265	2.10040
1	-0.47986	3.42448	-5.35446	24.6366	-0.47986	0.42448	-0.55579	0.96666

Table 2: Moments and cumulants of the standardized QoI in Eq. (24) with $N = 2$, for different values of the input standard deviation.

computed are displayed in Figure 3, namely GC3, GC4, GC5 and GC6. As discussed above, GC is not an asymptotic expansion and so it might be that, for instance, GC4 is a worse approximation than GC3 but GC5 is better than both GC3 and GC4. Hence, it makes sense to display all computed curves, to have an idea of the behavior of the truncated series. For ED, truncations up to four terms have been computed. However, not all curves are displayed in the figures, but only those that show a monotone reduction of the error. Such curves are typically ED1 and ED2. This scenario is expected considering the discussion in Section 5, and that $r = 1$ in Eq. (39). Figure 4 shows great agreement of the ED expansion with the histograms, for both the MCM and the SGM.

When the distribution of the quantity of interest is nearly Gaussian, our results suggest that the Gram-Charlier and Edgeworth expansions can effectively be used to describe the PDF of the quantity of interest. This is consistent with the nature of such expansions, that add successive corrections to a standard Gaussian distribution. The natural question that arises next concerns how well GC and ED can work when the PDF to approximate is not nearly Gaussian. Numerical results are presented in the next section to address this uncertainty.

| | Standardized QoI in Eq. (25) (SGM, $N = 2$) | | | | | | | |
| | Moments | | | | Cumulants | | | |
σ_γ	$m_{3_{STD}}$	$m_{4_{STD}}$	$m_{5_{STD}}$	$m_{6_{STD}}$	$\kappa_{3_{STD}}$	$\kappa_{4_{STD}}$	$\kappa_{5_{STD}}$	$\kappa_{6_{STD}}$
3	0.82751	4.31992	11.3622	51.2865	0.82751	1.31992	3.08712	9.63985
2.9	0.69515	3.89341	8.62298	37.3751	0.69515	0.89341	1.67138	4.14151
2.8	0.58557	3.62103	6.81178	29.6887	0.58557	0.62103	0.95604	1.94421
2.7	0.53517	3.51856	6.08253	27.0037	0.53517	0.51856	0.73080	1.36114

Table 3: Moments and cumulants of the standardized QoI in Eq. (25) with $N = 2$, for different values of the input standard deviation.
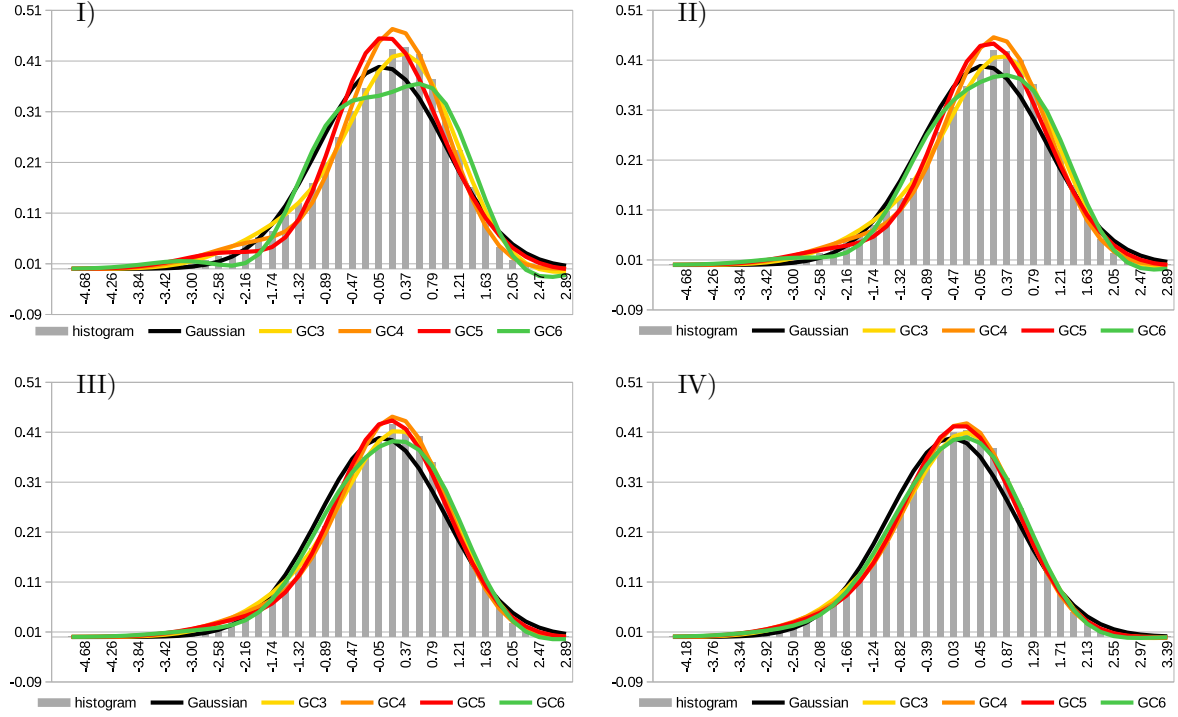
Figure 5: Gram-Charlier expansion of the standardized QoI in Eq. (24) for $N = 2$, obtained with SGM. I): $\sigma_\gamma = 1.6$, II): $\sigma_\gamma = 1.4$, III): $\sigma_\gamma = 1.2$, IV): $\sigma_\gamma = 1$.

7.2 Tests with general output distribution

We now investigate how well the GC and ED expansions can approximate a PDF that is not nearly Gaussian. This is achieved with larger values of the input standard deviation than those considered in the previous section. The QoI in (24) and QoI in (25) are considered. With the former, $\sigma_\gamma \in \{1, 1.2, 1.4, 1.6\}$, while with the latter $\sigma_\gamma \in \{2.7, 2.8, 2.9, 3\}$. Only the SGM is used, because for the values of σ_γ considered here, the sampling error of the MCM would produce results that are not accurate enough. The value $N = 2$ is chosen for the stochastic dimension. In Table 2 and Table 3, the values of the moments and cumulants are reported for the standardized QoIs. As the input standard deviation grows, all moments increase in magnitude, suggesting that larger values of σ_γ give increasingly non-Gaussian distributions. In Figure 5, the histograms for the QoI in (24) are reported: larger values of σ_γ produce an increasing negative skewness. The GC curves are in Figure 5. A stronger oscillating behavior with increasing σ_γ can be seen in the pictures. For all values of σ_γ, GC3 provides the best approximation of the PDF, among the GC curves that have been computed. The quality of the approximation slightly decreases as the input standard deviation grows but remains overall satisfactory, especially for lower values of σ_γ. GC4 and GC5 are also good approximations for $\sigma_\gamma = 1$ and $\sigma_\gamma = 1.2$, whereas GC6 may be considered good enough only for $\sigma_\gamma = 1$. In general, since the GC and ED series keep on correcting a standard Gaussian, it could be that more terms are required for a good approximation when the PDF is far from a Gaussian.

The ED results are shown in Figure 6. The best approximation among the curves computed is given by ED2 which well fits all histograms. The quality of the approximation slightly decreases as σ_γ grows, but the magnitude of this deterioration is much smaller than in the GC case. The ED expansions are
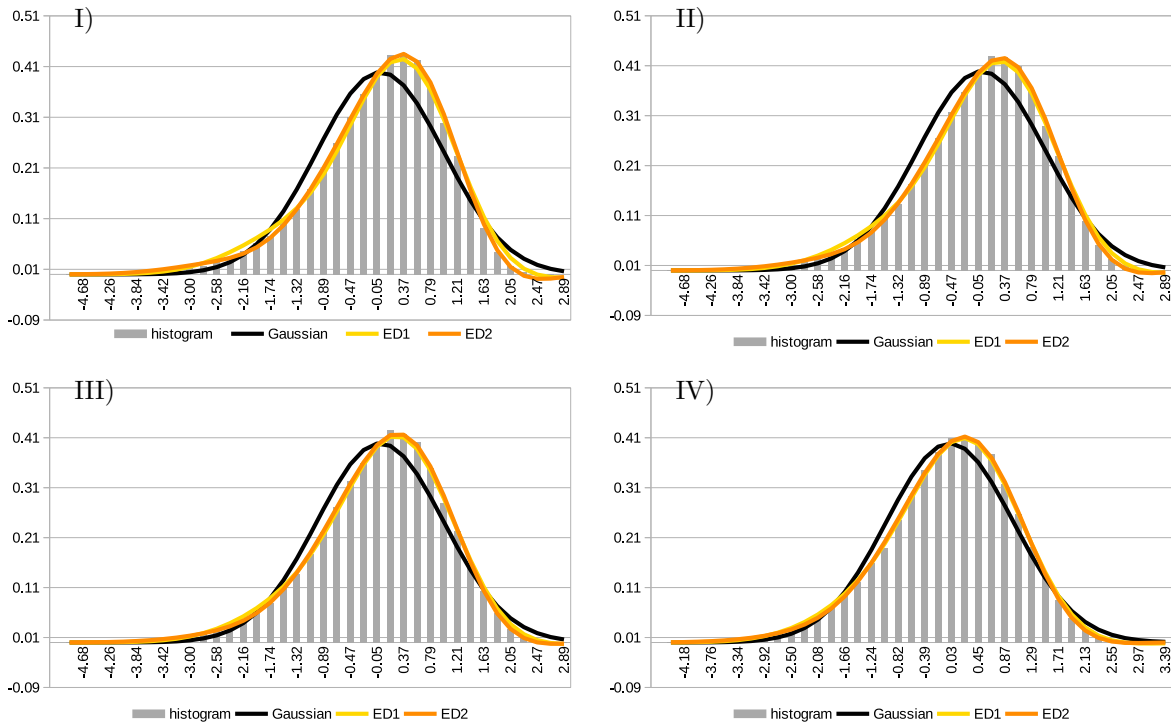
Figure 6: Edgeworth expansion of the standardized QoI in Eq. (24) for $N = 2$, obtained with SGM. I): $\sigma_\gamma = 1.6$, II): $\sigma_\gamma = 1.4$, III): $\sigma_\gamma = 1.2$, IV): $\sigma_\gamma = 1$.

qualitatively superior to the GC curves, because ED2 is a better approximation than GC3. Note that GC3 is by definition the same curve as ED1.

Next, we consider the quantity of interest given in Eq. (25). The histograms are in Figure 7. Greater values of σ_γ cause an increasing positive skewness, especially going from $\sigma_\gamma = 2.9$ to $\sigma_\gamma = 3$. The GC curves are in Figure 7. For $\sigma_\gamma = 2.7$, all computed GC curves well approximate the PDF, with GC4 and GC5 lying on top of each other. For all the other values of σ_γ, GC3 is again the best approximation, as the other computed curves progressively deviated from the histogram. For $\sigma_\gamma = 2.8$, GC4 and GC5 are still acceptable approximations, however they are not accurate enough for the subsequent values of σ_γ. GC6 is acceptable only for $\sigma_\gamma = 2.7$.

The ED curves are in Figure 8. ED2 well approximates the histogram for all values of σ_γ, similarly to what was observed in the previous example. Once again, due to the better approximation provided by ED2 compared to GC3, we conclude that the ED expansion is more valuable than the GC, for the examples considered.

We conclude this section with a comparison of the GC and ED expansions with the kernel density estimator. Given that our analysis is set in a univariate setting, the KDE is given by

$$f_K(x) = \frac{1}{h\,M} \sum_{m=1}^{M} \varphi\Big(\frac{x - \mathcal{Q}_u(\boldsymbol{y}_m)}{h}\Big), \tag{50}$$

where φ is the standard Gaussian distribution defined in Eq.(20), and M is the size of the sample set. The parameter h is called bandwidth and we selected it to be the same as the bin width used for the histograms in the previous figures. We chose a standard Gaussian kernel for the KDE because the GC and ED expansions also use a standard Gaussian kernel, hence the comparison is fair. In Figure 9, GC3 and ED2 are compared to the KDE estimator in Eq. (50) with $M = 10^5$: results for the QoI in (24) are visible in I) and II) using $\sigma_\gamma = 1.6$, whereas for the QoI in (25) they are in III) and IV) considering $\sigma_\gamma = 3$. The plots in figure 9 show that the GC and ED expansions are comparable to the KDE in terms of accuracy.

8   Computational times

Next, we compare the computational time required to construct a histogram approximation using $M = 10^6$ samples and the time required by the proposed method. For the latter, the moments in Eq. (26) are
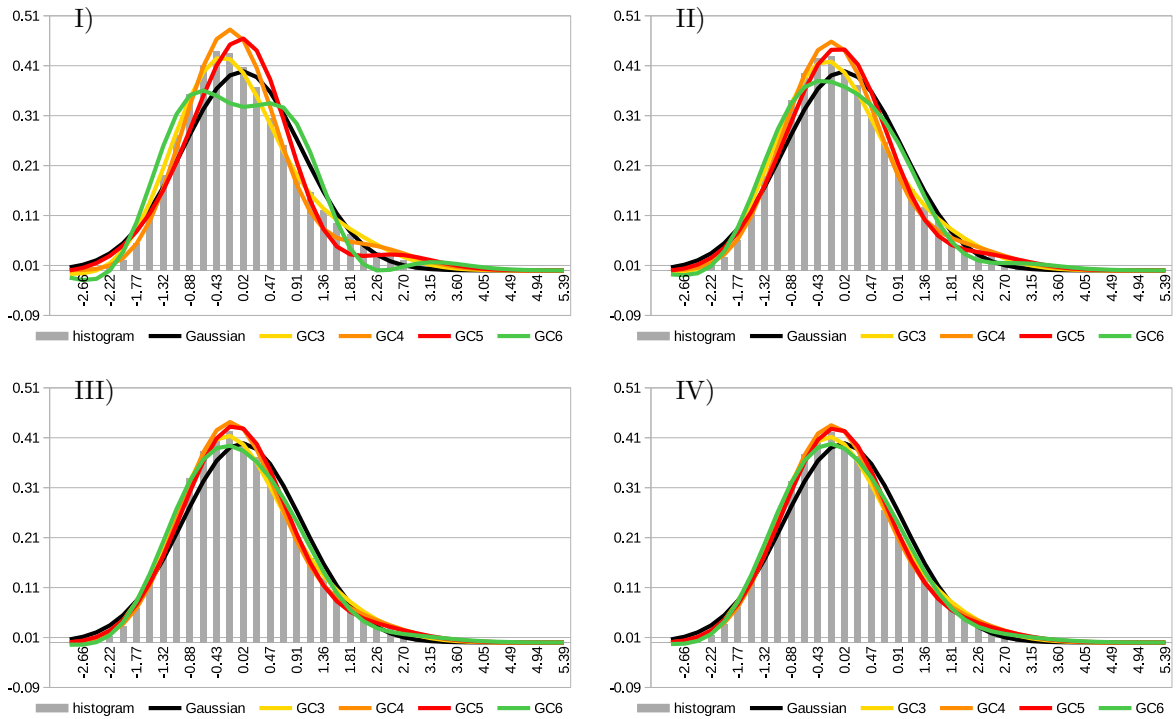
Figure 7: Gram-Charlier expansion of the standardized QoI in Eq. (25) for $N = 2$, obtained with SGM. I): $\sigma_\gamma = 3$, II): $\sigma_\gamma = 2.9$, III): $\sigma_\gamma = 2.8$, IV): $\sigma_\gamma = 2.7$.
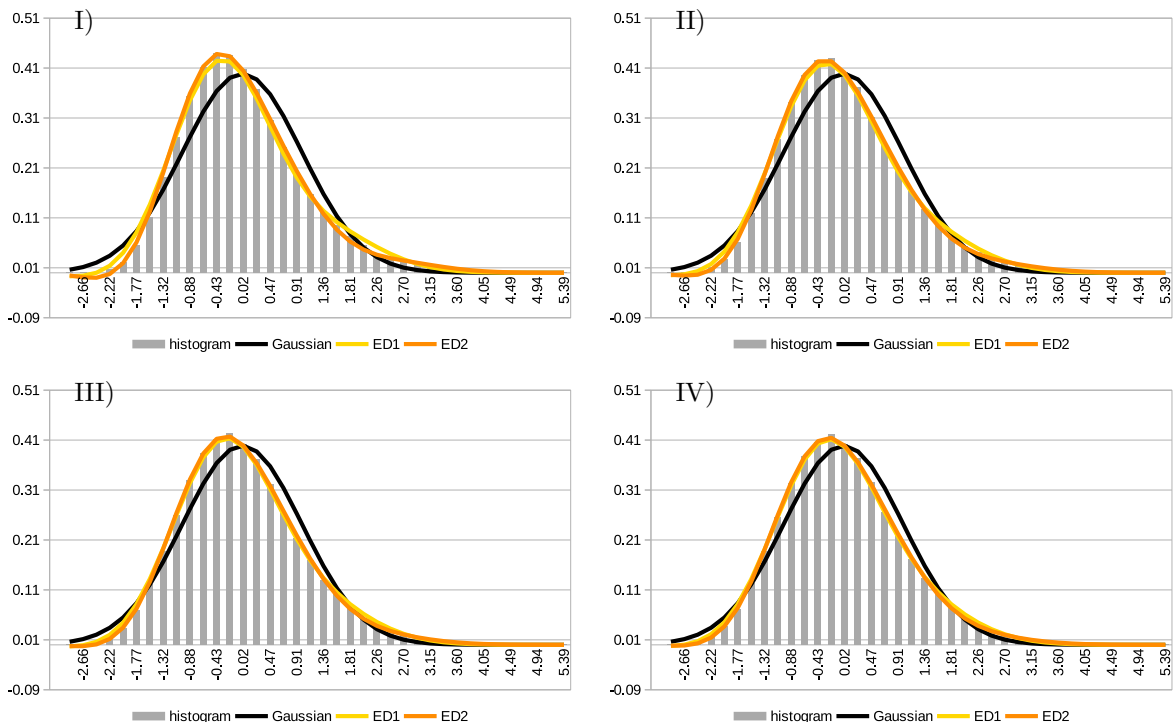

Figure 8: Edgeworth expansion of the standardized QoI in Eq. (25) for $N = 2$, obtained with SGM. I): $\sigma_\gamma = 3$, II): $\sigma_\gamma = 2.9$, III): $\sigma_\gamma = 2.8$, IV): $\sigma_\gamma = 2.7$.

evaluated with numerical quadrature, and a crude histogram approximation using $M = 10^4$ samples is computed. The comparison is carried out for different values of N, i.e. the dimension of the parameter space. The aim of this comparison is to show that the computational time required by the proposed approach is comparable to the one required to obtain an accurate histogram approximation. It is important to remark that the histogram is a discontinuous approximation, regardless of the number of samples employed, whereas the GC and ED expansions are continuous and infinitely differentiable. The CPU times
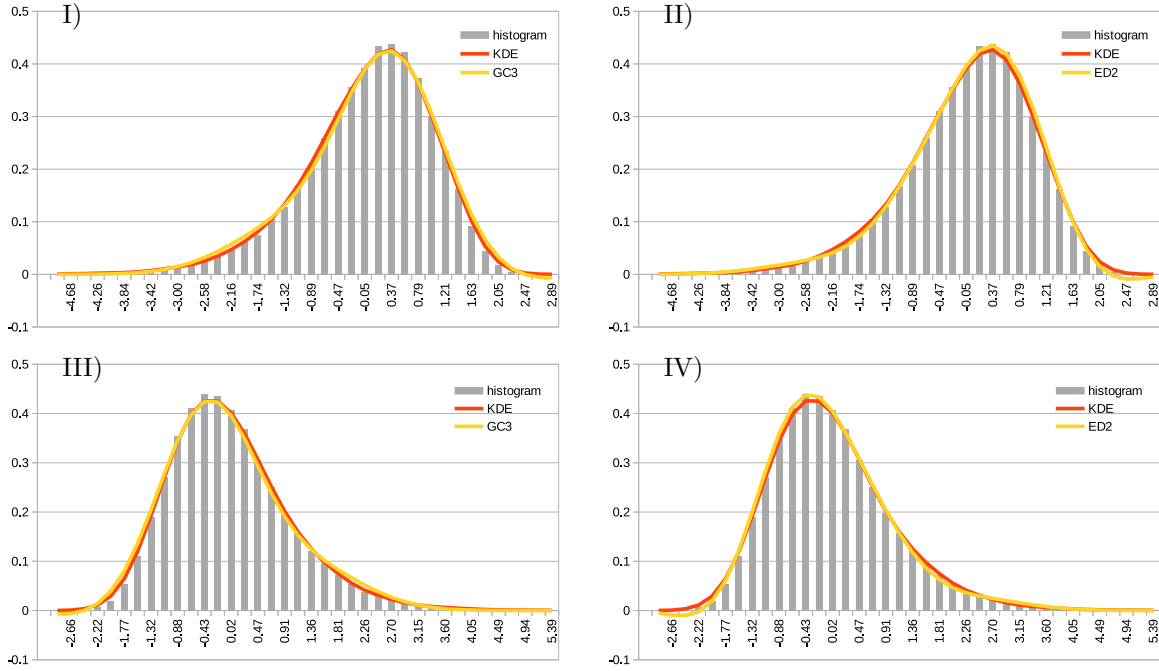
16

Figure 9: GC3 and ED2 expansions compared with the KDE estimator in Eq. (50). I) and II): QoI in (24) $\sigma_\gamma = 1.6$, III) and IV): QoI in (25) $\sigma_\gamma = 3$.

		Computational times [sec]			
	Histogram	Proposed			
N	TOT	Crude approx	Moments	Function evals	TOT
1	0.843663	0.007834	0.000115	0.000729	0.008678
2	1.251279	0.012149	0.001191	0.000477	0.013817
3	1.788778	0.017364	0.015390	0.000418	0.033172
4	2.533417	0.024813	0.308295	0.000497	0.333605

Table 4: Computational times of histogram and proposed method.

are shown in Table 4 and refer to the quantity of interest in Eq. (24) with $\sigma_\gamma = 1.6$. The other simulations parameters are as in the previous section, except N that is varied from 1 to 4. For the proposed method, the total CPU time reported is the sum of three different costs: the cost of the crude approximation with $M = 10^4$ samples, the cost of the numerical quadrature that does not involve any sampling, and the cost of performing enough function evaluations to plot the curves, that also does not involve any sampling. The results show that, for the values of N considered, the CPU times of the proposed method are lower than those of the histogram. It is true however that as the values of N increase, the number of quadrature points necessary for the evaluation of the moments in Eq. (26) will grow, likely causing the histogram to eventually become faster. In Table 5 we compare the CPU time of a function evaluation using the KDE with the CPU time of a function evaluation using the proposed method. The main difference between the two methods is that the dominant cost for the KDE is associated with online operations, whereas for the proposed method the most expensive operations are done offline. For the KDE, every function evaluation requires the computation of as many kernel values as the total number of samples. Function evaluations count as online operations for the estimation of the PDF. If the sample set is composed of M_s points and M_e function evaluations are required to estimate the PDF, the estimated cost of the KDE is proportional to the product $M_s M_e$ (the influence of the parameter space dimension is negligible). For the proposed method, the cost of a single function evaluation is given by the computation of the moments with numerical quadrature and by the crude histogram computation. These operations are all offline and can be performed once and for all regardless of the number of functions evaluations necessary to estimate the PDF. If $M_c < M_s$ is the number of samples used for the crude histogram and Q_p is the number of quadrature points required for the exact one-dimensional numerical quadrature, the estimated cost of the proposed method is proportional to $M_c + Q_p^N$, with N being the dimension of the parameter space. Table 5 shows CPU times results for $N = 1, 2, 3, 4$, considering $M_c = 10^4$ and $M_s = 10^6$ or $M_s = 500$ for the KDE. The KDE with $M_s = 500$ is the fastest approximation, however it also the most inaccurate,

CPU times for a single function evaluation [sec]						
	KDE ($M = 10^6$)		KDE ($M = 500$)		Proposed	
N	Offline	Online	Offline	Online	Offline	Online
1	negligible	0.048431	negligible	0.000034	0.007968	negligible
2	negligible	0.047788	negligible	0.000036	0.012642	negligible
3	negligible	0.052763	negligible	0.000050	0.044448	negligible
4	negligible	0.047178	negligible	0.000037	0.334770	negligible

Table 5: Computational times for a single evaluation of the KDE and of the proposed method (10^4 are used for the crude histogram).

as it can be seen from Figure 10. The KDE with $M_s = 10^6$, GC3 and ED2 have a comparable level of accuracy, hence we focus our analysis on the comparison between the proposed method and the KDE with $M_s = 10^6$. For $N = 1, 2, 3$ the proposed method is faster than a single function evaluation of the KDE with $M_s = 10^6$, whereas for $N = 4$, one KDE evaluation is approximately nine times faster than the proposed method. It is very likely that in general more than just nine function evaluations will be necessary to appropriately describe the approximated PDF. For instance, in Figure 10, the plots have been obtained with 46 function evaluations. Hence, the proposed method can compete with the KDE in terms of CPU time.
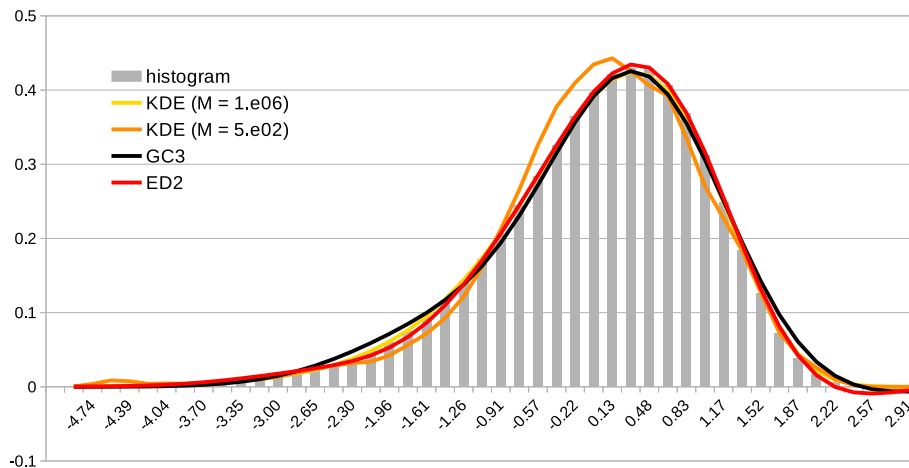

Figure 10: KDE and proposed method (GC3 and ED2) curves associated with the results in Table 5.

9 Discussion

It has been shown that the GC and ED truncated expansions represent a valid alternative to existing methods for the approximation of probability density functions associated with solutions of SPDEs. Our numerical results suggested that GC and ED provide an accurate estimate when the PDF is nearly Gaussian. This is consistent with the nature of the distributions. Even in the case of a general PDF, the truncated expansions well approximated the distributions. The asymptotic character of the ED expansion makes it more valuable than the GC due to a better monitoring of the error, as the truncation order is increased. Moreover, ED better approximated the distribution than GC, at least up to the truncation order considered in this work.

The proposed method is easy to implement and takes advantage of the fact that exact moments can be computed with the SGM, provided that enough quadrature points are considered. Moreover, all the computational burden is associated with offline computations, and point-wise evaluations have a negligible cost. A limitation is the lack of a rigorous procedure to determine the necessary number of terms in the truncated series to obtain the best possible approximation. This is inherent in the use of truncated series and the issue is present in all works on ED and GC that we were able to find in the literature. While the optimal number of terms to retain will likely be too large for the desired error in case of a convergent series, it could be theoretically determined a priori in case of an asymptotic expansion such as Edgeworth. Unfortunately, the lack of explicit error bounds make this unfeasible at the moment.

References

[1] Eugenio Aulisa, Simone Bná, and Giorgio Bornia. FEMuS Web page. https://github.com/eaulisa/MyFEMuS, 2017.

[2] Eugenio Aulisa, Giacomo Capodaglio, and Guoyi Ke. Construction of h-refined continuous finite element spaces with arbitrary hanging node configurations and applications to multigrid algorithms. *SIAM Journal on Scientific Computing*, 41(1):A480–A507, 2019.

[3] Ivo Babuška, Fabio Nobile, and Raul Tempone. A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 45(3):1005–1034, 2007.

[4] Ivo Babuska, Raúl Tempone, and Georgios E Zouraris. Galerkin finite element approximations of stochastic elliptic partial differential equations. *SIAM Journal on Numerical Analysis*, 42(2):800–825, 2004.

[5] Ivo Babuška, Raúl Tempone, and Georgios E Zouraris. Solving elliptic boundary value problems with uncertain coefficients by the finite element method: the stochastic formulation. *Computer methods in applied mechanics and engineering*, 194(12-16):1251–1294, 2005.

[6] Eric Temple Bell. Partition polynomials. *Annals of Mathematics*, pages 38–46, 1927.

[7] Mário N Berberan-Santos. Expressing a probability density function in terms of another PDF: A generalized Gram-Charlier expansion. *Journal of Mathematical Chemistry*, 42(3):585–594, 2007.

[8] Sergei Blinnikov and Richhild Moessner. Expansions for nearly gaussian distributions. *Astronomy and Astrophysics Supplement Series*, 130(1):193–205, 1998.

[9] Torgeir Brenn and Stian Normann Anfinsen. A revisit of the gram-charlier and edgeworth series expansions. 2017.

[10] Susanne Brenner and Ridgway Scott. *The mathematical theory of finite element methods*, volume 15. Springer Science & Business Media, 2007.

[11] Tan Bui-Thanh, Omar Ghattas, James Martin, and Georg Stadler. A computational framework for infinite-dimensional bayesian inverse problems part i: The linearized case, with application to global seismic inversion. *SIAM Journal on Scientific Computing*, 35(6):A2494–A2523, 2013.

[12] José E Chacón, Tarn Duong, et al. Data-driven density derivative estimation, with applications to nonparametric clustering and bump hunting. *Electronic Journal of Statistics*, 7:499–532, 2013.

[13] Peng Chen and Christoph Schwab. Model order reduction methods in computational uncertainty quantification. *Handbook of Uncertainty Quantification*, pages 1–53, 2016.

[14] Yizong Cheng. Mean shift, mode seeking, and clustering. *IEEE transactions on pattern analysis and machine intelligence*, 17(8):790–799, 1995.

[15] Philippe G Ciarlet. The finite element method for elliptic problems. *Classics in applied mathematics*, 40:1–511, 2002.

[16] K Andrew Cliffe, Mike B Giles, Robert Scheichl, and Aretha L Teckentrup. Multilevel monte carlo methods and applications to elliptic pdes with random coefficients. *Computing and Visualization in Science*, 14(1):3, 2011.

[17] Carlo R Contaldi, Rachel Bean, and Joao Magueijo. Photographing the wave function of the universe. *Physics Letters B*, 468(3-4):189–194, 1999.

[18] Harald Cramér. On the composition of elementary errors: First paper: Mathematical deductions. *Scandinavian Actuarial Journal*, 1928(1):13–74, 1928.

[19] Harald Cramér. *Mathematical methods of statistics (PMS-9)*, volume 9. Princeton university press, 2016.

[20] Masoumeh Dashti and Andrew M Stuart. The bayesian approach to inverse problems. *Handbook of Uncertainty Quantification*, pages 311–428, 2017.

[21] MB De Kock, HC Eggers, and J Schmiegel. Edgeworth versus Gram-Charlier series: x-cumulant and probability density tests. *Physics of Particles and Nuclei Letters*, 8(9):1023–1027, 2011.

[22] Valerio B Di Marco and G Giorgio Bombi. Mathematical functions for the representation of chromatographic peaks. *Journal of Chromatography A*, 931(1-2):1–30, 2001.

[23] Hans C Eggers, Michiel B de Kock, and Jurgen Schmiegel. Determining source cumulants in femtoscopy with Gram–Charlier and Edgeworth series. *Modern Physics Letters A*, 26(24):1771–1782, 2011.

[24] Miao Fan, Vijay Vittal, Gerald Thomas Heydt, and Raja Ayyanar. Probabilistic power flow studies for transmission systems with photovoltaic generation using cumulants. *IEEE Transactions on Power Systems*, 27(4):2251–2261, 2012.

[25] George Fishman. *Monte Carlo: concepts, algorithms, and applications*. Springer Science & Business Media, 2013.

[26] Philipp Frauenfelder, Christoph Schwab, and Radu Alexandru Todor. Finite elements for elliptic problems with stochastic coefficients. *Computer methods in applied mechanics and engineering*, 194(2-5):205–228, 2005.

[27] Keinosuke Fukunaga and Larry Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on information theory*, 21(1):32–40, 1975.

[28] Roger G Ghanem and Pol D Spanos. Stochastic finite element method: Response statistics. In *Stochastic Finite Elements: A Spectral Approach*, pages 101–119. Springer, 1991.

[29] Max D Gunzburger, Clayton G Webster, and Guannan Zhang. Stochastic finite element methods for partial differential equations with random input data. *Acta Numerica*, 23:521–650, 2014.

[30] Vicente Hernandez, Jose E Roman, and Vicente Vidal. SLEPc: A scalable and flexible toolkit for the solution of eigenvalue problems. *ACM Transactions on Mathematical Software (TOMS)*, 31(3):351–362, 2005.

[31] SP Huang, ST Quek, and KK Phoon. Convergence study of the truncated karhunen–loeve expansion for simulation of stochastic processes. *International journal for numerical methods in engineering*, 52(9):1029–1043, 2001.

[32] Alan Julian Izenman. Review papers: Recent developments in nonparametric density estimation. *Journal of the American Statistical Association*, 86(413):205–224, 1991.

[33] Eric Jondeau and Michael Rockinger. Gram–Charlier densities. *Journal of Economic Dynamics and Control*, 25(10):1457–1483, 2001.

[34] Roman Juszkiewicz, David Weinberg, Piotr Amsterdamski, Michal Chodorowski, and Francois Bouchet. Weakly non-linear gaussian fluctuations and the Edgeworth expansion. *arXiv preprint astro-ph/9308012*, 1993.

[35] Maurice G Kendall. *Advanced Theory Of Statistics Vol-I*. Charles Griffin: London, 1943.

[36] CF Li, YT Feng, DRJ Owen, DF Li, and IM Davis. A fourier–karhunen–loève discretization scheme for stationary random material properties in sfem. *International journal for numerical methods in engineering*, 73(13):1942–1965, 2008.

[37] Xiang Ma and Nicholas Zabaras. An efficient bayesian inference approach to inverse problems based on an adaptive sparse grid collocation method. *Inverse Problems*, 25(3):035013, 2009.

[38] Youssef M Marzouk, Habib N Najm, and Larry A Rahn. Stochastic spectral methods for efficient bayesian solution of inverse problems. *Journal of Computational Physics*, 224(2):560–586, 2007.

[39] Nicholas Metropolis and Stanislaw Ulam. The monte carlo method. *Journal of the American statistical association*, 44(247):335–341, 1949.

[40] Miloud Mihoubi. Bell polynomials and binomial type sequences. *Discrete Mathematics*, 308(12):2450–2459, 2008.

[41] Trino-Manuel Ñíguez and Javier Perote. Forecasting heavy-tailed densities with positive Edgeworth and Gram-Charlier expansions. *Oxford Bulletin of Economics and Statistics*, 74(4):600–627, 2012.

[42] Fabio Nobile, Raul Tempone, and Clayton G Webster. An anisotropic sparse grid stochastic colloca-tion method for partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 46(5):2411–2442, 2008.

[43] Fabio Nobile, Raúl Tempone, and Clayton G Webster. A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 46(5):2309–2345, 2008.

[44] Yung-Kyun Noh, Masashi Sugiyama, Song Liu, Marthinus C Plessis, Frank Chongwoo Park, and Daniel D Lee. Bias reduction and metric learning for nearest-neighbor estimation of kullback-leibler divergence. In *Artificial Intelligence and Statistics*, pages 669–677, 2014.

[45] MCM O'brien. Using the Gram-Charlier expansion to produce vibronic band shapes in strong coupling. *Journal of Physics: Condensed Matter*, 4(9):2347, 1992.

[46] Joaquim Olivé and Joan O Grimalt. Gram-Charlier and Edgeworth-Cramér series in the character-ization of chromatographic peaks. *Analytica chimica acta*, 249(2):337–348, 1991.

[47] Jamol Pender. Gram Charlier expansion for time varying multiserver queues with abandonment. *SIAM Journal on Applied Mathematics*, 74(4):1238–1265, 2014.

[48] Valentin V Petrov. Limit theorems of probability theory: sequences of independent random variables. Technical report, Oxford, New York, 1995.

[49] Valentin Vladimirovich Petrov. *Sums of independent random variables*, volume 82. Springer Science & Business Media, 2012.

[50] Ray Popovic and David Goldsman. Easy Gram-Charlier valuations of options. *Journal of Derivatives*, 20(2):79, 2012.

[51] JM Rickman, A Lawrence, AD Rollett, and MP Harmer. Calculating probability densities associated with grain-size distributions. *Computational Materials Science*, 101:211–215, 2015.

[52] Hiroaki Sasaki, Yung-Kyun Noh, and Masashi Sugiyama. Direct density-derivative estimation and its application in kl-divergence approximation. In *Artificial Intelligence and Statistics*, pages 809–818, 2015.

[53] Mattias Schevenels, Geert Lombaert, and Geert Degrande. Application of the stochastic finite element method for gaussian and non-gaussian systems. In *ISMA2004 International Conference on Noise and Vibration Engineering*, pages 3299–3314, 2004.

[54] Robert Sedgewick and Philippe Flajolet. *An introduction to the analysis of algorithms*. Pearson Education India, 2013.

[55] Andrew M Stuart. Inverse problems: a bayesian perspective. *Acta Numerica*, 19:451–559, 2010.

[56] Daniel M Tartakovsky and Svetlana Broyda. Pdf equations for advective–reactive transport in heterogeneous porous media with uncertain properties. *Journal of contaminant hydrology*, 120:129–140, 2011.

[57] David L Wallace. Asymptotic approximations to distributions. *The Annals of Mathematical Statis-tics*, 29(3):635–654, 1958.

[58] Xiaoliang Wan and George Em Karniadakis. An adaptive multi-element generalized polynomial chaos method for stochastic differential equations. *Journal of Computational Physics*, 209(2):617–642, 2005.

[59] Dongbin Xiu and Jan S Hesthaven. High-order collocation methods for differential equations with random inputs. *SIAM Journal on Scientific Computing*, 27(3):1118–1139, 2005.

[60] AS Zapevalov, AN Bol'shakov, and VE Smolov. Simulating of the probability density of sea surface elevations using the Gram-Charlier series. *Oceanology*, 51(3):407–414, 2011.