Multilevel higher order Quasi-Monte Carlo Bayesian Estimation

J. Dick and R. N. Gantner and Q. T. Le Gia and Ch. Schwab

Multilevel higher order Quasi-Monte Carlo Bayesian Estimation

Josef Dick, Robert N. Gantner, Quoc T. Le Gia, and Christoph Schwab

July 26, 2016

Abstract

We provide deterministic approximation algorithms for Bayesian inverse problems for operator equations with "noisy" input data. The algorithms use a multilevel (ML) approach based on deterministic, higher order quasi-Monte Carlo (HoQMC) quadrature for approximating the high-dimensional expectations, which arise in the Bayesian estimators, and a Petrov-Galerkin (PG) method for approximating the solution to the underlying partial differential equation (PDE). This extends the previous single-level approach from [J. Dick, R. N. Gantner, Q. T. Le Gia and Ch. Schwab, Higher order Quasi-Monte Carlo integration for Bayesian Estimation. Report 2016 Seminar for Applied Mathematics, ETH Zürich (in review)].

Compared to the single-level approach, the present convergence analysis of the multilevel method requires stronger assumptions on holomorphy and regularity of the countably-parametric uncertainty-to-observation maps of the forward problem. As in the single-level case and in the affine-parametric case analyzed in [J. Dick, F.Y. Kuo, Q. T. Le Gia and Ch. Schwab, Multi-level higher order QMC Galerkin discretization for affine parametric operator equations. Accepted for publication in SIAM J. Numer. Anal., 2016], we obtain sufficient conditions which allow us to achieve arbitrarily high, algebraic convergence rates in terms of work, which are independent of the dimension of the parameter space. The convergence rates are limited only by the spatial regularity of the forward problem, the discretization order achieved by the Petrov Galerkin discretization, and by the sparsity of the uncertainty parametrization.

We provide detailed numerical experiments for linear elliptic problems in two space dimensions, with $s = 1024$ parameters characterizing the uncertain input, confirming the theory and showing that the ML HoQMC algorithms outperform, in terms of error vs. computational work, both multilevel Monte Carlo (MLMC) methods and single-level (SL) HoQMC methods.

Key words: Higher order Quasi-Monte Carlo, parametric operator equations, infinite-dimensional quadrature, Bayesian inverse problems, Uncertainty Quantification, CBC construction, SPOD weights.

Contents

1 Introduction

In [9] we proposed and analyzed the convergence rates of higher order Quasi-Monte Carlo (HoQMC) approximations of conditional expectations which arise in Bayesian Inverse problems for partial differential equations (PDEs). We studied broad classes of forward problems with *distributed uncertain parametric input data.* Typical examples are elliptic or parabolic partial differential equations with uncertain, spatially inhomogeneous coefficients, but also differential and integral equations in uncertain domains of definition. Upon suitable *uncertainty parametrization,* and with a suitable Bayesian prior measure placed on the, in general, infinite-dimensional parameter space, the task of numerical evaluation of Bayesian estimates for quantities of interest (QoI's) becomes that of numerical computation of parametric, deterministic integrals over a high-dimensional parameter space. As an alternative to the Markov chain Monte Carlo (MCMC) method, in [20, 21] it was proposed to apply recently developed, dimension-adaptive Smolyak quadrature techniques to the evaluation of the corresponding integrals. In [9] we developed a convergence theory for HoQMC integration for the numerical evaluation of the corresponding integrals, based on our earlier work [6] on these methods in forward uncertainty quantification (UQ). In particular, we proved that *dimension-independent* convergence rates of order $> 1/2$ in terms of the number N of approximate solves of the forward problem can be achieved by replacing Monte Carlo or MCMC sampling of the Bayesian posterior with judiciously chosen, *deterministic* HoQMC quadratures. The achievable, dimension-independent rate in the proposed algorithms is, in principle, as high as the sparsity of the forward map permits, and "embarrassingly parallel": being QMC algorithms they access, unlike MCMC and sequential Monte Carlo (SMC) methods, the forward problem simultaneously and in parallel. The error analysis in [9] accounted for the quadrature error as well as for the errors incurred by a Petrov-Galerkin (PG) discretization and dimension truncation of the forward problem, but was performed for a single-level algorithm, i.e. the PG discretization of the forward model in all QMC quadrature points was based on the same subspace. As is well known in the context of Monte Carlo methods, *multilevel strategies* can lead to substantial gains in accuracy versus work. We refer to [13], and the references there, for a survey on multilevel Monte-Carlo methods. Multilevel discretizations for QMC integration were explored first for affine-parametric, linear forward problems in [18] and, in the context of HoQMC for parametric operator equations, in [7]. For the use of multilevel strategies in the context of MCMC methods for Bayesian inverse problems we refer to [10, 16] and the references there. The purpose of the present paper is to extend the convergence analysis of the HoQMC, Petrov-Galerkin approach from [9] to a multilevel algorithm.

As in our single-level HoQMC PG error analysis of Bayesian inversion in [9], we adopt the abstract setting of Bayesian inverse problems in infinite-dimensional function spaces from [4], and the convergence analysis of PG discretizations of abstract nonlinear problems from [19], as reviewed in [9]; *throughout the present work, we adopt the notations and terminology from* [9].

The principal contributions of the present work are as follows: we derive, for the possibly nonlinear, parametric operator equations with distributed, uncertain input considered in [9], multilevel extensions of high-order Petrov-Galerkin discretizations of the countably-parametric forward problem combined with HoQMC integration in the parameter domain. We provide a complete convergence analysis of the proposed algorithm, specifying, in particular, precise regularity and sparsity conditions on the forward response map which are sufficient to achieve a certain, dimension-independent convergence rate. Our analysis provides, in particular, information on the choice of algorithm parameters in applications: the convergence order of the PG discretization, also for functionals of the solution (not considered in [9]), the order of the interlaced polynomial lattice rule, the relation of HoQMC sample numbers N_ℓ and of truncation dimensions s_ℓ of the parameter space on the discretization level ℓ of the PG approximation of the forward problem. A major conclusion obtained with analytic continuation based error analysis from [8] is that *identical algorithmic steering parameters are admissible in our ML HoQMC algorithm for both, forward and inverse UQ.* With optimized HoQMC sample numbers, we also obtain asymptotic error vs. work bounds which indicate that the presently proposed algorithms outperform both, multilevel MC as well as the multilevel first order QMC algorithm considered in [18]; our analysis reveals that, analogous to sparse grid approximation, stronger regularity requirements on the parametric forward problem both in physical space and in parameter space are required.

In a suite of numerical experiments, we provide PDE examples in two space dimensions and with parameter spaces of dimension of several thousand where the presently proposed ML algorithms outperform all mentioned methods in terms of error vs. work, starting at relative errors as large as 10%.

The outline of this paper is as follows. In Section 2, we review the abstract, nonlinear parametric

operator equations with uncertain input data, the parametrization of the uncertain input data; and their PG approximations. The setting is analogous to that from [9] and the presentation is thus synoptic, and analogous and with references to [9, Sections 2, 3]. In Section 3, we review the general theory of well-posed Bayesian inverse problems in function spaces, from [4]. Again, the material is analogous to what we used in [9]; however, the error analysis of the ML HoQMC method, being a form of sparse grid approximation, requires the analog of mixed regularity, which we develop. Section 4 contains the core new mathematical results of the present paper: several multilevel computable Bayesian estimators and their error analysis. In particular it contains the error vs. work analysis of the combined ML HoQMC Petrov-Galerkin algorithms. Section 5 presents specific examples of parametric forward problems, and verifies that they satisfy all hypotheses of our foregoing error analysis; specifically, we consider linear, affine-parametric diffusion problems in two space dimensions, in primal variational formulation with space of continuous, piecewise linear functions on regular triangulations. Section 6 contains numerical tests of the proposed estimators for forward and inverse UQ for the PDE problems considered in Section 5. The numerical results are in agreement with the theory.

2 Forward UQ for parametric operator equations

We review the notation and mathematical setting of forward and inverse UQ for a class of smooth, possibly nonlinear, parametric operator equations, for which we developed the error analysis of the single-level algorithm in [9]. We develop here the error analysis for the multilevel extension of the algorithms in [9] for a general class of forward problems given by smooth, nonlinear operator equations with distributed uncertain input data u taking values in a separable Banach space X. Upon uncertainty parametrization with an unconditional basis of X such as, for example, the Karhunen-Loève basis, both forward and (Bayesian) inverse problems become countably parametric, deterministic operator equations. The problems of forward and inverse UQ are reformulated as countably-parametric integration problems. In the present work, we focus on the latter and analyze the use of *deterministic, higher order Quasi Monte-Carlo* integration methods, from [6, 7, 8] and the references there, in multilevel algorithms for Bayesian estimation in partial differential equations with uncertain input.

2.1 Uncertainty parametrization

As in [9], we parametrize the distributed uncertain input data u. To this end, X is assumed to be a separable, infinite-dimensional Banach space with norm $\|\cdot\|_X$, which has an unconditional basis $\{\psi_j\}_{j\geq 1}$: $X = \operatorname{span}\{\psi_j : j \geq 1\}$. Let $\langle u \rangle \in X$ be fixed. Any $u \in X$ has the representation

$$u = u(\boldsymbol{y}) \sim \langle u \rangle + \sum_{j\geq 1} y_j \psi_j \quad \text{for some } \boldsymbol{y} = (y_j)_{j\geq 1} \in \mathbb{R}^{\mathbb{N}}, \tag{2.1}$$

where $\sim$ means convergence in the norm $\|\cdot\|_X$ of X. The Karhunen-Loève expansions (see, e.g., [25, 24, 26, 4]) are examples of representations (2.1). We remark that the representation (2.1) is not unique: rescaling y_j and ψ_j will not change u.

Assume that a smoothness scale $\{X_r\}_{r\geq 0}$, with $X = X_0 \supset X_1 \supset X_2 \supset ...$, and a $t \geq 0$ is being given as part of the problem specification. We restrict the uncertain inputs u to sets X_t with "higher regularity" in order to obtain convergence rate estimates for the discretization of the forward problem. Note that $u \in X_t$ often corresponds to stronger decay of the ψ_j in (2.1). We assume that the $\{\psi_j\}_{j\geq 1}$ are scaled such that

$$\sum_{j\geq 1} \|\psi_j\|_{X_r}^{p_r} < \infty, \quad \text{for some } 0 < p_r < 1 \text{ and } r = 0, \ldots, t. \tag{2.2}$$

We define the set $\tilde{X}$ by

$$\tilde{X} = \{u(\boldsymbol{y}) \in X : u(\boldsymbol{y}) \sim \langle u \rangle + \sum_{j\geq 1} y_j \psi_j : \boldsymbol{y} = (y_j)_{j\geq 1} \in U\}, \tag{2.3}$$

where $U = (-1, 1)^{\mathbb{N}}$. Further we define the sequences $\boldsymbol{b}_r = (b_{j,r})_{j\geq 1}$ by $b_{j,r} := \|\psi_j\|_{X_r}$ for $r = 0, \ldots, t$. If $t = 0$, we write $X_0 = X$ and $\boldsymbol{b}_0 = \boldsymbol{b}$. From (2.2) we have $\sum_{j\geq 1} b_{j,r}^{p_r} < \infty$ for $r = 0, \ldots, t$. We also assume

that the ψ_j are enumerated so that

$$b_{1,t} \geq b_{2,t} \geq \cdots \geq b_{j,t} \geq \cdots . \tag{2.4}$$

With a given unconditional basis $\{\psi_j\}_{j\geq 1}$, every realization $u \in X$ can be identified in a one-to-one fashion with the pair $(\langle u \rangle, \boldsymbol{y})$, where $\langle u \rangle$ denotes the *nominal instance* of the uncertain datum u and $\boldsymbol{y}$ is the coordinate vector of the unique representation (2.1).

2.2 Operator equations with uncertain input

We consider the abstract forward problem:

$$\text{given } u \in \tilde{X} : \text{ find } q \in \mathcal{X} \text{ s.t. } {}_{\mathcal{Y}'}\langle \mathcal{R}(u; q), v \rangle_{\mathcal{Y}} = 0 \quad \forall v \in \mathcal{Y}. \tag{2.5}$$

Here, $X, \mathcal{X}$ and $\mathcal{Y}$ are real, separable Banach spaces and $\mathcal{R} : X \times \mathcal{X} \to \mathcal{Y}'$ is the residual of a *forward operator*.

A solution q_0 of (2.5) is called *regular at u*, for a given $u \in \tilde{X}$, if and only if the map $\mathcal{R}(u; \cdot)$ is Fréchet differentiable with respect to q at q_0 and if the differential is an isomorphism between $\mathcal{X}$ and $\mathcal{Y}'$. We assume the map $\mathcal{R}(u; \cdot) : \mathcal{X} \to \mathcal{Y}'$ admits a family of regular solutions *locally, in an open neighborhood of the nominal parameter instance* $\langle u \rangle \in X$, so that the operator equations involving $\mathcal{R}(u; q)$ are well-posed. A particular structural assumption on $\mathcal{R}$ is the representation

$$\mathcal{R}(u; q) = A(u; q) - F(u) \quad \text{in} \quad \mathcal{Y}' \tag{2.6}$$

with Frechet differentiable maps $A : X \times \mathcal{X} \to \mathcal{Y}'$ and $F : X \to \mathcal{Y}'$. The set $\{(u, q(u)) : u \in \tilde{X}\} \subset X \times \mathcal{X}$ is called *regular branch of solutions* of (2.5) (in the sense of [2]) if

$$\begin{aligned} &\tilde{X} \ni u \mapsto q(u) \text{ is continuous as mapping from } X \to \mathcal{X} , \\ &\mathcal{R}(u; q(u)) = 0 \quad \text{in} \quad \mathcal{Y}' . \end{aligned} \tag{2.7}$$

The regular branch of solutions (2.7) is called *nonsingular* if, in addition, the differential

$$(D_q \mathcal{R})(u; q(u)) \in \mathcal{L}(\mathcal{X}, \mathcal{Y}') \text{ is an isomorphism from } \mathcal{X} \text{ onto } \mathcal{Y}', \text{ for all } u \in \tilde{X} . \tag{2.8}$$

Conditions for well-posedness of (2.5) are stated in the following proposition: for regular branches of nonsingular solutions given by (2.5) - (2.8), the nonsingularity of the differential $(D_q\mathcal{R})(u, q(u))$ implies the *inf-sup conditions*. If $\mathcal{Y}$ is reflexive and for some nominal input $\langle u \rangle \in X$ of the uncertainty, the operator equation (2.5) admits a *regular branch of solutions* (2.7) with $q_0 \in \mathcal{X}$ denoting the "nominal" solution corresponding to input $\langle u \rangle \in X$, the differential $D_q\mathcal{R}$ at $(\langle u \rangle, q_0)$ given by the bilinear map

$$\mathcal{X} \times \mathcal{Y} \ni (\varphi, \psi) \mapsto {}_{\mathcal{Y}'}\langle D_q\mathcal{R}(\langle u \rangle; q_0)\varphi, \psi \rangle_{\mathcal{Y}}$$

is boundedly invertible, *uniformly with respect to* $u \in \tilde{X}$. For every instance $u \in \tilde{X}$, there exists a unique, isolated solution $q(u)$ of (2.5), which is uniformly bounded in the sense that there exists a constant $C(\mathcal{R}, \tilde{X}) > 0$ such that

$$\sup_{u \in \tilde{X}} \|q(u)\|_{\mathcal{X}} \leq C(\mathcal{R}, \tilde{X}) . \tag{2.9}$$

The set $\{(u, q(u)) : u \in \tilde{X}\} \subset \tilde{X} \times \mathcal{X}$ is called a *regular branch of nonsingular solutions*.

At every point of the regular branch $\{(u, q(u)) : u \in \tilde{X}\} \subset \tilde{X} \times \mathcal{X}$, if the nonlinear functional $\mathcal{R}$ is Fréchet differentiable with respect to $u \in \tilde{X}$ and Fréchet differentiable with respect to q, then the mapping relating u to $q(u)$ within the branch of nonsingular solutions is locally Lipschitz on $\tilde{X}$:

$$\forall u, v \in \tilde{X} : \quad \|q(u) - q(v)\|_{\mathcal{X}} \leq L(\mathcal{R}, \tilde{X})\|u - v\|_X . \tag{2.10}$$

In what follows, we consider the abstract setting (2.5) with the assumption that the mapping $\mathcal{R}(u; q)$ is uniformly continuously differentiable with boundedly invertible differential in a product of neighborhoods $B_X(\langle u \rangle; R) \times B_{\mathcal{X}}(q(\langle u \rangle); R) \subset X \times \mathcal{X}$, where B denotes the ball with radius R, of sufficiently small radius $R > 0$. Then $q_0 = q(\langle u \rangle) \in \mathcal{X}$ is the corresponding unique, isolated solution of (2.5) at the nominal

uncertain input $\langle u \rangle \in X$. This turns (2.5) into an *equivalent, deterministic, countably parametric operator equation*: given $\boldsymbol{y} \in U$, find $q(\boldsymbol{y}) \in \mathcal{X}$ such that

$$\mathcal{R}(\boldsymbol{y}; q(\boldsymbol{y})) = 0 \quad \text{in} \quad \mathcal{Y}' . \tag{2.11}$$

We refer to [9, Remark. 2.1] for further discussion. Under (2.9) and (2.10), the operator equation (2.5) will admit, for every $\boldsymbol{y} \in U$, a unique solution $q(\boldsymbol{y}; F)$. This solution is, due to (2.9) and (2.10), isolated, and depends Lipschitz continuously on the parameter sequence $\boldsymbol{y} \in U$: there holds

$$\sup_{\boldsymbol{y} \in U} \|q(\boldsymbol{y})\|_{\mathcal{X}} \leq C(\mathcal{R}, U), \tag{2.12}$$

and, if the local Lipschitz condition (2.10) holds, there exists a Lipschitz constant $L > 0$ such that

$$\forall \boldsymbol{y}, \boldsymbol{y}' \in U : \quad \|q(\boldsymbol{y}) - q(\boldsymbol{y}')\|_{\mathcal{X}} \leq L(\mathcal{R}, U)\|u(\boldsymbol{y}) - u(\boldsymbol{y}')\|_X . \tag{2.13}$$

The Lipschitz constant $L > 0$ in (2.13) is not, in general, equal to $L(\mathcal{R}, \tilde{X})$ in (2.10): it depends on the nominal instance $\langle u \rangle \in X$ and on the choice of basis $\{\psi_j\}_{j \geq 1}$. *Unless explicitly stated otherwise, throughout what follows, we shall identify $q_0 = q(\boldsymbol{0}) \in \mathcal{X}$ with the solution of (2.5) at the nominal input* $\langle u \rangle \in X$.

Note that $u \in X_t$ implies that $\mathcal{R}(\cdot) \in \mathcal{Y}'_t$ and $q(u) \in \mathcal{X}_t$, with corresponding subspaces $\mathcal{X}_t \subset \mathcal{X}$ and $\mathcal{Y}'_t \subset \mathcal{Y}'$ with extra regularity from suitable scales. If $t = 0$, we write $\mathcal{X}_0 = \mathcal{X}$, $\mathcal{Y}_0 = \mathcal{Y}$, etc..

2.3 Dimension truncation

Dimension truncation is equivalent to setting $y_j = 0$ for $j > s$, for a given $s \in \mathbb{N}$, in (2.1). The truncated uncertain datum is denoted by $u^s \in X$. We denote by $q^s(\boldsymbol{y})$ the solution of the corresponding parametric weak problem (2.11). For $\boldsymbol{y} \in U$, define $\boldsymbol{y}_{\{1:s\}} := (y_1, y_2, ..., y_s, 0, 0, ...)$. Then unique solvability of (2.11) implies $q^s(\boldsymbol{y}) = q(\boldsymbol{y}_{\{1:s\}})$. Consider the s-term truncated parametric problem: given $u^s = u(\boldsymbol{y}_{\{1:s\}}) \in \tilde{X}$,

$$\text{find } q^s \in \mathcal{X} : \quad {}_{\mathcal{Y}'}\langle \mathcal{R}(u^s; q^s), w \rangle_{\mathcal{Y}} = 0 \ \ \forall w \in \mathcal{Y} . \tag{2.14}$$

As shown in [9, Prop. 2.2], [17, Thm. 5.1], under Assumption (2.2), and under the uniform regularity shift in (2.18) ahead, for every $F \in \mathcal{Y}'_t$, for every $\boldsymbol{y} \in U$ and for every $s \in \mathbb{N}$, the parametric solution $q^s(\boldsymbol{y})$ of the dimensionally truncated, parametric weak problem (2.11) with s-term truncated parametric expansion (2.1) satisfies, with $b_{t,j}$ as defined in (2.2),

$$\sup_{\boldsymbol{y} \in U} \|q(\boldsymbol{y}) - q^s(\boldsymbol{y})\|_{\mathcal{X}_t} \leq C(\mathcal{R}, X) \sum_{j \geq s+1} b_{t,j} . \tag{2.15}$$

Moreover, for every $G \in \mathcal{X}'_t$, there exists $\theta \in \{1, 2\}$ such that

$$|I(G(q)) - I(G(q^s))| \leq \tilde{C} \left(\sum_{j \geq s+1} b_{t,j} \right)^{\theta} \tag{2.16}$$

where

$$I(G(q)) = \int_U G(q(\boldsymbol{y})) \, \boldsymbol{\pi}(\mathrm{d}\boldsymbol{y}) \text{ and } I(G(q^s)) = \int_{[-1,1]^s} G(q(y_1, \ldots, y_s, 0, \ldots)) \, \boldsymbol{\pi}(\mathrm{d}y_1 \cdots \mathrm{d}y_s),$$

for some constant $\tilde{C} > 0$ independent of s.

Remark 2.1. In (2.16), generally $\theta = 1$. There holds $\theta = 2$ if $\int_{-1}^{1} y_j \pi_j(\mathrm{d}y_j) = 0$, as, e.g. for the uniform measure (see, e.g. [17, Thm. 5.1]). If conditions (2.2) and (2.4) hold, then in (2.15) and (2.16)

$$\sum_{j \geq s+1} b_{t,j} \leq \min\left(\frac{1}{1/p_t - 1}, 1\right) \left(\sum_{j \geq 1} b_{t,j}^{p_t}\right)^{1/p_t} s^{-(1/p_t - 1)} . \tag{2.17}$$

2.4 Petrov-Galerkin discretization

Based on the theory in [14, Chap. IV.3] and in [19], an error analysis of Petrov-Galerkin discretizations of (2.11) for the approximation of regular branches of solutions of smooth, nonlinear forward problems (2.5) was developed in [9], and we review the general setting from [9]. As in [23, 6], we assume that we are given two sequences $\{\mathcal{X}^h\}_{h>0} \subset \mathcal{X}$ and $\{\mathcal{Y}^h\}_{h>0} \subset \mathcal{Y}$ of finite dimensional subspaces which are dense in $\mathcal{X}$ and in $\mathcal{Y}$, respectively. For the computational complexity analysis, we assume *regularity properties* and *approximation properties*:

uniform parametric regularity property (UPR): there are scales $\{\mathcal{X}_t\}_{t\geq 0}$ and $\{\mathcal{Y}_t\}_{t\geq 0}$ of function spaces such that $\mathcal{X}_{t'} \subset \mathcal{X}_t \subset \mathcal{X}_0 = \mathcal{X}$ and $\mathcal{X}'_{t'} \subset \mathcal{X}'_t \subset \mathcal{X}'_0 = \mathcal{X}'$ for any $0 < t < t' < \infty$ and analogously for $\mathcal{Y}_t$, such that there holds the uniform regularity shift: for $\mathcal{R}$ as in (2.6), for every $F \in \mathcal{Y}'_t$, $G \in \mathcal{X}'_t$, the parametric solutions $q(\boldsymbol{y}) = (A(\boldsymbol{y}))^{-1}F$ and $q^*(\boldsymbol{y}) = (A^*(\boldsymbol{y}))^{-1}G$, where A^* is the conjugate operator of A, satisfy regularity resp. adjoint regularity shifts which are uniform w.r. to $\boldsymbol{y}$, i.e.

$$\sup_{\boldsymbol{y} \in U} \|q(\boldsymbol{y})\|_{\mathcal{X}_t} \leq C(t)\|F\|_{\mathcal{Y}'_t} \ , \quad \sup_{\boldsymbol{y} \in U} \|q^*(\boldsymbol{y})\|_{\mathcal{Y}_t} \leq C(t)\|G\|_{\mathcal{X}'_t} \ . \tag{2.18}$$

We also assume **approximation properties**: for $0 < t, t' \leq \bar{t}$ and for $0 < h \leq h_0$ holds

$$\inf_{w^h \in \mathcal{X}^h} \|w - w^h\|_{\mathcal{X}} \ \leq \ C_t \, h^t \, \|w\|_{\mathcal{X}_t} \ , \quad \inf_{v^h \in \mathcal{Y}^h} \|v - v^h\|_{\mathcal{Y}} \ \leq \ C_{t'} \, h^{t'} \, \|v\|_{\mathcal{Y}_{t'}} \ . \tag{2.19}$$

Assume that the subspace range $\{\mathcal{X}^h\}_{h>0} \subset \mathcal{X}$ and $\{\mathcal{Y}^h\}_{h>0} \subset \mathcal{Y}$ are stable, i.e., there exist $\bar{\mu} > 0$ and $h_0 > 0$ such that for every $0 < h \leq h_0$, there hold the uniform (with respect to $\boldsymbol{y} \in U$) discrete inf-sup conditions

$$\forall \boldsymbol{y} \in U: \quad \inf_{0 \neq v^h \in \mathcal{X}^h} \sup_{0 \neq w^h \in \mathcal{Y}^h} \frac{_{\mathcal{Y}'}\langle (D_q\mathcal{R})(u(\boldsymbol{y}); q_0)v^h, w^h \rangle_{\mathcal{Y}}}{\|v^h\|_{\mathcal{X}}\|w^h\|_{\mathcal{Y}}} \geq \bar{\mu} > 0 \ , \tag{2.20}$$

$$\forall \boldsymbol{y} \in U: \quad \inf_{0 \neq w^h \in \mathcal{Y}^h} \sup_{0 \neq v^h \in \mathcal{X}^h} \frac{_{\mathcal{Y}'}\langle (D_q\mathcal{R})(u(\boldsymbol{y}); q_0)v^h, w^h \rangle_{\mathcal{Y}}}{\|v^h\|_{\mathcal{X}}\|w^h\|_{\mathcal{Y}}} \geq \bar{\mu} > 0 \ . \tag{2.21}$$

Then, for every $0 < h \leq h_0$ the Galerkin approximations: given $\boldsymbol{y} \in U$,

$$\text{find } q_h(\boldsymbol{y}) \in \mathcal{X}^h: \quad _{\mathcal{Y}'}\langle \mathcal{R}(\boldsymbol{y}; q_h(\boldsymbol{y})), w^h \rangle_{\mathcal{Y}} = 0 \quad \forall w^h \in \mathcal{Y}^h \ , \tag{2.22}$$

are uniquely defined and converge quasioptimally: there exists a constant $C > 0$ such that for all $\boldsymbol{y} \in U$

$$\|q(\boldsymbol{y}) - q_h(\boldsymbol{y})\|_{\mathcal{X}} \ \leq \ \frac{C}{\bar{\mu}} \inf_{0 \neq v^h \in \mathcal{X}^h} \|q(\boldsymbol{y}) - v^h\|_{\mathcal{X}} \ . \tag{2.23}$$

If $q(\boldsymbol{y}) \in \mathcal{X}_t$ for all $\boldsymbol{y} \in U$ and if (2.19) holds, then for every $\boldsymbol{y} \in U$

$$\|q(\boldsymbol{y}) - q_h(\boldsymbol{y})\|_{\mathcal{X}} \ \leq \ \frac{C}{\bar{\mu}} h^t \sup_{\boldsymbol{y} \in U} \|q(\boldsymbol{y})\|_{\mathcal{X}_t} \ . \tag{2.24}$$

Moreover, for sufficiently large truncation dimension $s \in \mathbb{N}$, for given $\boldsymbol{y}_{\{1:s\}} \in U$ the dimensionally truncated Galerkin approximations

$$\text{find } q_h(\boldsymbol{y}_{\{1:s\}}) \in \mathcal{X}^h: \quad _{\mathcal{Y}'}\langle \mathcal{R}(\boldsymbol{y}_{\{1:s\}}; q_h(\boldsymbol{y}_{\{1:s\}})), w^h \rangle_{\mathcal{Y}} = 0 \quad \forall w^h \in \mathcal{Y}^h \ , \tag{2.25}$$

admit unique solutions $q_h(\boldsymbol{y}_{\{1:s\}}) \in \mathcal{X}^h$ which converge, as $h \downarrow 0$, quasioptimally to $q(\boldsymbol{y}_{\{1:s\}}) \in \mathcal{X}$, i.e. (2.23) and (2.24) hold with $\boldsymbol{y}_{\{1:s\}}$ in place of $\boldsymbol{y}$, with the same constants $C > 0$ and $\bar{\mu}$ independent of s and of h.

3 Bayesian inverse UQ and HoQMC approximation

In this section, the abstract (possibly nonlinear) operator equation (2.5) is considered again. The system's forcing $F \in \mathcal{Y}'$ is allowed to depend on the uncertain input u and the uncertain operator $A(u; \cdot) \in \mathcal{L}(\mathcal{X}, \mathcal{Y}')$ is assumed to be boundedly invertible, at least locally for the uncertain input u, sufficiently close to a

nominal input $\langle u \rangle \in X$. We define the *forward response map*, which maps a given uncertain input u and a given forcing F in (2.6) to the response q in (2.5) by

$$G : X \times \mathcal{Y}' \to \mathcal{X} : \quad G(u, F) := q(u) .$$

In the general case (2.5), we omit F and denote the dependence of the forward solution on the uncertain input as $G(u) = q(u)$. A *bounded linear observation operator* on the space $\mathcal{X}$ of observed system responses in Y is given and denoted by $\mathcal{O} : \mathcal{X} \to Y$. Throughout the remainder of this paper, we assume that $Y = \mathbb{R}^K$ with $K < \infty$, then $\mathcal{O} \in \mathcal{L}(\mathcal{X}; Y) \simeq (\mathcal{X}^*)^K$. The space $Y = \mathbb{R}^K$ is equipped with the Euclidean norm, denoted by $|\cdot|$.

We assume the following form of observed data, composed of the observed system response and the additive noise η

$$\delta = \mathcal{O}(G(u)) + \eta \in Y . \tag{3.1}$$

The additive observation noise process η is assumed to be Gaussian, i.e. a random vector $\eta \sim \mathbb{Q}_0 = \mathcal{N}(0, \Gamma)$ with a positive definite covariance Γ on $\mathbb{R}^K$.

The *uncertainty-to-observation map* $\mathcal{G} : X \to \mathbb{R}^K$ then reads $\mathcal{G} = \mathcal{O} \circ G$, so that

$$\delta = \mathcal{G}(u) + \eta = \mathcal{O}(G(u)) + \eta. \tag{3.2}$$

When η varies randomly, δ varies in $L_\Gamma^2(\mathbb{R}^K)$, the space of random vectors taking values in $Y = \mathbb{R}^K$ which are square integrable with respect to the Gaussian measure with covariance matrix Γ. Bayes' formula [26, 4] yields a density of the Bayesian posterior with respect to the prior whose negative log-likelihood equals the observation noise covariance-weighted, least squares functional (also referred to as "potential" in what follows) $\Phi_\Gamma : X \times Y \to \mathbb{R}$ by

$$\Phi_\Gamma(u; \delta) = \frac{1}{2}|\delta - \mathcal{G}(u)|_\Gamma^2 := \frac{1}{2}\left((\delta - \mathcal{G}(u))^\top \Gamma^{-1}(\delta - \mathcal{G}(u))\right) . \tag{3.3}$$

3.1 Well-posedness and approximation

Unlike deterministic inverse problems where the data-to-solution maps can be severely ill-posed, for a positive definite covariance Γ, the expectations (3.19) are Lipschitz continuous with respect to the data δ, *provided that the potential Φ_Γ in (3.3) is locally Lipschitz with respect to the data δ* in the following sense.

Assumption 1. *[4, Assumption 4.2] Let $\tilde{X} \subseteq X$ **denote a subset of admissible uncertain input data,** and assume that $\Phi_\Gamma \in C(\tilde{X} \times Y; \mathbb{R})$ is Lipschitz on bounded sets.*

Assume also that there exist functions $\mathcal{M}_i : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$, $i = 1, 2$, (depending on $\Gamma > 0$) which are monotone, non-decreasing separately in each argument, and with $\mathcal{M}_2$ strictly positive, such that for all $u \in \tilde{X}$, and for all $\delta, \delta_1, \delta_2 \in B_Y(0, r)$

$$\Phi_\Gamma(u; \delta) \geq -\mathcal{M}_1(r, \|u\|_X), \tag{3.4}$$

and

$$|\Phi_\Gamma(u; \delta_1) - \Phi_\Gamma(u; \delta_2)| \leq \mathcal{M}_2(r, \|u\|_X)|\delta_1 - \delta_2| . \tag{3.5}$$

In [4, Thm. 4.3], the following infinite-dimensional version of Bayes' rule is shown.

Proposition 3.1. *Let Assumption 1 hold. Assume further that $\boldsymbol{\pi}(\tilde{X}) = 1$ and that $\boldsymbol{\pi}(\tilde{X} \cap B) > 0$ for some bounded set $B \subset X$. Assume additionally that, for every fixed $r > 0$,*

$$\exp(\mathcal{M}_1(r, \|u\|_X)) \in L_{\boldsymbol{\pi}}^1(X; \mathbb{R}) .$$

Then
(i) for $\mathbb{Q}_0$-a.e. data $\delta \in Y$,

$$Z := \int_X \exp\left(-\Phi_\Gamma(u; \delta)\right) \boldsymbol{\pi}(\mathrm{d}u) > 0 ,$$

(ii) the conditional distribution of $u|\delta$ (u given δ) exists and is denoted by $\boldsymbol{\pi}^\delta$. It is absolutely continuous with respect to $\boldsymbol{\pi}$ and there holds

$$\frac{d\boldsymbol{\pi}^\delta}{d\boldsymbol{\pi}}(u) = \frac{1}{Z} \exp\left(-\Phi_\Gamma(u; \delta)\right) . \tag{3.6}$$

In particular, the (Radon-Nikodym) derivative of the Bayesian posterior w.r.t. the prior measure admits a bounded density w.r.t. the prior $\boldsymbol{\pi}$, which is given by (3.6).

We remark that in the present context, (3.5) follows from (3.3); for convenient reference, we include (3.5) in Assumption 1. Under Assumption 1, the expectation (3.19) depends Lipschitz continuously on the data δ (see, e.g., [4, Sec. 4.1] for a proof):

$$\forall \phi \in L^2(\boldsymbol{\pi}^{\delta_1}, X; \mathbb{R}) \cap L^2(\boldsymbol{\pi}^{\delta_2}, X; \mathbb{R}) : \quad \|\mathbb{E}^{\boldsymbol{\pi}^{\delta_1}}[\phi] - \mathbb{E}^{\boldsymbol{\pi}^{\delta_2}}[\phi]\|_{\mathcal{Z}} \leq C(\Gamma, r)|\delta_1 - \delta_2| . \tag{3.7}$$

For $\delta \in Y$, the Bayesian posterior $\boldsymbol{\pi}_h^{\delta}$ with respect to the approximate potential $\Phi_{\Gamma}^{\ell} = \Phi_{\Gamma}^{(s_{\ell}, h_{\ell})}$ obtained from dimension truncation at dimension s_{ℓ} and PG discretization (2.25), with $O(h_{\ell}^{-d})$ degrees of freedom, is well-defined. In what follows, we investigate the impact of approximation errors in the forward response of the system (e.g. due to discretization and approximate numerical solution of system responses) on the Bayesian predictions (3.19). To bound the errors of the Bayesian expectations (3.19) as in [9], we work under

Assumption 2. *[4, Assumption 4.7] Let $\tilde{X} \subseteq X$ denote a subset of admissible uncertain input data, assume that $\boldsymbol{\pi}(\tilde{X}) = 1$, and let the Bayesian potential $\Phi_{\Gamma} \in C(\tilde{X}; \mathbb{R})$ be Lipschitz on bounded sets. Assume also that there exist functions $\mathcal{M}_i : \mathbb{R}_+ \to \mathbb{R}_+$, $i = 1, 2$, independent of the number $M = M_h = O(h^{-d})$ of degrees of freedom in the PG discretization of the forward problem, where the functions $\mathcal{M}_i$ are monotonically non-decreasing separately in each argument, and with $\mathcal{M}_2$ strictly positive, such that for all $u \in \tilde{X}$ and for all $\delta \in B_Y(0, r)$,*

$$\Phi_{\Gamma}(u; \delta) \geq -\mathcal{M}_1(\|u\|_X), \tag{3.8}$$

and there is a positive, monotonically decreasing $\varphi(\cdot)$ such that $\varphi(M) \to 0$ as $M \to \infty$, monotonically and uniformly w.r.t. $u \in \tilde{X}$ (resp. w.r.t. $\boldsymbol{y} \in U$) and such that

$$|\Phi_{\Gamma}(u; \delta) - \Phi_{\Gamma}^M(u; \delta)| \leq \mathcal{M}_2(\|u\|_X)\varphi(M) . \tag{3.9}$$

Proposition 3.2. *Suppose that Assumption 2 holds, and assume that for $\tilde{X} \subseteq X$ and for some bounded set $B \subset X$ we have $\boldsymbol{\pi}(\tilde{X} \cap B) > 0$ and*

$$X \ni u \mapsto \exp(\mathcal{M}_1(\|u\|_X))(1 + \mathcal{M}_2^2(\|u\|_X)) \in L_{\boldsymbol{\pi}}^1(X; \mathbb{R}) .$$

Then there holds, for every QoI $\phi : X \to \mathcal{Z}$ such that $\phi \in L_{\boldsymbol{\pi}^{\delta}}^2(X; \mathcal{Z}) \cap L_{\boldsymbol{\pi}_M^{\delta}}^2(X; \mathcal{Z})$ uniformly w.r.t. M and such that $Z > 0$ in (3.17), the consistency error bound

$$\|\mathbb{E}^{\boldsymbol{\pi}^{\delta}}[\phi] - \mathbb{E}^{\boldsymbol{\pi}_M^{\delta}}[\phi]\|_{\mathcal{Z}} \leq C(\Gamma, r)\varphi(M) . \tag{3.10}$$

A proof of Proposition 3.2 can be found in [4, Thm. 4.9, Rem. 4.10]. The consistency condition (3.9) in either of these cases follows from [9, Prop. 3.2]. Assume we are given a sequence $\{q^M\}_{M \geq 1}$ of approximations to the forward response $X \ni u \mapsto q(u) \in \mathcal{X}$ such that, with the parametrization (2.1),

$$\sup_{u \in \tilde{X}} \|(q - q^M)(\boldsymbol{y})\|_{\mathcal{X}} \leq \varphi(M) \tag{3.11}$$

with a consistency error bound $\varphi(M) \downarrow 0$ as in Assumption 2. Denote by G^M the corresponding (Petrov-Galerkin) approximations of the parametric forward maps. Then it was shown in [9, Prop. 3.2] that the approximate Bayesian potential

$$\Phi_{\Gamma}^M(u; \delta) = \frac{1}{2}(\delta - \mathcal{G}^M(u))^{\top}\Gamma^{-1}(\delta - \mathcal{G}^M(u)) : X \times Y \to \mathbb{R}, \tag{3.12}$$

where $\mathcal{G}^M := \mathcal{O} \circ G^M$, satisfies (3.9). In the ensuing error analysis of the combined HoQMC, PG approximation, q^M in (3.12) will denote the Petrov-Galerkin discretization (2.25) in Section 2.4, based on s-dimensional truncation of the parameter space, and on finite dimensional subspaces with $M = M_h = \dim(\mathcal{X}^h) = \dim(\mathcal{Y}^h)$ many degrees of freedom. Typically, for subspaces of piecewise polynomial functions obtained by isotropic mesh refinement in a bounded, physical domain $D \subset \mathbb{R}^d$, $M_h = O(h^{-d})$; in sparse grid discretizations in D (not considered in detail here, but covered by the present theory), $M_h = O(h^{-1}|\log h|^{d-1})$. In that case, $\mathcal{X}_t$ and $\mathcal{Y}_t$ are function spaces of mixed partial derivatives.

In applying QMC to Bayesian estimation, we shall be concerned with integration of parametric integrand functions with respect to a probability measure $\boldsymbol{\pi}$ on the set U of all parameters characterizing the uncertain input data. Typically, $\boldsymbol{\pi}$ will denote a Bayesian prior on U. Since U is a cartesian product of intervals, we *assume* that $\boldsymbol{\pi}$ is a product probability measure, i.e. $\boldsymbol{\pi}(\mathrm{d}\boldsymbol{y}) = \prod_{j \geq 1} \pi_j(\mathrm{d}y_j)$.

3.2 Parametric Bayesian posterior

As explained in [20, 21, 9] and in Section 2.1, we consider the parametric, deterministic forward problem in the probability space

$$(U, \mathcal{B}, \boldsymbol{\pi}) . \tag{3.13}$$

We assume throughout what follows that the prior $\boldsymbol{\pi}$ on the uncertain input $u \in X$, parametrized in the form (2.1), satisfies $\boldsymbol{\pi}(\tilde{X}) = 1$. We also assume the prior measure $\boldsymbol{\pi}$ being uniform, and the sequences $\boldsymbol{y}$ in (2.1) taking values in the parameter domain $U = [-1, 1]^{\mathbb{N}}$. With the parameter domain U as in (3.13) the parametric uncertainty-to-observation map $\Xi : U \to Y = \mathbb{R}^K$ reads

$$\Xi(\boldsymbol{y}) = \mathcal{G}(u)\Big|_{u=\langle u \rangle + \sum_{j \in \mathbb{N}} y_j \psi_j} . \tag{3.14}$$

In order to apply our QMC quadrature, we need a parametric version of Bayes' Theorem, as stated in Prop. 3.1, in terms of the uncertainty parametrization (2.1). We view U as the unit ball in $\ell^\infty([-1,1]^{\mathbb{N}}$, the Banach space of bounded sequences taking values in U.

Proposition 3.3. *Assume that $\Xi : U \to Y = \mathbb{R}^K$ is bounded and continuous. Then $\boldsymbol{\pi}^\delta(\mathrm{d}\boldsymbol{y})$, the distribution of $\boldsymbol{y} \in U$ given data $\delta \in Y$, is absolutely continuous with respect to $\boldsymbol{\pi}(\mathrm{d}\boldsymbol{y})$, i.e. there exists a parametric density $\Theta(\boldsymbol{y})$ such that for every $\boldsymbol{y} \in U$*

$$\frac{d\boldsymbol{\pi}^\delta}{d\boldsymbol{\pi}}(\boldsymbol{y}) = \frac{1}{Z}\Theta(\boldsymbol{y}) \tag{3.15}$$

with the posterior density $\Theta(\boldsymbol{y})$ given by

$$\Theta(\boldsymbol{y}) = \exp\left(-\Phi_\Gamma(u;\delta)\right)\Big|_{u=\langle u \rangle + \sum_{j \in \mathbb{N}} y_j \psi_j} . \tag{3.16}$$

Here the Bayesian potential Φ_Γ is as in (3.3) and

$$Z = \mathbb{E}^{\boldsymbol{\pi}}[1] = \int_U \Theta(\boldsymbol{y})\,\boldsymbol{\pi}(\mathrm{d}\boldsymbol{y}) > 0 . \tag{3.17}$$

Bayesian estimation is concerned with the approximation of a "most likely" Quantity of Interest (QoI) $\phi : X \to \mathcal{Z}$ (which may take values in a Banach space $\mathcal{Z}$), conditional on given (noisy) observation data $\delta \in Y$. In particular, the choice $\phi(u) = G(u) = q(u)$ (with $\mathcal{Z} = \mathcal{X}$) facilitates computation of the "most likely" (as expectation under the posterior, given data δ) system response. Based on Prop. 3.3 we associate with the QoI ϕ the deterministic, infinite-dimensional, parametric map

$$\Psi(\boldsymbol{y}) = \Theta(\boldsymbol{y})\phi(u)\big|_{u=\langle u \rangle + \sum_{j \in \mathbb{N}} y_j \psi_j} = \exp\left(-\Phi_\Gamma(u;\delta)\right)\phi(u)\Big|_{u=\langle u \rangle + \sum_{j \in \mathbb{N}} y_j \psi_j} : U \to \mathcal{Z} . \tag{3.18}$$

With the density $\Psi(\boldsymbol{y})$ in (3.18), the Bayesian estimate of the QoI ϕ, given noisy data δ, takes the form

$$\mathbb{E}^{\boldsymbol{\pi}^\delta}[\phi] = Z'/Z, \quad Z' := \mathbb{E}^{\boldsymbol{\pi}}[\Psi] = \int_U \Psi(\boldsymbol{y})\,\boldsymbol{\pi}(\mathrm{d}\boldsymbol{y}) . \tag{3.19}$$

3.3 Higher order QMC integration

In general, the integrals Z and Z' in (3.19) cannot be evaluated exactly. Hence we propose to approximate these integrals by a HoQMC quadrature rule. In the following, we assume the prior density $\boldsymbol{\pi}$ to be the uniform density. Supposing $\{\boldsymbol{y}_0, \ldots, \boldsymbol{y}_{N-1}\} \subset U = [0,1]^s$ is a collection of N QMC points, we define the QMC estimate of the integral of a function $f : U \to \mathcal{Z}$ with respect to the uniform measure $\boldsymbol{\pi}$ by the following equal weight quadrature rule in s dimensions,

$$Q_{N,s}[f] := \frac{1}{N}\sum_{j=0}^{N-1} f(\boldsymbol{y}_j) . \tag{3.20}$$

We shall analyze, in particular, $Q_{N,s}$ being *deterministic, interlaced higher order polynomial lattice rules* as introduced in [5] and as considered for affine-parametric operator equations in [7].

To generate a polynomial lattice rule with $N = b^m$ points (where b is a given prime number, m is a given positive integer), we need a *generating vector* of polynomials $\mathbf{g}(x) = (g_1(x), \ldots, g_s(x))$ where each $g_j(x)$ is a polynomial with degree $< m$ with its coefficients taken from a finite field $\mathbb{Z}_b$.

For each integer $n = 0, \ldots, b^m - 1$, we associate n with the polynomial

$$n(x) = \sum_{r=0}^{m-1} \eta_r x^r \quad \in \mathbb{Z}_b[x],$$

where $(\eta_{m-1}, \ldots, \eta_0)$ is the b-adic expansion of n, that is $n = \eta_0 + \eta_1 b + \ldots + \eta_{m-1} b^{m-1}$. We also need a map v_m which maps elements in $\mathbb{Z}_b(x^{-1})$ to the interval $[0, 1)$, defined for any integer w by

$$v_m \left(\sum_{\ell=w}^{\infty} t_\ell x^{-\ell} \right) = \sum_{\ell=\max(1,w)}^{m} t_\ell b^{-\ell}.$$

Let $P \in \mathbb{Z}[x]$ be an irreducible polynomial with degree m. The classical polynomial lattice rule $\mathcal{S}_{P,b,m,s}(\mathbf{g})$ associated with P and the generating vector $\mathbf{g}$ is comprised of the quadrature points

$$\boldsymbol{y}_n = \left(v_m \left(\frac{n(x)g_1(x)}{P(x)} \right), \ldots, v_m \left(\frac{n(x)g_s(x)}{P(x)} \right) \right) \in [0,1)^s, \quad n = 0, \ldots, b^m - 1.$$

Classical polynomial lattice rules give almost first order of convergence for integrands of bounded variation. To obtain higher order of convergence, an interlacing procedure described as follows is needed. The *digit interlacing function* with digit interlacing factor $\alpha \in \mathbb{N}$ is given by

$$\mathscr{D}_\alpha : [0,1)^\alpha \quad \to \quad [0,1) \; : \; (x_1, \ldots, x_\alpha) \mapsto \sum_{a=1}^{\infty} \sum_{j=1}^{\alpha} \xi_{j,a} b^{-j-(a-1)\alpha} , \tag{3.21}$$

where $x_j = \xi_{j,1} b^{-1} + \xi_{j,2} b^{-2} + \cdots$ for $1 \leq j \leq \alpha$. For vectors, we set

$$\begin{aligned} \mathscr{D}_\alpha : [0,1)^{\alpha s} \quad &\to \quad [0,1)^s \\ (x_1, \ldots, x_{\alpha s}) \quad &\mapsto \quad (\mathscr{D}_\alpha(x_1, \ldots, x_\alpha), \ldots, \mathscr{D}_\alpha(x_{(s-1)\alpha+1}, \ldots, x_{s\alpha})) . \end{aligned} \tag{3.22}$$

Then, an *interlaced polynomial lattice rule of order α with b^m points in s dimensions* is a QMC rule using $\mathscr{D}_\alpha(\mathcal{S}_{P,b,m,\alpha s}(\mathbf{g}))$ as quadrature points, for some given modulus P and generating vector $\mathbf{g}$.

3.4 Holomorphic parameter dependence

In the QMC quadrature error analysis, for each parameter $\boldsymbol{y}_j$, for $j = 1, \ldots, N$, we assume the parametric family of solutions admits a holomorphic extension into the complex domain $\mathbb{C}$ and these extensions must satisfy some uniform bounds. We recall from [15, 3] the notion of $(\boldsymbol{b}, p, \varepsilon)$-*holomorphy of parametric solutions* introduced to this end. *In the remainder of Section 3.4 all spaces X, Y, $\mathcal{X}$ and $\mathcal{Y}$ will be understood as Banach spaces over $\mathbb{C}$, without notationally indicating so.*

Definition 3.1. (($\boldsymbol{b}, p, \varepsilon$)-holomorphy) Let $\varepsilon > 0$ and $0 < p < 1$ be given. For a positive sequence $\boldsymbol{b} = (b_j)_{j\geq 1} \in \ell^p(\mathbb{N})$, a parametric mapping $g : U \to \mathcal{X}$ satisfies the $(\boldsymbol{b}, p, \varepsilon)$-*holomorphy assumption* if and only if all of the following conditions hold:

1. The map $\boldsymbol{y} \mapsto g(\boldsymbol{y})$ from U to $\mathcal{X}$, for each $\boldsymbol{y} \in U$, is uniformly bounded with respect to the parameter sequence $\boldsymbol{y}$, i.e. there is a bound $C_0 > 0$ such that

$$\sup_{\boldsymbol{y} \in U} \|g(\boldsymbol{y})\|_{\mathcal{X}} \leq C_0 . \tag{3.23}$$

2. For any sequence $\boldsymbol{\rho} := (\rho_j)_{j\geq 1}$ of numbers $\rho_j > 1$ that is $(\boldsymbol{b}, \varepsilon)$-*admissible*, i.e.

$$\sum_{j\geq 1} (\rho_j - 1)b_j \leq \varepsilon, \tag{3.24}$$

for sufficiently small $\varepsilon > 0$, the parametric solution map $U \ni \boldsymbol{y} \mapsto g(\boldsymbol{y})$ admits a continuous extension $\boldsymbol{z} \mapsto g(\boldsymbol{z})$ to the complex domain with respect to each variable z_j that is holomorphic in a cylindrical set of the form $O_{\boldsymbol{\rho}} := \bigotimes_{j\geq 1} O_{\rho_j}$ where, for every $j \geq 1$, $[-1,1] \subset O_{\rho_j} \subset \mathbb{C}$ is an open set $O_{\rho_j} \subset \mathbb{C}$.

3. For any poly-radius $\boldsymbol{\rho}$ satisfying (3.24), there is a second family $\tilde{O}_{\boldsymbol{\rho}} := \bigotimes_{j\geq 1} \tilde{O}_{\rho_j}$ of open, cylindrical sets

$$[-1,1] \subset O_{\rho_j} \subset \tilde{O}_{\rho_j} \subset \mathbb{C}$$

(strict inclusions), such that the extension is bounded on $\overline{\tilde{O}_{\boldsymbol{\rho}}}$ according to

$$\sup_{\boldsymbol{z}\in\tilde{O}_\rho} \|g(\boldsymbol{z})\|_{\mathcal{X}} \leq C_\varepsilon \,, \tag{3.25}$$

where the bounds $C_\varepsilon > 0$ depend on ε, but are independent of $\boldsymbol{\rho}$.

The notion of $(\boldsymbol{b}, p, \varepsilon)$-holomorphy depends implicitly on the choice of sets $\tilde{O}_{\rho_j}$. For the derivative bounds which arise in connection with the HoQMC error analysis (see, e.g., [6, 17]), as in [8], we use the following family of continuation domains: for $\kappa > 1$, consider by $\mathcal{T}_\kappa$ the set

$$\mathcal{T}_\kappa = \{z \in \mathbb{C} | \mathrm{dist}(z, [-1,1]) \leq \kappa - 1\} = \bigcup_{-1 \leq y \leq 1} \{z \in \mathbb{C} | |z - y| \leq \kappa - 1\} \,. \tag{3.26}$$

Then, once more, for a poly-radius $\boldsymbol{\rho}$ satisfying (3.24), we denote by $\mathcal{T}_{\boldsymbol{\rho}}$ the corresponding cylindrical set $\mathcal{T}_{\boldsymbol{\rho}} := \bigotimes_{j\geq 1} \mathcal{T}_{\rho_j} \subset \mathbb{C}^{\mathbb{N}}$. We refer to [15, 3] and the references there for further examples of parametric forward problems whose solutions admit $(\boldsymbol{b}, p, \varepsilon)$-holomorphic extensions.

3.5 HoQMC convergence for $(\boldsymbol{b}, p, \varepsilon)$-holomorphic integrands

In this section, we study the dependence of the solution $\boldsymbol{y} \mapsto q(\boldsymbol{y}) \in \mathcal{X}$ of the parametric, variational forward problem (2.5) on the parameter sequence $\boldsymbol{y}$. Precise bounds on the growth of the partial derivatives of $q(\boldsymbol{y})$ with respect to $\boldsymbol{y}$ imply, as in [17], dimension independent convergence rates for higher order QMC quadratures.

We are interested in the case where the integrand g is a composition of a continuous, linear functional $\mathcal{O}(\cdot) \in \mathcal{X}'$ with the (Petrov-)Galerkin approximation $q_h^s(2\boldsymbol{y} - \mathbf{1})$ of the dimension-truncated, parametric and $(\boldsymbol{b}, p, \varepsilon)$-holomorphic, operator equation (2.5). For every $s \in N$, the dimension-truncated integrand functions $g(\boldsymbol{y}) := (\mathcal{O} \circ q^s)(\boldsymbol{y}_{\{1:s\}})$ are $(\boldsymbol{b}, p, \varepsilon)$-holomorphic *uniformly w.r.t.* $s \in \mathbb{N}$ (see Section 2.3). Based on [6, Sec. 3] and [8, Prop. 4.1], the following result provides dimension-independent convergence rates of QMC quadratures based on higher order digital nets for integrand functions g which satisfy certain derivative bounds.

Proposition 3.4. *Let $s \geq 1$ and $N = b^m$ for $m \geq 1$ be integers and b be prime. Let $\boldsymbol{\beta} = (\beta_j)_{j\geq 1}$ be a sequence of positive numbers, and denote by $\boldsymbol{\beta}_s = (\beta_j)_{1\leq j \leq s}$ its truncation after s terms. Assume that $\boldsymbol{\beta} \in \ell^p(\mathbb{N})$ for some $0 < p < 1$, i.e. that*

$$\exists\, 0 < p < 1 : \quad \sum_{j=1}^\infty \beta_j^p < \infty \,. \tag{3.27}$$

Define, for $0 < p \leq 1$ as in (3.27), the digit interlacing parameter

$$\alpha := \lfloor 1/p \rfloor + 1 \,. \tag{3.28}$$

Consider parametric integrand functions $g : U \to \mathcal{Z}$ taking values in a separable Hilbert space $\mathcal{Z}$ which are $(\boldsymbol{\beta}, p, \varepsilon)$-holomorphic (cp. Definition 3.1).

Then, for every $N \in \mathbb{N}$, one can construct an interlaced polynomial lattice rule of order α with N points using a fast component-by-component algorithm, using $\mathcal{O}(\alpha^2 sN \log N)$ operations, plus $\mathcal{O}(\alpha^2 s^2 N)$ update cost, plus $\mathcal{O}(\alpha s N)$ memory cost, such that there holds the error bound

$$\forall s, N \in \mathbb{N} : \quad \|I_s(g) - Q_{N,s}(g)\|_{\mathcal{Z}} \leq C_{\alpha,\boldsymbol{\beta},b,p} \|g\|_{\boldsymbol{\rho},\mathcal{Z}} N^{-1/p} \,, \tag{3.29}$$

where $C_{\alpha,\boldsymbol{\beta},b,p} < \infty$ is a constant independent of s and N, and where $\|g\|_{\boldsymbol{\rho},\mathcal{Z}}$ denotes the (likewise independent of s and of N) maximum of the modulus for the integrand function over $\mathcal{T}_{\boldsymbol{\rho}}$ for every $(\boldsymbol{\beta}, \varepsilon)$-admissible poly-radius (i.e., satisfying (3.24)) $\boldsymbol{\rho} = (\rho_j)_{j\geq 1}$:

$$\|g\|_{\boldsymbol{\rho},\mathcal{Z}} := \sup_{\boldsymbol{z}\in\mathcal{T}_{\boldsymbol{\rho}}} \|g(\boldsymbol{z})\|_{\mathcal{Z}} \,. \tag{3.30}$$

Proof. The Hilbert space $\mathcal{Z}$ being isomorphic to $\mathcal{Z}'$, for every $g \in \mathcal{Z}$ the norm $\|g\|_{\mathcal{Z}}$ can be written as

$$\|g\|_{\mathcal{Z}} = \sup_{\chi \in \mathcal{Z}: \|\chi\|_{\mathcal{Z}} \leq 1} |(\chi, g)_{\mathcal{Z}}| \;,$$

where $(\cdot, \cdot)_{\mathcal{Z}}$ denotes the $\mathcal{Z}$ inner product. By separability of $\mathcal{Z}$, the Bochner-integral $I_s[g]$ and the QMC rules $Q_{N,s}[g]$ are well-defined also for measurable integrand functions taking values in $\mathcal{Z}$. For integrand functions g being a $(\boldsymbol{b}, p, \varepsilon)$-holomorphic map taking values in $\mathcal{Z}$, for every $\chi \in \mathcal{Z}$ the mapping $\boldsymbol{z} \mapsto (\chi, g(\boldsymbol{z}))_{\mathcal{Z}}$ is a $(\boldsymbol{b}, p, \varepsilon)$-holomorphic map with values in $\mathbb{C}$. Therefore, the complex variable HoQMC error bounds obtained in [8, Prop. 4.1] apply for this integrand: there exists a constant $C_{\alpha, \boldsymbol{\beta}, \boldsymbol{b}, p} > 0$ which is independent of the parametric integrand $\boldsymbol{y} \mapsto (\chi, g(\boldsymbol{y}))_{\mathcal{Z}}$ such that for every N holds the HoQMC error bound

$$|(I_s - Q_{N,s})[(\chi, g(\cdot))_{\mathcal{Z}}]| \leq C_{\alpha, \boldsymbol{\beta}, b, p} \|(\chi, g(\cdot))_{\mathcal{Z}}\|_{\boldsymbol{\rho}, \mathbb{C}} N^{-1/p} \;. \tag{3.31}$$

Here, for every fixed χ and any $(\boldsymbol{\beta}, \varepsilon)$-admissible poly-radius $\boldsymbol{\rho}$,

$$\|(\chi, g(\cdot))_{\mathcal{Z}}\|_{\boldsymbol{\rho}, \mathbb{C}} = \sup_{\boldsymbol{z} \in \mathcal{T}_{\boldsymbol{\rho}}} |(\chi, g(\boldsymbol{z}))_{\mathcal{Z}}| \;.$$

The linearity of $I_s[\cdot]$ and of $Q_{N,s}[\cdot]$ imply that for every $\chi \in \mathcal{Z}$ holds

$$(I_s - Q_{N,s})[(\chi, g(\cdot))_{\mathcal{Z}}] = (\chi, (I_s - Q_{N,s})[g(\cdot)])_{\mathcal{Z}} \;.$$

We may therefore estimate with (3.31)

$$
\begin{aligned}
\|(I_s - Q_{N,s})[g(\cdot)]\|_{\mathcal{Z}} &= \sup_{\chi \in \mathcal{Z}: \|\chi\|_{\mathcal{Z}} \leq 1} \left| (\chi, (I_s - Q_{N,s})[g(\cdot)])_{\mathcal{Z}} \right| \\
&= \sup_{\chi \in \mathcal{Z}: \|\chi\|_{\mathcal{Z}} \leq 1} |(I_s - Q_{N,s})[(g(\cdot), \chi)_{\mathcal{Z}}]| \\
&\leq C_{\alpha, \boldsymbol{\beta}, b, p} N^{-1/p} \sup_{\chi \in \mathcal{Z}: \|\chi\|_{\mathcal{Z}} \leq 1} \|(g(\cdot), \chi)_{\mathcal{Z}}\|_{\boldsymbol{\rho}, \mathbb{C}} \\
&= C_{\alpha, \boldsymbol{\beta}, b, p} N^{-1/p} \sup_{\chi \in \mathcal{Z}: \|\chi\|_{\mathcal{Z}} \leq 1} \sup_{\boldsymbol{z} \in \mathcal{T}_{\boldsymbol{\rho}}} |(g(\boldsymbol{z}), \chi)_{\mathcal{Z}}| \\
&= C_{\alpha, \boldsymbol{\beta}, b, p} N^{-1/p} \sup_{\boldsymbol{z} \in \mathcal{T}_{\boldsymbol{\rho}}} \sup_{\chi \in \mathcal{Z}: \|\chi\|_{\mathcal{Z}} \leq 1} |(g(\boldsymbol{z}), \chi)_{\mathcal{Z}}| \\
&= C_{\alpha, \boldsymbol{\beta}, b, p} N^{-1/p} \sup_{\boldsymbol{z} \in \mathcal{T}_{\boldsymbol{\rho}}} \|g(\boldsymbol{z})\|_{\mathcal{Z}} \;.
\end{aligned}
$$

$\square$

4 Multilevel HoQMC Bayesian estimators

4.1 Multilevel discretization of the forward problem

Multilevel Bayesian estimators are based on a *hierarchy of discretizations* of the forward problem (2.5). Below we design *increasing sequences* $\{s_\ell\}_{\ell \geq 0}$ and $\{N_\ell\}_{\ell \geq 0}$ of truncation dimensions and QMC sample numbers at discretization level ℓ, and a *decreasing sequence* $\{h_\ell\}_{\ell \geq 0}$ of discretization parameters. With this latter sequence, we associate *nested, dense sequences* $\{\mathcal{X}_\ell\}_{\ell \geq 0} = \{\mathcal{X}^{h_\ell}\}_{\ell \geq 0}$ and $\{\mathcal{Y}_\ell\}_{\ell \geq 0} = \{\mathcal{X}^{h_\ell}\}_{\ell \geq 0}$ of discretization spaces of increasing, finite dimension $M_\ell = \dim(\mathcal{X}_\ell) = \dim(\mathcal{Y}_\ell)$. We refer to the pairs $(s_\ell, h_\ell)_{\ell \geq 0}$ as *PG discretization parameters*, and also set $s_{-1} := 1$ and $h_{-1} := 1$. We denote by $q^{(\ell)} = q^{h_\ell}(\boldsymbol{y}_{\{1:s_\ell\}}) \in \mathcal{X}_\ell$ the dimensionally truncated, PG approximated parametric forward solution in (2.25). The consistency error bounds for approximations of the Bayesian posterior, (3.10), imply that the Bayesian estimates incur an error which is bounded by the error of the PG discretization in the forward problem, *provided posterior expectations are evaluated exactly.*

The potential Φ_Γ based on the solution of (2.5) on discretization level ℓ is denoted as Φ_Γ^ℓ, and is given in (3.12). The corresponding approximations to the posterior density $\Theta(\boldsymbol{y})$ from (3.16) are then

$$\Theta_\ell(\boldsymbol{y}) := \exp(-\Phi_\Gamma^\ell(u(\boldsymbol{y}_{\{1:s_\ell\}}); \delta)) \;. \tag{4.1}$$

The *exact normalization constant* for the forward model at discretization level ℓ is then

$$Z_\ell := \int_U \Theta_\ell(\boldsymbol{y}) \, \boldsymbol{\pi}(\mathrm{d}\boldsymbol{y}) \;. \tag{4.2}$$

Note that $\Theta_\ell(\boldsymbol{y})$, as defined in (4.1), includes also a dimension truncation to dimension s_ℓ.

In our ensuing error analysis of the Bayesian estimators, we shall use the bound (3.29) for various choices of $g(\boldsymbol{z})$; in particular, with $g(\boldsymbol{z}) = (\phi(q)\Theta - \phi(q^{(\ell)})\Theta_\ell)(\boldsymbol{z})$ both with $\phi \equiv 1$, $\mathcal{Z} = \mathbb{R}$ and with $\phi(\cdot) : \mathcal{X} \to \mathcal{Z}$ denoting the QoI.

Given a QoI $\phi : \mathcal{X} \to \mathcal{Z}$ and noisy observation data $\delta \in Y$ as in (3.2), we approximate the Bayesian estimate $\mathbb{E}^{\boldsymbol{\pi}^\delta}[\phi]$ defined in (3.19) by the expectation

$$\mathbb{E}^{\boldsymbol{\pi}^\delta_\ell}[\phi] = \frac{1}{Z_\ell} \int_U \phi(q^{(\ell)}(\boldsymbol{y}))\Theta_\ell(\boldsymbol{y})\,\boldsymbol{\pi}(\mathrm{d}\boldsymbol{y}) = \int_U \phi(q^{(\ell)}(\boldsymbol{y}))\pi^\delta_\ell(\mathrm{d}\boldsymbol{y}) = \frac{Z'_\ell}{Z_\ell} \,, \tag{4.3}$$

where π^δ_ℓ is given by the Radon-Nikodym derivative $\mathrm{d}\pi^\delta_\ell/\mathrm{d}\pi = \Theta_\ell/Z_\ell$.

4.2 Multilevel Bayesian estimators

We now describe two multilevel estimators, the ratio estimator proposed in [20, 21] and the splitting estimator, inspired by [1]. The ratio estimator is based on expanding the two integrals Z'_L and Z_L separately, and the splitting estimator is based on a joint telescopic expansion of Z'_L/Z_L. In conjunction with deterministic higher order QMC integration (3.20), either estimator will result in deterministic algorithms for Bayesian estimation of PDEs with distributed, uncertain input with high convergence rates which are independent of the dimension of the parameter space.

4.2.1 Multilevel ratio estimator

It was proposed in [20, 21] to numerically evaluate the Bayesian estimates (3.19) by "direct integration", i.e. by applying a deterministic quadrature rule to the (formally) infinite-dimensional, iterated integrals in the normalization constant Z and the QoI Z'. In [20, 21], dimension adaptive Smolyak quadrature was used to compute Z'/Z. Here, as in [9], we evaluate Z'_L/Z_L by approximating the integrals Z and Z' by HoQMC integration. In [9] we used single-level estimators for numerator and denominator; here, we develop a MLHoQMC estimator. To derive it, we write, using the telescopic sum identity and denoting integration with respect to the prior $\boldsymbol{\pi}$ by $I[\cdot]$,

$$Z'_L = Z'_0 + \sum_{\ell=1}^L (Z'_\ell - Z'_{\ell-1}) = I[\phi(q^{(0)})\Theta_0] + \sum_{\ell=1}^L I\Big[\phi(q^{(\ell)})\Theta_\ell - \phi(q^{(\ell-1)})\Theta_{\ell-1}\Big] \tag{4.4}$$

$$= \int_U \phi(q^{(0)}(\boldsymbol{y}))\Theta_0(\boldsymbol{y})\,\boldsymbol{\pi}(\mathrm{d}\boldsymbol{y}) + \sum_{\ell=1}^L \int_U \Big(\phi(q^{(\ell)}(\boldsymbol{y}))\Theta_\ell(\boldsymbol{y}) - \phi(q^{(\ell-1)}(\boldsymbol{y}))\Theta_{\ell-1}(\boldsymbol{y})\Big)\,\boldsymbol{\pi}(\mathrm{d}\boldsymbol{y}) \,,$$

and the corresponding ML approximation of the normalization constant Z,

$$Z_L = Z_0 + \sum_{\ell=1}^L (Z_\ell - Z_{\ell-1}) = I[\Theta_0] + \sum_{\ell=1}^L I\Big[\Theta_\ell - \Theta_{\ell-1}\Big] \tag{4.5}$$

$$= \int_U \Theta_0(\boldsymbol{y})\,\boldsymbol{\pi}(\mathrm{d}\boldsymbol{y}) + \sum_{\ell=1}^L \int_U \Big(\Theta_\ell(\boldsymbol{y}) - \Theta_{\ell-1}(\boldsymbol{y})\Big)\,\boldsymbol{\pi}(\mathrm{d}\boldsymbol{y}) \,.$$

We approximate the integrals in (4.4) and (4.5) on discretization level $\ell = 0,\dots,L$ with a higher order QMC rule based on N_ℓ points, resulting in the Multilevel Higher Order QMC estimators $Q^\star_L[Z']$ and $Q^\star_L[Z]$, respectively. Note that both estimators (for Z and Z') can be computed simultaneously, without having to reevaluate the forward model. The *multilevel ratio estimator* is thus given by

$$Q^\star_{L,ratio} := Q^\star_L[Z']/Q^\star_L[Z] \,, \tag{4.6}$$

with the individual multilevel approximations (omitting the variables $\boldsymbol{y}$) given by the *multilevel algorithm* $Q^\star_L$ of [18, 7], i.e.

$$Z'_L \approx Q^\star_L[Z'] := Q_{N_0,s_0}[\phi(q^{(0)})\Theta_0] + \sum_{\ell=1}^L Q_{N_\ell,s_\ell}\Big[\phi(q^{(\ell)})\Theta_\ell - \phi(q^{(\ell-1)})\Theta_{\ell-1}\Big] \,, \tag{4.7}$$

$$Z_L \approx Q^\star_L[Z] := Q_{N_0,s_0}[\Theta_0] + \sum_{\ell=1}^L Q_{N_\ell,s_\ell}\Big[\Theta_\ell - \Theta_{\ell-1}\Big] \,. \tag{4.8}$$

4.2.2 Multilevel splitting estimator

We consider approximations of the form (4.3) on each discretization level $\ell = 0, \ldots, L$ and write

$$\frac{Z'_L}{Z_L} = \frac{Z'_0}{Z_0} + \sum_{\ell=1}^{L} \left(\frac{Z'_\ell}{Z_\ell} - \frac{Z'_{\ell-1}}{Z_{\ell-1}} \right) . \tag{4.9}$$

We now apply QMC quadrature to the integrals in Z_ℓ, Z'_ℓ for each $\ell = 0, \ldots, L$, where we approximate all terms on the same discretization level ℓ (i.e. all terms inside the parenthesis) by a QMC rule with N_ℓ points, resulting in the Multilevel High Order QMC estimator Q^*_L. Denoting by $Z_{\ell,k} = Q_{N_k}[Z_\ell]$ and $Z'_{\ell,k} = Q_{N_k}[Z'_\ell]$ the HoQMC approximations of Z_ℓ and Z'_ℓ on quadrature level k, the *multilevel QMC splitting estimator* reads

$$Q^\star_{L,split} := \frac{Z'_{0,0}}{Z_{0,0}} + \sum_{\ell=1}^{L} \left(\frac{Z'_{\ell,\ell}}{Z_{\ell,\ell}} - \frac{Z'_{\ell-1,\ell}}{Z_{\ell-1,\ell}} \right) . \tag{4.10}$$

Thus, on discretization level $\ell = 0$ we must compute the two approximations $Z_{0,0} = Q_{N_0}[\Theta_0]$ and $Z'_{0,0} = Q_{N_0}[\phi(q^0)\Theta_0]$, and for each discretization level $\ell = 1, \ldots, L$ we evaluate four approximations

$$Z'_{\ell,\ell} = Q_{N_\ell,s_\ell}\big[\phi(q^{(\ell)}(\boldsymbol{y}))\Theta_\ell(\boldsymbol{y})\big] \qquad\qquad Z'_{\ell-1,\ell} = Q_{N_\ell,s_\ell}\big[\phi(q^{(\ell-1)}(\boldsymbol{y}))\Theta_{\ell-1}(\boldsymbol{y})\big]$$
$$Z_{\ell,\ell} = Q_{N_\ell,s_\ell}\big[\Theta_\ell(\boldsymbol{y})\big] \qquad\qquad Z_{\ell-1,\ell} = Q_{N_\ell,s_\ell}\big[\Theta_{\ell-1}(\boldsymbol{y})\big] . \tag{4.11}$$

Since all four approximations involve the same QMC quadrature points, the additional quadratures do not require extra solutions of the forward model on level ℓ or $\ell - 1$. Therefore, they do not increase the work significantly, assuming the cost of representing elements of $\mathcal{Z}$ in an implementation are negligible. In the case where different truncation dimensions $s_\ell > s_{\ell-1}$ are used, the "quadrature point" $\boldsymbol{y}_{s_\ell}$ on level ℓ can be truncated at dimension $s_{\ell-1}$: $\boldsymbol{y}_{s_{\ell-1}} := (\boldsymbol{y}_{s_\ell})_{\{1:s_{\ell-1}\}} \in [-\frac{1}{2}, \frac{1}{2}]^{s_{\ell-1}}$ on level $\ell - 1$.

Remark 4.1. The differences in the formula (4.9) can be written as the following expression (bearing in mind that index ℓ signifies implied dimension truncation to dimension s_ℓ)

$$\frac{Z'_\ell}{Z_\ell} - \frac{Z'_{\ell-1}}{Z_{\ell-1}} = \int_{\boldsymbol{y} \in [-1/2, 1/2]^{s_\ell}} \left(\phi(q^{(\ell)}(\boldsymbol{y})) - \frac{\Theta_{\ell-1}(\boldsymbol{y})Z_\ell}{\Theta_\ell(\boldsymbol{y})Z_{\ell-1}} \phi(q^{(\ell-1)}(\boldsymbol{y})) \right) \boldsymbol{\pi}^\delta_\ell(\mathrm{d}\boldsymbol{y}) , \tag{4.12}$$

see (4.3). This form corresponds to the splitting which is customary in MCMC and SMC methods, see, e.g. [1] and the references there. We expect the application of QMC quadratures to the integrand function in (4.12) to be advantageous for small observation covariance Γ. The integrand function of (4.12) is, *upon suitable (Γ-dependent) rescaling of coordinates as in [22, Theorem 4.1], $(\boldsymbol{b}, p, \varepsilon)$-holomorphic uniformly with respect to $\Gamma > 0$.*

4.3   Error analysis

Each of the proposed computable estimators (4.6), (4.10) is based on approximating posterior expectations of PG discretizations and dimensionally truncated forward models by HoQMC integration. The ensuing error analysis is based on the dimension independent error bounds for these quadratures which we developed in [6, 7]. These bounds are based on estimates of higher order derivatives of the integrand functions with respect to the integration variables. In [8, Section 3], we developed estimates for these derivatives based on analytic continuation of integrand functions into the complex domain and Cauchy's integral formula; this is feasible for holomorphic forward problems considered in [3], and for general, holomorphic potentials Φ_Γ. As is by now well-known, and in contrast to our single-level QMC error analysis for the quotient estimator in [9], multilevel QMC error estimates involve bounding *quadrature errors of PG forward problem discretization errors*. To use the higher order QMC error bounds of [8] in the present multilevel context therefore requires *bounds on the dimension truncation and PG discretization errors of analytic continuations of the integrand functions.* These bounds are to hold uniformly with respect to the truncation dimension s_ℓ. The present, analytic continuation approach should be contrasted with the "real-variable" approach used in the error analysis of QMC integration in [7], which is based on bootstrapping arguments and induction.

13

4.3.1 Holomorphy assumptions on the PG approximation

To ensure holomorphy of countably parametric families of integrand functions in the Bayesian estimators (4.6), (4.10), as required by the QMC error bounds stated in Proposition 3.4, we impose corresponding holomorphy assumptions on the parametric forward problem (2.11) as well as on its PG discretization (2.25). We formalize these conditions (which are versions of the uniform parametric regularity **UPR** and the parametric stability) in the following:

(i) **HCP**$_t$ Holomorphic continuation of the parametric forward problem in $\mathcal{X}_t$: there is a $(\boldsymbol{b}, p, \varepsilon)$-holomorphic extension of the parametric forward problem (2.14) such that, for any truncation dimension $s \in \mathbb{N}$, and for every $\boldsymbol{z} \in O_{\boldsymbol{\rho}}$, the extended problem

$$\text{find } q^s(\boldsymbol{z}_{\{1:s\}}) \in \mathcal{X}: \quad _{\mathcal{Y}'}\langle \mathcal{R}(u^s(\boldsymbol{z}); q^s(\boldsymbol{z})), w \rangle_{\mathcal{Y}} = 0 \ \forall w \in \mathcal{Y}, \tag{4.13}$$

admits a unique, parametric solution $q^s(\boldsymbol{z}_{\{1:s\}}) \in \mathcal{X}_t$ which is $(\boldsymbol{b}, p, \varepsilon)$-holomorphic.

(ii) **PGStabC** Stability and Quasioptimality of the PG discretization for the complex parametric extension: for every $\boldsymbol{z} \in O_{\boldsymbol{\rho}}$, and for every $0 < h \leq h_0$ the Galerkin approximations:

$$\text{find } q_h(\boldsymbol{z}) \in \mathcal{X}^h: \quad _{\mathcal{Y}'}\langle \mathcal{R}(\boldsymbol{z}; q_h(\boldsymbol{z})), w^h \rangle_{\mathcal{Y}} = 0 \quad \forall w^h \in \mathcal{Y}^h, \tag{4.14}$$

are uniquely defined and converge quasioptimally: there exists a constant $C > 0$ such that for all $\boldsymbol{z} \in O_{\boldsymbol{\rho}}$, with a $(\boldsymbol{b}_t, \varepsilon)$-admissible poly-radius $\boldsymbol{\rho}$,

$$\|q(\boldsymbol{z}) - q_h(\boldsymbol{z})\|_{\mathcal{X}} \leq \frac{C}{\bar{\mu}} \inf_{0 \neq v^h \in \mathcal{X}^h} \|q(\boldsymbol{z}) - v^h\|_{\mathcal{X}}. \tag{4.15}$$

We remark that **HCP**$_t$ and **PGStabC** imply, via the approximation property (2.19), the convergence rate $O(h^t)$ in $\mathcal{X}$ for the Galerkin solution $q_h(\boldsymbol{z}) \in \mathcal{X}^h$, with implied constant independent of the truncation dimension s.

(iii) **ANPGC**$_{t'}$ Aubin-Nitsche argument for the PG discretization of the complex parametric extension: There exists a constant $C > 0$ such that for every $G \in \mathcal{X}'_{t'}$ there holds a superconvergence estimate: for every $\boldsymbol{\rho}$ which is $(\boldsymbol{b}_t, \varepsilon)$-admissible for sufficiently small $\varepsilon > 0$,

$$\forall \boldsymbol{z} \in \mathcal{O}_{\boldsymbol{\rho}}: \quad |G(q(\boldsymbol{z})) - G(q_h(\boldsymbol{z}))| \leq Ch^{\tau}, \quad \tau = t + t'. \tag{4.16}$$

Examples with valid conditions (i) - (iii) will be presented in the numerical experiments Section 5 ahead; the hypotheses will be verified in particular for affine-parametric operator equations in Section 5.1.

4.3.2 Error bounds for the multilevel ratio estimator

Numerator and denominator of the estimator (4.6) are approximated separately by MLHoQMC estimators, (4.7), (4.8). We first analyze the error of these approximations, thereby also generalizing the single-level results in [8]. *Throughout, we assume that in all densities which occur in the integrals (3.17) - (3.19) the parametric forward problems are dimension-truncated to finite parameter dimensions $\{s_\ell\}_{\ell=0}^L$ for discretization levels $\ell = 0, 1, ..., L$, which we assume to be strictly increasing* (the ensuing error analysis remains valid with obvious modifications, if several or all truncation dimensions are equal) i.e.

$$0 < s_0 < s_1 < ... < s_L < \infty.$$

Throughout the error analysis, we assume $L \geq 2$ and $\theta \in \{1, 2\}$ is as in (2.16).

Theorem 4.1. *Assume that $0 \leq t \leq \bar{t}$, for some $\bar{t} > 0$ and that, in addition, there exists $C_0 > 0$ such that $|Z_\ell| > C_0$, $|Z_{\ell,\ell}| > C_0$ uniformly with respect to $\ell = 0, 1, 2, \ldots$. Assume also* **HCP**$_t$, **PGStabC** *hold for some $t > 0$.*

Then, for every $p_t < \lambda, q < 1$, with $\theta \in \{1, 2\}$ as in (2.16),

$$\left\| \frac{Q_L^*[Z_L']}{Q_L^*[Z_L]} - \mathbb{E}^{\boldsymbol{\pi}^\delta}[\phi] \right\|_{\mathcal{Z}} \leq C \left[h_L^t + s_L^{-\theta(1/p_0 - 1)} + N_0^{-1/\lambda} + \sum_{\ell=1}^L N_\ell^{-1/\lambda}(h_{\ell-1}^t + s_{\ell-1}^{-\theta(1/p_0 - 1/q)}) \right]. \tag{4.17}$$

Proof. Since $\mathbb{E}^{\boldsymbol{\pi}^{\delta}}[\phi] = Z'/Z$, we estimate

$$Z' - Q_L^*[Z_L'] = Z' - Z_L' + Z_L' - Q_L^*[Z_L'] =: I_L + II_L . \tag{4.18}$$

The first term I_L is the exact posterior expectation on the approximate forward model, i.e. a pure discretization error. It can be estimated using Proposition 3.2 and the approximation property (2.19) with property $\mathbf{HCP_t}$. There exists a constant $C > 0$ which is independent of h_L and of s_L such that

$$\|Z' - Z_L'\|_{\mathcal{Z}} \leq C(h_L^t + s_L^{-\theta(1/p_0 - 1)}) . \tag{4.19}$$

To bound term II_L in (4.18), we write it as telescoping sum

$$Z_L' - Q_L^*[Z_L'] = (I - Q_{N_0, s_0})[\phi(q^0)\Theta_0] + \sum_{\ell=1}^{L} \left[(I - Q_{N_\ell, s_\ell})[\phi(q^{(\ell)})\Theta_\ell - \phi(q^{(\ell-1)})\Theta_{\ell-1}] \right] . \tag{4.20}$$

The first term in (4.20) is a pure QMC integration error, on the coarsest level PG discretization; it can be estimated with (3.29). The remaining terms are QMC quadrature errors of dimension truncation and Petrov-Galerkin errors. A typical term reads

$$(I - Q_{N_\ell, s_\ell})[\phi(q^{(\ell)})\Theta_\ell - \phi(q^{(\ell-1)})\Theta_{\ell-1}] , \quad \ell = 1, ..., L . \tag{4.21}$$

We rewrite a generic term in the telescoping sum (4.20) as (parameter $\boldsymbol{z}$ not indicated)

$$\phi(q^{(\ell)})\Theta_\ell - \phi(q^{\ell-1})\Theta_{-1} = \phi(q^{(\ell)-q^{(\ell-1)}})\Theta_\ell + \phi(q^{(\ell-1)})(\Theta_\ell - \Theta_{\ell-1}) =: D_\phi^\ell \Theta_\ell + \phi(q^{(\ell-1)})D_\Theta^\ell .$$

Here, we used the linearity of the QoI $\phi()$.

To apply the QMC quadrature error bound (3.29) with $g(\boldsymbol{z}) = (\phi(q^{(\ell)})\Theta_\ell - \phi(q^{(\ell-1)})\Theta_{\ell-1})(\boldsymbol{z})$, we need to estimate (cp. (3.30)) for $\boldsymbol{z} \in \mathcal{T}_{\boldsymbol{\rho}}$ with $\boldsymbol{\rho}$ being $(\boldsymbol{\beta}_t, \varepsilon)$-admissible

$$
\begin{aligned}
\|g\|_{\boldsymbol{\beta}, \mathcal{Z}} &:= \sup_{\boldsymbol{z} \in \mathcal{T}_{\boldsymbol{\rho}}} \|(\phi(q^{(\ell)})\Theta_\ell - \phi(q^{(\ell-1)})\Theta_{-1})(\boldsymbol{z})\|_{\mathcal{Z}} \\
&\leq \sup_{\boldsymbol{z} \in \mathcal{T}_{\boldsymbol{\rho}}} \|D_\phi^\ell(\boldsymbol{z})\Theta_\ell(\boldsymbol{z})\|_{\mathcal{Z}} + \sup_{\boldsymbol{z} \in \mathcal{T}_{\boldsymbol{\rho}}} \|\phi(q^{(\ell-1)})(\boldsymbol{z})D_\Theta^\ell(\boldsymbol{z})\|_{\mathcal{Z}} .
\end{aligned}
\tag{4.22}
$$

We estimate each of the two differences in essentially the same way. Consider D_ϕ^ℓ. Then

$$
\begin{aligned}
\|D_\phi^\ell(\boldsymbol{z})\|_{\mathcal{Z}} &\leq \|\phi\|_{\mathcal{L}(\mathcal{X}, \mathcal{Z})} \|q^{(\ell)}(\boldsymbol{z}) - q^{(\ell-1)}(\boldsymbol{z})\|_{\mathcal{X}} \\
&= \|\phi\|_{\mathcal{L}(\mathcal{X}, \mathcal{Z})} \|q_{h_\ell}^{s_\ell}(\boldsymbol{z}) - q_{h_{\ell-1}}^{s_\ell}(\boldsymbol{z}) + q_{h_{\ell-1}}^{s_\ell}(\boldsymbol{z}) - q_{h_{\ell-1}}^{s_{\ell-1}}(\boldsymbol{z})\|_{\mathcal{X}} \\
&\leq \|\phi\|_{\mathcal{L}(\mathcal{X}, \mathcal{Z})} \left[\|q^{s_\ell}(\boldsymbol{z}) - q_{h_\ell}^{s_\ell}(\boldsymbol{z})\|_{\mathcal{X}} + \|q^{s_\ell}(\boldsymbol{z}) - q_{h_{\ell-1}}^{s_\ell}(\boldsymbol{z})\|_{\mathcal{X}} + \|q_{h_{\ell-1}}^{s_\ell}(\boldsymbol{z}) - q_{h_{\ell-1}}^{s_{\ell-1}}(\boldsymbol{z})\|_{\mathcal{X}} \right] .
\end{aligned}
$$

The first two terms are PG discretization errors for the complex-parametric extension of the operator equation. Property $\mathbf{HCP_t}$ (holomorphic continuability of the parametric forward problem) implies the well-posedness of the complex-parametric problem, and assumption $\mathbf{PGStabC}$ (stability of the PG discretization in the complex-parametric problem) implies that the PG discretization errors can be bounded as

$$\sup_{\boldsymbol{z} \in \mathcal{T}_{\boldsymbol{\rho}}} \left[\|q^{s_\ell}(\boldsymbol{z}) - q_{h_\ell}^{s_\ell}(\boldsymbol{z})\|_{\mathcal{X}} + \|q^{s_\ell}(\boldsymbol{z}) - q_{h_{\ell-1}}^{s_\ell}(\boldsymbol{z})\|_{\mathcal{X}} \right] \leq C h_{\ell-1}^t \sup_{\boldsymbol{z} \in \mathcal{T}_{\boldsymbol{\rho}}} \|q(\boldsymbol{z})\|_{\mathcal{X}_t} , \tag{4.23}$$

provided that $\boldsymbol{\rho}$ is $(\boldsymbol{\beta}_t, \varepsilon)$-admissible. The third term is a dimension truncation error. It vanishes if $s_{\ell-1} = s_\ell$, and by the stability of the PG discretization is bounded for any $p_t \leq q \leq 1$ by

$$\sup_{\boldsymbol{z} \in \mathcal{T}_{\boldsymbol{\rho}}} \|q_{h_{\ell-1}}^{s_\ell}(\boldsymbol{z}) - q_{h_{\ell-1}}^{s_{\ell-1}}(\boldsymbol{z})\|_{\mathcal{X}} \leq C s_{\ell-1}^{-\theta(1/p_0 - 1/q)} , \quad \ell = 1, 2, ..., L - 1 . \tag{4.24}$$

As the modulus of the posterior density, $|\Theta_\ell(\boldsymbol{z})|$, is uniformly bounded w.r.t. ℓ and w.r.t. $\boldsymbol{z} \in \mathcal{T}_{\boldsymbol{\rho}}$, for any $(\boldsymbol{\beta}_0, \varepsilon)$-admissible $\boldsymbol{\rho}$, the first term in the bound (4.22) can be estimated for $L \geq 2$ by an absolute multiple of

$$h_{\ell-1}^t + s_{\ell-1}^{-\theta(1/p_0 - 1/q)} , \quad \ell = 1, 2, ..., L - 1 . \tag{4.25}$$

For a fixed $(\boldsymbol{\beta}, \varepsilon)$-admissible poly-radius $\boldsymbol{\rho}$, and for any $\boldsymbol{z} \in \mathcal{T}_{\boldsymbol{\rho}}$ we estimate the second term in (4.22) by

$$\sup_{\boldsymbol{z} \in \mathcal{T}_{\boldsymbol{\rho}}} \|\phi(q^{(\ell-1)})(\boldsymbol{z})D_\Theta^\ell(\boldsymbol{z})\|_{\mathcal{Z}} = \sup_{\boldsymbol{z} \in \mathcal{T}_{\boldsymbol{\rho}}} |D_\Theta^\ell(\boldsymbol{z})| \sup_{\boldsymbol{z} \in \mathcal{T}_{\boldsymbol{\rho}}} \|\phi(q^{(\ell-1)})(\boldsymbol{z})\|_{\mathcal{Z}} . \tag{4.26}$$

15

Here, the latter supremum is bounded uniformly w.r.t. ℓ, due to condition **PGStabC**: the uniform stability of the PG discretization for the complex-parametric problem implies also the existence of a bound $B(\boldsymbol{\beta}, \varepsilon)$, independent of the discretization level ℓ such that

$$\forall \boldsymbol{\rho} \text{ which are } (\boldsymbol{\beta}, \varepsilon)\text{-admissible}: \quad \sup_{\boldsymbol{z} \in \mathcal{T}_{\boldsymbol{\rho}}} \|q^{(\ell)}(\boldsymbol{z})\|_{\mathcal{X}} \leq B(\boldsymbol{\beta}, \varepsilon) < \infty \, .$$

We focus on the first supremum in (4.26).

$$
\begin{aligned}
\sup_{\boldsymbol{z} \in \mathcal{T}_{\boldsymbol{\rho}}} |D_{\Theta}^{\ell}(\boldsymbol{z})| &= \sup_{\boldsymbol{z} \in \mathcal{T}_{\boldsymbol{\rho}}} |\Theta_{h_{\ell}}^{s_{\ell}}(\boldsymbol{z}) - \Theta_{h_{\ell-1}}^{s_{\ell}}(\boldsymbol{z}) + \Theta_{h_{\ell-1}}^{s_{\ell}}(\boldsymbol{z}) - \Theta_{h_{\ell-1}}^{s_{\ell-1}}(\boldsymbol{z})| \\
&\leq \sup_{\boldsymbol{z} \in \mathcal{T}_{\boldsymbol{\rho}}} \left[|\Theta^{s_{\ell}}(\boldsymbol{z}) - \Theta_{h_{\ell}}^{s_{\ell}}(\boldsymbol{z})| + |\Theta^{s_{\ell}}(\boldsymbol{z}) - \Theta_{h_{\ell-1}}^{s_{\ell}}(\boldsymbol{z})| + |\Theta_{h_{\ell-1}}^{s_{\ell}}(\boldsymbol{z}) - \Theta_{h_{\ell-1}}^{s_{\ell-1}}(\boldsymbol{z})| \right] \, .
\end{aligned}
$$
(4.27)

We deal with three errors in the posterior density, which arise due to different approximations in the forward maps.

To reduce the error in the posterior density to an error bound in the corresponding potentials Φ_{Γ}, we recall definition (3.3) (applied with transposition to the complex-parametric forward map) and write

$$\left| \int_{-\Phi_{\Gamma}}^{-\Phi_{\Gamma}^M} e^{\zeta} d\zeta \right| = \left| \exp(-\Phi_{\Gamma}^M) - \exp(-\Phi_{\Gamma}) \right| \leq |\Phi_{\Gamma}^M - \Phi_{\Gamma}| \max_{\zeta \in \text{conv}(-\Phi_{\Gamma}^M, -\Phi_{\Gamma})} |e^{\zeta}|$$

, where the complex integral is along a straight line in $\mathbb{C}$ connecting $-\Phi_{\Gamma}^M$ and $-\Phi_{\Gamma}$.

The difference $\Phi_{\Gamma} - \Phi_{\Gamma}^{h_{\ell}}$ in the Bayesian potentials due to the PG discretization (at equal truncation dimension s_{ℓ}) of the forward problem is estimated using (a complex-parametric extension of) (3.11) again in terms of the PG forward discretization error.

We obtain for every $\boldsymbol{\rho}$ which is $(\boldsymbol{\beta}_t, \varepsilon)$-admissible the bound

$$\sup_{\boldsymbol{z} \in \mathcal{T}_{\boldsymbol{\rho}}} \left[|\Theta^{s_{\ell}}(\boldsymbol{z}) - \Theta_{h_{\ell}}^{s_{\ell}}(\boldsymbol{z})| + |\Theta^{s_{\ell}}(\boldsymbol{z}) - \Theta_{h_{\ell-1}}^{s_{\ell}}(\boldsymbol{z})| \right] \leq C(t, \varepsilon) h_{\ell-1}^t \, .$$

The third term in (4.27) is bounded in the same way in terms of the error (4.24), resulting also for (4.26) in the upper bound (4.25). Collecting all bounds, we have shown that for any poly-radius $\boldsymbol{\rho}$ which is $(\boldsymbol{\beta}_t, \varepsilon)$-admissible, there exists a constant $C(\boldsymbol{\beta}_t, \varepsilon) > 0$ such that for any $p_t \leq \lambda < 1$ and for all $\boldsymbol{z} \in \mathcal{T}_{\boldsymbol{\rho}}$

$$
\begin{aligned}
\|Z'_L - Q_L^*[Z'_L]\|_{\mathcal{Z}} &\leq \|(I - Q_{N_0, s_0})[\phi(q^0)\Theta_0]\|_{\mathcal{Z}} + \sum_{\ell=1}^{L} \left\| (I - Q_{N_\ell, s_\ell})[\phi(q^{(\ell)})\Theta_\ell - \phi(q^{(\ell-1)})\Theta_{\ell-1}] \right\|_{\mathcal{Z}} \\
&\leq C \left[N_0^{-1/p_0} + \sum_{\ell=1}^{L} N_\ell^{-1/\lambda} (h_{\ell-1}^t + s_{\ell-1}^{-\theta(1/p_0 - 1/q)}) \right] \, .
\end{aligned}
$$

Combining this with (4.19), we obtain from (4.18) the error estimate

$$\|Z' - Q_L^*[Z'_L]\|_{\mathcal{Z}} \leq C \left[h_L^t + s_L^{-\theta(1/p_0 - 1)} + N_0^{-1/p_0} + \sum_{\ell=1}^{L} N_\ell^{-1/\lambda} (h_{\ell-1}^t + s_{\ell-1}^{-\theta(1/p_0 - 1/q)}) \right] \, .$$
(4.28)

By choosing $\phi \equiv 1$, and $\mathcal{Z} = \mathbb{R}$, the bound (4.28) applies also for the approximation $Q_L^*[Z_L]$ of Z.

We may now estimate the error in the quotient estimator (4.6). To this end, we write

$$
\begin{aligned}
\left\| \frac{Z'}{Z} - \frac{Q_L^*[Z']}{Q_L^*[Z]} \right\|_{\mathcal{Z}} &= \left\| \frac{Z' Q_L^*[Z] - Q_L^*[Z'] Z}{Z Q_L^*[Z]} \right\|_{\mathcal{Z}} \\
&\leq \frac{1}{Z Q_L^*[Z]} \left\{ \|Z'\|_{\mathcal{Z}} |Z - Q_L^*[Z]| + |Z| \|Z' - Q_L^*[Z']\|_{\mathcal{Z}} \right\} \, .
\end{aligned}
$$

Using (4.28) twice, we arrive at the error bound for the ratio estimator,

$$\left\| \frac{Z'}{Z} - \frac{Q_L^*[Z']}{Q_L^*[Z]} \right\|_{\mathcal{Z}} \leq C \left[h_L^t + s_L^{-\theta(1/p_0 - 1)} + N_0^{-1/p_0} + \sum_{\ell=1}^{L} N_\ell^{-1/\lambda} (h_{\ell-1}^t + s_{\ell-1}^{-\theta(1/p_0 - 1/q)}) \right] \, .$$
(4.29)

$\square$

4.3.3 Error bounds for the multilevel splitting estimator

Theorem 4.2. *Assume that $0 \leq t \leq \bar{t}$, for some $\bar{t} > 0$ and that, in addition, there exists $C_0 > 0$ such that $|Z_\ell| > C_0$, $|Z_{\ell,\ell}| > C_0$ uniformly with respect to $\ell = 0, 1, 2, \ldots$. Assume $\mathbf{HCP_t}$, $\mathbf{PGStabC}$ hold for some $t > 0$. Then, for every $p_t < \lambda, q < 1$, with $\theta \in \{1, 2\}$ as in (2.16),*

$$\left\| Q_{L,split}^*[\phi] - \mathbb{E}^{\boldsymbol{\pi}^\delta}[\phi] \right\|_{\mathcal{Z}} \leq C \left[h_L^t + s_L^{-\theta(1/p_0 - 1)} + N_0^{-1/\lambda} + \sum_{\ell=1}^L N_\ell^{-1/\lambda} (h_{\ell-1}^t + s_{\ell-1}^{-\theta(1/p_0 - 1/q)}) \right]. \quad (4.30)$$

Proof. For given sequences $\{s_\ell\}_{\ell=0}^L$, $\{N_\ell\}_{\ell=0}^L$ of truncation dimensions s_ℓ and numbers N_ℓ of QMC points, and for $\mathbb{E}^{\pi_\ell^\delta}$ given by (4.3), we estimate

$$\left\| \mathbb{E}^{\boldsymbol{\pi}^\delta}[\phi] - Q_L^*[\phi] \right\|_{\mathcal{Z}} \leq \left\| \mathbb{E}^{\boldsymbol{\pi}^\delta}[\phi] - \mathbb{E}^{\boldsymbol{\pi}_L^\delta}[\phi] \right\|_{\mathcal{Z}} + \left\| \mathbb{E}^{\boldsymbol{\pi}_L^\delta}[\phi] - Q_L^*[\phi] \right\|_{\mathcal{Z}}.$$

To bound the first term, we use Proposition 3.2 (see (3.10)) and the approximation property (2.19) and the dimension truncation estimate (2.15) with the discretization parameters $\{(s_\ell, N_\ell)\}_{\ell=0,\ldots,L}$,

$$\left\| \mathbb{E}^{\boldsymbol{\pi}^\delta}[\phi] - \mathbb{E}^{\boldsymbol{\pi}_L^\delta}[\phi] \right\|_{\mathcal{Z}} \leq C(\Gamma, R, r) \left(h_L^t + s_L^{-\theta(1/p_0 - 1)} \right).$$

To bound the second term, we write, after subtracting (4.9) and (4.10)

$$\mathbb{E}^{\pi_L^\delta}[\phi] - Q_{L,split}^* = \frac{Z_0'}{Z_0} - \frac{Z_{0,0}'}{Z_{0,0}} + \sum_{\ell=1}^L \left[\left(\frac{Z_\ell'}{Z_\ell} - \frac{Z_{\ell,\ell}'}{Z_{\ell,\ell}} \right) - \left(\frac{Z_{\ell-1}'}{Z_{\ell-1}} - \frac{Z_{\ell-1,\ell}'}{Z_{\ell-1,\ell}} \right) \right].$$

We rewrite the difference between the ℓ-th term in the sum over $\ell = 1, \ldots, L$ in (4.9) and (4.10) as

$$\left(\frac{Z_\ell'}{Z_\ell} - \frac{Z_{\ell,\ell}'}{Z_{\ell,\ell}} \right) - \left(\frac{Z_{\ell-1}'}{Z_{\ell-1}} - \frac{Z_{\ell-1,\ell}'}{Z_{\ell-1,\ell}} \right) = \underbrace{\frac{1}{Z_\ell}(Z_\ell' - Z_{\ell,\ell}') - \frac{1}{Z_{\ell-1}}(Z_{\ell-1}' - Z_{\ell-1,\ell}')}_{A_\ell}$$

$$+ \underbrace{Z_{\ell,\ell}' \left(\frac{1}{Z_\ell} - \frac{1}{Z_{\ell,\ell}} \right) - Z_{\ell-1,\ell}' \left(\frac{1}{Z_{\ell-1}} - \frac{1}{Z_{\ell-1,\ell}} \right)}_{B_\ell}.$$

We rewrite A_ℓ upon adding and subtracting the term $(Z_{\ell-1}' - Z_{\ell-1,\ell}')/Z_\ell$ as

$$A_\ell = A_\ell + (Z_{\ell-1}' - Z_{\ell-1,\ell}')/Z_\ell - (Z_{\ell-1}' - Z_{\ell-1,\ell}')/Z_\ell$$

$$= \frac{1}{Z_\ell}[(Z_\ell' - Z_{\ell,\ell}') - (Z_{\ell-1}' - Z_{\ell-1,\ell}')] + (Z_{\ell-1}' - Z_{\ell-1,\ell}') \left(\frac{1}{Z_\ell} - \frac{1}{Z_{\ell-1}} \right). \quad (4.31)$$

For the first term in the right hand side of (4.31), we observe that

$$(Z_\ell' - Z_{\ell,\ell}') - (Z_{\ell-1}' - Z_{\ell-1,\ell}') = (I[\phi(q^{(\ell)})\Theta_\ell] - Q_{N_\ell}[\phi(q^{(\ell)})\Theta_\ell]) - (I[\phi(q^{(\ell-1)})\Theta_{\ell-1}] - Q_{N_\ell}[\phi(q^{(\ell-1)})\Theta_{\ell-1}])$$

$$= (I - Q_{N_\ell})[\phi(q^{(\ell)})\Theta_\ell - \phi(q^{(\ell-1)})\Theta_{\ell-1}].$$

Using the estimates of (4.21) in the previous section we obtain

$$\left\| (Z_\ell' - Z_{\ell,\ell}') - (Z_{\ell-1}' - Z_{\ell-1,\ell}') \right\|_{\mathcal{Z}} \leq C N_\ell^{-1/\lambda} (h_{\ell-1}^t + s_{\ell-1}^{-\theta(1/p_0 - 1/q)})$$

where $C > 0$ is independent of $\{s_\ell\}_{\ell \geq 0}$ and of N_ℓ. For the second term in (4.31), we have

$$\|Z_{\ell-1}' - Z_{\ell-1,\ell}'\|_{\mathcal{Z}} \leq C N_\ell^{-1/\lambda}$$

and

$$\left| \frac{1}{Z_\ell} - \frac{1}{Z_{\ell-1}} \right| = \frac{|Z_{\ell-1} - Z_\ell|}{|Z_\ell Z_{\ell-1}|} \leq C(h_{\ell-1}^t + s_{\ell-1}^{-\theta(1/p_0 - 1/q)}).$$

Adding and subtracting the term $Z'_{\ell-1,\ell}(1/Z_\ell - 1/Z_{\ell,\ell})$, we may rewrite B_ℓ as

$$B_\ell = B_\ell - Z'_{\ell-1,\ell}\left(\frac{1}{Z_{\ell-1}} - \frac{1}{Z_{\ell,\ell}}\right) + Z'_{\ell-1,\ell}\left(\frac{1}{Z_{\ell-1}} - \frac{1}{Z_{\ell,\ell}}\right)$$

$$= (Z'_{\ell,\ell} - Z'_{\ell-1,\ell})\left(\frac{1}{Z_\ell} - \frac{1}{Z_{\ell,\ell}}\right) + Z'_{\ell-1,\ell}\left(\frac{1}{Z_\ell} - \frac{1}{Z_{\ell,\ell}} - \left(\frac{1}{Z_{\ell-1}} - \frac{1}{Z_{\ell-1,\ell}}\right)\right). \tag{4.32}$$

Using Proposition 3.4 and the fact that $|\Theta_\ell(\mathbf{z})|$ is uniformly bounded with respect to ℓ and $\mathbf{z} \in \mathcal{T}_\rho$, we have

$$|Z_{\ell,\ell} - Z_\ell| = |Q_{N_\ell}[\Theta_\ell] - I[\Theta_\ell]| \leq CN_\ell^{-1/\lambda}. \tag{4.33}$$

If $|Z_\ell| > C_0$, $|Z_{\ell,\ell}| > C_0$ uniformly with respect to $\ell = 0,1,2,\ldots$, for some $C_0 > 0$, then there exists $C > 0$ (depending on C_0, but independent of ℓ or N_ℓ) such that

$$\left|\frac{1}{Z_\ell} - \frac{1}{Z_{\ell,\ell}}\right| \leq \frac{|Z_{\ell,\ell} - Z_\ell|}{|Z_\ell||Z_{\ell,\ell}|} \leq CN_\ell^{-1/\lambda}. \tag{4.34}$$

Using similar techniques as in the estimates of (4.27),

$$\|Z'_{\ell,\ell} - Z'_{\ell-1,\ell}\|_{\mathcal{Z}} \leq C(h^t_{\ell-1} + s_{\ell-1}^{-\theta(1/p_0 - 1/q)}).$$

Thus, for any sequence $\{h_\ell\}_{\ell \geq 0}$ or $\{s_\ell\}_{\ell \geq 0}$, the $\|\cdot\|_{\mathcal{Z}}$-norm of the first term of (4.32) is bounded by $O(N_\ell^{-1/\lambda}(h^t_{\ell-1} + s_{\ell-1}^{-\theta(1/p_0 - 1/q)}))$, for any $p_t \leq \lambda, q < 1$.

For the second term of (4.32), adding and subtracting the same term $(Z_{\ell-1,\ell} - Z_{\ell-1})/(Z_\ell Z_{\ell,\ell})$

$$\frac{1}{Z_\ell} - \frac{1}{Z_{\ell,\ell}} - \left(\frac{1}{Z_{\ell-1}} - \frac{1}{Z_{\ell-1,\ell}}\right) = \frac{Z_{\ell,\ell} - Z_\ell}{Z_\ell Z_{\ell,\ell}} - \frac{Z_{\ell-1,\ell} - Z_{\ell-1}}{Z_\ell Z_{\ell,\ell}} + \frac{Z_{\ell-1,\ell} - Z_{\ell-1}}{Z_\ell Z_{\ell,\ell}} - \frac{Z_{\ell-1,\ell} - Z_{\ell-1}}{Z_{\ell-1} Z_{\ell-1,\ell}}$$

$$= \frac{1}{Z_\ell Z_{\ell,\ell}}[(Z_{\ell,\ell} - Z_\ell) - (Z_{\ell-1,\ell} - Z_{\ell-1})] \tag{4.35}$$

$$+ \left(\frac{1}{Z_\ell Z_{\ell,\ell}} - \frac{1}{Z_{\ell-1} Z_{\ell-1,\ell}}\right)(Z_{\ell-1,\ell} - Z_{\ell-1}). \tag{4.36}$$

We bound (4.35) by similar techniques as in the estimates (4.21) to arrive at the bound

$$CN_\ell^{-1/\lambda}(h^t_{\ell-1} + s_{\ell-1}^{-\theta(1/p_0 - 1/q)}).$$

For the remaining term (4.36), we rewrite the first factor as

$$\frac{1}{Z_\ell Z_{\ell,\ell}} - \frac{1}{Z_{\ell-1} Z_{\ell-1,\ell}} = \frac{Z_{\ell-1} Z_{\ell-1,\ell} - Z_\ell Z_{\ell,\ell}}{Z_\ell Z_{\ell,\ell} Z_{\ell-1} Z_{\ell-1,\ell}} = \frac{Z_{\ell-1,\ell}(Z_{\ell-1} - Z_\ell) + Z_\ell(Z_{\ell-1,\ell} - Z_{\ell,\ell})}{Z_\ell Z_{\ell,\ell} Z_{\ell-1} Z_{\ell-1,\ell}}. \tag{4.37}$$

Using similar estimates as in (4.27), we conclude that there exists $C > 0$ independent of $\{s_\ell\}_{\ell \geq 0}$ and of $\{h_\ell\}_{\ell \geq 0}$ such that

$$\begin{aligned}|Z_{\ell-1} - Z_\ell| &= |I(\Theta_{\ell-1} - \Theta_\ell)| \leq C(h^t_{\ell-1} + s_{\ell-1}^{-\theta(1/p_0 - 1/q)}),\\ |Z_{\ell-1,\ell} - Z_{\ell,\ell}| &= |Q_{N_\ell}(\Theta_{\ell-1} - \Theta_\ell)| \leq \sup_{\mathbf{z} \in \mathcal{T}_\rho}|\Theta_{\ell-1}(\mathbf{z}) - \Theta_\ell(\mathbf{z})| \leq C(h^t_{\ell-1} + s_{\ell-1}^{-\theta(1/p_0 - 1/q)}).\end{aligned} \tag{4.38}$$

All $|Z_\ell|$ are bounded above and below away from zero for $\ell = 0,1,2,\ldots$, and the first factor of (4.36) is bounded by the RHS of (4.38). The second factor of (4.36) is estimated by

$$|Z_{\ell-1,\ell} - Z_{\ell-1}| = |Q_{N_\ell}[\Theta_{\ell-1}] - I[\Theta_{\ell-1}]| \leq CN_\ell^{-1/\lambda}.$$

Combining all estimates for (4.35) and (4.36), we conclude that there exists a constant $C > 0$ such that, for every $p_t \leq \lambda, q \leq 1$, and for all sequences $\{h_\ell\}_{\ell \geq 0}$, $\{s_\ell\}_{\ell \geq 0}$ and $\{N_\ell\}_{\ell \geq 0}$ holds

$$\left|\left(\frac{1}{Z_\ell} - \frac{1}{Z_{\ell,\ell}}\right) - \left(\frac{1}{Z_{\ell-1}} - \frac{1}{Z_{\ell-1,\ell}}\right)\right| \leq CN_\ell^{-1/\lambda}(h^t_{\ell-1} + s_{\ell-1}^{-\theta(1/p_0 - 1/q)}).$$

Therefore, both A_ℓ and B_ℓ in (4.31), (4.32) are bounded by $C(N_\ell^{-1/\lambda}(h^t_{\ell-1} + s_{\ell-1}^{-(1/p_0 - 1/q)}))$ (with constant $C > 0$ independent of $\{h_\ell\}_{\ell \geq 0}$, $\{s_\ell\}_{\ell \geq 0}$ or $\{N_\ell\}_{\ell \geq 0}$). Summing the preceding estimates over all discretization levels, we arrive at the error bound (4.30) for the splitting estimator. $\qquad\square$

4.3.4 Selection of truncation dimensions $\{s_\ell\}_{\ell \geq 0}$ and sample numbers $\{N_\ell\}_{\ell \geq 0}$

We assume in the following that assumption **ANPGC**$_{t'}$ holds for some $t' \geq 0$, and define $\tau = t + t'$. The error analysis of the computable MLHoQMC Bayesian estimators in Sections 4.3.2 and 4.3.3 lead to error bounds (4.29) and (4.30), respectively, which are of the generic form

$$
h_L^\tau + s_L^{-\theta(1/p_0 - 1)} + N_0^{-1/p_0} + \sum_{\ell=1}^{L} N_\ell^{-1/\lambda}(h_{\ell-1}^\tau + s_{\ell-1}^{-\theta(1/p_0 - 1/q)}),
\tag{4.39}
$$

with $t' = 0$, and which is analogous to the bounds proved for MLHoQMC algorithms for forward UQ in [7] under **ANPGC**$_{t'}$ with some $t' > 0$. In (4.39), the maximal rate in N_ℓ is obtained for the choice $\lambda = p_t$, implying $\alpha = \lfloor 1/p_t \rfloor + 1$ for the digit interlacing factor. Moreover, to obtain the maximal rate for $s_{\ell-1}^{-\theta(1/p_0 - 1/q)}$ we choose q as small as possible, which yields $q = p_t$ since $p_t \leq q < 1$.

Choice of truncation dimension. In a first step, we balance the dimension truncation error on level L with the discretization error on level L, and similarly for the increment levels $\ell = 1, \ldots, L$. We assume in the following the behavior $h_\ell = 2^{-(\ell+\ell_0)}$ for $\ell = 0, \ldots, L$ and $\ell_0 \in \mathbb{N}_0$. For the terms on level L, we have $s_L^{-\theta(1/p_0 - 1)} = \mathcal{O}(h_L^\tau)$, which motivates the choice $s_L = 2^{\frac{p_0 \tau (L+\ell_0)}{\theta(1-p_0)}}$. Using $\frac{1}{p_t} = \frac{1}{p_0} - \frac{t}{d}$ (cp. [7, Eqn. (1.12)]), we have for the increments $s_\ell^{-\theta(1/p_0 - 1/p_t)} = s_\ell^{-\theta t/d} = \mathcal{O}(h_\ell^\tau)$, which leads to $s_\ell = 2^{\tau d(\ell+\ell_0)/(\theta t)}$. Since the discretization error on mesh level L limits the accuracy of the entire computation, increasing the parameter space dimension beyond s_L is unnecessary; we therefore choose the truncation dimensions

$$
s_\ell = \left\lceil \min(2^{\tau d(\ell+\ell_0)/(\theta t)}, 2^{\frac{p_0 \tau (L+\ell_0)}{\theta(1-p_0)}}) \right\rceil .
\tag{4.40}
$$

Error and work models. We now use $h_{\ell-1}/h_\ell = 2$ (any sequence $\{h_\ell\}_{\ell \geq 0}$ with uniformly bounded ratios would be admissible in the error analysis) and the preceding equilibration of the dimension truncation and FEM discretization errors to obtain an expression for the total error. We replace N_0^{-1/p_0} by N_0^{-1/p_t}, since for $p_0 \leq p_t$ holds $N_0^{-1/p_0} \leq N_0^{-1/p_t}$ for $N_0 \in \mathbb{N}$,

$$
E_{tot} = \mathcal{O}\left(h_L^\tau + \sum_{\ell=0}^{L} N_\ell^{-1/p_t} h_\ell^\tau \right).
\tag{4.41}
$$

The total work (cost) is given by

$$
W_{tot} = \mathcal{O}\left(\sum_{\ell=0}^{L} N_\ell h_\ell^{-d} s_\ell \right).
\tag{4.42}
$$

Optimal number of samples. We now seek to minimize the error (4.41) given a fixed work budget for (4.42), which can be found using a Lagrange multiplier Λ. We assume all constants in the asymptotic error bounds equal 1. We consider the Lagrangian

$$
\mathcal{L}(\{N_\ell\}_{\ell=0}^{L}, \Lambda) := h_L^\tau + \sum_{\ell=0}^{L} N_\ell^{-1/p_t} h_\ell^\tau + \Lambda \sum_{\ell=0}^{L} N_\ell h_\ell^{-d} s_\ell .
$$

Now, we consider the necessary condition that $\partial \mathcal{L}/\partial N_\ell = 0$ for all $\ell = 0, \ldots, L$. For now, we consider N_0 to be a free parameter (to be determined), and use this to find an expression for the multiplier Λ from the following condition:

$$
\frac{\partial \mathcal{L}}{\partial N_0} = -\frac{N_0^{-1/p_t - 1}}{p_t} h_0^\tau + \Lambda h_0^{-d} s_0 = 0 \quad \Rightarrow \quad \Lambda = \frac{h_0^{\tau+d} N_0^{-(1+p_t)/p_t}}{p_t s_0}.
\tag{4.43}
$$

Note that the assumption $h_0 = s_0 = 1$ can lead to a large increase in the sample numbers in the MC case, if it is not fulfilled. In the numerical experiments ahead, this assumption does not hold, since we use $h_\ell = 2^{-(\ell+\ell_0)}$ with $\ell_0 = 1$.

Inserting (4.43) into the remaining conditions for $\ell = 1, \ldots, L$ yields

$$N_\ell \sim N_0 \left(\frac{h_\ell^{\tau+d} s_0}{h_0^{\tau+d} s_\ell} \right)^{p_t/(1+p_t)} , \quad \ell = 1, \ldots, L .$$

Inserting this expression for N_ℓ into (4.41), we set $C = (h_0^{\tau+d} s_0^{-1})^{1/(1+p_t)}$ and write the total error as

$$E_{tot} = \mathcal{O}(h_L^\tau + C N_0^{-1/p_t} E), \quad \text{with } E := \sum_{\ell=0}^L \left(s_\ell h_\ell^{\tau p_t - d} \right)^{1/(1+p_t)},$$

and determine N_0 by equilibrating the two contributions $C N_0^{-1/p_t} E = \mathcal{O}(h_L^\tau)$, which yields

$$N_0 = \mathcal{O}\big((h_L^\tau (EC)^{-1})^{-p_t}\big) = \mathcal{O}(2^{(L-\ell_0) t p_t} (EC)^{p_t}) .$$

Since we consider interlaced polynomial lattice rules with $N_\ell = b^{m_\ell}$ points, we choose m_ℓ for $b = 2$ as

$$m_\ell = \begin{cases} \lceil p_t(\tau(L + \ell_0) + \log_2 E) - \frac{p_t}{1+p_t}(\ell_0(\tau + d) + \log_2 s_0) \rceil & \ell = 0 \\ \lceil m_0 - \frac{p_t}{1+p_t}(\ell(\tau + d) + \log_2(s_\ell/s_0)) \rceil & \ell = 1, \ldots, L \end{cases} . \quad (4.44)$$

4.3.5 Multilevel Monte Carlo (MLMC) sampling

An alternative to QMC evaluation of the ratio and splitting estimators is to approximate the prior expectations Z and Z' by MC sampling. In MC, the convergence rate is $O(N^{-1/2})$ (in mean square, however) in terms of the number N of MC samples, independent of the summability exponents p_0, p_t. The MLMC variant of MC appears as a particular case of the above algorithms; a general optimization of the level-dependent sample numbers N_ℓ yields the same truncation dimensions s_ℓ as in (4.40) above and the following choice of number of samples. Note, however, that in MC sampling we do not simply set $p_0 = p_t = 2$, since the truncation dimensions s_ℓ still depend on the value of the summability exponent $0 < p_0 < 1$. We obtain with $E_{MC} := \sum_{\ell=0}^L (s_\ell h_\ell^{2\tau-d})^{1/3}$ and $C_{MC} = (h_0^{\tau+d} s_0^{-1})^{1/3}$ the sample numbers

$$N_0 = 2^{2Lt} \lceil (C_{MC} E_{MC})^2 \rceil , \quad N_\ell = N_0 \lceil \big(h_\ell^{\tau+d} h_0^{-\tau-d} s_\ell^{-1} s_0 \big)^{2/3} \rceil , \quad \ell = 1, \ldots, L . \quad (4.45)$$

5 Model forward problems

We present examples of forward problems and discretizations which verify the abstract hypotheses of the foregoing error analysis. In Section 5.1 we verify that properties $\mathbf{HCP}_t$, $\mathbf{PGStabC}$ and $\mathbf{ANPGC}_{t'}$ of Section 4.3.1 hold for rather general classes of linear, affine-parametric operator equations. Next, in Section 5.2, we consider a concrete linear elliptic problem in two space dimensions with a particular, explicit selection of the uncertainty parametrization (2.1). We remark that analogous hypotheses are valid for a host of more general problems with high-dimensional uncertainty parametrization; we mention only parametric systems of nonlinear ODEs [15], and problems with domain uncertainty.

5.1 Affine-parametric linear operator equations

We verify the abstract hypotheses of the QMC-PG error analysis in Section 4.3.1 for the model problem (5.8) below; in doing so, rather than considering the particular problem (5.8), we consider the more general linear, affine parametric operator equations as considered in [23].

For a regularity parameter $0 \le t \le \bar{t} \le \infty$, let $\{A_j\}_{j\ge 0} \subset \mathcal{L}(\mathcal{X}_t, \mathcal{Y}_t')$ be a sequence of bounded, linear operators which satisfy the following hypotheses.

$\mathbf{H1}_t$ A_0 is boundedly invertible, i.e. $A_0^{-1} \in \mathcal{L}(\mathcal{Y}_t', \mathcal{X}_t)$.
$\mathbf{H2}_t$ For $r = 0, 1, \ldots, t$, the sequences $\boldsymbol{b}_r = (b_{j,r})_{j\ge 1}$ of norms $b_{j,r} := \max\{\|A_0^{-1} A_j\|_{\mathcal{L}(\mathcal{X}_r)}, \|(A_0^*)^{-1} A_j^*\|_{\mathcal{L}(\mathcal{Y}_r)}\}$ satisfy $\boldsymbol{b}_r \in \ell^{p_r}(\mathbb{N})$ for summability exponents $0 < p_0 \le p_1 \le \ldots \le p_t < 1$.
$\mathbf{H3}_t$ There holds $\|\boldsymbol{b}_t\|_{\ell^{p_t}(\mathbb{N})} < 1$.

Under $\mathbf{H1}$ - $\mathbf{H3}$, we consider for parameter sequences $\boldsymbol{y} = (y_j)_{j\ge 1} \in U = [-1,1]^{\mathbb{N}}$, and for $\{A_j\}_{j\ge 0} \subset \mathcal{L}(\mathcal{X}, \mathcal{Y}')$ the linear, affine-parametric operators $A(\boldsymbol{y}; q)$ given by

$$A(\boldsymbol{y}; q) = A(\boldsymbol{y})q = A_0 q + \sum_{j\ge 1} y_j A_j q . \quad (5.1)$$

Evidently, then, the differential $D_q\mathcal{R}$ of the residual map $\mathcal{R}(\boldsymbol{y};q) := A(\boldsymbol{y})q - f$ equals $A(\boldsymbol{y})$, and the (uniform w.r.t. $\boldsymbol{y} \in U$) bounded invertibility of $A(\boldsymbol{y})$ implies a uniform (w.r.t. $\boldsymbol{y}$) inf-sup condition for $A(\boldsymbol{y})$. Hypotheses $\mathbf{H1}_t$ - $\mathbf{H3}_t$ imply, with a Neumann series argument, properties $\mathbf{HCP}_t$ and $\mathbf{ANPGC}_t$ in Section 4.3.1 for the affine-parametric family (5.1) of linear operators. Specifically, for every $f \in \mathcal{Y}_t'$ and for every $\boldsymbol{y} \in U$ holds in $\mathcal{X}_t$ (with $B_j := A_0^{-1}A_j \in \mathcal{L}(\mathcal{X}_t)$ for $j \geq 1$)

$$q(\boldsymbol{y}) = (A(\boldsymbol{y}))^{-1}f = (I + \sum_{j\geq 1} y_j B_j)^{-1}A_0^{-1}f . \tag{5.2}$$

Denoting by $\mathcal{X}_t$ and $\mathcal{Y}_t$ "complexifications" of the function spaces, the affine-parametric operator family $\{A(\boldsymbol{y}) : \boldsymbol{y} \in U\}$ in (5.1) extends to the polydisc $\mathcal{D} = \{\boldsymbol{z} : |z_j| \leq 1\}$, and the Neumann series inversion formula (5.2) remains valid for $\boldsymbol{z} = (z_j)_{j\geq 1} \in \mathcal{D}$. The extended family $A(\boldsymbol{z})$ being affine-parametric is holomorphic with respect to each variable z_j and, by $\mathbf{H3}_t$, is also boundedly invertible in $\mathcal{L}(\mathcal{X}_t, \mathcal{Y}_t')$ for all $\boldsymbol{z} \in \mathcal{D}_{\boldsymbol{\rho}_t}$, provided the poly-radius $\boldsymbol{\rho}_t = (\rho_{t,j})_{j\geq 1}$ is $(\boldsymbol{b}_t, \varepsilon)$-admissible, i.e. provided that (3.24) holds for some $\varepsilon > 0$. Since the mapping $A \mapsto A^{-1}$ is analytic at any $A \in \mathcal{L}_{iso}(\mathcal{X}, \mathcal{Y}')$, (5.2) defines a parametric solution family $\{q(\boldsymbol{z}) : \boldsymbol{z} \in \mathcal{D}_{\boldsymbol{\rho}_t}\}$, which, for $\boldsymbol{z} \in \mathcal{D}_{\boldsymbol{\rho}_t}$ and some $\varepsilon > 0$, is $(\boldsymbol{b}_t, p_t, \varepsilon)$-holomorphic with holomorphy domains $\mathcal{O}_{\rho_j}$ in Definition 3.1 being the polydiscs $\mathcal{D}_{\rho_j}$. For every $(\boldsymbol{b}_t, \varepsilon)$-admissible $\boldsymbol{\rho}_t$ there holds the *holomorphic continuability of the parametric solutions* $\mathbf{HCP}_t$ for some $t > 0$, so that

$$\sup_{\boldsymbol{z}\in\mathcal{D}_{\boldsymbol{\rho}_t}} \|q(\boldsymbol{z})\|_{\mathcal{X}_t} \leq C(\boldsymbol{b}_t, \varepsilon)\|f\|_{\mathcal{Y}_t'} . \tag{5.3}$$

As for every $\kappa > 1$, $\mathcal{D}_\kappa \supset \mathcal{T}_\kappa$, Proposition 3.4 is applicable to $g(\boldsymbol{y}) := \phi(q(\boldsymbol{y}))$ for any QoI $\phi(\cdot) \in \mathcal{L}(\mathcal{X}_{t'}, \mathcal{Z})$.

The Neumann series argument used to verify (5.3) also implies condition $\mathbf{PGStabC}$ (Stability and Quasioptimality of PG for the complex parametric extension). To see this, we assume at hand a nested sequence $\{(\mathcal{X}^\ell, \mathcal{Y}^\ell)\}_{\ell\geq 0}$ of (dense in $\mathcal{X} \times \mathcal{Y}$) finite-dimensional PG trial- and test function spaces, which satisfy the *nominal discrete inf-sup condition*: there exists $\mu_0 > 0$ such that

$$\inf_{0\neq w\in\mathcal{X}^\ell} \sup_{0\neq v\in\mathcal{Y}^\ell} \frac{|_{\mathcal{Y}'}\langle A_0 w, v\rangle_{\mathcal{Y}}|}{\|w\|_{\mathcal{X}}\|v\|_{\mathcal{Y}}} \geq \mu_0 , \quad \inf_{0\neq v\in\mathcal{Y}^\ell} \sup_{0\neq w\in\mathcal{X}^\ell} \frac{|_{\mathcal{Y}'}\langle A_0 w, v\rangle_{\mathcal{Y}}|}{\|w\|_{\mathcal{X}}\|v\|_{\mathcal{Y}}} \geq \mu_0 . \tag{5.4}$$

With (5.4), assumption $\mathbf{H3}_0$ implies $\mathbf{PGStabC}$, i.e. Stability and Quasioptimality of the PG discretizations for the complex parametric extension to $\mathcal{D}_{\boldsymbol{\rho}_0}$ for any $(\boldsymbol{b}_0, \varepsilon)$-admissible polyradius $\boldsymbol{\rho}_0$: for such $\boldsymbol{z} \in \mathcal{D}_{\boldsymbol{\rho}_0}$ holds

$$|_{\mathcal{Y}'}\langle A(\boldsymbol{z})w, v\rangle_{\mathcal{Y}}| = \left|_{\mathcal{Y}'}\left\langle \left(A_0 + \sum_{j\geq 1} z_j A_j\right)w, v\right\rangle_{\mathcal{Y}}\right| \geq |_{\mathcal{Y}'}\langle A_0 w, v\rangle_{\mathcal{Y}}| - \left|_{\mathcal{Y}'}\left\langle \left(\sum_{j\geq 1} z_j A_j\right)w, v\right\rangle_{\mathcal{Y}}\right|,$$

which implies with (5.4) and with $|z_j| \leq \rho_j$ that

$$\forall \boldsymbol{z} \in \mathcal{D}_{\boldsymbol{\rho}} : \quad \inf_{0\neq v\in\mathcal{Y}^\ell} \sup_{0\neq w\in\mathcal{X}^\ell} \frac{|_{\mathcal{Y}'}\langle A(\boldsymbol{z})w, v\rangle_{\mathcal{Y}}|}{\|w\|_{\mathcal{X}}\|v\|_{\mathcal{Y}}} \geq \mu_0(1 - \sum_{j\geq 1} \mu_0^{-1}\rho_j\|A_j\|_{\mathcal{L}(\mathcal{X},\mathcal{Y}')}) . \tag{5.5}$$

Assume that $\boldsymbol{\rho}$ is $(\boldsymbol{\beta}, \varepsilon)$-admissible for the sequence $\boldsymbol{\beta} = (\beta_j)_{j\geq 1}$ given by $\beta_j := \mu_0^{-1}\|A_j\|_{\mathcal{L}(\mathcal{X},\mathcal{Y}')}$ for $j \geq 1$, and for sufficiently small $\varepsilon > 0$. Hypothesis $\mathbf{H1}_0$, i.e. that $A_0 \in \mathcal{L}(\mathcal{X}, \mathcal{Y}')$ is an isomorphism, implies $b_j = \|A_0^{-1}A_j\|_{\mathcal{L}(\mathcal{X})} \simeq \|A_j\|_{\mathcal{L}(\mathcal{X},\mathcal{Y}')} \simeq \beta_j$. By $\mathbf{H2}$ and $\mathbf{H3}$, $\boldsymbol{b} \in \ell^{p_0}(\mathbb{N})$ implies with $\mathbf{H1}$ that $\boldsymbol{\beta} \in \ell^{p_0}(\mathbb{N})$. The $(\boldsymbol{\beta}, \varepsilon)$-admissibility of $\boldsymbol{\rho}$ and $\mathbf{H3}_t$ with $t = 0$ imply

$$1 - \sum_{j\geq 1} \mu_0^{-1}\rho_j\|A_j\|_{\mathcal{L}(\mathcal{X},\mathcal{Y}')} = 1 - \sum_{j\geq 1} \beta_j - \sum_{j\geq 1}(\rho_j - 1)\beta_j \geq 1 - \varepsilon - \sum_{j\geq 1} \beta_j > 0,$$

which implies in (5.5) the uniform (w.r.t. $\boldsymbol{z} \in \mathcal{D}_{\boldsymbol{\rho}}$) parametric discrete inf-sup conditions

$$\inf_{0\neq v\in\mathcal{Y}^\ell} \sup_{0\neq w\in\mathcal{X}^\ell} \frac{|_{\mathcal{Y}'}\langle A(\boldsymbol{z})w, v\rangle_{\mathcal{Y}}|}{\|w\|_{\mathcal{X}}\|v\|_{\mathcal{Y}}} \geq \mu, \quad \inf_{0\neq w\in\mathcal{X}^\ell} \sup_{0\neq v\in\mathcal{Y}^\ell} \frac{|_{\mathcal{Y}'}\langle A(\boldsymbol{z})w, v\rangle_{\mathcal{Y}}|}{\|w\|_{\mathcal{X}}\|v\|_{\mathcal{Y}}} \geq \mu, \tag{5.6}$$

with $\mu = \mu_0(1 - \varepsilon - \sum_{j\geq 1} \beta_j) > 0$ independent of ℓ. The uniform parametric inf-sup conditions (5.6) imply uniform w.r.t. $\boldsymbol{z} \in \mathcal{D}_{\boldsymbol{\rho}}$ stability of the PG discretization and the existence, uniqueness and the

quasioptimality of the complex-parametric PG solution $q^{(\ell)}(z) \in \mathcal{X}^\ell$, for every $z \in \mathcal{D}_\rho$ with (β, ε)-admissible ρ. It also implies the (uniform w.r.t. the discretization level ℓ) $(\beta, p_0, \varepsilon)$-holomorphy of the PG solutions $q^{(\ell)}(z)$. The quasioptimality and the uniform parametric regularity (5.3) imply, with the approximation property (2.19), the (uniform w.r.t. z) asymptotic error bounds

$$\sup_{z \in \mathcal{D}_\rho} \|q(z) - q^{(\ell)}(z)\|_{\mathcal{X}} \leq C h_\ell^t \|f\|_{\mathcal{Y}_t'} , \tag{5.7}$$

where the constant $C > 0$ is independent of ℓ, f and of $z \in \mathcal{D}_\rho$, but depends on β and ε in the (β, ε)-admissibility of the poly-radius ρ. Choosing $z = (z_1, ..., z_{s_\ell}, 0, ...)$, (5.7) implies (4.23). For the linear, affine-parametric operators (5.1), property $\mathbf{ANPGC}_{t'}$ for some $0 \leq t' \leq \bar{t}$ was shown in [7], for the affine-parametric, linear operator family (5.1).

5.2 Affine-parametric linear test problem

We test the proposed ML algorithms for a model parametric, linear diffusion problem which was already considered in [7].

For a parameter $y \in U = [-\frac{1}{2}, \frac{1}{2}]^\mathbb{N}$, in the bounded spatial domain $D \subset \mathbb{R}^d$, we consider the second order, linear and affine-parametric elliptic PDE

$$A(u; q) = -\nabla \cdot (u(y)\nabla q(y)) = f , \quad q(y)|_{\partial D} = 0 , \quad u(y) = u_0(\cdot) + \sum_{j \geq 1} y_j \, \psi_j(\cdot) . \tag{5.8}$$

This is a particular (linear) forward problem (2.5). The usual (symmetric) primal variational formulation of (5.8) is based on $\mathcal{X} = \mathcal{Y} = H_0^1(D)$. In the particular case that $D = (0,1)^d$, we parametrize the uncertain diffusion coefficient u with the explicit, separable Karhunen-Loève basis

$$\lambda_{\boldsymbol{k}} = \pi^2(k_1^2 + \cdots + k_d^2), \quad \tilde{\psi}_{\boldsymbol{k}}(\boldsymbol{x}) = \prod_{i=1}^d \sin(\pi k_i x_i) . \tag{5.9}$$

Enumerating $\{\lambda_{\boldsymbol{k}}\}_{\boldsymbol{k} \in \mathbb{N}^d}$ in non-decreasing order $\{\lambda_j\}_{j \geq 1}$, there holds

$$\lambda_j \sim j^{2/d} \quad \text{as} \quad j \to \infty . \tag{5.10}$$

We consider in the ensuing numerical experiments the case $d = 2$, $D = (0,1)^2$, and

$$u(\boldsymbol{y})(\boldsymbol{x}) = u_0(\boldsymbol{x}) + \sum_{k_1, k_2 = 1}^\infty y_{k_1, k_2} \frac{1}{(k_1^2 + k_2^2)^2} \sin(k_1 \pi x_1) \sin(k_2 \pi x_2)$$

$$= u_0(\boldsymbol{x}) + \sum_{j=1}^\infty y_j \, \mu_j \, \sin(k_{1,j} \, \pi x_1) \sin(k_{2,j} \, \pi x_2) . \tag{5.11}$$

We enumerate the sequence of pairs $((k_{1,j}, k_{2,j}))_{j \in \mathbb{N}}$ such that $k_{1,j}^2 + k_{2,j}^2 \leq k_{1,j+1}^2 + k_{2,j+1}^2$ for all $j \in \mathbb{N}$ (with ties due to equality broken in an arbitrary manner). This ordering yields $\mu_j = (k_{1,j}^2 + k_{2,j}^2)^{-2} \asymp \lambda_j^{-2} \sim j^{-2}$ (cf. (5.10)). We take $u_0(\boldsymbol{x}) \equiv 1/2$. In (5.8), we use the forcing term $f(\boldsymbol{x}) = 100 x_1$, and consider the quantity of interest to be the integral of the parametric solution $u(\boldsymbol{y})$ over the subdomain $\bar{D} = (0.5, 1)^2 \subset D$, i.e., $G(q(\boldsymbol{y})) = \int_{\bar{D}} q(\boldsymbol{y})(\boldsymbol{x}) \, d\boldsymbol{x}$. This affine-parametric forward problem fits into the abstract MLHoQMC framework considered in [7] with symmetric, affine-parametric bilinear form $\mathfrak{a}(\boldsymbol{y}; \cdot, \cdot)$, and with $\mathcal{X} = \mathcal{Y} = H_0^1(D)$, and with

$$d = 2, \quad \theta = 2, \quad t = t' = 1, \quad \tau = t + t' = 2, \quad \text{and any} \quad \frac{1}{2} < p_0 \leq 1 . \tag{5.12}$$

As explained in [7] this implies the summability exponent $p_1 = p_0/(1 - p_0/2) > 2/3$. The regularity spaces $\mathcal{X}_t$ in the convergence rate estimate (2.24) are $\mathcal{X}_1 = (H_0^1 \cap H^2)(D)$ and $\mathcal{Y}_1' = L^2(D)$.

For Bayesian inversion, we consider as observation functional the integral over $\hat{D} = (0, 0.5)^2$. This scalar value is perturbed by a realization of a normally distributed random variable $\eta \sim \mathcal{N}(0, \Gamma)$ as in (3.1) to generate a measurement, which is fixed for each value of Γ before applying the various SL and ML HoQMC methods. For prior mean approximations, we compare the multilevel QMC method to

the single-level QMC approach and to the multilevel Monte Carlo method. For approximations to the posterior expectation, we compare the performance of the analyzed multilevel QMC estimators (4.6), (4.10) with both the single-level QMC ratio estimator of [9] as well as the two multilevel estimators (4.6), (4.10) combined with standard Monte Carlo sampling. In all considered algorithms, we solve (5.8) by the finite element method with continuous, piecewise linear elements on a family of uniform quadrilateral meshes with mesh width $h_\ell = h_0 2^{-\ell}$ for $\ell = 0, 1, 2 \ldots$, and we use interlaced polynomial lattice rules with $N = 2^m$ points, $m \in \mathbb{N}$, constructed by the fast CBC algorithm for SPOD weights from [6].

In the single-level HoQMC ratio estimator, the meshwidth is $h = h_L = h_0 2^{-L}$. The PG discretization error for regular functionals is, asymptotically, as $h \to 0$, $\mathcal{O}(h^2)$. We balance this $\mathcal{O}(h^2)$ discretization error with the dimension truncation error of $\mathcal{O}(s^{-2})$ and the HoQMC quadrature error of $\mathcal{O}(N^{-2})$. These asymptotic error bounds yield the choices $s = h^{-1} = 2^{L+1}$ and $N = h^{-1}$, so that $m = \log_2(h^{-1}) \simeq L+1$. Ignoring logarithmic factors, this yields a combined error of $\mathcal{O}(h^2) = \mathcal{O}(\varepsilon)$ and overall cost of $\mathcal{O}(Nh^{-2}s) = \mathcal{O}(h^{-4}) = \mathcal{O}(\varepsilon^{-2})$, with the constants implied in $\mathcal{O}(\cdot)$ being independent of $\{s_\ell\}_{\ell \geq 0}$.

The HoQMC rules will be based on SPOD weights from [17]. A major finding of the single-level theory in [9] and of the multilevel error analysis in the present paper is that *HoQMC rules which are efficient for forward UQ will perform equally well for the corresponding Bayesian Inverse UQ*, due to preservation of holomorphy domains. We therefore use, in the affine parametric forward problem, the HoQMC weights derived in [6]. They are given by [6, Equation (3.32) with (3.17)] and

$$\text{interlacing factor } \alpha = \lfloor 1/p_0 \rfloor + 1 = 2, \text{ and } \beta_j = \beta_{0,j} = \lambda_j = \frac{1}{(k_{1,j}^2 + k_{2,j}^2)^2} \ .$$

The generating vectors were computed by the fast CBC construction from [6] with Walsh constant $C = 0.1$. (computations with $C = 1$ yielded different generating vectors, and led to slight artefacts on high levels $L \geq 7$ in this example). For the presently used base $b = 2$, the choice $C = 1.0$ holds [27].

In the multilevel algorithm Q_L^* in (4.7), (4.8), for given maximal discretization level L, we take regular bisection refinement of the quadrilateral mesh in D, resulting in a sequence of regular, quadrilateral meshes of D with meshwidths $h_\ell = h_0 2^{-\ell}$ for $\ell = 0, 1, \ldots, L$. We select the truncation dimension as $s_\ell = \min(2^{2\ell}, 2^L)$ as in (4.40), and m_ℓ as in (4.44), where for this particular case N_ℓ is given in Table 1. Using formally the limiting values $p_0 = 1/2$ and $p_1 = 2/3$, the total error is $\mathcal{O}(h_L^2) = \mathcal{O}(\varepsilon)$ at cost

L	MLQMC	MLMC
0	(1)	(1)
1	(3,1)	(5,2)
2	(5,3,1)	(10,7,4)
3	(7,5,3,1)	(15,11,8,5)
4	(9,7,5,3,1)	(19,15,12,9,7)
5	(11,9,6,5,3,2)	(24,20,16,13,10,8)
6	(13,11,8,6,5,3,2)	(28,24,20,17,14,11,9)
7	(15,13,10,8,6,5,3,2)	–
8	(17,15,12,10,8,6,5,3,2)	–

Table 1: Logarithm of the number of samples per level $m_\ell = \log_2 N_\ell$ for the multilevel methods used in the results below, i.e. $N_\ell = 2^{m_\ell}$ with m_ℓ given by (4.44) for QMC for both ratio and splitting estimators; for MLMC, we show $m_\ell^{MC} = \lfloor \log_2 N_\ell \rfloor$ with N_ℓ from (4.45). Note that for MLMC we use N_ℓ directly, m_ℓ^{MC} is specified for ease of comparison.

of $\mathcal{O}\left(\sum_{\ell=0}^{L} N_\ell h_\ell^{-2} s_\ell\right) = \mathcal{O}(\varepsilon^{-3/2})$, ignoring logarithmic factors. For $\ell = 0$, we use the SPOD weights from the single-level case with $\beta_{0,j}$ from above. For $\ell > 0$, the SPOD weights that enter the fast CBC construction are different from those for the single-level algorithm; as indicated above, we use the choices derived for this problem in [6] for the affine-parametric forward problem also for the computation of integrals in the inverse problem. We take base $b = 2$, Walsh constant $C_{\alpha,b} = 0.1$, and

$$\text{digit interlacing factor } \alpha = \lfloor 1/p_1 \rfloor + 1 = 2, \text{ and } \beta_j = \beta_{1,j} = \lambda_j \pi \max(k_{1,j}, k_{2,j}) \ .$$

6 Implementation and numerical results

We now present numerical experiments which validate the choices of the algorithm steering parameters from Section 4.3.4 for their implementation, for the fast, *deterministic* solution of Bayesian inverse problems for PDEs with uncertain random field inputs. We present in particular experiments for the model linear, parametric elliptic forward problems from Section 5. The problems are set in the domain $D = (0,1)^2$, with homogeneous Dirichlet boundary conditions, and with parametric coefficients given by an s-term truncated Karhunen-Loève expansion in D, with exactly known eigenfunctions. The first problem has an affine-parametric coefficient, the second problem is a nonlinear (holomorphic) transformation of this coefficient. The purpose of the ensuing numerical experiments is to illustrate the preceding convergence analysis and to show that the proposed MLHoQMC algorithms outperform other methods, such as MLMC, in terms of error vs. work. For the implementation, we use the `gMLQMC` library from [11].

We compute the forward solution up to maximal discretization level $L = 8$, yielding $s = 512$ active dimensions. No exact solution for this problem is available, so we verify convergence rates by testing accuracy with respect to a numerically computed reference solution. This reference solution was computed on level $L = 9$ with maximal truncation dimension $s = 1024$ and the MLQMC method. For the posterior expectation, the splitting estimator was used. In the error vs. work plot in the figures ahead, we used the work measures

$$W_{\mathrm{SL}} := h_L^{-2} s N \,, \quad \text{and} \quad W_{\mathrm{ML}} := \sum_{\ell=0}^{L} N_\ell h_\ell^{-2} s_\ell \,. \tag{6.1}$$

The MLMC runs which are provided here for comparison purposes were performed with identical discretizations of the forward problems, and with the optimized MC sample numbers (4.45); in order to reduce the (inherent in MC sampling) scatter in the convergence rate plots, in the ensuing graphs the MLMC convergence was obtained by averaging 5 MLMC runs. We emphasize that the presently considered MLHoQMC are entirely deterministic: for the MLHoQMC algorithms, the computation of each convergence plot required only one single run.

6.1 Affine parametric, linear elliptic test problem

In the results below, the work measures (6.1) were used for the multilevel methods. The expected convergence rate is $-1/2$ for the single-level algorithm and $-2/3$ for the multilevel algorithm (for both forward and inverse approximations), both of which are confirmed by the results. In the MLMC runs, $R = 5$ repetitions were used to average out sampling noise in the estimated L^2 error. The MLQMC (splitting estimator) result with finest discretization level $L = 9$ was used as a reference in the error computation.
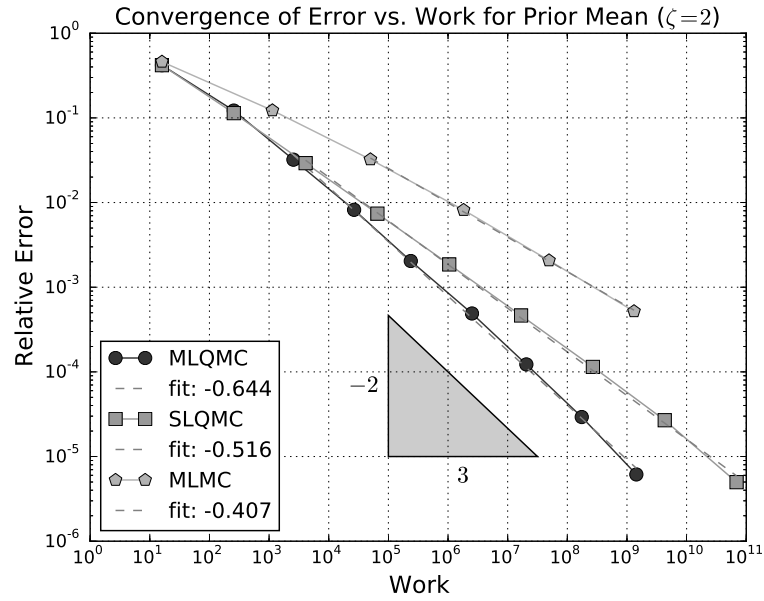
Figure 1: Convergence of forward UQ estimates for the affine-parametric 2d diffusion equation model. The slope of the lines was estimated by a least-squares fit omitting the first three points.
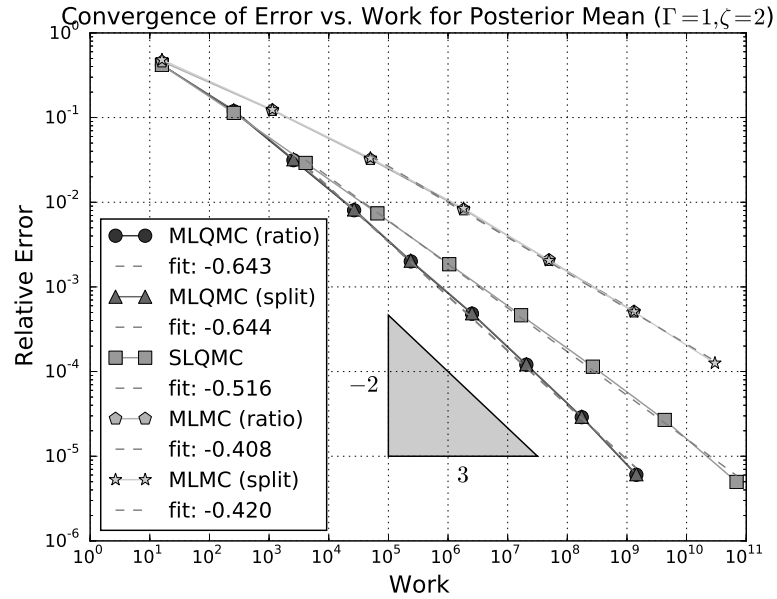

Figure 2: Convergence of estimates to the Bayesian inverse problem for the affine-parametric 2d diffusion equation model, with covariance $\Gamma = 1$. The slope of the lines was estimated by a least-squares fit omitting the first three points.
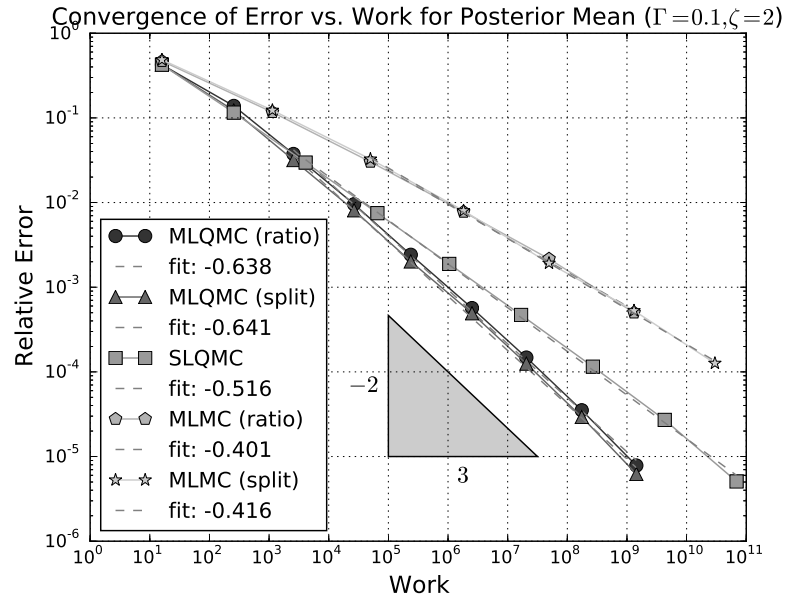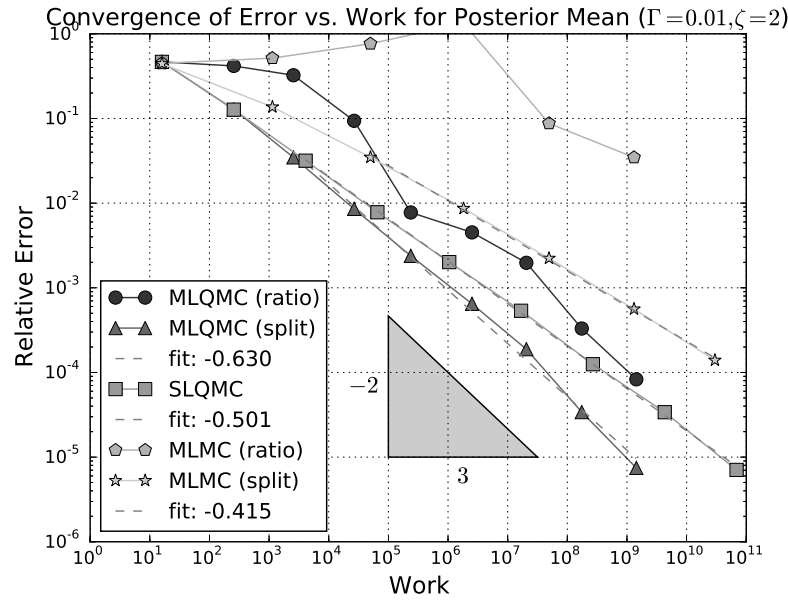
Figure 3: Convergence of estimates to the Bayesian inverse problem for the affine-parametric 2d diffusion equation model, with covariance $\Gamma = 0.1$. The slope of the lines was estimated by a least-squares fit omitting the first three points.

Figure 4: Convergence of estimates to the Bayesian inverse problem for the affine-parametric 2d diffusion equation model, with covariance $\Gamma = 0.01$. The slope of the lines was estimated by a least-squares fit omitting the first three points. In this case, the MLQMC splitting estimator was used as a reference. In the small noise case, the convergence of the ratio estimator is negatively impacted, both for MLMC and for MLQMC. The splitting estimator seems to be more resilient for small Γ, as indicated in Remark 4.1.

6.2 Nonaffine-parametric, linear test problem

We consider once more the linear operator equation (5.8), however, now with diffusion coefficient modelled for $y_j \in [-\frac{1}{2}, \frac{1}{2}]$ by the nonlinear expression

$$u(\cdot, \boldsymbol{y}) := \exp\left(\sum_{j \geq 1} y_j \psi_j(\cdot)\right) , \tag{6.2}$$

i.e. simply the exponential of the affine-parametric coefficient model from (5.8) with $u_0(\cdot) = 0$, yielding for the new model the nominal value $u(\boldsymbol{0}) \equiv 1$. When considering the exact same QMC parameters as for the above results (α, C, β_j) and maximal parameter dimension in the ML experiments being $s = 1024$, we observe the results in Figures 5 to 7.
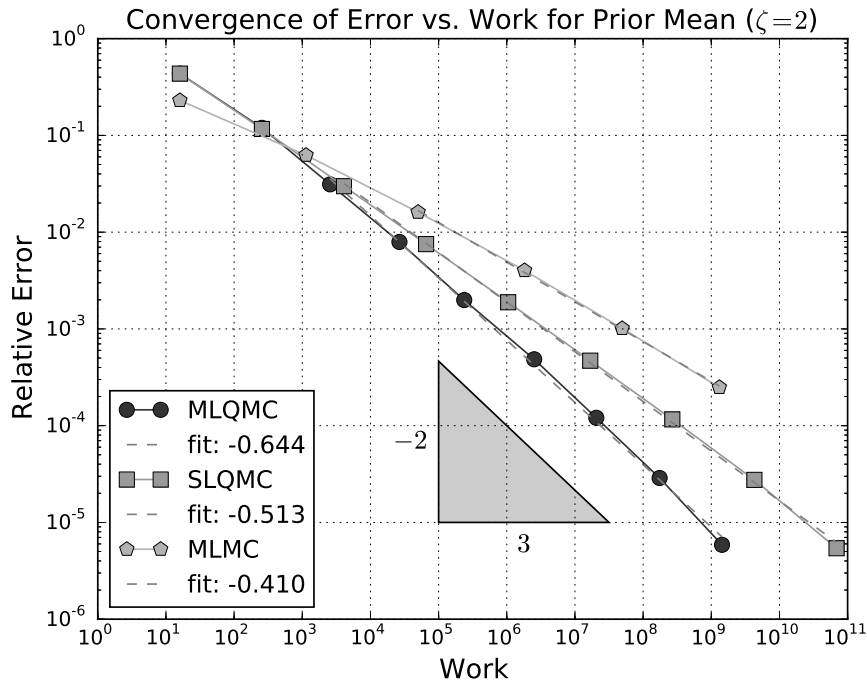
Figure 5: Convergence of forward UQ estimates for the nonaffine-parametric 2d diffusion equation model with coefficient (6.2). The slope of the lines was estimated by a least-squares fit omitting the first three points.

7 Conclusions

We extended [9] to a class of deterministic, multilevel Petrov-Galerkin, higher order Quasi-Monte Carlo integration algorithms for forward and Bayesian inverse computational uncertainty quantification of possibly nonlinear, well-posed operator equations. Novel, computable deterministic multilevel estimators have been proposed for "distributed" uncertain input data in a separable Banach space X. Upon parametrizing the uncertain input data in terms of a countable basis of X (as, e.g., through a Karhunen-Loève expansion), and upon multilevel Petrov-Galerkin discretization of the forward problems, the forward and Bayesian inverse uncertainty quantification problem is reduced to numerical evaluation of high-dimensional, parametric integrals of nonlinear functionals depending on the likelihood function of the responses from the parametric forward problem and the observation data. The numerical integration is conducted with the deterministic, higher order QMC integrations from [8]. The present results generalize, in particular, the HoQMC PG error analysis of [7] to smooth, nonlinear operator equations with holomorphic-parametric dependence of their responses on the parameters. They apply to broad classes
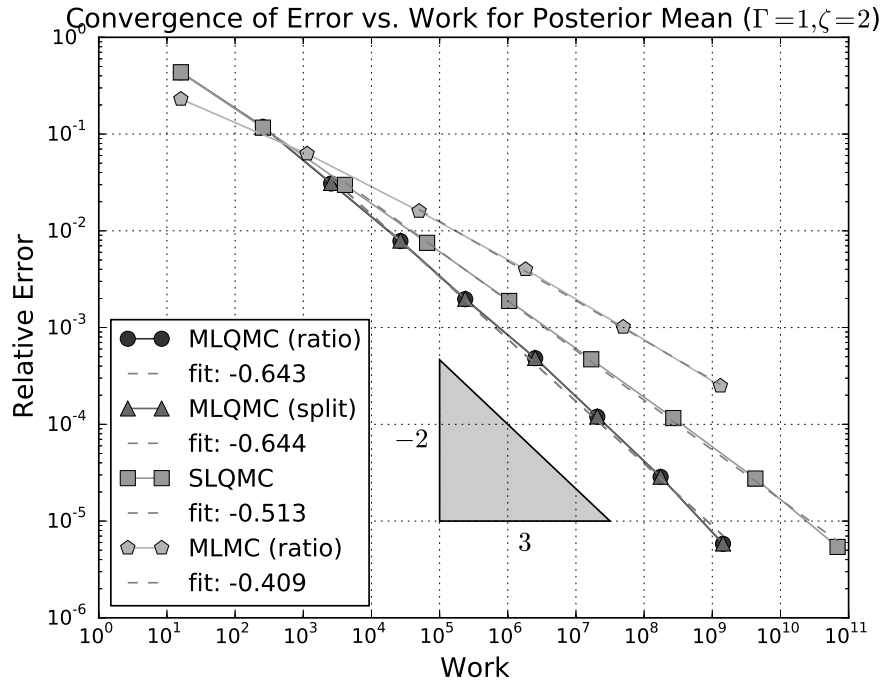
Figure 6: Convergence of estimates to the Bayesian inverse problem for the 2d diffusion equation model with coefficient (6.2), with $\Gamma = 1$. The slope of the lines was estimated by a least-squares fit omitting the first three points.

of forward equations, with possibly indefinite or saddle point variational formulations, and nonlinear, analytic dependence on the parameters.

In several numerical experiments for Bayesian inversion of linear, elliptic forward problems in two space dimensions, the presently proposed, multilevel higher order Quasi Monte-Carlo strategy consistently outperformed the corresponding single-level algorithms from [9], and corresponding MLMC methods in both forward as well as Bayesian inverse UQ on parametric inputs with proper sparsity: to reach one percent accuracy in the Bayesian estimate, the MLHoQMC strategy achieves a speedup of a factor 10 of error versus total work as compared to the SL strategy and to the MLMC approach. Higher efficiency is expected on input data with higher smoothness in the data space, i.e. $u \in X_t$, implies higher sparsity; having said this, we admit that for problems whose parametric input data does not afford sufficient sparsity, the presently proposed methods will not outperform MLMC and ML versions of first order QMC methods.

The presented numerical experiments also confirm the dimension-independence of the QMC convergence rates, which are only limited by input sparsity and by the digit interlacing order of the polynomial lattice rule. They also indicate the expected deterioration of the algorithms' performance for small observation noise covariance Γ; in this respect, the splitting estimator was found to be less sensitive than the ratio estimator. The presently introduced algorithms allow us to handle uncertainties with several hundred to thousands of parameters, in two space dimensions, with moderate computational effort. The parametric sparsity of the countably parametric model problem considered in the numerical experiments was moderate ($p_0 = 1/2$ in (5.12)); for classes of uncertain input data u with higher sparsity, i.e. smaller values of p_0, the gains of the presently proposed, HoQMC-based algorithms over MLMC are predicted to be correspondingly higher, as a consequence of the present theoretical results, and supported by extensive numerical experiments in [12].

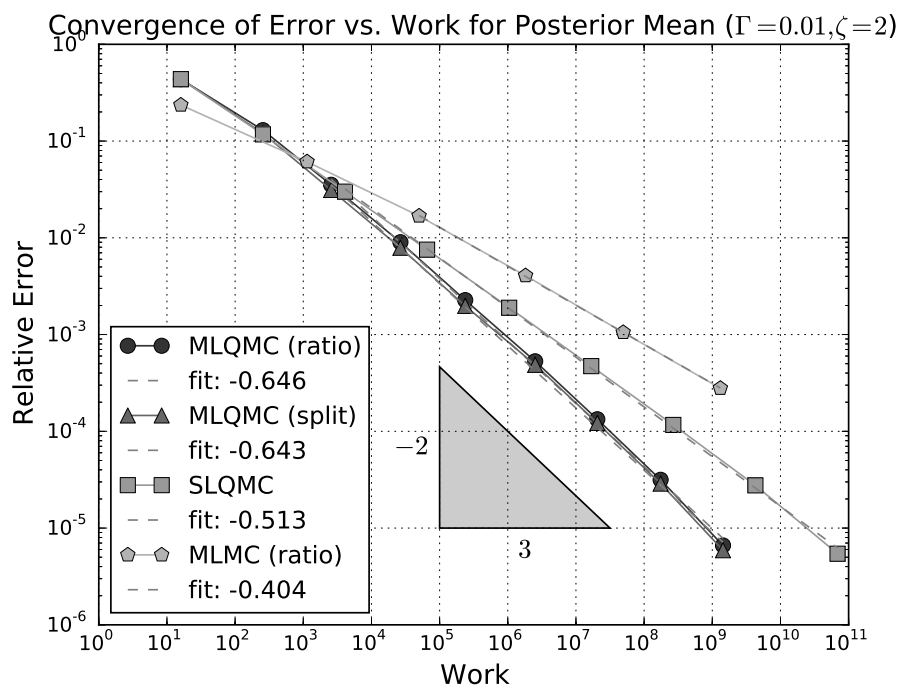Figure 7: Convergence of estimates to the Bayesian inverse problem for the 2d diffusion equation model with coefficient (6.2), with $\Gamma = 0.01$ The slope of the lines was estimated by a least-squares fit omitting the first three points.

References

[1] A. Beskos, A. Jasra, K. Law, R. Tempone and Y. Zhou, Multilevel Sequential Monte Carlo Samplers, Available at arXiv1503.07259v1.

[2] F. Brezzi, J. Rappaz and P. A. Raviart, Finite dimensional approximation of nonlinear problmes I: Branches of nonsingular solutions, Numer. Math., 36 (1980), pp. 1–25.

[3] A. Chkifa, A. Cohen and Ch. Schwab, Breaking the curse of dimensionality in sparse polynomial approximation of parametric PDEs. Journ. Math. Pures et Appliquees, 103, 400–428, 2015.

[4] M. Dashti and A. M. Stuart, The Bayesian Approach to Inverse Problems (to appear in Handbook of Uncertainty Quantification, Springer Publ. 2016) Available at arXiv:1302.6989v4.

[5] J. Dick, Walsh spaces containing smooth functions and Quasi-Monte Carlo rules of arbitrary high order. SIAM J. Numer. Anal., 46, 1519–1553, 2008.

[6] J. Dick, F.Y. Kuo, Q. T. Le Gia, D. Nuyens and Ch. Schwab, Higher order QMC Galerkin discretization for parametric operator equations. SIAM J. Numer. Anal., 52, 2676–2702, 2014.

[7] J. Dick, F.Y. Kuo, Q. T. Le Gia and Ch. Schwab, Multi-level higher order QMC Galerkin discretization for affine parametric operator equations. Report 2014-14 Seminar for Applied Mathematics, ETH Zürich, Switzerland (to appear in SIAM Journ. Numer. Analysis 2016). Available at arXiv:1406.4432

[8] J. Dick, Q. T. Le Gia and Ch. Schwab, Higher order Quasi Monte Carlo integration for holomorphic parametric operator equations. Report 2014-23, Seminar for Applied Mathematics, ETH Zürich, Switzerland (to appear in SIAM Journ. Unc. Quantification 2016). Available at arXiv:1409.2180.

[9] J. Dick, R. N. Gantner, Q. T. Le Gia and Ch. Schwab, Higher order Quasi-Monte Carlo integration for Bayesian Estimation. Report 2016-13, Seminar for Applied Mathematics, ETH Zürich, Switzerland (in review).

[10] T.J. Dodwell, C. Ketelsen, R. Scheichl and A.L. Teckentrup, A Hierarchical Multilevel Markov Chain Monte Carlo Algorithm with Applications to Uncertainty Quantification in Subsurface Flow. SIAM/ASA Journal on Uncertainty Quantification, 3, 1075–1108, 2015.

[11] R. N. Gantner, A Generic C++ Library for Multilevel Quasi-Monte Carlo. In *Proceedings of the Platform for Advanced Scientific Computing Conference*, PASC '16, pages 11:1–11:12, New York, NY, USA, 2016. ACM.

[12] R. N. Gantner and Ch. Schwab, Computational Higher Order Quasi-Monte Carlo Integration. In Monte Carlo and Quasi-Monte Carlo Methods 2014, R. Cools and D. Nuyens (eds.), Springer, Cham, pp. 271–288, 2016.

[13] M. B. Giles, Multilevel Monte Carlo methods. Acta Numer., 24, 259–328, 2015.

[14] V. Girault and P.A. Raviart, *Finite Element Methods for Navier-Stokes Equations.* Springer Verlag, Berlin, 1986.

[15] M. Hansen and Ch. Schwab, Analytic regularity and best N-term approximation of high dimensional, parametric initial value problems. Vietnam Journal of Mathematics, 41, 181–215, 2013.

[16] V. H. Hoang, Ch. Schwab and A.M. Stuart, *Complexity analysis of accelerated MCMC methods for Bayesian inversion.* Inverse Problems, 29, (8), 2013.

[17] F. Y. Kuo, Ch. Schwab and I. H. Sloan, Quasi-Monte Carlo finite element methods for a class of elliptic partial differential equations with random coefficient. SIAM J. Numer. Anal., 50, 3351–3374, 2012.

[18] F. Y. Kuo, Ch. Schwab and I. H. Sloan, Multi-Level Quasi-Monte Carlo finite element methods for a class of elliptic partial differential equations with random coefficient, Found. Comp. Math., 15, 411–449, 2015.

[19] J. Pousin and J. Rappaz, Consistency, stability, apriori and aposteriori errors for Petrov-Galerkin methods applied to nonlinear problems. Numer. Math., 69, 213–231, 1994.

[20] Cl. Schillings and Ch. Schwab, Sparse, adaptive Smolyak quadratures for Bayesian inverse problems. Inverse Problems, 29, 065011, 28 pp, 2013.

[21] Cl. Schillings and Ch. Schwab, Sparsity in Bayesian Inversion of Parametric Operator Equations. Inverse Problems, 30, 065007, 30 pp., 2014.

[22] Cl. Schillings and Ch. Schwab, Scaling Limits in Computational Bayesian Inversion. Report 2014-26, Seminar for Applied Mathematics, ETH Zürich, (to appear in M2AN, 2016).

[23] Ch. Schwab, QMC Galerkin discretizations of parametric operator equations. In J. Dick, F. Y. Kuo, G. W. Peters and I. H. Sloan (eds.), Monte Carlo and Quasi-Monte Carlo methods 2012, Springer Verlag, Berlin, 2013, pp. 613–630.

[24] Ch. Schwab and C.J. Gittelson, Sparse tensor discretizations of high-dimensional parametric and stochastic PDEs. Acta Numerica, 20, 291–467, 2011.

[25] Ch. Schwab and R. A. Todor, Karhunen-Loève approximation of random fields by generalized fast multipole methods. J. Comput. Phys., 217, 100–122, 2006.

[26] A. M. Stuart, Inverse problems: a Bayesian perspective. Acta Numerica, 19, 451–559, 2010.

[27] T. Yoshiki, Bounds on the Walsh coefficients by dyadic difference and a new Koksma-Hlawka type inequality for Quasi-Monte Carlo integration. Available at arXiv:1504.03175.