Available online at www.sciencedirect.com
ScienceDirect
Comput. Methods Appl. Mech. Engrg. 306 (2016) 95–122
Computer methods in applied mechanics and engineering
www.elsevier.com/locate/cma

# Multi-Index Stochastic Collocation for random PDEs

Abdul-Lateef Haji-Ali[a,*], Fabio Nobile[b], Lorenzo Tamellini[b,c], Raúl Tempone[a]

[a] *CEMSE, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia*
[b] *CSQI - MATHICSE, École Polytechnique Fédérale de Lausanne, Station 8, CH 1015, Lausanne, Switzerland*
[c] *Dipartimento di Matematica "F. Casorati", Università di Pavia, Via Ferrata 5, 27100 Pavia, Italy*

Received 29 August 2015; received in revised form 18 January 2016; accepted 18 March 2016
Available online 28 March 2016

**Abstract**

In this work we introduce the Multi-Index Stochastic Collocation method (MISC) for computing statistics of the solution of a PDE with random data. MISC is a combination technique based on mixed differences of spatial approximations and quadratures over the space of random data. We propose an optimization procedure to select the most effective mixed differences to include in the MISC estimator: such optimization is a crucial step and allows us to build a method that, provided with sufficient solution regularity, is potentially more effective than other multi-level collocation methods already available in literature. We then provide a complexity analysis that assumes decay rates of product type for such mixed differences, showing that in the optimal case the convergence rate of MISC is only dictated by the convergence of the deterministic solver applied to a one dimensional problem. We show the effectiveness of MISC with some computational tests, comparing it with other related methods available in the literature, such as the Multi-Index and Multilevel Monte Carlo, Multilevel Stochastic Collocation, Quasi Optimal Stochastic Collocation and Sparse Composite Collocation methods.


*MSC:* 41A10; 65C20; 65N30; 65N05

*Keywords:* Uncertainty Quantification; Random PDEs; Sparse grids; Stochastic Collocation methods; Multilevel methods; Combination technique

## 1. Introduction

Uncertainty Quantification (UQ) is an interdisciplinary, fast-growing research area that focuses on devising mathematical techniques to tackle problems in engineering and natural sciences in which only a probabilistic description of the parameters of the governing equations is available, due to measurement errors, intrinsic non-measurability/non-predictability, or incomplete knowledge of the system of interest. In this context, "parameters" is a term used in broad sense to refer to constitutive laws, forcing terms, domain shapes, boundary and initial conditions, etc.

UQ methods can be divided into deterministic and randomized methods. While randomized techniques, which include the Monte Carlo sampling method, are essentially based on random sampling and ensemble averaging,

* Corresponding author.
*E-mail addresses:* abdullateef.hajiali@kaust.edu.sa (A.-L. Haji-Ali), fabio.nobile@epfl.ch (F. Nobile), lorenzo.tamellini@unipv.it (L. Tamellini), raul.tempone@kaust.edu.sa (R. Tempone).

http://dx.doi.org/10.1016/j.cma.2016.03.029
0045-7825/

deterministic methods proceed by building a surrogate of the system's response function over the parameter space, which is then processed to obtain the desired information. Typical goals include computing statistical moments (expected value, variance, higher moments, correlations) of some quantity of interest of the system at hand, typically functionals of the state variables (forward problem), or updating the statistical description of the random parameters given some observations of the system at hand (inverse problem). In any case, multiple resolutions of the governing equations are needed to explore the dependence of the state variables on the random parameters. The computational method used should therefore be carefully designed to minimize the computational effort.

In this work, we focus on the case of PDEs with random data, for which both deterministic and randomized approaches have been extensively explored in recent years. As for the deterministic methods, we mention here the methods based on polynomial expansions computed either by global Galerkin-type projections [1–5] or collocation strategies based on sparse grids (see e.g. [6–9]), low-rank techniques [10–13] and reduced basis methods (see e.g. [14,15]). All these approaches have been found to be particularly effective when applied to problems with a moderate number of random parameters (low-dimensional probability space) and smooth response functions. Although significant effort has been expended on increasing the efficiency of such deterministic methods with respect to the number of random parameters (see, e.g., [16], the seminal work on infinite dimensional polynomial approximation of elliptic PDEs with random coefficients), Monte Carlo-type approximations remain the primary choice for problems with non-smooth response functions and/or those that depend on a high number of random parameters, despite their slow convergence with respect to sample size.

A very promising methodology that builds on the classical Monte Carlo method and enhances its performance is offered by the so-called *Multilevel Monte Carlo* (MLMC). It was first proposed in [17] for applications in parametric integration and extended to weak approximation of stochastic differential equations in [18], which also provided a full complexity analysis. Let $\{h_\ell\}_{\ell=0}^{L}$ be a (scalar) sequence of spatial/temporal resolution levels that can be used for the numerical discretization of the PDE at hand and $\{F_\ell\}_{\ell=0}^{L}$ be the corresponding approximations of the quantity of interest, and suppose that the final goal of the UQ analysis is to compute the expected value of $F$, $\mathbb{E}[F]$. While a classic Monte Carlo approach simply approximates the expected value by using an ensemble average over a sample of independent replicas of the random parameters, the MLMC method relies on the simple observation that, by linearity of expectation,

$$\mathbb{E}[F] \approx \mathbb{E}[F_L] = \mathbb{E}[F_0] + \sum_{\ell=1}^{L} \mathbb{E}[F_\ell - F_{\ell-1}], \tag{1}$$

and computes by independent Monte Carlo samplers each expectation in the sum. Indeed, if the discretization of the underlying differential model is converging with respect to the discretization level, $\ell$, the variance of $(F_\ell - F_{\ell-1})$ will be smaller and smaller as $\ell$ increases, i.e., when the spatial/temporal resolution increases. Dramatic computational saving can thus be obtained by approximating the quantities $\mathbb{E}[F_\ell - F_{\ell-1}]$ with a smaller and smaller sample size, since most of the variability of $F$ will be captured with coarse simulations and only a few resolutions over the finest discretization levels will be performed. The MLMC estimator is therefore given by

$$\mathbb{E}[F] \approx \sum_{\ell=0}^{L} \frac{1}{M_\ell} \sum_{m=1}^{M_\ell} \left(F_\ell(\omega_{m,\ell}) - F_{\ell-1}(\omega_{m,\ell})\right), \quad \text{with } F_{-1}(\cdot) = 0, \tag{2}$$

where $\omega_{m,\ell}$ are the i.i.d. replicas of the random parameters. The application of MLMC methods to UQ problems involving PDEs with random data has been investigated from the mathematical point of view in a number of recent publications, see e.g. [19–23]. Recent works [24–27] have explored the possibility of replacing the Monte Carlo sampler on each level by other quadrature formulas such as sparse grids or quasi-Monte Carlo quadrature, obtaining the so-called Multilevel Stochastic Collocation (MLSC) or Multilevel Quasi-Monte Carlo (MLQCM) methods. See also [28] for a related approach where the Multilevel Monte Carlo method is combined with a control variate technique.

The starting point of this work is instead the so-called Multi-Index Monte Carlo method (MIMC), recently introduced in [29], that differs from the Multilevel Monte Carlo method in that the telescoping idea presented in Eqs. (1)–(2) is applied to discretizations indexed by a multi-index rather than a scalar index, thus allowing each discretization parameter to vary independently of the others. Analogously to what done in [24–26] in the context of

stochastic collocation, here we propose to replace the Monte Carlo quadrature with a sparse grid quadrature at each telescopic level, obtaining in our case the Multi-Index Stochastic Collocation method (MISC). In other words, MISC can be seen as a multi-index version of MLSC, or a stochastic collocation version of MIMC. From a slightly different perspective, MISC is also closely related to the combination technique developed for the solution of (deterministic) PDEs in [30,7,31–33]; in this work, the combination technique is used with respect to both the deterministic and stochastic variables.

One key difference between the present work and [24–26] is that the number of problem solves to be performed at each discretization level is not determined by balancing the spatial and stochastic components of the error (based, e.g., on convergence error estimates), but rather suitably extending the knapsack-problem approach that we employed in [34–36] to derive the so-called Quasi-Optimal Sparse Grids method (see also [37]). A somewhat analogous approach was proposed in [38], where the number of solves per discretization level is prescribed *a-priori* based on a standard sparsification procedure (we will give more details on the comparison between these different methods later on). In this work, we provide a complexity analysis of MISC and illustrate its performance improvements, comparing it to other methods by means of numerical examples.

The remainder of this paper is organized as follows. In Section 2, we introduce the problem to be solved and the approximation schemes that will be used. The Multi-Index Stochastic Collocation method is introduced in Section 3, and our main theorem detailing the complexity of MISC for a particular choice of an index set is presented in Section 4. Finally, Section 5 presents some numerical tests, while Section 6 offers some conclusions and final remarks. The Appendix contains the technical proof of the main theorem. Throughout the rest of this work we use the following notation:

- $\mathbb{N}$ denotes the set of integer numbers including zero;
- $\mathbb{N}_+$ denotes the set of positive integer numbers, i.e. excluding zero;
- $\mathbb{R}_+$ denotes the set of positive real numbers, $\mathbb{R}_+ = \{r \in \mathbb{R} : r > 0\}$;
- $\mathbf{1}$ denotes a vector whose components are always equal to one;
- $\boldsymbol{e}_\ell^\kappa$ denotes the $\ell$th canonical vector in $\mathbb{R}^\kappa$, i.e., $(\boldsymbol{e}_\ell^\kappa)_i = 1$ if $\ell = i$ and zero otherwise; however, for the sake of clarity, we often omit the superscript $\kappa$ when obvious from the context. For instance, if $\boldsymbol{v} \in \mathbb{R}^N$, we will write $\boldsymbol{v} - \boldsymbol{e}_1$ instead of $\boldsymbol{v} - \boldsymbol{e}_1^N$;
- given $\boldsymbol{v} \in \mathbb{R}^N$, $|\boldsymbol{v}| = \sum_{n=1}^N v_n$, $\max(\boldsymbol{v}) = \max_{n=1,\dots,N} v_n$ and $\min(\boldsymbol{v}) = \min_{n=1,\dots,N} v_n$;
- given $\boldsymbol{v} \in \mathbb{R}^N$ and $f : \mathbb{R} \to \mathbb{R}$, $f(\boldsymbol{v})$ denotes the vector obtained by applying $f$ to each component of $\boldsymbol{v}$, $f(\boldsymbol{v}) = [f(v_1), f(v_2), \dots, f(v_N)] \in \mathbb{R}^N$;
- given $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^N$, the inequality $\boldsymbol{v} > \boldsymbol{w}$ holds true if and only if $v_n > w_n \ \forall n = 1, \dots, N$.
- given $\boldsymbol{v} \in \mathbb{R}^D$ and $\boldsymbol{w} \in \mathbb{R}^N$, $[\boldsymbol{v}, \boldsymbol{w}] = (v_1, \dots, v_D, w_1, \dots, w_N) \in \mathbb{R}^{D+N}$.

## 2. Problem setting

Let $\mathcal{B} \subset \mathbb{R}^d$, $d = 1, 2, 3$, be an open hyper-rectangular domain (referred to hereafter as the "physical domain") and let $\boldsymbol{y} = (y_1, y_2, \dots, y_N)$ be a $N$-dimensional random vector whose components are mutually independent and uniformly distributed random variables with support $\Gamma_n \subset \mathbb{R}$ and probability density function $\rho_n(y_n) = \frac{1}{|\Gamma_n|}$. Denoting $\Gamma = \Gamma_1 \times \Gamma_2 \cdots \times \Gamma_N$ (referred to hereafter as the "stochastic domain" or "parameter space") and by $\sigma_B(\Gamma)$ the Borel $\sigma$-algebra over $\Gamma$, $\rho(\boldsymbol{y})d\boldsymbol{y} = \prod_{n=1}^N \rho_n(y_n)dy_n$ is therefore a probability measure on $\Gamma$, due to the independence of $y_n$, and $(\Gamma, \sigma_B(\Gamma), \rho(\boldsymbol{y})d\boldsymbol{y})$ is a complete probability space. Consider the following generic PDE, together with the assumption stated next:

**Problem 1.** Find $u : \mathcal{B} \times \Gamma \to \mathbb{R}$ such that for $\rho$-almost every $\boldsymbol{y} \in \Gamma$

$$\begin{cases} \mathcal{L}(u; \boldsymbol{x}, \boldsymbol{y}) = \mathcal{F}(\boldsymbol{x}) & \boldsymbol{x} \in \mathcal{B}, \\ u(\boldsymbol{x}, \boldsymbol{y}) = 0 & \boldsymbol{x} \in \partial\mathcal{B}. \end{cases}$$

**Assumption 1** (*Well Posedness*)**.** Problem 1 is well posed in some Hilbert space $V$ for $\rho$-almost every $\boldsymbol{y} \in \Gamma$.

The solution of Problem 1 can be seen as an $N$-variate Hilbert-space valued function $u(\boldsymbol{y}) : \Gamma \to V$. The random variables, $y_n$, can represent scalar values whose exact value is unknown, or they can stem from a spectral decomposition of a random field, like a Karhunen–Loève or Fourier expansion, possibly truncated after a finite number of terms, see, e.g., [6,36]. It is also useful to introduce the Bochner space $L^2_\rho(\Gamma; V) = \left\{u : \Gamma \to V \text{ strongly measurable s.t. } \int_\Gamma \|u(\boldsymbol{y})\|_V^2 \rho(\boldsymbol{y}) d\boldsymbol{y} < \infty\right\}$. Finally, given some functional of the solution $u$, $\Theta : V \to \mathbb{R}$, we denote by $F : \Gamma \to \mathbb{R}$ the $N$-variate real-valued function assigning to each realization $\boldsymbol{y} \in \Gamma$ the corresponding value of $\Theta[u]$ (quantity of interest), i.e., $F(\boldsymbol{y}) = \Theta[u(\cdot, \boldsymbol{y})]$, and we aim at estimating its expected value,

$$\mathbb{E}[F] = \int_\Gamma F(\boldsymbol{y})\rho(\boldsymbol{y})d\boldsymbol{y}.$$

**Example 1.** As a motivating example, consider the following elliptic problem: find $u : \mathcal{B} \times \Gamma \to \mathbb{R}$ such that for $\rho$-almost every $\boldsymbol{y} \in \Gamma$

$$\begin{cases} -\text{div}(a(\boldsymbol{x}, \boldsymbol{y})\nabla u(\boldsymbol{x}, \boldsymbol{y})) = \mathcal{F}(\boldsymbol{x}) & \boldsymbol{x} \in \mathcal{B}, \\ u(\boldsymbol{x}, \boldsymbol{y}) = h(\boldsymbol{x}) & \boldsymbol{x} \in \partial\mathcal{B}, \end{cases} \tag{3}$$

holds, where div and $\nabla$ denote differentiation with respect to the physical variables, $\boldsymbol{x}$, only, and the function $a : \mathcal{B} \times \Gamma \to \mathbb{R}$ is bounded away from 0 and $\infty$, i.e., there exist two constants, $a_{\min}$, $a_{\max}$, such that

$$0 < a_{\min} \le a(\boldsymbol{x}, \boldsymbol{y}) \le a_{\max} < \infty, \quad \forall \boldsymbol{x} \in \mathcal{B} \text{ and for } \rho\text{-almost every } \boldsymbol{y} \in \Gamma. \tag{4}$$

This boundedness condition guarantees that Assumption 1 is satisfied, i.e. the equation is well posed for $\rho$-almost every $\boldsymbol{y} \in \Gamma$, thanks to a straightforward application of the Lax–Milgram lemma; moreover, the equation is well posed in $L^2_\rho(\Gamma; V)$, where $V$ is the classical Sobolev space $H^1_0(\mathcal{B})$, see, e.g., [6]. This is the example we will focus on in Section 5, where we will test numerically the performance of the Multi-Index Stochastic Collocation method that we will detail in Section 3.

**Remark 1.** The method that we present in the following sections can be also applied to more general problems than Problem 1 in which the forcing terms, boundary conditions and possibly domain shape are also modeled as uncertain; the extension to time-dependent problems with uncertain initial conditions is also straightforward. Other probability measures can also be considered; the very relevant case in which the random variables, $y_n$, are normally distributed is an example.

**Remark 2.** As will be clearer in a moment, the methodology we propose uses tensorized solvers for deterministic PDEs. Although for ease of exposition we have assumed that the spatial domain, $\mathcal{B}$, is a hyper-rectangle, it is important to remark that the methodology proposed in this work can also be applied to non hyper-rectangular domains: this can be achieved by introducing a mapping from a reference hyper-rectangle to the generic domain of interest (with techniques such as those proposed in the context of Isogeometric Analysis [39] or Transfinite Interpolation [40]) or by a Domain Decomposition approach [41] if the domain can be obtained as a union of hyper-rectangles.

### 2.1. Approximation along the deterministic and stochastic dimensions

In practice, we can only access the value of $F$ via a numerical solver yielding a numerical approximation of the solution $u$ of Problem 1, which depends on a set of $D$ discretization parameters, such as the mesh-size, the time-step, the tolerances of the numerical solvers, and others, which we denote by $h_i$, $i = 1, \dots, D$; we remark that in general $D$, the number of parameters, might be different from $d$, the number of spatial dimensions. For each of those parameters, we introduce a sequence of discretization levels, $h_{i,\alpha}$, $\alpha = 1, 2, \dots$, and for each multi-index $\boldsymbol{\alpha} \in \mathbb{N}_+^D$, we denote by $u^{\boldsymbol{\alpha}}(\boldsymbol{x}, \boldsymbol{y})$ the approximation of $u$ obtained from setting $h_i = h_{i,\alpha_i}$, with the implicit assumption that $u^{\boldsymbol{\alpha}}(\boldsymbol{x}, \boldsymbol{y}) \to u(\boldsymbol{x}, \boldsymbol{y})$ as $\min_{1 \le i \le D} \alpha_i \to \infty$ for $\rho$-almost every $\boldsymbol{y} \in \Gamma$; similarly, we also write $F^{\boldsymbol{\alpha}}(\boldsymbol{y}) = \Theta[u^{\boldsymbol{\alpha}}(\cdot, \boldsymbol{y})]$. For instance, we could solve the problem stated in Example 1 by a finite differences scheme with grid-sizes $h_{i,\alpha_i} = h_0 2^{-\alpha_i}$ in direction $i = 1, \dots, D$, for some $h_0 > 0$.

The discretization of $F^{\boldsymbol{\alpha}}$ over the random parameter space $\Gamma$ will consist of a suitable linear combination of tensor interpolants over $\Gamma$ based on Lagrangian polynomials. Observe that this approach is sound only if $F^{\boldsymbol{\alpha}}$ is at least

a continuous function over $\Gamma$ (the smoother $F^{\boldsymbol{\alpha}}$ is, the more effective the Lagrangian approximation will be); for instance, for the problem stated in Example 1, it can be shown under moderate assumptions on $a(\boldsymbol{x}, \boldsymbol{y})$ that $F$ and $F^{\boldsymbol{\alpha}}$ are $\boldsymbol{y}$-analytic, see, e.g., [35,16]; we will return to this point in Section 5.

To derive a generic tensor Lagrangian interpolation of $F^{\boldsymbol{\alpha}}$, we first introduce the set $\mathcal{C}^0(\Gamma_n)$ of real-valued continuous functions over $\Gamma_n$, and the subspace of polynomials of degree at most $q$ over $\Gamma_n$, $\mathbb{P}^q(\Gamma_n) \subset \mathcal{C}^0(\Gamma_n)$. Next, we consider a sequence of univariate Lagrangian interpolant operators in each dimension $Y_n$, i.e., $\{\mathcal{U}_n^{m(\beta_n)}\}_{\beta_n \in \mathbb{N}_+}$, where we refer to the value $\beta_n$ as the "interpolation level". Each interpolant is built over a set of $m(\beta_n)$ collocation points, $\mathcal{H}_n^{m(\beta_n)} = \{y_n^1, y_n^2 \dots y_n^{m(\beta_n)}\} \subset \Gamma_n$, where $m$ is a strictly increasing function, with $m(0) = 0$ and $m(1) = 1$, that we call the "level-to-nodes function"; thus, the interpolant yields a polynomial approximation,

$$\mathcal{U}_n^{m(\beta_n)} : \mathcal{C}^0(\Gamma_n) \to \mathbb{P}^{m(\beta_n)-1}(\Gamma_n), \qquad \mathcal{U}_n^{m(\beta_n)}[f](y_n) = \sum_{j=1}^{m(\beta_n)} \left( f(y_n^j) \prod_{k=1, k\neq j}^{m(\beta_n)} \frac{y_n - y_n^k}{y_n^j - y_n^k} \right),$$

with the convention that $\mathcal{U}_n^0[f] = 0 \ \forall f \in \mathcal{C}^0(\Gamma_n)$.

The $N$-variate Lagrangian interpolant can then be built by a tensorization of univariate interpolants: denote by $\mathcal{C}^0(\Gamma)$ the space of real-valued $N$-variate continuous functions over $\Gamma$ and by $\mathbb{P}^{\boldsymbol{q}}(\Gamma) = \bigotimes_{n=1}^N \mathbb{P}^{q_n}(\Gamma_n)$ the subspace of polynomials of degree at most $q_n$ over $\Gamma_n$, with $\boldsymbol{q} = (q_1, \dots, q_N) \in \mathbb{N}^N$, and consider a multi-index $\boldsymbol{\beta} \in \mathbb{N}_+^N$ assigning the interpolation level in each direction, $y_n$; the multivariate interpolant can then be written as

$$\mathcal{U}^{m(\boldsymbol{\beta})} : \mathcal{C}^0(\Gamma) \to \mathbb{P}^{m(\boldsymbol{\beta})-\mathbf{1}}(\Gamma), \qquad \mathcal{U}^{m(\boldsymbol{\beta})}[F^{\boldsymbol{\alpha}}](\boldsymbol{y}) = \left( \mathcal{U}_1^{m(\beta_1)} \otimes \cdots \otimes \mathcal{U}_N^{m(\beta_N)} \right) [F^{\boldsymbol{\alpha}}](\boldsymbol{y}).$$

The set of collocation points needed to build the tensor interpolant $\mathcal{U}^{m(\boldsymbol{\beta})}[u](\boldsymbol{y})$ is the tensor grid $\mathcal{T}^{m(\boldsymbol{\beta})} = \times_{n=1}^N \mathcal{H}_n^{m(\beta_n)}$ with cardinality $\#\mathcal{T}^{m(\boldsymbol{\beta})} = \prod_{n=1}^N m(\beta_n)$. Observe that the Lagrangian interpolant immediately induces an $N$-variate quadrature formula, $\mathcal{Q}^{m(\boldsymbol{\beta})} : \mathcal{C}^0(\Gamma) \to \mathbb{R}$,

$$\mathcal{Q}^{m(\boldsymbol{\beta})}[F^{\boldsymbol{\alpha}}] = \mathbb{E}[\mathcal{U}^{m(\boldsymbol{\beta})}[F^{\boldsymbol{\alpha}}](\boldsymbol{y})] = \sum_{j=1}^{\#\mathcal{T}^{m(\boldsymbol{\beta})}} F^{\boldsymbol{\alpha}}(\widehat{\boldsymbol{y}}_j) \varpi_j,$$

where $\widehat{\boldsymbol{y}}_j \in \mathcal{T}^{m(\boldsymbol{\beta})}$ and the quadrature weights $\varpi_j$ are the expected values of the Lagrangian polynomials centered in $\widehat{\boldsymbol{y}}_j$, which can be computed exactly for most of the common interpolation knots and probability measures of the random variables.

It is recommended that the collocation points $\mathcal{H}_n^{m(\beta_n)}$ to be used in each direction are chosen according to the underlying probability measure, $\rho(y_n)dy_n$, to ensure good approximation properties of the interpolant and quadrature operators, $\mathcal{U}^{m(\boldsymbol{\beta})}$ and $\mathcal{Q}^{m(\boldsymbol{\beta})}$. Common choices are Gaussian quadrature points like Gauss–Legendre for uniform measures or Gauss–Hermite for Gaussian measures, cf. e.g., [42], which are however *not nested*, i.e., $\mathcal{H}_n^{m(\beta_n)} \not\subset \mathcal{H}_n^{m(\beta_n+1)}$. This means that they are not optimal for successive refinements of the interpolation/quadrature, and we will not consider them in this work. Instead, we will work with *nested* collocation points, and specifically with Clenshaw–Curtis points [34,43], that are a classical choice for the uniform measure that we are considering here; other choices of nested points are available for uniform random variables, e.g., the Leja points [34,44], whose performance is somehow equivalent to that of Clenshaw–Curtis for quadrature purposes, see [45,46]. Clenshaw–Curtis points are defined as

$$y_n^j = \cos\left( \frac{(j-1)\pi}{m(i_n) - 1} \right), \quad 1 \le j \le m(i_n), \tag{5}$$

together with the following level-to-nodes relation, $m(i_n)$, that ensures their nestedness:

$$m(0) = 0, \qquad m(1) = 1, \qquad m(i_n) = 2^{i_n - 1} + 1. \tag{6}$$

We conclude this section by introducing the following operator norm, which acts as a "Lebesgue constant" from $\mathcal{C}^0(\Gamma)$ to $L^2_\rho(\Gamma)$:

$$\mathbb{M}^{m(\boldsymbol{\beta})} = \prod_{n=1}^{N} \mathbb{M}_n^{m(\beta_n)}, \quad \text{with } \mathbb{M}_n^{m(\beta_n)} = \sup_{\|f\|_{L^\infty(\Gamma_n)}=1} \|\mathcal{U}_n^{m(\beta_n)} f\|_{L^2_\rho(\Gamma_n)}. \tag{7}$$

In particular, for the Clenshaw–Curtis points, it is possible to bound $\mathbb{M}_n^{m(\beta_n)}$ as:

$$\mathbb{M}^{m(\boldsymbol{\beta})} \leq \mathbb{M}_{est}^{m(\boldsymbol{\beta})} = \prod_{n=1}^{N} \mathbb{M}_{n,est}^{m(\beta_n)}, \qquad \mathbb{M}_{n,est}^{q} = \begin{cases} 1 & \text{for } q = 1 \\ \frac{2}{\pi}\log(q-1)+1 & \text{for } q \geq 2. \end{cases} \tag{8}$$

See [34] and references therein.

**Remark 3.** Nested collocation points have been studied also for other probability measures than uniform probability measures. In the very relevant case of a normal distribution, one possible choice is the Genz–Keister points [47,36]; we mention also the recent work [46] on generalized Leja points that can be used for arbitrary measures on unbounded domains.

## 3. Multi-index stochastic collocation

It is easy to see that an accurate approximation of $\mathbb{E}[F]$ by a direct tensor technique as the one just introduced, $\mathbb{E}[F] \approx \mathcal{Q}^{m(\boldsymbol{\beta})}[F^{\boldsymbol{\alpha}}]$, might require a prohibitively large computational effort even for moderate values of $D$ and $N$ (what is referred to as the "curse of dimensionality"). In this work, following the setting that was presented in [34,29], we propose the Multi-Index Stochastic Collocation as an alternative. It can be seen as a generalization of the telescoping sum presented in the introduction, see Eqs. (1) and (2). Denoting $\mathcal{Q}^{m(\boldsymbol{\beta})}[F^{\boldsymbol{\alpha}}] = F_{\boldsymbol{\alpha},\boldsymbol{\beta}}$, the building blocks of such a telescoping sum are the first-order difference operators for the deterministic and stochastic discretization parameters, denoted respectively by $\Delta_i^{\text{det}}$ with $1 \leq i \leq D$ and $\Delta_j^{\text{stoc}}$ with $1 \leq j \leq N$:

$$\Delta_i^{\text{det}}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}] = \begin{cases} F_{\boldsymbol{\alpha},\boldsymbol{\beta}} - F_{\boldsymbol{\alpha}-\boldsymbol{e}_i,\boldsymbol{\beta}}, & \text{if } \alpha_i > 1, \\ F_{\boldsymbol{\alpha},\boldsymbol{\beta}} & \text{if } \alpha_i = 1, \end{cases} \tag{9}$$

$$\Delta_j^{\text{stoc}}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}] = \begin{cases} F_{\boldsymbol{\alpha},\boldsymbol{\beta}} - F_{\boldsymbol{\alpha},\boldsymbol{\beta}-\boldsymbol{e}_j}, & \text{if } \beta_j > 1, \\ F_{\boldsymbol{\alpha},\boldsymbol{\beta}} & \text{if } \beta_j = 1. \end{cases} \tag{10}$$

We then define the first-order tensor difference operators,

$$\boldsymbol{\Delta}^{\text{det}}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}] = \bigotimes_{i=1}^{D} \Delta_i^{\text{det}}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}] = \Delta_1^{\text{det}}\left[\Delta_2^{\text{det}}\left[\dots \Delta_D^{\text{det}}\left[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}\right]\right]\right] = \sum_{\boldsymbol{j}\in\{0,1\}^D} (-1)^{|\boldsymbol{j}|} F_{\boldsymbol{\alpha}-\boldsymbol{j},\boldsymbol{\beta}}, \tag{11}$$

$$\boldsymbol{\Delta}^{\text{stoc}}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}] = \bigotimes_{j=1}^{N} \Delta_j^{\text{stoc}}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}] = \sum_{\boldsymbol{j}\in\{0,1\}^N} (-1)^{|\boldsymbol{j}|} F_{\boldsymbol{\alpha},\boldsymbol{\beta}-\boldsymbol{j}} \tag{12}$$

with the convention that $F_{\boldsymbol{\nu},\boldsymbol{w}} = 0$ whenever a component of $\boldsymbol{\nu}$ or $\boldsymbol{w}$ is zero. Observe that computing $\Delta^{\text{det}}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]$ actually requires up to $2^D$ solver calls, and analogously applying $\Delta^{\text{stoc}}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]$ requires interpolating $F^{\boldsymbol{\alpha}}$ on up to $2^N$ tensor grids; for instance, if $D = N = 2$ and $\boldsymbol{\alpha}, \boldsymbol{\beta} > \mathbf{1}$, we have

$$\boldsymbol{\Delta}^{\text{det}}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}] = \Delta_2^{\text{det}}\left[\Delta_1^{\text{det}}\left[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}\right]\right] = \Delta_2^{\text{det}}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}} - F_{\boldsymbol{\alpha}-\boldsymbol{e}_1,\boldsymbol{\beta}}] = F_{\boldsymbol{\alpha},\boldsymbol{\beta}} - F_{\boldsymbol{\alpha}-\boldsymbol{e}_1,\boldsymbol{\beta}} - F_{\boldsymbol{\alpha}-\boldsymbol{e}_2,\boldsymbol{\beta}} + F_{\boldsymbol{\alpha}-\mathbf{1},\boldsymbol{\beta}},$$
$$\boldsymbol{\Delta}^{\text{stoc}}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}] = F_{\boldsymbol{\alpha},\boldsymbol{\beta}} - F_{\boldsymbol{\alpha},\boldsymbol{\beta}-\boldsymbol{e}_1} - F_{\boldsymbol{\alpha},\boldsymbol{\beta}-\boldsymbol{e}_2} + F_{\boldsymbol{\alpha},\boldsymbol{\beta}-\mathbf{1}}.$$

Finally, letting $\boldsymbol{\Delta}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}] = \boldsymbol{\Delta}^{\text{stoc}}[\boldsymbol{\Delta}^{\text{det}}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]]$, we define the Multi-Index Stochastic Collocation (MISC) estimator of $\mathbb{E}[F]$ as

$$\mathcal{M}_{\mathcal{I}}[F] = \sum_{[\boldsymbol{\alpha},\boldsymbol{\beta}]\in\mathcal{I}} \boldsymbol{\Delta}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}] = \sum_{[\boldsymbol{\alpha},\boldsymbol{\beta}]\in\mathcal{I}} c_{\boldsymbol{\alpha},\boldsymbol{\beta}} F_{\boldsymbol{\alpha},\boldsymbol{\beta}}, \tag{13}$$

where $\mathcal{I} \subset \mathbb{N}_+^{D+N}$ and $c_{\boldsymbol{\alpha},\boldsymbol{\beta}} \in \mathbb{Z}$. Observe that many of the coefficients in (13), $c_{\boldsymbol{\alpha},\boldsymbol{\beta}}$, may be zero: in particular, $c_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ is zero whenever $[\boldsymbol{\alpha}, \boldsymbol{\beta}] + \boldsymbol{j} \in \mathcal{I}\, \forall \boldsymbol{j} \in \{0, 1\}^{N+D}$. Similarly to the analogous sparse grid construction [34,48,7], we shall require that the multi-index set $\mathcal{I}$ be downward closed, i.e.,

$$\forall\, [\boldsymbol{\alpha}, \boldsymbol{\beta}] \in \mathcal{I}, \quad \begin{cases} \boldsymbol{\alpha} - \boldsymbol{e}_i \in \mathcal{I} & \text{for } 1 \le i \le D \text{ and } \alpha_i > 1, \\ \boldsymbol{\beta} - \boldsymbol{e}_j \in \mathcal{I} & \text{for } 1 \le j \le N \text{ and } \beta_j > 1. \end{cases}$$

**Remark 4.** In theory, a MISC approach could also be developed to approximate the entire solution $u(\boldsymbol{x}, \boldsymbol{y})$ and not just the expectation of functionals, considering differences between consecutive interpolant operators, $\mathcal{U}^{m(\boldsymbol{\beta})}$, on the stochastic domain rather than differences of the quadrature operators, $\mathcal{Q}^{m(\boldsymbol{\beta})}$, as a building block for the $\boldsymbol{\Delta}^{\text{stoc}}$ operators, as well as considering the discretized solution $u^{\boldsymbol{\alpha}}$ rather than just the quantity of interest, $F^{\boldsymbol{\alpha}}$, in the construction of the $\boldsymbol{\Delta}^{\text{det}}$ operators.

### *3.1. A knapsack-like construction of the set $\mathcal{I}$*

The efficiency of the MISC method in Eq. (13) will heavily depend on the specific choice of the index set, $\mathcal{I}$; in the following, we will first propose a general strategy to derive quasi-optimal sets and then prove in Section 4 a convergence result for such sets under some reasonable assumptions.

To derive an efficient set, $\mathcal{I}$, we recast the problem of its construction as an optimization problem, in the same spirit of [29,34,35,7,37]. We begin by introducing the concepts of "work contribution", $\Delta W_{\boldsymbol{\alpha},\boldsymbol{\beta}}$, and "error contribution", $\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}}$, for each hierarchical surplus operator, $\Delta[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]$. The work contribution measures the computational cost (measured, e.g., as a function of the total number of degrees of freedom, or in terms of computational time) required to add $\boldsymbol{\Delta}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]$ to $\mathcal{M}_{\mathcal{I}}[F]$, i.e., to solve the associated deterministic problems and to compute the corresponding interpolants over the parameter space, cf. Eqs. (11) and (12); the error contribution measures instead how much the error $|\mathbb{E}[F] - \mathcal{M}_{\mathcal{I}}[F]|$ would decrease once the operator $\boldsymbol{\Delta}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]$ has been added to $\mathcal{M}_{\mathcal{I}}[F]$. In formulas, we define

$$\Delta W_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \text{Work}\left[\mathcal{M}_{\mathcal{I}\cup\{[\boldsymbol{\alpha},\boldsymbol{\beta}]\}}[F]\right] - \text{Work}\left[\mathcal{M}_{\mathcal{I}}[F]\right] = \text{Work}\left[\boldsymbol{\Delta}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]\right],$$

so that

$$\text{Work}\left[\mathcal{M}_{\mathcal{I}}[F]\right] = \sum_{[\boldsymbol{\alpha},\boldsymbol{\beta}]\in\mathcal{I}} \Delta W_{\boldsymbol{\alpha},\boldsymbol{\beta}}. \tag{14}$$

Observe that this work definition is sharp only if we think of building the MISC estimator with an incremental approach, i.e., we assume that adding the multi-index $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ to the index set $\mathcal{I}$ would not reduce the work that has to be done to evaluate the MISC estimator on the index set. This implies that one cannot take advantage of the fact that some of the coefficients in (13), $c_{\boldsymbol{\alpha},\boldsymbol{\beta}}$, that are non-zero when considering the set $\mathcal{I}$ could become zero if the MISC estimator is instead built considering the set $\mathcal{I} \cup \{[\boldsymbol{\alpha}, \boldsymbol{\beta}]\}$, hence it would be possible not to compute the corresponding approximations $F_{\boldsymbol{\alpha},\boldsymbol{\beta}}$. This approach is discussed in Section 5.3.

Similarly, we define

$$\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \left|\mathcal{M}_{\mathcal{I}\cup\{[\boldsymbol{\alpha},\boldsymbol{\beta}]\}}[F] - \mathcal{M}_{\mathcal{I}}[F]\right| = \left|\boldsymbol{\Delta}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]\right|.$$

Thus, by construction, the error of the MISC estimator (13) can be bounded as the sum of the error contributions not included in the estimator $\mathcal{M}_{\mathcal{I}}[F]$,

$$\begin{aligned} \text{Error}\left[\mathcal{M}_{\mathcal{I}}[F]\right] = |\mathbb{E}[F] - \mathcal{M}_{\mathcal{I}}[F]| &= \left|\sum_{[\boldsymbol{\alpha},\boldsymbol{\beta}]\notin\mathcal{I}} \boldsymbol{\Delta}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]\right| \\ &\le \sum_{[\boldsymbol{\alpha},\boldsymbol{\beta}]\notin\mathcal{I}} \left|\boldsymbol{\Delta}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]\right| = \sum_{[\boldsymbol{\alpha},\boldsymbol{\beta}]\notin\mathcal{I}} \Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}}. \end{aligned} \tag{15}$$

Consequently, a quasi-optimal set $\mathcal{I}$ can be computed by solving the following "binary knapsack problem" [49]:

$$\text{maximize} \sum_{[\boldsymbol{\alpha},\boldsymbol{\beta}]\in\mathbb{N}_+^{D+N}} \Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}} x_{\boldsymbol{\alpha},\boldsymbol{\beta}}$$

$$\text{such that} \sum_{[\boldsymbol{\alpha},\boldsymbol{\beta}]\in\mathbb{N}_+^{D+N}} \Delta W_{\boldsymbol{\alpha},\boldsymbol{\beta}} x_{\boldsymbol{\alpha},\boldsymbol{\beta}} \leq W_{\max}, \tag{16}$$
$$x_{\boldsymbol{\alpha},\boldsymbol{\beta}} \in \{0, 1\},$$

and setting $\mathcal{I} = \{[\boldsymbol{\alpha}, \boldsymbol{\beta}] \in \mathbb{N}_+^{D+N} : x_{\boldsymbol{\alpha},\boldsymbol{\beta}} = 1\}$. Observe that such a set is only "quasi" optimal since the error decomposition (15) is not an exact representation but rather an upper bound. The optimization problem above is well known to be computationally intractable. Still, an approximate greedy solution (which coincides with the exact solution under certain hypotheses that will be clearer in a moment) can be found if one instead allows the variables $x_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ to assume fractional values, i.e., it is possible to include fractions of multi-indices in $\mathcal{I}$. For this simplified problem, the resulting problem can be solved analytically by the so-called Dantzig algorithm [49]:

1. compute the "profit" of each hierarchical surplus, i.e., the quantity
$$P_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \frac{\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}}}{\Delta W_{\boldsymbol{\alpha},\boldsymbol{\beta}}};$$
2. sort the hierarchical surpluses by decreasing profit;
3. add the hierarchical surpluses to $\mathcal{M}_{\mathcal{I}}[F]$ according to such order until the constraint on the maximum work is fulfilled.

Note that by construction $x_{\boldsymbol{\alpha},\boldsymbol{\beta}} = 1$ for all the multi-indices included in the selection except for the last one, for which $x_{\boldsymbol{\alpha},\boldsymbol{\beta}} < 1$ might hold; in other words, the last multi-index is the only one that might not be taken entirely. However, if this is the case, we assume that we could slightly adjust the value $W_{\max}$, so that all $x_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ have integer values (see also [7]); observe that this integer solution is also the solution of the original binary knapsack problem (16) with the new value of $W_{\max}$ in the work constraint. Thus, if the quantities $\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ and $\Delta W_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ were available, the quasi-optimal index set for the MISC estimator could be computed as

$$\mathcal{I} = \mathcal{I}(\epsilon) = \left\{ [\boldsymbol{\alpha}, \boldsymbol{\beta}] \in \mathbb{N}_+^{D+N} : \frac{\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}}}{\Delta W_{\boldsymbol{\alpha},\boldsymbol{\beta}}} \geq \epsilon \right\}, \tag{17}$$

for a suitable $\epsilon > 0$.

**Remark 5.** The MISC setting could in principle include the Multilevel Stochastic Collocation method proposed in [24] as a special case, by simply considering a discretization of the spatial domain on regular meshes, and letting the diameter of each element (the mesh-size) be the only discretization parameter, i.e., $D = 1$.

However, the sparse grids to be used at each level are determined in [24] by computing the minimal number of collocation points needed to balance the stochastic and spatial error. This is done by relying on sparse grid error estimates; yet, since in general it is not possible to generate a sparse grid with a predefined number of points, some rounding strategy to the sparse grid with the nearest cardinality must be devised, which may affect the optimality of the multilevel strategy. In the present work, we overcome this issue by relying instead on profit estimates to build a set of multi-indices that simultaneously prescribe the spatial discretization and the associated tensor grid in the stochastic variables. Furthermore, only standard isotropic Smolyak sparse grids are considered in the actual numerical experiments in [24] (although in principle anisotropic sparse grids could be used as well, provided that good convergence estimates for such sparse grids are available), while our implementation naturally uses anisotropic stochastic collocation methods at each spatial level.

The MISC approach also includes as a special case the "Sparse Composite Collocation Method" developed in [38], by considering again only one deterministic discretization parameter, i.e., $D = 1$, and then setting

$$\mathcal{I} = \left\{ [\alpha, \boldsymbol{\beta}] \in \mathbb{N}_+^{1+N} : \alpha + \sum_{n=1}^{N} \beta_n \leq w \right\}, \tag{18}$$

with $w \in \mathbb{N}_+$. In other words, the approach in [38] is based neither on profit nor on error balancing.

## 4. Complexity analysis of the MISC method

In this section, we assume suitable models for the error and work contributions, $\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ and $\Delta W_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ (which are numerically verified in Section 5 for the problem in Example 1) and then state our main convergence theorem for the

MISC method built using a particular index set, $\mathcal{I}^*$, which can be regarded as an approximation of the quasi-optimal set introduced in the previous section.

**Assumption 2.** The discretization parameters, $h_i$, for the deterministic solver depend exponentially on the discretization level $\alpha_i$, and the number of collocation points over the parameter space grows exponentially with the level $\beta_i$:

$$h_{i,\alpha_i} = h_0 2^{-\alpha_i} \quad \text{and} \quad C_{m,low} 2^{\beta_i} \leq m(\beta_i) \leq C_{m,up} 2^{\beta_i}.$$

**Assumption 3.** The error and work contributions, $\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ and $\Delta W_{\boldsymbol{\alpha},\boldsymbol{\beta}}$, can be bounded as products of two terms,

$$\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}} \leq \Delta E^{\text{det}}_{\boldsymbol{\alpha}} \Delta E^{\text{stoc}}_{\boldsymbol{\beta}} \quad \text{and} \quad \Delta W_{\boldsymbol{\alpha},\boldsymbol{\beta}} \leq \Delta W^{\text{det}}_{\boldsymbol{\alpha}} \Delta W^{\text{stoc}}_{\boldsymbol{\beta}},$$

where $\Delta W^{\text{det}}_{\boldsymbol{\alpha}}$ and $\Delta E^{\text{det}}_{\boldsymbol{\alpha}}$ denote the cost and the error contributions due to the deterministic difference operator, $\Delta^{\text{det}}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]$, and similarly $\Delta W^{\text{stoc}}_{\boldsymbol{\beta}}$ and $\Delta E^{\text{stoc}}_{\boldsymbol{\beta}}$ denote the cost and the error contribution due to the stochastic difference operator, $\Delta^{\text{stoc}}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]$, cf. Eqs. (11)–(12).

**Assumption 4.** The following bounds hold true for the factors appearing in Assumption 3:

$$\Delta W^{\text{det}}_{\boldsymbol{\alpha}} \leq C^{\text{det}}_{\text{work}} \prod_{i=1}^{D} (h_{i,\alpha_i})^{-\widetilde{\gamma}_i}, \tag{19}$$

$$\Delta E^{\text{det}}_{\boldsymbol{\alpha}} \leq C^{\text{det}}_{\text{error}} \prod_{i=1}^{D} (h_{i,\alpha_i})^{\widetilde{r}_i}, \tag{20}$$

$$\Delta W^{\text{stoc}}_{\boldsymbol{\beta}} \leq \widetilde{C}^{\text{stoc}}_{\text{work}} \prod_{n=1}^{N} m(\beta_n) \leq C^{\text{stoc}}_{\text{work}} \prod_{n=1}^{N} 2^{\beta_n}, \tag{21}$$

$$\Delta E^{\text{stoc}}_{\boldsymbol{\beta}} \leq C^{\text{stoc}}_{\text{error}} e^{-\sum_{i=1}^{N} \widetilde{g}_i m(\beta_i)}, \tag{22}$$

for some rates $\widetilde{\gamma}_i, \widetilde{r}_i, \widetilde{g}_i > 0$.

With these assumptions, we are now ready to state our main theorem. The proof is technical and we therefore place it in the Appendix. The proof is based on summing the error contributions outside a particular index set, $\mathcal{I}^*$, and the work contributions inside the same index set. This can be seen as a weighted cardinality argument in finite dimensions. See also [50,51] for similar arguments for different choices of finite and infinite dimensional index sets.

**Theorem 1** (*MISC Computational Complexity*)**.** *Under Assumptions 2–4, the bounds for the factors appearing in Assumption 3 can be equivalently rewritten as*

$$\Delta W_{\boldsymbol{\alpha},\boldsymbol{\beta}} \leq C_{\text{work}} e^{\sum_{i=1}^{D} \gamma_i \alpha_i} e^{\delta |\boldsymbol{\beta}|}, \tag{23a}$$

$$\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}} \leq C_{\text{error}} e^{-\sum_{i=1}^{D} r_i \alpha_i} e^{-\sum_{j=1}^{N} g_j \exp(\delta \beta_j)}, \tag{23b}$$

*with* $\gamma_i = \widetilde{\gamma}_i \log 2$, $r_i = \widetilde{r}_i \log 2$, $\delta = \log 2$ *and* $g_i = \widetilde{g}_i C_{m,low}$. *Define the following set*

$$\mathcal{I}^*(L) = \left\{ [\boldsymbol{\alpha}, \boldsymbol{\beta}] \in \mathbb{N}_+^{D+N} : \sum_{i=1}^{D} (r_i + \gamma_i)\alpha_i + \sum_{i=1}^{N} (\delta \beta_i + g_i e^{\delta \beta_i}) \leq L \right\} \quad \text{with } L \in \mathbb{R}_+. \tag{24}$$

*Then there exists a constant* $\mathcal{C}_W$ *such that, for any* $W_{\text{max}}$ *satisfying*

$$W_{\text{max}} \geq \mathcal{C}_W \exp(\chi), \tag{25}$$

*and choosing $L$ as*

$$L = L(W_{\max}) = \frac{1}{\chi}\left(\log\left(\frac{W_{\max}}{\mathcal{C}_W}\right) - (\mathfrak{z}-1)\log\left(\frac{1}{\chi}\log\left(\frac{W_{\max}}{\mathcal{C}_W}\right)\right)\right), \tag{26}$$

*with* $\boldsymbol{\Xi} = \left(\frac{\gamma_1}{\gamma_1+r_1}, \dots, \frac{\gamma_D}{\gamma_D+r_D}\right)$, $\chi = \max(\boldsymbol{\Xi})$, $\zeta = \min_{i=1,\dots,D} \frac{r_i}{\gamma_i}$ *and* $\mathfrak{z} = \#\{i = 1, \dots, D : \frac{r_i}{\gamma_i} = \zeta\}$, *the MISC estimator* $\mathcal{M}_{\mathcal{I}^*(L(W_{\max}))}$ *satisfies*

$$\text{Work}\left[\mathcal{M}_{\mathcal{I}^*(L(W_{\max}))}\right] \leq W_{\max}, \tag{27a}$$

$$\limsup_{W_{\max}\uparrow\infty} \frac{\text{Error}\left[\mathcal{M}_{\mathcal{I}^*(L(W_{\max}))}\right]}{W_{\max}^{-\zeta}\left(\log\left(W_{\max}\right)\right)^{(\zeta+1)(\mathfrak{z}-1)}} = \mathcal{C}_E < \infty. \tag{27b}$$

**Remark 6.** The set $\mathcal{I}^*$ proposed in Theorem 1 can be obtained by assuming that the bounds in Eqs. (23a) and (23b) are actually equalities and by using the definition of the quasi-optimal set (18):

$$\begin{aligned}
\mathcal{I}^* &= \left\{[\boldsymbol{\alpha},\boldsymbol{\beta}] \in \mathbb{N}_+^{D+N} : \frac{\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}}}{\Delta W_{\boldsymbol{\alpha},\boldsymbol{\beta}}} \geq \epsilon\right\} \\
&= \left\{[\boldsymbol{\alpha},\boldsymbol{\beta}] \in \mathbb{N}_+^{D+N} : \frac{e^{-\sum_{i=1}^{D} r_i\alpha_i}\; e^{-\sum_{j=1}^{N} g_j \exp(\delta\beta_j)}}{e^{\sum_{i=1}^{D}\gamma_i\alpha_i}\; e^{\delta|\boldsymbol{\beta}|}} \geq \epsilon\right\} \\
&= \left\{[\boldsymbol{\alpha},\boldsymbol{\beta}] \in \mathbb{N}_+^{D+N} : \sum_{i=1}^{D}(r_i+\gamma_i)\alpha_i + \sum_{i=1}^{N}(\delta\beta_i + g_i e^{\delta\beta_i}) \leq L\right\},
\end{aligned}$$

where the last equality holds with $L = -\log\epsilon$.

**Remark 7.** Refining along the spatial or the stochastic variables has different effects on the error of the MISC estimator. Indeed, due to the double exponential $e^{-\sum_{j=1}^{N} g_j \exp(\delta\beta_j)}$ in (23b), the stochastic contribution to the error will quickly fade to zero, which in turn implies that most of the work will be used to reduce the deterministic error. This is confirmed by the fact that the error convergence rate in Theorem 1 only depends on $\gamma_i$ and $r_i$, i.e., the cost and error rates of the deterministic solver, respectively. This observation coincides with that in [38, page 2299]: "since the stochastic error decreases exponentially, the convergence rate should tend towards the algebraic rate of the spatial discretization [...]; see Proposition 3.8". Compared with the method proposed in [38], MISC takes greater advantage of this fact, since it is based on an optimization procedure, cf. Eq. (17); this performance improvement is well documented by the comparison between the two methods shown in the next section. Fig. 1 shows the multi-indices included in $\mathcal{I}$ according to (24) for increasing values of $L$, for a problem with $N = D = 1$, $\gamma_i = 1$, $r = 2$, and $g = 1.5$: as expected, the shape of $\mathcal{I}$ becomes more and more curved as $L$ grows, due to this lack of balance between the stochastic and deterministic directions.

**Remark 8.** Theorem 1 is only valid in case Assumptions 1–4 are true. In the next section we motivate these assumption for a specific elliptic problem that we use for numerically testing the MISC method. However, we stress that deriving bounds on the error and work contributions is problem-dependent and the corresponding analysis must be carried out in each case. Moreover, under a different set of assumption the complexity theorem would have to be rewritten accordingly.

## 5. Example and numerical evidence

In this section, we test the effectiveness of the MISC approximation on some instances of the general elliptic equation (3) in Example 1; more precisely, we consider a problem with one physical dimension ($d = 1$) as well as a more challenging problem with three dimensions ($d = 3$); in both cases, we set $\mathcal{B} = [0,1]^d$, $\mathcal{F}(\boldsymbol{x}) = 1$. As for the

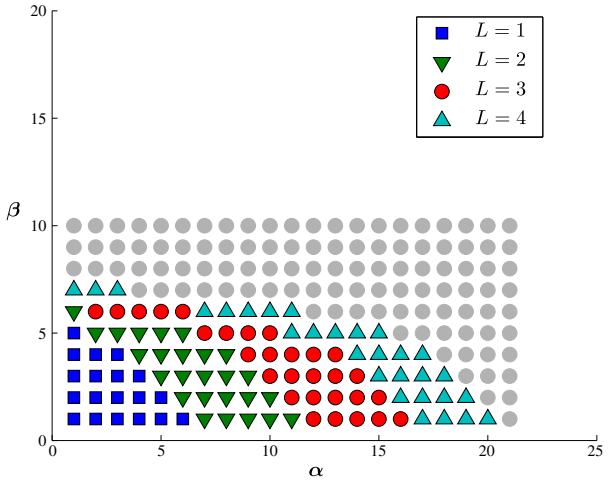



Fig. 1. Index sets $\mathcal{I}(L)$ for $D = N = 1$, according to Eq. (24).

Table 1
Included functions for $d = 3$ in (28). Here $\psi_n(\boldsymbol{x}) = \phi_{i(n)}(x_1)\phi_{j(n)}(x_2)\phi_{k(n)}(x_3)$.

| $n$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $i(n)$ | 1 | 2 | 1 | 1 | 3 | 2 | 2 | 1 | 1 | 1 |
| $j(n)$ | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 3 | 2 | 1 |
| $k(n)$ | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 3 |

random diffusion coefficient, we set

$$a(\boldsymbol{x},\boldsymbol{y}) = e^{\gamma_N(\boldsymbol{x},\boldsymbol{y})}, \quad \gamma_N(\boldsymbol{x},\boldsymbol{y}) = \sum_{n=1}^{N} \lambda_n \psi_n(\boldsymbol{x}) y_n, \tag{28}$$

where $y_n$ are uniform random variables over $\Gamma_n = [-1, 1]$, $\lambda_n = \sqrt{3}\exp(-n)$ and take $\psi_n$ to be a tensorization of trigonometric functions. More precisely, we define the function

$$\phi_n(x) = \begin{cases} \sin\left(\frac{n}{2}\pi x\right) & \text{if } n \text{ is even} \\ \cos\left(\frac{n-1}{2}\pi x\right) & \text{if } n \text{ is odd} \end{cases}$$

and set $\psi_n(x) = \phi_n(x)$ if $d = 1$. If $d = 3$, we take $\psi_n(\boldsymbol{x}) = \phi_{i(n)}(x_1)\phi_{j(n)}(x_2)\phi_{k(n)}(x_3)$ for some indices $i(n), j(n), k(n)$ detailed in Table 1. Observe that the boundedness of the supports of the random variables $y_n$ guarantees the existence of the two bounding constants in Eq. (4), $a_{\min}$ and $a_{\max}$, that in turn assures the well posedness of the problem. Finally, the quantity of interest is defined as

$$F(\boldsymbol{y}) = \int_{\mathcal{B}} u(\boldsymbol{x},\boldsymbol{y}) Q(\boldsymbol{x}) d\boldsymbol{x}, \quad Q(\boldsymbol{x}) = \frac{1}{(\sigma\sqrt{2\pi})^d} \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{x}_0\|_2^2}{2\sigma^2}\right) \tag{29}$$

with $\sigma = 0.16$ and locations $\boldsymbol{x}_0 = 0.3$ for $d = 1$ and $\boldsymbol{x}_0 = [0.3, 0.2, 0.6]$ for $d = 3$. We also make the choice $h_0 = 1/3$ in Assumption 2. With these values and using the coarsest discretization, $h_0 = 1/3$, in all dimensions, the coefficient of variation of the quantity of interest can be approximated to be between 90% and 100% depending on the number of dimensions, $d$, and the number of random variables, $N$, that we consider below.

### 5.1. Verifying bounds on work and error contributions

In this subsection we discuss the validity of Assumptions 2–4, upon which the MISC convergence theorem is based. To this end, we analyze separately the properties of the deterministic solver and of the collocation method applied to the problem just introduced.

*Deterministic solver*. The deterministic solver considered in this work consists of a tensorized finite difference solver, with the grid size along each direction, $x_1, \dots, x_d$, defined by $h_{i,\alpha_i} = h_0 2^{-\alpha_i}$ and no other discretization parameters are considered: therefore, $D = d$, Assumption 2 is satisfied, and, due to the Dirichlet boundary conditions prescribed for $u$, the overall number of degrees of freedom of the corresponding finite difference solution is $\prod_{i=1}^{D}\left(\frac{1}{h_{i,\alpha_i}} - 1\right) \leq \prod_{i=1}^{D}\left(\frac{1}{h_{i,\alpha_i}}\right)$. The associated linear system is solved with the GMRES method. We have numerically fitted the parameters, $\vartheta$ and $C_{\text{GMRES}}$, in the model:

$$\text{Work}\left[F_{\boldsymbol{\alpha}}\right] \leq C_{\text{GMRES}} \prod_{i=1}^{D} (h_{i,\alpha_i})^{-\vartheta},$$

for each individual tensor grid solve and found that $\vartheta = 1$ gives a good fit in our numerical experiments. From this we can recover the rates $\{\tilde{\gamma}_i\}_{i=1}^{D}$ and the constant $C_{\text{work}}^{\text{det}}$ in (19) with the following argument: since computing $\boldsymbol{\Delta}^{\text{det}}[F^{\boldsymbol{\alpha}}]$ requires up to $2^D$ solver calls, each on a different grid (cf. Eq. (11)), we have

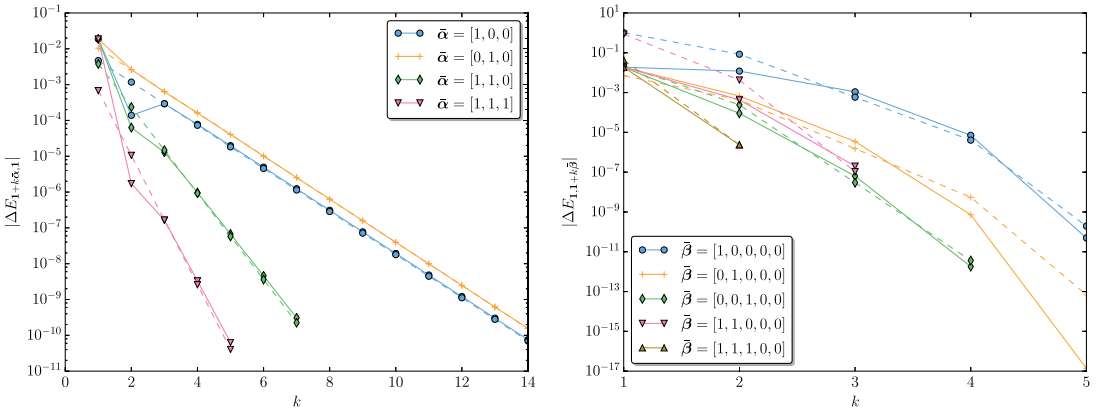
$$\begin{aligned}
\Delta W_{\boldsymbol{\alpha}}^{\text{det}} &= \text{Work}\left[\boldsymbol{\Delta}^{\text{det}}[F_{\boldsymbol{\alpha}}]\right] = \sum_{\boldsymbol{j}\in\{0,1\}^D} \text{Work}\left[F_{\boldsymbol{\alpha}-\boldsymbol{j}}\right] \\
&\leq C_{\text{GMRES}} \sum_{\boldsymbol{j}\in\{0,1\}^D} \prod_{i=1}^{D} \left(h_0 2^{-(\alpha_i - j_i)}\right)^{-\vartheta} \\
&= C_{\text{GMRES}} \left(\prod_{i=1}^{D} \left(h_0 2^{-\alpha_i}\right)^{-\vartheta}\right) \sum_{\boldsymbol{j}\in\{0,1\}^D} \prod_{i=1}^{D} 2^{-j_i\vartheta} \\
&= C_{\text{GMRES}} (1 + 2^{-\vartheta})^D \prod_{i=1}^{D} (h_{i,\alpha_i})^{-\vartheta},
\end{aligned}$$

i.e., bound (19) is verified with $\widetilde{\gamma}_i = \vartheta, \forall i = 1, \dots, D$ and $C_{\text{work}}^{\text{det}} = C_{\text{GMRES}}(1 + 2^{-\vartheta})^D$. Hence, the sum of costs of the solver calls is proportional to the cost of the call on the finest grid.

Concerning the error contribution $\Delta E_{\boldsymbol{\alpha}}^{\text{det}}$, we observe numerically that bound (20) holds true in practice with $\widetilde{r}_i = 2, i = 1, \dots, D$, due to the fact that $a \in C^{\infty}(\mathcal{B})$ for $\rho$-almost every $\boldsymbol{y} \in \Gamma$, $\mathcal{F} \in C^{\infty}(\mathcal{B})$ and the function $Q$ appearing in the quantity of interest (29) is also infinitely differentiable, confined in a small region inside the domain and zero up to machine precision on the boundary. In more detail, assuming for a moment that Assumption 3 is valid (we will numerically verify it later in this section), in Fig. 2(a) we show the value of $\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \Delta E_{\boldsymbol{\alpha}}^{\text{det}} \Delta E_{\boldsymbol{\beta}}^{\text{stoc}}$ for fixed $\boldsymbol{\beta} = \mathbf{1}$ and variable $\boldsymbol{\alpha} = j\bar{\boldsymbol{\alpha}} + \mathbf{1}$, $j = 1, 2, \dots$, as well as the corresponding value of the bound (20) for $\Delta E_{\boldsymbol{\alpha}}^{\text{det}}$. The line obtained by choosing $\bar{\boldsymbol{\alpha}} = [1,\ 0,\ 0]$ confirms that the size of $\Delta E_{\boldsymbol{\alpha}}^{\text{det}}$ indeed decreases exponentially fast with respect to $\alpha_1$, and by fitting the computed values of $\Delta E_{\boldsymbol{\alpha}}^{\text{det}}$ we obtain that the convergence rate is $\widetilde{r}_j = 2$, as previously mentioned; analogous conclusions can be obtained for $\alpha_2$ and $\alpha_3$ by setting $\bar{\boldsymbol{\alpha}} = [0,\ 1,\ 0]$ (shown in Fig. 2(a)) and $\bar{\boldsymbol{\alpha}} = [0,\ 0,\ 1]$ (not shown). Most importantly, confirmation of the product structure of $\Delta E_{\boldsymbol{\alpha}}^{\text{det}}$ can be obtained by observing, e.g., the decay of $\Delta E_{\boldsymbol{\alpha}}^{\text{det}}$ for $\bar{\boldsymbol{\alpha}} = [1,\ 1,\ 0]$ and $\bar{\boldsymbol{\alpha}} = [1,\ 1,\ 1]$.

*Stochastic discretization*. The interpolation over the parameter space is based on the tensorized Lagrangian interpolation technique with Clenshaw–Curtis points explained in Section 2.1, cf. Eqs. (5) and (6). In particular, due to the nestedness of the Clenshaw–Curtis points, adding the operator $\Delta^{\text{stoc}}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]$ to the MISC estimator will require $\Delta W_{\boldsymbol{\beta}}^{\text{stoc}} = \prod_{j=1}^{N}\left(m(\beta_j) - m(\beta_j - 1)\right)$ new collocation points, which in view of Eq. (6) can be bounded as

$$m(\beta_j) - m(\beta_j - 1) = \begin{cases} 1 & \text{if } \beta_j = 1 \\ 2 & \text{if } \beta_j = 2 \\ 2^{\beta_j - 2}, & \text{if } \beta_j > 2, \end{cases} \leq 2^{\beta_j - 1}, \quad \forall j = 1, 2, \dots,$$




(a) For fixed $\boldsymbol{\beta} = \mathbf{1}$ and variable $\boldsymbol{\alpha} = k\bar{\boldsymbol{\alpha}} + \mathbf{1}$.

(b) For fixed $\boldsymbol{\alpha} = \mathbf{1}$ and variable $\boldsymbol{\beta} = k\bar{\boldsymbol{\beta}} + \mathbf{1}$.

Fig. 2. Verifying the validity of the bound (23b) for the value of $|\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}}|$ for the test case with $D = 3$ and $N = 5$. The *dashed* lines are based on the model in (23b) with $\tilde{r}_i = 2$ for all $i = 1, 2, 3$ and $g_j$ as in Table 2 for $j = 1, \ldots, 5$. The *solid* lines are based on computed values.

provided that the set $\mathcal{I}$ is downward closed: Assumption 2 and bound (21) in Assumption 4 are thus verified. Observe that the nestedness of the Clenshaw–Curtis knots is a key property here: indeed, if the nodes are not nested $\Delta W_{\boldsymbol{\beta}}^{\text{stoc}}$ is not uniquely defined, i.e., it depends on the set $\mathcal{I}$ to which $\Delta^{\text{stoc}}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]$ is being added, see, e.g., [34, Example 1 in Section 3].

Finally, to discuss the validity of bound Eq. (22) for $\Delta E_{\boldsymbol{\beta}}^{\text{stoc}}$, we rely on the theory developed in our previous works [35,34]. We begin by introducing the Chebyshev polynomials of the first kind $\Psi_q(y)$ for $y \in [-1, 1]$, which are defined by the relation

$$\Psi_q(\cos(\theta)) = \cos(q\theta), \quad 0 \le \theta \le \pi,\ q \in \mathbb{N}.$$

Then, for any multi-index $\boldsymbol{q} \in \mathbb{N}^N$, we consider the $N$-variate Chebyshev polynomials $\boldsymbol{\Psi}_{\boldsymbol{q}}(\boldsymbol{y}) = \prod_{n=1}^{N} \Psi_{q_n}(y_n)$ and introduce the spectral expansion of $f : [-1, 1]^N \to \mathbb{R}$ over $\{\boldsymbol{\Psi}_{\boldsymbol{q}}\}_{\boldsymbol{q}\in\mathbb{N}^N}$,

$$f(\boldsymbol{y}) = \sum_{\boldsymbol{q}\in\mathbb{N}^N} f_{\boldsymbol{q}}\, \boldsymbol{\Psi}_{\boldsymbol{q}}(\boldsymbol{y}), \quad f_{\boldsymbol{q}} = \int_{\Gamma} f(\boldsymbol{y})\, \boldsymbol{\Psi}_{\boldsymbol{q}}(\boldsymbol{y}) \prod_{n=1}^{N} \frac{1}{\sqrt{1 - y_n^2}} d\boldsymbol{y},$$

Next, given any $\xi_1, \xi_2, \ldots, \xi_N > 1$ we introduce the Bernstein polyellipse $\mathcal{E}_{\xi_1,\ldots,\xi_N} = \prod_{n=1}^{N} \mathcal{E}_{n,\xi_n}$, where $\mathcal{E}_{n,\xi_n}$ denotes the ellipses in the complex plane defined as

$$\mathcal{E}_{n,\xi_n} = \left\{ z_n \in \mathbb{C} :\ \mathfrak{Re}\,(z) \le \frac{\xi_n + \xi_n^{-1}}{2} \cos\phi,\ \ \mathfrak{Im}\,(z) \le \frac{\xi_n - \xi_n^{-1}}{2} \sin\phi,\ \ \phi \in [0, 2\pi) \right\},$$

and recall the following lemma (see [34, Lemma 2] for a proof).

**Lemma 1.** *Let $f : [-1, 1]^N \to \mathbb{R}$, and assume that there exist $\xi_1, \xi_2, \ldots, \xi_N > 1$ such that $f$ admits a complex continuation $f^* : \mathbb{C}^N \to \mathbb{R}$ holomorphic in the Bernstein polyellipse $\mathcal{E}_{\xi_1,\ldots,\xi_N}$ with $\sup_{z\in\mathcal{E}_{\xi_1,\ldots,\xi_N}} |f^*(z)| \le B$ and $B = B(\xi_1, \xi_2, \ldots, \xi_n) < \infty$. Then $f$ admits a Chebyshev expansion that converges in $C^0([-1, 1]^N)$, and whose coefficients $f_{\boldsymbol{q}}$ are such that*

$$|f_{\boldsymbol{q}}| \le C_{\text{Cheb}}(\boldsymbol{q}) \prod_{n=1}^{N} e^{-g_n^* q_n}, \quad g_n^* = \log \xi_n \tag{30}$$

*with $C_{\text{Cheb}}(\boldsymbol{q}) = 2^{\|\boldsymbol{q}\|_0} B$, where $\|\boldsymbol{q}\|_0$ denotes the number of non-zero elements of $\boldsymbol{q}$.*

The following lemma then shows that the region of analyticity of $F(\boldsymbol{y})$ indeed contains a Bernstein ellipse, so that a decay of exponential type can be expected for its Chebyshev coefficients.

**Lemma 2.** *The quantity of interest $F(\boldsymbol{y}) = \Theta[u(\cdot,\boldsymbol{y})]$ is analytic in a Bernstein polyellipse with parameters $\xi_n = \tau_n + \sqrt{\tau_n^2+1}$, for any $\tau_n < \frac{\pi}{2N\lambda_n}$.*

**Proof.** Eq. (3) can be extended in the complex domain by replacing $\boldsymbol{y}$ with $\boldsymbol{z} \in \mathbb{C}^N$, and is analytic in the set $\Sigma = \left\{\boldsymbol{z} \in \mathbb{C}^N : \Re e[a(\boldsymbol{x},\boldsymbol{z})] > 0\right\}$, see, e.g., [6]. By writing $z_n = b_n + ic_n$, we have

$$\begin{aligned} a(\boldsymbol{x},\boldsymbol{z}) &= \exp\left(\sum_n z_n\lambda_n\psi_n(\boldsymbol{x})\right) = \exp\left(\sum_n b_n\lambda_n\psi_n(\boldsymbol{x})\right)\exp\left(\sum_n ic_n\lambda_n\psi_n(\boldsymbol{x})\right) \\ &= \exp\left(\sum_n b_n\lambda_n\psi_n(\boldsymbol{x})\right)\left[\cos\left(\sum_n c_n\lambda_n\psi_n(\boldsymbol{x})\right) + i\sin\left(\sum_n c_n\lambda_n\psi_n(\boldsymbol{x})\right)\right] \end{aligned}$$

so that the region $\Sigma$ can be rewritten as

$$\Sigma = \left\{\boldsymbol{z} = \boldsymbol{b} + i\boldsymbol{c} \in \mathbb{C}^N : \cos\left(\sum_n c_n\lambda_n\psi_n(\boldsymbol{x})\right) > 0, \forall \boldsymbol{x} \in \mathcal{B}\right\}.$$

Such a region includes the smaller region

$$\Sigma_2 = \left\{\boldsymbol{z} = \boldsymbol{b} + i\boldsymbol{c} \in \mathbb{C}^N : \left\|\sum_n c_n\lambda_n\psi_n\right\|_{L^\infty(\mathcal{B})} < \frac{\pi}{2}\right\},$$

which in turn includes

$$\Sigma_3 = \left\{\boldsymbol{z} = \boldsymbol{b} + i\boldsymbol{c} \in \mathbb{C}^N : \sum_n \lambda_n|c_n| < \frac{\pi}{2}\right\},$$

where the last equality is due to the fact that, by construction, $\|\psi_n\|_{L^\infty(\mathcal{B})} = 1$, cf. Eq. (28). Next we let $\tau_n = \frac{\pi}{2N\lambda_n}$ and define the following subregion of $\Sigma_3$:

$$\Sigma_4 = \left\{\boldsymbol{z} = \boldsymbol{b} + i\boldsymbol{c} \in \mathbb{C}^N : |c_n| < \tau_n\right\} \subset \Sigma_3.$$

$\Sigma_4$ is actually a polystrip in the complex plain that it in turn contains the Bernstein ellipse with parameters $\xi_n$ such that

$$\frac{\xi_n - \xi_n^{-1}}{2} = \tau_n \Rightarrow \xi_n^2 - 1 - 2\tau_n\xi_n = 0 \Rightarrow \xi_n = \tau_n + \sqrt{\tau_n^2+1}$$

in which $u(\boldsymbol{x},\boldsymbol{y})$ is analytic. Finally, the quantity of interest, $F = \Theta[u]$, is also analytic in the same Bernstein polyellipse due to the linearity of the operator $\Theta$. □

**Remark 9.** Incidentally, we remark that the choice of $\tau_n$ considered in Lemma 2 degenerates for $N \to \infty$. In this case, if we know that $\sum_{n=0}^{\infty}(\lambda_n\|\psi_n\|_{L^\infty(\mathcal{B})})^p < \infty$ for some $p < 1$, then we could set $\tau_n = \frac{\pi}{2}(\lambda_n\|\psi_n\|_{L^\infty(\mathcal{B})})^{p-1}$, which does not depend on $N$.

**Lemma 3.**

$$\Delta E_{\boldsymbol{\beta}}^{\text{stoc}} \leq C_E e^{-\sum_{n=1}^{N} g_n^* m(\beta_n - 1)} \mathbb{M}^{m(\boldsymbol{\beta})} \tag{31}$$

*holds, where $C_E = 4^N B \prod_{n=1}^N \frac{1}{1-e^{-g_n^*}}$, $B$ as in Lemma 1, $g_n^* = \log\xi_n$ with $\xi_n$ as in Lemma 2, and $\mathbb{M}^{m(\boldsymbol{\beta})}$ has been defined in Eq. (7).*

**Proof.** Combining Lemmas 1 and 2, we obtain that the Chebyshev coefficients of $F$ can be bounded as

$$|F_{\boldsymbol{q}}| \leq C_{\text{Cheb}}(\boldsymbol{q}) \prod_{n=1}^{N} e^{-g_n^* q_n},$$

Table 2
Values of rates $g$ for the test cases considered.

| $g_1$ | $g_2$ | $g_3$ | $g_4$ | $g_5$ | $g_6$ | $g_7$ | $g_8$ | $g_9$ | $g_{10}$ |
|---|---|---|---|---|---|---|---|---|---|
| 2.4855 | 2.8174 | 4.5044 | 4.1938 | 4.7459 | 6.8444 | 7.1513 | 7.8622 | 8.6584 | 9.4545 |

with $g_n^* = \log \xi_n = \log(\tau_n + \sqrt{\tau_n^2 + 1})$ and $\tau_n$ as in Lemma 2. Then, the result can be obtained following the same argument of [34, Lemma 5]. □

To conclude, we first observe that $\mathbb{M}^{m(\boldsymbol{\beta})}$ grows logarithmically with respect to $m(\boldsymbol{\beta})$, see Eq. (8), so it is asymptotically negligible in the estimate above, i.e. we can write

$$\Delta E_{\boldsymbol{\beta}}^{\text{stoc}} \leq C_{E_2}(\epsilon) \prod_{n=1}^{N} e^{-g_n^*(1-\epsilon_E)m(\beta_n-1)}$$

for an arbitrary $\epsilon_E > 0$ and with $C_{E_2}(\epsilon_E) > C_E$, and furthermore that the definition of $m(i)$ in (6) implies that $m(i-1) \geq \frac{m(i)-1}{2}$. We can finally write

$$\Delta E_{\boldsymbol{\beta}}^{\text{stoc}} \leq C_{E_2}(\epsilon) \prod_{n=1}^{N} e^{-g_n^*(1-\epsilon_E)\frac{m(\beta_n)-1}{2}} = C_{\text{error}}^{\text{stoc}} \prod_{n=1}^{N} e^{-\widetilde{g}_n m(\beta_n)},$$

with $C_{\text{error}}^{\text{stoc}} = C(\epsilon) \prod_{n=1}^{N} e^{\frac{g_n^*}{2}(1-\epsilon)}$ and $\widetilde{g}_n = \frac{g_n^*}{2}(1-\epsilon_E)$. The latter bound actually shows that bound (22) in Assumption 4 is valid for the test we are considering. Finally, we point out that in practice we work with the expression (23b), whose rates $g_n$ are actually better estimated numerically, using the same procedure used to obtain the deterministic rates $\tilde{r}_j = 2$: we choose a sufficiently fine spatial resolution level $\boldsymbol{\alpha}$, consider a variable $\boldsymbol{\beta} = j\bar{\boldsymbol{\beta}} + \mathbf{1}$ and fit the (simplified) model $\Delta E_{\boldsymbol{\beta}}^{\text{stoc}} \leq C \prod_{n=1}^{N} e^{-g_n 2^{\beta_n}}$. The values obtained are reported in Table 2, and they are found to be equal for the case $d = 1$ and $d = 3$ (see also [52,35,8]). To make sure that the estimated value of $g_n$ does not depend on the spatial discretization, one could repeat the procedure for a few different values of $\boldsymbol{\alpha}$ and verify that the estimate is robust with respect to the spatial discretization: we note, however, that a rough estimate of $g_n$ will also be sufficient, since the convergence of the method is in practice dictated by the deterministic solver, as we have already discussed in Remark 7. Fig. 2(b) then shows the validity of the bound $\Delta E_{\boldsymbol{\beta}}^{\text{stoc}} \leq C \prod_{n=1}^{N} e^{-g_n 2^{\beta_n}}$ comparing for fixed $\boldsymbol{\alpha} = \mathbf{1}$ and $\boldsymbol{\beta} = j\bar{\boldsymbol{\beta}} + \mathbf{1}$ the value of $\Delta E_{\boldsymbol{\alpha}}^{\text{det}} \Delta E_{\boldsymbol{\beta}}^{\text{stoc}}$ and the corresponding estimate.

*Stochastic–deterministic product structure*. We conclude this section by verifying Assumption 3, i.e., the fact that the error contribution can be factorized as $\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \Delta E_{\boldsymbol{\alpha}}^{\text{det}} \Delta E_{\boldsymbol{\beta}}^{\text{stoc}}$ and that an analogous decomposition holds for $\Delta W_{\boldsymbol{\alpha},\boldsymbol{\beta}}$. While the latter is trivial, to verify the former we employ the same strategy used to verify the models for $\Delta E_{\boldsymbol{\alpha}}^{\text{det}}$ and $\Delta E_{\boldsymbol{\beta}}^{\text{stoc}}$, this time letting both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ change for every point, i.e., $\boldsymbol{\alpha} = j\bar{\boldsymbol{\alpha}} + \mathbf{1}$ and $\boldsymbol{\beta} = j\boldsymbol{\beta}_0 + \mathbf{1}$. Fig. 3 shows the comparison between the computed value of $\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ and their estimated counterpart and confirms the validity of the product structure assumption.

### 5.2. Test setup

In our numerical tests, we compare MISC with the methods listed below. For each of them we show (for both test cases considered) plots of the convergence of the error in the computation of $\mathbb{E}[F]$ with respect to the computational work, taking as a reference value the result obtained using a well-resolved MISC solution. To avoid discrepancies in running time due to implementation details, the computational work is estimated in terms of the total number of degrees of freedom, i.e., using (14) and (23a). The names used here for the methods are also used in the legends of the figures showing the convergence plots.

**"a-priori" MISC** refers to the MISC method with index set $\mathcal{I}$ defined by (17), where $\Delta W_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ and $\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ are taken to equal their upper bounds in (23a) and (23b), respectively. The resulting set is explicitly written in (24). The convergence rate of this set is predicted by Theorem 1, cf. Remark 6. Note that we do not need to determine the value of the constants $C_{\text{work}}$ and $C_{\text{error}}$ since they can be absorbed in the parameter $\epsilon$ in (17).

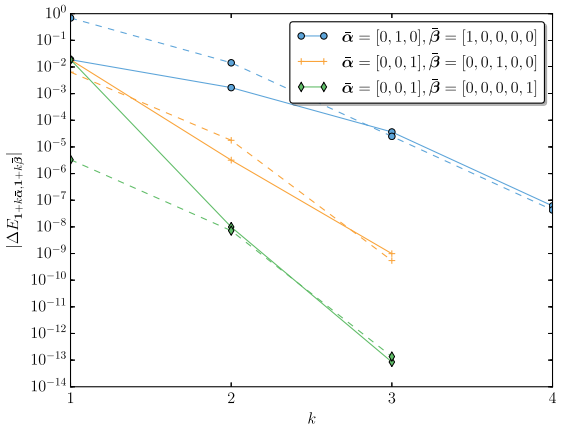



Fig. 3. Comparison of $|\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}}|$ for $\boldsymbol{\beta} = j\bar{\boldsymbol{\beta}} + \mathbf{1}$ and $\boldsymbol{\alpha} = j\bar{\boldsymbol{\alpha}} + \mathbf{1}$ for the test case with $D = 3$ and $N = 5$. The *dashed* lines are based on the model in (23b) with $\tilde{r}_i = 2$ for all $i = 1, 2, 3$ and $g_j$ as in Table 2 for $j = 1, \ldots, 5$. The *solid* lines are based on computed values.

**"a-posteriori" MISC** refers to the MISC method with index set $\mathcal{I}$ defined by (17), where $\Delta W_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ is taken to equal its upper bound in (23a), and $\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ is instead computed explicitly as $\left|\Delta[F_{\alpha,\beta}]\right|$. Notice that this method is not practical since the cost of constructing set $\mathcal{I}$ would dominate the cost of the MISC estimator by far. However, this method would produce the best possible convergence and serve as a benchmark for both "a-priori" MISC and the bound (23b).

**MLSC** (only in the case $d > 1$) refers to the Multilevel Stochastic Collocation obtained by setting $\alpha_1 = \cdots = \alpha_D$ (i.e. considering the mesh-size as the only discretization parameter), as already mentioned in Remark 5; we recall this is not exactly the MLSC method that was implemented in [24], see again Remark 5. Just as with MISC, we consider both the "a-priori" and "a-posteriori" version of MLSC, where $\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ is taken to be equal to its upper (23b) in the former case and assessed by direct numerical evaluation in the latter case.

**SCC** refers to the "Sparse Composite Collocation method" in Remark 5, see Eq. (18).

**MIMC** refers to the Multi-Index Monte Carlo method as detailed in [29], for which the complexity $\mathcal{O}\left(W_{\max}^{-0.5}\right)$ can be estimated for the test case at hand and as long as $d < 4$.

**SGSC** refers to the quasi-optimal Sparse Grids Stochastic Collocation (SGSC) with fixed spatial discretization as proposed in [35,34]. To determine the needed spatial discretization for a given work and for a fair comparison against MISC, we actually compute the convergence curves of SGSC for all relevant levels of spatial discretizations and then show in the plots only the lower envelope of the corresponding convergence curves, ignoring the possible spurious reductions of error that might happen due to non-asymptotic, unpredictable cancellations, cf. Fig. 4. In this way, we ensure that the error shown for such "single-level methods" has been obtained with the smallest computational error possible. Again, this is not a computationally practical method but is taken as a reference for what a sparse grids Stochastic Collocation method with optimal balancing of the space and stochastic discretization errors could achieve.

### 5.3. Implementation details

To implement MISC, we need two components:

1. Given a profit level parameter, $\epsilon$, we build the quasi-optimal set $\mathcal{I}$ based on (17) and (23a) and (23b). One method to achieve this is to exploit the fact that this set is downward closed and use the following recursive algorithm.

```
FUNCTION BuildSet(epsilon, multiIndex)
    FOR i = 1 to (D+N)
        IF Profit(multiIndex + e_i) > epsilon
        THEN
            ADD multiIndex+e_i to FinalSet
            CALL BuildSet(epsilon, multiIndex+e_i)
```

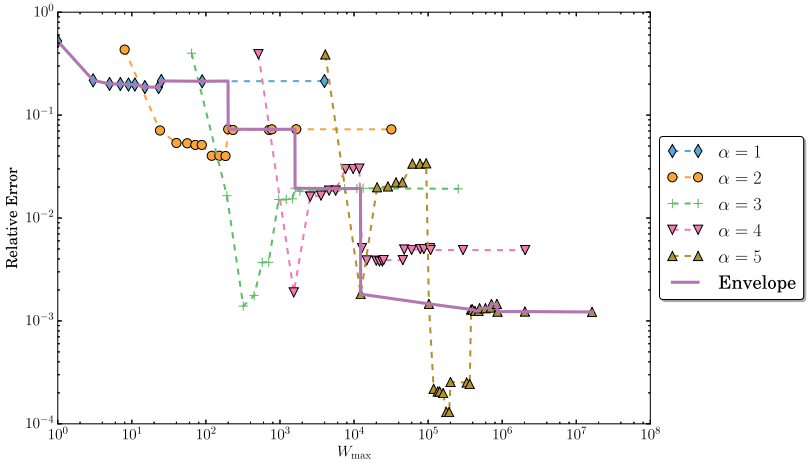



Fig. 4. Envelope of SGSC convergence curves for the test case with $d = 3$ and $N = 10$.

```
        END IF
    END FOR
END FUNCTION
```

2. Given the set, $\mathcal{I}(L)$, we evaluate (13). Here we have two choices:
   - Evaluate the individual terms $\boldsymbol{\Delta}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]$ for every $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathcal{I}$. To do so, we use the operator defined in (9) along each stochastic and spatial direction. By storing the values of these terms, we can evaluate the MISC with different index sets (contained in $\mathcal{I}(L)$), which might be required to test the convergence of the MISC method. Moreover, this implementation is suitable for adaptive methods that expand the index set based on some criteria and reevaluate the MISC estimator. On the other hand, this implementation has a computational overhead since most computed values of $F_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ will actually not contribute to the final value of the estimator. However, this computational overhead is only a fraction of the minimum time required to evaluate the estimator.
   - Use the combination form of (13) and only compute the terms that have $c_{\boldsymbol{\alpha},\boldsymbol{\beta}} \neq 0$. This would remove the overhead of computing terms that make zero contribution to the estimator. This implementation is more efficient but less flexible as we cannot evaluate the estimator on sets contained in $\mathcal{I}(L)$ or build the set adaptively.

### 5.4. Test with $D = 1$

Here we consider three different numbers of stochastic variables, namely $N = 1, 5, 10$. Results are shown in Fig. 5. As expected, a-posteriori MISC shows the best convergence, with a-priori MISC being slightly worse and the single level methods following. Finally, we verify the accuracy of the estimated asymptotic convergence rate provided by Theorem 1: in this case, $\zeta = \frac{r_1}{\gamma_1} = \frac{\widetilde{r}_1 \log 2}{\widetilde{\gamma}_1 \log 2} = 2$ and $\mathfrak{z} = 1$ holds. Hence, the predicted convergence rate is $W^{-\zeta}(\log W)^{(\zeta+1)(\mathfrak{z}-1)} = W^{-2}$, which appears to be in good agreement with the experimental convergence rate.

### 5.5. Test with $D = 3$

In this case, we obtain the convergence curves shown in Fig. 6, where the Multilevel Stochastic Collocation method has also been included. The hierarchy between the methods is in agreement with the case $d = 1$, with the Multilevel Stochastic Collocation being comparable or slightly better than single level methods, but worse than the MISC approaches as expected.

Concerning the accuracy of the theoretical estimate: since for this test $\widetilde{r}_1 = \widetilde{r}_2 = \widetilde{r}_2 = 2$ and $\widetilde{\gamma}_1 = \widetilde{\gamma}_2 = \widetilde{\gamma}_3 = 1$, $\zeta = 2$ still holds, while this time $\mathfrak{z} = 3$; hence, the predicted convergence rate is $W^{-\zeta}(\log W)^{(\zeta+1)(\mathfrak{z}-1)} = W^{-2}(\log W)^6$. The plots suggest that the theoretical estimates might be slightly too optimistic when $N$ increases but it is important to recall that Theorem 1 gives only an asymptotic result, and the plot could be negatively influenced by

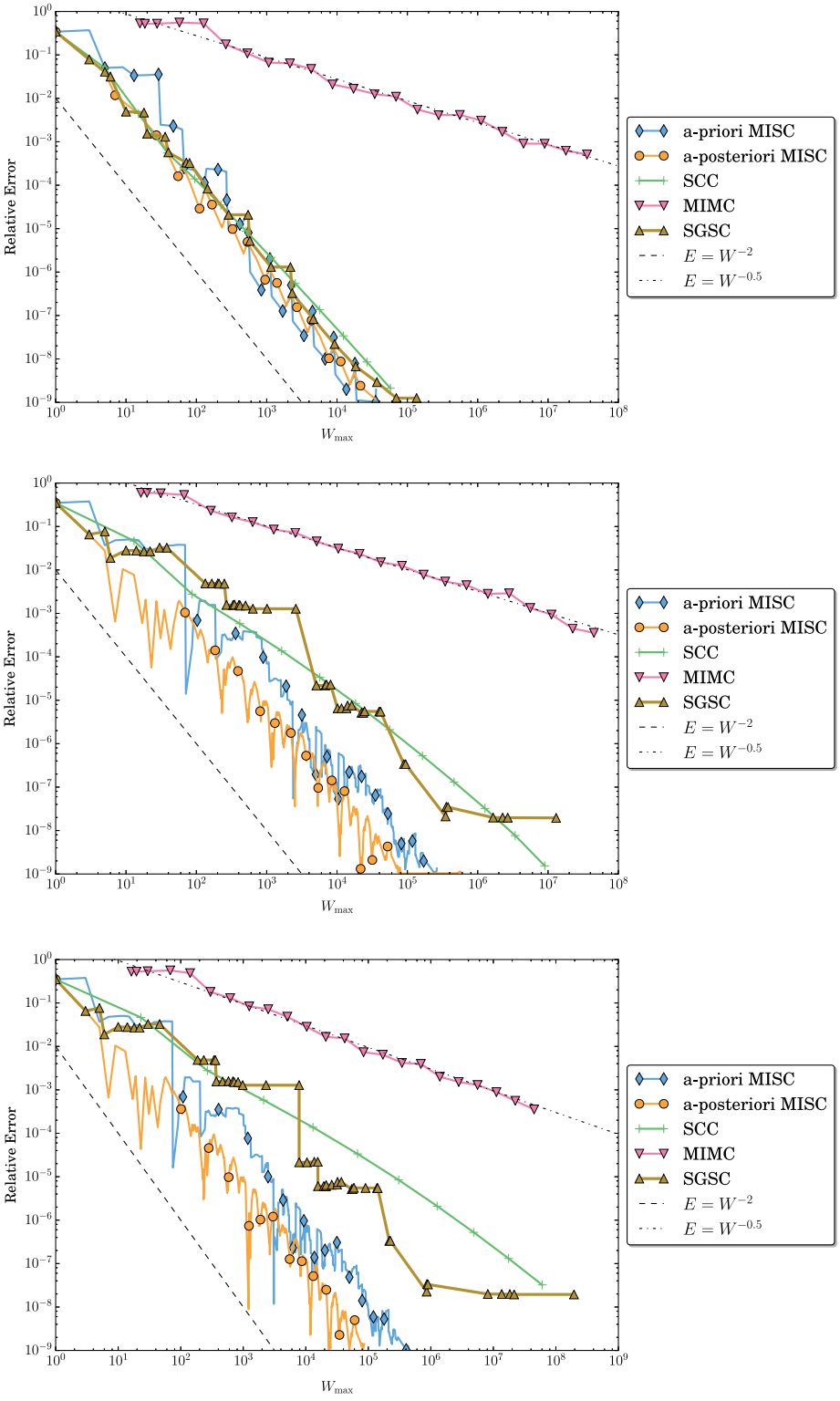



Fig. 5. Results for test $D = 1$, case $N = 1$ (top), $N = 5$ (center) and $N = 10$ (bottom).

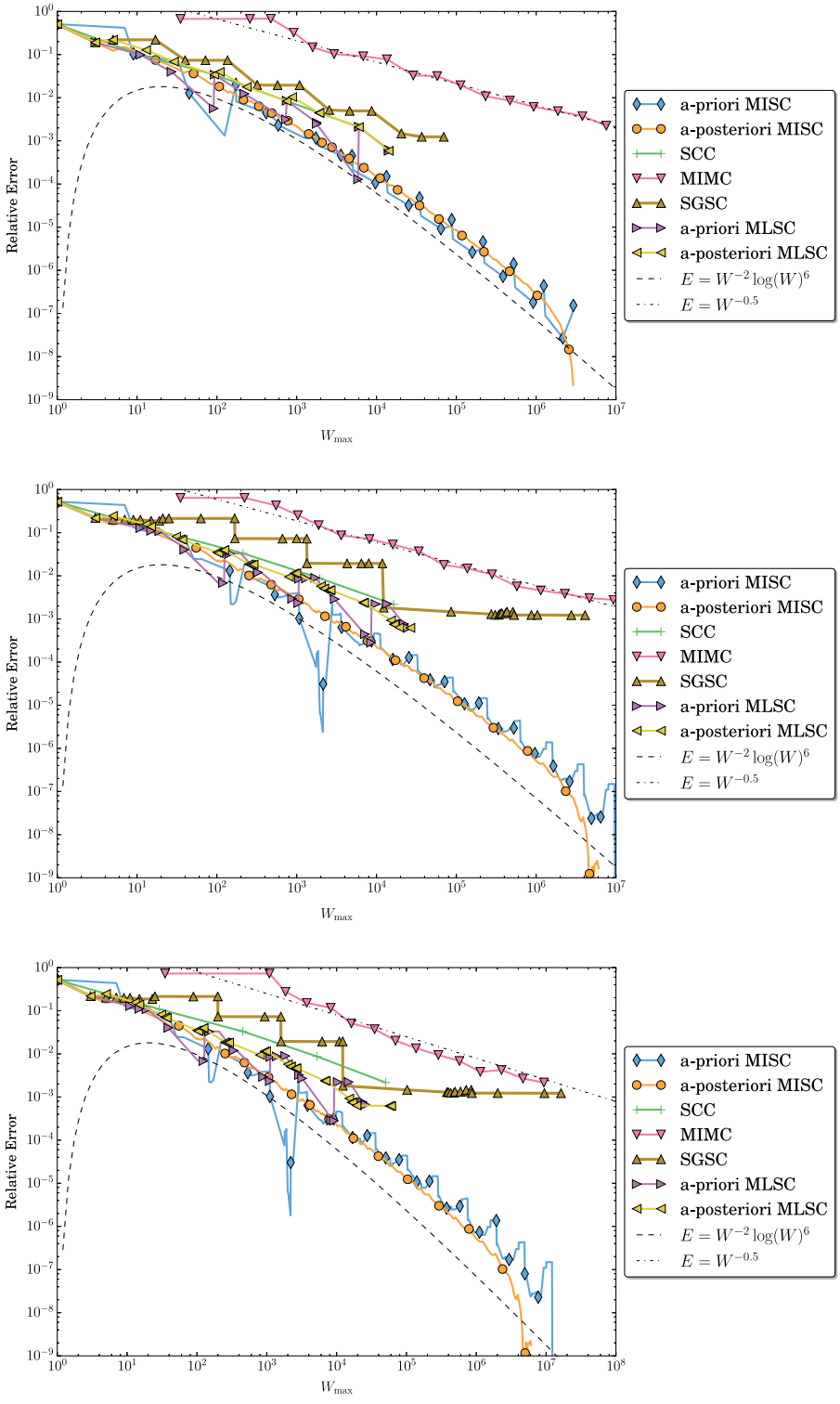



Fig. 6. Results for test $D = 3$, case $N = 1$ (top), $N = 5$ (center) and $N = 10$ (bottom).

pre-asymptotic effects. Observe also that in this case there are a few data points where a-posteriori MISC is not better than a-priori MISC; this observation can be ascribed to the fact that a-posteriori MISC is optimal only with respect to

the upper bound in (15). In other words, a-posteriori MISC selects the contributions according to the absolute value of the contributions but then the MISC estimator is computed by summing signed contributions. Hence, cancellations between contributions with similar sizes and opposite signs will occur.

Finally, we remark that, in our calculations, MLSC and SGSC were not able to achieve very small errors, unlike MISC. This is due to a limitation in the linear solver we are using that allows systems with only up to $2^{17}$ degrees of freedom (around 1 GB of memory) to be solved. These "single-level" methods hit that limit sooner than MISC since they entail solving a very large system that comes from isotropically discretizing all three spatial dimensions.

## 6. Conclusions

In this work, we have proposed MISC, a combination technique method to solve UQ problems, optimizing both the deterministic and stochastic resolution levels simultaneously to minimize the computational cost. A distinctive feature of MISC is that its construction is based on the notion of profit of the mixed differences composing it, rather than balancing the total error contributions arising from the deterministic and stochastic components. We have detailed a complexity analysis and derived a convergence theorem showing that in certain cases the convergence of the method is essentially dictated by the convergence properties of the deterministic solver. We have then verified the effectiveness of the method proposed on a couple of numerical test cases, comparing its performance with other methods available in the literature. The results obtained are encouraging, as they suggest that the proposed methodology is more effective than the other methods considered here. The theoretical results have been also found to be consistent with the numerical results to a satisfactory extent.

As a final remark, we observe that the methodology presented here is not limited to the spatial or temporal discretization parameters of the deterministic problem, but could also be applied to other discretization parameters, such as smoothing parameters or artificial viscosities.

## Acknowledgments

F. Nobile and L. Tamellini received support from the Center for ADvanced MOdeling Science (CADMOS) and partial support by the Swiss National Science Foundation under the Project No. 140574 "Efficient numerical methods for flow and transport phenomena in heterogeneous random porous media". R. Tempone is a member of the KAUST Strategic Research Initiative, Center for Uncertainty Quantification in Computational Sciences and Engineering.

## Appendix. Proof of Theorem 1

The following technical lemmas are needed in the convergence proof.

**Lemma 4.** *For $\boldsymbol{x} \in (1,\infty)^D$, define $\lfloor \boldsymbol{x} \rfloor = (\lfloor x_i \rfloor)_{i=1}^D$. For any $f : (1,\infty)^D \to \mathbb{R}$ and $g : (1,\infty)^D \to \mathbb{R}_+$,*

$$\sum_{\{\boldsymbol{\alpha}\in\mathbb{N}_+^D : f(\boldsymbol{\alpha})\le 0\}} g(\boldsymbol{\alpha}) = \int_{\{\boldsymbol{x}\in(1,\infty)^D : f(\lfloor \boldsymbol{x}\rfloor)\le 0\}} g(\lfloor \boldsymbol{x}\rfloor)\, d\boldsymbol{x}$$

*holds. Moreover, if $g$ and $f$ are increasing, then*

$$\sum_{\{\boldsymbol{\alpha}\in\mathbb{N}_+^D : f(\boldsymbol{\alpha})\le 0\}} g(\boldsymbol{\alpha}) \le \int_{\{\boldsymbol{x}\in(1,\infty)^D : f(\boldsymbol{x}-1)\le 0\}} g(\boldsymbol{x})\, d\boldsymbol{x},$$

*and if $g$ and $f$ are decreasing, then*

$$\sum_{\{\boldsymbol{\alpha}\in\mathbb{N}_+^D : f(\boldsymbol{\alpha})\le 0\}} g(\boldsymbol{\alpha}) \le \int_{\{\boldsymbol{x}\in(1,\infty)^D : f(\boldsymbol{x})\le 0\}} g(\boldsymbol{x}-1)\, d\boldsymbol{x}.$$

**Proof.** We have

$$\begin{aligned}\sum_{\{\boldsymbol{\alpha}\in\mathbb{N}_+^D: f(\boldsymbol{\alpha})\le 0\}} g(\boldsymbol{\alpha}) &= \sum_{\{\boldsymbol{\alpha}\in\mathbb{N}_+^D: f(\boldsymbol{\alpha})\le 0\}} g(\boldsymbol{\alpha}) \int_{\boldsymbol{x}\in[0,1]^D} \mathrm{d}\boldsymbol{x} \\ &= \sum_{\{\boldsymbol{\alpha}\in\mathbb{N}_+^D: f(\boldsymbol{\alpha})\le 0\}} \int_{\boldsymbol{x}\in[0,1]^D} g(\lfloor \boldsymbol{\alpha}+\boldsymbol{x}\rfloor)\, \mathrm{d}\boldsymbol{x} \\ &= \int_{\{\boldsymbol{x}\in(1,\infty)^D: f(\lfloor \boldsymbol{x}\rfloor)\le 0\}} g(\lfloor \boldsymbol{x}\rfloor)\, \mathrm{d}\boldsymbol{x}.\end{aligned}$$

Combining these inequalities with $x-1\le \lfloor x\rfloor \le x$ finishes the proof. □

**Lemma 5.** *Assume $\boldsymbol{a}\in\mathbb{R}_+^D$, $\boldsymbol{b}\in\mathbb{R}_+^D$ and $L>|\boldsymbol{a}|$. Then,*

$$\sum_{\{\boldsymbol{x}\in\mathbb{N}_+^D:\sum_{i=1}^D a_i e^{b_i x_i}+b_i x_i>L\}} \exp\left(-\sum_{i=1}^D a_i e^{b_i x_i}\right) \le \left(\prod_{i=1}^D \frac{\exp(2a_i)}{a_i^2}\right)\exp(-L)\,(L+1)^{2D+1}.$$

**Proof.** Define the set

$$\mathcal{P}=\left\{\left(e^{b_i x_i}\right)_{i=1}^D : \boldsymbol{x}\in\mathbb{N}_+^D\right\},$$

and define $\lfloor \boldsymbol{y}\rfloor=(\lfloor y_i\rfloor)_{i=1}^D$. Then

$$\begin{aligned}\sum_{\{\boldsymbol{x}\in\mathbb{N}_+^D:\sum_{i=1}^D a_i e^{b_i x}+b_i x_i>L\}} \exp\left(-\sum_{i=1}^D a_i e^{b_i x_i}\right) &= \sum_{\{\boldsymbol{y}\in\mathcal{P}:\boldsymbol{a}\cdot\boldsymbol{y}+|\log(\boldsymbol{y})|>L\}} \exp(-\boldsymbol{a}\cdot\boldsymbol{y}) \\ &\le \sum_{\{\boldsymbol{y}\in\mathcal{P}:\boldsymbol{a}\cdot(\lfloor\boldsymbol{y}\rfloor+\boldsymbol{1})+|\log(\lfloor\boldsymbol{y}\rfloor+\boldsymbol{1})|>L\}} \exp(-\boldsymbol{a}\lfloor\boldsymbol{y}\rfloor) \\ &\le \sum_{\{\boldsymbol{y}\in\mathbb{N}_+^D:\boldsymbol{a}\cdot\boldsymbol{y}+|\log(\boldsymbol{y}+\boldsymbol{1})|>L-|\boldsymbol{a}|\}} \exp(-\boldsymbol{a}\cdot\boldsymbol{y}) \\ &\le \int_{\{\boldsymbol{y}\in(1,\infty)^D:\boldsymbol{a}\cdot\boldsymbol{y}+|\log(\boldsymbol{y}+\boldsymbol{1})|>L-|\boldsymbol{a}|\}} \exp(-\boldsymbol{a}\cdot(\boldsymbol{y}-\boldsymbol{1}))\,\mathrm{d}\boldsymbol{y}.\end{aligned}$$

Letting $z_i=a_i y_i+\log(y_i+1)$ and $p(z_i)=y_i\le \frac{z_i}{a_i}$, then

$$\begin{aligned}&\sum_{\{\boldsymbol{x}\in\mathbb{N}_+^D:\sum_{i=1}^D a_i e^{b_i x}+b_i x_i>L\}} \exp\left(-\sum_{i=1}^D a_i e^{b_i x_i}\right) \\ &= \exp(|\boldsymbol{a}|)\int_{\{\boldsymbol{y}\in(1,\infty)^D:\boldsymbol{a}\cdot\boldsymbol{y}+|\log(\boldsymbol{y}+\boldsymbol{1})|>L-|\boldsymbol{a}|\}} \exp(-\boldsymbol{a}\cdot\boldsymbol{y}-|\log(\boldsymbol{y}+\boldsymbol{1})|+|\log(\boldsymbol{y}+\boldsymbol{1})|)\,\mathrm{d}\boldsymbol{y} \\ &= \exp(|\boldsymbol{a}|)\int_{\{\boldsymbol{z}\in\otimes_{i=1}^D(a_i+\log(2),\infty):|\boldsymbol{z}|>L-|\boldsymbol{a}|\}} \exp(-|\boldsymbol{z}|)\prod_{i=1}^D\left(\frac{p(z_i)+1}{a_i+\frac{1}{p(z_i)+1}}\right)\mathrm{d}\boldsymbol{z} \\ &\le \exp(|\boldsymbol{a}|)\int_{\{\boldsymbol{z}\in\otimes_{i=1}^D(a_i+\log(2),\infty):|\boldsymbol{z}|>L-|\boldsymbol{a}|\}} \exp(-|\boldsymbol{z}|)\prod_{i=1}^D\left(\frac{z_i+a_i}{a_i^2}\right)\mathrm{d}\boldsymbol{z} \\ &\le \left(\prod_{i=1}^D\frac{\exp(a_i)}{a_i^2}\right)\int_{\{\boldsymbol{z}\in\otimes_{i=1}^D(a_i+\log(2),\infty):|\boldsymbol{z}|>L-|\boldsymbol{a}|\}} \exp(-|\boldsymbol{z}|+|\log(\boldsymbol{z}+\boldsymbol{a})|)\,\mathrm{d}\boldsymbol{z}\end{aligned}$$

$$= \left(\prod_{i=1}^{D} \frac{\exp(2a_i)}{a_i^2}\right) \int_{\{\boldsymbol{x}\in\otimes_{i=1}^{D}(\log(2),\infty):|\boldsymbol{x}|>L\}} \exp(-|\boldsymbol{x}| + |\log(\boldsymbol{x})|)\, \mathrm{d}\boldsymbol{x}$$

$$\leq \left(\prod_{i=1}^{D} \frac{\exp(2a_i)}{a_i^2}\right) \int_{\{\boldsymbol{z}\in\otimes_{i=1}^{D}(0,\infty):|\boldsymbol{z}|>L\}} \exp(-|\boldsymbol{z}| + |\log(\boldsymbol{z})|)\, \mathrm{d}\boldsymbol{z}.$$

Now let us prove, by induction on $D$, that we have

$$\int_{\{\boldsymbol{z}\in\mathbb{R}_+^D:|\boldsymbol{z}|>L\}} \exp(-|\boldsymbol{z}| + |\log(\boldsymbol{z})|)\, \mathrm{d}\boldsymbol{z} \leq \exp(-L)(L+1)^{2D-1}.$$

For $D = 1$, the inequality is a trivial equality that can be obtained with integration by parts. Assume the inequality is true for $D$ and let us prove it for $D+1$:

$$\begin{aligned}
\int_{\{\boldsymbol{z}\in\mathbb{R}_+^{D+1}:|\boldsymbol{z}|>L\}} \exp(-|\boldsymbol{z}| + \log(\boldsymbol{z}))\, \mathrm{d}\boldsymbol{z} &= \int_L^\infty y\exp(-y) \int_{\{\boldsymbol{x}\in\mathbb{R}_+^D\}} \exp(-|\boldsymbol{x}| + \log(\boldsymbol{x}))\, \mathrm{d}\boldsymbol{x}\, \mathrm{d}y \\
&\quad + \int_0^L y\exp(-y) \int_{\{\boldsymbol{x}\in\mathbb{R}_+^D:|\boldsymbol{x}|>L-y\}} \exp(-|\boldsymbol{x}| + \log(\boldsymbol{x}))\, \mathrm{d}\boldsymbol{x}\, \mathrm{d}y \\
&\leq \exp(-L)(L+1) \\
&\quad + \int_0^L y\exp(-y)\exp(-L+y)(L-y+1)^{2D-1}\, \mathrm{d}y \\
&\leq \exp(-L)\left(L+1+L^2(L+1)^{2D-1}\right) \\
&\leq \exp(-L)\,(L+1)\left(1+L(L+1)^{2D-1}\right) \\
&\leq \exp(-L)(L+1)^{2(D+1)-1}.
\end{aligned}$$

Finally, substituting back, we get the result. □

**Definition 1.** Given $\boldsymbol{a} \in \mathbb{R}_+^D$ and $A > 0$, let $\mathfrak{n}(\boldsymbol{a}, A)$ denote the number of occurrences of $A$ in $\boldsymbol{a}$,

$$\mathfrak{n}(\boldsymbol{a}, A) = \#\{i = 1, \ldots, d : a_i = A\}.$$

**Lemma 6.** *Assume $k \in \mathbb{N}$, $\boldsymbol{a} \in \mathbb{R}_+^D$, $L > |\boldsymbol{a}|$. Then, the following bounds hold true:*

$$\int_{\{\boldsymbol{x}\in\mathbb{R}_+^D:|\boldsymbol{x}|>L\}} \exp(-\boldsymbol{a}\cdot\boldsymbol{x})\, d\boldsymbol{x} \leq \mathfrak{B}_D(\boldsymbol{a}) \exp(-\min(\boldsymbol{a})L) L^{\mathfrak{n}(\boldsymbol{a},\min(\boldsymbol{a}))-1}$$

*where $\mathfrak{B}_D(\boldsymbol{a})$ is a positive constant independent of $L$.*

**Proof.** See [29, Lemma B.3] for a proof of the inequality and the value of $\mathfrak{B}_D(\boldsymbol{a})$. Moreover, a proof of a consistent equality for the case $\boldsymbol{a} = \mathbf{1}$ can be found in [51, Proposition 2.3]. □

**Lemma 7.** *Assume $k \in \mathbb{N}$, $\boldsymbol{a} \in \mathbb{R}_+^D$, $L > |\boldsymbol{a}|$. Then, the following bound holds:*

$$\int_{\{\boldsymbol{x}\in\mathbb{R}_+^D:|\boldsymbol{x}|\leq L\}} \exp(\boldsymbol{a}\cdot\boldsymbol{x})\,(L-|\boldsymbol{x}|)^k\, d\boldsymbol{x} \leq \mathfrak{A}_D(\boldsymbol{a},k) \exp(\max(\boldsymbol{a})L) L^{\mathfrak{n}(\boldsymbol{a},\max(\boldsymbol{a}))-1},$$

*where*

$$\mathfrak{A}_D(\boldsymbol{a},k) = \frac{k!}{(\mathfrak{n}(\boldsymbol{a},\max(\boldsymbol{a}))-1)!\max(\boldsymbol{a})^{k+1}} \left(\prod_{a_i<\max(\boldsymbol{a})} \frac{1}{\max(\boldsymbol{a})-a_i}\right).$$

**Proof.** Without loss of generality, assume that $a_i \geq a_{i+1}$ for all $i = 1 \dots D$, such that $a_1 = \max(\boldsymbol{a})$. We prove the result by induction on $D$. For $D = 1$, we have

$$\begin{aligned}\int_0^L \exp(ax)\,(L-x)^k \,\mathrm{d}x &= \frac{k!}{a^{k+1}}\left(\exp(aL) - \sum_{i=0}^{k}\frac{a^i L^i}{i!}\right)\\ &\leq \frac{k!\exp(aL)}{a^{k+1}}.\end{aligned}$$

Next, assume that the result is valid for a given $D > 1$ and $\boldsymbol{a} \in \mathbb{R}_+^D$ where $a_i \geq a_{i+1}$ for all $i = 1 \dots D$, such that $a_1 = \max(\boldsymbol{a})$. Let $b \leq a_1$ and define a new vector $\widetilde{\boldsymbol{a}} = (\boldsymbol{a}, b) \in \mathbb{R}_+^{D+1}$. We have

$$\begin{aligned}&\int_{\{(\boldsymbol{x},y)\in\mathbb{R}_+^{D+1}:\, y+|\boldsymbol{x}|\leq L\}} \exp(by + \boldsymbol{a}\cdot\boldsymbol{x})\,(L-y-|\boldsymbol{x}|)^k \,\mathrm{d}y\,\mathrm{d}\boldsymbol{x}\\ &\quad= \int_0^L \exp(by)\int_{\{\boldsymbol{x}\in\mathbb{R}_+^D:\,|\boldsymbol{x}|\leq L-y\}} \exp(\boldsymbol{a}\cdot\boldsymbol{x})\,(L-y-|\boldsymbol{x}|)^k \,\mathrm{d}\boldsymbol{x}\,\mathrm{d}y\\ &\quad\leq \mathfrak{A}_D(\boldsymbol{a},k)\exp(a_1 L)\int_0^L \exp((b-a_1)y)\,(L-y)^{\mathfrak{n}(\boldsymbol{a},a_1)-1}\,\mathrm{d}y.\end{aligned}$$

We distinguish between two cases:

1. $b < a_1$ then $\mathfrak{n}(\widetilde{\boldsymbol{a}}, a_1) = \mathfrak{n}(\boldsymbol{a}, a_1)$ and

$$\begin{aligned}\int_0^L \exp(-(a_1-b)y)\,(L-y)^{\mathfrak{n}(\boldsymbol{a},a_1)-1}\,\mathrm{d}y &\leq L^{\mathfrak{n}(\boldsymbol{a},a_1)-1}\int_0^\infty \exp(-(a_1-b)y)\,\mathrm{d}y\\ &\leq L^{\mathfrak{n}(\widetilde{\boldsymbol{a}},a_1)-1}\frac{1}{a_1-b},\end{aligned}$$

   and in this case

$$\begin{aligned}\mathfrak{A}_D(\boldsymbol{a},k)\frac{1}{a_1-b} &= \frac{k!}{(\mathfrak{n}(\boldsymbol{a},a_1)-1)!a_1^{k+1}}\left(\prod_{a_i<a_1}\frac{1}{a_1-a_i}\right)\left(\frac{1}{a_1-b}\right)\\ &= \mathfrak{A}_{D+1}(\widetilde{\boldsymbol{a}},k).\end{aligned}$$

2. $b = a_1$ then $\mathfrak{n}(\widetilde{\boldsymbol{a}}, a_1) = \mathfrak{n}(\boldsymbol{a}, a_1) + 1$ and

$$\int_0^L (L-y)^{\mathfrak{n}(\boldsymbol{a},a_1)-1}\,\mathrm{d}y = \frac{L^{\mathfrak{n}(\boldsymbol{a},a_1)}}{\mathfrak{n}(\boldsymbol{a},a_1)} = \frac{L^{\mathfrak{n}(\widetilde{\boldsymbol{a}},a_1)-1}}{\mathfrak{n}(\widetilde{\boldsymbol{a}},a_1)-1},$$

   and again

$$\begin{aligned}\frac{1}{\mathfrak{n}(\widetilde{\boldsymbol{a}},a_1)-1}\mathfrak{A}_D(\boldsymbol{a},k) &= \frac{1}{\mathfrak{n}(\widetilde{\boldsymbol{a}},a_1)-1}\cdot\frac{k!}{(\mathfrak{n}(\boldsymbol{a},a_1)-1)!a_1^{k+1}}\left(\prod_{a_i<a_1}\frac{1}{a_1-a_i}\right)\\ &= \mathfrak{A}_{D+1}(\widetilde{\boldsymbol{a}},k). \quad \square\end{aligned}$$

**Theorem 1** (*MISC Computational Complexity*)**.** *Under Assumptions 2–4, the bounds for the factors appearing in Assumption 3 can be equivalently rewritten as*

$$\Delta W_{\boldsymbol{\alpha},\boldsymbol{\beta}} \leq C_{\text{work}} e^{\sum_{i=1}^{D}\gamma_i\alpha_i} e^{\delta|\boldsymbol{\beta}|}, \tag{23a}$$

$$\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}} \leq C_{\text{error}} e^{-\sum_{i=1}^{D} r_i\alpha_i} e^{-\sum_{j=1}^{N} g_j\exp(\delta\beta_j)}, \tag{23b}$$

*with* $\gamma_i = \widetilde{\gamma}_i\log 2$, $r_i = \widetilde{r}_i\log 2$, $\delta = \log 2$ *and* $g_i = \widetilde{g}_i C_{m,low}$. *Define the following set*

$$\mathcal{I}^*(L) = \left\{[\boldsymbol{\alpha},\boldsymbol{\beta}] \in \mathbb{N}_+^{D+N} : \sum_{i=1}^{D}(r_i+\gamma_i)\alpha_i + \sum_{i=1}^{N}(\delta\beta_i + g_i e^{\delta\beta_i}) \leq L\right\} \quad \text{with } L \in \mathbb{R}_+. \tag{24}$$

*Then there exists a constant* $\mathcal{C}_W$ *such that, for any* $W_{\max}$ *satisfying*

$$W_{\max} \geq \mathcal{C}_W \exp(\chi), \tag{25}$$

*and choosing* $L$ *as*

$$L = L(W_{\max}) = \frac{1}{\chi}\left(\log\left(\frac{W_{\max}}{\mathcal{C}_W}\right) - (\mathfrak{z}-1)\log\left(\frac{1}{\chi}\log\left(\frac{W_{\max}}{\mathcal{C}_W}\right)\right)\right), \tag{26}$$

*with* $\boldsymbol{\Xi} = \left(\frac{\gamma_1}{\gamma_1+r_1}, \dots, \frac{\gamma_D}{\gamma_D+r_D}\right)$, $\chi = \max(\boldsymbol{\Xi})$, $\zeta = \min_{i=1,\dots,D} \frac{r_i}{\gamma_i}$ *and* $\mathfrak{z} = \#\{i = 1, \dots, D : \frac{r_i}{\gamma_i} = \zeta\}$, *the MISC estimator* $\mathcal{M}_{\mathcal{I}^*(L(W_{\max}))}$ *satisfies*

$$\text{Work}\left[\mathcal{M}_{\mathcal{I}^*(L(W_{\max}))}\right] \leq W_{\max}, \tag{27a}$$

$$\limsup_{W_{\max}\uparrow\infty} \frac{\text{Error}\left[\mathcal{M}_{\mathcal{I}^*(L(W_{\max}))}\right]}{W_{\max}^{-\zeta}\left(\log(W_{\max})\right)^{(\zeta+1)(\mathfrak{z}-1)}} = \mathcal{C}_E < \infty. \tag{27b}$$

**Proof.** The bounds (23a) and (23b) can be obtained by elementary algebraic operations combining Assumptions 2 and 4; for instance,

$$\begin{aligned}
\Delta W_{\boldsymbol{\alpha}}^{\text{det}} &\leq C_{\text{work}}^{\text{det}} \prod_{i=1}^{D} h_i^{-\widetilde{\gamma}_i} = C_{\text{work}}^{\text{det}} \prod_{i=1}^{D} (h_0 2^{-\alpha_i})^{-\widetilde{\gamma}_i} = C_{\text{work}}^{\text{det}} h_0^{-|\widetilde{\boldsymbol{\gamma}}|} \prod_{i=1}^{D} 2^{\widetilde{\gamma}_i \alpha_i} = C_{\text{work}}^{\text{det}} h_0^{-|\widetilde{\boldsymbol{\gamma}}|} \prod_{i=1}^{D} e^{\widetilde{\gamma}_i \alpha_i \log 2},\\
\Delta W_{\boldsymbol{\beta}}^{\text{stoc}} &\leq C_{\text{work}}^{\text{stoc}} \prod_{n=1}^{N} 2^{\beta_n} = C_{\text{work}}^{\text{stoc}} \prod_{n=1}^{N} e^{\beta_n \log 2},
\end{aligned}$$

from which (23a) follows by setting $C_{\text{work}} = C_{\text{work}}^{\text{det}} h_0^{-|\widetilde{\boldsymbol{\gamma}}|} C_{\text{work}}^{\text{stoc}}$. The proof is then divided into two steps.

*Step* 1: *Work estimate*. Observe that $\Xi_i = \frac{\gamma_i}{\gamma_i+r_i} < 1$ for all $i = 1, \dots, D$ and that $\mathfrak{z} = \mathfrak{n}(\boldsymbol{\Xi}, \chi)$. Thanks to Eqs. (14) and (23a), and using Lemma 4, the total work satisfies

$$\begin{aligned}
\text{Work}\left[\mathcal{I}^*(L)\right] &= \sum_{(\boldsymbol{\alpha},\boldsymbol{\beta})\in\mathcal{I}^*(L)} \Delta W_{\boldsymbol{\alpha},\boldsymbol{\beta}}\\
&\leq C_{\text{work}} \sum_{\left\{(\boldsymbol{\alpha},\boldsymbol{\beta})\in\mathbb{N}_+^{D+N} : \sum_{i=1}^{D}(r_i+\gamma_i)\alpha_i + \sum_{j=1}^{N}\delta\beta_j + g_j e^{\delta\beta_j} \leq L\right\}} \exp\left(\sum_{i=1}^{D}\gamma_i\alpha_i + \delta|\boldsymbol{\beta}|\right)\\
&\leq C_{\text{work}} \int_{\left\{(\boldsymbol{\alpha},\boldsymbol{\beta})\in(1,\infty)^{D+N} : \sum_{i=1}^{D}(r_i+\gamma_i)(\alpha_i-1) + \sum_{j=1}^{N}\delta(\beta_j-1) + g_j e^{\delta(\beta_j-1)} \leq L\right\}} \exp\left(\sum_{i=1}^{D}\gamma_i\alpha_i + \delta|\boldsymbol{\beta}|\right) \mathrm{d}\boldsymbol{\alpha}\,\mathrm{d}\boldsymbol{\beta}.
\end{aligned}$$

Next, let $\overline{\beta_j} = g_j e^{\delta(\beta_j-1)}$ and $\overline{\alpha_i} = (r_i+\gamma_i)(\alpha_i-1)$. We have

$$\begin{aligned}
\text{Work}\left[\mathcal{I}^*(L)\right] \leq C_{\text{work}} &\left(\prod_{j=1}^{N}\frac{2}{g_j\delta}\right)\left(\prod_{i=1}^{D}\frac{\exp(\gamma_i)}{r_i+\gamma_i}\right)\\
&\int_{\left\{(\overline{\boldsymbol{\alpha}},\overline{\boldsymbol{\beta}})\in\mathbb{R}_+^D\times(\otimes_{j=1}^{N}(g_j,\infty)) : |\overline{\boldsymbol{\alpha}}|+|\overline{\boldsymbol{\beta}}|+|\log\overline{\boldsymbol{\beta}}| \leq L+|\log\boldsymbol{g}|\right\}} \exp\left(\boldsymbol{\Xi}\cdot\overline{\boldsymbol{\alpha}}\right)\,\mathrm{d}\overline{\boldsymbol{\alpha}}\,\mathrm{d}\overline{\boldsymbol{\beta}}.
\end{aligned}$$

Dropping the over-line notation and defining $\widetilde{L} = L + |\log\boldsymbol{g}|$ and $\mathcal{C}_{\text{W},1}$ to be the constant factor, we obtain

$$\begin{aligned}
\text{Work}\left[\mathcal{I}^*(L)\right] &\leq \mathcal{C}_{\text{W},1} \int_{\left\{(\boldsymbol{\alpha},\boldsymbol{\beta})\in\mathbb{R}_+^D\times(\otimes_{j=1}^{N}(g_j,\infty)) : |\boldsymbol{\alpha}|+|\boldsymbol{\beta}|+|\log\boldsymbol{\beta}| \leq \widetilde{L}\right\}} \exp\left(\boldsymbol{\Xi}\cdot\boldsymbol{\alpha}\right)\,\mathrm{d}\boldsymbol{\alpha}\,\mathrm{d}\boldsymbol{\beta}\\
&= \mathcal{C}_{\text{W},1} \int_{\left\{\boldsymbol{\beta}\in\otimes_{j=1}^{N}(g_j,\infty) : |\boldsymbol{\beta}|+|\log\boldsymbol{\beta}| \leq \widetilde{L}\right\}} \int_{\left\{\boldsymbol{\alpha}\in\mathbb{R}_+^D : |\boldsymbol{\alpha}| \leq \widetilde{L}-|\boldsymbol{\beta}|-|\log\boldsymbol{\beta}|\right\}} \exp\left(\boldsymbol{\Xi}\cdot\boldsymbol{\alpha}\right)\,\mathrm{d}\boldsymbol{\alpha}\,\mathrm{d}\boldsymbol{\beta}\\
&\leq \mathcal{C}_{\text{W},1}\mathfrak{A}_D(\boldsymbol{\Xi},0) \int_{\left\{\boldsymbol{\beta}\in\otimes_{j=1}^{N}(g_j,\infty) : |\boldsymbol{\beta}|+|\log\boldsymbol{\beta}| \leq \widetilde{L}\right\}} \exp\Big(\chi\left(\widetilde{L}-|\boldsymbol{\beta}|-|\log\boldsymbol{\beta}|\right)\Big)\left(\widetilde{L}-|\boldsymbol{\beta}|-|\log\boldsymbol{\beta}|\right)^{\mathfrak{z}-1}\,\mathrm{d}\boldsymbol{\beta}.
\end{aligned}$$

Define $\mathcal{C}_{\mathrm{W},2} = \mathcal{C}_{\mathrm{W},1}\mathfrak{A}_D\left(\boldsymbol{\Xi}, 0\right)\exp(\chi\widetilde{L})$, then

$$\begin{aligned}\mathrm{Work}\left[\mathcal{I}^*(L)\right] &\leq \mathcal{C}_{\mathrm{W},2}\int_{\left\{\boldsymbol{\beta}\in\otimes_{j=1}^{N}(g_j,\infty):|\boldsymbol{\beta}|+|\log\boldsymbol{\beta}|\leq\widetilde{L}\right\}}\exp\left(-\chi\left(|\boldsymbol{\beta}|+|\log\boldsymbol{\beta}|\right)\right)\left(\widetilde{L}-|\boldsymbol{\beta}|-|\log\boldsymbol{\beta}|\right)^{\mathfrak{z}-1}\,\mathrm{d}\boldsymbol{\beta}\\ &\leq \mathcal{C}_{\mathrm{W},2}\left(\widetilde{L}-|\boldsymbol{g}|-|\log\boldsymbol{g}|\right)^{\mathfrak{z}-1}\int_{\left\{\boldsymbol{\beta}\in\otimes_{j=1}^{N}(g_j,\infty):|\boldsymbol{\beta}|+|\log\boldsymbol{\beta}|\leq\widetilde{L}\right\}}\exp\left(-\chi\left(|\boldsymbol{\beta}|+|\log\boldsymbol{\beta}|\right)\right)\,\mathrm{d}\boldsymbol{\beta}.\end{aligned}$$

Since $\chi > 0$, the previous integral is bounded for all $\widetilde{L}$ and we have

$$\mathrm{Work}\left[\mathcal{I}^*(L)\right] \leq \mathcal{C}_{\mathrm{W}}\exp(\chi L)\left(L-|\boldsymbol{g}|\right)^{\mathfrak{z}-1} \leq \mathcal{C}_{\mathrm{W}}\exp(\chi L)L^{\mathfrak{z}-1},$$

where

$$\mathcal{C}_{\mathrm{W}} = C_{\mathrm{work}}\left(\prod_{j=1}^{N}\frac{2g_i^{\chi}}{g_j\log 2}\right)\left(\prod_{i=1}^{D}\frac{\exp(\gamma_i)}{r_i+\gamma_i}\right)\mathfrak{A}_D\left(\boldsymbol{\Xi}, 0\right)\int_{\left\{\boldsymbol{\beta}\in\otimes_{j=1}^{N}(g_j,\infty)\right\}}\exp\left(-\chi\left(|\boldsymbol{\beta}|+|\log\boldsymbol{\beta}|\right)\right)\,\mathrm{d}\boldsymbol{\beta}.$$

Substituting (26) yields

$$\mathrm{Work}\left[\mathcal{I}^*(L)\right] \leq W_{\max}\left(1-\frac{(\mathfrak{z}-1)\log\left(\frac{\log\left(\frac{W_{\max}}{\mathcal{C}_{\mathrm{W}}}\right)}{\chi}\right)}{\log\left(\frac{W_{\max}}{\mathcal{C}_{\mathrm{W}}}\right)}\right)^{\mathfrak{z}-1}.$$

From here it is easy to see that if (25) is satisfied, then (27a) follows.

*Step* 2: *Error estimate*. Thanks to Eqs. (15) and (23b), the total error satisfies

$$\begin{aligned}\mathrm{Error}\left[\mathcal{I}^*(L)\right] &\leq \sum_{(\boldsymbol{\alpha},\boldsymbol{\beta})\notin\mathcal{I}^*}\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}}\\ &\leq C_{\mathrm{error}}\sum_{\left\{(\boldsymbol{\alpha},\boldsymbol{\beta})\in\mathbb{N}_+^{D+N}:\Sigma_{i=1}^{D}(r_i+\gamma_i)\alpha_i+\Sigma_{j=1}^{N}\delta\beta_j+g_je^{\delta\beta_j}>L\right\}}\exp\left(-\sum_{i=1}^{D}r_i\alpha_i-\sum_{j=1}^{N}g_je^{\delta\beta_j}\right)\\ &= C_{\mathrm{error}}\sum_{\left\{(\boldsymbol{\alpha},\boldsymbol{\beta})\in\mathbb{N}_+^{D+N}:\Sigma_{i=1}^{D}(r_i+\gamma_i)\alpha_i>L\right\}}\exp\left(-\sum_{i=1}^{D}r_i\alpha_i-\sum_{j=1}^{N}g_je^{\delta\beta_j}\right)\\ &\quad + C_{\mathrm{error}}\sum_{\left\{\boldsymbol{\alpha}\in\mathbb{N}_+^{D}:\Sigma_{i=1}^{D}(r_i+\gamma_i)\alpha_i\leq L\right\}}\sum_{\left\{\boldsymbol{\beta}\in\mathbb{N}_+^{N}:\Sigma_{j=1}^{N}\delta\beta_j+g_je^{\delta\beta_j}>L-\Sigma_{i=1}^{D}(r_i+\gamma_i)\alpha_i\right\}}\exp\left(-\sum_{i=1}^{D}r_i\alpha_i-\sum_{j=1}^{N}g_je^{\delta\beta_j}\right).\end{aligned}$$

Looking at the first term, let $\eta_i = \frac{r_i}{\gamma_i+r_i} < 1$ and $\boldsymbol{\eta} = (\eta_i)_{i=1}^{D}$ and note that $\mathfrak{z} = \#\left\{i = 1\ldots D : \eta_i = \min(\boldsymbol{\eta})\right\}$. Then

$$\begin{aligned}&\sum_{\left\{(\boldsymbol{\alpha},\boldsymbol{\beta})\in\mathbb{N}_+^{D+N}:\sum_{i=1}^{D}(r_i+\gamma_i)\alpha_i>L\right\}}\exp\left(-\sum_{i=1}^{D}r_i\alpha_i-\sum_{j=1}^{N}g_je^{\delta\beta_j}\right)\\ &= \left(\sum_{\boldsymbol{\beta}\in\mathbb{N}_+^{N}}\exp\left(-\sum_{j=1}^{N}g_je^{\delta\beta_j}\right)\right)\left(\sum_{\left\{\boldsymbol{\alpha}\in\mathbb{N}_+^{D}:\sum_{i=1}^{D}(r_i+\gamma_i)\alpha_i>L\right\}}\exp\left(-\sum_{i=1}^{D}r_i\alpha_i\right)\right)\\ &\leq \mathcal{C}_{\mathrm{E},1}\int_{\left\{\boldsymbol{\alpha}\in(1,\infty)^{D}:\sum_{i=1}^{D}(r_i+\gamma_i)\alpha_i>L\right\}}\exp\left(-\sum_{i=1}^{D}r_i\left(\alpha_i-1\right)\right)\,\mathrm{d}\boldsymbol{\alpha}\end{aligned}$$

$$= \mathcal{C}_{\mathrm{E},1} \left( \prod_{i=1}^{D} \frac{\exp(r_i)}{r_i + \gamma_i} \right) \int_{\{\boldsymbol{x} \in \otimes_{i=1}^{D} (r_i+\gamma_i, \infty) : |\boldsymbol{x}| > L\}} \exp\left( - \sum_{i=1}^{D} \frac{r_i}{r_i + \gamma_i} x_i \right) \mathrm{d}\boldsymbol{x}$$

$$\leq \mathcal{C}_{\mathrm{E},2} \exp\left(-\min(\boldsymbol{\eta}) L\right) L^{\mathfrak{z}-1},$$

where

$$\mathcal{C}_{\mathrm{E},2} = \mathfrak{B}_D(\boldsymbol{\eta}) \left( \prod_{i=1}^{D} \frac{\exp(r_i)}{r_i + \gamma_i} \right) \sum_{\boldsymbol{\beta} \in \mathbb{N}_+^N} \exp\left( - \sum_{j=1}^{N} g_j e^{\delta \beta_j} \right).$$

For the second term, letting $H = L - \sum_{i=1}^{D} (r_i + \gamma_i)\alpha_i$, we can bound the sum using Lemma 5:

$$\sum_{\left\{\boldsymbol{\beta} \in \mathbb{N}_+^N : \sum_{j=1}^{N} \delta\beta_j + g_j e^{\delta\beta_j} > H\right\}} \exp\left( - \sum_{j=1}^{N} g_j e^{\delta \beta_j} \right) \leq \left( \prod_{j=1}^{N} \frac{\exp(2g_j)}{g_j^2} \right) \exp(-H)(H+1)^{2N-1}.$$

Defining $\mathcal{C}_{\mathrm{E},3} = \prod_{j=1}^{N} \exp(2g_j) g_j^{-2}$ and substituting back

$$\sum_{\left\{\boldsymbol{\alpha} \in \mathbb{N}_+^D : \sum_{i=1}^{D} (r_i+\gamma_i)\alpha_i \leq L\right\}} \exp\left( - \sum_{i=1}^{D} r_i \alpha_i \right) \sum_{\left\{\boldsymbol{\beta} \in \mathbb{N}_+^N : \sum_{j=1}^{N} \delta\beta_j + g_j e^{\delta\beta_j} > L - \sum_{i=1}^{D} (r_i+\gamma_i)\alpha_i\right\}} \exp\left( - \sum_{j=1}^{N} g_j e^{\delta \beta_j} \right)$$

$$\leq \mathcal{C}_{\mathrm{E},3} \sum_{\left\{\boldsymbol{\alpha} \in \mathbb{N}_+^D : \sum_{i=1}^{D} (r_i+\gamma_i)\alpha_i \leq L\right\}} \exp\left( -L + \sum_{i=1}^{D} \gamma_i \alpha_i \right) \left( L + 1 - \sum_{i=1}^{D} (r_i + \gamma_i)\alpha_i \right)^{2N-1}$$

$$= \mathcal{C}_{\mathrm{E},3} \int_{\left\{\boldsymbol{\alpha} \in (1,\infty)^D : \sum_{i=1}^{D} (r_i+\gamma_i)\lfloor \alpha_i \rfloor \leq L\right\}} \exp\left( -L + \sum_{i=1}^{D} \gamma_i \lfloor \alpha_i \rfloor \right) \left( L + 1 - \sum_{i=1}^{D} (r_i + \gamma_i)\lfloor \alpha_i \rfloor \right)^{2N-1} \mathrm{d}\boldsymbol{\alpha}$$

$$\leq \mathcal{C}_{\mathrm{E},3} \int_{\left\{\boldsymbol{\alpha} \in (1,\infty)^D : \sum_{i=1}^{D} (r_i+\gamma_i)(\alpha_i - 1) \leq L\right\}} \exp\left( -L + \sum_{i=1}^{D} \gamma_i \alpha_i \right) \left( L + 1 - \sum_{i=1}^{D} (r_i + \gamma_i)(\alpha_i - 1) \right)^{2N-1} \mathrm{d}\boldsymbol{\alpha}$$

$$= \mathcal{C}_{\mathrm{E},3} \left( \prod_{i=1}^{D} \frac{\exp(\gamma_i)}{\gamma_i + r_i} \right) \exp(-L) \int_{\{\boldsymbol{\alpha} \in \mathbb{R}_+^D : |\boldsymbol{x}| \leq L\}} \exp\left( \boldsymbol{\Xi} \cdot \boldsymbol{x} \right) (L + 1 - |\boldsymbol{x}|)^{2N-1} \mathrm{d}\boldsymbol{x}$$

$$\leq \mathcal{C}_{\mathrm{E},4} \exp((\chi - 1)L) L^{\mathfrak{z}-1},$$

where

$$\mathcal{C}_{\mathrm{E},4} = \left( \prod_{j=1}^{N} \frac{\exp(2g_j)}{g_j^2} \right) \left( \prod_{i=1}^{D} \frac{\exp(\gamma_i)}{\gamma_i + r_i} \right) \mathfrak{A}_D(\boldsymbol{\Xi}, 2N-1).$$

Finally, noting that

$$\chi - 1 = -\min(\boldsymbol{\eta}),$$

we have the error estimate

$$\mathrm{Error}\left[\mathcal{I}^*(L)\right] \leq C_{\mathrm{error}} \left( \mathcal{C}_{\mathrm{E},2} + \mathcal{C}_{\mathrm{E},4} \right) \exp(-\min(\boldsymbol{\eta}) L) L^{\mathfrak{z}-1}.$$

Then, substituting $L$ from (26) and evaluating the limit gives (27b). □

## References

[1] R.G. Ghanem, P.D. Spanos, Stochastic Finite Elements: A Spectral Approach, Springer-Verlag, New York, 1991.

[2] O.P. Le Maître, O.M. Knio, Spectral methods for uncertainty quantification, in: Scientific Computation, Springer, New York, 2010, with applications to computational fluid dynamics.

[3] H.G. Matthies, A. Keese, Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations, Comput. Methods Appl. Mech. Engrg. 194 (12–16) (2005) 1295–1331.
[4] R.A. Todor, C. Schwab, Convergence rates for sparse chaos approximations of elliptic problems with stochastic coefficients, IMA J. Numer. Anal. 27 (2) (2007) 232–261.
[5] D. Xiu, G. Karniadakis, The Wiener-Askey polynomial chaos for stochastic differential equations, SIAM J. Sci. Comput. 24 (2) (2002) 619–644.
[6] I. Babuška, F. Nobile, R. Tempone, A stochastic collocation method for elliptic partial differential equations with random input data, SIAM Rev. 52 (2) (2010) 317–355.
[7] H. Bungartz, M. Griebel, Sparse grids, Acta Numer. 13 (2004) 147–269.
[8] F. Nobile, R. Tempone, C. Webster, An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data, SIAM J. Numer. Anal. 46 (5) (2008) 2411–2442.
[9] D. Xiu, J. Hesthaven, High-order collocation methods for differential equations with random inputs, SIAM J. Sci. Comput. 27 (3) (2005) 1118–1139.
[10] B. Khoromskij, C. Schwab, Tensor-structured galerkin approximation of parametric and stochastic elliptic pdes, SIAM J. Sci. Comput. 33 (1) (2011) 364–385.
[11] B. Khoromskij, I. Oseledets, Quantics-tt collocation approximation of parameter-dependent and stochastic elliptic pdes, Comput. Methods Appl. Math. 10 (4) (2010) 376–394.
[12] A. Nouy, Generalized spectral decomposition method for solving stochastic finite element equations: invariant subspace problem and dedicated algorithms, Comput. Methods Appl. Mech. Engrg, 197 (51) (2008) 4718–4736.
[13] J. Ballani, L. Grasedyck, Hierarchical tensor approximation of output quantities of parameter-dependent PDEs, SIAM/ASA J. Uncertain. Quantif. 3 (1) (2015) 852–872. http://dx.doi.org/10.1137/140960980.
[14] S. Boyaval, C. Le Bris, T. Lelièvre, Y. Maday, N. Nguyen, A. Patera, Reduced basis techniques for stochastic problems, Arch. Comput. Methods Eng. 17 (4) (2010) 435–454.
[15] P. Chen, A. Quarteroni, G. Rozza, Comparison between reduced basis and stochastic collocation methods for elliptic problems, J. Sci. Comput. 59 (1) (2014) 187–216.
[16] A. Cohen, R. Devore, C. Schwab, Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDE'S, Anal. Appl. (Singap.) 9 (1) (2011) 11–47.
[17] S. Heinrich, Multilevel Monte Carlo methods, in: Large-Scale Scientific Computing, in: Lecture Notes in Computer Science, vol. 2179, Springer, Berlin, Heidelberg, 2001, pp. 58–67.
[18] M.B. Giles, Multilevel Monte Carlo path simulation, Oper. Res. 56 (3) (2008) 607–617.
[19] A. Barth, C. Schwab, N. Zollinger, Multi-level Monte Carlo finite element method for elliptic PDEs with stochastic coefficients, Numer. Math. 119 (1) (2011) 123–161.
[20] A. Barth, A. Lang, C. Schwab, Multilevel Monte Carlo method for parabolic stochastic partial differential equations, BIT 53 (1) (2013) 3–27.
[21] J. Charrier, R. Scheichl, A. Teckentrup, Finite element error analysis of elliptic pdes with random coefficients and its application to multilevel Monte Carlo methods, SIAM J. Numer. Anal. 51 (1) (2013) 322–352.
[22] K. Cliffe, M. Giles, R. Scheichl, A. Teckentrup, Multilevel Monte Carlo methods and applications to elliptic pdes with random coefficients, Comput. Vis. Sci. 14 (1) (2011) 3–15.
[23] S. Mishra, C. Schwab, J. Sukys, Multi-level Monte Carlo finite volume methods for nonlinear systems of conservation laws in multi-dimensions, J. Comput. Phys. 231 (8) (2012) 3365–3388.
[24] A.L. Teckentrup, P. Jantsch, C.G. Webster, M. Gunzburger, A multilevel stochastic collocation method for partial differential equations with random input data, SIAM/ASA J. Uncertain. Quantif. 3 (1) (2015) 1046–1074.
[25] H.W. van Wyk, Multilevel sparse grid methods for elliptic partial differential equations with random coefficients, 2014, arXiv preprint arXiv:1404.0963.
[26] H. Harbrecht, M. Peters, M. Siebenmorgen, On multilevel quadrature for elliptic stochastic partial differential equations, in: Sparse Grids and Applications, in: Lecture Notes in Computational Science and Engineering, vol. 88, Springer, 2013, pp. 161–179.
[27] F.Y. Kuo, C. Schwab, I. Sloan, Multi-level quasi-Monte Carlo finite element methods for a class of elliptic PDEs with random coefficients, Found. Comput. Math. 15 (2) (2015) 411–449.
[28] F. Nobile, F. Tesei, A multi level Monte Carlo method with control variate for elliptic PDEs with log-normal coefficients, Stoch. Partial Differ. Equ. Anal. Comput. (2015) 1–47.
[29] A.-L. Haji-Ali, F. Nobile, R. Tempone, Multi-index Monte Carlo: when sparsity meets sampling, Numer. Math. 132 (2015) 767–806. http://dx.doi.org/10.1007/s00211-015-0734-5.
[30] H. Bungartz, M. Griebel, D. Röschke, C. Zenger, Pointwise convergence of the combination technique for the Laplace equation, East-West J. Numer. Math. 2 (1994) 21–45.
[31] M. Griebel, M. Schneider, C. Zenger, A combination technique for the solution of sparse grid problems, in: P. de Groen, R. Beauwens (Eds.), Iterative Methods in Linear Algebra, IMACS, Elsevier, North Holland, 1992, pp. 263–281.
[32] M. Hegland, J. Garcke, V. Challis, The combination technique and some generalisations, Linear Algebra Appl. 420 (2–3) (2007) 249–275.
[33] M. Griebel, H. Harbrecht, On the convergence of the combination technique, in: J. Garcke, D. Pflüger (Eds.), Sparse Grids and Applications—Munich 2012, in: Lecture Notes in Computational Science and Engineering, vol. 97, Springer International Publishing, 2014, pp. 55–74. http://dx.doi.org/10.1007/978-3-319-04537-5_3.
[34] F. Nobile, L. Tamellini, R. Tempone, Convergence of quasi-optimal sparse-grid approximation of Hilbert-space-valued functions: application to random elliptic PDEs, Numer. Math. (2015) In print. http://dx.doi.org/10.1007/s00211-015-0773-y.
[35] J. Beck, F. Nobile, L. Tamellini, R. Tempone, On the optimal polynomial approximation of stochastic PDEs by Galerkin and collocation methods, Math. Models Methods Appl. Sci. 22 (09) (2012) 1250023.

[36] J. Beck, F. Nobile, L. Tamellini, R. Tempone, A quasi-optimal sparse grids procedure for groundwater flows, in: Spectral and High Order Methods for Partial Differential Equations—ICOSAHOM 2012, in: Lecture Notes in Computational Science and Engineering, vol. 95, Springer, 2014, pp. 1–16.
[37] M. Griebel, S. Knapek, Optimized general sparse grid approximation spaces for operator equations, Math. Comp. 78 (268) (2009) 2223–2257.
[38] M. Bieri, A sparse composite collocation finite element method for elliptic SPDEs, SIAM J. Numer. Anal. 49 (6) (2011) 2277–2301.
[39] T. Hughes, J. Cottrell, Y. Bazilevs, Isogeometric analysis: Cad, finite elements, nurbs, exact geometry and mesh refinement, Comput. Methods Appl. Mech. Engrg. 194 (39–41) (2005) 4135–4195.
[40] W.J. Gordon, C.A. Hall, Construction of curvilinear co-ordinate systems and applications to mesh generation, Internat. J. Numer. Methods Engrg. 7 (4) (1973) 461–477.
[41] A. Quarteroni, A. Valli, Numerical Mathematics and Scientific Computation, Clarendon Press, 1999.
[42] L. Trefethen, Approximation Theory and Approximation Practice, Society for Industrial and Applied Mathematics, 2013.
[43] L.N. Trefethen, Is Gauss quadrature better than Clenshaw-Curtis? SIAM Rev. 50 (1) (2008) 67–87.
[44] A. Chkifa, On the lebesgue constant of leja sequences for the complex unit disk and of their real projection, J. Approx. Theory 166 (0) (2013) 176–200.
[45] F. Nobile, L. Tamellini, R. Tempone, Comparison of Clenshaw–Curtis and Leja quasi-optimal sparse grids for the approximation of random PDEs, in: R.M. Kirby, M. Berzins, J.S. Hesthaven (Eds.), Spectral and High Order Methods for Partial Differential Equations—ICOSAHOM 2014, in: Lecture Notes in Computational Science and Engineering, vol. 106, Springer, 2015, pp. 475–482.
[46] A. Narayan, J.D. Jakeman, Adaptive leja sparse grid constructions for stochastic collocation and high-dimensional approximation, SIAM J. Sci. Comput. 36 (6) (2014) A2952–A2983.
[47] A. Genz, B.D. Keister, Fully symmetric interpolatory rules for multiple integrals over infinite regions with Gaussian weight, J. Comput. Appl. Math. 71 (2) (1996) 299–309.
[48] G. Wasilkowski, H. Wozniakowski, Explicit cost bounds of algorithms for multivariate tensor product problems, J. Complexity 11 (1) (1995) 1–56.
[49] S. Martello, P. Toth, Knapsack problems: algorithms and computer implementations, in: Wiley-Interscience Series in Discrete Mathematics and Optimization, J. Wiley & Sons, 1990.
[50] D. Dũng, M. Griebel, Hyperbolic cross approximation in infinite dimensions, J. Complexity (2015) In print. http://dx.doi.org/10.1016/j.jco.2015.09.006.
[51] M. Griebel, J. Oettershagen, On tensor product approximation of analytic functions, J. Approx. Theory (2016) In print. http://dx.doi.org/10.1016/j.jat.2016.02.006.
[52] J. Bäck, F. Nobile, L. Tamellini, R. Tempone, Stochastic spectral Galerkin and collocation methods for PDEs with random coefficients: a numerical comparison, in: Spectral and High Order Methods for Partial Differential Equations, in: Lecture Notes in Computational Science and Engineering, vol. 76, Springer, 2011, pp. 43–62.