www.oeaw.ac.at

Johann Radon Institute for
Computational and Applied Mathematics
Austrian Academy of Sciences (ÖAW)

# Constructing QMC finite element methods for elliptic PDEs with random coefficients by a reduced CBC construction

A. Ebert, P. Kritzer, D. Nuyens

RICAM-Report 2019-01

# Constructing QMC Finite Element Methods for Elliptic PDEs with Random Coefficients by a Reduced CBC Construction

Adrian Ebert Peter Kritzer* Dirk Nuyens

January 3, 2019

**Abstract**

In the analysis of using quasi-Monte Carlo (QMC) methods to approximate expectations of a linear functional of the solution of an elliptic PDE with random diffusion coefficient the sensitivity w.r.t. the parameters is often stated in terms of product-and-order-dependent (POD) weights. The (offline) fast component-by-component (CBC) construction of an $N$-point QMC method making use of these POD weights leads to a cost of $\mathcal{O}(sN\log(N)+s^2N)$ with $s$ the parameter truncation dimension. When $s$ is large this cost is prohibitive. As an alternative Herrmann and Schwab [8] introduced an analysis resulting in product weights to reduce the construction cost to $\mathcal{O}(sN\log(N))$. We here show how the reduced CBC method can be used for POD weights to reduce the cost to $\mathcal{O}(\sum_{j=1}^{\min\{s,s^*\}}(m-w_j+j)\,b^{m-w_j})$, where $N=b^m$ with prime $b$, $w_1\le\ldots\le w_s$ are nonnegative integers and $s^*$ can be chosen much smaller than $s$ depending on the regularity of the random field expansion as such making it possible to use the POD weights directly. We show a total error estimate for using randomly shifted lattice rules constructed through the reduced CBC construction.

## 1 Introduction and Problem Setting

We consider the parametric elliptic Dirichlet problem given by

$$-\nabla\cdot(a(\boldsymbol{x},\boldsymbol{y})\,\nabla u(\boldsymbol{x},\boldsymbol{y}))=f(\boldsymbol{x})\quad\text{for } \boldsymbol{x}\in D\subset\mathbb{R}^d,\quad u(\boldsymbol{x},\boldsymbol{y})=0\quad\text{for } \boldsymbol{x} \text{ on } \partial D,\tag{1.1}$$

for $D\subset\mathbb{R}^d$ a bounded, convex Lipschitz polyhedron domain with boundary $\partial D$ and fixed spatial dimension $d\in\{1,2,3\}$. The function $f$ lies in $L^2(D)$, the parametric variable $\boldsymbol{y}=(y_j)_{j\ge1}$ belongs to a domain $U$, and the differential operators are understood to be with respect to the physical variable $\boldsymbol{x}\in D$. Here we study the "uniform case", i.e., we assume that $\boldsymbol{y}$ is uniformly distributed on $U:=\left[-\frac{1}{2},\frac{1}{2}\right]^{\mathbb{N}}$ with uniform probability measure $\mu(\mathrm{d}\boldsymbol{y})=\bigotimes_{j\ge1}\mathrm{d}y_j=\mathrm{d}\boldsymbol{y}$. The parametric diffusion coefficient $a(\boldsymbol{x},\boldsymbol{y})$ is assumed to depend linearly on the parameters $y_j$ in the following way,

$$a(\boldsymbol{x},\boldsymbol{y})=a_0(\boldsymbol{x})+\sum_{j\ge1}y_j\,\psi_j(\boldsymbol{x}),\quad \boldsymbol{x}\in D,\quad \boldsymbol{y}\in U.\tag{1.2}$$

*Supported by the Austrian Science Fund (FWF) Project F5506-N26, part of the Special Research Program "Quasi-Monte Carlo Methods: Theory and Applications".

For the variational formulation of (1.1), we consider the Sobolev space $V = H_0^1(D)$ of functions $v$ which vanish on the boundary $\partial D$ with norm

$$\|v\|_V := \left( \int_D \sum_{j=1}^d |\partial_{x_j} v(\boldsymbol{x})|^2 \,\mathrm{d}\boldsymbol{x} \right)^{\frac{1}{2}} = \|\nabla v\|_{L^2(D)}.$$

The corresponding dual space of bounded linear functionals on $V$ with respect to the pivot space $L^2(D)$ is further denoted by $V^* = H^{-1}(D)$. Then, for given $f \in V^*$ and $\boldsymbol{y} \in U$, the weak (or variational) formulation of (1.1) is to find $u(\cdot, \boldsymbol{y}) \in V$ such that

$$A(\boldsymbol{y}; u(\cdot,\boldsymbol{y}), v) = \langle f, v\rangle_{V^* \times V} = \int_D f(\boldsymbol{x}) v(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x} \quad \text{for all} \quad v \in V, \tag{1.3}$$

with parametric bilinear form $A : U \times V \times V \to \mathbb{R}$ given by

$$A(\boldsymbol{y}; w, v) := \int_D a(\boldsymbol{x}, \boldsymbol{y})\, \nabla w(\boldsymbol{x}) \cdot \nabla v(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x} \quad \text{for all} \quad w, v \in V, \tag{1.4}$$

and duality pairing $\langle\cdot,\cdot\rangle_{V^* \times V}$ between $V^*$ and $V$. We will often identify elements $\varphi \in V$ with dual elements $L_\varphi \in V^*$. Indeed, for $\varphi \in V$ and $v \in V$, a bounded linear functional is given via $L_\varphi(v) := \int_D \varphi(x) v(x) dx = \langle \varphi, v\rangle_{L^2(D)}$ and by the Riesz representation theorem there exists a unique representer $\widetilde{\varphi} \in V$ such that $L_\varphi(v) = \langle \widetilde{\varphi}, v\rangle_{L^2(D)}$ for all $v \in V$. Hence, the definition of the canonical duality pairing yields that $\langle L_\varphi, v\rangle_{V^* \times V} = L_\varphi(v) = \langle \varphi, v\rangle_{L^2(D)}$.

Our quantity of interest is the expected value, with respect to $\boldsymbol{y} \in U$, of a given bounded linear functional $G \in V^*$ applied to the solution $u(\cdot, \boldsymbol{y})$ of the PDE. We therefore seek to approximate this expectation by numerically integrating $G$ applied to a finite element approximation $u_h^s(\cdot, \boldsymbol{y})$ of the solution $u^s(\cdot, \boldsymbol{y}) \in H_0^1(D) = V$ of (1.3) with truncated diffusion coefficient $a(\boldsymbol{x}, (\boldsymbol{y}_{\{1:s\}}; 0))$; that is,

$$\mathbb{E}[G(u)] := \int_U G(u(\cdot, \boldsymbol{y}))\, \mu(\,\mathrm{d}\boldsymbol{y}) = \int_U G(u(\cdot, \boldsymbol{y})) \,\mathrm{d}\boldsymbol{y} \approx Q_N(G(u_h^s)), \tag{1.5}$$

where $Q_N(\cdot)$ is a linear quadrature rule using $N$ function evaluations and we write $(\boldsymbol{y}_{\{1:s\}}; 0) = (\tilde{y}_j)_{j \geq 1}$ with $\tilde{y}_j = y_j$ for $j \in \{1 : s\}$ and $\tilde{y}_j = 0$ otherwise. The infinite-dimensional integral $\mathbb{E}[G(u)]$ in (1.5) is defined as

$$\mathbb{E}[G(u)] = \int_U G(u(\cdot, \boldsymbol{y})) \,\mathrm{d}\boldsymbol{y} := \lim_{s \to \infty} \int_{\left[-\frac{1}{2}, \frac{1}{2}\right]^s} G(u(\cdot, (y_1, \ldots, y_s, 0, 0, \ldots))) \,\mathrm{d}y_1 \cdots \mathrm{d}y_s$$

such that our integrands of interest are of the form $F(\boldsymbol{y}) = G(u(\cdot, \boldsymbol{y}))$ with $\boldsymbol{y} \in U$. In this article, we will employ (randomized) QMC methods of the form

$$Q_N(f) = \frac{1}{N} \sum_{k=1}^N F(\boldsymbol{t}_k),$$

i.e., equal-weight quadrature rules with (randomly shifted) deterministic points $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_N \in \left[-\frac{1}{2}, \frac{1}{2}\right]^s$. This elliptic PDE is a standard problem considered in the numerical analysis of computational methods in uncertainty quantification, see, e.g., [1, 2, 5, 7–11].

## 1.1 Existence of solutions of the variational problem

To assure that a unique solution to the weak problem (1.3) exists, we need certain conditions on the diffusion coefficient $a$. We assume $a_0 \in L^\infty(D)$ and $\operatorname*{ess\,inf}_{\boldsymbol{x} \in D} a_0(\boldsymbol{x}) > 0$, which is equivalent to the existence of two constants $0 < a_{0,\min} \le a_{0,\max} < \infty$ such that a.e. on $D$ we have

$$a_{0,\min} \le a_0(\boldsymbol{x}) \le a_{0,\max}, \tag{1.6}$$

and that there exists a $\overline{\kappa} \in (0,1)$ such that

$$\left\| \sum_{j\ge1} \frac{|\psi_j|}{2a_0} \right\|_{L^\infty(D)} \le \overline{\kappa} < 1. \tag{1.7}$$

Via (1.7), we obtain that $|\sum_{j\ge1} y_j\psi_j(\boldsymbol{x})| \le \overline{\kappa}\, a_0(\boldsymbol{x})$ and hence, using (1.6), almost everywhere on $D$ and for any $\boldsymbol{y} \in U$

$$0 < (1-\overline{\kappa})\, a_{0,\min} \le a_0(\boldsymbol{x}) + \sum_{j\ge1} y_j\psi_j(\boldsymbol{x}) = a(\boldsymbol{x},\boldsymbol{y}) \le (1+\overline{\kappa})\, a_{0,\max}. \tag{1.8}$$

These estimates yield the continuity and coercivity of $A(\boldsymbol{y},\cdot,\cdot)$ defined in (1.4) on $V\times V$, uniformly for all $\boldsymbol{y}\in U$. The Lax–Milgram theorem then ensures the existence of a unique solution $u(\cdot,\boldsymbol{y})$ of the weak problem in (1.3).

## 1.2 Parametric regularity

Having established the existence of unique weak parametric solutions $u(\cdot,\boldsymbol{y})$, we investigate their regularity in terms of the behaviour of their mixed first-order derivatives. Our analysis combines multiple techniques which can be found in the literature, see, e.g., [1, 2, 7–9]. In particular we want to point out that our POD form bounds can take advantage of wavelet like expansions of the random field, a technique introduced in [1] and used to the advantage of QMC constructions by [8] to deliver product weights to save on the construction compared to POD weights. Although we end up again with POD weights, we will save on the construction cost by making use of a special construction method, called the reduced CBC construction, which we will introduce in Section 2.4. We introduce some notation. Let $\boldsymbol{\nu} = (\nu_j)_{j\ge1}$ with $\nu_j \in \mathbb{N}_0 := \{0,1,2,\dots\}$ be a sequence of positive integers which we will refer to as a multi-index. We define the order $|\boldsymbol{\nu}|$ and the support $\mathrm{supp}(\boldsymbol{\nu})$ as

$$|\boldsymbol{\nu}| := \sum_{j\ge1} \nu_j \quad \text{and} \quad \mathrm{supp}(\boldsymbol{\nu}) := \{j \ge 1 : \nu_j > 0\}$$

and introduce the sets $\mathcal{F}$ and $\mathcal{F}_1$ of finitely supported multi-indices as

$$\mathcal{F} := \{\boldsymbol{\nu} \in \mathbb{N}_0^{\mathbb{N}} : \mathrm{supp}(\boldsymbol{\nu}) < \infty\} \quad \text{and} \quad \mathcal{F}_1 := \{\boldsymbol{\nu} \in \{0,1\}^{\mathbb{N}} : \mathrm{supp}(\boldsymbol{\nu}) < \infty\},$$

where $\mathcal{F}_1 \subseteq \mathcal{F}$ is the restriction containing only $\boldsymbol{\nu}$ with $\nu_j \in \{0,1\}$. Then, for $\boldsymbol{\nu}\in\mathcal{F}$ denote the $\boldsymbol{\nu}$-th partial derivative with respect to the parametric variables $\boldsymbol{y}\in U$ by

$$\partial^{\boldsymbol{\nu}} = \frac{\partial^{|\boldsymbol{\nu}|}}{\partial y_1^{\nu_1}\partial y_2^{\nu_2}\cdots},$$

and for a sequence $\boldsymbol{b} = (b_j)_{j\ge1} \subset \mathbb{R}^{\mathbb{N}}$, set $\boldsymbol{b}^{\boldsymbol{\nu}} := \prod_{j\ge1} b_j^{\nu_j}$. We further write $\boldsymbol{\omega} \le \boldsymbol{\nu}$ if $\omega_j \le \nu_j$ for all $j \ge 1$ and denote by $\boldsymbol{e}_i \in \mathcal{F}_1$ the multi-index with components $e_j = \delta_{i,j}$. For a fixed $\boldsymbol{y}\in U$, we introduce the energy norm $\|\cdot\|_{a_{\boldsymbol{y}}}^2$ in the space $V$ via

$$\|v\|_{a_{\boldsymbol{y}}}^2 := \int_D a(\boldsymbol{x},\boldsymbol{y})\, |\nabla v(\boldsymbol{x})|^2 \,\mathrm{d}\boldsymbol{x}$$

for which it holds true by (1.8) that

$$(1-\overline{\kappa})\, a_{0,\min} \|v\|_V^2 \le \|v\|_{a_{\boldsymbol{y}}}^2 \quad \text{for all} \quad v \in V. \tag{1.9}$$

Consequently, we have that $(1-\overline{\kappa})\, a_{0,\min}\|u(\cdot,\boldsymbol{y})\|_V^2 \le \|u(\cdot,\boldsymbol{y})\|_{a_{\boldsymbol{y}}}^2$ and hence applying the Cauchy–Schwarz inequality yields the following initial estimate,

$$
\begin{aligned}
\|u(\cdot,\boldsymbol{y})\|_{a_{\boldsymbol{y}}}^2 &= \int_D a(\boldsymbol{x},\boldsymbol{y})\,|\nabla u(\boldsymbol{x},\boldsymbol{y})|^2 \,\mathrm{d}\boldsymbol{x} = \int_D f(\boldsymbol{x})u(\boldsymbol{x},\boldsymbol{y})\,\mathrm{d}\boldsymbol{x} \\
&\le \|f\|_{V^*}\|u(\cdot,\boldsymbol{y})\|_V \le \frac{\|f\|_{V^*}\|u(\cdot,\boldsymbol{y})\|_{a_{\boldsymbol{y}}}}{\sqrt{(1-\overline{\kappa})a_{0,\min}}}
\end{aligned}
$$

which gives in turn

$$
\|u(\cdot,\boldsymbol{y})\|_{a_{\boldsymbol{y}}}^2 \le \frac{\|f\|_{V^*}^2}{(1-\overline{\kappa})\,a_{0,\min}}. \tag{1.10}
$$

In order to exploit the decay of the norm sequence $(\|\psi_j\|_{L^\infty(D)})_{j\ge1}$ of the basis functions, we extend condition (1.7) as follows. To characterize the smoothness of the random field, we assume that there exist a sequence of reals $\boldsymbol{b}=(b_j)_{j\ge1}$ with $0<b_j\le1$ for all $j$, a constant $\kappa\in(0,1)$ and therefore also constants $\widetilde{\kappa}(\boldsymbol{\nu})\le\kappa$ for all $\boldsymbol{\nu}\in\mathcal{F}_1$ such that

$$
\kappa := \left\|\sum_{j\ge1}\frac{|\psi_j|/b_j}{2a_0}\right\|_{L^\infty(D)} < 1, \qquad \widetilde{\kappa}(\boldsymbol{\nu}) = \left\|\sum_{j\in\mathrm{supp}(\boldsymbol{\nu})}\frac{|\psi_j|/b_j}{2a_0}\right\|_{L^\infty(D)}. \tag{1.11}
$$

We remark that condition (1.7) is included in this assumption by letting $b_j=1$ for all $j\ge1$ and that $0<\overline{\kappa}\le\kappa<1$. Using the above estimations we can derive the following theorem for the mixed first-order partial derivatives.

**Theorem 1.** *Let $\boldsymbol{\nu}\in\mathcal{F}_1$ be a multi-index of finite support and let $k\in\{0,1,\dots,|\boldsymbol{\nu}|\}$. Then, for every $f\in V^*$ and every $\boldsymbol{y}\in U$,*

$$
\sum_{\substack{\boldsymbol{\omega}\le\boldsymbol{\nu}\\|\boldsymbol{\omega}|=k}} \boldsymbol{b}^{-2\boldsymbol{\omega}}\|\partial^{\boldsymbol{\omega}}u(\cdot,\boldsymbol{y})\|_V^2 \le \left(\left(\frac{2\,\widetilde{\kappa}(\boldsymbol{\nu})}{1-\overline{\kappa}}\right)^k \frac{\|f\|_{V^*}}{(1-\overline{\kappa})\,a_{0,\min}}\right)^2,
$$

*with $\widetilde{\kappa}(\boldsymbol{\nu})$ as in* (1.11)*. Moreover, for $k=|\boldsymbol{\nu}|$ we obtain*

$$
\|\partial^{\boldsymbol{\nu}}u(\cdot,\boldsymbol{y})\|_V \le \boldsymbol{b}^{\boldsymbol{\nu}}\left(\frac{2\widetilde{\kappa}(\boldsymbol{\nu})}{1-\overline{\kappa}}\right)^{|\boldsymbol{\nu}|}\frac{\|f\|_{V^*}}{(1-\overline{\kappa})\,a_{0,\min}}.
$$

*Proof.* For the special case $\boldsymbol{\nu}=\boldsymbol{0}$, the claim follows by combining (1.9) and (1.10). For $\boldsymbol{\nu}\in\mathcal{F}_1$ with $|\boldsymbol{\nu}|>0$, as is known from, e.g., [2] and [10, Appendix], the linearity of $a(\boldsymbol{x},\boldsymbol{y})$ gives rise to the following identity for any $\boldsymbol{y}\in U$:

$$
\|\partial^{\boldsymbol{\nu}}u(\cdot,\boldsymbol{y})\|_{a_{\boldsymbol{y}}}^2 = -\sum_{j\in\mathrm{supp}(\boldsymbol{\nu})}\int_D \psi_j(\boldsymbol{x})\,\nabla\partial^{\boldsymbol{\nu}-\boldsymbol{e}_j}u(\boldsymbol{x},\boldsymbol{y})\cdot\nabla\partial^{\boldsymbol{\nu}}u(\boldsymbol{x},\boldsymbol{y})\,\mathrm{d}\boldsymbol{x}. \tag{1.12}
$$

For sequences of $L^2(D)$-integrable functions $\boldsymbol{f}=(f_{\boldsymbol{\omega},j})_{\boldsymbol{\omega}\in\mathcal{F},j\ge1}$ with $f_{\boldsymbol{\omega},j}:D\to\mathbb{R}$, we define the inner product $\langle\boldsymbol{f},\boldsymbol{g}\rangle_{\boldsymbol{\nu},k}$ as follows,

$$
\langle\boldsymbol{f},\boldsymbol{g}\rangle_{\boldsymbol{\nu},k} := \sum_{\substack{\boldsymbol{\omega}\le\boldsymbol{\nu}\\|\boldsymbol{\omega}|=k}}\int_D\sum_{j\in\mathrm{supp}(\boldsymbol{\omega})} f_{\boldsymbol{\omega},j}(\boldsymbol{x})\,g_{\boldsymbol{\omega},j}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}.
$$

We can then apply the Cauchy–Schwarz inequality to $\boldsymbol{f}=(f_{\boldsymbol{\omega},j})$ and $\boldsymbol{g}=(g_{\boldsymbol{\omega},j})$ with $f_{\boldsymbol{\omega},j}=\boldsymbol{b}^{-\boldsymbol{e}_j/2}|\psi_j|^{\frac12}\,\boldsymbol{b}^{-(\boldsymbol{\omega}-\boldsymbol{e}_j)}\nabla\partial^{\boldsymbol{\omega}-\boldsymbol{e}_j}u(\cdot,\boldsymbol{y})$ and $g_{\boldsymbol{\omega},j}=\boldsymbol{b}^{-\boldsymbol{e}_j/2}|\psi_j|^{\frac12}\,\boldsymbol{b}^{-\boldsymbol{\omega}}\nabla\partial^{\boldsymbol{\omega}}u(\cdot,\boldsymbol{y})$ to obtain, with the help of (1.12),

$$
\sum_{\substack{\boldsymbol{\omega}\le\boldsymbol{\nu}\\|\boldsymbol{\omega}|=k}} \boldsymbol{b}^{-2\boldsymbol{\omega}}\|\partial^{\boldsymbol{\omega}}u(\cdot,\boldsymbol{y})\|_{a_{\boldsymbol{y}}}^2
$$

$$= -\sum_{\substack{\boldsymbol{\omega}\le\boldsymbol{\nu}\\|\boldsymbol{\omega}|=k}} \int_D \sum_{j\in\mathrm{supp}(\boldsymbol{\omega})} \boldsymbol{b}^{-\boldsymbol{e}_j}\boldsymbol{b}^{-(\boldsymbol{\omega}-\boldsymbol{e}_j)}\boldsymbol{b}^{-\boldsymbol{\omega}}\psi_j(\boldsymbol{x})\,\nabla\partial^{\boldsymbol{\omega}-\boldsymbol{e}_j}u(\boldsymbol{x},\boldsymbol{y})\cdot\nabla\partial^{\boldsymbol{\omega}}u(\boldsymbol{x},\boldsymbol{y})\;\mathrm{d}\boldsymbol{x}$$

$$\le \left(\int_D \sum_{\substack{\boldsymbol{\omega}\le\boldsymbol{\nu}\\|\boldsymbol{\omega}|=k}} \sum_{j\in\mathrm{supp}(\boldsymbol{\omega})} \boldsymbol{b}^{-\boldsymbol{e}_j}|\psi_j(\boldsymbol{x})|\,\left|\boldsymbol{b}^{-(\boldsymbol{\omega}-\boldsymbol{e}_j)}\nabla\partial^{\boldsymbol{\omega}-\boldsymbol{e}_j}u(\boldsymbol{x},\boldsymbol{y})\right|^2\;\mathrm{d}\boldsymbol{x}\right)^{\frac12}$$

$$\times\left(\int_D \sum_{\substack{\boldsymbol{\omega}\le\boldsymbol{\nu}\\|\boldsymbol{\omega}|=k}} \sum_{j\in\mathrm{supp}(\boldsymbol{\omega})} \boldsymbol{b}^{-\boldsymbol{e}_j}|\psi_j(\boldsymbol{x})|\,\left|\boldsymbol{b}^{-\boldsymbol{\omega}}\nabla\partial^{\boldsymbol{\omega}}u(\boldsymbol{x},\boldsymbol{y})\right|^2\;\mathrm{d}\boldsymbol{x}\right)^{\frac12}.$$

The first of the two factors above is then bounded as follows,

$$\int_D \sum_{\substack{\boldsymbol{\omega}\le\boldsymbol{\nu}\\|\boldsymbol{\omega}|=k}} \sum_{j\in\mathrm{supp}(\boldsymbol{\omega})} \boldsymbol{b}^{-\boldsymbol{e}_j}|\psi_j(\boldsymbol{x})|\,\left|\boldsymbol{b}^{-(\boldsymbol{\omega}-\boldsymbol{e}_j)}\nabla\partial^{\boldsymbol{\omega}-\boldsymbol{e}_j}u(\boldsymbol{x},\boldsymbol{y})\right|^2\;\mathrm{d}\boldsymbol{x}$$

$$= \int_D \sum_{\substack{\boldsymbol{\omega}\le\boldsymbol{\nu}\\|\boldsymbol{\omega}|=k-1}} \left(\sum_{\substack{j\in\mathrm{supp}(\boldsymbol{\nu})\\\boldsymbol{\omega}+\boldsymbol{e}_j\le\boldsymbol{\nu}}} \boldsymbol{b}^{-\boldsymbol{e}_j}|\psi_j(\boldsymbol{x})|\right)\left|\boldsymbol{b}^{-\boldsymbol{\omega}}\nabla\partial^{\boldsymbol{\omega}}u(\boldsymbol{x},\boldsymbol{y})\right|^2\;\mathrm{d}\boldsymbol{x}$$

$$\le \left\|\sum_{j\in\mathrm{supp}(\boldsymbol{\nu})}\frac{|\psi_j|/b_j}{a(\cdot,\boldsymbol{y})}\right\|_{L^\infty(D)} \sum_{\substack{\boldsymbol{\omega}\le\boldsymbol{\nu}\\|\boldsymbol{\omega}|=k-1}} \boldsymbol{b}^{-2\boldsymbol{\omega}}\int_D a(\boldsymbol{x},\boldsymbol{y})\,|\nabla\partial^{\boldsymbol{\omega}}u(\boldsymbol{x},\boldsymbol{y})|^2\;\mathrm{d}\boldsymbol{x}$$

$$= \left\|\sum_{j\in\mathrm{supp}(\boldsymbol{\nu})}\frac{|\psi_j|/b_j}{a(\cdot,\boldsymbol{y})}\right\|_{L^\infty(D)} \sum_{\substack{\boldsymbol{\omega}\le\boldsymbol{\nu}\\|\boldsymbol{\omega}|=k-1}} \boldsymbol{b}^{-2\boldsymbol{\omega}}\|\partial^{\boldsymbol{\omega}}u(\cdot,\boldsymbol{y})\|_{a_{\boldsymbol{y}}}^2,$$

while the other factor can be bounded trivially. Furthermore, using (1.8), we have for any $\boldsymbol{y}\in U$

$$\left\|\sum_{j\in\mathrm{supp}(\boldsymbol{\nu})}\frac{|\psi_j|/b_j}{a(\cdot,\boldsymbol{y})}\right\|_{L^\infty(D)} \le \frac{1}{1-\overline{\kappa}}\left\|\sum_{j\in\mathrm{supp}(\boldsymbol{\nu})}\frac{|\psi_j|/b_j}{a_0}\right\|_{L^\infty(D)} := \frac{2\widetilde{\kappa}(\boldsymbol{\nu})}{1-\overline{\kappa}},$$

so that, combining these three estimates, we obtain

$$\sum_{\substack{\boldsymbol{\omega}\le\boldsymbol{\nu}\\|\boldsymbol{\omega}|=k}} \boldsymbol{b}^{-2\boldsymbol{\omega}}\|\partial^{\boldsymbol{\omega}}u(\cdot,\boldsymbol{y})\|_{a_{\boldsymbol{y}}}^2$$

$$\le \frac{2\widetilde{\kappa}(\boldsymbol{\nu})}{1-\overline{\kappa}}\left(\sum_{\substack{\boldsymbol{\omega}\le\boldsymbol{\nu}\\|\boldsymbol{\omega}|=k-1}} \boldsymbol{b}^{-2\boldsymbol{\omega}}\|\partial^{\boldsymbol{\omega}}u(\cdot,\boldsymbol{y})\|_{a_{\boldsymbol{y}}}^2\right)^{\frac12}\left(\sum_{\substack{\boldsymbol{\omega}\le\boldsymbol{\nu}\\|\boldsymbol{\omega}|=k}} \boldsymbol{b}^{-2\boldsymbol{\omega}}\|\partial^{\boldsymbol{\omega}}u(\cdot,\boldsymbol{y})\|_{a_{\boldsymbol{y}}}^2\right)^{\frac12}.$$

Therefore, we finally obtain that

$$\sum_{\substack{\boldsymbol{\omega}\le\boldsymbol{\nu}\\|\boldsymbol{\omega}|=k}} \boldsymbol{b}^{-2\boldsymbol{\omega}}\|\partial^{\boldsymbol{\omega}}u(\cdot,\boldsymbol{y})\|_{a_{\boldsymbol{y}}}^2 \le \left(\frac{2\widetilde{\kappa}(\boldsymbol{\nu})}{1-\overline{\kappa}}\right)^2 \sum_{\substack{\boldsymbol{\omega}\le\boldsymbol{\nu}\\|\boldsymbol{\omega}|=k-1}} \boldsymbol{b}^{-2\boldsymbol{\omega}}\|\partial^{\boldsymbol{\omega}}u(\cdot,\boldsymbol{y})\|_{a_{\boldsymbol{y}}}^2$$

which inductively gives

$$\sum_{\substack{\boldsymbol{\omega}\le\boldsymbol{\nu}\\|\boldsymbol{\omega}|=k}} \boldsymbol{b}^{-2\boldsymbol{\omega}}\|\partial^{\boldsymbol{\omega}}u(\cdot,\boldsymbol{y})\|_{a_{\boldsymbol{y}}}^2 \le \left(\frac{2\widetilde{\kappa}(\boldsymbol{\nu})}{1-\overline{\kappa}}\right)^{2k}\|u(\cdot,\boldsymbol{y})\|_{a_{\boldsymbol{y}}}^2 \le \left(\frac{2\widetilde{\kappa}(\boldsymbol{\nu})}{1-\overline{\kappa}}\right)^{2k}\frac{\|f\|_{V^*}^2}{(1-\overline{\kappa})\,a_{0,\min}},$$

where the last inequality follows from the initial estimate (1.10). The estimate (1.9) then gives

$$\begin{aligned}\sum_{\substack{\boldsymbol{\omega}\le\boldsymbol{\nu}\\|\boldsymbol{\omega}|=k}} \boldsymbol{b}^{2\boldsymbol{\omega}}\|\partial^{\boldsymbol{\omega}}u(\cdot,\boldsymbol{y})\|_V^2 &\le \frac{1}{(1-\overline{\kappa})\,a_{0,\min}}\sum_{\substack{\boldsymbol{\omega}\le\boldsymbol{\nu}\\|\boldsymbol{\omega}|=k}} \boldsymbol{b}^{-2\boldsymbol{\omega}}\|\partial^{\boldsymbol{\omega}}u(\cdot,\boldsymbol{y})\|_{a_{\boldsymbol{y}}}^2\\ &\le \left(\frac{2\widetilde{\kappa}(\boldsymbol{\nu})}{1-\overline{\kappa}}\right)^{2k}\frac{\|f\|_{V^*}^2}{(1-\overline{\kappa})^2\,a_{0,\min}^2},\end{aligned}$$

which yields the first claim. The second claim follows since the sum over the $\boldsymbol{\omega}\le\boldsymbol{\nu}$ with $|\boldsymbol{\omega}|=|\boldsymbol{\nu}|$ and $\boldsymbol{\nu}\in\mathcal{F}_1$ consists only of the term corresponding to $\boldsymbol{\omega}=\boldsymbol{\nu}$. □

**Corollary 1.** *Under the assumptions of Theorem 1, there exists a number $\kappa(k)$ for each $k\in\mathbb{N}$, given by*

$$\kappa(k) := \sup_{\substack{\boldsymbol{\nu}\in\mathcal{F}_1\\|\boldsymbol{\nu}|=k}}\widetilde{\kappa}(\boldsymbol{\nu}),$$

*such that $\widetilde{\kappa}(\boldsymbol{\nu})\le\kappa(k)\le\kappa<1$ for all $\boldsymbol{\nu}\in\mathcal{F}_1$ with $|\boldsymbol{\nu}|=k$. Then for $\boldsymbol{\nu}\in\mathcal{F}_1$, every $f\in V^*$, and every $\boldsymbol{y}\in U$, the solution $u(\cdot,\boldsymbol{y})$ satisfies*

$$\|\partial^{\boldsymbol{\nu}}u(\cdot,\boldsymbol{y})\|_V \le \boldsymbol{b}^{\boldsymbol{\nu}}\left(\frac{2\kappa(|\boldsymbol{\nu}|)}{1-\overline{\kappa}}\right)^{|\boldsymbol{\nu}|}\frac{\|f\|_{V^*}}{(1-\overline{\kappa})\,a_{0,\min}}. \tag{1.13}$$

Note that since $0<\overline{\kappa}\le\kappa<1$, the results of Theorem 1 and Corollary 1 remain also valid for $\overline{\kappa}$ replaced by $\kappa$.

The obtained bounds on the mixed first-order derivatives turn out to be of product and order-dependent (so-called POD) form; that is, they are of the general form

$$\|\partial^{\boldsymbol{\nu}}u(\cdot,\boldsymbol{y})\|_V \le C\,\boldsymbol{b}^{\boldsymbol{\nu}}\,\Gamma(|\boldsymbol{\nu}|)\,\|f\|_{V^*} \tag{1.14}$$

with a map $\Gamma:\mathbb{N}_0\to\mathbb{R}$, a sequence of reals $\boldsymbol{b}=(b_j)_{j\ge1}\in\mathbb{R}^{\mathbb{N}}$ and some constant $C\in\mathbb{R}_+$. This finding motivates us to consider this special type of bounds in the following error analysis.

# 2 Quasi-Monte Carlo finite element error

We analyze the error $\mathbb{E}[G(u)]-Q_N(G(u_h^s))$ obtained by applying QMC rules to the finite element approximation $u_h^s$ to approximate the expected value

$$\mathbb{E}[G(u)] = \int_U G(u(\cdot,\boldsymbol{y}))\,\mathrm{d}\boldsymbol{y}.$$

To this end, we introduce the finite element approximation $u_h^s(\boldsymbol{x},\boldsymbol{y}) := u_h(\boldsymbol{x},(\boldsymbol{y}_{\{1:s\}};0))$ of a solution of (1.3) with truncated diffusion coefficient $a(\boldsymbol{x},(\boldsymbol{y}_{\{1:s\}};0))$, where $u_h$ is a finite element approximation as defined in (2.2) and $(\boldsymbol{y}_{\{1:s\}};0)=(y_1,\ldots,y_s,0,0,\ldots)$. The overall absolute QMC finite element error is then bounded as follows

$$\begin{aligned}&|\mathbb{E}[G(u)]-Q_N(G(u_h^s))|\\ &\quad= |\mathbb{E}[G(u)]-\mathbb{E}[G(u^s)]+\mathbb{E}[G(u^s)]-\mathbb{E}[G(u_h^s)]+\mathbb{E}[G(u_h^s)]-Q_N(G(u_h^s))|\end{aligned}$$

$$\leq |\mathbb{E}[G(u - u^s)]| + |\mathbb{E}[G(u^s - u_h^s)]| + |\mathbb{E}[G(u_h^s)] - Q_N(G(u_h^s))|. \tag{2.1}$$

The first term on the right hand side of (2.1) will be referred to as (dimension) truncation error, the second term is the finite element discretization error and the last term is the QMC quadrature error for the integrand $u_h^s$. In the following sections we will analyze these different error terms separately.

## 2.1 Finite Element Approximation

Here, we consider the approximation of the solution $u(\cdot, \boldsymbol{y})$ of (1.3) by a finite element approximation $u_h(\cdot, \boldsymbol{y})$ and assess the finite element discretization error. More specifically, denote by $\{V_h\}_{h>0}$ a family of subspaces $V_h \subset V$ of finite dimension $M_h$ such that $V_h \to V$ as $h \to 0$. We define the parametric finite element (FE) approximation as follows: for $f \in V^*$ and given $\boldsymbol{y} \in U$, find $u_h(\cdot, \boldsymbol{y}) \in V_h$ such that

$$A(\boldsymbol{y}; u_h(\cdot, \boldsymbol{y}), v_h) = \langle f, v_h \rangle_{V^* \times V} = \int_D f(\boldsymbol{x}) v_h(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \quad \text{for all} \quad v_h \in V_h. \tag{2.2}$$

To establish convergence of the finite element approximations, we need some further conditions on $a(\boldsymbol{x}, \boldsymbol{y})$. To this end, we define the space $W^{1,\infty}(D) \subseteq L^\infty(D)$ endowed with the norm $\|v\|_{W^{1,\infty}(D)} = \max\{\|v\|_{L^\infty(D)}, \|\nabla v\|_{L^\infty(D)}\}$ and require that

$$a_0 \in W^{1,\infty}(D) \quad \text{and} \quad \sum_{j \geq 1} \|\psi_j\|_{W^{1,\infty}(D)} < \infty. \tag{2.3}$$

Under these conditions and using that $f \in L^2(D)$, it was proven in [11, Theorems 7.1 and 7.2] that for any $\boldsymbol{y} \in U$ the approximations $u_h(\cdot, \boldsymbol{y})$ satisfy

$$\|u(\cdot, \boldsymbol{y}) - u_h(\cdot, \boldsymbol{y})\|_V \leq C_1 \, h \, \|f\|_{L^2}.$$

In addition, if (the representer of) the bounded linear functional $G \in V^*$ lies in $L^2(D)$ we have for any $\boldsymbol{y} \in U$, as $h \to 0$,

$$\begin{aligned} |G(u(\cdot, \boldsymbol{y})) - G(u_h(\cdot, \boldsymbol{y}))| &\leq C_2 \, h^2 \, \|f\|_{L^2} \, \|G\|_{L^2}, \\ |\mathbb{E}[G(u(\cdot, \boldsymbol{y}) - u_h(\cdot, \boldsymbol{y}))]| &\leq C_3 \, h^2 \, \|f\|_{L^2} \, \|G\|_{L^2}, \end{aligned} \tag{2.4}$$

where the constants $C_1, C_2, C_3 > 0$ are independent of $h$ and $\boldsymbol{y}$. Since the above statements hold true for any $\boldsymbol{y} \in U$, they remain also valid for $u^s(\boldsymbol{x}, \boldsymbol{y}) := u(\boldsymbol{x}, (\boldsymbol{y}_{\{1:s\}}; 0))$ and $u_h^s(\boldsymbol{x}, \boldsymbol{y}) := u_h(\boldsymbol{x}, (\boldsymbol{y}_{\{1:s\}}; 0))$.

## 2.2 Dimension Truncation

For every $s \in \mathbb{N}$ and $\boldsymbol{y} \in U$, we formally define the solution of the parametric weak problem (1.3) corresponding to the diffusion coefficient $a(\boldsymbol{x}, (\boldsymbol{y}_{\{1:s\}}; 0))$ with sum truncated to $s$ terms as

$$u^s(\cdot, \boldsymbol{y}) := u(\cdot, (\boldsymbol{y}_{\{1:s\}}; 0)). \tag{2.5}$$

In [7, Proposition 5.1] it was shown that for the solution $u^s$ the following error estimates are satisfied.

**Theorem 2.** *Let $\overline{\kappa} \in (0, 1)$ be such that (1.7) is satisfied and assume furthermore that there exists a sequence of reals $\boldsymbol{b} = (b_j)_{j \geq 1}$ with $0 < b_j \leq 1$ for all $j$ and a constant $\kappa \in [\overline{\kappa}, 1)$ as defined in (1.11). Then, for every $\boldsymbol{y} \in U$ and each $s \in \mathbb{N}$*

$$\|u(\cdot, \boldsymbol{y}) - u^s(\cdot, \boldsymbol{y})\|_V \leq \frac{a_{0,\max} \, \|f\|_{V^*}}{(a_{0,\min}(1 - \overline{\kappa}))^2} \sup_{j \geq s+1} b_j.$$

*Moreover, if it holds for $\kappa$ that $\frac{\kappa\, a_{0,\max}}{(1-\overline{\kappa})\, a_{0,\min}} \sup_{j\geq s+1} b_j < 1$, then for every $G \in V^*$ we have*

$$\left| \mathbb{E}[G(u)] - \int_{[-\frac{1}{2},\frac{1}{2}]^s} G(u^s(\cdot,(\boldsymbol{y}_{\{1:s\}};0)))\, \mathrm{d}\boldsymbol{y}_{\{1:s\}} \right| \\ \leq \frac{\|G\|_{V^*}\, \|f\|_{V^*}}{(1-\overline{\kappa})\, a_{0,\min} - a_{0,\max}\, \kappa \sup_{j\geq s+1} b_j} \left( \frac{a_{0,\max}}{(1-\overline{\kappa})\, a_{0,\min}} \kappa \sup_{j\geq s+1} b_j \right)^2 . \tag{2.6}$$

In the following subsection, we will discuss how to approximate the finite-dimensional integral of solutions of the form (2.5) by means of QMC methods.

## 2.3 Quasi-Monte Carlo Integration

For a real-valued function $F : [-\frac{1}{2},\frac{1}{2}]^s \to \mathbb{R}$ defined over the $s$-dimensional unit cube centered at the origin, we consider the approximation of the integral $I_s(F)$ by $N$-point QMC rules $Q_N(F)$, i.e.,

$$I_s(F) := \int_{[-\frac{1}{2},\frac{1}{2}]^s} F(\boldsymbol{y})\, \mathrm{d}\boldsymbol{y} \approx \frac{1}{N} \sum_{k=1}^{N} F(\boldsymbol{t}_k) =: Q_N(F),$$

with quadrature points $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_N \in [-\frac{1}{2},\frac{1}{2}]^s$. As a quality criterion of such a rule, we define the worst-case error for QMC integration in some Banach space $\mathcal{H}$ as

$$e^{\mathrm{wor}}(\boldsymbol{t}_1, \ldots, \boldsymbol{t}_N) := \sup_{\substack{F\in\mathcal{H} \\ \|F\|_{\mathcal{H}}\leq 1}} |I_s(F) - Q_N(F)|.$$

In this article, we consider randomly shifted rank-1 lattice rules as randomized QMC rules, with underlying points of the form

$$\widetilde{\boldsymbol{t}}_k(\boldsymbol{\Delta}) = \{(k\boldsymbol{z})/N + \boldsymbol{\Delta}\} - (1/2, \ldots, 1/2)\,, \quad k = 1, \ldots, N,$$

with generating vector $\boldsymbol{z} \in \mathbb{Z}^s$, uniform random shift $\boldsymbol{\Delta} \in [0,1]^s$ and component-wise applied fractional part, denoted by $\{\boldsymbol{x}\}$. For simplicity, we denote the worst-case error using a shifted lattice rule with generating vector $\boldsymbol{z}$ and shift $\boldsymbol{\Delta}$ by $e_{N,s}(\boldsymbol{z}, \boldsymbol{\Delta})$.

For randomly shifted QMC rules, the probabilistic error bound

$$\sqrt{\mathbb{E}_{\boldsymbol{\Delta}}\left[|I_s(F) - Q_N(F)|^2\right]} \leq \widehat{e}_{N,s}(\boldsymbol{z})\, \|F\|_{\mathcal{H}},$$

holds for all $F \in \mathcal{H}$, with shift-averaged worst-case error

$$\widehat{e}_{N,s}(\boldsymbol{z}) := \left( \int_{[0,1]^s} e^2_{N,s}(\boldsymbol{z}, \boldsymbol{\Delta})\, \mathrm{d}\boldsymbol{\Delta} \right)^{1/2} .$$

As function space $\mathcal{H}$ for our integrands $F$, we consider the weighted, unanchored Sobolev space $\mathcal{W}_{s,\boldsymbol{\gamma}}$, which is a Hilbert space of functions defined over $[-\frac{1}{2},\frac{1}{2}]^s$ with square integrable mixed first derivatives and general non-negative weights $\boldsymbol{\gamma} = (\gamma_{\mathfrak{u}})_{\mathfrak{u}\subseteq\{1:s\}}$. More precisely, the norm for $F \in \mathcal{W}_{s,\boldsymbol{\gamma}}$ is given by

$$\|F\|_{\mathcal{W}_{s,\boldsymbol{\gamma}}} := \left( \sum_{\mathfrak{u}\subseteq\{1:s\}} \gamma_{\mathfrak{u}}^{-1} \int_{[-\frac{1}{2},\frac{1}{2}]^{|\mathfrak{u}|}} \left( \int_{[-\frac{1}{2},\frac{1}{2}]^{s-|\mathfrak{u}|}} \frac{\partial^{|\mathfrak{u}|} F}{\partial \boldsymbol{y}_{\mathfrak{u}}} (\boldsymbol{y}_{\mathfrak{u}}; \boldsymbol{y}_{-\mathfrak{u}})\, \mathrm{d}\boldsymbol{y}_{-\mathfrak{u}} \right)^2 \mathrm{d}\boldsymbol{y}_{\mathfrak{u}} \right)^{1/2} , \tag{2.7}$$

where $\{1:s\} := \{1, \ldots, s\}$, $\frac{\partial^{|\mathfrak{u}|} F}{\partial \boldsymbol{y}_{\mathfrak{u}}}$ denotes the mixed first derivative with respect to the variables $\boldsymbol{y}_{\mathfrak{u}} = (y_j)_{j\in\mathfrak{u}}$ and we set $\boldsymbol{y}_{-\mathfrak{u}} = (y_j)_{j\in\{1:s\}\setminus\mathfrak{u}}$.

For the efficient construction of good lattice rule generating vectors $\boldsymbol{z}$, we consider the so-called reduced component-by-component (CBC) construction studied in [4]. For $m \in \mathbb{N}$, we define the group of units of integers modulo $b^m$ via

$$\mathbb{Z}_{b^m}^{\times} := \{z \in \{1, 2, \ldots, b^m - 1\} : \gcd(z, b^m) = 1\},$$

and note that $\mathbb{Z}_{b^0}^{\times} = \mathbb{Z}_1^{\times} = \{0\}$ since $\gcd(0,1) = 1$. Furthermore, we let $Y_j := b^{w_j}$ for $j \in \{1 : s\}$ and recall that $|\mathbb{Z}_{b^m}^{\times}| = \varphi(b^m) = b^{m-1}(b-1)$, where $\varphi$ denotes Euler's totient function. Let $N$ be a prime power, i.e., $N = b^m$, where $b$ is a prime number and $m \in \mathbb{N}$; let $\boldsymbol{w} := (w_j)_{j \geq 1}$ be a non-decreasing sequence of integers in $\mathbb{N}_0$, the elements of which we will refer to as reduction indices.

In [4], the reduced CBC construction was introduced to construct rank-1 lattice rules for 1-periodic functions in a weighted Korobov space $\mathcal{H}(K_{s,\alpha,\boldsymbol{\gamma}})$ of smoothness $\alpha$ (see, e.g., [13]). We denote the worst-case error in $\mathcal{H}(K_{s,\alpha,\boldsymbol{\gamma}})$ using a rank-1 lattice rule with generator $\boldsymbol{z}$ by $e_{N,s}(\boldsymbol{z})$. Following [4], the reduced CBC construction is then given in Algorithm 1.

**Algorithm 1** The reduced component-by-component construction

Let $N$, $w_1 \leq \cdots \leq w_s$, and $Y_1, \ldots, Y_s$ be as above. Construct $\boldsymbol{z} = (Y_1 z_1, \ldots, Y_s z_s)$ as follows.

For $j$ from 1 to $s$ and as long as $w_j < m$ do:

- Calculate $e_{N,j}^2((Y_1 z_1, \ldots, Y_{j-1} z_{j-1}, Y_j\, z))$ for all $z \in \mathbb{Z}_{b^{m-w_j}}^{\times}$.
- Select $z_j \in \mathbb{Z}_{b^{m-w_j}}^{\times}$ such that
$$z_j = \operatorname*{argmin}_{z \in \mathbb{Z}_{b^{m-w_j}}^{\times}} e_{N,j}^2((Y_1 z_1, \ldots, Y_{j-1} z_{j-1}, Y_j\, z)).$$

For $j \in \{1, \ldots, s\}$ with $w_j \geq m$: Set $z_j = 0$.

Then the following theorem, proven in [4], which states that the algorithm yields generating vectors $\boldsymbol{z}$ with a small integration error, holds true for general weights $\gamma_{\mathfrak{u}}$.

**Theorem 3.** *Let $\boldsymbol{z} = (Y_1 z_1, \ldots, Y_s z_s) \in \{1, \ldots, N-1\}^s$ be constructed according to Algorithm 1. Then for every $d \in \{1 : s\}$ it is true that, for $\lambda \in (1/\alpha, 1]$,*

$$e_{N,d}^2((Y_1 z_1, \ldots, Y_d z_d)) \leq \left( \sum_{\emptyset \neq \mathfrak{u} \subseteq \{1:d\}} \gamma_{\mathfrak{u}}^{\lambda} \frac{2(2\zeta(\alpha\lambda))^{|\mathfrak{u}|}}{b^{\max\{0, m - \max_{j \in \mathfrak{u}} w_j\}}} \right)^{\frac{1}{\lambda}}.$$

This theorem can be extended to the weighted unanchored Sobolev space $\mathcal{W}_{s,\boldsymbol{\gamma}}$ using randomly shifted lattice rules as follows.

**Theorem 4.** *Let $N = b^m$, where $b$ is a prime number and $m \in \mathbb{N}$, and let $F \in \mathcal{W}_{s,\boldsymbol{\gamma}}$ belong to the weighted, unanchored Sobolev space defined over $[-\frac{1}{2}, \frac{1}{2}]^s$ with weights $\boldsymbol{\gamma} = (\gamma_{\mathfrak{u}})_{\mathfrak{u} \subseteq \{1:s\}}$. A randomly shifted lattice rule can be constructed by the reduced CBC algorithm, see Algorithm 1, such that for all $\lambda \in (1/2, 1]$,*

$$\sqrt{\mathbb{E}_{\boldsymbol{\Delta}}\left[|I_s(F) - Q_N(F)|^2\right]} \leq \left( \sum_{\emptyset \neq \mathfrak{u} \subseteq \{1:s\}} \gamma_{\mathfrak{u}}^{\lambda}\, \varrho^{|\mathfrak{u}|}(\lambda)\, b^{\min\{m, \max_{j \in \mathfrak{u}} w_j\}} \right)^{1/(2\lambda)} \left( \frac{2}{N} \right)^{1/(2\lambda)} \|F\|_{\mathcal{W}_{s,\boldsymbol{\gamma}}},$$

*with reduction indices $w_1, \ldots, w_s \in \mathbb{N}_0$ such that $w_1 \leq w_2 \leq \cdots \leq w_s$ and $\varrho(\lambda) = 2\zeta(2\lambda)(2\pi^2)^{-\lambda}$.*

*Proof.* Using Theorem 3, and the well-known connection to the worst-case error in the Korobov space $\mathcal{H}(K_{s,2,\boldsymbol{\gamma}})$ with $\alpha = 2$ yields the following bound on the shift-averaged worst-case error for the weighted Sobolev space,

$$\widehat{e}_{N,s}(\boldsymbol{z}) \le \left( \sum_{\emptyset \neq \mathfrak{u} \subseteq \{1:s\}} \left( \frac{\gamma_{\mathfrak{u}}}{(2\pi^2)^{|\mathfrak{u}|}} \right)^{\lambda} \frac{2(2\zeta(2\lambda))^{|\mathfrak{u}|}}{b^{\max\{0, m - \max_{j\in\mathfrak{u}} w_j\}}} \right)^{1/(2\lambda)} .$$

By using the identity $b^{-\max\{0,m-\max_{j\in\mathfrak{u}} w_j\}} = b^{\min\{0,\max_{j\in\mathfrak{u}} w_j - m\}} = \frac{1}{N} b^{\min\{m,\max_{j\in\mathfrak{u}} w_j\}}$, and

$$\sqrt{\mathbb{E}_{\boldsymbol{\Delta}}\left[|I(F) - Q_N(F)|^2\right]} \le \sqrt{\mathbb{E}_{\boldsymbol{\Delta}}\left[e^2_{N,s}(\boldsymbol{z}, \boldsymbol{\Delta}) \, \|F\|^2_{\mathcal{W}_{s,\boldsymbol{\gamma}}}\right]} = \widehat{e}_{N,s}(\boldsymbol{z}) \, \|F\|_{\mathcal{W}_{s,\boldsymbol{\gamma}}},$$

we obtain the claimed result. □

## 2.4 Implementation of the reduced CBC algorithm

The squared worst-case error for POD weights $\boldsymbol{\gamma} = (\gamma_{\mathfrak{u}})_{\mathfrak{u} \subseteq \{1:s\}}$ with $\gamma_{\mathfrak{u}} = \Gamma(|\mathfrak{u}|) \prod_{j \in \mathfrak{u}} \gamma_j$ in the weighted Korobov space $\mathcal{H}(K_{s,\alpha,\boldsymbol{\gamma}})$ can be written as

$$e^2_{N,s}(\boldsymbol{z}) = \frac{1}{N} \sum_{k=0}^{N-1} \sum_{\ell=1}^{s} \sum_{\substack{\mathfrak{u} \subseteq \{1:s\} \\ |\mathfrak{u}| = \ell}} \Gamma(\ell) \prod_{j \in \mathfrak{u}} \left( \gamma_j \omega \left( \left\{ \frac{k z_j}{N} \right\} \right) \right),$$

where $\omega(x) = \sum_{h \in \mathbb{Z}_*} e^{2\pi i h x} / |h|^{\alpha}$, see, e.g., [6, 12], and for $n \in \mathbb{N}$ we set $\Omega_n$ as

$$\Omega_n := \left[ \omega \left( \frac{kz \bmod n}{n} \right) \right]_{\substack{z \in \mathbb{Z}_n^{\times} \\ k \in \mathbb{Z}_n}} \in \mathbb{R}^{\varphi(n) \times n}.$$

Similar to other variants of CBC, we here derive a fast version of the reduced CBC method for POD weights in Algorithm 2. The derivation of Algorithm 2 is explained and motivated in the Appendix in Section 5. Note the algorithm as presented only constructs the generating vector $\boldsymbol{z} = (Y_1 z_1, \ldots, Y_s z_s)$ and does not compute the associated worst-case error $e_{N,s}(\boldsymbol{z})$ (which is possible given some extra calculations). The standard fast CBC algorithm for POD weights has a complexity of $\mathcal{O}(s\, N \log N + s^2 N)$, see, e.g., [6, 12]. The cost of our new algorithm, derived from the derivation in [6], can be substantially lower as is stated in the following theorem. We stress that the presented algorithm is the first realization of the reduced CBC construction for POD weights and the construction technique is superior to the one considered in [4].

**Theorem 5.** *Given a sequence of integers $w_1 \le \ldots \le w_s$ with $w_1 = 0$, the reduced CBC algorithm for $N = b^m$ points in $s$ dimensions as specified in Algorithm 2 can construct a lattice rule with near optimal worst-case error as in Theorem 4 with an arithmetic cost of*

$$\mathcal{O}\left( \sum_{j=1}^{\min\{s,s^*\}} (m - w_j + j)\, b^{m - w_j} \right),$$

*where $s^*$ is defined to be the largest integer such that $w_{s^*} < m$.*

*Proof.* The precomputations in step (1.a) require $\mathcal{O}(m b^m)$ calculations. Moreover, the partitioning into the summands in the definition of the $\boldsymbol{q}_{j-1,\ell}$ in step 2 can be done in $\mathcal{O}(j\, b^{m - w_{j-1}})$ operations and the calculation of $\boldsymbol{q}_j$ in step 3 requires $\mathcal{O}(j\, b^{m-w_j})$ calculations for each step. Using the fast matrix-vector multiplication as in, e.g., [3], the calculation of $T_j(z)$ in step 4 for all $z \in \mathbb{Z}^{\times}_{b^{m-w_j}}$ is possible in $\mathcal{O}((m - w_j)\, b^{m-w_j})$ operations in each step. Lastly, the update of the vectors $\boldsymbol{q}_{j,\ell}$ for $\ell = 1, \ldots, j$ in step 6 uses $\mathcal{O}(j\, b^{m-w_j})$ calculations in each step. Since we assumed that $w_1 = 0$, this gives the claimed result. □

**Algorithm 2** The fast reduced CBC construction for POD weights

(1) Precomputation:

(a) Compute $\omega\left(\frac{k}{b^m}\right)$ for $k = 0, 1, \ldots, b^m - 1$ and store the results.

(b) Initialize the vectors $\boldsymbol{q}_{0,0} = \mathbf{1}_{b^m}$ and $\boldsymbol{q}_{0,1} = \mathbf{0}_{b^m}$ and set $w_0 = 0$ and $\Gamma(0) = 1$.

(c) Set $s^*$ to be the largest integer such that $w_{s^*} < m$.

For $j$ from 1 to $s$ and as long as $w_j < m$:

(2) For $\ell = 0, 1, \ldots, j$ partition each vector $\boldsymbol{q}_{j-1,\ell}$ into $b^{w_j - w_{j-1}}$ vectors $\boldsymbol{q}_{j-1,\ell}^{(1)}, \ldots, \boldsymbol{q}_{j-1,\ell}^{(b^{w_j - w_{j-1}})}$ of length $b^{m-w_j}$, where, for $t = 1, \ldots, b^{w_j - w_{j-1}}$,

$$\boldsymbol{q}_{j-1,\ell}^{(t)} = (q_{j-1,\ell}(1 + (t-1)\, b^{m-w_j}), \ldots, q_{j-1,\ell}(t\, b^{m-w_j})) \in \mathbb{R}^{m-w_j}$$

and set $\boldsymbol{q}_{j-1,\ell} = \boldsymbol{q}_{j-1,\ell}^{(1)} + \cdots + \boldsymbol{q}_{j-1,\ell}^{(b^{w_j - w_{j-1}})} \in \mathbb{R}^{m-w_j}$ for $\ell = 0, 1, \ldots, j$.

(3) Set $\boldsymbol{q}_j = \sum_{\ell=1}^{j} \frac{\Gamma(\ell)}{\Gamma(\ell-1)} \boldsymbol{q}_{j-1,\ell-1} \in \mathbb{R}^{m-w_j}$.

(4) Calculate $T_j = \Omega_{b^{m-w_j}}\, \boldsymbol{q}_j$ by exploiting the block-circulant structure of the matrix $\Omega_{b^{m-w_j}}$ using FFTs.

(5) Set $z_j = \arg\min_{z \in \mathbb{Z}^\times_{b^{m-w_j}}} T_j(z)$, where $T_j(z)$ denotes the component of $T_j$ corresponding to $z$.

(6) Set $\boldsymbol{q}_{j,0} = \mathbf{1}_{b^{m-w_j}}$ and $\boldsymbol{q}_{j,j+1} = \mathbf{0}_{b^{m-w_j}}$ and for $\ell = 1, \ldots, j$ set the vectors $\boldsymbol{q}_{j,\ell}$ via

$$\boldsymbol{q}_{j,\ell} := \boldsymbol{q}_{j-1,\ell} + \frac{\Gamma(\ell)}{\Gamma(\ell-1)} \gamma_j\, \Omega_{b^{m-w_j}}(z_j, :) \,.\!*\, \boldsymbol{q}_{j-1,\ell-1} \in \mathbb{R}^{m-w_j}.$$

If $s > s^*$, then set $z_{s^*+1} = \cdots = z_s = 0$.

# 3 QMC finite element error analysis

We now combine the results of the previous subsections to analyze the overall QMC finite element error. We consider the root mean square error (RMSE) given by

$$e^{\mathrm{RMSE}}_{N,s,h}(G(u)) := \sqrt{\mathbb{E}_{\boldsymbol{\Delta}}\left[|\mathbb{E}[G(u)] - Q_N(G(u_h^s))|^2\right]}.$$

The error $\mathbb{E}[G(u)] - Q_N(G(u_h^s))$ can be written as

$$\mathbb{E}[G(u)] - Q_N(G(u_h^s)) = \mathbb{E}[G(u)] - I_s(G(u_h^s)) + I_s(G(u_h^s)) - Q_N(G(u_h^s))$$

such that due to the fact that $\mathbb{E}_{\boldsymbol{\Delta}}(Q_N(f)) = I_s(f)$ for any integrand $f$ we obtain

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\Delta}}\left[(\mathbb{E}[G(u)] - Q_N(G(u_h^s)))^2\right] &= (\mathbb{E}[G(u)] - I_s(G(u_h^s)))^2 + \mathbb{E}_{\boldsymbol{\Delta}}\left[(I_s - Q_N)^2(G(u_h^s))\right] \\
&\quad + 2(\mathbb{E}[G(u)] - I_s(G(u_h^s)))\, \mathbb{E}_{\boldsymbol{\Delta}}\left[(I_s - Q_N)(G(u_h^s))\right] \\
&= (\mathbb{E}[G(u)] - I_s(G(u_h^s)))^2 + \mathbb{E}_{\boldsymbol{\Delta}}\left[(I_s - Q_N)^2(G(u_h^s))\right].
\end{aligned}$$

Then, noting that $\mathbb{E}[G(u)] - I_s(G(u_h^s)) = \mathbb{E}[G(u)] - I_s(G(u^s)) + I_s(G(u^s)) - I_s(G(u_h^s))$,

$$\begin{aligned}(\mathbb{E}[G(u)] - I_s(G(u_h^s)))^2 &= (\mathbb{E}[G(u)] - I_s(G(u^s)))^2 + (I_s(G(u^s)) - I_s(G(u_h^s)))^2 \\ &\quad + 2(\mathbb{E}[G(u)] - I_s(G(u^s)))(I_s(G(u^s)) - I_s(G(u_h^s)))\end{aligned}$$

and since for general $x, y \in \mathbb{R}$ it holds that $2xy \leq x^2 + y^2$, we obtain furthermore

$$(\mathbb{E}[G(u)] - I_s(G(u_h^s)))^2 \leq 2(\mathbb{E}[G(u)] - I_s(G(u^s)))^2 + 2(I_s(G(u^s)) - I_s(G(u_h^s)))^2.$$

From the previous subsections we can then use (2.6) for the truncation part, (2.4), which holds for general $\boldsymbol{y} \in U$ and thus also for $y_{\{1:s\}}$, for the finite element error, and Theorem 4 for the QMC integration error to obtain the following error bound for the mean square error $\mathbb{E}_{\boldsymbol{\Delta}}[|\mathbb{E}[G(u)] - Q_N(G(u_h^s))|^2] =: e_{N,s,h}^{\mathrm{MSE}}(G(u))$,

$$\begin{aligned} e_{N,s,h}^{\mathrm{MSE}}(G(u)) \leq K_1 \|f\|_{V^*}^2 \|G\|_{V^*}^2 &\left( \frac{1}{(1-\overline{\kappa})\, a_{0,\min} - a_{0,\max}\, \kappa \sup_{j \geq s+1} b_j} \right)^2 \\ &\times \left( \frac{a_{0,\max}}{(1-\overline{\kappa})\, a_{0,\min}} \kappa \sup_{j\geq s+1} b_j \right)^4 + K_2 \|f\|_{L^2}^2 \|G\|_{L^2}^2\, h^4 \\ &+ \left( \sum_{\emptyset \neq \mathfrak{u} \subseteq \{1:s\}} \gamma_{\mathfrak{u}}^{\lambda}\, \varrho^{|\mathfrak{u}|}(\lambda)\, b^{\min\{m, \max_{j\in\mathfrak{u}} w_j\}} \right)^{1/\lambda} \left(\frac{2}{N}\right)^{1/\lambda} \|G(u_h^s)\|_{\mathcal{W}_{s,\boldsymbol{\gamma}}}^2 \end{aligned} \tag{3.1}$$

for some constants $K_1, K_2 \in \mathbb{R}_+$ and provided that $\frac{a_{0,\max}}{(1-\overline{\kappa})\, a_{0,\min}} \kappa \sup_{j \geq s+1} b_j < 1$.

## 3.1 Derivative bounds of POD form

In the following we assume that we have general bounds on the mixed partial derivatives $\partial^{\boldsymbol{\nu}} u(\cdot, \boldsymbol{y})$ which are of POD form; that is,

$$\|\partial^{\boldsymbol{\nu}} u(\cdot, \boldsymbol{y})\|_V \leq C\, \widetilde{\boldsymbol{b}}^{\boldsymbol{\nu}}\, \Gamma(|\boldsymbol{\nu}|)\, \|f\|_{V^*} \tag{3.2}$$

with a map $\Gamma : \mathbb{N}_0 \to \mathbb{R}$, a sequence of reals $\widetilde{\boldsymbol{b}} = (\widetilde{b}_j)_{j\geq 1} \in \mathbb{R}^{\mathbb{N}}$ and some constant $C \in \mathbb{R}_+$. Such bounds can be found in the literature and we provided a new derivation in Theorem 1 also leading to POD weights.

For bounding the norm $\|G(u_h^s)\|_{\mathcal{W}_{s,\boldsymbol{\gamma}}}$, we can then use (3.2) and the definition in (2.7) to proceed as outlined in [10], to obtain the estimate

$$\|G(u_h^s)\|_{\mathcal{W}_{s,\boldsymbol{\gamma}}} \leq C\, \|f\|_{V^*} \|G\|_{V^*} \left( \sum_{\mathfrak{u} \subseteq \{1:s\}} \frac{\Gamma(|\mathfrak{u}|)^2 \prod_{j\in\mathfrak{u}} \widetilde{b}_j^2}{\gamma_{\mathfrak{u}}} \right)^{1/2}. \tag{3.3}$$

Denoting $\boldsymbol{w} := (w_j)_{j\geq 1}$ and using (3.3), the contribution of the quadrature error to the mean square error $e_{N,h,s}^{\mathrm{MSE}}(G(u))$ can be upper bounded by

$$\begin{aligned} &\left( \sum_{\emptyset \neq \mathfrak{u} \subseteq \{1:s\}} \gamma_{\mathfrak{u}}^{\lambda}\, \varrho^{|\mathfrak{u}|}(\lambda)\, b^{\min\{m, \max_{j\in\mathfrak{u}} w_j\}} \right)^{1/\lambda} \left(\frac{2}{N}\right)^{1/\lambda} \|G(u_h^s)\|_{\mathcal{W}_{s,\boldsymbol{\gamma}}}^2 \\ &\qquad \leq C\, \|f\|_{V^*} \|G\|_{V^*}\, C_{\boldsymbol{\gamma},\boldsymbol{w},\lambda} \left(\frac{2}{N}\right)^{1/\lambda}, \end{aligned} \tag{3.4}$$

where we define

$$C_{\boldsymbol{\gamma},\boldsymbol{w},\lambda} := \left( \sum_{\emptyset \neq \mathfrak{u} \subseteq \{1:s\}} \gamma_{\mathfrak{u}}^{\lambda}\, \varrho^{|\mathfrak{u}|}(\lambda)\, b^{\min\{m, \max_{j\in\mathfrak{u}} w_j\}} \right)^{1/\lambda} \left( \sum_{\mathfrak{u} \subseteq \{1:s\}} \frac{\Gamma(|\mathfrak{u}|)^2 \prod_{j\in\mathfrak{u}} \widetilde{b}_j^2}{\gamma_{\mathfrak{u}}} \right).$$

The term $C_{\boldsymbol{\gamma},\boldsymbol{w},\lambda}$ can be bounded as

$$C_{\boldsymbol{\gamma},\boldsymbol{w},\lambda} \le \left( \sum_{\mathfrak{u}\subseteq\{1:s\}} \gamma_{\mathfrak{u}}^{\lambda}\, \varrho^{|\mathfrak{u}|}(\lambda)\, b^{\sum_{j\in\mathfrak{u}} w_j - \sum_{\ell=1}^{|\mathfrak{u}|-1} w_\ell} \right)^{1/\lambda} \left( \sum_{\mathfrak{u}\subseteq\{1:s\}} \frac{\Gamma(|\mathfrak{u}|)^2 \prod_{j\in\mathfrak{u}} \widetilde{b}_j^2}{\gamma_{\mathfrak{u}}} \right).$$

Due to [11, Lemma 6.2] the latter term is minimized by choosing the weights $\gamma_{\mathfrak{u}}$ as

$$\gamma_{\mathfrak{u}} := \left( \frac{\Gamma(|\mathfrak{u}|)^2 \prod_{j\in\mathfrak{u}} \widetilde{b}_j^2 \prod_{\ell=1}^{|\mathfrak{u}|-1} b^{w_\ell}}{\prod_{j\in\mathfrak{u}} \rho(\lambda)\, b^{w_j}} \right)^{1/(1+\lambda)}. \tag{3.5}$$

Then we set

$$A_\lambda := \sum_{\mathfrak{u}\subseteq\{1:s\}} \gamma_{\mathfrak{u}}^{\lambda}\, \varrho^{|\mathfrak{u}|}(\lambda)\, b^{\sum_{j\in\mathfrak{u}} w_j - \sum_{\ell=1}^{|\mathfrak{u}|-1} w_\ell} = \sum_{\mathfrak{u}\subseteq\{1:s\}} \left[ \left( \frac{\Gamma(|\mathfrak{u}|)^{2\lambda}}{\prod_{\ell=1}^{|\mathfrak{u}|-1} b^{w_\ell}} \right) \left( \prod_{j\in\mathfrak{u}} \rho(\lambda)\, \widetilde{b}_j^{2\lambda}\, b^{w_j} \right) \right]^{\frac{1}{1+\lambda}}$$

and easily see that also

$$\sum_{\mathfrak{u}\subseteq\{1:s\}} \gamma_{\mathfrak{u}}^{-1} \left( \Gamma(|\mathfrak{u}|)^2 \prod_{j\in\mathfrak{u}} \widetilde{b}_j^2 \right) = A_\lambda,$$

which implies that $C_{\boldsymbol{\gamma},\boldsymbol{w},\lambda} \le A_\lambda^{1+1/\lambda}$. We demonstrate how the term $A_\lambda$ can be estimated for the derivative bounds derived in Section 1.2.

In view of Theorem 1, assume in the following that

$$\Gamma(|\mathfrak{u}|) = \kappa^{|\mathfrak{u}|}, \quad \widetilde{b}_j = \frac{2\, b_j}{1-\kappa}, \quad \sum_{j=1}^{\infty} (b_j b^{w_j})^p < \infty \quad \text{for} \quad p \in (0,1). \tag{3.6}$$

Note that we could also choose $\Gamma(|\mathfrak{u}|) = \kappa(|\mathfrak{u}|)^{|\mathfrak{u}|}$ above, in which case the subsequent estimate of $A_\lambda$ can be done analogously, but to make the argument less technical, we consider the slightly coarser variant $\Gamma(|\mathfrak{u}|) = \kappa^{|\mathfrak{u}|}$ here. In this case,

$$A_\lambda = \sum_{\mathfrak{u}\subseteq\{1:s\}} \left[\kappa^{|\mathfrak{u}|}\right]^{\frac{2\lambda}{1+\lambda}} \left( \prod_{\ell=1}^{|\mathfrak{u}|-1} b^{\frac{-w_\ell}{2\lambda}} \right)^{\frac{2\lambda}{1+\lambda}} \prod_{j\in\mathfrak{u}} \left( \left( \frac{2\, b_j}{1-\kappa} \right)^{2\lambda} b^{w_j}\, \rho(\lambda) \right)^{\frac{1}{1+\lambda}}.$$

Note that, as $\lambda \le 1$, it holds that $b^{\frac{-w_\ell}{2\lambda}} \le b^{\frac{-w_\ell}{2}}$ and hence

$$A_\lambda \le \sum_{\mathfrak{u}\subseteq\{1:s\}} \left( \kappa^{|\mathfrak{u}|} \prod_{\ell=1}^{|\mathfrak{u}|-1} b^{\frac{-w_\ell}{2}} \right)^{\frac{2\lambda}{1+\lambda}} \prod_{j\in\mathfrak{u}} \left( \left( \frac{2\, b_j}{1-\kappa} \right)^{2\lambda} b^{w_j}\, \rho(\lambda) \right)^{\frac{1}{1+\lambda}}.$$

We now proceed similarly to the proof of Theorem 6.4 in [11]. Let $(\alpha_j)_{j\ge1}$ be a sequence of positive reals, to be specified below, which satisfies $\Sigma := \sum_{j=1}^{\infty} \alpha_j < \infty$. Dividing and multiplying by $\prod_{j\in\mathfrak{u}} \alpha_j^{(2\lambda)/(1+\lambda)}$, and applying Hölder's inequality with conjugate components $p = (1+\lambda)/(2\lambda)$ and $p^* = (1+\lambda)/(1-\lambda)$,

$$\begin{aligned} A_\lambda &\le \sum_{\mathfrak{u}\subseteq\{1:s\}} \left( \kappa^{|\mathfrak{u}|} \prod_{\ell=1}^{|\mathfrak{u}|-1} b^{\frac{-w_\ell}{2}} \right)^{\frac{2\lambda}{1+\lambda}} \left( \prod_{j\in\mathfrak{u}} \alpha_j^{\frac{2\lambda}{1+\lambda}} \right) \prod_{j\in\mathfrak{u}} \left( \left( \frac{2\, b_j}{1-\kappa} \right)^{2\lambda} b^{w_j}\, \rho(\lambda)/\alpha_j^{2\lambda} \right)^{\frac{1}{1+\lambda}} \\ &\le \left( \sum_{\mathfrak{u}\subseteq\{1:s\}} \kappa^{|\mathfrak{u}|} \left( \prod_{\ell=1}^{|\mathfrak{u}|-1} b^{\frac{-w_\ell}{2}} \right) \prod_{j\in\mathfrak{u}} \alpha_j \right)^{\frac{2\lambda}{1+\lambda}} \end{aligned}$$

$$\times\left(\sum_{\mathfrak{u}\subseteq\{1:s\}}\prod_{j\in\mathfrak{u}}\left(\left(\frac{2\,b_j}{1-\kappa}\right)^{2\lambda} b^{w_j}\,\rho(\lambda)/\alpha_j^{2\lambda}\right)^{\frac{1}{1-\lambda}}\right)^{\frac{1-\lambda}{1+\lambda}} = B^{\frac{2\lambda}{1+\lambda}}\cdot \widetilde{B}^{\frac{1-\lambda}{1+\lambda}},$$

where we define

$$B := \sum_{\mathfrak{u}\subseteq\{1:s\}}\kappa^{|\mathfrak{u}|}\left(\prod_{\ell=1}^{|\mathfrak{u}|-1} b^{\frac{-w_\ell}{2}}\right)\prod_{j\in\mathfrak{u}}\alpha_j,\quad \widetilde{B} := \sum_{\mathfrak{u}\subseteq\{1:s\}}\prod_{j\in\mathfrak{u}}\left(\left(\frac{2\,b_j}{1-\kappa}\right)^{2\lambda}\frac{b^{w_j}\,\rho(\lambda)}{\alpha_j^{2\lambda}}\right)^{\frac{1}{1-\lambda}}.$$

For the first factor we estimate

$$\begin{aligned} B &\leq \sum_{\mathfrak{u}\colon|\mathfrak{u}|<\infty}\kappa^{|\mathfrak{u}|}\prod_{\ell=1}^{|\mathfrak{u}|-1} b^{\frac{-w_\ell}{2}}\prod_{j\in\mathfrak{u}}\alpha_j = \sum_{k=1}^{\infty}\left(\kappa^k\prod_{\ell=1}^{k-1} b^{\frac{-w_\ell}{2}}\right)\sum_{\substack{\mathfrak{u}\colon|\mathfrak{u}|<\infty\\|\mathfrak{u}|=k}}\prod_{j\in\mathfrak{u}}\alpha_j \\ &\leq \sum_{k=1}^{\infty}\left(\kappa^k\prod_{\ell=1}^{k-1} b^{\frac{-w_\ell}{2}}\right)\frac{1}{k!}\sum_{\boldsymbol{u}\in\mathbb{N}^k}\prod_{i=1}^{k}\alpha_{u_i} = \sum_{k=1}^{\infty}\left(\kappa^k\prod_{\ell=1}^{k-1} b^{\frac{-w_\ell}{2}}\right)\frac{1}{k!}\Sigma^k. \end{aligned}$$

By the ratio test, the latter expression is finite if we choose $(\alpha_j)_{j\geq 1}$ such that $L := \sup_{k\in\mathbb{N}}\kappa\, b^{\frac{-w_k}{2}}(k+1)^{-1} = \kappa b^{\frac{-w_1}{2}}/2 < 1/\Sigma$. Hence we assume that $(\alpha_j)_{j\geq 1}$ is chosen such that indeed $L<1/\Sigma$. Note that $L$ is small if $\kappa$ is small, which means that $\Sigma$ can be allowed to be large in this case. Consider now the term

$$\begin{aligned} \widetilde{B} &\leq \sum_{\mathfrak{u}\colon|\mathfrak{u}|<\infty}\prod_{j\in\mathfrak{u}}\left(\left(\frac{2\,b_j}{1-\kappa}\right)^{2\lambda} b^{w_j}\,\rho(\lambda)/\alpha_j^{2\lambda}\right)^{\frac{1}{1-\lambda}} \\ &\leq \exp\left(\sum_{j=1}^{\infty}\left(\left(\frac{2\,b_j}{1-\kappa}\right)^{2\lambda} b^{w_j}\,\rho(\lambda)/\alpha_j^{2\lambda}\right)^{\frac{1}{1-\lambda}}\right) \\ &\leq \exp\left(\sum_{j=1}^{\infty}\left(\frac{1}{1-\kappa}\right)^{\frac{2\lambda}{1-\lambda}}(\rho(\lambda))^{\frac{1}{1-\lambda}}4^{\lambda}\left(b_j b^{w_j}\frac{1}{\alpha_j}\right)^{\frac{2\lambda}{1-\lambda}}\right) \\ &= \exp\left((1-\kappa)^{\frac{-2\lambda}{1-\lambda}}(\rho(\lambda))^{\frac{1}{1-\lambda}}4^{\lambda}\sum_{j=1}^{\infty}\left(b_j b^{w_j}\alpha_j^{-1}\right)^{\frac{2\lambda}{1-\lambda}}\right). \end{aligned}$$

We require

$$L < 1/\Sigma = 1/\sum_{j=1}^{\infty}\alpha_j \quad\text{and}\quad \sum_{j=1}^{\infty}\left(b_j b^{w_j}\alpha_j^{-1}\right)^{\frac{2\lambda}{1-\lambda}} < \infty. \tag{3.7}$$

To this end, we choose $\alpha_j := \frac{(b_j b^{w_j})^p}{\theta}$, where $\frac{\theta}{L} > \sum_{j=1}^{\infty}(b_j b^{w_j})^p$. Then,

$$\begin{aligned} A_\lambda &\leq \left(\sum_{k=1}^{\infty}\left(\kappa^k\prod_{\ell=1}^{k-1} b^{\frac{-w_\ell}{2}}\right)\frac{1}{k!}\Sigma^k\right)^{\frac{2\lambda}{1+\lambda}} \\ &\quad\times\exp\left(\frac{1-\lambda}{1+\lambda}\left(\frac{1}{1-\kappa}\right)^{\frac{2\lambda}{1-\lambda}}(\rho(\lambda))^{\frac{1}{1-\lambda}}4^{\lambda}\sum_{j=1}^{\infty}\left(b_j b^{w_j}\frac{1}{\alpha_j}\right)^{\frac{2\lambda}{1-\lambda}}\right) \end{aligned} \tag{3.8}$$

as long as we choose $\lambda$ such that

$$\sum_{j=1}^{\infty}\left(b_j b^{w_j}\alpha_j^{-1}\right)^{2\lambda/(1-\lambda)} < \infty. \tag{3.9}$$

We denote the upper bound in (3.8) by $\overline{A}(\lambda)$. Similarly to what is done in [11, Proof of Theorem 6.4], we see that Condition (3.9) is satisfied if $\lambda \geq \frac{p}{2-p}$. Again, similarly to [11, Proof of Theorem 6.4] we see that the latter can be achieved by choosing

$$\lambda_p = \begin{cases} 1/(2-2\delta) & \text{for some } \delta \in (0,1/2) \quad \text{if } p \in (0,2/3], \\ p/(2-p) & \text{if } p \in (2/3,1). \end{cases} \tag{3.10}$$

Hence by choosing $\lambda$ equal to $\lambda_p$, we get an efficient bound on $C_{\boldsymbol{\gamma},\boldsymbol{w},\lambda_p} = A_{\lambda_p}^{1+1/\lambda_p}$, as long as the $w_j$ are chosen to guarantee convergence of $\sum_{j=1}^{\infty} (b_j b^{w_j})^p$.

# 4 Combined error bound

The derivation in the previous section leads to the following result.

**Theorem 6.** *Given the PDE in* (1.1) *for which we characterized the regularity of the random field by a sequence of $b_j$ with sparsity $p \in (0,1)$ and determined a sequence of $w_j$ such that $\sum_{j=1}^{\infty}(b_j\, b^{w_j})^p < \infty$, we can construct the generating vector for an $N$-point randomized lattice rule using the reduced CBC algorithm (Algorithm 2), at the cost of $\mathcal{O}(\sum_{j=1}^{\min\{s,s^*\}}(m - w_j + j)\, b^{m-w_j})$ operations, such that, assuming that* (1.11), (2.3) *and* $\frac{\kappa\, a_{0,\max}}{(1-\kappa)\, a_{0,\min}} \sup_{j\geq s+1} b_j < 1$ *hold, we obtain an upper bound*

$$e_{N,s,h}^{MSE}(G(u)) \lesssim \left(\sup_{j\geq s+1} b_j\right)^2 + h^4 + \left(\frac{2}{N}\right)^{1/\lambda_p}, \tag{4.1}$$

*where the implied constant is independent of $s$, $h$ and $N$.*

Observe that if the $w_j$ increase sufficiently fast, the construction cost of Algorithm 2 does not depend anymore on the increasing dimensionality. Further note that the first term on the right-hand side of (4.1) is small if $\sup_{j\geq s+1} b_j$ is small, and, since we assumed that $b_j$ must tend to zero by assumption (3.6), we can shrink the first summand by choosing $s$ sufficiently large. By choosing $h$ sufficiently small, and $N$ sufficiently large, we can also make the other two summands in the overall error bound small.

Note that $\sup_{j\geq s+1} b_j \leq \sum_{j\geq s+1} b_j$, and that (3.6) yields $\sum_{j=1}^{\infty} b_j^p < \infty$, which implies that one can use the machinery developed in [11] to obtain a cost analysis similar to [11, Theorem 8.1]. Note, in particular, that it is sufficient to choose $N$ of order $\mathcal{O}(\varepsilon^{-\lambda_p/2})$, independently of $s$, to meet an error threshold of $\varepsilon$.

# 5 Appendix

In this appendix, we motivate the derivation of the fast reduced CBC algorithm for POD weights in Algorithm 2.

To this end, let $d \in \{1,\ldots,s\}$ and assume $z_1,\ldots,z_{d-1}$ have already been selected. For $N = b^m$ and $z_d \in \mathbb{Z}_{b^{m-w_d}}^{\times}$ with $w_d < m$, we denote the worst-case error $e_{N,d}(z_1,\ldots,z_d)$ by $e_{N,d}(z_d)$. Then we obtain the following expression

$$\begin{aligned} e_{N,d}^2(z_d) &= \frac{1}{b^m}\sum_{k=0}^{b^m-1}\sum_{\ell=1}^{d}\sum_{\substack{\mathfrak{u}\subseteq\{1:d\}\\|\mathfrak{u}|=\ell}}\Gamma_\ell\prod_{j\in\mathfrak{u}}\left(\gamma_j\omega\left(\frac{k\,b^{w_j}z_j \bmod b^m}{b^m}\right)\right) \\ &= \frac{1}{b^m}\sum_{k=0}^{b^m-1}\sum_{\ell=1}^{d}\underbrace{\sum_{\substack{\mathfrak{u}\subseteq\{1:d\}\\|\mathfrak{u}|=\ell}}\Gamma_\ell\prod_{j\in\mathfrak{u}}\left(\gamma_j\omega\left(\frac{k\,z_j \bmod b^{m-w_j}}{b^{m-w_j}}\right)\right)}_{=:q_{d,\ell}(k)} = \frac{1}{b^m}\sum_{k=0}^{b^m-1}\sum_{\ell=1}^{d} q_{d,\ell}(k) \end{aligned}$$

$$= \frac{1}{b^m} \sum_{k=0}^{b^m-1} \sum_{\ell=1}^{d} \Bigg[ \underbrace{\sum_{\substack{\mathfrak{u}\subseteq\{1:d-1\}\\|\mathfrak{u}|=\ell}} \Gamma_\ell \prod_{j\in\mathfrak{u}} \left(\gamma_j \omega\left(\frac{k\, z_j \bmod b^{m-w_j}}{b^{m-w_j}}\right)\right)}_{=q_{d-1,\ell}(k)}$$

$$+ \frac{\Gamma_\ell}{\Gamma_{\ell-1}} \gamma_d \underbrace{\omega\left(\frac{k\, z_d \bmod b^{m-w_d}}{b^{m-w_d}}\right)}_{=\Omega_{b^{m-w_d}}(z_d,k)} \underbrace{\sum_{\substack{\mathfrak{u}\subseteq\{1:d-1\}\\|\mathfrak{u}|=\ell-1}} \Gamma_{\ell-1} \prod_{j\in\mathfrak{u}} \left(\gamma_j \omega\left(\frac{k\, z_j \bmod b^{m-w_j}}{b^{m-w_j}}\right)\right)}_{=q_{d-1,\ell-1}(k)} \Bigg].$$

Since $\Omega_{b^{m-w_d}}(z_d,k) = \omega(\frac{k\,z_d \bmod b^{m-w_d}}{b^{m-w_d}}) = \Omega_{b^{m-w_d}}(z_d, k \bmod b^{m-w_d})$, we can rewrite the sum $\sum_{k=0}^{b^m-1} a(k)$ as $\sum_{t=1}^{b^{w_d}} \sum_{k=0}^{b^{m-w_d}-1} a(k+(t-1)b^{m-w_d})$ for suitably chosen map $a: \mathbb{N}_0 \to \mathbb{R}$ to obtain

$$\begin{aligned}
&e^2_{N,d}(z_d)\\
&= \frac{1}{b^m} \sum_{\ell=1}^{d} \Bigg[ \sum_{t=1}^{b^{w_d}} \sum_{k=0}^{b^{m-w_d}-1} \sum_{\substack{\mathfrak{u}\subseteq\{1:d-1\}\\|\mathfrak{u}|=\ell}} \Gamma_\ell \prod_{j\in\mathfrak{u}} \left(\gamma_j \omega\left(\frac{(k+(t-1)b^{m-w_d})\, z_j \bmod b^{m-w_j}}{b^{m-w_j}}\right)\right)\\
&\quad + \sum_{t=1}^{b^{w_d}} \sum_{k=0}^{b^{m-w_d}-1} \frac{\Gamma_\ell}{\Gamma_{\ell-1}} \gamma_d\, \omega\left(\frac{(k+(t-1)b^{m-w_d})\, z_d \bmod b^{m-w_d}}{b^{m-w_d}}\right)\\
&\quad \times \sum_{\substack{\mathfrak{u}\subseteq\{1:d-1\}\\|\mathfrak{u}|=\ell-1}} \Gamma_{\ell-1} \prod_{j\in\mathfrak{u}} \left(\gamma_j \omega\left(\frac{(k+(t-1)b^{m-w_d})\, z_j \bmod b^{m-w_j}}{b^{m-w_j}}\right)\right) \Bigg]\\
&= \frac{1}{b^m} \sum_{\ell=1}^{d} \Bigg[ \sum_{t=1}^{b^{w_d}} \sum_{k=0}^{b^{m-w_d}-1} \sum_{\substack{\mathfrak{u}\subseteq\{1:d-1\}\\|\mathfrak{u}|=\ell}} \Gamma_\ell \prod_{j\in\mathfrak{u}} \left(\gamma_j \omega\left(\frac{(k+(t-1)b^{m-w_d})\, z_j \bmod b^{m-w_j}}{b^{m-w_j}}\right)\right)\\
&\quad + \sum_{k=0}^{b^{m-w_d}-1} \frac{\Gamma_\ell}{\Gamma_{\ell-1}} \gamma_d\, \omega\left(\frac{k\, z_d \bmod b^{m-w_d}}{b^{m-w_d}}\right)\\
&\quad \times \sum_{t=1}^{b^{w_d}} \sum_{\substack{\mathfrak{u}\subseteq\{1:d-1\}\\|\mathfrak{u}|=\ell-1}} \Gamma_{\ell-1} \prod_{j\in\mathfrak{u}} \left(\gamma_j \omega\left(\frac{(k+(t-1)b^{m-w_d})\, z_j \bmod b^{m-w_j}}{b^{m-w_j}}\right)\right) \Bigg].
\end{aligned}$$

Defining now the vector $\boldsymbol{q}_{d,\ell,w} \in \mathbb{R}^{m-w_d}$ with $w < m$ and components

$$q_{d,\ell,w}(k) := \sum_{t=1}^{b^{w_d}} \sum_{\substack{\mathfrak{u}\subseteq\{1:d\}\\|\mathfrak{u}|=\ell}} \Gamma_\ell \prod_{j\in\mathfrak{u}} \left(\gamma_j \omega\left(\frac{(k+(t-1)b^{m-w})\, z_j \bmod b^{m-w_j}}{b^{m-w_j}}\right)\right)$$

and using the above derivation, we can rewrite the squared worst-case error as

$$\begin{aligned}
e^2_{N,d}(z_d) &= \frac{1}{b^m} \sum_{k=0}^{b^{m-w_d}-1} \sum_{\ell=1}^{d} q_{d,\ell,w_d}(k)\\
&= \frac{1}{b^m} \sum_{t=1}^{b^{w_d}} \sum_{k=0}^{b^{m-w_d}-1} \sum_{\ell=1}^{d} \sum_{\substack{\mathfrak{u}\subseteq\{1:d\}\\|\mathfrak{u}|=\ell}} \Gamma_\ell \prod_{j\in\mathfrak{u}} \left(\gamma_j \omega\left(\frac{(k+(t-1)b^{m-w_d})\, z_j \bmod b^{m-w_j}}{b^{m-w_j}}\right)\right)
\end{aligned}$$

$$= \frac{1}{b^m} \sum_{\ell=1}^{d} \left[ \sum_{k=0}^{b^{m-w_d}-1} q_{d-1,\ell,w_d}(k) + \sum_{k=0}^{b^{m-w_d}-1} \frac{\Gamma_\ell}{\Gamma_{\ell-1}} \gamma_d \, \Omega_{b^{m-w_d}}(z_d, k) \, q_{d-1,\ell-1,w_d}(k) \right]$$

$$= e^2_{N,d-1}(z_1, \ldots, z_{d-1}) + \frac{\gamma_d}{b^m} \sum_{k=0}^{b^{m-w_d}-1} \Omega_{b^{m-w_d}}(z_d, k) \left( \sum_{\ell=1}^{d} \frac{\Gamma_\ell}{\Gamma_{\ell-1}} q_{d-1,\ell-1,w_d}(k) \right). \quad (5.1)$$

In matrix-vector notation, for all $z \in \mathbb{Z}^\times_{b^{m-w_d}}$, this can be written as

$$\boldsymbol{e}^2_{N,d} = e^2_{N,d-1} \mathbf{1}_{\varphi(b^{m-w_d})} + \frac{\gamma_d}{b^m} \Omega_{b^{m-w_d}} \left( \sum_{\ell=1}^{d} \frac{\Gamma_\ell}{\Gamma_{\ell-1}} \boldsymbol{q}_{d-1,\ell-1,w_d} \right), \quad (5.2)$$

and furthermore gives rise to the following update formula

$$\boldsymbol{q}_{d,\ell,w_d} = \boldsymbol{q}_{d-1,\ell,w_d} + \frac{\Gamma_\ell}{\Gamma_{\ell-1}} \gamma_d \, \Omega_{b^{m-w_d}}(z^*_d, \cdot) \, \boldsymbol{q}_{d-1,\ell-1,w_d}, \quad (5.3)$$

where $z^*_d \in \mathbb{Z}^\times_{b^{m-w_d}}$ is the component which minimizes $e^2_{N,d}(z_d)$ as function of $z_d$. We note that the expression in (5.1) has the same structure as the original worst-case error with $q_{d,\ell}(k)$ replaced by $q_{d,\ell,w_d}(k)$. Due to the identities in (1.33) and (1.34), we can employ the above argumentation inductively for the step $d+1$, where then the summation ranges from $k = 0, 1, \ldots, b^{m-w_d} - 1$, to obtain the formulation of Algorithm 2. Here, equation (1.35) justifies the minimization and matrix-vector multiplication of Algorithm 2 while equation (1.36) justifies the update formula for the vectors $\boldsymbol{q}_{j,\ell}$ in Algorithm 2.

# References


[1] M. Bachmayr, A. Cohen, G. Migliorati. Sparse polynomial approximation of parametric elliptic PDEs. Part I: affine coefficients. *ESAIM: Mathematical Modelling and Numerical Analysis*, 51:321–339, 2017.

[2] A. Cohen, R. De Vore, Ch. Schwab. Convergence rates of best $N$-term Galerkin approximations for a class of elliptic sPDEs. *Foundations of Computational Mathematics*, 10:615–646, 2010.

[3] R. Cools, F.Y. Kuo, D. Nuyens. Constructing embedded lattice rules for multivariate integration. *SIAM Journal on Scientific Computing*, 28:2162–2188, 2006.

[4] J. Dick, P. Kritzer, G. Leobacher, F. Pillichshammer. A reduced fast component-by-component construction of lattice points for integration in weighted spaces with fast decreasing weights. *Journal of Computational Applied Mathematics*, 276:1–15, 2015.

[5] J. Dick, F.Y. Kuo, Q.T. Le Gia, D. Nuyens, Ch. Schwab. Higher order QMC Galerkin discretization for parametric operator equations. *SIAM Journal of Numerical Analysis*, 52:2676–2702, 2014.

[6] J. Dick, F.Y. Kuo, I.H. Sloan. High-dimensional integration: The quasi-Monte Carlo way. *Acta Numerica*, 22:133–288, 2013.

[7] R.N. Gantner, L. Herrmann, Ch. Schwab. Quasi-Monte Carlo integration for affine-parametric, elliptic PDEs: Local supports and product weights. *SIAM Journal on Numerical Analysis*, 56:111–135, 2018.

[8] L. Herrmann, Ch. Schwab. QMC integration for lognormal-parametric, elliptic PDEs: local supports and product weights. *Numerische Mathematik*, 2018.

[9] Y. Kazashi. Quasi-Monte Carlo integration with product weights for elliptic PDEs with log-normal coefficients. *IMA Journal of Numerical Analysis*, 2017

[10] F.Y. Kuo, D. Nuyens. Application of Quasi-Monte Carlo Methods to Elliptic PDEs with Random Diffusion Coefficients: A Survey of Analysis and Implementation. *Foundations of Computational Mathematics*, 16:1631–1696, 2016.

[11] F.Y. Kuo, Ch. Schwab, I.H. Sloan. Quasi-Monte Carlo finite element methods for a class of elliptic partial differential equations with random coefficients. *SIAM Journal of Numerical Analysis*, 50:3351–3374, 2012.

[12] D. Nuyens. The construction of good lattice rules and polynomial lattice rules. In *Uniform Distribution and Quasi-Monte Carlo Methods: Discrepancy, Integration and Applications*, (P. Kritzer, H. Niederreiter, F. Pillichshammer, A. Winterhof, eds), De Gruyter, pp. 223–255, 2014.

[13] I.H. Sloan, H. Woźniakowski. Tractability of multivariate integration for weighted Korobov classes. *Journal of Complexity*, 17:697–721, 2001.

**Authors' addresses:**

Adrian Ebert
Department of Computer Science
KU Leuven
Celestijnenlaan 200A, 3001 Leuven, Belgium.
`adrian.ebert@cs.kuleuven.be`

Peter Kritzer
Johann Radon Institute for Computational and Applied Mathematics (RICAM)
Austrian Academy of Sciences
Altenbergerstr. 69, 4040 Linz, Austria.
`peter.kritzer@oeaw.ac.at`

Dirk Nuyens
Department of Computer Science
KU Leuven
Celestijnenlaan 200A, 3001 Leuven, Belgium.
`dirk.nuyens@cs.kuleuven.be`