MATHICSE
Mathematics Institute of Computational Science and Engineering
School of Basic Sciences - Section of Mathematics

MATHICSE Technical Report
Nr. 29.2015
November 2015

Multi-index stochastic collocation convergence rates for random PDEs with parametric regularity

Abdul-Lateef Haji-Ali, Fabio Nobile, Lorenzo Tamellini, Raúl Tempone

http://mathicse.epfl.ch

Phone: +41 21 69 37648

Address:
EPFL - SB - MATHICSE (Bâtiment MA)
Station 8 - CH-1015 - Lausanne - Switzerland

Fax: +41 21 69 32545

Multi-index Stochastic Collocation convergence rates for random PDEs with parametric regularity

Abdul-Lateef Haji-Ali (KAUST) Fabio Nobile (EPFL)
Lorenzo Tamellini (EPFL, UNIPV) Raúl Tempone (KAUST)

November 3, 2015

Abstract

We analyze the recent Multi-index Stochastic Collocation (MISC) method for computing statistics of the solution of a partial differential equation with random data, where the random coefficient is parametrized by means of a countable sequence of terms in a suitable expansion. MISC is a combination technique based on mixed differences of spatial approximations and quadratures over the space of random data and, naturally, the error analysis uses the joint regularity of the solution both with respect to the physical variables (the variables in the physical domain) and the parametric variables (the parameters corresponding to randomness). In MISC, the number of problem solutions performed at each discretization level is not determined by balancing the spatial and stochastic components of the error, but rather by suitably extending the knapsack-problem approach that we have employed in the construction of the quasi-optimal sparse-grids and Multi-index Monte Carlo methods. In this methodology, we use a greedy optimization procedure to select the most effective mixed differences to include in the MISC estimator and provide a complexity analysis based on a summability argument showing algebraic rates of convergence with respect to the overall computational work. We apply our theoretical estimates to a linear elliptic partial differential equation in which the diffusion coefficient is modeled as a random field whose realizations have spatial regularity determined by a scalar parameter (in the spirit of a Matérn covariance) and we estimate the rate of convergence in terms of the smoothness parameter, the physical dimension and the complexity of the linear solver. Numerical experiments show the effectiveness of MISC in this infinite-dimensional setting compared with Multi-index Monte Carlo, as well as the sharpness of the convergence result.

Keywords: Multilevel, Multi-index Stochastic Collocation, Infinite dimensional integration, Elliptic partial differential equations with random coefficients, Finite element method, Uncertainty Quantification, random partial differential equations, Multivariate approximation, Sparse grids, Stochastic Collocation methods, Multilevel methods, Combination technique.

AMS class: 41A10 (approx by polynomials), 65C20 (models, numerical methods), 65N30 (Finite elements) 65N05 (Finite differences)

1 Introduction

This work is concerned with the approximation of quantities of interest (outputs) from the solutions of partial differential equations (PDEs) with random coefficients. This kind of equations arise in many applications in which the coefficients of the PDE need be described in terms of random variables/fields due either to a lack of knowledge of the system or to its inherent non-predictability. Typical examples are the prediction of stresses on a structure under the action of random forces, such as wind and/or earthquakes, the forecasting of weather, or the design of groundwater management policies that take into account rainfall and the properties of the subsurface soil which are known at only a few drilling locations [1, 6, 39]. Here, we focus on the weak approximation of the solution of the following linear elliptic problem:

$$\begin{cases} -\mathrm{div}(a(\boldsymbol{x},\boldsymbol{y})\,\nabla u(\boldsymbol{x},\boldsymbol{y})) = f(\boldsymbol{x}) & \text{in} \quad \mathscr{B} \\ u(\boldsymbol{x},\boldsymbol{y}) \,=\, 0 & \text{on} \quad \partial\mathscr{B}, \end{cases} \tag{1}$$

where $\mathcal{B} = [0,1]^d$, $d = 1,2,3$ (hereafter referred to as the "physical domain"), the operators div and ∇ act with respect to the physical variable, $\boldsymbol{x}$, only, and $\boldsymbol{y} = \{y_n\}_{n\geq 1}$ is a random sequence whose components are independent and uniformly distributed random variables. More precisely, each y_n has support in $\Gamma_n = [-1,1]$ with measure $\frac{\mathrm{d}\lambda}{2}$, where $\mathrm{d}\lambda$ is the standard Lebesgue measure. We further define $\Gamma = \times_{n\geq 1}\Gamma_n$ (hereafter referred to as the "stochastic domain" or the "parameter space"), with measure $\mathrm{d}\mu = \bigotimes_{j\geq 1} \frac{\mathrm{d}\lambda}{2}$.

The right-hand side of (1), namely the deterministic function f, does not play a central role in this work and it is assumed to be a smooth function of class $C^\infty(\overline{\mathcal{B}})$. One can of course relax this regularity requirement but we keep it to ease the presentation and concentrate on tracking the regularity effect of the coefficient a in (1). Thanks to a straightforward application of the Lax-Milgram lemma, the well posedness of (1) in the classical Sobolev space $V = H_0^1(\mathcal{B})$ almost surely (a.s.) in Γ is guaranteed if the following assumption holds:

Assumption A1 (Boundedness of the diffusion coefficient)**.** *There exist two functions,* $a_{\min}, a_{\max} : \Gamma \to \mathbb{R}$*, such that*

$$0 < a_{\min}(\boldsymbol{y}) \leq a(\boldsymbol{x},\boldsymbol{y}) \leq a_{\max}(\boldsymbol{y}) < \infty, \quad \forall \boldsymbol{x} \in \mathcal{B},\ \textit{a.s. in } \Gamma.$$

Moreover, the equation is well posed in the Bochner space $L^q(\Gamma;V)$ for some $q \geq 1$ (see [1, 9] and Lemma 16), provided that sufficiently high moments of the functions $1/a_{\min}$ and $a_{\max}$ are bounded (we recall that, given $q \geq 1$, $L^q(\Gamma;V)$ is defined as $L^q(\Gamma;V) = \left\{v : \mathcal{B}\times\Gamma \to \mathbb{R} \text{ s.t. } \int_\Gamma \|u\|_V^q \ \mathrm{d}\mu < \infty\right\}$). The goal of our computation is the approximation of an expected value,

$$\mathbb{E}[F] = \mathbb{E}[\Theta(u)] \in \mathbb{R},$$

where Θ is a deterministic smooth functional, and $F(\boldsymbol{y}) = \Theta(u(\cdot,\boldsymbol{y}))$ is a real-valued random variable, $F : \Gamma \to \mathbb{R}$. To this end, we consider the Multi-index Stochastic Collocation method (MISC), which we have introduced in a previous work [22].

In MISC, we consider a decomposition in terms of tensorized univariate details (i.e., a tensorized hierarchical decomposition), for both the discrete space in which (1) is solved for a fixed value of $\boldsymbol{y} \in \Gamma$ and for the quadrature operator used to approximate the expected value of F, relying on the well-established theory of sparse-grid approximation of PDEs on the one hand [7, 8, 21, 25, 40] and of sparse-grid quadrature on the other hand [1, 7, 16, 32, 33, 39]. We use tensor products of such univariate details, obtaining combined deterministic-stochastic, first-order mixed differences to build the MISC estimator of $\mathbb{E}[F]$ by selecting the most effective mixed differences with an optimization approach inspired by the literature on the knapsack-problem (see, e.g., [28]); the same knapsack-approach was used in [31] to obtain the so-called quasi-optimal sparse grids for PDEs with stochastic coefficients and in [7, 20] in the context of sparse-grid resolution of high-dimensional PDEs.

The resulting method can be seen as an extension of the sparse-grid combination technique for PDEs with stochastic coefficients, as well as a fully sparse, non-randomized version of the Multilevel Monte Carlo method [2, 10, 17, 26]. In particular, MISC differs from other works in the literature that attempt to optimally combine spatial and stochastic resolution levels [5, 24, 27, 35, 37] in two aspects. First, MISC uses combined deterministic-stochastic, first-order differences, which allows us to exploit not only the regularity of the solution with respect to the spatial variables and the stochastic parameters, but also the mixed deterministic-stochastic regularity whenever available. Second, the MISC estimator is built upon an optimization procedure, whereas the above-mentioned works try to balance the error contributions arising from the deterministic and stochastic components of the method without taking into account the correspoding costs. Finally, MISC can also be seen as a sparse-grid quadrature version of the Multi-index Monte Carlo method we previously proposed and analyzed in [23].

In [22], we looked at easy cases of problems of type (1) depending on a <u>finite</u> number of random variables, $\boldsymbol{y} \in \Gamma \subset \mathbb{R}^N$. Here, we provide a complexity analysis of MISC in the challenging case of a countable sequence of random variables, $\{y_j\}_{j\geq 1}$. This new framework requires that the tools used to prove the complexity of the method be changed: while in [22] we used a "direct counting" argument, i.e., we derived a complexity estimate by explicitly summing the work and the error contributions associated with each mixed difference included in the MISC estimator, here we base our proof on a summability argument and on suitable interpolation estimates in mixed regularity spaces.

The rest of this work is organized as follows. Section 2 introduces a specific example of a random diffusion coefficient that we consider throughout the work as well as the functional analysis results that are needed for the subsequent analysis of the MISC method. The MISC method is presented in Section 3 and its complexity analysis is carried out in Sections 4 and 5, where we provide a general convergence theorem. We then discuss its application to the specific class of diffusion coefficients that we consider here: in particular, we track the dependence of the convergence rate on the regularity of the diffusion coefficient. Section 6 presents some numerical tests to verify the convergence analysis carried out in the previous section. Finally, Section 7 offers some conclusions and final remarks. We also provide an appendix that includes some technical results on the summability and regularity properties of certain random fields written in terms of their series expansion.

In the following, $\mathbb{N}$ denotes the set of integer numbers including zero, while $\mathbb{N}_+$ denotes the set of positive integer numbers, i.e., excluding zero. We refer to vectors in $\mathbb{N}^{\mathbb{N}}$ as "multi-indices". Moreover, we often use a vector notation for sequences, i.e., we formally treat sequences as vectors in $\mathbb{N}^{\mathbb{N}}$ (or $\mathbb{R}^{\mathbb{N}}$) and denote them with bold symbols. In particular, we use the following notation, with the understanding that $N < \infty$ for actual vectors and $N = \infty$ for sequences:

- $\mathbf{1}$ denotes a vector in $\mathbb{N}^N$ whose components are all equal to one;
- $\boldsymbol{e}_\ell^N$ denotes the ℓ-th canonical vector in $\mathbb{R}^N$, i.e., $\left(\boldsymbol{e}_\ell^N\right)_i = 1$ if $\ell = i$ and zero otherwise; however, for the sake of clarity, we often omit the superscript N whenever obvious from the context. For instance, if $\boldsymbol{v} \in \mathbb{R}^N$, we write $\boldsymbol{v} - \boldsymbol{e}_1$ instead of $\boldsymbol{v} - \boldsymbol{e}_1^N$;
- given $\boldsymbol{v} \in \mathbb{R}^N$, $|\boldsymbol{v}| = \sum_{n=1}^N v_n$, $|\boldsymbol{v}|_0$ denotes the number of non-zero components of $\boldsymbol{v}$, $\max(\boldsymbol{v}) = \max_{n=1,\dots N} v_n$ and $\min(\boldsymbol{v}) = \min_{n=1,\dots N} v_n$;
- $\mathfrak{L}$ denotes the set of finitely supported sequences, i.e., $\mathfrak{L} = \{\boldsymbol{p} \in \mathbb{N}^{\mathbb{N}} : |\boldsymbol{p}|_0 < \infty\}$;
- $\mathfrak{L}_+$ denotes the set of sequences with positive components with only finitely many elements larger than 1, i.e., $\mathfrak{L}_+ = \{\boldsymbol{\beta} \in \mathbb{N}_+^{\mathbb{N}} : |\boldsymbol{\beta} - \mathbf{1}|_0 < \infty\}$;
- 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 $\boldsymbol{v}$, $f(\boldsymbol{v}) = [f(v_1), f(v_2), \cdots, f(v_N)] \in \mathbb{R}^N$;
- 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 $\boldsymbol{w} \in \mathbb{R}^N$, $[\boldsymbol{v}, \boldsymbol{w}] = (v_1, \dots, v_D, w_1, \dots, w_N) \in \mathbb{R}^{D+N}$;
- given $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^N$, we denote by $\boldsymbol{v}^{\boldsymbol{w}}$ the product $\prod_{n=1}^N v_n^{w_n}$.

2 Functional setting

To ensure the necessary parametric regularity needed in our error analysis on the solution, u, to (1), we make the following assumption:

Assumption A2 (Expansion of the diffusion coefficient)**.** *The diffusion coefficient, $a(\boldsymbol{x}, \boldsymbol{y})$, has the following expression:*

$$a(\boldsymbol{x}, \boldsymbol{y}) = e^{\kappa(\boldsymbol{x}, \boldsymbol{y})}, \text{ with } \kappa(\boldsymbol{x}, \boldsymbol{y}) = \sum_{j \in \mathbb{N}_+} \psi_j(\boldsymbol{x}) y_j. \tag{2}$$

Here, $\{\psi_j\}_{j \in \mathbb{N}_+}$ is a sequence of functions $\psi_j \in C^\infty(\overline{\mathcal{B}})$, such that $\{\|\psi_j\|_{L^\infty(\mathcal{B})}\}_{j \in \mathbb{N}_+}$ is decreasing. Moreover, given the sequences

$$b_{0,j} = \|\psi_j\|_{L^\infty(\mathcal{B})}, \qquad j \geq 1, \tag{3}$$

$$b_{s,j} = \max_{\boldsymbol{s} \in \mathbb{N}^d : |\boldsymbol{s}| \leq s} \|D^{\boldsymbol{s}} \psi_j\|_{L^\infty(\mathcal{B})}, \qquad j \geq 1, \tag{4}$$

we assume that there exist $0 < p_0 \leq p_s < \frac{1}{2}$ such that $\{b_{0,j}\}_{j \in \mathbb{N}_+} \in \ell^{p_0}$ and $\{b_{s,j}\}_{j \in \mathbb{N}_+} \in \ell^{p_s}$, i.e.,

$$\|\boldsymbol{b}_0\|_{\ell^{p_0}}^{p_0} = \sum_{j \in \mathbb{N}_+} b_{0,j}^{p_0} < \infty \quad \text{and} \quad \|\boldsymbol{b}_s\|_{\ell^{p_s}}^{p_s} = \sum_{j \in \mathbb{N}_+} b_{s,j}^{p_s} < \infty. \tag{5}$$

We observe that with the above definitions, $b_{s,j} \to 0^+$ for $j \to \infty$ and $0 \leq b_{0,j} \leq b_{s,j}$. Moreover, given Assumption A2, we have that $\boldsymbol{b}_0 \in \ell^1$, which, together with the fact that $y_j \in [-1,1]$, guarantees that Assumption A1 holds true and therefore that (1) is well posed in V almost surely in Γ. However, the conditions in Assumption A2 are sufficient but not necessary for Assumption A1 to hold. Indeed, Assumption A1 holds for (2) if $\boldsymbol{b}_s \in \ell^2$ for any $s > 0$, see Lemma 16 (and Corollary 17 for a specific example of the diffusion coefficient, ⊣) in the appendix. The summability of the sequence $\boldsymbol{b}_s$ plays a central role in this work: indeed, if $\boldsymbol{b}_s$ is p_s-summable, u almost surely belongs to a subspace of V with additional regularity properties that allow us to to show convergence of the MISC method, with convergence rate dictated by p_0 and p_s. It is also important to remark that assumption $p_s < \frac{1}{2}$ could be relaxed to $p_s < 1$; yet, we work under this more stringent assumption since it considerably simplifies some technical steps in the following discussion without affecting the main part of the proof, as we make clearer later on (see Remark 5).

Example 1. *In the numerical section of this work, we consider the following form for $\kappa(\boldsymbol{x},\boldsymbol{y})$:*

$$\kappa(\boldsymbol{x},\boldsymbol{y}) = \sum_{\boldsymbol{k}\in\mathbb{N}^d} A_{\boldsymbol{k}} \sum_{\boldsymbol{\ell}\in\{0,1\}^d} y_{\boldsymbol{k},\boldsymbol{\ell}} \prod_{j=1}^{d} \left(\cos\left(\frac{\pi}{L}k_j x_j\right)\right)^{\ell_j} \left(\sin\left(\frac{\pi}{L}k_j x_j\right)\right)^{1-\ell_j}, \tag{6}$$

where the coefficients $A_{\boldsymbol{k}}$ are taken as

$$A_{\boldsymbol{k}} = \left(\sqrt{3}\right) 2^{\frac{|\boldsymbol{k}|_0}{2}} (1+|\boldsymbol{k}|^2)^{-\frac{\nu+d/2}{2}}, \tag{7}$$

for some $\nu > 0$. We observe that ν is a parameter dictating the $\boldsymbol{x}$-regularity of the realizations of κ, hence of a. Moreover, the parameters ν and d govern the p_s-summability of the sequence $\boldsymbol{b}_s$ (and as a consequence the overall convergence of the MISC method). Section 5 analyzes the the summability of the series (6)*.*

As already suggested, even though the problem is well posed in V, we need to make sure that realizations of u almost surely belong to more regular spaces to prove a convergence result for MISC. More specifically, due to the classic spatial spars-grid approximation theory, we need certain integrability conditions on the mixed derivatives of u. To this end, we introduce some suitable functional spaces and introduce a "shift" assumption, i.e., we assume that the diffusion coefficient, a, and the forcing, f, are such that realizations of u are sufficiently regular. This assumption needs to be stated in the complex domain, for reasons that should be clearer in a moment. Thus, we consider the diffusion coefficient

$$a(\boldsymbol{x},\boldsymbol{z}) = e^{\kappa(\boldsymbol{x},\boldsymbol{z})}, \qquad \kappa(\boldsymbol{x},\boldsymbol{z}) = \sum_{j\in\mathbb{N}_+} \psi_j(\boldsymbol{x}) z_j, \qquad z_j \in \mathbb{C},$$

and the corresponding solution of (1), $u(\boldsymbol{x},\boldsymbol{z})$, which is now a $H_0^1(\mathcal{B})$ function taking values in $\mathbb{C}$, $u(\cdot,\boldsymbol{z}) \in H_0^1(\mathcal{B},\mathbb{C})$.

Definition 1 (Fractional Sobolev spaces)**.** *For a given $q \geq 0$ and $r, r_1 \ldots, r_d$ positive real numbers, let*

$$\begin{aligned} &H^r(0,1) = \left\{ u \in H^{\lfloor r \rfloor}(0,1) : \int_{[0,1]}\int_{[0,1]} \frac{|D^{\lfloor r \rfloor}u(x) - D^{\lfloor r \rfloor}u(x')|^2}{|x-x'|^{1+2(r-\lfloor r \rfloor)}} dx dx' < \infty \right\}, \\ &H^{r_1,\ldots,r_d}(\mathcal{B}) = H^{r_1}(0,1) \otimes \ldots \otimes H^{r_d}(0,1), \\ &\mathcal{H}^{1+q}(\mathcal{B}) = H^{1+q,q,\ldots,q}(\mathcal{B}) \cap H^{q,1+q,\ldots,q}(\mathcal{B}) \cap H^{q,q,\ldots,1+q}(\mathcal{B}). \end{aligned}$$

In particular, we recall that $H^{r_1,\ldots,r_d}(\mathcal{B}) = H^{r_1}(0,1) \otimes \ldots \otimes H^{r_d}(0,1)$ is the completion of formal sums $v = \sum_{k=1}^{K} v_{1,k} v_{2,k} \cdots v_{d,k}$ with $v_{j,k} \in H^{r_j}(0,1)$ with respect to the norm induced by the inner product $(v,w) = \sum_{k,j} (v_{1,k}, w_{1,j})_{H^{r_1}(0,1)} (v_{2,k}, w_{2,j})_{H^{r_2}(0,1)} \cdots (v_{d,k}, w_{d,j})_{H^{r_d}(0,1)}$. It holds that $\mathcal{H}^1(\mathcal{B}) = H^1(\mathcal{B})$, and in general we have the following inclusion result:

$$u \in H^{1+r}(\mathcal{B}) \Rightarrow u \in \mathcal{H}^{1+q}(\mathcal{B}), \quad \text{for } 0 < q < r/d. \tag{8}$$

Since our approximation method will only be applied to finite subsets of parameters z_j, $j \in \mathbb{N}_+$, we introduce the following notation. Let $\mathcal{G} \subset \mathbb{N}_+$ be a finite subset of indices with cardinality $\#\mathcal{G} = G$:

we denote by $\mathbb{C}^{\mathcal{G}}$ the space of complex-valued sequences with zero components outside the set $\mathcal{G}$, i.e.,

$$\mathbb{C}^{\mathcal{G}} = \{\boldsymbol{z} \in \mathbb{C}^{\mathbb{N}_+} : z_j = 0, \quad \forall j \notin \mathcal{G}\}. \tag{9}$$

We observe that for any $\boldsymbol{z} \in \mathbb{C}^{\mathcal{G}}$, we have $\kappa(\cdot, \boldsymbol{z}) \in C^s(\overline{\mathcal{B}}, \mathbb{C})$, $\|\kappa(\cdot, \boldsymbol{z})\|_{C^s(\overline{\mathcal{B}},\mathbb{C})} \leq \sum_{j \in \mathcal{G}} |z_j| b_{s,j}$ and infer, from the multivariate Faà di Bruno formula (see Appendix A and [14]), that $a(\boldsymbol{z}) \in C^s(\overline{\mathcal{B}}, \mathbb{C})$ as well, with the estimate

$$\|a(\cdot, \boldsymbol{z})\|_{C^s(\overline{\mathcal{B}},\mathbb{C})} \leq \frac{s!}{(\log 2)^s} \|a(\cdot, \boldsymbol{z})\|_{C^0(\overline{\mathcal{B}},\mathbb{C})} (1 + \|\kappa(\cdot, \boldsymbol{z})\|_{C^s(\overline{\mathcal{B}})})^s, \quad \forall \boldsymbol{z} \in \mathbb{C}^{\mathcal{G}}. \tag{10}$$

Also, the complex-valued problem (1) is well posed as long as $\mathfrak{Re}\,[a(\boldsymbol{x}, \boldsymbol{z})] > 0$ for almost every (a.e.) $\boldsymbol{x} \in \mathcal{B}$. We therefore define the following region in $\mathbb{C}^{\mathcal{G}}$:

$$\Sigma_{\mathcal{G},\delta} = \{\boldsymbol{z} \in \mathbb{C}^{\mathcal{G}} : \mathfrak{Re}\,[a(\boldsymbol{x}, \boldsymbol{z})] \geq \delta > 0 \text{ for a.e. } \boldsymbol{x} \in \mathcal{B}\}. \tag{11}$$

Assumption A3 (Shift assumption). *We assume that $f \in C_0^\infty(\overline{\mathcal{B}})$ and $a(\cdot, \boldsymbol{z})$ are such that for any finite set $\mathcal{G}$ of random variables and for any $\boldsymbol{z} \in \Sigma_{\mathcal{G},\delta}$ the three following conditions hold*

1. *$u(\boldsymbol{z}) \in H^{1+s}(\mathcal{B}, \mathbb{C}) \cap H_0^1(\mathcal{B}, \mathbb{C})$;*
2. *$\frac{du}{dz_i} \in H^{1+s}(\mathcal{B}, \mathbb{C})$, $\forall i \in \mathcal{G}$, where $\frac{du}{dz_i}$ denotes the partial complex derivative of u;*
3. *$\|u(\cdot, \boldsymbol{z})\|_{H^{1+s}(\mathcal{B},\mathbb{C})} \leq C(\delta, s)(\|a(\cdot, \boldsymbol{z})\|_{C^s(\overline{\mathcal{B}},\mathbb{C})} + \|f\|_{H^{s-1}(\mathcal{B})})$, with $C(\delta, s) \to \infty$ for $\delta \to 0$.*

Remark 1 (Shift assumption in Example 1). *The family of random fields presented in Example 1 does not actually satisfy Assumption A3, because of the sine functions appearing in the expansion. An analogous expansion with only cosine terms would conversely satisfy Assumption A3 (cf. Appendix B).*

Finally, we state some summability and regularity results for u with respect to the random sequence, $\boldsymbol{z}$.

Definition 2 (Bernstein polyellipse). *For any $\zeta > 1$, let $\mathcal{E}_\zeta$ denote the ellipse in the complex plane*

$$\mathcal{E}_\zeta = \left\{ z \in \mathbb{C} : \ \mathfrak{Re}\,[z] \leq \frac{\zeta + \zeta^{-1}}{2} \cos\vartheta, \ \ \mathfrak{Im}\,[z] \leq \frac{\zeta - \zeta^{-1}}{2} \sin\vartheta, \ \vartheta \in [0, 2\pi) \right\}.$$

Then, for any sequence $\boldsymbol{\zeta} = \{\zeta_j\}_{j \in \mathbb{N}_+}$ with $\zeta_j > 1$ for every $j \in \mathbb{N}_+$ and for any finite set of random variables $\mathcal{G} = \{j_1, j_2, \ldots, j_G\}$, we introduce the Bernstein polyellipse:

$$\mathcal{E}_{\mathcal{G},\boldsymbol{\zeta}} = \{\boldsymbol{z} \in \mathbb{C}^{\mathbb{N}_+} : z_j \in \mathcal{E}_{\zeta_j} \ \ if \ j \in \mathcal{G}, z_j = 0 \ \ if \ j \notin \mathcal{G}\}.$$

For ease of notation, in the finite-dimensional case, i.e., $\mathcal{G} = \{1, 2, \ldots, N\}$ and $\boldsymbol{\zeta} \in \mathbb{R}^N$, we write $\mathcal{E}_{\boldsymbol{\zeta}}$ instead of $\mathcal{E}_{\mathcal{G},\boldsymbol{\zeta}}$, i.e., $\mathcal{E}_{\boldsymbol{\zeta}} = \{\boldsymbol{z} \in \mathbb{C}^N : z_j \in \mathcal{E}_{\zeta_j}$ for $j = 1, 2, \ldots, N\}$.

Lemma 1 (Holomorphic complex continuation of u in H_0^1 in a Bernstein polyellipse). *Consider the sequence $\boldsymbol{b}_0$ defined in equation (3). For any $\delta > 0$, let $E_\delta > 2$ be such that*

$$\frac{\pi}{E_\delta} = -\|\boldsymbol{b}_0\|_{\ell^1} - \log\delta + \log\cos\left(\frac{\pi}{E_\delta}\right),$$

and consider the sequence $\boldsymbol{\zeta}_0 = \{\zeta_{0,n}\}_{n \in \mathbb{N}_+}$, with

$$\zeta_{0,n} = \tau_{0,n} + \sqrt{\tau_{0,n}^2 + 1} > 1 \tag{12}$$

$$\tau_{0,n} = \frac{\pi}{E_\delta} \frac{(b_{0,n})^{p_0 - 1}}{\|\boldsymbol{b}_0\|_{\ell^{p_0}}^{p_0}}\ , \tag{13}$$

with p_0 as in equation (5). Then, for any finite set of random variables $\mathcal{G} = \{j_1, j_2, \ldots, j_G\}$, u admits a holomorphic complex continuation, $u : \mathbb{C}^{\mathcal{G}} \to H_0^1(\mathcal{B}, \mathbb{C})$, in the Bernstein polyellipse $\mathcal{E}_{\mathcal{G},\boldsymbol{\zeta}_0}$, with $\sup_{\boldsymbol{z} \in \mathcal{E}_{\mathcal{G},\boldsymbol{\zeta}_0}} \|u(\cdot, \boldsymbol{z})\|_{H^1(\mathcal{B})} \leq C_u = \frac{\|f\|_{H^{-1}(\mathcal{B})}}{\delta} < \infty$, with $\mathbb{C}^{\mathcal{G}}$ as in Equation (9) and C_u independent of $\mathcal{G}$.

Proof. It is well known in the literature that $u : \mathbb{C}^{\mathcal{G}} \to H^1_0(\mathcal{B},\mathbb{C})$ is holomorphic in $\Sigma_{\mathcal{G},\delta}$ in (11) (see, e.g., [1]). To compute the parameters $\zeta_{j_1}, \zeta_{j_2}, \ldots, \zeta_{j_G}$ of a Bernstein ellipse contained in $\Sigma_{\mathcal{G},\delta}$, we rewrite $a(\boldsymbol{x},\boldsymbol{z})$ as

$$\begin{aligned} a(\boldsymbol{x},\boldsymbol{z}) &= \exp\left(\sum_{j\in\mathcal{G}} z_j\psi_j(\boldsymbol{x})\right) = \exp\left(\sum_{j\in\mathcal{G}} \mathfrak{Re}\,[z_j]\psi_j(\boldsymbol{x})\right)\exp\left(\sum_{j\in\mathcal{G}} i\mathfrak{Im}\,[z_j]\,\psi_j(\boldsymbol{x})\right)\\ &= \exp\left(\sum_{j\in\mathcal{G}} \mathfrak{Re}\,[z_j]\psi_j(\boldsymbol{x})\right)\left[\cos\left(\sum_{j\in\mathcal{G}} \mathfrak{Im}\,[z_j]\,\psi_j(\boldsymbol{x})\right) + i\sin\left(\sum_{j\in\mathcal{G}} \mathfrak{Im}\,[z_j]\,\psi_j(\boldsymbol{x})\right)\right], \end{aligned}$$

so that $\Sigma_{\mathcal{G},\delta}$ can be rewritten as

$$\Sigma_{\mathcal{G},\delta} = \left\{\boldsymbol{z}\in\mathbb{C}^{\mathcal{G}} : \exp\left(\sum_{j\in\mathcal{G}} \mathfrak{Re}\,[z_j]\psi_j(\boldsymbol{x})\right)\cos\left(\sum_{j\in\mathcal{G}} \mathfrak{Im}\,[z_j]\,\psi_j(\boldsymbol{x})\right) \geq \delta \;\text{ for a.e. } \boldsymbol{x}\in\mathcal{B}\right\}.$$

Now, for some $E > 2$ that we choose in the following, the two following conditions on $\boldsymbol{z}$ imply that $\boldsymbol{z}\in\Sigma_{\mathcal{G},\delta}$:

$$\begin{cases} \cos\Big(\sum_{j\in\mathcal{G}} |\mathfrak{Im}\,[z_j]\,|\,b_{0,j}\Big) \geq \cos\left(\dfrac{\pi}{E}\right)\\ \exp\Big(-\sum_{j\in\mathcal{G}} |\mathfrak{Re}\,[z_j]|\,b_{0,j}\Big) \geq \dfrac{\delta}{\cos\left(\frac{\pi}{E}\right)}; \end{cases}$$

equivalently, we write

$$\begin{cases} \sum_{j\in\mathcal{G}} |\mathfrak{Im}\,[z_j]|\,b_{0,j} \leq \dfrac{\pi}{E}\\ \sum_{j\in\mathcal{G}} |\mathfrak{Re}\,[z_j]|\,b_{0,j} \leq -\log\delta + \log\cos\left(\dfrac{\pi}{E}\right). \end{cases}$$

For a fixed E, the equations above define a second region, Σ', included in $\Sigma_{\mathcal{G},\delta}$. In turn, these conditions are verified if the following conditions, which define an hyperrectangular region $\Sigma''\subset\Sigma'$, are verified

$$\begin{cases} |\mathfrak{Im}\,[z_j]| \leq \tau_{0,j} = \dfrac{\pi(b_{0,j})^{p_0-1}}{E\,\|\boldsymbol{b}_0\|^{p_0}_{\ell^{p_0}}},\\ |\mathfrak{Re}\,[z_j]| \leq 1 + w_{0,j}, \qquad \text{with } w_{0,j} = \dfrac{(b_{0,j})^{p_0-1}}{\|\boldsymbol{b}_0\|^{p_0}_{\ell^{p_0}}}\left(-\|\boldsymbol{b}_0\|_{\ell^1} - \log\delta + \log\cos\left(\dfrac{\pi}{E}\right)\right), \end{cases}$$

provided that δ and E are such that the quantity $-\|\boldsymbol{b}_0\|_{\ell^1} - \log\delta + \log\cos\left(\frac{\pi}{E}\right)$ remains positive: observe that such δ and E exist, since $f(E) = \log\cos\left(\frac{\pi}{E}\right)$ is a monotonically increasing function, with $f(E)\to-\infty$ for $E\to 2$ and $f(E)\to 0$ for $E\to\infty$, and $-\log\delta$ is positive for sufficiently small δ. In particular, for any $\delta > 0$, we choose $E = E_\delta$ such that $w_{0,j} = \tau_{0,j}$, which leads to

$$\frac{\pi}{E_\delta} = -\|\boldsymbol{b}_0\|_{\ell^1} - \log\delta + \log\cos\left(\frac{\pi}{E_\delta}\right).$$

We observe that with this choice, $\tau_{0,j}$ (and hence $w_{0,j}$) actually does not depend on the set of variables, $\mathcal{G}$, considered, and we can define the sequence $\boldsymbol{\tau}_0 = \{\tau_{0,n}\}_{n\in\mathbb{N}_+}$.

We are now in the position to compute the Bernstein ellipses that touch the boundary of Σ'' on the real and imaginary axis. For the real axis, we have to enforce

$$\frac{\zeta_{n,\text{real}} + \zeta^{-1}_{n,\text{real}}}{2} = 1 + \tau_{0,n} \Rightarrow \zeta_{n,\text{real}} = 1 + \tau_{0,n} + \sqrt{(1+\tau_{0,n})^2 - 1},$$

while for the imaginary axis we have to enforce

$$\frac{\zeta_{n,\text{imag}} - \zeta^{-1}_{n,\text{imag}}}{2} = \tau_{0,n} \Rightarrow \zeta_{n,\text{imag}} = \tau_{0,n} + \sqrt{\tau^2_{0,n} + 1}.$$

The proof is concluded by observing that $\zeta_{n,\text{imag}} \leq \zeta_{n,\text{real}}$, i.e., the only ellipse entirely contained in Σ'', and hence in $\Sigma_{\mathcal{G},\delta}$, is the one touching Σ'' on the imaginary axis, which also implies that the bound $\sup_{\boldsymbol{z}\in\mathcal{E}_{\mathcal{G},\boldsymbol{\zeta}}} \|u(\cdot,\boldsymbol{z})\|_{H^1(\mathcal{B})} \leq C_u = \frac{\|f\|_{H^{-1}(\mathcal{B})}}{\delta} < \infty$ holds independently of $\mathcal{G}$. □

Lemma 2 (Holomorphic complex continuation of u in H^{1+s} in a Bernstein polyellipse)**.** *Let* $\boldsymbol{\zeta}_s = \{\zeta_{s,n}\}_{n\in\mathbb{N}_+}$*, with*

$$\zeta_{s,n} = \tau_{s,n} + \sqrt{\tau_{s,n}^2 + 1} > 1,$$
$$\tau_{s,n} = \frac{\pi}{E_\delta}\frac{(b_{s,n})^{p_s-1}}{\|\boldsymbol{b}_s\|_{\ell^{p_s}}^{p_s}},$$

with $\boldsymbol{b}_s$ *as in equation* (4)*,* p_s *as in equation* (5)*, and* E_δ *as in Lemma 1. For any finite set of random variables* $\mathcal{G} = \{j_1, j_2, \ldots, j_G\}$*,* $u : \mathbb{C}^{\mathcal{G}} \to H^{1+s}(\mathcal{B}, \mathbb{C})$ *is holomorphic in the Bernstein polyellipse* $\mathcal{E}_{\mathcal{G},\boldsymbol{\zeta}_s}$*, with* $\mathbb{C}^{\mathcal{G}}$ *as in Equation* (9) *and*

$$\sup_{\boldsymbol{z}\in\mathcal{E}_{\mathcal{G},\boldsymbol{\zeta}_s}} \|u(\cdot,\boldsymbol{z})\|_{H^{1+s}(\mathcal{B})} \le C_{s,u} = C(\tilde\delta, s)(M + \|f\|_{H^{s-1}(\mathcal{B})}) < \infty$$

with $M = \frac{s!}{(\log 2)^s} e^{K\frac{\pi}{E_\delta}}\left(1 + K\frac{\pi}{E_\delta}\right)^s$*,* $K = \left(2 + \frac{1}{\min_{n\in\mathbb{N}_+}\tau_{s,n}^2}\right)^{1/2}$*,* $\tilde\delta = e^{-K\frac{\pi}{E_\delta}}$*,* $C(\tilde\delta, s)$ *as in Assumption A3, and* $C_{s,u}$ *independent of* $\mathcal{G}$*.*

Proof. From Assumption A3, $u : \mathbb{C}^{\mathcal{G}} \to H^{1+s}(\mathcal{B}, \mathbb{C})$ is complex-differentiable for every $\boldsymbol{z}$ in $\Sigma_{\mathcal{G},\delta}$ for any $\delta > 0$. It is therefore holomorphic in $\Sigma_{\mathcal{G},\delta}$. Similarly to the previous lemma, we look for a region in which we have an a-priori bound on the $H^{1+s}(\mathcal{B}, \mathbb{C})$ norm of u uniformly on $\boldsymbol{z}$. Again from Assumption A3, we have that this is true in the region

$$\Sigma' = \{\boldsymbol{z} \in \mathbb{C}^{\mathcal{G}} : \|a(\cdot,\boldsymbol{z})\|_{C^s(\overline{\mathcal{B}})} \le M\} \cap \Sigma_{\mathcal{G},\tilde\delta} \quad \text{for any } \tilde\delta > 0.$$

Contrary to the previous lemma, in this proof we do not derive the expression of an ellipse contained in Σ', but content ourselves with verifying that the ellipses $\mathcal{E}_{\mathcal{G},\boldsymbol{\zeta}_s}$ proposed in the statement of the lemma (that we have obtained simply by replacing $b_{0,n}$ with $b_{s,n}$ in Equation (13)) satisfy the requirement, i.e., $\mathcal{E}_{\mathcal{G},\boldsymbol{\zeta}_s} \subset \Sigma'$, for every finite set of random variables, $\mathcal{G}$. To this end, let us consider the univariate ellipse $\mathcal{E}_{\zeta_{s,n}}$. We first prove that this ellipse is contained in the rectangle in the complex domain

$$\mathcal{R}_n = \{z \in \mathbb{C} : |\mathfrak{Re}\,[z]| \le \sqrt{1 + \tau_{s,n}^2},\ |\mathfrak{Im}\,[z]\,| \le \tau_{s,n}\}.$$

The bound on the imaginary part of z is an immediate consequence of the choice of the ellipse. As for the real part, we compute the point z_0 where the ellipse intersects the real axis by equating

$$\begin{aligned} z_0 =& \frac{\zeta_{s,n} + \frac{1}{\zeta_{s,n}}}{2} = \frac{\zeta_{s,n}^2 + 1}{2\zeta_{s,n}} = \frac{\tau_{s,n}^2 + 1 + \tau_{s,n}\sqrt{\tau_{s,n}^2 + 1}}{\tau_{s,n} + \sqrt{\tau_{s,n}^2 + 1}} \\ =& \left(\tau_{s,n}^2 + 1 + \tau_{s,n}\sqrt{\tau_{s,n}^2 + 1}\right)\left(\sqrt{\tau_{s,n}^2 + 1} - \tau_{s,n}\right) = \sqrt{1 + \tau_{s,n}^2}. \end{aligned}$$

Furthermore, we observe that for every $z \in \mathcal{R}_n$, it holds that $|z| \le \sqrt{1 + 2\tau_{s,n}^2} \le K\tau_{s,n}$ for some $K > 0$; for instance, we could look for the smallest $\tau_{s,n}$, say τ_{s,n^*}, choose K accordingly, i.e., such that $(K^2 - 2)\tau_{s,n^*}^2 \ge 1$, and obtain the value in the statement of the lemma (recall that we have ordered variables according to $b_{0,n}$, hence $\tau_{0,n}$ is necessarily increasing, but $\tau_{s,n}$ is not necessarily so). Next, according to equation (10) and Assumption A2,

$$\|a(\cdot,\boldsymbol{z})\|_{C^s(\overline{\mathcal{B}},\mathbb{C})} \le \frac{s!}{(\log 2)^s}\|a(\cdot,\boldsymbol{z})\|_{C^0(\overline{\mathcal{B}},\mathbb{C})}(1 + \|\kappa(\cdot,\boldsymbol{z})\|_{C^s(\overline{\mathcal{B}},\mathbb{C})})^s \le \frac{s!}{(\log 2)^s} e^{\sum_{j\in\mathcal{G}} b_{0,j}|z_j|}\left(1 + \sum_{j\in\mathcal{G}} b_{s,j}|z_j|\right)^s,$$

holds. We then conclude by observing that for every $\boldsymbol{z} \in \mathcal{E}_{\mathcal{G},\boldsymbol{\zeta}_s}$, we have

$$\sum_{j\in\mathcal{G}} b_{0,j}|z_j| \le \sum_{j\in\mathcal{G}} b_{s,j}|z_j| \le K\sum_{j\in\mathcal{G}} b_{s,j}\tau_{s,j} = K\frac{\pi}{E_\delta}\sum_{j\in\mathcal{G}} b_{s,j}\frac{(b_{s,j})^{p_s-1}}{\|\boldsymbol{b}_s\|_{\ell^{p_s}}^{p_s}} \le K\frac{\pi}{E_\delta},$$

which gives uniform control of the norm of $\|a(\cdot,\boldsymbol{z})\|_{C^s(\overline{\mathcal{B}},\mathbb{C})}$ within $\mathcal{E}_{\mathcal{G},\boldsymbol{\zeta}_s}$ as required. More precisely, we have

$$\|a(\cdot,\boldsymbol{z})\|_{C^s(\overline{\mathcal{B}},\mathbb{C})} \le M = \frac{s!}{(\log 2)^s} e^{K\frac{\pi}{E_\delta}}\left(1 + K\frac{\pi}{E_\delta}\right)^s,$$

which, together with Assumption A3, gives the desired bound on $\|u(\cdot,\boldsymbol{z})\|_{H^{1+s}(\mathcal{B})}$. Moreover,

$$\mathfrak{Re}\left[a(\boldsymbol{x},\boldsymbol{z})\right] \geq e^{-K\frac{\pi}{E_\delta}} =: \tilde{\delta} > 0,$$

i.e., we can also control the coercivity of the problem. □

Lemma 3 (Chebyshev expansion of a holomorphic function)**.** *Let $\phi_{q_n}(y_n)$ be the family of Chebyshev polynomials of the first kind on $[-1,1]$ [1] and, for any $\boldsymbol{p} \in \mathbb{N}^N$, let $\Phi_{\boldsymbol{p}}(\boldsymbol{y}) = \prod_{n=1}^N \phi_{p_n}(y_n)$. If $f : [-1,1]^N \to \mathbb{R}$ admits an analytic complex extension in a Bernstein polyellipse $\mathcal{E}_{\boldsymbol{\zeta}}$ for some $\boldsymbol{\zeta} \in \mathbb{R}^N$ such that $\zeta_n > 1$ for all $n = 1,\dots,N$ and $\sup_{\boldsymbol{z}\in\mathcal{E}_{\boldsymbol{\zeta}}} |f(\boldsymbol{z})| \leq C_f$, then f admits the following Chebyshev expansion*

$$f(\boldsymbol{y}) = \sum_{\boldsymbol{p}\in\mathbb{N}^N} f_{\boldsymbol{p}} \Phi_{\boldsymbol{p}}(\boldsymbol{y}),$$

$$f_{\boldsymbol{p}} = \frac{1}{\int_{[-1,1]^N} \Phi_{\boldsymbol{p}}^2(\boldsymbol{y})\varrho_C(\boldsymbol{y})d\boldsymbol{y}} \int_{[-1,1]^N} f(\boldsymbol{y})\Phi_{\boldsymbol{p}}(\boldsymbol{y})\varrho_C(\boldsymbol{y})d\boldsymbol{y}, \quad \varrho_C(\boldsymbol{y}) = \prod_{n=1}^N \left(\sqrt{1-y_n^2}\right)^{-1},$$

which converges uniformely in $\mathcal{E}_{\boldsymbol{\zeta}}$. Moreover, if $\zeta_n > 2$ for any $n = 1,\dots,N$, the following bound on the coefficients $f_{\boldsymbol{p}}$ holds:

$$|f_{\boldsymbol{p}}| \leq \sup_{\boldsymbol{z}\in\mathcal{E}_{\boldsymbol{\zeta}}} |f(\boldsymbol{z})| \prod_{n=1}^N e^{-\tilde{g}_n p_n}, \quad \tilde{g}_n = \log\zeta_n - \log 2.$$

Proof. A straightforward extension to the N-dimensional case of the argument in [15, Chapter 7, Theorem 8.1] (see also [1]) allows us to write

$$|f_{\boldsymbol{p}}| \leq \sup_{\boldsymbol{z}\in\mathcal{E}_{\boldsymbol{\zeta}}} |f(\boldsymbol{z})| 2^{|\boldsymbol{p}|_0} \prod_{n=1}^N e^{-p_n \log\zeta_n},$$

where $|\boldsymbol{p}|_0$ denotes the number of non-zero elements of $\boldsymbol{p}$. By writing $2^{|\boldsymbol{p}|_0} = \prod_{n=1,p_n\neq 0}^N e^{\log 2}$ and setting $\tilde{g}_n = \log\zeta_n - \log 2$, we then obtain the final statement of the theorem. □

3 The Multi-index Stochastic Collocation method

In this section, we introduce approximations of $\mathbb{E}[F]$ along the deterministic and stochastic dimensions and their decomposition in terms of tensorizations of univariate difference operators. We then define the so-called mixed difference operators and build the MISC estimator by suitable sums of such operators, selecting them with a greedy optimization algorithm.

3.1 Approximation along the deterministic and stochastic variables

A tensorized deterministic solver. Consider a numerical method for the approximation of the solution of (1) for a fixed value of the random variables, $\boldsymbol{y}$, based on a quadrilateral/hexaedral mesh over $\mathcal{B}$ (e.g., finite differences, finite volumes, tensorized finite elements or h-refined bi- and tri-dimensonal splines, such as those used in the isogeometric context), and let $h_i, i = 1,\dots,d$ denote the mesh-size along each direction. The values of h_i are actually given as functions of an integer positive value, $\alpha \geq 1$, referred to as a "deterministic discretization level", i.e., $h_i = h_i(\alpha)$. Given a multi-index $\boldsymbol{\alpha} \in \mathbb{N}_+^d$, we denote by $u^{\boldsymbol{\alpha}}(\boldsymbol{x},\boldsymbol{y})$ the approximation of u obtained by setting $h_i = h_i(\alpha_i)$ and and use notation $F^{\boldsymbol{\alpha}}(\boldsymbol{y}) = \Theta[u^{\boldsymbol{\alpha}}(\cdot,\boldsymbol{y})]$.

It would be straightforward to extend this setting to discretization methods based on degree-elevation rather than on mesh-refinement, such as spectral methods, p-refined finite elements or p- and k-refined splines, however, in this work, we limit ourselves to the setting defined above for ease of presentation. It would also be possible to include in this framework time-dependent problems, but in this case we might need to take care of possible constraints on discretization parameters, such

[1] I.e., such that $|\phi_q(y)| \leq 1$ in $[-1,1]$ for every $q \in \mathbb{N}$.

as CFL conditions; a broader generalization could also include "non-physical" parameters such as tolerances for numerical solvers. Finally, more general problems, e.g., those depending on random variables with probability distributions other than uniform distributions, with uncertain boundary conditions and/or forcing terms, and, more importantly, defined on spatial domains that are not hyper-rectangles, could also be addressed with suitable modifications of the MISC methodology, as briefly mentioned in [22].

Tensorized quadrature formulae for expected value approximation. Similarly to what was done for the deterministic problem, we base our approximation of the expected value of $F^{\boldsymbol{\alpha}}(\boldsymbol{y})$ on a tensorization of quadrature formulae over the stochastic domain, Γ. Assumptions A2 and A3 guarantee that $F^{\boldsymbol{\alpha}}(\boldsymbol{y})$ is actually a continuous function (even analytic) over Γ. A quadrature approach is thus sound.

Let $C^0([-1,1])$ be the set of real-valued continuous functions over $[-1,1]$, $\beta \geq 1$ be an integer positive value referred to as a "stochastic discretization level", and $m : \mathbb{N} \to \mathbb{N}$ be a strictly increasing function with $m(0) = 0$ and $m(1) = 1$, that we call a "level-to-nodes function". At level β, we consider a set of $m(\beta)$ distinct quadrature points in $[-1,1]$, $\mathcal{H}^{m(\beta)} = \{y_\beta^1, y_\beta^2 \dots y_\beta^{m(\beta)}\} \subset [-1,1]$, and a set of quadrature weights, $\mathcal{W}^{m(\beta)} = \{\varpi_\beta^1, \varpi_\beta^2 \dots \varpi_\beta^{m(\beta)}\}$. We then define the quadrature operator as

$$Q^{m(\beta)} : C^0([-1,1]) \to \mathbb{R}, \qquad Q^{m(\beta)}[f] = \sum_{j=1}^{m(\beta)} f(y_\beta^j)\varpi_\beta^j. \tag{14}$$

The quadrature weights are selected such that

$$Q^{m(\beta)}[y^k] = \int_{-1}^{1} \frac{y^k}{2} dy, \quad k = 0, 1, \dots, m(\beta) - 1,$$

and the quadrature points are chosen to optimize the convergence properties of the quadrature error (the specific choice of quadrature points is discussed later in this section); in particular, for symmetry reasons, we define the trivial operator $Q^1[f] = f(0)\ \forall f \in C^0([-1,1])$.

Defining a quadrature operator over Γ is more delicate, since Γ is defined as a countable tensor product of intervals. To this end, we follow [34] and define, for any finitely supported multi-index $\boldsymbol{\beta} \in \mathfrak{L}_+$,

$$\mathcal{Q}^{m(\boldsymbol{\beta})} : \Gamma \to \mathbb{R}, \quad \mathcal{Q}^{m(\boldsymbol{\beta})} = \bigotimes_{n \geq 1} Q^{m(\beta_n)}$$

where the n-th quadrature operator is understood to act only on the n-th variable of f, and the tensor product is well defined since it is composed of finitely many non-trivial factors (see again [34]). In practice, the value of $\mathcal{Q}^{m(\boldsymbol{\beta})}[f]$ can be obtained by considering the tensor grid $\mathcal{T}^{m(\boldsymbol{\beta})} = \times_{n \geq 1} \mathcal{H}^{m(\beta_n)}$ with cardinality $\#\mathcal{T}^{m(\boldsymbol{\beta})} = \prod_{n \geq 1} m(\beta_n)$ and computing

$$\mathcal{Q}^{m(\boldsymbol{\beta})}[f] = \sum_{j=1}^{\#\mathcal{T}^{m(\boldsymbol{\beta})}} f(\widehat{\boldsymbol{y}}_j) w_j,$$

where $\widehat{\boldsymbol{y}}_j \in \mathcal{T}^{m(\boldsymbol{\beta})}$ and w_j are (infinite) products of weights of the univariate quadrature rules. Notice that it is essential in this construction that $m(1) = 1$ so that the cardinality of $\mathcal{T}^{m(\boldsymbol{\beta})}$ is finite for any $\boldsymbol{\beta} \in \mathfrak{L}_+$ and $\varpi_{\beta_n}^1 = 1$ whenever $\beta_n = 1$, so that all weights, w_j, are bounded.

Coming back to the choice of the univariate quadrature points, it is recommended, for optimal performance, that they are chosen according to the underlying measure, $\mathrm{d}\lambda/2$; moreover, since we aim at a hierarchical decomposition of the operator $\mathcal{Q}^{m(\boldsymbol{\beta})}$, it is useful (although not necessary, see e.g., [31]) that the nodes be nested collocation points, i.e. $\mathcal{H}^{m(\beta)} \subset \mathcal{H}^{m(\beta+1)}$ for any $\beta \geq 1$. Thus, in this work, we consider Clenshaw-Curtis points [31, 36] that are defined as:

$$y_\beta^j = \cos\left(\frac{(j-1)\pi}{m(\beta)-1}\right), \quad 1 \leq j \leq m(\beta), \tag{15}$$

and are nested provided that the level-to-nodes function is defined as

$$m(0) = 0,\ m(1) = 1,\ m(i_n) = 2^{i_n - 1} + 1. \tag{16}$$

We close this section by mentioning that another family of nested points for uniform measures available in the literature are the Leja points [11, 31], whose performance is equivalent to that of Clenshaw-Curtis for quadrature purposes (see [29, 30, 34]).

3.2 Construction of the MISC estimator

It is straightforward to see that a direct approximation $\mathbb{E}[F] \approx \mathcal{Q}^{m(\boldsymbol{\beta})}[F^{\boldsymbol{\alpha}}]$ is not a viable option in practical cases, due to the well-known "curse of dimensionality" effect. In [22], we proposed to use MISC as a computational strategy to combine spatial and stochastic discretizations in such a way as to obtain an effective approximation scheme for $\mathbb{E}[F]$.

MISC is based on a classic sparsification approach in which approximations like $\mathcal{Q}^{m(\boldsymbol{\beta})}[F^{\boldsymbol{\alpha}}]$ are decomposed in a hierarchy of operators. Only the most important of these operators are then retained in the approximation. In more detail, let us denote for brevity $\mathcal{Q}^{m(\boldsymbol{\beta})}[F^{\boldsymbol{\alpha}}] = F_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ and introduce the first-order difference operators for the deterministic and stochastic discretization operators, denoted respectively by Δ_i^{det} with $1 \le i \le d$ and Δ_j^{stoc} with $j \ge 1$:

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

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

As a second step, we define the so-called mixed difference operators,

$$\boldsymbol{\Delta}^{\text{det}}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}] = \bigotimes_{i=1}^{d} \Delta_i^{\text{det}}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}] = \Delta_1^{\text{det}}\left[\Delta_2^{\text{det}}\left[\cdots \Delta_d^{\text{det}}\left[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}\right]\right]\right] = \sum_{\boldsymbol{j}\in\{0,1\}^d} (-1)^{|\boldsymbol{j}|} F_{\boldsymbol{\alpha}-\boldsymbol{j},\boldsymbol{\beta}}, \tag{17}$$

$$\boldsymbol{\Delta}^{\text{stoc}}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}] = \bigotimes_{j\ge 1} \Delta_j^{\text{stoc}}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}] = \sum_{\boldsymbol{j}\in\{0,1\}^{\mathbb{N}}} (-1)^{|\boldsymbol{j}|} F_{\boldsymbol{\alpha},\boldsymbol{\beta}-\boldsymbol{j}}\,, \tag{18}$$

with the convention that $F_{\boldsymbol{v},\boldsymbol{w}} = 0$ whenever a component of $\boldsymbol{v}$ or $\boldsymbol{w}$ is zero. Notice that, since $\boldsymbol{\beta}$ has finitely many components larger than 1, the sum on the right-hand side of (18) contains only a finite number of terms. Finally, letting

$$\boldsymbol{\Delta}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}] = \boldsymbol{\Delta}^{\text{stoc}}[\boldsymbol{\Delta}^{\text{det}}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]], \tag{19}$$

we define the Multi-index Stochastic Collocation (MISC) estimator of $\mathbb{E}[F]$ as

$$\mathcal{M}_{\mathcal{I}}[F] = \sum_{[\boldsymbol{\alpha},\boldsymbol{\beta}]\in\mathcal{I}} \boldsymbol{\Delta}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}] = \sum_{[\boldsymbol{\alpha},\boldsymbol{\beta}]\in\mathcal{I}} c_{\boldsymbol{\alpha},\boldsymbol{\beta}} F_{\boldsymbol{\alpha},\boldsymbol{\beta}}, \quad c_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \sum_{\substack{[\boldsymbol{i},\boldsymbol{j}]\in\{0,1\}^{d+\mathbb{N}} \\ [\boldsymbol{\alpha}+\boldsymbol{i},\boldsymbol{\beta}+\boldsymbol{j}]\in\mathcal{I}}} (-1)^{|[\boldsymbol{i},\boldsymbol{j}]|_0}, \tag{20}$$

where $\mathcal{I} \subset \mathbb{N}_+^d \otimes \mathfrak{L}_+$. The second form of the MISC estimator is known as "combination technique", since it expresses the MISC approximation as a linear combination of a number of tensor approximations, $F_{\boldsymbol{\alpha},\boldsymbol{\beta}}$, and might be useful for the practical implementation of the method; we observe in particular that many of its coefficients, $c_{\boldsymbol{\alpha},\boldsymbol{\beta}}$, are zero.

The effectiveness of MISC crucially depends on the choice of the index set, $\mathcal{I}$. Given the hierarchical structure of MISC, a natural requirement is that $\mathcal{I}$ should be downward closed, i.e.,

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

(see also [7, 31, 38]). Beside this general constraint, in [22] we have detailed a procedure to derive an efficient set, $\mathcal{I}$, based on an optimization technique inspired by the Dantzig algorithm for the approximate solution of the knapsack problem (see [28]). In the following, we briefly summarize this procedure and refer to [22] as well as to [3, 7, 31] for a thorough discussion on the similarities between this procedure and the Dantzig algorithm.

The first step of our optimized construction consists of introducing the "work contribution", $\Delta W_{\boldsymbol{\alpha},\boldsymbol{\beta}}$, and "error contribution", $\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}}$, for each operator, $\boldsymbol{\Delta}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]$. 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.

$$\Delta W_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \text{Work}\left[\mathcal{M}_{\mathcal{I}\cup\{[\boldsymbol{\alpha},\boldsymbol{\beta}]\}}\right] - \text{Work}[\mathcal{M}_{\mathcal{I}}] = \text{Work}[\boldsymbol{\Delta}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]]. \tag{21}$$

Similarly, the error contribution measures how much the error, $|\mathbb{E}[F] - \mathcal{M}_{\mathcal{I}}[F]|$, would decrease if the operator $\boldsymbol{\Delta}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]$ were added to $\mathcal{M}_{\mathcal{I}}[F]$,

$$\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \left|\mathcal{M}_{\mathcal{I}\cup\{[\boldsymbol{\alpha},\boldsymbol{\beta}]\}}[F] - \mathcal{M}_{\mathcal{I}}[F]\right| = \left|\boldsymbol{\Delta}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]\right|. \tag{22}$$

We observe that the following decompositions of the total work and error of the MISC estimator hold:

$$\text{Work}[\mathcal{M}_{\mathcal{I}}] = \sum_{[\boldsymbol{\alpha},\boldsymbol{\beta}]\in\mathcal{I}} \Delta W_{\boldsymbol{\alpha},\boldsymbol{\beta}}, \tag{23}$$

$$|\mathbb{E}[F] - \mathcal{M}_{\mathcal{I}}[F]| = \left|\sum_{[\boldsymbol{\alpha},\boldsymbol{\beta}]\notin\mathcal{I}} \boldsymbol{\Delta}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]\right| \leq \sum_{[\boldsymbol{\alpha},\boldsymbol{\beta}]\notin\mathcal{I}} |\boldsymbol{\Delta}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]| = \sum_{[\boldsymbol{\alpha},\boldsymbol{\beta}]\notin\mathcal{I}} \Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}}. \tag{24}$$

Although it would be tempting to define $\mathcal{I}$ as the set of couples $[\boldsymbol{\alpha},\boldsymbol{\beta}]$ with the largest error contribution, this choice could be far from optimal in terms of computational cost; as suggested in the literature on the knapsack problem (see [28]), the benefit-over-cost ratio should rather be taken into account in the decision (see also [3, 7, 22, 31]). More precisely, we propose to build the MISC estimator by first assessing the so-called "profit" of each operator $\boldsymbol{\Delta}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]$, i.e., the quantity

$$P_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \frac{\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}}}{\Delta W_{\boldsymbol{\alpha},\boldsymbol{\beta}}};$$

and then we build an index set for the MISC estimator as

$$\mathcal{I} = \mathcal{I}(\epsilon) = \left\{[\boldsymbol{\alpha},\boldsymbol{\beta}] \in \mathbb{N}_+^d \otimes \mathfrak{L}_+ \;:\; P_{\boldsymbol{\alpha},\boldsymbol{\beta}} \geq \epsilon\right\}, \tag{25}$$

for a suitable $\epsilon > 0$. We observe that the set thus obtained is not necessarily downward-closed, and we have to enforce this condition a posteriori. Obviously, $\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ and $\Delta W_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ are not in general at our disposal. In practice, we base the construction of the MISC estimator on a-priori bounds for such quantities. More precisely, we derive a-priori ansatzes for these bounds from theoretical considerations and then fit the constants appearing in the ansatzes with some auxiliary computations. We could actually refer to the entire strategy as a-priori/posteriori.

Remark 2. *We remark that the general form of the MISC estimator* (20) *is quite broad and includes other related methods (i.e., methods that combine different spatial and stochastic discretization levels to optimize the computational effort) available in the literature, such as the "Multi Level Stochastic Collocation" [35, 37] and "Sparse Composite Collocation Method" [5]; see also [24]. The main novelty of the MISC estimator* (20)-(25) *with respect to such methods is the profit-oriented selection of difference operators. Also novel is the fact that difference operators are introduced in both the spatial and stochastic domains. See [22] for more details on the comparison between the above-mentioned methods and MISC.*

4 Error Analysis of the MISC method

In this section, we state and prove a convergence theorem for the profit-based MISC estimator based on the multi-index set (25). The theorem is based on a result from our previous work [31], which was proved in the context of sparse-grid approximation of Hilbert-space-valued functions. Since the sparse grid and the MISC constructions are identical, this theorem can be used verbatim here. In particular, it links the summability of the profits to the convergence rate of methods such as MISC and Sparse Grids Stochastic Collocation, i.e., based on a sum of difference operators.

To use this result, we have to assess the summability properties of the profits; therefore, in the following we introduce suitable estimates of the error and work contributions, $\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ and $\Delta W_{\boldsymbol{\alpha},\boldsymbol{\beta}}$, respectively. In particular, the estimate of $\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ depends on the spatial regularity of the solution, on the convergence rate of the Finite Element Method used to solve the deterministic problems, and on the summability property of the Chebyshev expansion of the solution over the parameter space.

Theorem 4 (Convergence of the profit-based MISC estimator, see [31])**.** *If the profits $P_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ satisfy the weighted summability condition*

$$\left(\sum_{[\boldsymbol{\alpha},\boldsymbol{\beta}] \in \mathbb{N}_+^d \otimes \mathfrak{L}_+} P_{\boldsymbol{\alpha},\boldsymbol{\beta}}^p \Delta W_{\boldsymbol{\alpha},\boldsymbol{\beta}} \right)^{1/p} = C_P(p) < \infty$$

for some $0 < p \leq 1$, then

$$\left| \mathbb{E}[F] - \mathcal{M}_{\mathcal{I}}[F] \right| \leq \mathrm{Work}[\mathcal{M}_{\mathcal{I}}]^{1-1/p} C_P(p),$$

where $\mathrm{Work}[\mathcal{M}_{\mathcal{I}}]$ *is given by* (23)*.*

We begin by introducing an estimate for the size of the work contribution, $\Delta W_{\boldsymbol{\alpha},\boldsymbol{\beta}}$. To this end, let $\Delta W_{\boldsymbol{\alpha}}^{\mathrm{det}} = \mathrm{Work}\left[\boldsymbol{\Delta}^{\mathrm{det}}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]\right]$, i.e., let it be the cost of computing $\boldsymbol{\Delta}^{\mathrm{det}}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]$ according to equation (17).

Assumption A4 (Spatial work contribution)**.** *Assume that there exist $\gamma \geq 1$ and $C_W > 0$ such that*

$$\Delta W_{\boldsymbol{\alpha}}^{\mathrm{det}} \leq C_W 2^{\gamma \sum_{j=1}^d \alpha_j}. \tag{26}$$

Lemma 5 (Stochastic work contribution)**.** *When using Clenshaw-Curtis points for the discretization over the parameter space, the work contribution $\Delta W_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ of each difference operator $\boldsymbol{\Delta}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]$ can be decomposed as*

$$\Delta W_{\boldsymbol{\alpha},\boldsymbol{\beta}} \leq C_W 2^{\gamma \sum_{j=1}^d \alpha_j + \sum_{j \geq 1} (\beta_j - 1)},$$

with γ and C_W as in Assumption A4.

Proof. Combining equations (21) and (19), we have

$$\Delta W_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \mathrm{Work}\left[\boldsymbol{\Delta}^{\mathrm{stoc}}[\boldsymbol{\Delta}^{\mathrm{det}}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]]\right] = \mathrm{Work}\left[\boldsymbol{\Delta}^{\mathrm{stoc}}[\boldsymbol{\Delta}^{\mathrm{det}}[\mathcal{Q}^{m(\boldsymbol{\beta})}[F^{\boldsymbol{\alpha}}(\cdot)]]]\right].$$

Since the Clenshaw-Curtis points are nested, computing $\Delta W_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ (i.e., adding $[\boldsymbol{\alpha},\boldsymbol{\beta}]$ to the set $\mathcal{I}$ that defines the current MISC estimator) amounts then to evaluating $F^{\boldsymbol{\alpha}}(\boldsymbol{y})$ in the set of "new" points added to the estimator by $\boldsymbol{\Delta}^{\mathrm{stoc}}[\cdot]$, i.e., $\times_{n:\beta_n>1} \left\{ \mathcal{H}^{m(\beta_n)} \setminus \mathcal{H}^{m(\beta_n - 1)} \right\}$, whose cardinality is $\prod_{j \geq 1}(m(\beta_j) - m(\beta_j - 1))$. The proof is then concluded by observing that the definition of $m(\beta)$ in equation (16) immediately gives $m(\beta_j) - m(\beta_j - 1) \leq 2^{\beta_j - 1}$ and recalling Assumption A4. □

Remark 3. *We observe that the exponent $\beta_j - 1$ guarantees that the directions along which no quadrature is actually performed (i.e., $\beta_j = 1$) do not contribute to the total work.*

Next, we prove a sequence of lemmas that allow us to conclude that an analogous estimate holds for the error contribution as well, i.e., that $\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ can be bounded as a product of a term related to the spatial discretization and a term related to the approximation over the parameter space. To this end, we need to introduce the quantity

$$\mathrm{Leb}_{m(\beta)} = \sup_{f \in C^0(\Gamma), \|f\|_{L^\infty(\Gamma)} = 1} \left| Q^{m(\beta)}[f] - Q^{m(\beta-1)}[f] \right| \quad \forall \beta \in \mathbb{N}_+,$$

where $Q^{m(\beta)}[\cdot]$ are the univariate quadrature operators introduced in Equation (14). We observe that $\mathrm{Leb}_1 = 1$ and $\mathrm{Leb}_{m(\beta)} \leq 2$ for $\beta \geq 2$, and that for nested points we can also bound $\mathrm{Leb}_{m(\beta)} \leq \widetilde{\mathrm{Leb}}_{m(\beta)}$, with

$$\widetilde{\mathrm{Leb}}_{m(\beta)} = \sum_{y_\beta^j \in \mathcal{H}^{m(\beta)} \cap \mathcal{H}^{m(\beta-1)}} \left| \varpi_\beta^j - \varpi_{\beta-1}^j \right| + \sum_{y_j \in \mathcal{H}^{m(\beta)} \setminus \mathcal{H}^{m(\beta-1)}} \left| \varpi_\beta^j \right|.$$

In particular, for the Clenshaw-Curtis points it can be verified numerically that $\widetilde{\mathrm{Leb}}_{m(\beta)}$ converges to 1 for $\beta \to \infty$, and that the maximum value is attained at $\beta = 3$, i.e., for $m(3) = 5$ points, with value $\widetilde{\mathrm{Leb}}_5 = \frac{16}{15} \approx 1.067$, cf. Figure 1.

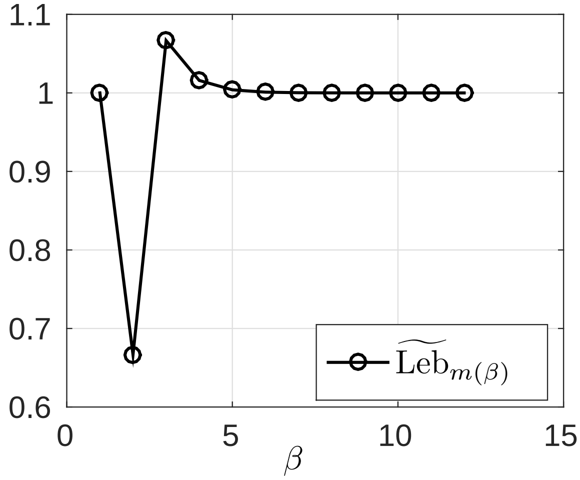


Figure 1: Numerical evaluation of $\widetilde{\mathrm{Leb}}_{m(i_n)}$ for Clenshaw-Curtis points.

Lemma 6 (Stochastic error contribution)**.** *Let* $f : \Gamma = [-1,1]^{\mathbb{N}_+} \to \mathbb{R}$ *and* $\boldsymbol{\beta} \in \mathfrak{L}_+$, *and assume that the quadrature operator,* $\mathcal{Q}^{m(\boldsymbol{\beta})}$, *is built with Clenshaw-Curtis abscissae. If there exists a sequence* $\boldsymbol{r} = \{r_n\}_{n\in\mathbb{N}_+}$ *such that*

1. $\log r_n > \log 2 + \frac{1}{3}\log\frac{16}{15}$ *for all* $n \in \mathbb{N}_+$,
2. $\sum_{n\geq 1} \frac{1}{r_n} < \infty$,
3. $f : \Gamma \to \mathbb{R}$ *admits an analytic complex extension in a Bernstein polyellipse,* $\mathcal{E}_{\mathcal{G},\boldsymbol{r}}$, *for any finite set of random variables,* $\mathcal{G} = \{j_1, \ldots, j_G\}$, *with* $\sup_{\boldsymbol{z}\in\mathcal{E}_{\mathcal{G},\boldsymbol{r}}} |f| \leq C_f$,

then

$$\left|\boldsymbol{\Delta}^{\mathrm{stoc}}[\mathcal{Q}^{m(\boldsymbol{\beta})} f]\right| \leq C_g \sup_{\boldsymbol{z}\in\mathcal{E}_{\mathcal{G},\boldsymbol{r}}} |f(\boldsymbol{z})| e^{-\sum_{n\geq 1} g_n m(\beta_n - 1)},$$

holds, where $g_n = \log r_n - \log 2 - \frac{1}{3}\log\frac{16}{15} > 0$ *and* $C_g = \prod_{n\geq 1} \frac{1}{1-e^{-(\log r_n - \log 2)}} < \infty$.

Proof. Let $\mathcal{G}$ be the support of $\boldsymbol{\beta} - \boldsymbol{1}$ with cardinality $\#\mathcal{G} = G$, $\boldsymbol{q} \in \mathbb{N}^G$, and let $\Phi_{\mathcal{G},\boldsymbol{q}}$ denote the Chebyshev polynomials of the first kind with the degrees $q_1, \ldots, q_G$ along the random variables, y_j, for $j \in \mathcal{G}$ and degree 0 along y_j for $j \notin \mathcal{G}$. For notational convenience, we index the components of $\boldsymbol{q}$ with $j \in \mathcal{G}$, i.e., $q_{j_1} = q_1, q_{j_2} = q_2$, etc. We observe that $\Phi_{\mathcal{G},\boldsymbol{q}}$ are equivalent to the G-variate Chebyshev polynomials over $[-1,1]^G$ thanks to the product structure of the multivariate Chebyshev polynomials and to the fact that $\phi_0(t) = 1$. Next, consider the analytic extension of $f : \mathbb{C}^{\mathcal{G}} \to \mathbb{C}$, and its Chebyshev expansion over $\Phi_{\mathcal{G},\boldsymbol{q}}$ introduced in Lemma 3, there

$$|\boldsymbol{\Delta}^{\mathrm{stoc}}[\mathcal{Q}^{m(\boldsymbol{\beta})} f]| = \left|\boldsymbol{\Delta}^{\mathrm{stoc}}\left[\mathcal{Q}^{m(\boldsymbol{\beta})}\Big[\sum_{\boldsymbol{q}\in\mathbb{N}^G} f_{\boldsymbol{q}}\Phi_{\mathcal{G},\boldsymbol{q}}\Big]\right]\right| = \left|\sum_{\boldsymbol{q}\in\mathbb{N}^G} f_{\boldsymbol{q}}\boldsymbol{\Delta}^{\mathrm{stoc}}\left[\mathcal{Q}^{m(\boldsymbol{\beta})}\left[\Phi_{\mathcal{G},\boldsymbol{q}}\right]\right]\right|,$$

holds. We observe now that by construction of hierarchical surplus there holds $\boldsymbol{\Delta}^{\mathrm{stoc}}[\mathcal{Q}^{m(\boldsymbol{\beta})}[\Phi_{\mathcal{G},\boldsymbol{q}}]] = 0$ for all Chebyshev polynomials $\Phi_{\mathcal{G},\boldsymbol{q}}$ such that $\exists\, j \in \mathcal{G} : q_j < m(\beta_j - 1)$ (i.e., for polynomials that are integrated exactly at least in one direction by both quadrature operators along that direction). Therefore, the previous sum reduces to the multi-index set $\boldsymbol{q} \geq m(\boldsymbol{\beta} - \boldsymbol{1})$, and furthermore by triangular inequality we have

$$|\boldsymbol{\Delta}^{\mathrm{stoc}}[\mathcal{Q}^{m(\boldsymbol{\beta})} f]| \leq \sum_{\boldsymbol{q}\geq m(\boldsymbol{\beta}-\boldsymbol{1})} |f_{\boldsymbol{q}}| \left|\boldsymbol{\Delta}^{\mathrm{stoc}}\left[\mathcal{Q}^{m(\boldsymbol{\beta})}\left[\Phi_{\mathcal{G},\boldsymbol{q}}\right]\right]\right|.$$

Next, using the definitions of $\mathbf{\Delta}^{\mathrm{stoc}}$ and of $\widetilde{\mathrm{Leb}}$, and recalling that Chebyshev polynomials of the first kind on $[-1,1]$ are bounded by 1 and that $\widetilde{\mathrm{Leb}}_{m(\beta)} \leq 1$ for $\beta = 1,2$ and $\widetilde{\mathrm{Leb}}_{m(\beta)} \leq \frac{16}{15}$ for $\beta \geq 3$, we have

$$\begin{aligned}\left|\mathbf{\Delta}^{\mathrm{stoc}}\left[\mathcal{Q}^{m(\boldsymbol{\beta})}\left[\Phi_{\mathcal{G},\boldsymbol{q}}\right]\right]\right| &= \left|\bigotimes_{j\in\mathcal{G}} \Delta[Q^{m(\beta_j)}[\phi_{q_j}]]\right| \qquad (27)\\ &\leq \prod_{j\in\mathcal{G}} \widetilde{\mathrm{Leb}}_{m(\beta_j)} \left\|\phi_{q_j}\right\|_{L^\infty([-1,1])} = \prod_{j\in\mathcal{G}} \widetilde{\mathrm{Leb}}_{m(\beta_j)} \leq \prod_{\substack{j\in\mathcal{G}\\ \beta_j\geq 3}} \frac{16}{15}.\end{aligned}$$

We then bound $|f_{\boldsymbol{q}}|$ by Lemma 3 denoting $\tilde{g}_j = \log r_j - \log 2$ for $j \in \mathcal{G}$. We obtain

$$\begin{aligned}|\mathbf{\Delta}^{\mathrm{stoc}}[\mathcal{Q}^{m(\boldsymbol{\beta})} f]| &\leq \sup_{\boldsymbol{z}\in\mathcal{E}_{\mathcal{G},\boldsymbol{r}}} |f(\boldsymbol{z})| \Bigg(\prod_{\substack{j\in\mathcal{G}\\ \beta_j\geq 3}} \frac{16}{15}\Bigg) \sum_{\boldsymbol{q}\geq m(\boldsymbol{\beta}-\boldsymbol{1})} \prod_{j\in\mathcal{G}} e^{-\tilde{g}_j q_j}\\ &= \sup_{\boldsymbol{z}\in\mathcal{E}_{\mathcal{G},\boldsymbol{r}}} |f(\boldsymbol{z})| \Bigg(\prod_{\substack{j\in\mathcal{G}\\ \beta_j\geq 3}} \frac{16}{15}\Bigg) \prod_{j\in\mathcal{G}} \Bigg(\sum_{q_j\geq m(\beta_j-1)} e^{-\tilde{g}_j q_j}\Bigg)\\ &= \sup_{\boldsymbol{z}\in\mathcal{E}_{\mathcal{G},\boldsymbol{r}}} |f(\boldsymbol{z})| \Bigg(\prod_{\substack{j\in\mathcal{G}\\ \beta_j\geq 3}} \frac{16}{15}\Bigg) \prod_{j\in\mathcal{G}} \frac{e^{-\tilde{g}_j m(\beta_j-1)}}{1-e^{-\tilde{g}_j}}.\end{aligned}$$

Next, we observe that, for $\beta \geq 3$, there holds $\frac{16}{15} = e^{\log\frac{16}{15}} \leq e^{\log\frac{16}{15}\frac{m(\beta-1)}{m(2)}} = e^{C_L m(\beta-1)}$with $C_L = \frac{1}{3}\log\frac{16}{15} \approx 0.0215$. Therefore, we have

$$\begin{aligned}|\mathbf{\Delta}^{\mathrm{stoc}}[\mathcal{Q}^{m(\boldsymbol{\beta})} f]| &\leq \sup_{\boldsymbol{z}\in\mathcal{E}_{\mathcal{G},\boldsymbol{r}}} |f(\boldsymbol{z})| \prod_{j\in\mathcal{G}} \frac{1}{1-e^{-\tilde{g}_j}} \prod_{\substack{j\in\mathcal{G}\\ \beta_j\geq 3}} e^{-(\tilde{g}_j - C_L)m(\beta_j-1)} \prod_{\substack{j\in\mathcal{G}\\ \beta_j\leq 2}} e^{-\tilde{g}_j m(\beta_j-1)}\\ &\leq \sup_{\boldsymbol{z}\in\mathcal{E}_{\mathcal{G},\boldsymbol{r}}} |f(\boldsymbol{z})| C_g \prod_{j\geq 1} e^{-(\tilde{g}_j - C_L)m(\beta_j-1)},\end{aligned}$$

where in the last step we have extended the product from $j \in \mathcal{G}$ to $j \geq 1$, defined $C_g = \prod_{j\geq 1}\frac{1}{1-e^{-\tilde{g}_j}}$, and introduced the correction term C_L along all random variables. We observe that after these operations the bound is independent of $\mathcal{G}$ and yet finite, since $\beta_j = 1$ hold for $j \notin \mathcal{G}$, which implies $m(0) = 0$ and moreover

$$C_g = \prod_{j\geq 1}\frac{1}{1-e^{-\tilde{g}_j}} < \infty \Leftrightarrow -\sum_{j\geq 1}\log(1-e^{-\tilde{g}_j}) < \infty \Leftrightarrow \sum_{j\geq 1} e^{-\tilde{g}_j} < \infty \Leftrightarrow \sum_{j\geq 1}\frac{2}{r_j} < \infty,$$

which is true by hypothesis. □

Remark 4. *Sharper estimates could be obtained by exploiting the structure of the Chebyshev polynomials when computing* $\Delta[Q^{m(\beta_j)}[\phi_{q_j}]]$ *in equation* (27) *(for instance, the fact that* $Q^{m(\beta_j)}[\phi_{q_j}] = 0$ *whenever* q_j *is odd and larger than 1) rather than using the general bound* $\Delta[Q^{m(\beta_j)}[\phi_{q_j}]] \leq \widetilde{\mathrm{Leb}}_{m(\beta_j)} \left\|\phi_{q_j}\right\|_{L^\infty([-1,1])}$.

Next, we state an assumption of convenience, which is not necessary but considerably simplifies some technical passages in the development of the theory, as we clarify later on.

Assumption A5 (Lower bounds for $\zeta_{0,n}$ and $\zeta_{s,n}$)**.** *We assume that* $\log\zeta_{s,n} - \log 2 - \frac{1}{3}\log\frac{16}{15} > 0$ *for every* $n \in \mathbb{N}_+$*, and that* $\log\zeta_{0,1} - \log 2 - \frac{1}{3}\log\frac{16}{15} > 0$*, where* $\zeta_{0,1}, \zeta_{s,n}$ *are the parameters specifying the size of the Bernstein ellypses in Lemmas 1 and 2.*

We observe that from the previous assumption, we also have that $\log\zeta_{0,n} - \log 2 - \frac{1}{3}\log\frac{16}{15} > 0$ for every $n \in \mathbb{N}_+$. We are now almost in the position to prove the estimate on the error contribution (see Lemma 8); before doing this, we need another auxiliary lemma that gives conditions for the summability of certain sequences that will be considered in the proof of Lemma 8 as well as in the proof of the main theorem on the convergence of MISC.

Lemma 7 (Summability of stochastic rates)**.** *Let*

- $g_{0,n} = \log \zeta_{0,n} - \log 2 - M$,
- $g_{s,n} = \log \zeta_{s,n} - \log 2 - M_s$,

with $\zeta_{0,n}$ as in Lemma 1 and $\zeta_{s,n}$ as in Lemma 2, and assume that M and M_s are positive numbers such that $g_{0,n}, g_{s,n} > 0$. Then,

- *the sequences $\{e^{-g_{0,n}}\}_{n\in\mathbb{N}_+}$ and $\left\{\frac{1}{\zeta_{0,n}}\right\}_{n\in\mathbb{N}_+}$ are ℓ^q-summable for $q > q_{0,min} = \frac{p_0}{1-p_0}$,*
- *the sequences $\{e^{-g_{s,n}}\}_{n\in\mathbb{N}_+}$ and $\left\{\frac{1}{\zeta_{s,n}}\right\}_{n\in\mathbb{N}_+}$ are ℓ^q-summable for $q > q_{s,min} = \frac{p_s}{1-p_s}$.*

Under Assumption A2, $q_{0,min}, q_{s,min} < 1$.

Proof. We have $1 > e^{-g_{0,n}} = e^{-(\log\zeta_{0,n} - \log 2 - M)} = e^M \left(\frac{\zeta_{0,n}}{2}\right)^{-1}$; from (12) and we can bound $2\tau_{0,n} \leq \zeta_{0,n}$. We can therefore obtain

$$\sum_{n\geq 1} e^{-g_{0,n}q} = e^{Mq} \sum_{n\geq 1} \left(\frac{\zeta_{0,n}}{2}\right)^{-q} \leq e^{Mq} \sum_{n\geq 1} \tau_{0,n}^{-q} = e^{Mq} \left(\frac{\pi}{E\,\|\boldsymbol{b}_0\|_{\ell^{p_0}}^{p_0}}\right)^{-q} \sum_{n\geq 1} b_{0,n}^{(1-p_0)q}$$

From Assumption A2, we know that $\boldsymbol{b}_0 \in \ell^{p_0}$. We therefore have the condition

$$(1-p_0)q \geq p_0 \Rightarrow q_{0,min} = \frac{p_0}{1-p_0},$$

and enforcing $q_{0,min} < 1$ results in $p_0 < \frac{1}{2}$, as stated in Assumption A5. The same argument applies to $\{e^{g_{s,n}}\}_{n\in\mathbb{N}_+}$. □

Lemma 8 (Total error contribution)**.** *Assume the deterministic problem is solved with piecewise linear finite elements with spatial discretization parameters $h_i = 2^{-\alpha_i}$. Then the error contribution $\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ of each difference operator $\boldsymbol{\Delta}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]$ can be decomposed as*

$$\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \min\left\{C_0 \Delta E_{\boldsymbol{\beta}}^{\text{stoc},0}, C_1 \Delta E_{\boldsymbol{\alpha}}^{\text{det},1} \Delta E_{\boldsymbol{\beta}}^{\text{stoc},1}\right\}, \tag{28}$$

where the factors can be bounded as

$$C_0 = C_{CT} \|\Theta\|_{H^{-1}(\mathcal{B})} C_u \prod_{n\geq 1} \frac{1}{1 - e^{-(\log\zeta_{0,n} - \log 2)}} < \infty$$

$$C_1 = C_{CT} \|\Theta\|_{H^{-1}(\mathcal{B})} C_{s,u} \prod_{n\geq 1} \frac{1}{1 - e^{-(\log\zeta_{s,n} - \log 2)}} < \infty$$

$$\Delta E_{\boldsymbol{\beta}}^{\text{stoc},0} \leq e^{-\sum_{n\geq 1} m(\beta_n - 1) g_{0,n}} \tag{29}$$

$$\Delta E_{\boldsymbol{\alpha}}^{\text{det},1} \leq 2^{-|\boldsymbol{\alpha}| r_{\text{FEM}}} \tag{30}$$

$$\Delta E_{\boldsymbol{\beta}}^{\text{stoc},1} \leq e^{-\sum_{n\geq 1} m(\beta_n - 1) g_{s,n}} \tag{31}$$

with C_u as in Lemma 1, $C_{s,u}$ as in Lemma 2, $g_n = \log\zeta_n - \log 2 - \frac{1}{3}\log\frac{16}{15}$, $g_{s,n} = \log\zeta_{s,n} - \log 2 - \frac{1}{3}\log\frac{16}{15}$, $r_{\text{FEM}} = \min\{1, s/d\}$ with s as in Assumption A3 and $C_{CT} > 0$.

Proof. Combining the definition of $\boldsymbol{\Delta}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]$, cf. equation (19), and the definition of $\Delta^{\text{det}}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]$, cf. equation (17), we have

$$\begin{aligned}\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}} = |\boldsymbol{\Delta}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]| = \boldsymbol{\Delta}^{\text{stoc}}[\boldsymbol{\Delta}^{\text{det}}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]] &= \boldsymbol{\Delta}^{\text{stoc}}[\sum_{\boldsymbol{j}\in\{0,1\}^d} (-1)^{|\boldsymbol{j}|} F_{\boldsymbol{\alpha}-\boldsymbol{j},\boldsymbol{\beta}}] \\ &= \boldsymbol{\Delta}^{\text{stoc}}[\sum_{\boldsymbol{j}\in\{0,1\}^d} (-1)^{|\boldsymbol{j}|} \mathcal{Q}^{m(\boldsymbol{\beta})}[\Theta[u^{\boldsymbol{\alpha}-\boldsymbol{j}}(\cdot,\boldsymbol{y})]] \\ &= \boldsymbol{\Delta}^{\text{stoc}}[\mathcal{Q}^{m(\boldsymbol{\beta})}\Theta[\sum_{\boldsymbol{j}\in\{0,1\}^d} (-1)^{|\boldsymbol{j}|} u^{\boldsymbol{\alpha}-\boldsymbol{j}}(\cdot,\boldsymbol{y})]] \\ &= \boldsymbol{\Delta}^{\text{stoc}}[\mathcal{Q}^{m(\boldsymbol{\beta})}\Theta[\boldsymbol{\Delta}^{\text{det}}[u^{\boldsymbol{\alpha}}(\cdot,\boldsymbol{y})]]]\end{aligned}$$

We observe that $f(\boldsymbol{y}) = \Theta[\boldsymbol{\Delta}^{\det}[u^{\boldsymbol{\alpha}}(\cdot,\boldsymbol{y})]]$ is a linear combination of some $u^{\boldsymbol{\alpha}}$ and that each of these $u^{\boldsymbol{\alpha}}$ is analytic, being the finite-element approximation of the analytic function u; hence, $f(\boldsymbol{y})$ is also analytic. Then, there exists an ellipse $\mathcal{E}_{\mathcal{G},\boldsymbol{r}}$ (with size specified by a suitable sequence $\boldsymbol{r}$ that we will choose later) such that we can apply Lemma 6 and obtain

$$\begin{aligned}\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}} &\leq \sup_{\boldsymbol{z}\in\mathcal{E}_{\mathcal{G},\boldsymbol{r}}} |\Theta[\boldsymbol{\Delta}^{\det}[u^{\boldsymbol{\alpha}}(\boldsymbol{x},\boldsymbol{z})]| \prod_{n\geq 1} \frac{1}{1-e^{-(\log r_n - \log 2)}} e^{-\sum_{n\geq 1} g_n m(\beta_n-1)} \\ &\leq \|\Theta\|_{H^{-1}(\mathcal{B})} \sup_{\boldsymbol{z}\in\mathcal{E}_{\mathcal{G},\boldsymbol{r}}} \left\|\boldsymbol{\Delta}^{\det}[u^{\boldsymbol{\alpha}}(\cdot,\boldsymbol{z})]\right\|_{H^1(\mathcal{B},\mathbb{C})} \prod_{n\geq 1} \frac{1}{1-e^{-(\log r_n - \log 2)}} e^{-\sum_{n\geq 1} g_n m(\beta_n-1)},\end{aligned}$$

where the rates g_n depend on the parameters r_n, $g_n = \log r_n - \log 2 - \frac{1}{3}\log\frac{16}{15}$. Next, assuming that the spatial discretization consists of piecewise linear finite elements with spatial discretization parameters $h_i = 2^{-\alpha_i}$, and combining (i) the a-priori bounds on the decay of the difference operators coming from the Combination Technique theory (see, e.g., [19]), (ii) the fact that $u \in H^{1+s}$ for some $s > 0$ (cf. Assumption A3), and (iii) equation (8), we have two valid bounds for every $\boldsymbol{z}$ in the ellipse, $\mathcal{E}_{\mathcal{G},\boldsymbol{r}}$, (which we recall is yet to be specified):

$$\begin{aligned}&\left\|\boldsymbol{\Delta}^{\det}[u^{\boldsymbol{\alpha}}(\cdot,\boldsymbol{z})]\right\|_{H^1(\mathcal{B})} \leq C_{CT} \|u(\cdot,\boldsymbol{z})\|_{H^1(\mathcal{B},\mathbb{C})}, \\ &\left\|\boldsymbol{\Delta}^{\det}[u^{\boldsymbol{\alpha}}(\cdot,\boldsymbol{z})]\right\|_{H^1(\mathcal{B})} \leq C_{CT} \|u(\cdot,\boldsymbol{z})\|_{\mathcal{H}^{1+s/d}(\mathcal{B},\mathbb{C})} 2^{-|\boldsymbol{\alpha}| r_{\mathrm{FEM}}} \leq C_{CT} \|u(\cdot,\boldsymbol{z})\|_{H^{1+s}(\mathcal{B},\mathbb{C})} 2^{-|\boldsymbol{\alpha}| r_{\mathrm{FEM}}},\end{aligned}$$

for some $C_{CT} > 0$. We then conclude using Lemmas 1 and 2, which guarantee the boundedness of $\|u(\cdot,\boldsymbol{z})\|_{H^1(\mathcal{B},\mathbb{C})}$ and $\|u(\cdot,\boldsymbol{z})\|_{H^{1+s}(\mathcal{B},\mathbb{C})}$ in the Bernstein ellypses with parameters $\boldsymbol{r} = \boldsymbol{\zeta}_0$ and $\boldsymbol{r} = \boldsymbol{\zeta}_s$, respectively. We observe that the hypotheses of Lemma 6 are verified by $\boldsymbol{\zeta}_0$ and $\boldsymbol{\zeta}_s$, thanks to Assumptions A2, A5 and Lemma 7. We then have the following two valid bounds:

$$\begin{aligned}\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}} &\leq \|\Theta\|_{H^{-1}(\mathcal{B})} \sup_{\boldsymbol{z}\in\mathcal{E}_{\mathcal{G},\boldsymbol{\zeta}_0}} \|u(\cdot,\boldsymbol{z})\|_{H^1(\mathcal{B},\mathbb{C})} \prod_{n\geq 1} \frac{1}{1-e^{-(\log \zeta_{0,n} - \log 2)}} e^{-\sum_n g_{0,n} m(\beta_n-1)}, \\ \Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}} &\leq \|\Theta\|_{H^{-1}(\mathcal{B})} \sup_{\boldsymbol{z}\in\mathcal{E}_{\mathcal{G},\boldsymbol{\zeta}_s}} \|u(\cdot,\boldsymbol{z})\|_{H^{1+s}(\mathcal{B},\mathbb{C})} \prod_{n\geq 1} \frac{1}{1-e^{-(\log \zeta_{s,n} - \log 2)}} e^{-\sum_n g_{s,n} m(\beta_n-1)} 2^{-|\boldsymbol{\alpha}| r_{\mathrm{FEM}}},\end{aligned}$$

where the quantities $\prod_{n\geq 1} \frac{1}{1-e^{-(\log \zeta_{0,n} - \log 2)}}$ and $\prod_{n\geq 1} \frac{1}{1-e^{-(\log \zeta_{0,n} - \log 2)}}$ are bounded, thanks again to Lemma 6. The proof is then concluded by recalling that $\sup_{\boldsymbol{z}\in\mathcal{E}_{\mathcal{G},\boldsymbol{\zeta}_0}} \|u(\cdot,\boldsymbol{z})\|_{H^1(\mathcal{B},\mathbb{C})} \leq C_u$ due to Lemma 1, and similarly $\sup_{\boldsymbol{z}\in\mathcal{E}_{\mathcal{G},\boldsymbol{\zeta}_s}} \|u(\cdot,\boldsymbol{z})\|_{H^{1+s}(\mathcal{B},\mathbb{C})} \leq C_{s,u}$ due to Lemma 2, with C_u and $C_{s,u}$ independent of $\mathcal{G}$. □

Lemma 9 (Total error contribution, restated)**.** *The bound for $\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ in Lemma 8 can be rewritten as*

$$\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}} \leq C_E 2^{-\sum_{j\geq 1} m(\beta_j-1)\chi_j - \max\left\{\sum_{j\geq 1} m(\beta_j-1)\theta_j,\, r_{\mathrm{FEM}}|\boldsymbol{\alpha}|\right\}},$$

with $C_E = \min\{C_0, C_1\}$, $\chi_j = g_{s,j}\log_2 e$, and $\theta_j = (g_{0,j} - g_{s,j})\log_2 e$.

Proof. The statement is a compact rewriting of the following expression, which is obtained by combining equations (28) to (31):

$$\begin{aligned}\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}} &\leq C_E \min_{q\in\{0,1\}} 2^{-q r_{\mathrm{FEM}}|\boldsymbol{\alpha}|} \prod_{j\geq 1} e^{-m(\beta_j-1)[g_{s,j}+(1-q)(g_{0,j}-g_{s,j})]} \\ &= C_E \min_{q\in\{0,1\}} 2^{-q r_{\mathrm{FEM}}|\boldsymbol{\alpha}|} \prod_{j\geq 1} 2^{-m(\beta_j-1)[g_{s,j}+(1-q)(g_{0,j}-g_{s,j})]\log_2 e} \\ &= C_E \min_{q\in\{0,1\}} 2^{-q r_{\mathrm{FEM}}|\boldsymbol{\alpha}|} \prod_{j\geq 1} 2^{-m(\beta_j-1)[\chi_j+(1-q)\theta_j]} \\ &= C_E 2^{-\sum_{j\geq 1} m(\beta_j-1)\chi_j - \max_{q\in\{0,1\}}\left(q r_{\mathrm{FEM}}|\boldsymbol{\alpha}| + \sum_{j\geq 1} m(\beta_j-1)(1-q)\theta_j\right)} \\ &= C_E 2^{-\sum_{j\geq 1} m(\beta_j-1)\chi_j - \max\left\{\sum_{j\geq 1} m(\beta_j-1)\theta_j,\, r_{\mathrm{FEM}}|\boldsymbol{\alpha}|\right\}}.\end{aligned}$$

□

Remark 5 (Relaxing the simplifying assumptions). *We now clarify why the assumptions $p_s < \frac{1}{2}$ and $\zeta_{s,n}, \zeta_{0,n} > \log 2 + \frac{1}{3}\log\frac{16}{15}$ are not essential and could be relaxed, at the expense of a more involved presentation. First, assuming $\zeta_{s,n}, \zeta_{0,n} > \log 2 + \frac{1}{3}\log\frac{16}{15}$ allowed us to use the simple bound provided by Lemma 3 for the coefficients of the Chebyshev expansion of $f(\boldsymbol{y}) = \Theta[\boldsymbol{\Delta}^{\det}[u^{\boldsymbol{\alpha}}(\boldsymbol{x},\boldsymbol{y})]]$ in Lemma 8. Without this assumption we would need to identify the first random variable, say N^*, such that $\zeta_{0,N^*} > \log 2 + \frac{1}{3}\log\frac{16}{15}$ and $\zeta_{s,N^*} > \log 2 + \frac{1}{3}\log\frac{16}{15}$ and treat separately the variables $y_1, y_2, \dots, y_{N^*-1}$, in the proof of the main theorem.*

Next, due to a suboptimal choice of the parameters $\zeta_{0,n}$ and $\zeta_{s,n}$ in Lemmas 1 and 2, the sequences $\{e^{-g_{0,n}}\}_{n\in\mathbb{N}_+}$ and $\{e^{-g_{s,n}}\}_{n\in\mathbb{N}_+}$ in Lemma 7 which are related to the solution u and which appear in the proof of the MISC convergence theorem, have worse summability than the sequences $\{b_{0,n}\}_{n\in\mathbb{N}_+}$ and $\{b_{s,n}\}_{n\in\mathbb{N}_+}$ related to the diffusion coefficient. As we will see in the main theorem, assuming $p_0, p_s < \frac{1}{2}$ is then needed to guarantee convergence of MISC. It would be possible to have $q_{0,min} = p_0, q_{s,min} = p_s$, with no restriction on $p_s < \frac{1}{2}$ for the method to converge, by choosing the ellipses in Lemmas 1 and 2 by the more elaborated strategy presented in [12]. However, for ease of exposition, we maintain the sub-optimal choice, which is enough for the purpose of presenting the argument that proves convergence of MISC. The restriction $p_0, p_s < \frac{1}{2}$ formally prevents us from applying the MISC convergence analysis to diffusion coefficients with low spatial regularity. In practice, we see in Section 6 that the convergence estimates are numerically valid beyond this restriction.

Before proving the main theorem of this section, we finally need the following technical lemma.

Lemma 10 (Bounding a sum of double exponentials). *For $a > 0$, $b \geq 2$ and $0 \leq c < ab$,*

$$\sum_{k=1}^{\infty} e^{-ab^k + ck} \leq e^{-ab + \varepsilon(a,b,c)},$$

holds, where for each fixed c and b, we have that $\varepsilon(\cdot, b, c)$ is a monotonically decreasing, strictly positive function with $\varepsilon(a,b,c) \to c$ as $a \to +\infty$.

Proof.

$$\begin{aligned}\sum_{k\geq 1} e^{-ab^k+ck} =& e^{-ab+c} + \sum_{k\geq 2} e^{-ab^k+ck} = e^{-ab+c} + \sum_{k\geq 1} e^{-ab^{k+1}+c(k+1)} \\ =& e^{-ab}\left(e^c + e^c \sum_{k\geq 1} e^{-ab(b^k-1)+ck}\right).\end{aligned}$$

We observe that for $b \geq 2$ we have $b^k - 1 \geq k$ for $k \geq 1$ integer. Therefore, $e^{-ab(b^k-1)} \leq e^{-abk}$ and we have

$$\sum_{k\geq 1} e^{-ab^k+ck} \leq e^{-ab}\left(e^c + e^c \sum_{k\geq 1} e^{k(c-ab)}\right) = e^{-ab}\left(e^c + \frac{e^{2c-ab}}{1-e^{c-ab}}\right).$$

Then,

$$\varepsilon(a,b,c) = \log\left(e^c + \frac{e^{2c-ab}}{1-e^{c-ab}}\right),$$

and we finish by verifying that the function ε has the required properties. □

Theorem 11 (MISC convergence theorem). *Under Assumptions A1 to A5, for the profit-based MISC estimator built using the set $\mathcal{I}$ defined in* (25)*, Stochastic Collocation over Clenshaw-Curtis points and piecewise linear finite elements for solving the deterministic problems,*

$$\left|\mathbb{E}[F] - \mathcal{M}_{\mathcal{I}}[F]\right| \leq C_P\left(\frac{1}{1-r_{\mathrm{MISC}}}\right)\mathrm{Work}[\mathcal{M}_{\mathcal{I}}]^{-r_{\mathrm{MISC}}},$$

holds, where $C_P\left(\frac{1}{1-r_{\mathrm{MISC}}}\right)$ is as in Theorem 4, $\mathrm{Work}[\mathcal{M}_{\mathcal{I}}]$ *is given by* (23) *and r_{MISC} is as follows:*

Case 1 *if $\frac{\gamma}{r_{\mathrm{FEM}}+\gamma} \geq \frac{p_s}{1-p_s}$, then $r_{\mathrm{MISC}} < \frac{r_{\mathrm{FEM}}}{\gamma}$,*

Case 2 *if* $\frac{\gamma}{r_{\mathrm{FEM}}+\gamma} \leq \frac{p_s}{1-p_s}$, *then*

$$r_{\mathrm{MISC}} < \left(\frac{1}{p_0} - 2\right)\left(\gamma \frac{p_s - p_0}{r_{\mathrm{FEM}} p_0 p_s} + 1\right)^{-1} .$$

Proof. In view of using Theorem 4, we need to estimate the p-summability of weighted profits. To this end, we use Lemma 9 and bound the sum of the profit as follows:

$$\begin{aligned}
&\sum_{[\boldsymbol{\alpha},\boldsymbol{\beta}]\in\mathbb{N}_+^d\otimes\mathfrak{L}_+} P^p_{\boldsymbol{\alpha},\boldsymbol{\beta}} \Delta W_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \sum_{[\boldsymbol{\alpha},\boldsymbol{\beta}]\in\mathbb{N}_+^d\otimes\mathfrak{L}_+} \Delta E^p_{\boldsymbol{\alpha},\boldsymbol{\beta}} \Delta W^{1-p}_{\boldsymbol{\alpha}} \Delta W^{1-p}_{\boldsymbol{\beta}} \\
&\leq C_E^p C_W^{1-p} \sum_{[\boldsymbol{\alpha},\boldsymbol{\beta}]\in\mathbb{N}_+^d\otimes\mathfrak{L}_+} \left(\frac{1}{2}\right)^{p[\max\{r_{\mathrm{FEM}}|\boldsymbol{\alpha}|, \sum_{j\geq 1} m(\beta_j-1)\theta_j\} + \sum_{j\ \geq 1} m(\beta_j-1)\chi_j] - (1-p)\gamma|\boldsymbol{\alpha}| - (1-p)\sum_j(\beta_j-1)} \\
&\leq C_E^p C_W^{1-p} \min_{\lambda\in[0,1]} \sum_{[\boldsymbol{\alpha},\boldsymbol{\beta}]\in\mathbb{N}_+^d\otimes\mathfrak{L}_+} \left(\frac{1}{2}\right)^{p[\lambda r_{\mathrm{FEM}}|\boldsymbol{\alpha}| + (1-\lambda)\sum_{j\geq 1} m(\beta_j-1)\theta_j + \sum_{j\geq 1} m(\beta_j-1)\chi_j] - (1-p)\gamma|\boldsymbol{\alpha}| - (1-p)\sum_{j\geq 1}(\beta_j-1)} \\
&= C_E^p C_W^{1-p} \min_{\lambda\in[0,1]} \left(\prod_{i=1}^d \sum_{k=1}^\infty \left(\frac{1}{2}\right)^{(p(\lambda r_{\mathrm{FEM}}+\gamma)-\gamma)k}\right) \left(\prod_{j=1}^\infty \sum_{k=1}^\infty \left(\frac{1}{2}\right)^{p((1-\lambda)m(k-1)\theta_j + m(k-1)\chi_j) - (1-p)(k-1)}\right),
\end{aligned} \tag{32}$$

and we investigate under what conditions each of the two factors are finite (the constants C_E, C_W are bounded, cf. Lemmas 5, 8 and 9). For the first term, we immediately have

$$p > \frac{\gamma}{\lambda r_{\mathrm{FEM}} + \gamma} . \tag{33}$$

For the second factor, by denoting the generic term of the inner sum as $a_{j,k}$ for brevity and observing that $a_{j,1} = 1$ for every j, we have

$$\prod_{j=1}^\infty \sum_{k=1}^\infty a_{j,k} \leq \prod_{j=1}^\infty \left(1 + \sum_{k=2}^\infty a_{j,k}\right) = \exp\left(\sum_{j=1}^\infty \log\left(1 + \sum_{k=2}^\infty a_{j,k}\right)\right) \leq \exp\left(\sum_{j=1}^\infty \sum_{k=2}^\infty a_{j,k}\right),$$

and we only have to discuss the convergence of the sum

$$\sum_{j=1}^\infty \sum_{k=2}^\infty 2^{-pm(k-1)[(1-\lambda)\theta_j + \chi_j] + (1-p)(k-1)} \leq \sum_{j=1}^\infty \sum_{k=2}^\infty 2^{-p\frac{2^k}{4}[(1-\lambda)\theta_j+\chi_j] + (1-p)(k-1)}, \tag{34}$$

where the last step is a consequence of the fact that for Clenshaw-Curtis points $m(k-1) \geq \frac{m(k)-1}{2}$ holds for $k \geq 1$ and moreover $\frac{m(k)-1}{2} = 2^{k-2}$ for $k \geq 2$, cf. equation (16). To study the summability of (34), we first use Lemma 10 to bound the inner sum in (34); we have

$$\begin{aligned}
\sum_{k=2}^\infty 2^{-p\frac{2^k}{4}[(1-\lambda)\theta_j+\chi_j] + (1-p)(k-1)} &= \sum_{k=2}^\infty 2^{-p\frac{2^{k-1}}{2}[(1-\lambda)\theta_j+\chi_j] + (1-p)(k-1)} \\
&= \sum_{\tilde{k}=1}^\infty 2^{-p\frac{2^{\tilde{k}}}{2}[(1-\lambda)\theta_j+\chi_j] + (1-p)\tilde{k}} \\
&\leq \sum_{k=1}^\infty \exp\left(-p\frac{\log 2}{2}[(1-\lambda)\theta_j + \chi_j]2^k + (1-p)k\log 2\right) \\
&\leq \sum_{k=1}^\infty \exp\left(-ab^k + ck\right) \\
&\text{with } a = p\frac{\log 2}{2}[(1-\lambda)\theta_j + \chi_j], \quad b = 2, \quad c = (1-p)\log 2,
\end{aligned}$$

where we have used the notation in Lemma 10. Note that this lemma holds true under the assumptions that $0 \leq c < ab$, $a > 0$ and $b \geq 2$, which have to be verified. We have

$$ab > c \Leftrightarrow p \log 2[(1-\lambda)\theta_j + \chi_j] > (1-p)\log 2 \Leftrightarrow (1-\lambda)\theta_j + \chi_j > \frac{(1-p)}{p},$$

which is true for sufficiently large j, say $j \geq j^*$, since $\lambda \leq 1$, χ_j is increasing with $\chi_j \to \infty$ as $j \to \infty$, and θ_j is always positive. The condition $c \geq 0$ is true for $0 < p < 1$, since $\log 2 > 0$. Finally, $a > 0$ is equivalent to $(1-\lambda)\theta_j + \chi_j > 0$, which is again true. Resuming from (34), we have

$$\begin{aligned}\sum_{j=1}^{\infty}\sum_{k=2}^{\infty} 2^{-p\frac{2^k}{4}[(1-\lambda)\theta_j+\chi_j]+(1-p)(k-1)} &\leq C(j^*) + \sum_{j=j^*}^{\infty}\sum_{k=1}^{\infty}\exp\left(-ab^k + ck\right) \\ &\leq C(j^*) + \sum_{j=j^*}^{\infty} e^{-ab+\varepsilon(a,b,c)}.\end{aligned}$$

Next, since according to Lemma 10 $\varepsilon(a,b,c)$ is a monotonically decreasing function with limit $c = (1-p)\log 2$ independent of j, the previous series converges if and only if

$$\sum_{j\geq j^*}^{\infty} e^{-ab} = \sum_{j\geq j^*}^{\infty} e^{-p\log 2[(1-\lambda)\theta_j+\chi_j]} = \sum_{j\geq j^*}^{\infty} 2^{-p[(1-\lambda)\theta_j+\chi_j]}$$

converges. Inserting the expression of θ_j and χ_j, cf. Lemma 9, we get:

$$\begin{aligned}\sum_{j\geq j^*}^{\infty} 2^{-p[(1-\lambda)\theta_j+\chi_j]} &= \sum_{j\geq j^*}^{\infty} 2^{-p[(1-\lambda)(g_{0,j}-g_{s,j})+g_{s,j}]\log_2 e} = \sum_{j\geq j^*}^{\infty} e^{-p[(1-\lambda)(g_{0,j}-g_{s,j})+g_{s,j}]} \\ &= \sum_{j\geq j^*}^{\infty} e^{-p(1-\lambda)g_{0,j}} e^{-p\lambda g_{s,j}}.\end{aligned}$$

After applying Hölder inequality in the previous summation with exponents $q_1^{-1} + q_2^{-1} = 1$, we need to simultaneously ensure the boundedness of the following sums:

$$\sum_{j\geq j^*}^{\infty} e^{-p(1-\lambda)g_{0,j}q_2} \quad \text{and} \quad \sum_{j\geq j^*}^{\infty} e^{-p\lambda g_{s,j}q_1}.$$

Recalling the summability result in Lemma 7, we have that the following two conditions must hold:

$$\begin{cases} p(1-\lambda)q_2 > \dfrac{p_0}{1-p_0} \\ p\lambda q_1 > \dfrac{p_s}{1-p_s} \end{cases} \Rightarrow \begin{cases} p > \dfrac{p_0}{1-p_0}\dfrac{1}{1-\lambda}\dfrac{1}{q_2} \\ p > \dfrac{p_s}{1-p_s}\dfrac{1}{\lambda}\left(1-\dfrac{1}{q_2}\right) \end{cases},$$

which closes the discussion on the summability of the second factor of (32). Recalling the constraint (33) coming from the first factor of (32), we finally have to solve the following optimization problem:

$$p > \min_{\lambda\in[0,1],1\leq q_2} \max\left\{\frac{\gamma}{\lambda r_{\mathrm{FEM}}+\gamma}, \frac{p_s}{1-p_s}\frac{1}{\lambda}\left(1-\frac{1}{q_2}\right), \frac{p_0}{1-p_0}\frac{1}{1-\lambda}\frac{1}{q_2}\right\}$$

i.e., to choose q_2 and λ to minimize the lower bound on p above. We first optimally select q_2 given λ, i.e. take $q_2 = q_2^*$ such that

$$\frac{p_s}{1-p_s}\frac{1}{\lambda}\left(1-\frac{1}{q_2^*}\right) = \frac{p_0}{1-p_0}\frac{1}{1-\lambda}\frac{1}{q_2^*} \;\Rightarrow\; q_2^* = 1 + \frac{1-p_s}{p_s}\frac{p_0}{1-p_0}\frac{\lambda}{1-\lambda}$$

Substituting back, we obtain

$$\begin{aligned}\frac{p_s}{1-p_s}\frac{1}{\lambda}\frac{q_2^*-1}{q_2^*} &= \frac{p_s}{1-p_s}\frac{1}{\lambda}\frac{\frac{1-p_s}{p_s}\frac{p_0}{1-p_0}\frac{\lambda}{1-\lambda}}{1+\frac{1-p_s}{p_s}\frac{p_0}{1-p_0}\frac{\lambda}{1-\lambda}} = \frac{p_0}{1-p_0}\frac{1}{(1-\lambda)+\frac{1-p_s}{p_s}\frac{p_0}{1-p_0}\lambda} \\ &= \frac{\frac{p_0}{1-p_0}\frac{p_s}{1-p_s}}{(1-\lambda)\frac{p_s}{1-p_s}+\frac{p_0}{1-p_0}\lambda} = \frac{\frac{p_0}{1-p_0}\frac{p_s}{1-p_s}}{\frac{p_s}{1-p_s}+\lambda\left(\frac{p_0}{1-p_0}-\frac{p_s}{1-p_s}\right)},\end{aligned}$$

so that the minimization problem reads

$$p > \min_{\lambda\in[0,1]} \max\left\{f_1(\lambda), f_2(\lambda)\right\}, \qquad f_1(\lambda) = \frac{\gamma}{\lambda r_{\mathrm{FEM}} + \gamma}, \quad f_2(\lambda) = \frac{\frac{p_0}{1-p_0}\frac{p_s}{1-p_s}}{\frac{p_s}{1-p_s} + \lambda\left(\frac{p_0}{1-p_0} - \frac{p_s}{1-p_s}\right)}. \tag{35}$$

Now recall that $p_0 \leq p_s$, which implies that $\frac{p_0}{1-p_0} \leq \frac{p_s}{1-p_s}$, hence $f_2(\lambda)$ is increasing with λ; conversely, $f_1(\lambda)$ is instead decreasing with λ, since γ, r_{FEM} are positive numbers. Furthermore, notice that we cannot have $f_1(\lambda) < f_2(\lambda)$ for all $\lambda \in [0,1]$. Indeed, the previous condition is equivalent to $f_1(0) \leq f_2(0)$, i.e., $1 \leq \frac{p_0}{1-p_0} \Rightarrow p_0 \geq \frac{1}{2}$, which does not satisfy Assumption A2. Note that, in this case, the lower bound for p in (35) would have been minimized for $\lambda = 0$, implying that $p > \frac{p_0}{1-p_0} \geq 1$, i.e., the method would not converge (cf. the statement of Theorem 4). Thus, we have only two cases:

Case 1 $f_1(\lambda) > f_2(\lambda)$ for all $\lambda \in [0,1]$, which means that the convergence of the method is dictated by the spatial discretization. Given that f_1 is decreasing and f_2 is increasing, the previous condition is equivalent to $f_1(1) \geq f_2(1)$, i.e., $\frac{\gamma}{r_{\mathrm{FEM}}+\gamma} \geq \frac{p_s}{1-p_s}$. In this case, the lower bound for p (35) is minimized for $\lambda = 1$, and we have $p > \frac{\gamma}{r_{\mathrm{FEM}}+\gamma}$.

Case 2 There exist $\lambda^* \in (0,1)$ s.t. $f_1(\lambda^*) = f_2(\lambda^*)$. This condition is equivalent to the two conditions

$$\begin{cases} f_1(0) \geq f_2(0) \\ f_1(1) \leq f_2(1) \end{cases} \Rightarrow \begin{cases} 1 \geq \frac{p_0}{1-p_0} \\ \frac{\gamma}{r_{\mathrm{FEM}}+\gamma} \leq \frac{p_s}{1-p_s}. \end{cases}$$

Letting $c = \frac{p_0}{1-p_0}$ and $\bar{c} = \frac{p_s}{1-p_s}$, we derive λ^* by equating

$$\begin{aligned}
\frac{\gamma}{\lambda r_{\mathrm{FEM}} + \gamma} &= \frac{c\bar{c}}{\bar{c} + \lambda(c - \bar{c})} \\
[\bar{c} + \lambda(c-\bar{c})]\gamma &= c\bar{c}(\lambda r_{\mathrm{FEM}} + \gamma) \\
\lambda[c\gamma - \bar{c}\gamma - c\bar{c} r_{\mathrm{FEM}}] &= c\bar{c}\gamma - \bar{c}\gamma \\
\lambda\left[\frac{1}{\bar{c}}\gamma - \frac{1}{c}\gamma - r_{\mathrm{FEM}}\right] &= \gamma - \frac{\gamma}{c} \\
\lambda^* &= \frac{\gamma - \frac{\gamma}{c}}{\frac{1}{\bar{c}}\gamma - \frac{1}{c}\gamma - r_{\mathrm{FEM}}},
\end{aligned}$$

which yields

$$p > \frac{\gamma}{\gamma + r_{\mathrm{FEM}}\left(\frac{\gamma}{\frac{p_0}{1-p_0}} - \gamma\right)\left(r_{\mathrm{FEM}} + \frac{\gamma}{\frac{p_0}{1-p_0}} - \frac{\gamma}{\frac{p_s}{1-p_s}}\right)^{-1}}.$$

We can now apply Theorem 4 and derive the convergence estimate,

$$\left|\mathbb{E}[F] - \mathcal{M}_{\mathcal{I}}[F]\right| \leq C_P(p)\mathrm{Work}[\mathcal{M}_{\mathcal{I}}]^{1-1/p},$$

which we reformulate as

$$\left|\mathbb{E}[F] - \mathcal{M}_{\mathcal{I}}[F]\right| \leq C_P\left(\frac{1}{1-r_{\mathrm{MISC}}}\right)\mathrm{Work}[\mathcal{M}_{\mathcal{I}}]^{-r_{\mathrm{MISC}}},$$

with $r_{\mathrm{MISC}} = \frac{1}{p} - 1$. The results above, stated in terms of $p > K$, translate to $r_{\mathrm{MISC}} < \frac{1}{K} - 1$. Elementary algebra then allows us to derive the final statement of the theorem. □

5 Analysis of Example 1

In this section, we determine the values of s, p_0 and p_s for Example 1. Let us define

$$\Upsilon_{\boldsymbol{k},\boldsymbol{\ell}}(\boldsymbol{x}) = \prod_{j=1}^{d}\left(\cos\left(\frac{\pi}{L}k_j x_j\right)\right)^{\ell_j}\left(\sin\left(\frac{\pi}{L}k_j x_j\right)\right)^{1-\ell_j},$$

so that κ from (6) can be written as

$$\begin{aligned}\kappa(\boldsymbol{x},\boldsymbol{y}) &= \sum_{\boldsymbol{k}\in\mathbb{N}^d} A_{\boldsymbol{k}} \sum_{\boldsymbol{\ell}\in\{0,1\}^d} y_{\boldsymbol{k},\boldsymbol{\ell}} \Upsilon_{\boldsymbol{k},\boldsymbol{\ell}}(\boldsymbol{x}) \\ &= \sum_{j=0}^{\infty} \sum_{\{\boldsymbol{k}\in\mathbb{N}^d\,:\,|\boldsymbol{k}|=j\}} A_{\boldsymbol{k}} \sum_{\boldsymbol{\ell}\in\{0,1\}^d} y_{\boldsymbol{k},\boldsymbol{\ell}} \Upsilon_{\boldsymbol{k},\boldsymbol{\ell}}(\boldsymbol{x}).\end{aligned}$$

Based on this expression, we analyze the summability of $\left\{A_{\boldsymbol{k}} \|D^{\boldsymbol{s}}\Upsilon_{\boldsymbol{k},\boldsymbol{\ell}}\|_{L^\infty(\mathcal{B})}\right\}$ for $|\boldsymbol{s}| = 0$ and $|\boldsymbol{s}| \le s$ to determine the values of p_0 and p_s, respectively. First note that for $|\boldsymbol{s}| \le s$ we can show for a constant c independent of $\boldsymbol{k}$ that

$$\|D^{\boldsymbol{s}}\Upsilon_{\boldsymbol{k},\boldsymbol{\ell}}(\boldsymbol{x})\|_{L^\infty(\mathcal{B})} = \prod_{j=1}^{d} \left(\frac{\pi}{L} k_j\right)^{s_j} \le c|\boldsymbol{k}|^s.$$

Then

$$\begin{aligned}\sum_{j=0}^{\infty} \sum_{\{\boldsymbol{k}\in\mathbb{N}^d\,:\,|\boldsymbol{k}|=j\}} \sum_{\boldsymbol{\ell}\in\{0,1\}^d} A_{\boldsymbol{k}}^{p_s} \|D^{\boldsymbol{s}}\Upsilon_{\boldsymbol{k},\boldsymbol{\ell}}\|_{L^\infty(\mathcal{B})}^{p_s} &\le c2^d \sum_{j=0}^{\infty} \sum_{\{\boldsymbol{k}\in\mathbb{N}^d\,:\,|\boldsymbol{k}|=j\}} 2^{p_s \frac{|\boldsymbol{k}|_0}{2}} |\boldsymbol{k}|^{p_s s} (1+|\boldsymbol{k}|^2)^{-\frac{p_s\left(\nu+\frac{d}{2}\right)}{2}} \\ &\le c2^d + c2^{d+p_s\frac{d}{2}} \sum_{j=1}^{\infty} \sum_{\{\boldsymbol{k}\in\mathbb{N}^d\,:\,|\boldsymbol{k}|=j\}} j^{-p_s\left(\nu+\frac{d}{2}-s\right)} \\ &= c2^d + c\frac{2^{d+p_s\frac{d}{2}}}{(d-1)!} \sum_{j=1}^{\infty} j^{-p_s\left(\nu+\frac{d}{2}-s\right)} \prod_{i=1}^{d-1}(j+i) \\ &= c2^d + c\frac{2^{d+p_s\frac{d}{2}}}{(d-1)!} \sum_{j=1}^{\infty} j^{-p_s\left(\nu+\frac{d}{2}-s\right)+d-1} \left(1+\frac{d-1}{j}\right)^{d-1} \\ &\le c2^d + \frac{c2^{d+p_s\frac{d}{2}} d^{d-1}}{(d-1)!} \sum_{j=1}^{\infty} j^{-p_s\left(\nu+\frac{d}{2}-s\right)+d-1}.\end{aligned}$$

From here we obtain the bounds

$$p_0 > \left(\frac{\nu}{d}+\frac{1}{2}\right)^{-1} \qquad \text{and} \qquad p_s > \left(\frac{\nu}{d}+\frac{1}{2}-\frac{s}{d}\right)^{-1}. \tag{36}$$

Moreover, imposing $p_0 < \frac{1}{2}$ and $p_s < \frac{1}{2}$ gives the bounds

$$\nu > \frac{3d}{2} \qquad \text{and} \qquad s < \nu - \frac{3d}{2}, \tag{37}$$

respectively. Since $\Upsilon_{\boldsymbol{k},\boldsymbol{\ell}} \in C^\infty(\mathcal{B})$, this is the only bound on the value of s. To determine the optimal value of s, we substitute $r_{\text{FEM}} = \min(1, \frac{s}{d})$ and the lower bounds of p_0 and p_s in Theorem 11 and simplify to obtain

$$r_{\text{MISC}} < \begin{cases} \frac{s}{d\gamma} & \frac{\nu}{d}-\frac{3}{2} \ge \frac{s}{d}\left(1+\frac{1}{\gamma}\right) \text{ and } s \le d, \\ \frac{1}{\gamma} & \frac{\nu}{d}-\frac{3}{2} \ge \frac{s}{d}+\frac{1}{\gamma} \text{ and } s \ge d, \\ \left(\frac{\nu}{d}-\frac{3}{2}\right)\left(\frac{1}{\gamma+1}\right) & \frac{\nu}{d}-\frac{3}{2} \le \frac{s}{d}\left(1+\frac{1}{\gamma}\right) \text{ and } s \le d, \\ \left(\frac{\nu}{d}-\frac{3}{2}\right)\left(\frac{d}{s\gamma+d}\right) & \frac{\nu}{d}-\frac{3}{2} \le \frac{s}{d}+\frac{1}{\gamma} \text{ and } s \ge d. \end{cases}$$

From here it is clear that the optimal value is $s = d$. Hence, to satisfy the bound (37) we make the choice

$$s = \min\left(d, \nu - \frac{3d}{2}\right) > 0. \tag{38}$$

In Figure 2, we plot the upper bound of the rate of MISC work complexity, r_{MISC}, based on Theorem 11 and the following cases:

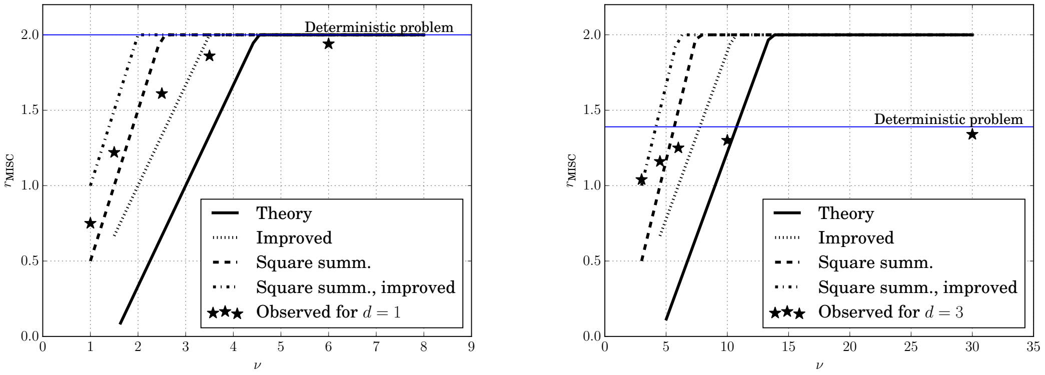


Figure 2: The upper bound of the MISC rate r_{MISC} as predicted in Theorem 11 versus the observed rates when running the example detailed in Section 6. Refer to Section 5 for an explanation of the different curves. Also included are the observed convergence rate for a few values of ν and the observed convergence rate of MISC with no random variable and constant diffusion coefficient, a, as a horizontal line. The latter is referred to as "deterministic problem" and shows more clearly the non-asymptotic effect of the logarithmic factor in the work for $d > 1$, as shown in Figure 7 and proved [22, Theorem 1].

Theory. This is based on the previous analysis, i.e., by considering the summability of $\left\{A_{\boldsymbol{k}} \| D^{\boldsymbol{s}} \Upsilon_{\boldsymbol{k},\boldsymbol{\ell}} \|_{L^\infty(\mathcal{B})}\right\}$ and plugging in the lower bounds of p_0 and p_s in (36) and s in (38). We also use the value $r_{\mathrm{FEM}} = 2\min\left(1, \frac{s}{d}\right)$. This is motivated by the fact that we expect to double the convergence rate of the underlying FEM method since we are considering convergence of a smooth functional of the solution.

Square summability. Motivated by the arguments in Lemma 16 in the appendix, we believe that our results can be improved by taking the values of p_0 and p_s as summability exponents of $\left\{A_{\boldsymbol{k}}^2 \| D^{\boldsymbol{s}} \Upsilon_{\boldsymbol{k},\boldsymbol{\ell}}^2 \|_{L^\infty(\mathcal{B})}\right\}$ for $|\boldsymbol{s}| = 0$ and $|\boldsymbol{s}| \le s$, respectively. Similar calculations to above yield the bounds

$$p_0 > \left(\frac{2\nu}{d} + 1\right)^{-1} \qquad \text{and} \qquad p_s > \left(\frac{2\nu}{d} + 1 - \frac{2s}{d}\right)^{-1}, \tag{39}$$

and the corresponding conditions $\nu > \frac{d}{2}$ and $s > \nu - \frac{d}{2}$, respectively. Here, we also use the value $s = \min\left(d, \nu - \frac{d}{2}\right) > 0$.

Improved. As mentioned previously in Remark 5, we could in principle make our results sharper by taking $q_{0,\min} = p_0$ instead of $q_{0,\min} = \frac{p_0}{1-p_0}$ and similarly for $q_{s,\min}$. The modifications of Theorem 11 to account for these rates are straightforward. Moreover, when considering square summability, the conditions become $\nu > 0$ and $s > \nu$ so that we can make the choice $s = \min(d, \nu) > 0$.

We also include in Figure 2 the observed convergence rates that were obtained numerically by running MISC with the example discussed in Section 6, and the convergence rate of MISC when applied to the same problem with no random variables and a constant diffusion coefficient a. In this case MISC reduces to a deterministic combination technique [8] and the observed convergence rate of MISC for any value of ν is necessarily less than the observed convergence rate of the deterministic combination technique.

From this figure, we can clearly see that the predicted rates in out theory are pessimistic when compared to the observed rates and that the suggested analysis of using the square summability or using the improved rates $q_{0,\min}$ and $q_{s,\min}$ might yield sharper bounds for the predicted work rates. On the other hand, we know from our previous work [22, Theorem 1] that the work degrades with increasing d with a log factor and in fact the expected work rate for maximum regularity when the number of random variables is finite is of $\mathcal{O}\left(W_{\max}^{-2} \log(W_{\max})^{d-1}\right)$. This can be seen Figure 2 and $d = 3$, since in this case observed work rate seems to be converging to a value less that 2.

6 Numerical experiments

We now verify the effectiveness of the MISC approximation on some instances of the general elliptic equation (1), as well as the validity of the convergence analysis detailed in the previous sections. In particular, we consider the family of random diffusion coefficients specified in Example 1. As already mentioned in Remark 1, this choice of random fields in not part of our theory since the Shift Theorem in Assumption A3 does not hold. Nonetheless, we work in this setting to obtain numerical evidence of the fact that the predicted results are valid for this choice of random fields as well, which means that the results are rather robust and apply to a broader class of problems.

In more detail, we consider a problem with one physical dimension ($d = 1$) and another with three dimensions ($d = 3$); in both cases, we set $f(\boldsymbol{x}) = 1$ and $h(\boldsymbol{x}) = 0$, and model the diffusion coefficient by the expansion (6) with different values of ν. Finally, the quantity of interest is a local average defined as

$$F(\boldsymbol{y}) = \int_{\mathcal{B}} u(\boldsymbol{x}, \boldsymbol{y}) q(\boldsymbol{x}) d\boldsymbol{x}, \quad q(\boldsymbol{x}) = \frac{10}{(\sigma\sqrt{2\pi})^d} \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{x}_0\|_2^2}{2\sigma^2}\right) \tag{40}$$

with $\sigma = 0.2$ and location $\boldsymbol{x}_0 = 0.3$ for $d = 1$ and $\boldsymbol{x}_0 = [0.3, 0.2, 0.6]$ for $d = 3$. The deterministic problems are discretized with a second-order centered finite differences scheme (for which we expect to recover in the numerical experiments the same rate that we would obtain with piece-wise bi-linear finite elements on a structured mesh) with mesh-sizes $h_i(\boldsymbol{\alpha}) = h_0 \times 2^{-\alpha_i}$ and $h_0 = 1/12$, and the resulting linear system is solved with GMRES with sufficient accuracy. The quadrature points on the stochastic domain are the already-mentioned Clenshaw-Curtis points (see eq. (15) and (16)).

In the plots below, to avoid discrepancies in running time due to implementation details, the computational work is compared in terms of the total number of degrees of freedom, i.e., using (23) and Assumption A4. Moreover, we set $\gamma = 1$ in (26), which is motivated by the fact that, for the tolerances we are interested in, we estimate that the cost of solving a linear system with GMRES is linear with respect to the number of degrees of freedom.

In order to evaluate the MISC estimator, we need to build the index set (25). To do that, we must be able to evaluate two quantities for every $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$: the work contribution, $\Delta W_{\boldsymbol{\alpha},\boldsymbol{\beta}}$, and the error contribution, $\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}}$. Evaluating the work contribution is straightforward, thanks to Assumption A4 and using $\gamma = 1$. On the other hand, evaluating the error contribution is more involved. We look at two options:

"a-posteriori" evaluation. We compute $\boldsymbol{\Delta}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]$ for all $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ withing some "universe" index set and set $\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}} = |\boldsymbol{\Delta}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]|$. Notice that this method is not practical since the cost of constructing the set $\mathcal{I}$ would far dominate the cost of the MISC estimator. However, within some "universe" index-set, this method would produce the best possible convergence and serve as a benchmark for other MISC sets within that universe.

"a-priori" evaluation. We use Lemma 8 to bound $\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}}$. Using these bounds instead of exact values produces quasi-optimal index sets (cf. [4]). This method in turn requires the estimation of the parameters $r_{\text{FEM}}, \{g_{0,n}\}_{n\geq 1}$ and $\{g_{s,n}\}_{n\geq 1}$. Since we use a second-order centered finite differences scheme and consider the convergence of a quantity of interest, we expect $r_{\text{FEM}} = 2$ for large enough ν. This can also be validated numerically in the usual way by fixing all random variables to their expected value and checking the decay of $\Delta E_{\boldsymbol{\alpha},\boldsymbol{1}}$ with respect to $\boldsymbol{\alpha}$.

On the other hand, estimating $\{g_{0,n}\}_{n\geq 1}$ and $\{g_{s,n}\}_{n\geq 1}$ is more difficult since, in principle, we do not know a priori if $\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ is decaying with rate $g_{0,n}$ or $g_{s,n}$. Instead, we use a "simplified" model that was used in [22]:

$$\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}} \leq C e^{-\sum_{n\geq 1} m(\beta_n - 1)\tilde{g}_n} 2^{-|\boldsymbol{\alpha}| r_{\text{FEM}}}, \tag{41}$$

where $\tilde{g}_n$ is some unknown function of $g_{0,n}$ and $g_{s,n}$. Estimating $\tilde{g}_n$ can be done given r_{FEM} and a set of evaluations of $|\boldsymbol{\Delta}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]|$ for some $(\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \mathcal{J}$ by solving a least-squares problem to fit the linear model

$$\sum_{n\geq 1} \tilde{g}_n m(\beta_n - 1) = -\log\left(|\boldsymbol{\Delta}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]|\right) - |\boldsymbol{\alpha}| r_{\text{FEM}}, \qquad \text{for all } (\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \mathcal{J}.$$

For our example, these rates are plotted in Figures 3(a) and 5(a) for $d = 1$ and $d = 3$, respectively. In our current implementation, the construction of the optimal MISC set, $\mathcal{I}$, is separate from the set $\mathcal{J}$. However, it is possible in principle to construct an algorithm in which the optimal MISC set, $\mathcal{I}$, is constructed iteratively by alternating between estimating rates given a set of indices and evaluating the MISC estimator.

Note that, in the current work, there are certain operations whose costs we do not track or compare. The first operation is the estimation of the stochastic rates, $\{\tilde{g}_n\}_{n\geq 1}$. The second operation is the construction of the optimal set given estimates of error and work contribution. We believe that the cost of these two operations can be reduced by using the previously mentioned iterative algorithm, so that the cost of these operations is dominated by the cost of evaluating the MISC estimator. The third operation is the assembly of the stiffness matrix, especially since it scales linearly with the number of random variables. While the cost of this operation is relevant to our discussion, it is usually dominated by the cost of the linear solver, at least for fine enough discretizations.

Finally, we also compare MISC to Multi-index Monte Carlo (MIMC) method as detailed in [23], for which $\mathcal{O}\left(W_{\max}^{-0.5}\right)$ convergence can be proved for Example 1 with $\gamma = 1, d \leq 3$ and sufficiently large ν (see Appendix A). Moreover, when computing errors, we use the result obtained using a well-resolved MISC solution as a reference value.

Figures 3(b–d) and Figures 5(b–d) compare some computed values of $|\boldsymbol{\Delta}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]|$ versus the model (41) using the estimated rates $r_{\text{FEM}} = 2$ and $\{\tilde{g}_n\}_{n\geq 1}$. These plots show that the model (41) is a good fit for the case $d = 1, \nu = 2.5$ and $d = 3, \nu = 4.5$, respectively. Moreover, similar plots were produced for other values of d and ν that are not reported here but also show good fit.

Figures 4 and 6 show

- the maximum space discretization level, $\max_{(\boldsymbol{\alpha},\boldsymbol{\beta})\in\mathcal{I}} \max(\boldsymbol{\alpha})$,
- the maximum quadrature level, $\max_{(\boldsymbol{\alpha},\boldsymbol{\beta})\in\mathcal{I}} \max(\boldsymbol{\beta})$,
- the index of the last activated random variable, $\max_{(\boldsymbol{\alpha},\boldsymbol{\beta})\in\mathcal{I}} \max_{\beta_j>1} j$,
- and the maximum number of jointly activated variables, $\max_{(\boldsymbol{\alpha},\boldsymbol{\beta})\in\mathcal{I}} |\boldsymbol{\beta} - 1|_0$.

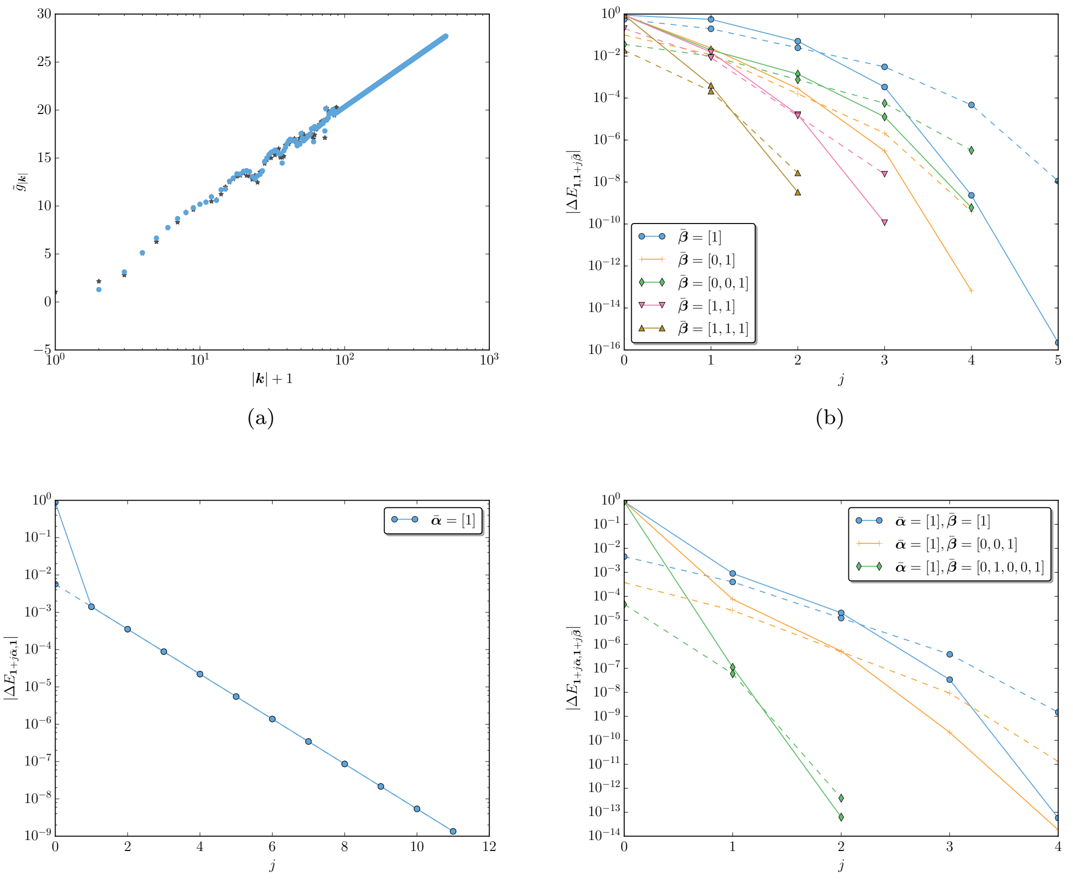
These values convey the size of the used index set, $\mathcal{I}$, for different values of $W_{\max}$.

As previously discussed, Figure 2 shows the observed convergence rates of MISC vs MIMC for the cases $d = 1, \nu = 2.5$ and $d = 3, \nu = 4.5$. Other examples for different values of ν were also validated and all observed MISC convergence rates are included in Figure 2. This figure shows that the observed rates are better than what is predicted by the theory developed in this work, which suggests that further improvement in the theory is possible as suggested in Remark 5. Figures 7 and 8 show in greater details a few of the observed convergence curves and their respective linear fit in log-log scale.

We recall that, as shown in [22, Theorem 1], the convergence rate of MISC with a finite number of random variables is $\mathcal{O}\left(W_{\max}^{-2} \log(W_{\max})^{d-1}\right)$. Compare this to the theory presented here that predicts, as $\nu \to \infty$, a convergence of $\mathcal{O}\left(W_{\max}^{-2+\epsilon}\right)$ for any $\epsilon > 0$. However, Figure 7 shows that even for a problem with $d = 3$ and no random variables, MISC (which, in this case, becomes equivalent to a deterministic combination technique [8]) has an observed convergence rate that is closer to -1.38. This is due to the non-asymptotic effect of the logarithmic term that is nonzero for $d > 1$. Based on this, we should not expect a better convergence rate for $d = 3$ and any finite ν. This is also numerically validated in Figure 8 that shows the full convergence curves for $d = 1, \nu = 2.5$ and $d = 3, \nu = 4.5$.

7 Conclusions

In this work, we analyzed the performance of the MISC method when applied to problems depending on a countable sequence of random variables. We proved a convergence result using a summability argument, showing that in certain cases the convergence of the method is essentially dictated by the convergence properties of the deterministic solver. We have then applied the convergence theorem to derive convergence rates for the approximation of the expected value of a functional of the



(a)

(b)

(c)

(d)

Figure 3: Example 1, $d = 1$ and $\nu = 2.5$. (a) The estimated stochastic rates, $\tilde{g}_n$, that are used in (41). Different markers correspond to different modes multiplying the same value of $A_{\boldsymbol{k}}$. (b–d) The computed $\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}} = |\boldsymbol{\Delta}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]|$ with solid lines versus the model in (41) with dashed lines.

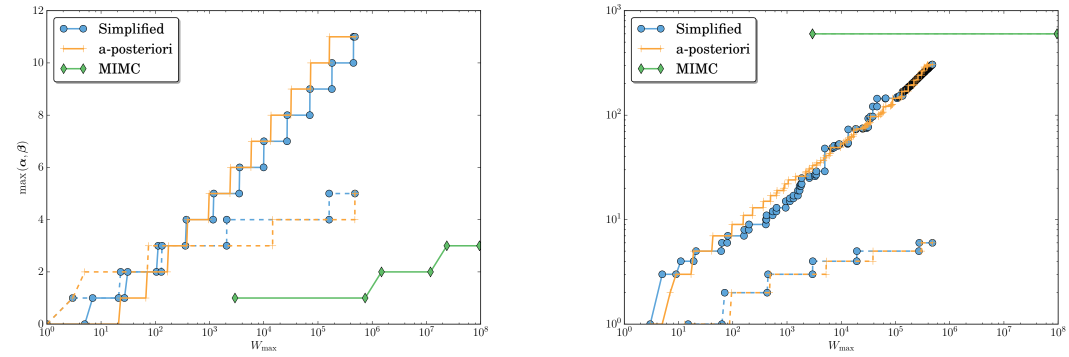
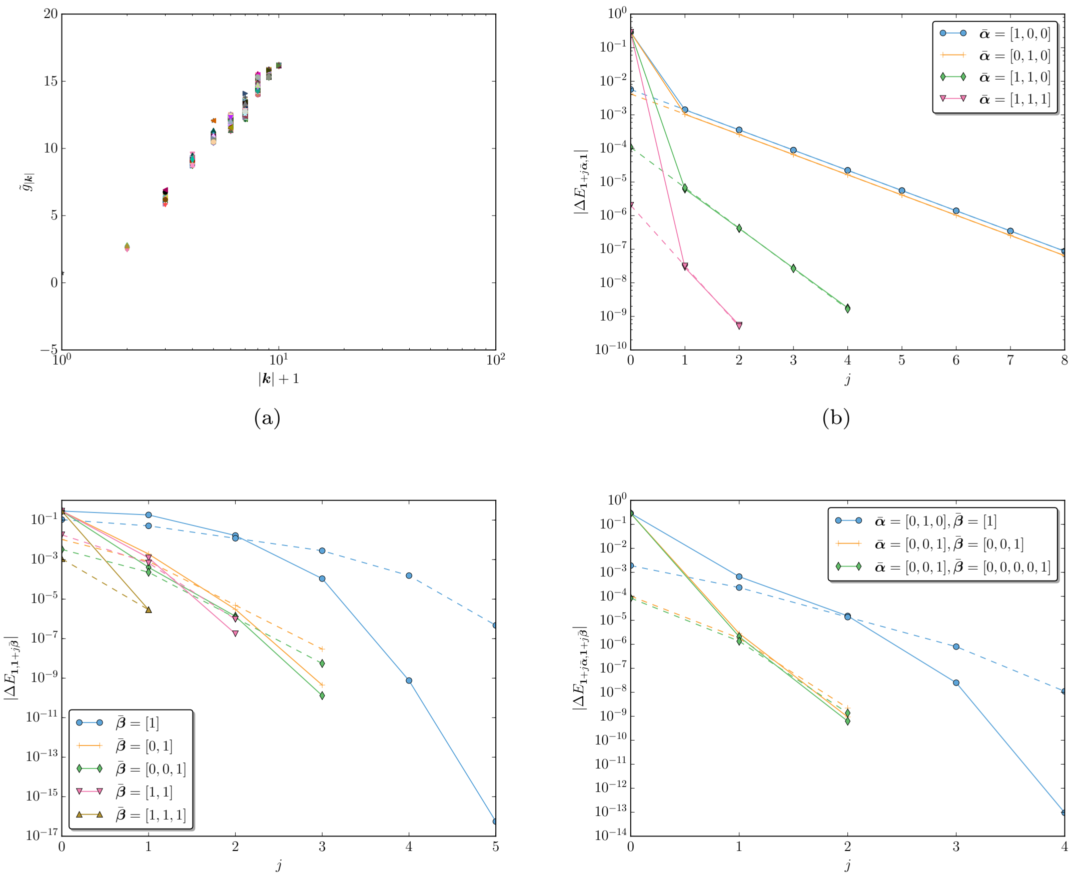


Figure 4: Example 1, $d = 1$ and $\nu = 2.5$. This figure shows extreme values of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ included in the MISC set, $\mathcal{I}$. Specifically, left-solid is the maximum space discretization level, $\max_{(\boldsymbol{\alpha},\boldsymbol{\beta})\in\mathcal{I}}\left(\max\left(\boldsymbol{\alpha}\right)\right)$, left-dashed is the maximum quadrature level, $\max_{(\boldsymbol{\alpha},\boldsymbol{\beta})\in\mathcal{I}}\left(\max\left(\boldsymbol{\beta}\right)\right)$, right-solid is the index of the last activated random variable, $\max_{(\boldsymbol{\alpha},\boldsymbol{\beta})\in\mathcal{I}}\left(\max_{\beta_j>1} j\right)$, and right-dashed is the maximum number of jointly activated variables, $\max_{(\boldsymbol{\alpha},\boldsymbol{\beta})\in\mathcal{I}}\left(|\boldsymbol{\beta}-1|_0\right)$.



(a)

(b)

(c)

(d)

Figure 5: Example 1, $d=3$ and $\nu=4.5$. (a) The estimated stochastic rates, $\tilde{g}_n$, that are used in (41). Here different markers correspond to different modes multiplying the same value of $A_{\boldsymbol{k}}$. (b–d) The computed $\Delta E_{\boldsymbol{\alpha},\boldsymbol{\beta}} = |\boldsymbol{\Delta}[F_{\boldsymbol{\alpha},\boldsymbol{\beta}}]|$ with solid lines versus the model in (41) with dashed lines.

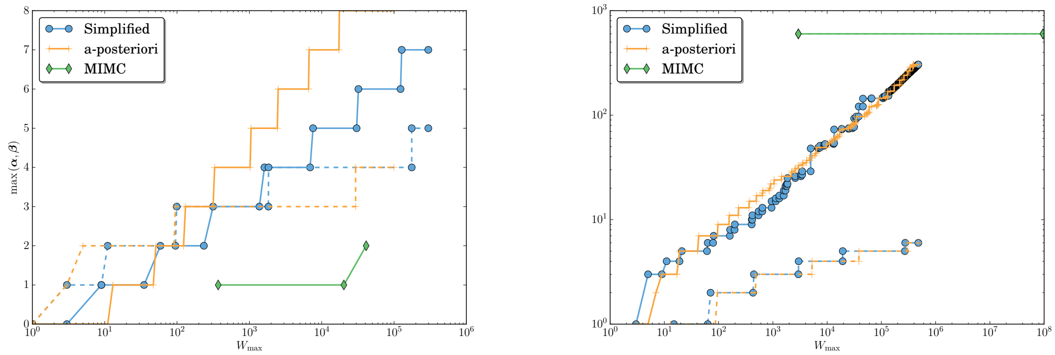
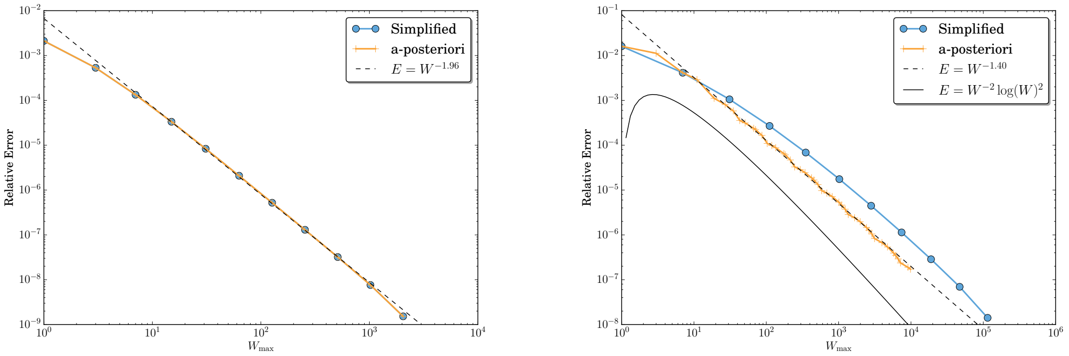


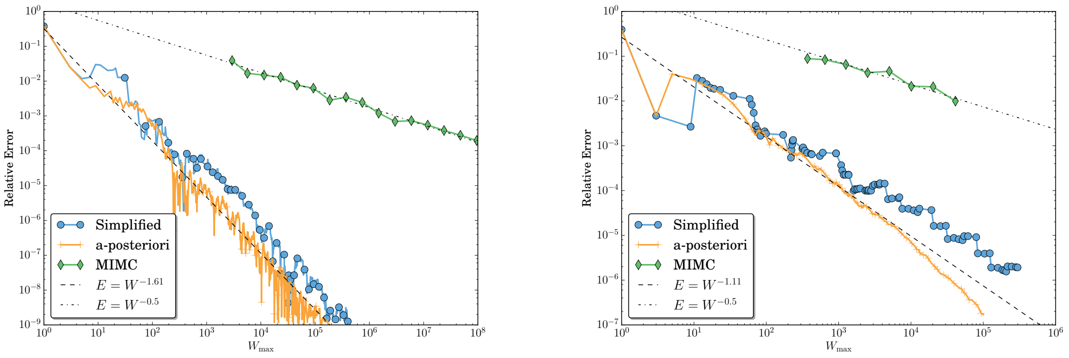
Figure 6: Example 1, $d=3$ and $\nu=4.5$. This figure shows extreme values of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ included in the MISC set $\mathcal{I}$. Specifically, left-solid is the maximum space discretization level, $\max_{(\boldsymbol{\alpha},\boldsymbol{\beta})\in\mathcal{I}}(\max(\boldsymbol{\alpha}))$, left-dashed is the maximum quadrature level, $\max_{(\boldsymbol{\alpha},\boldsymbol{\beta})\in\mathcal{I}}(\max(\boldsymbol{\beta}))$, right-solid is the index of the last activated random variable, $\max_{(\boldsymbol{\alpha},\boldsymbol{\beta})\in\mathcal{I}}\left(\max_{\beta_j>1} j\right)$, and right-dashed is the maximum number of jointly activated variables, $\max_{(\boldsymbol{\alpha},\boldsymbol{\beta})\in\mathcal{I}}(|\boldsymbol{\beta}-1|_0)$.



(a) $d = 1$

(b) $d = 3$

Figure 7: Convergence results of MISC Example 1 with a constant diffusion coefficient, a. In this case, MISC is equivalent to a deterministic combination technique [8]. These plots shows the non-asymptotic effect of the logarithmic factor for $d > 1$ (as discussed in [22, Theorem 1]) on the linear convergence fit in log-log scale.



(a) $d = 1, \nu = 2.5$

(b) $d = 3, \nu = 4.5$

Figure 8: Convergence results of MISC vs MIMC when applied to Example 1.

solution of an elliptic PDE with the diffusion coefficient described by a random field, tracking the dependence of the convergence rate on the spatial regularity of the realizations of the random field. The theoretical findings are backed up by numerical experiments that show the dependence of the convergence rate on the regularity parameter. Future works includes extending the convergence analysis to higher-order finite element solvers and improving the estimates of the error contribution of each difference operator by taking into account the factorial terms appearing in the estimates for the size of the Chebyshev coefficients, cf. [3, 12]. Moreover, the ideas in [13] can be extended to design an algorithm that iteratively estimates the optimal MISC set by alternating between optimizing the set and evaluating the estimator such as to ensure that the work to optimize the set is dominated by the work to evaluate the MISC estimator.

Acknowledgement 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.

A Summability of series expansion

We start recalling a useful multivariate Faà di Bruno formula taken from [14, Theorem 2.1].

Lemma 12. *Let $\mathcal{B} \subset \mathbb{R}^d$ be an open domain, $g : \mathcal{B} \to \mathbb{R}$ and $f : \mathbb{R} \to \mathbb{R}$ be functions of class C^s and denote $h = f \circ g : \mathcal{B} \to \mathbb{R}$. For any multi-index $\boldsymbol{i} \in \mathbb{N}^d$, $|\boldsymbol{i}| \leq s$, and any $\boldsymbol{x} \in \mathcal{B}$,*

$$D^{\boldsymbol{i}} h(g(\boldsymbol{x})) = \boldsymbol{i}! \sum_{\lambda=1}^{|\boldsymbol{i}|} f^{(\lambda)}(g(\boldsymbol{x})) \sum_{r=1}^{\lambda} \sum_{p_r(\boldsymbol{i},\lambda)} \prod_{j=1}^{r} \frac{(D^{\boldsymbol{\ell}_j} g(\boldsymbol{x}))^{k_j}}{k_j!(\boldsymbol{\ell}_j!)^{k_j}}, \tag{42}$$

holds, where

$$p_r(\boldsymbol{i},\lambda) = \{(k_j, \boldsymbol{\ell}_j) \in \mathbb{N} \times \mathbb{N}_0^d,\ j = 1, \ldots, r :\ \mathbf{0} \prec \boldsymbol{\ell}_1 \prec \boldsymbol{\ell}_2 \prec \cdots \prec \boldsymbol{\ell}_r,\ \sum_{j=1}^{r} k_j = \lambda,\ \sum_{j=1}^{r} k_j \boldsymbol{\ell}_j = \boldsymbol{i}\}$$

and $\prec$ denotes the lexicographic ordering of multi-indices. The set $p_r(\boldsymbol{i},\lambda)$ denotes the set of possible decompositions of $\boldsymbol{i}$ as a sum of λ multi-indices with $r \leq \lambda$ distinct multi-indices $\boldsymbol{\ell}_j$ taken with multiplicity k_j such that $\sum_{j=1}^{r} k_j = \lambda$.

Still from [14, Corollary 2.9], we have that, for any $\boldsymbol{i} \in \mathbb{N}^d$,

$$\boldsymbol{i}! \sum_{r=1}^{\lambda} \sum_{p_r(\boldsymbol{i},\lambda)} \prod_{j=1}^{r} \frac{1}{k_j!(\boldsymbol{\ell}_j!)^{k_j}} = S_{|\boldsymbol{i}|,\lambda},$$

where $S_{n,k}$ is the Stirling number of the second kind, which counts the number of ways to partition a set of n objects into k non-empty subsets. Similarly, the Bell number, $B_n = \sum_{k=0}^{n} S_{n,k}$, counts the number of partitions of a set of n objects, whereas the ordered Bell numbers are defined by $\tilde{B}_n = \sum_{k=0}^{n} k! S_{n,k}$ and satisfy the recursive relation $\tilde{B}_n = \sum_{k=0}^{n-1} \binom{n}{k} \tilde{B}_k$. Clearly, $B_n \leq \tilde{B}_n$. Moreover, the bound $B_n \leq \tilde{B}_n \leq n!/(\log 2)^n$ has been given in [3, Lemma A.3].

We now use these results to show the following result

Lemma 13. *Let $\mathcal{B} \subset \mathbb{R}^d$ be an open-bounded domain and $\kappa \in C^s(\mathcal{B})$ (real or complex valued). Then, $a = e^{\kappa} \in C^s(\mathcal{B})$ and*

$$\|a\|_{C^s(\mathcal{B})} \leq \frac{s!}{(\log 2)^s} \|a\|_{C^0(\mathcal{B})} (1 + \|\kappa\|_{C^s(\mathcal{B})})^s.$$

Proof. Using formula (42) we have for any $\boldsymbol{i} \in \mathbb{N}^d$, $|\boldsymbol{i}| \leq s$ and any $\boldsymbol{x} \in \mathcal{B}$

$$\begin{aligned}|D^{\boldsymbol{i}} e^{\kappa(\boldsymbol{x})}| &= \boldsymbol{i}! \sum_{\lambda=1}^{|\boldsymbol{i}|} e^{\kappa(\boldsymbol{x})} \sum_{r=1}^{\lambda} \sum_{p_r(\boldsymbol{i},\lambda)} \prod_{j=1}^{r} \frac{|D^{\boldsymbol{\ell}_j} \kappa(\boldsymbol{x})|^{k_j}}{k_j!(\boldsymbol{\ell}_j!)^{k_j}} \leq \|a\|_{C^0(\mathcal{B})} \sum_{\lambda=1}^{|\boldsymbol{i}|} \|\kappa\|_{C^s(\mathcal{B})}^{\lambda} S_{|\boldsymbol{i}|,\lambda} \\ &\leq \|a\|_{C^0(\mathcal{B})} (1 + \|\kappa\|_{C^s(\mathcal{B})})^{|\boldsymbol{i}|} B_n.\end{aligned}$$

The result then follows from the recalled bound on the Bell numbers. □

A.1 $L^p(\Gamma)$ summability, pointwise in space

We consider now a diffusion coefficient as in Assumption A2:

$$a(\boldsymbol{x}, \boldsymbol{y}) = \exp\left\{ \sum_{j \in \mathbb{N}_+} \psi_j(\boldsymbol{x}) y_j \right\} = \prod_{j=1}^{\infty} e^{y_j \psi_j(\boldsymbol{x})}, \qquad \boldsymbol{x} \in \mathcal{B},$$

with y_j, $j \in \mathbb{N}_+$, i.i.d. uniformly distributed in $[-1, 1]$ and recall the definition of the sequence $\boldsymbol{b}_s = \{b_{s,j}\}_{j \in \mathbb{N}_+}$, $s \in \mathbb{N}$ in (3)-(4):

$$b_{s,j} \;=\; \max_{\boldsymbol{s} \in \mathbb{N}^d : |\boldsymbol{s}| \leq s} \|D^{\boldsymbol{s}} \psi_j\|_{L^\infty(\mathcal{B})} \;, \qquad j \geq 1.$$

Lemma 14. *If $\boldsymbol{b}_0 \in \ell^2$ then $\mathbb{E}[a(\boldsymbol{x})^p] < \infty$ for all $0 < p < \infty$ and $\forall \boldsymbol{x} \in \mathcal{B}$.*

Proof. For any $\boldsymbol{x} \in \mathcal{B}$ we estimate the p-th moment of $a(\boldsymbol{x}, \boldsymbol{y})$, exploiting independence of the random variables y_j:

$$\mathbb{E}[a(\boldsymbol{x})^p] = \mathbb{E}\left[\prod_{j=1}^{\infty} e^{p y_j \psi_j(\boldsymbol{x})}\right] = \prod_{j=1}^{\infty} \mathbb{E}\left[e^{p y_j \psi_j(\boldsymbol{x})}\right] = \prod_{j=1}^{\infty} \frac{\sinh(p\psi_j(\boldsymbol{x}))}{p\psi_j(\boldsymbol{x})} = \exp\left\{\sum_{j=1}^{\infty} \log\left(\frac{\sinh(p\psi_j(\boldsymbol{x}))}{p\psi_j(\boldsymbol{x})}\right)\right\}$$

where in the last two equalities we have implicitly assumed that $\sinh(z)/z = 1$ for $z = 0$. Setting $\theta_0(p; \boldsymbol{x}) = \prod_{j=1}^{\infty} \frac{\sinh(p\psi_j(\boldsymbol{x}))}{p\psi_j(\boldsymbol{x})}$ and observing that $\log(\sinh(z)/z) \sim z^2/6$, we have

$$\mathbb{E}[a(\boldsymbol{x})^p] = \theta_0(p; \boldsymbol{x}) < \infty \quad \forall \boldsymbol{x} \in \mathcal{B}, \;\; 0 < p < \infty \qquad \Longleftrightarrow \qquad \sum_{j=1}^{\infty} \psi_j(\boldsymbol{x})^2 < \infty.$$

Since $\sum_{j=1}^{\infty} b_{0,j}^2 < \infty$ implies $\sum_{j=1}^{\infty} \psi_j(\boldsymbol{x})^2 < \infty$ for any $\boldsymbol{x} \in \mathcal{B}$, this concludes the proof. □

A similar result holds for higher order derivatives of a.

Lemma 15. *Let $s \in \mathbb{N}_+$. If $\boldsymbol{b}_s \in \ell^2$ then for any $\boldsymbol{i} \in \mathbb{N}^d$, $|\boldsymbol{i}| = s$, $\mathbb{E}\left[(D^{\boldsymbol{i}} a(\boldsymbol{x}))^{2p}\right] < \infty$ for all $0 < p < \infty$ and $\forall \boldsymbol{x} \in \mathcal{B}$.*

Proof. Since the calculations are very tedious, we prove the result only for $s = 1$. Using the chain rule, we have

$$\begin{aligned}(\partial_{x_i} a(\boldsymbol{x}, \boldsymbol{y}))^{2p} &= \left(\sum_{j \in \mathbb{N}_+} a(\boldsymbol{x}, \boldsymbol{y}) \partial_{x_i} \psi_j(\boldsymbol{x}) y_j\right)^{2p} = a(\boldsymbol{x}, \boldsymbol{y})^{2p} \sum_{\substack{\boldsymbol{q} \in \mathbb{N}^{\mathbb{N}_+} \\ |\boldsymbol{q}| = 2p}} (2p)! \prod_{j=1}^{\infty} \frac{1}{q_j!} (\partial_{x_i} \psi_j(\boldsymbol{x}) y_j)^{q_j} \\ &= \sum_{\substack{\boldsymbol{q} \in \mathbb{N}^{\mathbb{N}_+} \\ |\boldsymbol{q}| = 2p}} (2p)! \prod_{j=1}^{\infty} \frac{1}{q_j!} (\partial_{x_i} \psi_j(\boldsymbol{x}) y_j)^{q_j} e^{2p y_j \psi_j(\boldsymbol{x})}\end{aligned}$$

Hence

$$\mathbb{E}\left[(\partial_{x_i} a(\boldsymbol{x}, \boldsymbol{y}))^{2p}\right] = \sum_{\substack{\boldsymbol{q} \in \mathbb{N}^{\mathbb{N}_+} \\ |\boldsymbol{q}| = 2p}} (2p)! \prod_{j=1}^{\infty} (\partial_{x_i} \psi_j(\boldsymbol{x}))^{q_j} \mathbb{E}\left[\frac{1}{q_j!} y_j^{q_j} e^{2p y_j \psi_j(\boldsymbol{x})}\right].$$

We now distinguish between q_j even or odd. For q_j even we have

$$\mathbb{E}\left[\frac{1}{q_j!}y_j^{q_j}e^{2py_j\psi_j(\boldsymbol{x})}\right] \leq \mathbb{E}\left[\frac{1}{q_j!}e^{2py_j\psi_j(\boldsymbol{x})}\right] = \frac{1}{q_j!}\frac{\sinh(2p\psi_j(\boldsymbol{x}))}{2p\psi_j(\boldsymbol{x})}$$

while for q_j odd we have

$$\begin{aligned}\mathbb{E}\left[\frac{1}{q_j!}y_j^{q_j}e^{2py_j\psi_j(\boldsymbol{x})}\right] &= \frac{1}{q_j!}\int_{-1}^{1}\frac{1}{2}y^{q_j}e^{2py\psi_j(\boldsymbol{x})}dy = \frac{1}{q_j!}\int_0^1 y^{q_j}\sinh(2py\psi_j(\boldsymbol{x}))dy \\ &= \frac{1}{q_j!}\sum_{n=0}^{\infty}\frac{(2p\psi_j(\boldsymbol{x}))^{2n+1}}{(2n+1)!}\int_0^1 y^{2n+1+q_j}dy = \frac{1}{q_j!}\sum_{n=0}^{\infty}\frac{(2p\psi_j(\boldsymbol{x}))^{2n+1}}{(2n+1)!(2n+2+q_j)} \\ &\leq \frac{1}{(q_j+1)!}\sinh(2p|\psi_j(\boldsymbol{x})|) \leq \frac{2pb_{1,j}}{(q_j+1)!}\frac{\sinh(2p\psi_j(\boldsymbol{x}))}{2p\psi_j(\boldsymbol{x})}\end{aligned}$$

Hence, defining the function

$$f(q_j) = \begin{cases}\frac{1}{q_j!} & \text{for } q_j \text{ even} \\ \frac{2pb_{1,j}}{(q_j+1)!} & \text{for } q_j \text{ odd}\end{cases}$$

we have

$$\begin{aligned}\mathbb{E}\left[(\partial_{x_i}a(\boldsymbol{x},\boldsymbol{y}))^{2p}\right] &\leq \sum_{\substack{\boldsymbol{q}\in\mathbb{N}^{\mathbb{N}}+\\|\boldsymbol{q}|=2p}}(2p)!\prod_{j=1}^{\infty}b_{1,j}^{q_j}f(q_j)\frac{\sinh(2p\psi(\boldsymbol{x}))}{2p\psi_j(\boldsymbol{x})} = \theta_0(2p;\boldsymbol{x})\sum_{\substack{\boldsymbol{q}\in\mathbb{N}^{\mathbb{N}}+\\|\boldsymbol{q}|=2p}}(2p)!\prod_{j=1}^{\infty}b_{1,j}^{q_j}f(q_j) \\ &\leq \theta_0(2p;\boldsymbol{x})\sum_{\substack{\boldsymbol{q}\in\mathbb{N}^{\mathbb{N}}+\\|\boldsymbol{q}|=2p,\boldsymbol{q}\text{ even}}}(2p)!(1+2p)^{|\boldsymbol{q}|_0}\prod_{j=1}^{\infty}\frac{b_{1,j}^{q_j}}{q_j!} \\ &\leq (1+2p)^p\theta_0(2p;\boldsymbol{x})\sum_{\substack{\boldsymbol{q}\in\mathbb{N}^{\mathbb{N}}+\\|\boldsymbol{q}|=p}}(2p)!\prod_{j=1}^{\infty}\frac{b_{1,j}^{2q_j}}{(2q_j)!} \\ &\leq (1+2p)^p(2p)^p\theta_0(2p;\boldsymbol{x})\sum_{\substack{\boldsymbol{q}\in\mathbb{N}^{\mathbb{N}}+\\|\boldsymbol{q}|=p}}p!\prod_{j=1}^{\infty}\frac{(b_{1,j}^2)^{q_j}}{q_j!} = (1+2p)^p(2p)^p\theta_0(2p;\boldsymbol{x})\left(\sum_{j\in\mathbb{N}_+}b_{1,j}^2\right)\end{aligned}$$

from which we see that $\mathbb{E}\left[(\partial_{x_i}a(\boldsymbol{x},\boldsymbol{y}))^{2p}\right]$ is bounded for any $0 \leq p < \infty$ and any $\boldsymbol{x} \in \mathcal{B}$ if $\boldsymbol{b}_1 \in \ell^2$. □

A.2 $L^p(\Gamma)$ summability, uniform in space

Assuming now that $\boldsymbol{b}_s \in \ell^2$ so that the random field, a, is s-times differentiable in an $L^p(\Gamma)$ sense according to Lemma 15, we show that this implies some uniform $L^p(\Gamma)$ summability as detailed in the next lemma.

Lemma 16. *Let $s \in \mathbb{N}_+$. If $\boldsymbol{b}_s \in \ell^2$ then $\mathbb{E}\left[\|a\|^p_{W^{\upsilon,\infty}(\mathcal{B})}\right] < \infty$ for all $1 \leq p < \infty$ and $\upsilon < s$.*

Proof. We exploit the Sobolev embedding $W^{\upsilon+\frac{d}{2q},2q}(\mathcal{B}) \subseteq W^{\upsilon,\infty}(\mathcal{B})$ for all $\upsilon \geq 0$ and $q \geq 1$. For $q \geq \max\{d/2(s-\upsilon), p/2\}$, we have

$$\begin{aligned}\mathbb{E}\left[\|a\|^p_{W^{\upsilon,\infty}(\mathcal{B})}\right] &\leq \mathbb{E}\left[\|a\|^{2q}_{W^{s-\frac{d}{2q},\infty}(\mathcal{B})}\right] \lesssim \mathbb{E}\left[\|a\|^{2q}_{W^{s,2q}(\mathcal{B})}\right] = \mathbb{E}\left[\sum_{|\boldsymbol{i}|\leq s}\int_{\mathcal{B}}(D^{\boldsymbol{i}}a(\boldsymbol{x}))^{2q}d\boldsymbol{x}\right] \\ &= \sum_{|\boldsymbol{i}|\leq s}\int_{\mathcal{B}}\mathbb{E}\left[(D^{\boldsymbol{i}}a(\boldsymbol{x}))^{2q}\right]d\boldsymbol{x} < \infty\end{aligned}$$

the last term being bounded from Lemma 15. □

Now we directly observe by taking $\upsilon = 0$ in the previous result that $a_{\max} = \|a\|_{L^\infty(\mathcal{B})}$ has bounded moments,

$$\mathbb{E}[a_{\max}^p] < \infty,$$

for all $1 \leq p < \infty$ and $0 < s$. Finally, by observing that due to the construction (2) in Assumption A2 we have that $a_{\min} = \frac{1}{\|a^{-1}\|_{L^\infty(\mathcal{B})}}$ has the same distribution as $a_{\max}$. As a consequence, $a_{\min}$ has bounded moments as well. This implies in turn that Assumption A1 holds and thus problem (1) is well posed in the Bochner space $L^p\left(\Gamma; H_0^1(\mathcal{B})\right)$, namely

Corollary 17 (Well posedness with log uniform coefficient)**.** *We have for $0 < \nu$ that the problem in Example 1 is well posed in the Bochner space $L^q\left(\Gamma; H_0^1(\mathcal{B})\right)$. The corresponding solution, u, satisfies*

$$\|u\|_{L^p(\Gamma; H_0^1(\mathcal{B}))} \leq C\mathbb{E}\left[\frac{1}{a_{\min}^p}\right]^{1/p} \|f\|_{H^{-1}(\mathcal{B})}.$$

We observe that higher regularity of the solution, u, can be obtained by using larger values of s in Lemma 16. This in turn yields control on moments of $W^{\upsilon,\infty}(\mathcal{B})$ norms of the coefficient, a, and following, for instance, estimates similar to (2.10) in [18, Theorem 2.4] we can estimate moments of the $H^{1+s}(\mathcal{B})$ norm of the solution, u. These regularity estimates, once combined with pathwise error estimates for the combination technique, can be further used to then show the corresponding ν-dependent convergence rates of MIMC [23] for Example 1, similar to what was done in Section 5 in the current work for MISC.

B Shift theorem for problem (1)

In this appendix, we seek to establish a shift theorem for the problem

$$\begin{cases} -\operatorname{div}(a(\boldsymbol{x})\, \nabla u(\boldsymbol{x})) = f(\boldsymbol{x}) & \text{in} \quad \mathcal{B} = [0,1]^d \\ u(\boldsymbol{x}) \, = \, 0 & \text{on} \quad \partial\mathcal{B}, \end{cases} \tag{43}$$

under suitable assumptions on a and f.

With respect to problem (1), for convenience we have dropped the dependence on the parameter vector, $\boldsymbol{y}$. We consider an odd periodic extension of f, on $[-1,1]^d$, and an even periodic extension of the coefficient a on $[-1,1]^d$, named, respectively, $\tilde{f}$, $\tilde{a}$. More precisely, for $\boldsymbol{j} = \{0,1\}^d$, we denote by $\boldsymbol{x_j} = ((-1)^{j_1}x_1, \ldots, (-1)^{j_d}x_d)$ and

$$\tilde{f}(\boldsymbol{x_j} + 2\boldsymbol{k}) = (-1)^{|\boldsymbol{j}|} f(\boldsymbol{x}), \qquad \tilde{a}(\boldsymbol{x_j} + 2\boldsymbol{k}) = a(\boldsymbol{x}), \qquad \forall \boldsymbol{x} \in [0,1]^2, \;\; \boldsymbol{j} \in \{0,1\}^d, \;\; \boldsymbol{k} \in \mathbb{N}^d$$

The following Shift theorem holds for problem (43).

Lemma 18. *If the coefficient a is such that its periodic extension satisfies $\tilde{a} \in W^{s,\infty}(\mathbb{R}^d)$, $s \geq 0$ and $f \in C_0^\infty(\mathcal{B})$ then $u \in H^{s+1}(\mathcal{B})$.*

Proof. We define the extended problem

$$\begin{cases} -\operatorname{div}(\tilde{a}(\boldsymbol{x})\, \nabla \tilde{u}(\boldsymbol{x})) = \tilde{f}(\boldsymbol{x}) & \text{in} \quad \tilde{\mathcal{B}} = [-1,1]^d \\ \int_{\tilde{\mathcal{B}}} u(\boldsymbol{x}) \, = \, 0 & \\ \text{periodic boundary conditions on } \partial\tilde{\mathcal{B}} & \end{cases}$$

Since by assumption $\tilde{a} \in L^\infty(\mathbb{R}^d)$ and $\tilde{f} \in L^2(\tilde{\mathcal{B}})$ this problem has a unique solution $\tilde{u} \in H^1_{per}(\tilde{\mathcal{B}})\backslash\mathbb{R}$, where we denote with $H^s_{per}(\tilde{\mathcal{B}})$ the space of periodic functions with (periodic) square integrable derivatives up to order s. It is easy to check that the solution $\tilde{u}$ is odd, that is $\tilde{u}(\boldsymbol{x_j}) = (-1)^{\boldsymbol{j}}\tilde{u}(\boldsymbol{x})$, $\forall \boldsymbol{x} \in [0,1]^d$, hence $\tilde{u} = 0$ (in the sense of traces) on $\partial\mathcal{B}$ and it coincides with the (unique) solution of (43) on $\mathcal{B}$. Moreover, standard elliptic regularity arguments allow us to say that $\tilde{u} \in H^s_{per}(\tilde{\mathcal{B}})$, hence $u \in H^s(\mathcal{B})$. □

References

[1] I. Babuška, F. Nobile, and R. Tempone. A stochastic collocation method for elliptic partial differential equations with random input data. SIAM Review, 52(2):317–355, June 2010.

[2] A. Barth, C. Schwab, and N. Zollinger. Multi-level Monte Carlo finite element method for elliptic PDEs with stochastic coefficients. Numerische Mathematik, 2011.

[3] J. Beck, F. Nobile, L. Tamellini, and R. Tempone. On the optimal polynomial approximation of stochastic PDEs by Galerkin and collocation methods. Mathematical Models and Methods in Applied Sciences, 22(09), 2012.

[4] J. Beck, F. Nobile, L. Tamellini, and R. Tempone. Convergence of quasi-optimal Stochastic Galerkin methods for a class of PDEs with random coefficients. Computers & Mathematics with Applications, 67(4):732 – 751, 2014.

[5] M. Bieri. A sparse composite collocation finite element method for elliptic SPDEs. SIAM Journal on Numerical Analysis, 49(6):2277–2301, 2011.

[6] M. Bieri and C. Schwab. Sparse high order FEM for elliptic sPDEs. Computer Methods in Applied Mechanics and Engineering, 198(1314):1149 – 1170, 2009.

[7] H.J Bungartz and M. Griebel. Sparse grids. Acta Numer., 13:147–269, 2004.

[8] H.J. Bungartz, M. Griebel, D. Röschke, and C. Zenger. Pointwise convergence of the combination technique for the Laplace equation. East-West J. Numer. Math., 2:21–45, 1994.

[9] J. Charrier. Strong and weak error estimates for elliptic partial differential equations with random coefficients. SIAM J. Numer. Anal., 50(1), 2012.

[10] J. Charrier, R. Scheichl, and A. Teckentrup. Finite element error analysis of elliptic pdes with random coefficients and its application to multilevel monte carlo methods. SIAM Journal on Numerical Analysis, 51(1):322–352, 2013.

[11] A. Chkifa. On the lebesgue constant of leja sequences for the complex unit disk and of their real projection. Journal of Approximation Theory, 166(0):176 – 200, 2013.

[12] A. Cohen, R. Devore, and C. Schwab. Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDE'S. Anal. Appl. (Singap.), 9(1):11–47, 2011.

[13] Nathan Collier, Abdul-Lateef Haji-Ali, Fabio Nobile, Erik von Schwerin, and Ral Tempone. A continuation multilevel monte carlo algorithm. BIT Numerical Mathematics, 55(2):399–432, 2015.

[14] G. M. Constantine and T. H. Savits. A multivariate Faà di Bruno formula with applications. Trans. Amer. Math. Soc., 348(2):503–520, 1996.

[15] R. A. DeVore and G. G. Lorentz. Constructive Approximation. Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen. Springer, 1993.

[16] B. Ganapathysubramanian and N. Zabaras. Sparse grid collocation schemes for stochastic natural convection problems. Journal of Computational Physics, 225(1):652–685, 2007.

[17] Michael B. Giles. Multilevel monte carlo path simulation. Operations Research, 56(3):607–617, 2008.

[18] I.G. Graham, R. Scheichl, and E. Ullmann. Mixed finite element analysis of lognormal diffusion and multilevel monte carlo methods. Stochastic Partial Differential Equations: Analysis and Computations, pages 1–35, 2015.

[19] M. Griebel and H. Harbrecht. On the convergence of the combination technique. In Jochen Garcke and Dirk Pflger, editors, Sparse Grids and Applications - Munich 2012, volume 97 of Lecture Notes in Computational Science and Engineering, pages 55–74. Springer International Publishing, 2014.

[20] M. Griebel and S. Knapek. Optimized general sparse grid approximation spaces for operator equations. Math. Comp., 78(268):2223–2257, 2009.

[21] M. Griebel, M. Schneider, and C. Zenger. A combination technique for the solution of sparse grid problems. In P. de Groen and R. Beauwens, editors, Iterative Methods in Linear Algebra, pages 263–281. IMACS, Elsevier, North Holland, 1992.

[22] A.-L. Haji-Ali, F. Nobile, L. Tamellini, and R. Tempone. Multi-index stochastic collocation for random PDEs. MATHICSE report 22/2015, EPFL, 2015. Also available as arXiv:1508.07467.

[23] A.-L. Haji-Ali, F. Nobile, and R. Tempone. Multi-index Monte Carlo: when sparsity meets sampling. Numerische Mathematik, pages 1–40, 2015.

[24] H. Harbrecht, M. Peters, and M. Siebenmorgen. On multilevel quadrature for elliptic stochastic partial differential equations. In Sparse Grids and Applications, volume 88 of Lecture Notes in Computational Science and Engineering, pages 161–179. Springer, 2013.

[25] M. Hegland, J. Garcke, and V. Challis. The combination technique and some generalisations. Linear Algebra and its Applications, 420(23):249 – 275, 2007.

[26] S. Heinrich. Multilevel monte carlo methods. In Large-Scale Scientific Computing, volume 2179 of Lecture Notes in Computer Science, pages 58–67. Springer Berlin Heidelberg, 2001.

[27] F. Y. Kuo, C. Schwab, and I. Sloan. Multi-level Quasi-Monte Carlo Finite Element Methods for a Class of Elliptic PDEs with Random Coefficients. Foundations of Computational Mathematics, 15(2):411–449, 2015.

[28] S. Martello and P. Toth. Knapsack problems: algorithms and computer implementations. Wiley-Interscience series in discrete mathematics and optimization. J. Wiley & Sons, 1990.

[29] Akil Narayan and John D. Jakeman. Adaptive Leja Sparse Grid Constructions for Stochastic Collocation and High-Dimensional Approximation. SIAM Journal on Scientific Computing, 36(6):A2952–A2983, 2014.

[30] F. Nobile, L. Tamellini, and R. Tempone. Comparison of Clenshaw–Curtis and Leja quasi-optimal sparse grids for the approximation of random PDEs. In Spectral and High Order Methods for Partial Differential Equations - ICOSAHOM '14, volume 106 of Lecture Notes in Computational Science and Engineering. Springer, 2015. To appear. Also available as MATHICSE report 41/2014.

[31] F. Nobile, L. Tamellini, and R. Tempone. Convergence of quasi-optimal sparse-grid approximation of Hilbert-space-valued functions: application to random elliptic PDEs. Numerische Mathematik, pages 1–46, 2015.

[32] F. Nobile, R. Tempone, and C.G. Webster. An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data. SIAM J. Numer. Anal., 46(5):2411–2442, 2008.

[33] F. Nobile, R. Tempone, and C.G. Webster. A sparse grid stochastic collocation method for partial differential equations with random input data. SIAM J. Numer. Anal., 46(5):2309–2345, 2008.

[34] C. Schillings and C. Schwab. Sparse, adaptive Smolyak quadratures for Bayesian inverse problems. Inverse Problems, 29(6), 2013.

[35] A. L. Teckentrup, P. Jantsch, C. G. Webster, and M. Gunzburger. A Multilevel Stochastic Collocation Method for Partial Differential Equations with Random Input Data. SIAM/ASA Journal on Uncertainty Quantification, 3(1):1046–1074, 2015.

[36] L. N. Trefethen. Is Gauss quadrature better than Clenshaw-Curtis? SIAM Rev., 50(1):67–87, 2008.

[37] H. W. van Wyk. Multilevel sparse grid methods for elliptic partial differential equations with random coefficients. arXiv arXiv:1404.0963, e-print, 2014.

[38] G.W. Wasilkowski and H. Wozniakowski. Explicit cost bounds of algorithms for multivariate tensor product problems. Journal of Complexity, 11(1):1 – 56, 1995.

[39] D. Xiu and J.S. Hesthaven. High-order collocation methods for differential equations with random inputs. SIAM J. Sci. Comput., 27(3):1118–1139, 2005.

[40] Christoph Zenger. Sparse grids. In Wolfgang Hackbusch, editor, Parallel Algorithms for Partial Differential Equations, volume 31 of Notes on Numerical Fluid Mechanics, pages 241–251. Vieweg, 1991.

Recent publications:

MATHEMATICS INSTITUTE OF COMPUTATIONAL SCIENCE AND ENGINEERING
Section of Mathematics
Ecole Polytechnique Fédérale
CH-1015 Lausanne

16.2015 Simone Deparis, Davide Forti, Alfio Quarteroni:
A fluid-structure interaction algorithm using radial basis function interpolation between non-conforming interfaces

17.2015 Assyr Abdulle, Ondrej Budac:
A reduced basis finite element heterogeneous multiscale method for Stokes flow in porous media

18.2015 Daniel Kressner, Michael Steinlechner, Bart Vandereycken:
Preconditioned low-rank Riemannian optimization for linear systems with tensor product structure

19.2015 Alessandro S. Patelli, Luca Dedè, Toni Lassila, Andrea Bartezzaghi, Alfio Quarteroni:
Isogeometric approximation of cardiac electrophysiology models on surfaces: an accuracy study with application to the human left atrium

20.2015 Matthieu Wilhelm, Luca Dedè, Laura M. Sangalli, Pierre Wilhelm:
IGS: an IsoGeometric approach for Smoothing on surfaces

21.2015 Simone Deparis, Davide Forti, Paola Gervasio, Alfio Quarteroni:
INTERNODES: an accurate interpolation-based method for coupling the Galerkin solutions of PDEs on subdomains featuring non-conforming interfaces

22.2015 Abdul-Lateef Haji-Ali, Fabio Nobile, Lorenzo Tamellini, Raùl Tempone:
Multi-index stochastic collocation for random PDEs

23.2015 Simone Brugiapaglia, Fabio Nobile, Stefano Micheletti, Simona Perotto:
A theoretical study of COmpRessed SolvING for advection-diffusion-reaction problems

24.2015 Ana šušnjara, Nathanaël Perraudin, Daniel Kressner, Pierre Vandergheynst:
Accelerate filtering on graphs using Lanczos mehtod

25.2015 Francesco Ballarin, Elena Faggiano, Sonia Ippolito, Andrea Manzoni, Alfio Quarteroni, Gianluigi Rozza, Roberto Scrofani:
Fast simulation of patient-specific haemodynamics of coronary artery bypass grafts based on a Pod-Galerkin method and a vascular shape parametrization

26.2015 Francisco Macedo:
Benchmark problems on stochastic automata networks in tensor train format

27.2015 Jonas Ballani, Daniel Kressner:
Reduced basis methods: from low-rank matrices to low-rank tensors

28.2015 Albert Cohen, Giovanni Migliorati, Fabio Nobile:
Discrete least-squares approximations over optimized downward closed polynomial spaces in arbitrary dimension

29.2015 Abdul-Lateef Haji-Ali, Fabio Nobile, Lorenzo Tamellini, Raúl Tempone:
Multi-index stochastic collocation convergence rates for random PDEs with parametric regularity