Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Seminar for Applied Mathematics

Multilevel QMC Uncertainty Quantification for Advection-Reaction-Diffusion

L. Herrmann and Ch. Schwab

Research Report No. 2019-06
January 2019

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

Funding SNF: 159940

Multilevel QMC Uncertainty Quantification for Advection-Reaction-Diffusion

Lukas Herrmann and Christoph Schwab

Abstract We survey the numerical analysis of a class of deterministic, higher-order QMC integration methods in forward and inverse uncertainty quantification algorithms for advection-reaction-diffusion (ARD) equations in polygonal domains $D \subset \mathbb{R}^2$ with distributed uncertain inputs. We admit spatially heterogeneous material properties. For the parametrization of the uncertainty, we assume at hand systems of functions which are locally supported in D. Distributed uncertain inputs are written in countably parametric, deterministic form with locally supported representation systems. Parametric regularity and sparsity of solution families and of response functions in scales of weighted Kontrat'ev spaces in D is quantified using analytic continuation.

1 Introduction

Computational uncertainty quantification (UQ) addresses the efficient, quantitative numerical treatment of differential- and integral equation models in engineering and in the sciences. In the simplest setting, such models need to be analyzed for *parametric* input data with sequences $\boldsymbol{y} = (y_j)_{j\geq 1}$ of parameters y_j which range in a compact, metric space U. In [15] the authors 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 parametric operator equations with *distributed uncertain parametric input data*. Typical examples are elliptic or parabolic partial differential equations with uncertain, spatially heterogeneous coefficients, but also differential and integral equations in uncertain physical domains of definition. Upon suitable *uncertainty parametrization* and, in inverse uncertainty quantification, with

Lukas Herrmann · Christoph Schwab
Seminar for Applied Mathematics, ETH Zürich, Rämistrasse 101, 8092 Zurich, Switzerland e-mail: lukas.herrmann@sam.math.ethz.ch, e-mail: christoph.schwab@sam.math.ethz.ch

a suitable Bayesian prior measure placed on the, in general, infinite-dimensional parameter space, the task of numerical evaluation of statistical estimates for quantities of interest (QoI's) becomes *numerical computation of parametric, deterministic integrals over a high-dimensional parameter space*.

The method of choice in many current inverse computational UQ is the Markov chain Monte Carlo (MCMC) method and its variants ([27, 5]). Due to its Monte Carlo character, it affords a generally low convergence rate. As an alternative to the MCMC method, in [40, 41, 7] recently developed, dimension-adaptive Smolyak quadrature techniques were applied to the evaluation of the corresponding integrals. In [15, 14] a convergence theory for HoQMC integration for the numerical evaluation of the corresponding integrals was developed, based on earlier work [16] on these methods in forward UQ. In particular, it was shown in [14] that convergence rates of order $> 1/2$ in terms of the number N of approximate solves of the forward problem that are independent of the dimension can be achieved with judiciously chosen, *deterministic* HoQMC quadratures instead of Monte Carlo or MCMC sampling of the Bayesian posterior. The achievable, dimension-independent rate of HoQMC is, in principle, only limitted by the sparsity of the forward problem. Moreover, the execution of the algorithm is "embarrassingly parallel", since for QMC algorithms, unlike MCMC and sequential Monte Carlo (SMC) methods, the forward problem may be solved simultaneously and in parallel. The error analysis in [14] was extended in [15] to the multilevel setting. As is well known in the context of Monte Carlo methods, *multilevel strategies* can lead to substantial gains in accuracy versus computational cost, see also the survey [24] on MLMC methods. Multilevel discretizations for QMC integration were explored first for parametric, linear forward problems in [31, 29] and, in the context of HoQMC for parametric operator equations, in [15]. For the use of multilevel strategies in the context of MCMC methods for Bayesian inverse problems we refer to [20, 27] and the references there. The purpose of the present paper is to extend the convergence analysis of deterministic Bayesian inversion algorithms for forward problems given by PDEs with distributed random input data, which are based on Quasi-Monte Carlo integration from [15] and the references there, to uncertainty parametrization with basis functions which are locally supported in the physical domain D. Let us mention in passing that while we consider here conforming Finite Element (FEM) discretization, other discretizations in D could equally be considered. We mention only discontinuous Galerkin FEM which have been introduced for advection-diffusion-reaction ("ADR") equations as considered here in [28]. The duality argument in weighted function spaces for these methods has been developed in [32].

The principal contributions of the present work are as follows: we prove, for a class of linear ADR problems in a polygon D with uncertain diffusion coefficients, drift coefficient and reaction coefficient, the well-posedness of the corresponding Bayesian inverse problem. We establish optimal convergence of FE discretizations of the forward problem, with judicious mesh refinement towards the corners $\mathscr{C}$ of D, allowing in particular also corner singularities in the uncertain input data; these appear typically in Karhunen-Loève eigenfunctions corresponding to principal components of covariance operators which are negative fractional powers of the precision operator given as the Dirichlet Laplacean.

We show that a singularity-adapted uncertainty parametrization with locally supported in D spline-wavelet functions allows for optimal (in the sense of convergence rate) parametrization of the uncertain input data.

The structure of this paper is as follows. In Section 2, we present a class of linear, second order diffusion problems in bounded, polygonal domains. Particular attention is paid to regularity in weighted function spaces which account for possible singularities at the corners of the physical domain; we base our presentation on the recent reference [4] where the corresponding regularity theory has been developed.

In Section 5, we review the general theory of well-posed Bayesian inverse problems in function spaces, from [12]. The presentation and the setup is analogous to what was used in [15], but in technical details, there are important differences: unlike the development in [15], the *uncertainty parametrization* employed in the present paper will be achieved by *locally supported functions* ψ_j in the physical domain D. In particular, we shall admit biorthogonal, piecewise polynomial multiresolution analyses in D. These allow, as we show, to resolve uncertain inputs with corner and interface singularities at optimal rates, and their local supports enable the use of Higher order QMC integration with so-called SPROD ("smoothness driven, product") weights. To this end, and as in [15], we require a novel, combined regularity theory of the parametric forward maps in weighted Kondrat'ev-Sobolev spaces in D. In particular, we present an error vs. work analysis of the combined ML HoQMC Petrov–Galerkin algorithms.

2 UQ for advection-reaction-diffusion equations in polygons

We review the notation and mathematical setting of forward and inverse UQ for a class of smooth, parametric operator equations. We develop here the error analysis for the multilevel extension of the algorithms in [22] for general linear, second order advection-reaction-diffusion problems in an open, polygonal domain $D \subset \mathbb{R}^2$, see also [21]. We assume that the uncertain inputs comprise the operators' coefficients $u = ((a_{ij}(x), b_i(x), c(x))$ to belong to a separable Banach space X being a weighted or Hölder space in the physical domain D. As in [15], uncertainty parametrization with an unconditional basis of X will result in a countably-parametric, deterministic boundary value problem. Unlike the Karhunen-Loève basis which is often used for uncertainty parametrization in UQ, we consider here the use of *representation systems whose elements have well-localized supports contained in D*; one example are spline wavelets.

Upon adopting such representations, both forward and (Bayesian) inverse problems become countably parametric, deterministic operator equations. In [40], Bayesian inverse UQ was expressed as countably-parametric, deterministic quadrature problem, with the integrand functions appearing in the Bayesian estimation problems stemming from a (Petrov–)Galerkin discretization of the parametric forward problem in the physical domain. Contrary to widely used MCMC algorithms (e.g. [20] and the references there), high-dimensional, deterministic quadratures of Smolyak type for numerical integration against the (Bayesian) posterior were proposed in [40, 41]. In the present paper, we review this approach for forward and (Bayesian) inverse UQ for ADR in planar, polygonal domains D. We consider in detail high order FEM discretization of the ADR problem on meshes with local corner-refinement in D. We review the use of *deterministic*, HoQMC integration methods, from [16, 17, 19] and the references there, in multilevel algorithms for Bayesian estimation in ADR models with uncertain input.

2.1 Model advection-reaction-diffusion problem in D

We present the parametric ADR model problem in a plane, polygonal domain D and recapitulate its well-posedness and regularity, following [4]. There, in particular, regularity in weighted function spaces in D and holomorphy of the data-to-solution map for this problem in these weighted spaces was established. Optimal FE convergence rates result for FEM in D with locally refined meshes near the singular points of the solution (being either corners of D or boundary points where the nature of the boundary condition changes) [1] and references there.

In the bounded, polygonal domain D with J corners $\mathscr{C} = \{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_J\}$, for some $J \in \mathbb{N}$, we consider the *forward problem* being the mixed boundary value problem for the linear, second order divergence form differential operator

$$\begin{aligned} \mathscr{L}(u)q &:= -\sum_{i,j=1}^{2} \partial_i(a_{ij}\partial_j q) + \sum_{i=1}^{2} b_i \partial_i q - \sum_{i=1}^{2} \partial_j(b_{2+i}q) + cq = f \quad \text{in } D, \\ q\Big|_{\Gamma_1} &= 0, \quad \sum_{i=1}^{2}\sum_{j=1}^{2} \left(a_{ij}\partial_j q + b_{d+j} q\right) n_i\Big|_{\Gamma_2} = 0, \end{aligned} \tag{1}$$

where n denotes the unit normal vector of the domain D and $\emptyset \neq \Gamma_1 \subset \partial D$ denotes the Dirichlet boundary and $\Gamma_2 = \partial D \backslash \Gamma_1$ denotes the Neumann boundary. We shall assume that $\mathscr{C} \subset \overline{\Gamma_1}$.

Define further

$$V := \{v \in H^1(D) : v|_{\Gamma_1} = 0\},$$

where $v|_{\Gamma_1} \in H^{1/2}(\Gamma_1)$ has to be understood in the sense of a trace of $v \in H^1(D)$. In (1), the differential operator $\mathscr{L}$ depends on the uncertain, parametric coefficients

$$u(\boldsymbol{y}) := ((a_{ij}(\boldsymbol{y}^0)_{1\le i,j\le 2}, (b_i(\boldsymbol{y}^1))_{1\le i\le 4}, c(\boldsymbol{y}^2)), \quad \boldsymbol{y}^i \in \left[-\frac{1}{2}, \frac{1}{2}\right]^{\mathbb{N}}, i = 0,1,2,$$

where we have used the notation $\boldsymbol{y} := (\boldsymbol{y}^0, \boldsymbol{y}^1, \boldsymbol{y}^2)$ and further introduce the parameter set

$$U := \prod_{i=0,1,2} \left[-\frac{1}{2}, \frac{1}{2}\right]^{\mathbb{N}} .$$

The uncertain coefficient functions $u(\boldsymbol{y})$ may also depend on the spatial coordinate $x \in D$, and for each $\boldsymbol{y} \in U$ are assumed to belong to weighted Sobolev spaces $\mathscr{W}^{m,\infty}(D)$ of integer order $m \geq 0$ being given by

$$\mathscr{W}^{m,\infty}(D) := \{v : D \to \mathbb{C} : r_D^{|\alpha|} \partial^\alpha v \in L^\infty(D), |\alpha| \leq m\} . \tag{2}$$

Specifically, for $m \in \mathbb{N}_0$, we assume that

$$u \in X_m := \{u : a_{ij} \in \mathscr{W}^{m,\infty}(D),\ r_D b_i \in \mathscr{W}^{m,\infty}(D),\ r_D^2 c \in \mathscr{W}^{m,\infty}(D)\} . \tag{3}$$

Here, $D \ni x \mapsto r_D(x)$ denotes a "regularized" distance to the corners $\mathscr{C}$ of D, i.e. $r_D(x) \simeq \mathrm{dist}(x, \mathscr{C})$ for $x \in D$. We equip X_m, $m \in \mathbb{N}_0$, with the norm

$$\|u\|_{X_m} := \max\{\|a_{ij}\|_{\mathscr{W}^{m,\infty}(D)}, \|r_D b_i\|_{\mathscr{W}^{m,\infty}(D)}, \|r_D^2 c\|_{\mathscr{W}^{m,\infty}(D)}\} . \tag{4}$$

We introduce the parametric bilinear form

$$A(u(\boldsymbol{y}))(w,v) := \langle \mathscr{L}(u(\boldsymbol{y}))w, v\rangle_{V^*,V} , \quad \forall w,v \in V.$$

The variational formulation of the parametric, deterministic problem reads: given $\boldsymbol{y} \in U$, find $q(\boldsymbol{y}) \in V$ such that

$$A(u(\boldsymbol{y}))(q(\boldsymbol{y}),v) = \langle f, v\rangle_{V^*,V} , \quad \forall v \in V .$$

This parametric problem is well-posed if $u(\boldsymbol{y}) \in X_0$ is such that there exists a positive constant $c > 0$

$$\inf_{\boldsymbol{y} \in U} \Re(A(u(\boldsymbol{y}))(v,v)) > c\|v\|_V^2 , \quad \forall v \in V, \tag{5}$$

where $\Re(z)$ denotes the real part of $z \in \mathbb{C}$. We observe that (5) precludes implicitly that the ADR operator in (1) is singularly perturbed. This, in turn, obviates in the ensuing FE approximation theory in Section 3 the need for boundary layer resolution or anisotropic mesh refinements. As a consequence of (5) and of the Lax–Milgram lemma, for every $\boldsymbol{y} \in U$ the parametric solution $q(\boldsymbol{y}) \in V$ exists and satisfies the *uniform a-priori estimate*

$$\sup_{\boldsymbol{y} \in U} \|q(\boldsymbol{y})\|_V \leq c^{-1}\|f\|_{V^*}. \tag{6}$$

We introduce *weighted Sobolev spaces of Kondrat'ev type* $\mathscr{K}_a^m(D)$, $m \in \mathbb{N}_0 \cup \{-1\}$, $a \in \mathbb{R}$, as closures of $C^\infty(\overline{D};\mathbb{C})$ with respect to the *homogeneous weighted norm* given by

$$\|v\|^2_{\mathscr{K}_a^m(D)} := \sum_{|\alpha| \leq m} \|r_D^{|\alpha|-a} \partial^\alpha v\|^2_{L^2(D)}. \tag{7}$$

We observe that (up to equivalence of norms) $V = \{v \in \mathscr{K}^1_1(D) : v|_{\Gamma_1} = 0\}$, which is a consequence of the Hardy inequality (see e.g. [37, Theorem 21.3]). In [4], the authors proved regularity shifts of $\mathscr{L}(u)$. Specifically, if $A(u)$ satisfies (5) and if $u \in X_m$, then by [4, Corollary 4.5 and Theorem 4.4] there exists a constant $C > 0$, $a_0 > 0$ such that for every $a \in (-a_0, a_0)$, $f \in \mathscr{K}^{m-1}_{a-1}(D)$, $q(\boldsymbol{y}) \in \mathscr{K}^{m+1}_{a+1}(D)$ and

$$\sup_{\boldsymbol{y}^i \in \left[-\frac{1}{2},\frac{1}{2}\right]^{\mathbb{N}}, i=0,1,2} \|q(\boldsymbol{y})\|_{\mathscr{K}^{m+1}_{a+1}(D)} \leq C \|f\|_{\mathscr{K}^{m-1}_{a-1}(D)}. \tag{8}$$

Note that the dependence of the constant C on the coefficients u can also be made explicit, see cp. [4].

2.2 Uncertainty parametrization

For uncertainty parametrization, the data space X is assumed to be a separable, infinite-dimensional Banach space with norm $\|\cdot\|_X$ (separably valued data u in a otherwise non-separable space are equally admissible). We suppose that we have at hand representation systems $(\psi^i_j)_{j\geq 1} \subset L^\infty(D;\mathbb{R}^{k_i})$, $i \in \{0,1,2\}$ of locally supported functions in D which parametrize the uncertain coefficient functions $u = (a,b,c)$ for integers $k_0, k_1, k_2 \in \mathbb{N}$.

The smoothness scale $\{X_m\}_{m\geq 0}$ defined in (3) for $m \geq 1$ with $X = L^\infty(D)^8 = X_0 \supset X_1 \supset X_2 \supset \ldots$, and a smoothness order $t \geq 1$ 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 for $u \in X_t$ and which ψ_j being Fourier or multiresolution analyses, higher values of t correspond to stronger decay of the ψ^i_j, $i \in \{0,1,2\}$.

For the numerical analysis of a FE discretization in D, we have to slightly strengthen the norm of X_m. To this end, we define the spaces $W^{m,\infty}_\delta(D)$, $\delta \in [0,1]$, $m \in \mathbb{N}$, as a subspace of $\mathscr{W}^{m,\infty}(D)$ equipped with the norm

$$\|v\|_{W^{m,\infty}_\delta(D)} := \max_{|\alpha|\leq m} \{\|r_D^{\max\{0,\delta+|\alpha|-1\}} \partial^\alpha v\|_{L^\infty(D)}\}.$$

Note that $W^{m,\infty}_1(D) = \mathscr{W}^{m,\infty}(D)$. For $\delta \in [0,1)$, we define $X_{m,\delta}$ by the norm

$$\|u\|_{X_{m,\delta}} := \max\{\|a_{ij}\|_{W^{m,\infty}_\delta(D)}, \|r_D b_i\|_{W^{m,\infty}_\delta(D)}, \|r_D^2 c\|_{W^{m,\infty}_\delta(D)}\} .$$

It is easy to see that the embedding $X_{m,\delta} \subset X_m$ is continuous, $m \geq 1$, $\delta \in [0,1)$. We assume that the $\{\psi^i_j\}_{j\geq 1}$, $i \in \{0,1,2\}$, are scaled such that for some $\delta \in [0,1)$, $\tau \in \mathbb{N}$, and positive sequences $(\rho^i_{r,j})_{j\geq 1}$,

$$\max_{|\boldsymbol{\alpha}|\leq r} \left\| \sum_{j\geq 1} \rho^i_{r,j} r_D^{\max\{0,\delta+|\alpha|-1\}} |\partial^{\boldsymbol{\alpha}}_x ((r_D)^i \psi^i_j)| \right\|_{L^\infty(D)} < \infty, \quad r = 0,\ldots,\tau. \tag{9}$$

Lemma 1. *Let $w \in \mathscr{W}^{m,\infty}(D;\mathbb{C}^k)$ for some $m,k \in \mathbb{N}$ and let $v : D \times \mathbb{C}^k \supset D \times \overline{w(D)} \to \mathbb{C}$ be a function that is $\mathscr{W}^{m,\infty}$-regular in the first argument and analytic in the second. Then, $[x \mapsto v(x,w(x))] \in \mathscr{W}^{m,\infty}(D)$.*

Proof. Let $\widetilde{v} := [\overline{w(D)} \ni z \mapsto v(x,z)]$ for arbitrary $x \in D$ such that $\widetilde{v}$ is well-defined. By an application of the Faà di Bruno formula [11, Theorem 2.1],

$$r_D^{|\boldsymbol{\alpha}|}\partial_x^{\boldsymbol{\alpha}}(\widetilde{v}\circ w) = \sum_{1\le|\boldsymbol{\lambda}|\le n}\partial_y^{\boldsymbol{\lambda}}\widetilde{v}\sum_{\iota=1}^{n}\sum_{p_\iota(\boldsymbol{\alpha},\boldsymbol{\lambda})}\boldsymbol{\alpha}!\prod_{j=1}^{\iota}\frac{1}{\boldsymbol{\nu}(j)!(\boldsymbol{\nu}(j)!)^{|\boldsymbol{\nu}(j)|}}\prod_{i=1}^{k}(r_D^{|\boldsymbol{\nu}(j)|}\partial^{\boldsymbol{\nu}(j)}w_i)^{\nu(j)_i},$$

where $n = |\boldsymbol{\alpha}|$, $w = (w_1,\ldots,w_k)$, and

$$p_\iota(\boldsymbol{\alpha},\boldsymbol{\lambda}) = \Bigg\{(\boldsymbol{\nu}(1),\ldots,\boldsymbol{\nu}(\iota);\boldsymbol{\nu}(1),\ldots,\boldsymbol{\nu}(\iota)) : |\boldsymbol{\nu}(i)|>0,$$
$$\mathbf{0}\prec\boldsymbol{\nu}(1)\prec\ldots\prec\boldsymbol{\nu}(\iota), \sum_{i=1}^{\iota}\boldsymbol{\nu}(i)=\boldsymbol{\lambda}, \text{and } \sum_{i=1}^{\iota}|\boldsymbol{\nu}(i)|\boldsymbol{\nu}(i)=\boldsymbol{\alpha}\Bigg\},$$

where the multiindices $\boldsymbol{\nu}$ are k-dimensional and the multiindices $\boldsymbol{\nu}$ are d-dimensional (here $d=2$, since the domain D is a polygon). The symbol $\prec$ for multiindices $\boldsymbol{\nu}$ and $\widetilde{\boldsymbol{\nu}}$ is defined by $\boldsymbol{\nu}\prec\widetilde{\boldsymbol{\nu}}$ (here for $d=2$) if either (i) $|\boldsymbol{\nu}|<|\widetilde{\boldsymbol{\nu}}|$ or (ii) $|\boldsymbol{\nu}|=|\widetilde{\boldsymbol{\nu}}|$ and $\nu_1<\widetilde{\nu}_1$. Since $L^\infty(D)$ is an algebra and $|\boldsymbol{\alpha}|=\sum_{i=1}^{\iota}|\boldsymbol{\nu}(i)||\boldsymbol{\nu}(i)|$, $(\widetilde{v}\circ w)\in\mathscr{W}^{m,\infty}(D)$. The claim of the lemma now follows by another application of the Faà di Bruno formula. □

Remark 1. The statement of Lemma 1 also holds if we replace $\mathscr{W}^{m,\infty}(D)$ with $W_\delta^{m,\infty}(D)$, $\delta\in[0,1)$, at all places.

Define the complex-parametric sets of admissible data

$$U^i := \left\{\sum_{j\ge1} z_j|\psi_j^i(x)| : \boldsymbol{z}\in\mathbb{C}^{\mathbb{N}}, |z_j|\le\rho_{0,j}^i, j\ge1, x\in D\right\}\subset\mathbb{C}^{k_i}, i\in\{0,1,2\},$$

where $|\cdot|$ denotes component-wise absolute value. Let $g : D\times U^0\times U^1\times U^2\to\mathbb{C}^8$ be a function such that $(z^1,z^2,z^2)\mapsto g(x,z^0,z^1,z^2)$ is holomorphic for almost every $x\in D$ and such that $[x\mapsto g(x,z^0,z^1,z^2)]\in X_m$ for some $m\ge1$ and every $(z^0,z^1,z^2)\in U^0\times U^1\times U^2$. The uncertain coefficient $u=(a,b,c)$ is then parametrized by

$$\begin{aligned} u(x,\boldsymbol{y}) &= \big(a(x,\boldsymbol{y}^0), b(x,\boldsymbol{y}^1), c(x,\boldsymbol{y}^2)\big) \\ &= g\left(x, \sum_{j\ge1}y_j^0\psi_j^0, \sum_{j\ge1}y_j^1\psi_j^1, \sum_{j\ge1}y_j^2\psi_j^2\right), \text{ a.e. } x\in D, \boldsymbol{y}^i\in U, i=0,1,2. \end{aligned} \tag{10}$$

Hence, $u=(a,b,c)$ is given through the coordinates of the function g via $a_{11}=g_1$, $a_{22}=g_2$, $a_{21}=a_{12}=g_3$, $b_i=g_{i+3}$, $i=1,\ldots,4$, $c=g_8$.

Elements in the space $X_{m,\delta}$ may have singularities in the corners, but can be approximated in the X_0-norm at optimal rates for example by biorthogonal wavelets with suitable refinements near vertices of D.

Proposition 1. *Let $\delta \in [0,1)$ and $m \in \mathbb{N}$ be given. Assume further at hand a biorthogonal, compactly supported spline wavelet basis with sufficiently large number (depending on m) vanishing moments and compactly supported dual basis. Then, there exists a constant $C > 0$ and, for every $L \in \mathbb{N}$ projection operators P_L into a biorthogonal wavelet basis such that*

$$\|w - P_L w\|_{X_0} \leq C N_L^{-m/2} \|w\|_{X_{m,\delta}}, \quad \forall w \in X_{m,\delta},$$

where N_L denotes the number of terms in the expansion $P_L w$.

The proof of this (in principle, well-known) proposition is given in Section 6.2, where also further details on biorthogonal wavelets are presented.

3 Finite Element discretization

We introduce conforming Finite Element discretizations in the physical domain D and review an approximation property as a preparation for the analysis of the impact of numerical integration on locally refined meshes in D. Let $\mathfrak{T}$ denote a family of regular, simplicial triangulations of the polygon D. We assume that $\mathfrak{T}$ is obtained from a coarse, initial triangulation by *newest vertex bisection*, cp. [23]. In this section we will omit the parameter vector $\boldsymbol{y}$ in our notation with the understanding that all estimates depend on the parameter vector $\boldsymbol{y}$ only via dependencies on the coefficients $u = (a,b,c)$. We assume there exists a constant $C > 0$ independent of h such that for every $\mathscr{T} \in \mathfrak{T}$ and every $K \in \mathscr{T}$:

$$\begin{aligned} &\text{(i) If } \overline{K} \cap \mathscr{C} = \emptyset, \text{ then } C^{-1} h r_D^{\beta}(x) \leq h_K \leq C h r_D^{\beta}(x) \text{ for every } x \in K. \\ &\text{(ii) If } \overline{K} \cap \mathscr{C} \neq \emptyset, \text{ then } C^{-1} h \sup_{x \in K}\{r_D^{\beta}(x)\} \leq h_K \leq C h \sup_{x \in K}\{r_D^{\beta}(x)\}, \end{aligned} \tag{11}$$

where

$$h_K := \operatorname{diam}(K), K \in \mathscr{T}, \quad \text{and} \quad h := \max_{K \in \mathscr{T}}\{h_K\}.$$

Such a mesh can be achieved with the algorithm proposed in [23, Section 4.1] with input values the global meshwidth h, the polynomial degree k, and the weight exponent $\gamma = (1+k)(1-\beta)$, assuming $(1+k)(1-\beta) < 1$. There are also *graded meshes* that satisfy (11), which were introduced in [2]. We define the Finite Element spaces of order $k \in \mathbb{N}$ by

$$V_{\mathscr{T}}^k := \{v \in V : v|_K \in \mathbb{P}_k(K), K \in \mathscr{T}\}, \quad \mathscr{T} \in \mathfrak{T},$$

where $\mathbb{P}_k(K)$ denotes the polynomials of degree smaller or equal to k on element $K \in \mathscr{T}$.

Proposition 2. *Let $k \in \mathbb{N}$ and $0 < \delta < \beta < 1$ be such that $(1-\delta)/(1-\beta) > k$ and set $a = 1-\delta$. There exists a constant $C > 0$ independent of the global mesh width h such that for every $w \in \mathscr{K}^{k+1}_{a+1}(D)$ there exist $w^{\mathscr{T}} \in \widetilde{V}^k_{\mathscr{T}} \subset V^k_{\mathscr{T}}$ such that*

$$\|w - w^{\mathscr{T}}\|_{H^1(D)} \leq Ch^k \|w\|_{\mathscr{K}^{k+1}_{a+1}(D)}.$$

This result is, in principle, known; e.g. [1, 34] and references there.

3.1 Numerical integration

An essential component in the numerical analysis of the considered class of problems is the efficient numerical evaluation of the mass and stiffness matrices which contain the inhomogeneous, parametric coefficients. Owing to their origin as pathwise realizations of random fields, these coefficients have, in general, only finite Sobolev regularity. Furthermore, for covariances in bounded domains which result from precision operators which include boundary conditions such as the Dirichlet Laplacean, these realizations can exhibit singular behaviour near corners of D. This is accomodated by the weighted Sobolev spaces $\mathscr{W}^{m,\infty}(D)$ comprising the data spaces X_m as defined above in (3). Efficient numerical quadrature for the evaluation of the stiffness and mass matrices which preserves the FE approximation properties on locally refined meshes is therefore needed. The numerical analysis of the impact of quadrature on FEM on locally refined meshes for uncertain coefficients in X_m is therefore required.

The impact of numerical integration in approximate computation of the stiffness matrix and load vector on the convergence rates of the FE solution is well understood for uniform mesh refinement, cp. for example [9, Section 4.1]. We extend this theory to regular, simplicial meshes with local refinement toward the singular points, and to possibly singular coefficients which belong to weighted spaces, i.e., $u = (a,b,c) \in X_{m,\delta}$, $\delta \in [0,1)$, as defined in (3). We provide a strategy to numerically approximate the stiffness matrix by quadrature so that the resulting additional consistency error is consistent with the FE approximation error, *uniformly with respect to the parameter sequences which characterize the uncertain inputs*. We denote by $\widetilde{A}$ on $V_{\mathscr{T}} \times V_{\mathscr{T}}$ the bilinear form, which has been obtained with numerical integration, i.e., for quadrature weights and nodes $(\omega_{K,k}, x_{K,k})_{K\in\mathscr{T},k\in\mathscr{I}} \subset (0,\infty) \times \overline{D}$

$$\widetilde{A}(w,v) := \sum_{K\in\mathscr{T}_\ell} \sum_{k\in\mathscr{I}} \omega_{K,k} \left(\sum_{i,j=1}^{2} a_{ij}\partial_j w \partial_i v + \sum_{i=1}^{2} b_i \partial_i w v + \sum_{i=1}^{2} b_{2+i} w \partial_j v + cwv \right)(x_{K,k}),$$

for every $w, v \in V_{\mathscr{T}}$. Let us denote by $F_K : \widehat{K} \to K$ the affine element mappings that are given by $\xi \mapsto F_K(\xi) = B_K \xi + b_K$, $K \in \mathscr{T}$. Let $(\widehat{\omega}_k, \widehat{x}_k)_{k \in \mathscr{I}}$ be a set of positive weights and nodes for quadrature on the reference element $\widehat{K}$. Then, $\omega_{K,k} := \det(B_K)\omega_k > 0$ and $x_{K,k} := F_K(x_k) \in K$, $K \in \mathscr{T}$, $k \in \mathscr{I}$. We define the element quadrature error for every $K \in \mathscr{T}$ and integrable ϕ such that point evaluation is well-defined by

$$E_K(\phi) = \int_K \phi \, \mathrm{d}x - \sum_{k \in \mathscr{I}} \omega_{K,k} \phi(x_{K,k}).$$

The quadrature error $E_{\widehat{K}}$ on the reference element $\widehat{K}$ is defined analogously.

Under these assumptions, it can be shown as in the proof of [9, Theorem 4.1.2]] (which covers the case that $b_i \equiv 0$ and $c \equiv 0$ in u in (3)), that the corresponding approximate sesquilinear form $\widetilde{A}(u)(\cdot,\cdot) : V^k_{\mathscr{T}} \times V^k_{\mathscr{T}} \to \mathbb{C}$ satisfies coercivity (5) with a positive coercivity constant $\tilde{c}$, possibly smaller than $c > 0$ in (5) but still independent of $\boldsymbol{y}^i \in [-1/2, 1/2]^{\mathbb{N}}$, $i = 0, 1, 2$.

Let us denote the FE solution with respect to the bilinear form $\widetilde{A}(u) : V^k_{\mathscr{T}} \times V^k_{\mathscr{T}} \to \mathbb{C}$ by $\widetilde{q}^{\mathscr{T}} \in V^k_{\mathscr{T}}$, i.e.,

$$\widetilde{A}(u)(\widetilde{q}, v) = \langle f, v \rangle_{V^*,V}, \quad \forall v \in V^k_{\mathscr{T}}.$$

The error incurred by employing numerical quadrature is consistent with the FE approximation rate, as demonstrated in the following theorem.

Theorem 1. *For $k \geq 1$, suppose that $E_{\widehat{K}}(\widehat{\phi}) = 0$ for every $\widehat{\phi} \in \mathbb{P}_{2k-1}$. Let $0 < \delta < \beta < 1$ satisfy $(1-\delta)/(1-\beta) > k$. There exists a constant $C > 0$ independent of h and of $u = (a, b, c) \in X_{k,\delta}$ such that*

$$\|q - \widetilde{q}^{\mathscr{T}}\|_V \leq C h^k \left(1 + \|u\|_{X_{k,\delta}}\right) \|q\|_{\mathscr{K}^{k+1}_{a+1}(D)}.$$

The impact of numerical integration in linear functionals of the solution shall has been studied in the case of uniform refinement and unweighted Sobolev space in [3]. We shall extend their result to solutions to ADR problems in polygons in the following corollary.

Corollary 1. *Let $0 \leq k' \leq k$ be integers. Suppose that $E_{\widehat{K}}(\widehat{\phi}) = 0$ for every $\widehat{\phi} \in \mathbb{P}_{2k}$. Let $0 < \delta < \beta < 1$ satisfy $(1-\delta)/(1-\beta) > k + k'$. There exists a constant $C > 0$ that does not depend on h such that for every $G \in \mathscr{K}^{k'-1}_{a-1}(D)$,*

$$|G(q) - G(\widetilde{q}^{\mathscr{T}})| \leq C h^{k+k'} \left(1 + \|u\|_{X_{k,\delta}}\right) \left(1 + \|u\|_{X_{k+k',\delta}}\right) \|q\|_{\mathscr{K}^{k+1}_{a+1}(D)} \|G\|_{\mathscr{K}^{k'-1}_{a-1}(D)}.$$

The proofs of Theorem 1 and Corollary 1 are given in Section 6.1.

3.2 Finite Element approximation of parametric solution

To this end we suppose that we have a sequence of FE triangulations $\{\mathscr{T}_\ell\}_{\ell\geq 0}$ such that $\mathscr{T}_\ell$ satisfies the assumption in (11) with constants that are uniform in $\ell \geq 0$. The global mesh widths are denoted by $(h_\ell)_{\ell\geq 0}$. We denote by V_ℓ^k, $\ell \geq 0$, the respective FE spaces for some $k \geq 1$ and define

$$M_\ell := \dim(V_\ell^k), \quad \ell \geq 0.$$

Recall the bilinear form $\widetilde{A}(u(\boldsymbol{y}))$ on $V_\ell^k \times V_\ell^k$ that results from the application of numerical integration in the previous section. The Galerkin approximation $\widetilde{q}^{\mathscr{T}_\ell}(\boldsymbol{y}) \in V_\ell^k$ is the unique solution to

$$\widetilde{A}(u(\boldsymbol{y}))(\widetilde{q}^{\mathscr{T}_\ell}(\boldsymbol{y}), v) = \langle f, v\rangle_{V^*,V}, \quad \forall v \in V_\ell^k. \tag{12}$$

We recall the sparsity assumption in (9) for postive sequences $(\rho_{r,j}^i)_{j\geq 1}$, $r = 0,\dots,\tau$. This assumption and Lemma 1 imply that $a_{i,j}(\boldsymbol{y}^0), b_j(\boldsymbol{y}^1), c(\boldsymbol{y}^2) \in W_\delta^{\tau,\infty}(D)$ for every $\boldsymbol{y}^0, \boldsymbol{y}^1, \boldsymbol{y}^2 \in [-1/2, 1/2]^{\mathbb{N}}$ and admissible i, j. We assume that $f \in \mathscr{K}_{a-1}^{t-1}(D)$, $G \in \mathscr{K}_{a-1}^{t'-1}(D)$ for integers $t, t' \geq 0$ satisfying $t + t' \leq \tau$. Then, by Corollary 1 and (8),

$$\sup_{\boldsymbol{y}\in U} |G(q(\boldsymbol{y})) - G(\widetilde{q}^{\mathscr{T}_\ell}(\boldsymbol{y}))| \leq C M_\ell^{-(\min\{t,k\}+\min\{t',k\})/2} \|f\|_{\mathscr{K}_{a-1}^{t-1}(D)} \|G\|_{\mathscr{K}_{a-1}^{t'-1}(D)}, \tag{13}$$

where we applied that $M_\ell = \mathscr{O}(h_\ell^{-d})$, $\ell \geq 0$.

The parametric solution may be approximated consistently up to any order of h_ℓ by preconditioned conjugate gradient in work $\mathscr{O}(M_\ell \log(M_\ell))$. Admissible preconditioners in the symmetric case, i.e., $b_i(\boldsymbol{y}^{(i)}) \equiv 0$, for $\Gamma_2 = \emptyset$ are so called BPX and symmetric V-cycle. Respective condition numbers for local mesh refinement by newest vertex bisection for BPX and symmetric V-cycle has been studied for the Dirichlet Laplacean in [6]. These results are applicable, since the Dirichlet Laplacean is spectrally equivalent to $\mathscr{L}(u)$. For notational convenience, the PCG approximation of $\widetilde{q}^{\mathscr{T}_\ell}(\boldsymbol{y})$ will be denoted by the same symbol. Since PCG is an iterative method, holomorphic dependence on the parameters $\boldsymbol{y}$ is preserved.

We also consider parameter dimension truncation to obtain a finite dimensional parameter set and denote by $s^0, s^1, s^2 \in \mathbb{N}$ the corresponding parameter numbers. We denote the triple of those by $\boldsymbol{s} := (s^0, s^1, s^2)$. Let us introduce

$$\widetilde{q}^{\boldsymbol{s},\mathscr{T}_\ell}(\boldsymbol{y}) := \widetilde{q}^{\mathscr{T}_\ell}(\boldsymbol{y}^0_{\{1:s^0\}}, \boldsymbol{y}^1_{\{1:s^1\}}, \boldsymbol{y}^2_{\{1:s^2\}}), \quad \boldsymbol{y}^i \in \left[-\frac{1}{2}, \frac{1}{2}\right]^{\mathbb{N}}, i = 0,1,2,$$

where we have used the notation $(\boldsymbol{y}^i_{\{1:s^1\}})_j = y^i_j$ for $j \in \{1 : s^1\} := \{1,\dots,s^1\}$ and zero otherwise, $i = 0,1,2$. We define $u^{\boldsymbol{s}}$ and $q^{\boldsymbol{s}}$ analogously.

Lemma 2. *Let $u_1 = (a^1, b^1, c^1), u_2 = (a^2, b^2, c^2) \in X_0$ and $q_1, q_2 \in V$ satisfy $\mathscr{L}(u_i)q_i = f$, $i = 1,2$. Assume that the bilinear forms $A(u_1)(\cdot,\cdot), A(u_2)(\cdot,\cdot)$ are coercive with* *coercivity constants* $c_1, c_2 > 0$ *in the sense of* (5). *Then, there exists a constant* $C > 0$ *independent of* q_1, q_2, u_1, u_2 *such that*

$$\|q_1 - q_2\|_V \leq \frac{C}{c_1 c_2} \|u_1 - u_2\|_{X_0} \|f\|_{V^*}$$

Proof. We observe that $\|q_1 - q_2\|_V^2 c_1^{-1} \leq A(u_1)(q_1 - q_2, q_1 - q_2) = A(u_2 - u_1)(q_2, q_1 - q_2)$. By the Hardy inequality (see e.g. [37, Theorem 21.3]) there exists a constant $C > 0$ such that for every $v \in V$

$$\|r_D^{-1} v\|_{L^2(D)} \leq C \| |\nabla v| \|_{L^2(D)}. \tag{14}$$

As a consequence,

$$\begin{aligned} &|A(u_2 - u_1)(q_2, q_1 - q_2)| \\ &\quad \leq C \left(\sum_{i,j=1}^{2} \|a_{ij}^1 - a_{ij}^2\|_{L^\infty(D)} + \sum_{j=1}^{4} \|r_D(b_j^1 - b_j^2)\|_{L^\infty(D)} + \|r_D^2(c^1 - c^2)\|_{L^\infty(D)} \right) \\ &\qquad \times \|q_2\|_V \|q_1 - q_2\|_V , \end{aligned}$$

where $C > 0$ is the constant from the Hardy inequality. In the previous step, we used multiplication by one, i.e., by $r_D r_D^{-1}$ for the advection terms and by $r_D^2 r_D^{-2}$ for the reaction term. The claim now follows with (6). □

It is easy to see that since g as introduced in (10) is in particular locally Lipschitz continuous, by Lemma 2 there exists a constant $C >$ such that

$$\sup_{\boldsymbol{y} \in U} \|u(\boldsymbol{y}) - u^{\boldsymbol{s}}(\boldsymbol{y})\|_{L^\infty(D)} \leq C \left(\sup_{j > s^0} \{(\rho_{0,j}^0)^{-1}\} + \sup_{j > s^1} \{(\rho_{0,j}^1)^{-1}\} + \sup_{j > s^2} \{(\rho_{0,j}^2)^{-1}\} \right).$$

Thus, by (13)

$$\sup_{\boldsymbol{y} \in U} |G(q(\boldsymbol{y})) - G(\widetilde{q}^{\boldsymbol{s}, \mathscr{T}_\ell}(\boldsymbol{y}))| \leq C \left(M_\ell^{-(\min\{t,k\} + \min\{t',k\})/2} + \max_{i=0,1,2} \sup_{j > s^i} \{(\rho_{0,j}^i)^{-1}\} \right). \tag{15}$$

4 Forward UQ

In this section we discuss the consistent approximation of the expectation of $G(q)$, where $G \in V^*$ is a linear functional. The expectation is taken with respect to the uniform product measure on U, which is denoted by $\mathrm{d}\boldsymbol{y} := \bigotimes_{i=0,1,2} \bigotimes_{j \geq 1} \mathrm{d}y_j$. The expectation of $G(q)$ will be denoted by

$$\mathbb{E}(G(q)) := \int_U G(q(\boldsymbol{y})) \mathrm{d}\boldsymbol{y}.$$

4.1 Higher order QMC integration

Quadrature by QMC approximates integrals over the $s \in \mathbb{N}$ dimensional unit cube with equal quadrature weights, i.e., for a suitable integrand function F (possibly Banach space valued) and judiciously chosen, deterministic QMC points $\{\boldsymbol{y}^{(0)}, \ldots, \boldsymbol{y}^{(N-1)}\} \subset [0,1]^s$

$$I_s(F) := \int_{[-\frac{1}{2},\frac{1}{2}]^s} F(\boldsymbol{y}) \mathrm{d}\boldsymbol{y} \approx \frac{1}{N} \sum_{i=0}^{N-1} F\left(\boldsymbol{y}^{(i)} - \boldsymbol{\frac{1}{2}}\right) =: Q_{s,N}(F),$$

where $(\boldsymbol{\frac{1}{2}})_j = 1/2$, $j = 1, \ldots, s$.

Integration by QMC methods is able to achieve convergence rates that are independent of the dimension of integration and are higher than Monte Carlo sampling; we refer to the surveys [30, 18]. In particular, *interlaced polynomial lattice rules* are able to achieve even convergence rates, which can be of arbitrary order if the integrand satisfies certain conditions, cp. [16]. The analysis in this work will be for QMC by interlaced polynomial lattice rules. As in our previous works [22, 21], we will justify the application of interlaced polynomial lattice rules with *product weights*, which implies that the construction cost of the respective QMC points by the fast CBC construction is $\mathscr{O}(sN\log(N))$, where s is the dimension of integration and N the number of QMC points, cp. [16, 35, 36]. We state the main approximation result for interlaced polynomial lattice rules from [16] for product weights given in [16, Equation (3.18)].

Theorem 2 ([16, Theorem 3.2]). *Let $s \in \mathbb{N}$ and $N = b^m$ for $m \in \mathbb{N}$ and b a prime number. Let $\boldsymbol{\beta} = (\beta_j)_{j\geq 1}$ be a sequence of positive numbers and assume that $\boldsymbol{\beta} \in \ell^p(\mathbb{N})$ for some $p \in (0,1]$. Define the integer $\alpha = \lfloor 1/p \rfloor + 1 \geq 2$. Suppose the partial derivatives of the integrand $F : [-1/2,1/2]^s \to \mathbb{R}$ satisfy*

$$\forall \boldsymbol{y} \in [-1/2,1/2]^s,\ \forall \boldsymbol{\nu} \in \{0,\ldots,\alpha\}^s : \quad |\partial_{\boldsymbol{y}}^{\boldsymbol{\nu}} F(\boldsymbol{y})| \leq c\boldsymbol{\nu}! \prod_{j=1}^{s} \beta_j^{\tau_j},$$

for some constant $c > 0$ which is independent of s and of $\boldsymbol{\nu}$.

Then, there exists an interlaced polynomial lattice rule which can be constructed with the CBC algorithm and product weights $(\gamma_{\mathfrak{u}})_{\mathfrak{u}\subset\mathbb{N}}$ given by $\gamma_{\emptyset} = 1$ and

$$\gamma_{\mathfrak{u}} = \prod_{j\in\mathfrak{u}} \left(C_{\alpha,b} b^{\alpha(\alpha-1)/2} \sum_{\nu=1}^{\alpha} 2^{\delta(\nu,\alpha)} \beta_j^{\nu} \right), \quad \mathfrak{u} \subset \mathbb{N}, |\mathfrak{u}| < \infty, \tag{16}$$

($\delta(\nu,\alpha) = 1$ if $\nu = \alpha$ and zero otherwise) in $\mathscr{O}(\alpha s N \log(N))$ operations such that

$$\forall N \in \mathbb{N} : \quad |I_s(F) - Q_{s,N}(F)| \leq C_{\alpha,\boldsymbol{\beta},b,p} N^{-1/p},$$

where $C_{\alpha,\boldsymbol{\beta},b,p} < \infty$ is independent of s and N.

The value for the Walsh constant $C_{\alpha,b}$ is given in [16, Equation (3.11)]. An improved bound for $C_{\alpha,b}$ is derived in [46].

4.2 Parametric regularity

For the applicability of higher order integration methods such as QMC for UQ, the assumption on the partial derivatives with respect to the parameter $\boldsymbol{y}$ in Theorem 2 of the solution $q(\boldsymbol{y})$ or of functionals composed with $q(\boldsymbol{y})$ have to be verified. In [4] the authors proved analytic dependence of the solution on the coefficient in the complex valued setting. Hence, holomorphy is a direct consequence. By (5), (9), and Lemma 1, for every truncation $\boldsymbol{s} = (s^0, s^1, s^2)$ the coefficients

$$u : \mathscr{D}^{\boldsymbol{s}}_{\boldsymbol{\rho}_r} \to X_r$$

is holomorphic for $r = 0, \dots, t$, where

$$\mathscr{D}^{\boldsymbol{s}}_{\boldsymbol{\rho}_r} := \{\boldsymbol{z} = (\boldsymbol{z}^0, \boldsymbol{z}^1, \boldsymbol{z}^2) : \boldsymbol{z}^i \in \mathbb{C}^{s_i}, |z^i_j| \le \rho^i_{r,j}/2, i = 0,1,2\}.$$

As a composition of holomorphic mappings by [4, Corollary 5.1], the map

$$u : \mathscr{D}^{\boldsymbol{s}}_{\boldsymbol{\rho}_r} \to \mathscr{K}^{r+1}_{a+1}(D), \quad r = 0, \dots, t,$$

is holomorphic and

$$\sup_{\boldsymbol{s}\in\mathbb{N}^3} \sup_{\boldsymbol{z}\in\mathscr{D}^{\boldsymbol{s}}_{\boldsymbol{\rho}_r}} \|q(\boldsymbol{z})\|_{\mathscr{K}^{r+1}_{a+1}(D)} < \infty, \quad r = 0, \dots, t.$$

The following lemma is a version of [19, Lemma 3.1].

Lemma 3 ([19, Lemma 3.1]). *For a Banach space B and $\boldsymbol{\rho} = (\rho_j)_{j\ge1} \in (1,\infty)^{\mathbb{N}}$, $s \in \mathbb{N}$, let $F : \mathscr{D}^s_{\boldsymbol{\rho}} \to B$ be holomorphic, where $\mathscr{D}^s_{\boldsymbol{\rho}} := \{\boldsymbol{z} \in \mathbb{C}^s : |z_j| \le \rho_j, j = 1,\dots,s\}$. Then, for every $\boldsymbol{y} \in [-1,1]^{\mathbb{N}}$,*

$$\forall \boldsymbol{\nu} \in \mathbb{N}_0^{\mathbb{N}}, |\boldsymbol{\nu}| < \infty : \quad \|\partial^{\boldsymbol{\nu}}_{\boldsymbol{y}} F(\boldsymbol{y})\|_B \le \sup_{\boldsymbol{z}\in\mathscr{D}^s_{\boldsymbol{\rho}}} \{\|F(\boldsymbol{z})\|_B\} \prod_{j\ge1} \frac{\rho_j}{(\rho_j - 1)^{\nu_j+1}}.$$

The argument used in the proof of this lemma is based on the Cauchy integral formula for holomorphic functions (see also [8, 10]).

Theorem 3. *There exists a constant $C > 0$ such that for every $\boldsymbol{\nu} = (\boldsymbol{\nu}^0, \boldsymbol{\nu}^1, \boldsymbol{\nu}^2)$, $\boldsymbol{\nu}^i \in \mathbb{N}_0^{\mathbb{N}}$, $|\boldsymbol{\nu}^i| < \infty$, $\boldsymbol{s} = (s^0, s^1, s^2)$, $s^i \in \mathbb{N}$, $i = 0,1,2$, $\ell \ge 0$, $0 \le t' \le t \le k$, $\theta \in [0,1]$, and every $\boldsymbol{y} \in U$,*

$$\begin{aligned}
&|\partial^{\boldsymbol{\nu}}_{\boldsymbol{y}} (G(q(\boldsymbol{y})) - G(\widetilde{q}^{\boldsymbol{s},\mathscr{T}_\ell}(\boldsymbol{y})))| \\
&\quad \le C\|G\|_{V^*}\|f\|_{V^*} \max_{i=0,1,2} \sup_{j>s^i} \{(\rho^i_{0,j})^{-\theta}\} \prod_{i=0,1,2} \prod_{j\ge1} (\rho^i_{0,j}/2)^{-\nu^i_j(1-\theta)} \\
&\qquad + C\|G\|_{\mathscr{K}^{t'-1}_{a-1}} \|f\|_{\mathscr{K}^{t-1}_{a-1}} M_\ell^{-(t+t')/d} \prod_{i=0,1,2} \prod_{j\ge1} (\rho^i_{t+t',j}/2)^{-\nu^i_j}.
\end{aligned}$$

Proof. The estimate will follow by a twofold application of Lemma 3 and the holomorphic dependence of the solution on the parametric input. By the triangle inequality,

$$|G(q(\boldsymbol{y})) - G(\widetilde{q}^{\boldsymbol{s},\mathscr{T}_\ell}(\boldsymbol{y}))| \le |G(q(\boldsymbol{y})) - G(q^{\boldsymbol{s}}(\boldsymbol{y}))| + |G(q^{\boldsymbol{s}}(\boldsymbol{y})) - G(\widetilde{q}^{\boldsymbol{s},\mathscr{T}_\ell}(\boldsymbol{y}))|.$$

By the assumption in (9) and [4, Corollary 5.1], the mapping $\boldsymbol{z} \mapsto G(q(\boldsymbol{z})) - G(q^{\boldsymbol{s}}(\boldsymbol{z}))$ is holomorphic on $\mathscr{D}^{\boldsymbol{s}}_{(\boldsymbol{\rho}_0)^{1-\theta}}$ and by Lemma 2 it holds that

$$\sup_{\boldsymbol{z}\in\mathscr{D}^{\boldsymbol{s}}_{(\boldsymbol{\rho}_0)^{1-\theta}}} |G(q(\boldsymbol{z})) - G(q^{\boldsymbol{s}}(\boldsymbol{z}))| \le C\|G\|_{V^*}\|f\|_{V^*} \max_{i=0,1,2} \sup_{j>s^i}\{(\rho^i_{0,j})^{-\theta}\}.$$

Hence, by Lemma 3, where we scale the parameter vectors by a factor of $1/2$

$$\begin{aligned} &|\partial^{\boldsymbol{\nu}}_{\boldsymbol{y}}(G(q(\boldsymbol{y})) - G(q^{\boldsymbol{s}}(\boldsymbol{y})))| \\ &\qquad \le C\|G\|_{V^*}\|f\|_{V^*} \max_{i=0,1,2} \sup_{j>s^i}\{(\rho^i_{0,j})^{-\theta}\} \prod_{i=0,1,2}\prod_{j\ge 1}(\rho^i_{0,j}/2)^{-\tau^i_j(1-\theta)}. \end{aligned}$$

Furthermore by the assumption in (9) and [4, Corollary 5.1], the mapping $\boldsymbol{z} \mapsto q^{\boldsymbol{s}}(\boldsymbol{z})$ is holomorphic from $\mathscr{D}^{\boldsymbol{s}}_{\boldsymbol{\rho}_t}$ to $\mathscr{K}^{t+1}_{a+1}(D)$ and

$$\sup_{\boldsymbol{z}\in\mathscr{D}^{\boldsymbol{s}}_{\boldsymbol{\rho}_t}} \|q(\boldsymbol{z})\|_{\mathscr{K}^{t+1}_{a+1}(D)} < \infty.$$

Thus, by (13), there exists $C > 0$ such that for all $\boldsymbol{s}$ and all ℓ holds

$$\sup_{\boldsymbol{z}\in\mathscr{D}^{\boldsymbol{s}}_{\boldsymbol{\rho}_t}} |G(q^{\boldsymbol{s}}(\boldsymbol{z})) - G(\widetilde{q}^{\boldsymbol{s},\mathscr{T}_\ell}(\boldsymbol{z}))| \le C\|G\|_{\mathscr{K}^{t'-1}_{a-1}}\|f\|_{\mathscr{K}^{t-1}_{a-1}} M_\ell^{-(t+t')/d}.$$

The second part of the estimate now also follows by Lemma 3. ◻

4.3 Multilevel QMC error estimates

Multilevel integration schemes offer a reduction in the overall computational cost, subject to suitable regularity (see, e.g., [17, 29, 26]). For $\boldsymbol{s}_{\ell=0,\dots,L}$, $N_{\ell=0,\dots,L}$, $L \in \mathbb{N}_0$, define the multilevel QMC quadrature

$$Q_L(G(\widetilde{q}^L)) := \sum_{\ell=0}^{L} Q_{|\boldsymbol{s}|_\ell,N_\ell}(G(\widetilde{q}^\ell) - G(\widetilde{q}^{\ell-1})),$$

where we used the notation $\widetilde{q}^\ell := \widetilde{q}^{\boldsymbol{s}_\ell,\mathscr{T}_\ell}$, $\ell = 0,\dots,L$, and $\widetilde{q}^{-1} := 0$. The QMC weights are obtained from (16) with input

$$\beta_{j(k,i)} := 2\max\{(\rho^i_{0,k})^{-(1-\theta)}, (\rho^i_{\tau,k})^{-1}\}, \tag{17}$$

where $\tau = t + t'$ and $j(k,i) := 3k - i$, $k \in \mathbb{N}$, $i = 0,1,2$, is an enumeration of $\mathbb{N}$ with elements in $\mathbb{N} \times \{0,1,2\}$.

Theorem 4. *Suppose that the weight sequence in* (17) *satisfies* $\boldsymbol{\beta} = (\beta_j)_{j\geq 1} \in \ell^p(\mathbb{N})$ *for some* $p \in (0,1]$. *Then, with an interlaced polynomial lattice rules of order* $\alpha = \lfloor 1/p \rfloor + 1$ *and product weights* (16) *with weight sequence* (17) *there exists a constant* $C > 0$ *such that for* $\boldsymbol{s}_{\ell=0,\dots,L}$, $N_{\ell=0,\dots,L}$, $L \in \mathbb{N}_0$,

$$|\mathbb{E}(G(q)) - Q_L(G(\widetilde{q}^L))| \leq C \left(M_L^{-(t'+t)/d} + \max_{i=0,1,2} \sup_{j>s^i} \{(\rho^i_{0,j})^{-1}\} \right.$$
$$\left. + \sum_{\ell=0}^{L} N_\ell^{-1/p} \left(M_{\ell-1}^{-(t'+t)/d} + \max_{i=0,1,2} \sup_{j>s^i_{\ell-1}} \{(\rho^i_{0,j})^{-\theta}\} \right) \right).$$

Proof. By the triangle inequality, we obtain the deterministic error estimate

$$|\mathbb{E}(G(q)) - Q_L(G(\widetilde{q}^L))|$$
$$\leq |\mathbb{E}(G(q)) - Q_L(G(\widetilde{q}^L))| + \sum_{\ell=0}^{L} |(I_{|\boldsymbol{s}_\ell|} - Q_{|\boldsymbol{s}_\ell|,N_\ell})(G(\widetilde{q}^\ell) - G(\widetilde{q}^{\ell-1}))|,$$

where $|\boldsymbol{s}_\ell| = s^0_\ell + s^1_\ell + s^2_\ell$. Then, Theorems 2 and 3 and (15) imply the claim. □

4.4 Error vs. work analysis

In this section we analyze the overall computational complexity of the multilevel QMC algorithm with product weights for function systems $(\psi^i_j)_{j\geq 1}$, $i = 0,1,2$. Let us assume that the function systems $(\psi^i_j)_{j\geq 1}$, $i = 0,1,2$, have a multilevel structure with control of the overlaps of the supports. Suppose for $i = 0,1,2$, there exist enumerations $j_i : \bigtriangledown \to \mathbb{N}$, where elements of $\lambda \in \bigtriangledown^i$ are tuples of the form $\lambda = (\ell,k)$, where $k \in \bigtriangledown^i_\ell$. Also define $|\lambda| = |(\ell,k)| = \ell$ for every $\lambda \in \bigtriangledown$. We assume that $|\bigtriangledown^i_\ell| = \mathscr{O}(2^{d\ell})$, $|\mathrm{supp}(\psi_\lambda)| = \mathscr{O}(2^{-d\ell})$, $\lambda = (\ell,k) \in \bigtriangledown$, and there there exists $K > 0$ such that for every $x \in D$ and every $\ell \in \mathbb{N}_0$,

$$\left|\left\{\lambda \in \bigtriangledown^i : |\lambda| = \ell, \psi^i_\lambda(x) \neq 0\right\}\right| \leq K. \tag{18}$$

Moreover, we assume that

$$\rho^i_{r,\lambda} \lesssim 2^{-|\lambda|(\widehat{\alpha}-r)}, \quad \lambda \in \bigtriangledown^i, r = 0,\dots,t, i = 0,1,2,$$

for $\widehat{\alpha} > t$. Note that $\rho^i_{r,j(\lambda)} \lesssim j^{-(\widehat{\alpha}-r)/d}$, $j \geq 1$. We equilibrate the sparsity contribution of the sequences $(\rho^i_{0,\lambda})_{\lambda\in\bigtriangledown^i}$ and $(\rho^i_{t,\lambda})_{\lambda\in\bigtriangledown^i}$ in the weight sequence in (17). Hence, we choose $\theta = \tau/\widehat{\alpha}$. Furthermore, with this choice of θ we also equilibrate the errors in the multilevel QMC estimate from Theorem 4, where the truncation dimension $\boldsymbol{s}_\ell$

is still a free parameter. The error contributions in Theorem 4 are equilibrated for the choice

$$s_\ell^i \sim M_\ell, \quad i = 0,1,2.$$

In conclusion, the overall error of multilevel QMC with $L \in \mathbb{N}_0$ levels satisfies for every $p > d/(\widehat{\alpha} - \tau)$,

$$\text{error}_L = \mathscr{O}\left(M_L^{-(t'+t)/d} + \sum_{\ell=0}^{L} N_\ell^{-1/p} M_{\ell-1}^{-(t'+t)d} \right). \tag{19}$$

We assume that we have a procedure at hand that approximates the solution of a parameter instance up to accurary, which is consistent with the discretization error, in computational cost

$$\text{work}_{\text{PDE solver},\ell} = \mathscr{O}(M_\ell \log(M_\ell)), \quad \ell \geq 0.$$

Recall from Section 3.2 that in the self-adjoint case with homogeneous Dirichlet boundary conditions, i.e., $b_i(\boldsymbol{y}^1) = 0$ and $\Gamma_2 = \emptyset$, this can be achieved by PCG with BPX or symmetric V-cycle as preconditioners. The total computational cost of the multilevel QMC algorithm is the sum of the cost of the CBC construction, the cost of assembling the stiffness matrix plus the cost of approximating the solution of the linear systems multiplied by the number of QMC points. Specifically,

$$\text{work}_L = \mathscr{O}\left(\sum_{\ell=0}^{L} M_\ell N_\ell \log(N_\ell) + N_\ell (M_\ell \log(M_\ell) + M_\ell) \right),$$

where we remind that the dimension of integration on each level ℓ is $\mathscr{O}(M_\ell)$. Since the QMC convergence rate $1/p$ satisfies the strict inequality $\chi := 1/p < (\widehat{\alpha} - \tau)/d$, also the rate $\chi(1+\varepsilon)$ is admissible for sufficiently small $\varepsilon > 0$. This way the sample numbers can be reduced to $N_\ell^{1/(1+\varepsilon)}$, which allows us to estimate $N_\ell^{1/(1+\varepsilon)} \log(N_\ell) \leq N_\ell^{1/(1+\varepsilon)} N_\ell^{\varepsilon/(1+\varepsilon)} (1+\varepsilon)/(e\varepsilon) \leq N_\ell (1+\varepsilon) e/(\varepsilon)$, where we used the elementary estimate $\log(N) \leq N^{\varepsilon'}/(e\varepsilon')$ for every $N \geq 1$, $\varepsilon' > 0$. Thus, we obtain the estimate of the work

$$\text{work}_L = \mathscr{O}\left(\sum_{\ell=0}^{L} M_\ell \log(M_\ell) N_\ell \right). \tag{20}$$

By [23, Lemma 4.9], it holds that $M_\ell = \mathscr{O}(2^{d\ell})$. The sample numbers are now obtained by optimizing the error versus the computational work, cp. [31, Section 3.7]. For the error and work estimates in (19) and in (20), sample numbers are derived in [21, Section 6]. Specifically, by [21, Equations (26) and (27)],

$$N_\ell := \left\lceil N_0 (M_\ell^{-1-\tau/d} \log(M_\ell)^{-1})^{p/(1+p)} \right\rceil, \quad \ell = 1, \dots, L, \tag{21}$$

where

$$N_0 := \begin{cases} M_L^{\tau p/d} & \text{if } d < p\tau, \\ M_L^{\tau p/d} \log(M_L)^{p(p+2)/(p+1)} & \text{if } d = p\tau, \\ M_L^{(1+\tau/d)p/(p+1)} \log(M_L)^{p/(p+1)} & \text{if } d > p\tau. \end{cases} \quad (22)$$

The corresponding work satisfies (see for example [21, p. 22])

$$\text{work}_L = \begin{cases} \mathscr{O}(M_L^{\tau p/d}) & \text{if } d < p\tau, \\ \mathscr{O}(M_L^{\tau p/d} \log(M_L)^{p+2}) & \text{if } d = p\tau, \\ \mathscr{O}(M_L \log(M_L)) & \text{if } d > p\tau. \end{cases}$$

We summarize the preceding discussion in the following theorem stating the ε-complexity of the multilevel QMC algorithm.

Theorem 5. *For $p \in (d/(\widehat{\alpha}-\tau), 1]$, assuming $d < \widehat{\alpha}-\tau$, an error threshold $\varepsilon > 0$, i.e.,*

$$|\mathbb{E}(G(q)) - Q_L(G(\tilde{q}^L))| = \mathscr{O}(\varepsilon)$$

can be achieved with

$$\text{work}_L = \begin{cases} \mathscr{O}(\varepsilon^{-p}) & \text{if } d < p\tau, \\ \mathscr{O}(\varepsilon^{-p} \log(\varepsilon^{-1})^{p+2}) & \text{if } d = p\tau, \\ \mathscr{O}(\varepsilon^{-d/\tau} \log(\varepsilon^{-1})) & \text{if } d > p\tau. \end{cases}$$

5 Bayesian inverse UQ

The preceding considerations pertained to so-called *forward UQ* for the ADR problem (1) with uncertain input data $u = ((a_{ij}), (b_j), c)$ taking values in certain subsets of the function spaces X_m in (3). The goal of computation is the efficient evaluation of *ensemble averages*, i.e. the expected response over all *parametric inputs u* as in (10) with respect to a probability measure on the parameter domains U^i.

In Bayesian inverse UQ, we are interested in similar expectations of a QoI of the forward response of the ADR PDE, *conditional to noisy observations of functional of the responses*. Again, a (prior) probability measure on the uncertain (and assumed nonobservable) *parametric ADR PDE inputs u* in (10) is prescribed. As explained in [12, 40], in this setting Bayes' theorem provides a formula for the conditional expectation as high-dimensional, parametric deterministic integral which, as shown in [39, 15, 17], is amenable to deterministic HoQMC integration affording convergence rates which are superior to those of, e.g., MCMC methods [27, 20].

5.1 Formulation of the BIP

Specifically, assume at hand noisy observations of the ADR PDE response $q = (\mathscr{L}(u))^{-1}f$ subject to *additive Gaussian observation noise* $\boldsymbol{\eta}$, i.e.

$$\boldsymbol{\delta} = \boldsymbol{G}(q) + \boldsymbol{\eta}. \tag{23}$$

In (23), q denotes the response of the uncertain input u, $\boldsymbol{G} = (G_1, \ldots, G_K)$ is a vector of K (linear) observation functionals, i.e., $G_i \in V^*$, the additive noise $\boldsymbol{\eta}$ is assumed centered and normally distributed with positive covariance Γ, i.e.,$\boldsymbol{\eta} \sim \mathscr{N}(0,\Gamma)$, and the data $\boldsymbol{\delta} \in \mathbb{R}^K$ is supposed to be available. We introduce the so-called prior measure $\boldsymbol{\pi}$ on X_0 as the law of $U \ni \boldsymbol{y} \mapsto u(\boldsymbol{y}) \in X_0$ with respect to the uniform product measure $\mathrm{d}\boldsymbol{y}$ on U. The density of the posterior distribution with respect to the prior is given by [12, Theorem 14]

$$U \ni \boldsymbol{y} \mapsto \frac{1}{Z} \exp\left(-\Phi_\Gamma(q(\boldsymbol{y});\boldsymbol{\delta})\right), \tag{24}$$

where the negative *log-likelihood* Φ_Γ is given by

$$\Phi_\Gamma(q(\boldsymbol{y});\boldsymbol{\delta}) := \frac{1}{2}(\boldsymbol{\delta} - \boldsymbol{G}(q(\boldsymbol{y})))^\top \Gamma^{-1}(\boldsymbol{\delta} - \boldsymbol{G}(q(\boldsymbol{y}))) \quad \forall \boldsymbol{y} \in U.$$

Since (6) implies $\sup_{\boldsymbol{y}\in U} \|q(\boldsymbol{y})\|_V < \infty$, the *normalization constant* in (24) satisfies

$$Z := \int_{X_0} \exp\left(-\Phi_\Gamma(q;\boldsymbol{\delta})\right) \boldsymbol{\pi}(\mathrm{d}u) = \int_U \exp\left(-\Phi_\Gamma(q(\boldsymbol{y});\boldsymbol{\delta})\right) \mathrm{d}\boldsymbol{y} > 0.$$

The posterior measure will be denoted by $\boldsymbol{\pi}^{\boldsymbol{\delta}}$ and the posterior with respect to $\widetilde{q}^{\boldsymbol{s},\mathscr{T}_\ell}$ will be denoted by $\widetilde{\boldsymbol{\pi}}^{\boldsymbol{\delta}}_{\boldsymbol{s},\mathscr{T}_\ell}$. The QoI being assumed bounded linear functionals applied to $q \in V$ (which could be weakened [20]) admit a unique representer $\phi \in V^*$. For any QoI $\phi \in V^*$, denote the expectation with respect to the posterior of ϕ by

$$\mathbb{E}^{\boldsymbol{\pi}^{\boldsymbol{\delta}}}(\phi) := \int_{X_0} \phi(q) \boldsymbol{\pi}^{\boldsymbol{\delta}}(\mathrm{d}u) = \frac{1}{Z} \int_U \phi(q(\boldsymbol{y})) \exp\left(-\Phi_\Gamma(q(\boldsymbol{y});\boldsymbol{\delta})\right) \mathrm{d}\boldsymbol{y}.$$

Here, $\Phi_\Gamma(q(\boldsymbol{y});\boldsymbol{\delta})$ is Lipschitz continuous with respect to $\boldsymbol{\delta}$ and with respect to $q(\boldsymbol{y})$, $\boldsymbol{y} \in U$. As a consequence of (15), for every $\boldsymbol{s} \in \mathbb{N}^3$ and $\ell \geq 0$,

$$|\mathbb{E}^{\boldsymbol{\pi}^{\boldsymbol{\delta}}}(\phi) - \mathbb{E}^{\widetilde{\boldsymbol{\pi}}^{\boldsymbol{\delta}}_{\boldsymbol{s},\mathscr{T}_\ell}}(\phi)| \leq C \left(M_\ell^{-(\min\{t,k\}+\min\{t',k\})/2} + \max_{i=0,1,2} \sup_{j>s^i} \{(\rho^i_{0,j})^{-1}\} \right), \tag{25}$$

where we also used that $\Phi_\Gamma(q(\boldsymbol{y});\boldsymbol{\delta})$ and $\Phi_\Gamma(\widetilde{q}^{\boldsymbol{s},\mathscr{T}_\ell}(\boldsymbol{y});\boldsymbol{\delta})$ are uniformly upper bounded with respect to $\boldsymbol{y} \in U$. See also the discussion in [14, Section 3.3]. Here, the abstract assumptions made in [14, Section 3.3], stemming from [12], may be verified concretely. The estimate in (25) is not just a restatement of the results of [14, 15]. Here, a general parametric ADR forward problem on polygonal domains is considered and higher order FE is admitted based on regularity in weighted spaces of Kondrat'ev type. The corresponding FE approximation results are proved in Section 6.

5.2 Multi-Level HoQMC-FE Discretization

The expectation with respect to the posterior measure $\widetilde{\boldsymbol{\pi}}^{\boldsymbol{\delta}}_{\boldsymbol{s},\mathscr{T}_\ell}$ is an integral over a $|\boldsymbol{s}|$-dimensional parameter space and may therefore be approximated by multilevel QMC. We recall the FE spaces $\mathscr{T}_\ell$ and suppose a sequence of $\boldsymbol{s}_\ell$ of dimension truncations, $\ell = 0,\dots,L$, where $L \in \mathbb{N}$ is the maximal level. The error analysis will be along the lines of [15, Section 4], see also [40, 41]. Following the notation in [15], define for $\ell = 0,\dots,L$,

$$\mathbb{E}^{\widetilde{\boldsymbol{\pi}}^{\boldsymbol{\delta}}_{\boldsymbol{s}_\ell,\mathscr{T}_\ell}}(\phi) = \frac{\int_{[-1/2,1/2]^{|\boldsymbol{s}|}} \phi(\widetilde{q}^{\boldsymbol{s}_\ell,\mathscr{T}_\ell}(\boldsymbol{y}))\Theta_\ell(\boldsymbol{y})\mathrm{d}\boldsymbol{y}}{\int_{[-1/2,1/2]^{|\boldsymbol{s}|}} \Theta_\ell(\boldsymbol{y})\mathrm{d}\boldsymbol{y}} =: \frac{Z'_\ell}{Z_\ell},$$

where $\Theta_\ell(\boldsymbol{y}) := \exp(-\Phi_\Gamma(\widetilde{q}^{\boldsymbol{s}_\ell,\mathscr{T}_\ell}(\boldsymbol{y});\boldsymbol{\delta}))$. In [15, Sec.4.2], multilevel QMC *ratio* and *splitting* estimators were proposed for the deterministic approximation of Z'_L/Z_L. They are, for a sequence of numbers of QMC points $(N_\ell)_{\ell=0,\dots,L}$ and dimension truncations $(\boldsymbol{s}_\ell)_{\ell=0,\dots,L}$ defined by

$$Q_{L,\mathrm{ratio}} := \frac{Q_L(\phi(\widetilde{q}^L)\Theta_L)}{Q_L(\Theta_L)} \tag{26}$$

and

$$Q_{L,\mathrm{split}} := \frac{Q_{|\boldsymbol{s}|_0,N_0}(\phi(\widetilde{q}^0)\Theta_0)}{Q_{|\boldsymbol{s}|_0,N_0}(\Theta_0)} + \sum_{\ell=1}^{L} \frac{Q_{|\boldsymbol{s}|_\ell,N_\ell}(\phi(\widetilde{q}^\ell)\Theta_\ell)}{Q_{|\boldsymbol{s}|_\ell,N_\ell}(\Theta_\ell)} - \frac{Q_{|\boldsymbol{s}|_\ell,N_\ell}(\phi(\widetilde{q}^{\ell-1})\Theta_{\ell-1})}{Q_{|\boldsymbol{s}|_\ell,N_\ell}(\Theta_{\ell-1})}. \tag{27}$$

The error analysis of these estimators requires that the integrands satisfy certain parametric regularity estimates. In Section 4.2, parametric regularity estimates of $q(\boldsymbol{y}) - \widetilde{q}^{\boldsymbol{s},\mathscr{T}_\ell}(\boldsymbol{y})$ were shown using analytic continuation. The integrands $\phi(\widetilde{q}^\ell)\Theta_\ell$ and Θ_ℓ depend analytically on $\widetilde{q}^\ell$ and are as compositions and products of holomorphic mappings again holomorphic.

The error of the ratio and splitting estimators are analyzed in [15, Sections 4.3.2 and 4.3.3] in the setting of globally supported function systems. However, the proofs of [15, Theorem 4.1 and Theorem 4.2] are applicable.

Proposition 3. *Let the assumptions and the setting of steering parameters θ, p, t', t, $\boldsymbol{s}_\ell$, N_ℓ, M_ℓ, $\ell = 0,\dots,L$ of Q_L in Theorem 4 be satisfied. Suppose that there exists a constant $C_0 > 0$, which does not depend on L, such that $Q_L(\Theta_L) \geq C_0$. Then,*

$$|\mathbb{E}^{\boldsymbol{\pi}^{\boldsymbol{\delta}}}(\phi) - Q_{L,\mathrm{ratio}}| \leq C \left(M_L^{-(t'+t)/d} + \max_{i=0,1,2} \sup_{j>s^i} \{(\rho^i_{0,j})^{-1}\} \right.$$
$$\left. + \sum_{\ell=0}^{L} N_\ell^{-1/p} \left(M_{\ell-1}^{-(t'+t)/d} + \max_{i=0,1,2} \sup_{j>s^i_{\ell-1}} \{(\rho^i_{0,j})^{-\theta}\} \right) \right) .$$

Proof. The assertion follows as [15, Theorem 4.1]. As mentioned above, $\phi(\widetilde{q}^\ell)\Theta_\ell$ maybe analytically extended to a suitable polydisc as in the proof of Theorem 3. The same line of argument use in the proof of [15, Theorem 4.1] may be applied here. Further details are left to reader. ☐

The error estimate from Proposition 3 for Bayesian estimation can also be shown for the splitting estimator $Q_{L,\mathrm{split}}$ along the lines of the proof of [15, Theorem 4.2].

Since the posterior density depends analytically on the response q, the QMC sample numbers for ratio and splitting estimators $Q_{L,\mathrm{ratio}}$ and $Q_{L,\mathrm{split}}$ are the same as those for forward UQ in (21) and (22). In particular, also the same ε-complexity estimates from Theorem 5 hold under the same assumptions on the steering parameters.

Remark 2. Forward and Bayesian inverse UQ for uncertain domains is by pullbacks to a polygonal *nominal* or reference domain a straightforward extension of the presented theory. This requires the extension of the regularity theory to parametric right hand sides $f(\boldsymbol{y})$. Since this dependence is inherited by the parametric solution due to linearity, we did not explicitly consider it for the sake of a concise presentation.

6 Proofs

We provide proofs of several results in the main text. They were postponed to this section to increase readability of the main text.

6.1 Numerical integration

In the following proofs, we require a nodal interpolant. As a preparation, for $k \geq 3$, we introduce certain subsets of $\mathscr{T}$

$$\mathscr{T}^{k'} := \left\{ K \in \mathscr{T} \backslash \mathscr{T}^{k'-1} : \overline{K} \cap \bigcup_{K' \in \mathscr{T}^{k'-1}} \overline{K'} \neq \emptyset \right\}, \quad k' = 2, \ldots, k-1,$$

where $\mathscr{T}^1 := \{K \in \mathscr{T} : \overline{K} \cap \mathscr{C} \neq \emptyset\}$ and $\mathscr{T}^k := \mathscr{T} \backslash \mathscr{T}^{k-1}$. For $k = 2$, $\mathscr{T}^2 := \mathscr{T} \backslash \mathscr{T}^1$ and $\mathscr{T}^1$ is defined as above. For $k = 1$, $\mathscr{T}^1 := \mathscr{T}$. We define a FE space such that in the elements abutting at a vertex, $\mathbb{P}_1$ FE is used and for the remaining, "interior" elements, $\mathbb{P}_k$ FE is used such that the polynomial degree of neighboring elements only differs by one, i.e.,

$$\widetilde{V}^k_{\mathscr{T}} := \{v \in V : v|_K \in \mathbb{P}_{k'}(K), K \in \mathscr{T}^{k'}, k' = 1, \dots, k\}.$$

The potential change of the polynomial degree in neighboring elements near the singular points constitutes a difficulty in defining a nodal interpolant for $k \geq 2$. Let $K_1 \in \mathscr{T}^{k'-1}, K_2 \in \mathscr{T}^{k'}$ be neighboring triangles such that $\overline{K_1} \cap \overline{K_2} =: e$ denotes the common edge. To avoid a discontinuity across the edge e, the usual nodal interpolant $I^{k'}_{K_2}$ may need to be corrected. For $v \in C^0(\overline{K}_2)$, the discontinuity $(I^{k'}_{K_2} v|_e - I^{k'-1}_e v)$ is equal to zero at the endpoints of the edge e. By [43, Lemma 4.55], there exists $(I^{k'}_{K_2} v|_e - I^{k'-1}_e v)_{\mathrm{lift},k',e} \in \mathbb{P}^{k'}(K_2)$ such that $(I^{k'}_{K_2} v|_e - I^{k'-1}_e v)_{\mathrm{lift},k,e} = (I^{k'}_{K_2} v|_e - I^{k'-1}_e v)$ on the edge e, $(I^{k'}_{K_2} v|_e - I^{k'-1}_e v)_{\mathrm{lift},k',e} = 0$ on the remaining edges of K_2, and it holds

$$\begin{aligned} \|(I^{k'}_{K_2} v|_e - I^{k'-1}_e v)_{\mathrm{lift},k',e}\|^2_{H^1(K_2)} &\leq C h_{K_2} \|(I^{k'}_{K_2} v)|_e - I^{k'}_e v|_e\|^2_{H^1(e)} \\ &\leq C' h^{2k'-1}_{K_2} |I^{k'}_e v|^2_{H^{k'}(e)} \leq C' h^{2k'-2}_{K_2} |I^{k'}_{K_2} v|^2_{H^{k'}(K_2)}, \end{aligned} \tag{28}$$

where we applied the approximation property in dimension $d-1=1$, cp. [9, Theorem 3.1.6], the shape regularity of $\mathscr{T}$, and the fact that k'-th order partial derivatives of $I^{k'}_{K_2} v$ are constant on K_2.

We will define an interpolant $I_{\mathscr{T}} : \mathscr{K}^{k+1}_{a+1} \to V^k_{\mathscr{T}}$ by

$$I_{\mathscr{T}} v = \begin{cases} I^1_K v & \text{if } K \in \mathscr{T}^1, \\ I^{k'}_K v - (I^{k'}_K v|_e - I^{k'-1}_e v)_{\mathrm{lift},k',e} & \text{if } K \in \mathscr{T}^{k'}, e := \overline{K} \cap \overline{\mathscr{T}^{k'-1}} \neq \emptyset, \\ & k' = 2, \dots, k, \\ I^k_K v & \text{if } K \in \mathscr{T}^k, \overline{K} \cap \overline{\mathscr{T}^{k-1}} = \emptyset, \end{cases}$$

where $I^{k'}_K$ is the usual nodal interpolant of order $k' \in \mathbb{N}$ on the element K and we introduced the notation $\overline{\mathscr{T}^{k'}} := \bigcup_{K' \in \mathscr{T}^{k'}} \overline{K'}$, $k' = 1, \dots, k$. This first paragraph of Section 6.1 originates from [25, Section 3.2], where also a proof of Proposition 2 is given.

Proposition 4. *Suppose that for some integer $k \in \mathbb{N}$, $k' \in \mathbb{N}_0$,*

$$E^k_{\widehat{K}}(\widehat{\phi}) = 0, \quad \forall \widehat{\phi} \in \mathbb{P}_{k'+k-1}(\widehat{K}).$$

Then, there exists a constant $C > 0$ such that for every $K \in \mathscr{T}$, $a \in W^{k,\infty}(K)$, $v \in \mathbb{P}_k(K)$, $w \in \mathbb{P}_{k'}(K)$,

$$|E^k_K(avw)| \leq C h^k_K \left(\sum_{j=0}^{k} |a|_{W^{k-j,\infty}(K)} |v|_{H^j(K)} \right) \|w\|_{L^2(K)}.$$

Proof. This is a version of [9, Theorem 4.1.4]. The claimed estimate follows by [9, Equations (4.1.47) and (4.1.46), Theorems 3.1.2 and 3.1.3]. We note that we did not assume here $v \in \mathbb{P}_{k-1}(K)$, which results in the sum over $j = 0,\dots,k$. However, if $v \in \mathbb{P}_{k-1}(K)$ for some $k \geq 1$, then $|v|_{H^k(K)} = 0$. □

Lemma 4. *Let $K \in \mathscr{T}$ be such that $\boldsymbol{c}_i \in \overline{K}$, $i \in \{1,\dots,J\}$. There exists a constant $C > 0$ independent of K such that for every $v \in \mathbb{P}_1(K)$ satisfying $v(\boldsymbol{c}_i) = 0$*

$$\|r_D^{-1} v\|_{L^\infty(K)} \leq C \|r_D^{-1} v\|_{L^2(K)} \det(B_K)^{-1/2}.$$

Proof. We will prove the main step on the reference element $\widehat{K}$. It is easy to see that $\|r_D^{-1} v\|_{L^\infty(K)} = \|\widehat{r_D}^{-1}\widehat{v}\|_{L^\infty(\widehat{K})}$ and $\|\widehat{r_D}^{-1}\widehat{v}\|_{L^2(\widehat{K})} = \|r_D^{-1} v\|_{L^2(K)} \det(B_K)^{-1/2}$.

Suppose that $\widehat{K} := \{\widehat{x} \in (0,1)^2 : 0 < \widehat{x}_1 + \widehat{x}_2 < 1\}$ and that $F_K^{-1}(\boldsymbol{c}_i) = 0$. The space $\{\widehat{v} \in \mathbb{P}_1(\widehat{K}) : v(0) = 0\}$ is spanned by the monomials $\{\widehat{x}_1, \widehat{x}_2\}$. By [9, Theorem 3.1.3] and shape regularity of $\mathscr{T}$, there exists constants $C, C' > 0$ independent of K such that $\|B_K\| \leq C h_K$ and $\|B_K^{-1}\| \leq C' h_K^{-1}$. Note that $\min_{i=1,2}\{\int_{\widehat{K}} \frac{|\widehat{x}_i|}{|\widehat{x}|} \mathrm{d}\widehat{x}\} =: C'' > 0$. This implies by elementary manipulations and the Cauchy–Schwarz inequality

$$\begin{aligned}\sup_{\widehat{x} \in \widehat{K}} \frac{|\widehat{x}|}{|B_K \widehat{x}|} = \|B_K^{-1}\| \leq 2\frac{C'}{C'' h_K} \int_{\widehat{K}} \frac{|\widehat{x}_i|}{|\widehat{x}|} \mathrm{d}\widehat{x} \leq 2\frac{CC'}{C''} \int_{\widehat{K}} \frac{|\widehat{x}_i|}{\|B_K\| |\widehat{x}|} \mathrm{d}\widehat{x} \\ \leq 2^{3/2} \frac{CC'}{C''} \left(\int_{\widehat{K}} \frac{|\widehat{x}_i|^2}{|B_K \widehat{x}|^2} \mathrm{d}\widehat{x} \right)^{1/2}.\end{aligned} \tag{29}$$

On $\widehat{K}$, $\widehat{r}_D(\widehat{x}) = r_D(F_K(\widehat{x})) \sim |B_K \widehat{x}|$. Let $\widehat{v} = \widehat{v}_1 \widehat{x}_1 + \widehat{v}_2 \widehat{x}_2$. Thus, by (29) there exist constants $C, C' > 0$ independent of K such that

$$\|\widehat{r_D}^{-1}\widehat{v}\|_{L^\infty(K)} \leq C \left(|\widehat{v}_1| \sup_{\widehat{x} \in \widehat{K}} \frac{|\widehat{x}_1|}{|B_K \widehat{x}|} + |\widehat{v}_2| \sup_{\widehat{x} \in \widehat{K}} \frac{|\widehat{x}_2|}{|B_K \widehat{x}|} \right) \leq C' \left(\int_{\widehat{K}} \frac{|\widehat{v}(\widehat{x})|^2}{|B_K \widehat{x}|^2} \mathrm{d}\widehat{x} \right)^{1/2}.$$

The proof of the lemma is complete, since $\widehat{r}_D(\widehat{x}) \sim |B_K \widehat{x}|$ on $\widehat{K}$. □

Proposition 5. *Let $K \in \mathscr{T}$ be such that $\boldsymbol{c}_i \in \overline{K}$, for some $i \in \{1,\dots,J\}$. Let $E_K^1(\cdot)$ denote the error from a one point quadrature in the barycenter $\bar{x}$ of K. Let $\delta_1, \delta_2, \delta_3, \delta_4 \in [0,1)$. Then there exists a constant $C > 0$ such that for every $(r_D^{\delta_3+\delta_4} a) \in L^\infty(K)$ satisfying $r_D^{\delta_1+\delta_2} |\nabla a| \in L^\infty(K)$ such that point evaluation at $\bar{x}$ is well defined and for every $v, w \in \mathbb{P}_k(K)$ for some $k \geq 0$*

$$\begin{aligned}&|E_K^1(avw)| \\ &\quad \leq C h_K^{1-\delta_1} \|r_D^{\delta_1+\delta_2} |\nabla a|\|_{L^\infty(K)} \|v\|_{L^2(K)} \|r_D^{-\delta_2} w\|_{L^2(K)} \\ &\qquad + C h_K^{1-\delta_3} \|r_D^{\delta_3+\delta_4} a\|_{L^\infty(K)} \left(|v|_{H^1(K)} \|r_D^{-\delta_4} w\|_{L^2(K)} + \|r_D^{-\delta_4} v\|_{L^2(K)} |w|_{H^1(K)} \right).\end{aligned}$$

If additionally $v, w \in \mathbb{P}_1(K)$ satisfy that $v(\boldsymbol{c}_i) = 0 = w(\boldsymbol{c}_i)$, the above assumption can be relaxed to $r_D^{\delta_3+i} a \in L^\infty(K)$ and $r_D^{\delta_1+1+i} |\nabla a| \in L^\infty(K)$, $i = 0, 1$, and it holds that

$$|E_K^1(avw)| \le Ch_K^{1-\delta_1}\|r_D^{\delta_1+1+i}|\nabla a|\|_{L^\infty(K)}\|r_D^{-i}v\|_{L^2(K)}\|r_D^{-1}w\|_{L^2(K)}$$
$$+Ch_K^{1-\delta_3}\|r_D^{\delta_3+i}a\|_{L^\infty(K)}(|v|_{H^1(K)}\|r_D^{-i}w\|_{L^2(K)}+\|r_D^{-i}v\|_{L^2(K)}|w|_{H^1(K)}).$$

Proof. We observe that

$$|E_K^1(avw)| \le \int_K |a(x)-a(\bar{x})||v(x)w(x)|\mathrm{d}x + \int_K |a(x)|\left(|v(x)-v(\bar{x})||w(x)|+|v(\bar{x})||w(x)-w(\bar{x})|\right)\mathrm{d}x. \tag{30}$$

For any $f \in W^{1,\infty}(\widetilde{K})$ and any $x \in \widetilde{K}$ ($\widetilde{K}$ a compact subset of K),

$$|f(x)-f(\bar{x})| \le \sup_{\widetilde{x}\in\gamma_{x,\bar{x}}([0,1])}\{|\nabla f(\widetilde{x})|\}|x-\bar{x}|,$$

where $\gamma_{x,\bar{x}}$ is a suitable smooth path such that $\gamma_{x,\bar{x}}(1)=x$ and $\gamma_{x,\bar{x}}(0)=\bar{x}$. We will estimate the two integrals in (30) separately. Since $\boldsymbol{c}_i \in \overline{K}$, the weight function is locally $r_D(x) \simeq |x-\boldsymbol{c}_i|$. Due to the radial monotonicity of $x \mapsto |x-\boldsymbol{c}_i|$, $\gamma_{x,\bar{x}}$ can be chosen such that

$$\inf_{\widetilde{x}\in\gamma_{x,\bar{x}}([0,1])}\{|\widetilde{x}-\boldsymbol{c}_i|\} \in \{|x-\boldsymbol{c}_i|,|\bar{x}-\boldsymbol{c}_i|\}.$$

Hence, there exists a constant $C>0$ independent of K such that for every $x \in K$

$$\frac{|a(x)-a(\bar{x})|}{|x-\bar{x}|}\min\{r_D^{\delta_1+\delta_2}(x),r_D^{\delta_1+\delta_2}(\bar{x})\} \le C\|r_D^{\delta_1+\delta_2}|\nabla a|\|_{L^\infty(K)}. \tag{31}$$

Since all norms on $\mathbb{P}_k(\widehat{K})$ are equivalent, there exists a constant $C>0$ independent of K such that $\|v\|_{L^\infty(K)}=\|\widehat{v}\|_{L^\infty(\widehat{K})}\le C\|\widehat{v}\|_{L^2(\widehat{K})}=C\|v\|_{L^2(K)}\det(B_K)^{-1/2}$. Moreover, since there exists a constant $C>0$ independent of K such that for every $x \in K$, $r_D(x)\le Cr_D(\bar{x})$, there exists a constant $C>0$ independent of K such that

$$\|1/\min\{r_D^{\delta_1},r_D^{\delta_1}(\bar{x})\}\|_{L^2(K)} \le Ch_K^{1-\delta_1}.$$

Similarly, $\|w/\min\{r_D^{\delta_2},r_D^{\delta_2}(\bar{x})\}\|_{L^2(K)}\le C\|r_D^{-\delta_2}w\|_{L^2(K)}$. It also holds that $|x-\bar{x}|\le Ch_K$ and $\det(B_K)\sim h_K^2$. Hence, for a constant $C>0$ independent of K,

$$\begin{aligned}
&\int_K |a(x)-a(\bar{x})||v(x)w(x)|\mathrm{d}x \\
&\quad\le C\|r_D^{\delta_1+\delta_2}|\nabla a|\|_{L^\infty(K)}\int_K |v(x)|\frac{|w(x)|}{r_D^{\delta_2}}\frac{|x-\bar{x}|}{r_D^{\delta_1}}\mathrm{d}x \\
&\quad\le Ch_k\|r_D^{\delta_1+\delta_2}|\nabla a|\|_{L^\infty(K)}\|v\|_{L^\infty(K)}\|r_D^{-\delta_2}w\|_{L^2(K)}\|r_D^{-\delta_1}\|_{L^2(K)} \\
&\quad\le Ch_K^{1-\delta_1}\|r_D^{\delta_1+\delta_2}|\nabla a|\|_{L^\infty(K)}\|v\|_{L^2(K)}\|r_D^{-\delta_2}w\|_{L^2(K)}.
\end{aligned} \tag{32}$$

For the first summand in the second integral in (30), we obtain similarly

$$\begin{aligned}\int_K |a(x)||v(x)-v(\bar{x})||w(x)|\mathrm{d}x &= \int_K r_D^{\delta_3+\delta_4}(x)|a(x)||v(x)-v(\bar{x})||w(x)|\mathrm{d}x \\ &\leq \|r_D^{\delta_3+\delta_4}a\|_{L^\infty(K)}\int_K \frac{|v(x)-v(\bar{x})|}{|x-\bar{x}|}\frac{|w(x)|}{r_D^{\delta_4}(x)}\frac{|x-\bar{x}|}{r_D^{\delta_3}(x)}\mathrm{d}x \\ &\leq C\|r_D^{\delta_3+\delta_4}a\|_{L^\infty(K)}\|\nabla v\|_{L^\infty(K)}\int_K \frac{|w(x)|}{r_D^{\delta_4}(x)}\frac{|x-\bar{x}|}{r_D^{\delta_3}(x)}\mathrm{d}x \\ &\leq Ch_K^{1-\delta_3}\|r_D^{\delta_3+\delta_4}a\|_{L^\infty(K)}|v|_{H^1(K)}\|r_D^{-\delta_4}w\|_{L^2(K)}\end{aligned}$$

using that there is a constant $C>0$ independent of K such that $\|\partial_{x_i}v\|_{L^\infty(K)} = Ch_K^{-1}\|\partial_{\widehat{x}_i}\widehat{v}\|_{L^\infty(\widehat{K})} \leq Ch_K^{-1}\|\partial_{\widehat{x}_i}\widehat{v}\|_{L^2(\widehat{K})} = C\|\partial_{x_i}v\|_{L^2(K)}\det(B_K)^{-1/2}$. Also note that $\det(B_K)\sim h_K^2$. The second summand in the second integral in (30) is estimated analogously.

The second estimate follows since $\|r_D^{-1}v\|_{L^\infty(K)}<\infty$ and $\|r_D^{-1}w\|_{L^2(K)}<\infty$, which allows us to conclude similarly as in (32)

$$\begin{aligned}\int_K |a(x)-a(\bar{x})||v(x)w(x)|\mathrm{d}x &\leq C\|r_D^{\delta_1+1+i}|\nabla a|\|_{L^\infty(K)}\int_K \frac{|v(x)|}{r_D^i(x)}\frac{|w(x)|}{r_D(x)}\frac{|x-\bar{x}|}{r_D^{\delta_1}(x)}\mathrm{d}x \\ &\leq Ch_k^{2-\delta_1}\|r_D^{\delta_1+2}|\nabla a|\|_{L^\infty(K)}\|r_D^{-i}v\|_{L^\infty(K)}\|r_D^{-1}w\|_{L^2(K)} \\ &\leq Ch_K^{1-\delta_1}\|r_D^{\delta_1+2}|\nabla a|\|_{L^\infty(K)}\|r_D^{-i}v\|_{L^2(K)}\|r_D^{-1}w\|_{L^2(K)},\end{aligned}$$

where we used that $\|r_D^{-1}v\|_{L^\infty(K)} \leq Ch_K^{-1}\|r_D^{-1}v\|_{L^2(K)}$ for a constant $C>0$ that neither depend on K nor on v, which follows by Lemma 4. □

Proof.[of Theorem 1] The proof generalizes [9, Theorem 4.1.6] to the case of local mesh refinement and singularities of the solution and the coefficients. We recall the *first Strang lemma*, see for example [9, Theorem 4.1.1]

$$\begin{aligned}&\|q-\widetilde{q}^{\mathscr{T}}\|_V \\ &\quad\leq \frac{u_{\max}}{u_{\min}}\inf_{v^{\mathscr{T}}\in V^k_{\mathscr{T}}}\left\{\|q-v^{\mathscr{T}}\|_V + \sup_{0\neq w^{\mathscr{T}}\in V^k_{\mathscr{T}}}\frac{|A(u)(v^{\mathscr{T}},w^{\mathscr{T}})-\widetilde{A}(u)(v^{\mathscr{T}},w^{\mathscr{T}})|}{\|w^{\mathscr{T}}\|_V}\right\},\end{aligned}$$

where $u_{\max}$ and $u_{\min}$ are continuity and coercivity constants of $A(u),\widetilde{A}(u)$. The right hand side of the first Strang lemma will be upper bounded by choosing $v^{\mathscr{T}} := I_{\mathscr{T}}q \in \widetilde{V}^k_{\mathscr{T}}$.

We will treat the second, first, and zero order terms separately and start with the second order term. We decompose $A(u)=\sum_{i,j=1}^2 A(a_{ij})+\sum_{j=1}^4 A(b_j)+A(c)$ and $\widetilde{A}(u)=\sum_{i,j=1}^2\widetilde{A}(a_{ij})+\sum_{j=1}^4\widetilde{A}(b_j)+\widetilde{A}(c)$. As in the proof of Proposition 2, we work on the elements. There, we have to treat three cases, $K\in\mathscr{T}^1$, $K\in\mathscr{T}^{k'}$ and $\overline{K}\cap\overline{\mathscr{T}^{k'-1}}\neq\emptyset$, $k'=2,\dots,k$, and $K\in\mathscr{T}^k$ and $\overline{K}\cap\overline{\mathscr{T}^{k-1}}=\emptyset$. We observe

$$
\begin{aligned}
&|\sum_{i,j=1}^{2} A(a_{ij})(I_{\mathscr{T}}q, w^{\mathscr{T}}) - \sum_{i,j=1}^{2} \widetilde{A}(a_{ij})(I_{\mathscr{T}}q, w^{\mathscr{T}})| \\
&\leq \sum_{K\in\mathscr{T}^1} \sum_{i,j=1}^{2} |E_K^1(a_{ij}\partial_j I_K^1 q \partial_i w^{\mathscr{T}})| + \sum_{K\in\mathscr{T}^k, \overline{K}\cap\overline{\mathscr{T}^{k-1}}=\emptyset} \sum_{i,j=1}^{2} |E_K^k(a_{ij}\partial_j I_K^k q \partial_i w^{\mathscr{T}})| \\
&\quad + \sum_{k'=2}^{k} \sum_{K\in\mathscr{T}^{k'}, e:=\overline{K}\cap\overline{\mathscr{T}^{k'-1}}\neq\emptyset} \sum_{i,j=1}^{2} |E_K^{k'}(a_{ij}\partial_j (I_K^{k'}q - (I_K^{k'}q|_e - I_e^{k'-1}q)_{\mathrm{lift},k',e})\partial_i w^{\mathscr{T}})|.
\end{aligned}
$$

By (11), for $K \in \mathscr{T}^1$,

$$h_K^{1-\delta} \leq Ch^{(1-\delta)/(1-\beta)} \leq Ch^k. \tag{33}$$

For $K \in \mathscr{T}^1$, by Proposition 5 (with $\delta_1 = \delta_3 = \delta$, $\delta_2 = \delta_4 = 0$) and (33)

$$|E_K^1(a_{ij}\partial_j I_K^1 q \partial_i w^{\mathscr{T}})| \leq Ch^k \|a_{ij}\|_{W_\delta^{1,\infty}(K)} \|I_K^1 q\|_{H^1(K)} \|w^{\mathscr{T}}\|_{H^1(K)}$$

and [43, Lemma 4.16] implies with the triangle inequality $\|I_K^1 q\|_{H^1(K)} \leq C(\|q\|_{H^1(K)} + h_K^{1-\delta}|q|_{H_\delta^2(K)})$. For $K \in \mathscr{T}^k$ such that $\overline{K}\cap\overline{\mathscr{T}^{k-1}} = \emptyset$, by Proposition 4

$$
\begin{aligned}
&|E_K^k(a_{ij}\partial_j I_K^k q \partial_i w^{\mathscr{T}})| \\
&\quad \leq Ch^k \sum_{\ell=0}^{k-1} \inf_{x\in K} r_D^{\beta(k-\ell)}(x) |a_{ij}|_{W^{k-\ell,\infty}(K)} \inf_{x\in K} r_D^{\beta\ell}(x) |I_K^k q|_{H^{\ell+1}(K)} |w^{\mathscr{T}}|_{H^1(K)}.
\end{aligned} \tag{34}
$$

It follows directly from (11),

$$h_K \leq Ch^{1/(1-(\beta-\alpha))} r_D^{\alpha/(1-(\beta-\alpha))}(x) \quad \forall x \in K, \forall \alpha \in (0,\beta).$$

We choose $\alpha := (1-\beta)(k'-2+\delta)/(1-\delta)$ and apply $(1-\delta)/(1-\beta) > k$,

$$h_K \leq Ch^{k/(k'-1)} r_D^{(\delta+k'-2)/(k'-1)}(x) \quad \forall x \in K, k' = 2,\dots,k. \tag{35}$$

For $K \in \mathscr{T}^{k'}$ such that $e := \overline{K}\cap\overline{\mathscr{T}^{k'-1}} \neq \emptyset$, $k' = 2,\dots,k$, by Proposition 4 and (35)

$$
\begin{aligned}
&|E_K^{k'}(a_{ij}\partial_j (I_K^{k'}q - (I_K^{k'}q|_e - I_e^{k'-1}q)_{\mathrm{lift},k',e})\partial_i w^{\mathscr{T}})| \\
&\quad \leq Ch_K^{k'} \sum_{\ell=0}^{k'-1} |a_{ij}|_{W^{k'-\ell,\infty}(K)} |I_K^{k'}q - (I_K^{k'}q|_e - I_e^{k'-1}q)_{\mathrm{lift},k',e}|_{H^{\ell+1}(K)} |w^{\mathscr{T}}|_{H^1(K)} \\
&\quad \leq Ch^k \sum_{\ell=0}^{k'-1} \inf_{x\in K} r_D^{\delta+k'-1}(x) |a_{ij}|_{W^{k'-\ell,\infty}(K)} \\
&\qquad\qquad \times |I_K^{k'}q - (I_K^{k'}q|_e - I_e^{k'-1}q)_{\mathrm{lift},k',e}|_{H^{\ell+1}(K)} |w^{\mathscr{T}}|_{H^1(K)}.
\end{aligned}
$$

Note that $(1-\delta)/(1-\beta) > k$ implies that $\beta k' > \delta + k' - 1$, $k' = 1,\dots,k$. We observe with [9, Theorem 3.1.6]

$$|I_K^k q|_{H^{\ell+1}(K)} \leq C(|q|_{H^{\ell+1}(K)} + h_K^{k'-1}|q|_{H^{k'+1}(K)}), \quad \ell = 0,\dots,k'-1,$$

and by a similar argument as in the proof of Proposition 2 for $\ell = 0,\dots,k'-1$,

$$|I_K^{k'} q - (I_K^{k'} q|_e - I_e^{k'-1} q)_{\mathrm{lift},k',e}|_{H^{\ell+1}(K)} \leq C(|q|_{H^{\ell+1}(K)} + h_K^{k'-1}|q|_{H^{k'+1}(K)}).$$

By the Cauchy–Schwarz inequality we conclude with the previous inequalities

$$\begin{aligned} &|\sum_{i,j=1}^{2} A(a_{ij})(I_{\mathscr{T}} q, w^{\mathscr{T}}) - \sum_{i,j=1}^{2} \widetilde{A}(a_{ij})(I_{\mathscr{T}} q, w^{\mathscr{T}})| \\ &\qquad \leq C h^k \sum_{i,j=1}^{2} \|a_{ij}\|_{W_\delta^{k,\infty}(K)} \|q\|_{\mathscr{K}_{a+1}^{k+1}(D)} \|w^{\mathscr{T}}\|_{H^1(D)}. \end{aligned}$$

The argument for the advection and reaction terms $\sum_{j=1}^{4} A(b_j)$, $A(c)$ is similar. Here, the additional weight r_D for the advection terms and r_D^2 for the reaction term needs to be accommodated. For the advection term, by the second part of Proposition 5 (with $\delta_1 = \delta_3 = \delta$, $i = 0$) and $K \in \mathscr{T}_1$ for $j = 1,2$

$$|E_K^1(b_j(\partial_j I_K^1 q) w^{\mathscr{T}})| \leq C h^k \|r_D^\delta b_j\|_{W_\delta^{1,\infty}(K)} \|I_K^1 q\|_{H^1(K)} \left(\|r_D^{-1} w^{\mathscr{T}}\|_{L^2(K)} + |w^{\mathscr{T}}|_{H^1(K)} \right) \tag{36}$$

and for $j = 3,4$

$$|E_K^1(b_j I_K^1 q \partial_j w^{\mathscr{T}})| \leq C h^k \|r_D^\delta b_j\|_{W_\delta^{1,\infty}(K)} \left(\|r_D^{-1} I_K^1 q\|_{L^2(K)} + |I_K^1 q|_{H^1(K)} \right) \|w^{\mathscr{T}}\|_{H^1(K)}.$$

For the interior elements $K \in \mathscr{T} \setminus \mathscr{T}^1$, the additional weight r_D can be accommodated by compensating it with $\|r_D^{-1} w^{\mathscr{T}}\|_{L^2(K)}$ as in (36) for $j = 1,2$. If the partial derivative is on the trial function, i.e., $j = 3,4$, the order of the Sobolev semi-norm as for example in (34) is reduced by one to $|I_K^k q|_{H^\ell(K)}$. Here, the weight $r_D^{\beta(k-\ell+1)}$ is assigned to $|b_j|_{W^{k-\ell,\infty}(K)}$, if $\ell \geq 1$. For $\ell = 0$, the additional weight r_D can be compensated by $\|r_D^{-1} I_K^k q\|_{L^2(K)}$. We recall the Hardy inequality from (14), i.e., there exists a constant $C > 0$ such that for every $v \in V$

$$\|r_D^{-1} v\|_{L^2(D)} \leq C \| |\nabla v| \|_{L^2(D)}.$$

Thus, $\|r_D^{-1} w^{\mathscr{T}}\|_{L^2(D)} \leq C \|w^{\mathscr{T}}\|_V$ and $\|r_D^{-1} I_K^k q\|_{L^2(D)} \leq C \|I_K^k q\|_V$. The rest of the proof for the advection terms in analogous to the diffusion terms, which were proved in detail. The argument for the reaction term uses the second part of Proposition 5 with $i = 1$, we omit the details. □

Proof.[of Corollary 1] The solution $g \in V$ to the adjoint problem is characterized by

$$A(u)(w,g) = \langle G, w \rangle_{V^*,V} \quad \forall w \in V.$$

The respective FE approximation $g^{\mathscr{T}}$ is characterized by $A(u)(w^{\mathscr{T}}, g-g^{\mathscr{T}})=0$ for every $w^{\mathscr{T}}\in V^k_{\mathscr{T}}$. Generally, we will follow the proof of [3, Theorem 3.6)]. By a version of [3, Lemma 3.1] for non-symmetric blinear forms $A(u)(\cdot,\cdot)$,

$$G(q)-G(q^{\mathscr{T}})=A(u)(q-q^{\mathscr{T}},g-g^{\mathscr{T}})+\widetilde{A}(u)(\widetilde{q}^{\mathscr{T}},g^{\mathscr{T}})-A(u)(\widetilde{q}^{\mathscr{T}},g^{\mathscr{T}}).$$

As in the previous proof, we begin by estimating the diffusion terms related to a_{ij}. By a similarly argument that we used to show (34) also using Proposition 5, we obtain

$$\begin{aligned}
&|\widetilde{A}(a_{ij})(\widetilde{q}^{\mathscr{T}},g^{\mathscr{T}})-A(a_{ij})(\widetilde{q}^{\mathscr{T}},g^{\mathscr{T}})|\\
&\leq Ch^{k+k'}\left(\sum_{K\in\mathscr{T}^1}\|a_{ij}\|_{W^{1,\infty}_\delta(D)}\|\widetilde{q}^{\mathscr{T}}\|_{H^1(K)}\|g^{\mathscr{T}}\|_{H^1(K)}\right.\\
&\left.+\sum_{K\in\mathscr{T}\setminus\mathscr{T}^1}\sum_{i,j=1}^{2}\sum_{\ell=0}^{k+k'-1}\inf_{x\in K}r_D^{\beta(k+k'-\ell)}(x)|a_{ij}|_{W^{k+k'-\ell,\infty}(K)}\inf_{x\in K}r_D^{\beta\ell}|\partial_j\widetilde{q}^{\mathscr{T}}\partial_i g^{\mathscr{T}}|_{H^\ell(K)}\right).
\end{aligned}$$

Note that $(\partial_j\widetilde{q}^{\mathscr{T}})|_K,(\partial_i g^{\mathscr{T}})|_K\in\mathbb{P}_{k-1}(K)$, which implies that

$$\partial^{\boldsymbol{\alpha}}(\partial_j\widetilde{q}^{\mathscr{T}})|_K=0=\partial^{\boldsymbol{\alpha}}(\partial_i g^{\mathscr{T}})|_K\quad\forall\boldsymbol{\alpha}\in\mathbb{N}_0^2,|\boldsymbol{\alpha}|>k-1.$$

By the product rule and by the Cauchy–Schwarz inequality

$$|\partial_j\widetilde{q}^{\mathscr{T}}\partial_i g^{\mathscr{T}}|_{H^\ell(K)}\leq C\sum_{\ell'=0}^{\ell}|\partial_j\widetilde{q}^{\mathscr{T}}|_{H^{\ell'}(K)}|\partial_i g^{\mathscr{T}}|_{H^{\ell-\ell'}(K)}.$$

By the inverse inequality and the element-wise approximation property of the nodal interpolant, e.g. [9, Theorem 3.1.6] we observe that there exist constants $C,C'>0$ such that for every $K\in\mathscr{T}\setminus\mathscr{T}^1$,

$$\begin{aligned}
|\partial_j\widetilde{q}^{\mathscr{T}}|_{H^{\ell'}(K)}&\leq|q|_{H^{\ell'+1}(K)}+|\partial_j q-I^k_K\partial_j q|_{H^{\ell'}(K)}+|I^k_K\partial_j q-\partial_j\widetilde{q}^{\mathscr{T}}|_{H^{\ell'}(K)}\\
&\leq C|q|_{H^{\ell'+1}(K)}+Ch_K^{-\ell'}\|I^k_K\partial_j q-\partial_j\widetilde{q}^{\mathscr{T}}\|_{L^2(K)}\\
&\leq C|q|_{H^{\ell'+1}(K)}+Ch_K^{-\ell'}(\|I^k_K\partial_j q-\partial_j q\|_{L^2(K)}+\|\partial_j q-\partial_j\widetilde{q}^{\mathscr{T}}\|_{L^2(K)})\\
&\leq C'(|q|_{H^{\ell'+1}(K)}+h_K^{-\ell'}|q-\widetilde{q}^{\mathscr{T}}|_{H^1(K)}).
\end{aligned}$$

Similarly, it holds that $|\partial_i g^{\mathscr{T}}|_{H^{\ell-\ell'}(K)}\leq C|g^{\mathscr{T}}|_{H^{\ell-\ell'+1}(K)}$. The previous element-wise estimates allow us to conclude with the Cauchy–Schwarz inequality

$$\begin{aligned}
&|\widetilde{A}(a_{ij})(\widetilde{q}^{\mathscr{T}},g^{\mathscr{T}})-A(a_{ij})(\widetilde{q}^{\mathscr{T}},g^{\mathscr{T}})|\\
&\quad\leq Ch^{k+k'}\|a_{ij}\|_{W^{k+k',\infty}_\delta(D)}(\|q\|_{\mathscr{K}^{k+1}_{a+1}(D)}+h^{-k}\|q-\widetilde{q}^{\mathscr{T}}\|_V)\|g\|_{\mathscr{K}^{k'+1}_{a+1}(D)}\\
&\quad\leq Ch^{k+k'}\|a_{ij}\|_{W^{k+k',\infty}_\delta(D)}(1+\|u\|_{X_{k,\delta}})\|q\|_{\mathscr{K}^{k+1}_{a+1}(D)}\|g\|_{\mathscr{K}^{k'+1}_{a+1}(D)},
\end{aligned}$$

where we used Theorem 1 in the second step. The argument for the advection and reaction terms $A(b_j)$, $j = 1,\dots,4$, and $A(c)$ is similar. See also the proof of Theorem 1. Since Proposition 2 and (8) imply with Céa's lemma

$$|A(u)(q - q^{\mathscr{T}}, g - g^{\mathscr{T}})| \le Ch^{k+k'} \|f\|_{\mathscr{K}^{k-1}_{a-1}(D)} \|G\|_{\mathscr{K}^{k'-1}_{a-1}(D)},$$

the assertion follows. □

6.2 Approximation of functions with point singularities

In this section we analyze FE approximation rates by biorthogonal wavelet expansions for functions in D with point singularities. We consider regularity in weighted Hölder spaces $\mathscr{W}^{m,\infty}_{\delta}(D)$ and more generally in $X_{m,\delta}$ for $\delta \in [0,1)$. We explicitly define *a-priori truncation* of infinite bioorthogonal wavelet expansions of these functions, mimicking in this way FE mesh refinement in D as in [33] (see also [42]).

Let $(\psi_\lambda)_{\lambda\in\triangledown}$ be a biorthogonal spline wavelet basis of $L^2(D)$ with dual wavelet system $(\widetilde{\psi}_\lambda)_{\lambda\in\triangledown}$, we refer to [45, 13, 38, 44] for concrete constructions. We suppose that $(\psi_\lambda)_{\lambda\in\triangledown}$ and $(\widetilde{\psi}_\lambda)_{\lambda\in\triangledown}$ have the following properties.

1. (biorthogonality) $\int_D \psi_\lambda \widetilde{\psi}_{\lambda'} \mathrm{d}x = \delta_{\lambda\lambda'}$, $\lambda, \lambda' \in \triangledown$,
2. (normalization) $\|\psi_\lambda\|_{L^\infty(D)} \lesssim 2^{d|\lambda|/2}$ and $\|\widetilde{\psi}_\lambda\|_{L^\infty(D)} \lesssim 2^{d|\lambda|/2}$ for every $\lambda \in \triangledown$,
3. (compact support) $|\mathrm{supp}(\psi_\lambda)| = \mathscr{O}(2^{-|\lambda|d})$ and $|\mathrm{supp}(\widetilde{\psi}_\lambda)| = \mathscr{O}(2^{-|\lambda|d})$ for every $\lambda \in \triangledown$,
4. (vanishing moments of order k) $\int_D x^{\boldsymbol{\alpha}} \psi_\lambda \mathrm{d}x = 0$ and $\int_D x^{\boldsymbol{\alpha}} \widetilde{\psi}_\lambda \mathrm{d}x = 0$ for all multi-indices $\boldsymbol{\alpha} \in \mathbb{N}_0^2$ such that $|\boldsymbol{\alpha}| \le k$ and for every $\lambda \in \triangledown$.

We also suppose that $(\psi_\lambda)_{\lambda\in\triangledown}$ satisfies the finite overlap property in (18). Denoting the $L^2(D)$ inner product by $(\cdot,\cdot)_{L^2(D)}$, for $L \in \mathbb{N}_0$ and $\beta \in [0,1)$, define the index sets

$$\Lambda_{L,\beta} := \left\{ \lambda \in \triangledown : r_D^{\beta}(x_\lambda) \le 2^{L-|\lambda|} \right\},$$

where x_λ is the barycenter of $\mathrm{supp}(\psi_\lambda)$, $\lambda \in \triangledown$. Every function $w \in L^2(D)$ can be represented as $u = \sum_{\lambda\in\triangledown} (w, \widetilde{\psi}_\lambda)_{L^2(D)} \psi_\lambda$ with equality in $L^2(D)$. With the index set $\Lambda_{L,\beta}$, we define the interpolant $P_{L,\beta}$ by

$$P_{L,\beta} w := \sum_{\lambda\in\Lambda_{L,\beta}} (w, \widetilde{\psi}_\lambda)_{L^2(D)} \psi_\lambda. \tag{37}$$

Proposition 6. *For $m \in \mathbb{N}$, suppose $m > k$ and $0 < \delta < \beta < 1$ satisfy $(1-\delta)/(1-\beta) > m$. There exists a constant $C > 0$ such that for every $w \in W^{m,\infty}_{\delta}(D)$*

$$\|w - P_{L,\beta} w\|_{L^\infty(D)} \le C 2^{-\min\{k+1,m\}L} \|w\|_{W^{m,\infty}_{\delta}(D)}.$$

Proof. Without loss of generality we assume that $k+1=m$. We distinguish the cases that $\inf_{x\in\mathrm{supp}(\widetilde{\psi}_\lambda)} r_D(x)=0$ and $\inf_{x\in\mathrm{supp}(\widetilde{\psi}_\lambda)} r_D(x)>0$. In the latter case $w\in W^{m,\infty}(\mathrm{supp}(\widetilde{\psi}_\lambda))$. The Taylor sum $\sum_{|\boldsymbol{\alpha}|\le k} w_{\boldsymbol{\alpha}} x^{\boldsymbol{\alpha}}$ of w in $\mathrm{supp}(\widetilde{\psi}_\lambda)$ satisfies that there exists a constant $C>0$ independent of w such that for every $\lambda\in\triangledown$,

$$\operatorname{ess\,sup}_{x\in\mathrm{supp}(\psi_\lambda)}\left|w(x)-\sum_{|\boldsymbol{\alpha}|\le k} w_{\boldsymbol{\alpha}} x^{\boldsymbol{\alpha}}\right| \le C|\mathrm{diam}(\mathrm{supp}(\psi_\lambda))|^{k+1}\|w\|_{W^{k+1,\infty}(\mathrm{supp}(\psi_\lambda))}. \tag{38}$$

By the *vanishing moments property*, the $L^\infty(D)$ bounds and the support property of $\widetilde{\psi}_\lambda$, the Cauchy–Schwarz inequality, and (38),

$$|(w,\widetilde{\psi}_\lambda)_{L^2(D)}|\le C2^{-(k+1)|\lambda|}2^{-d|\lambda|/2}\|w\|_{W^{k+1,\infty}(\mathrm{supp}(\widetilde{\psi}_\lambda))}. \tag{39}$$

This estimate is suitable if $\inf_{x\in\mathrm{supp}(\psi_\lambda)} r_D(x)>0$. If λ is such that $\inf_{x\in\mathrm{supp}(\psi_\lambda)} r_D(x)=0$, which essentially implies that $\mathrm{supp}(\psi_\lambda)$ abuts at a corner of D, by the estimate in (31), there exists a constant $C>0$ (independent of w and λ) such that

$$\operatorname{ess\,sup}_{x\in\mathrm{supp}(\widetilde{\psi}_\lambda)}\{r_D^\delta(x)|w(x)-w(x_\lambda)|\}\le C2^{-|\lambda|}\|r_D^\delta|\nabla w|\|_{L^\infty(\mathrm{supp}(\widetilde{\psi}_\lambda))}.$$

Thus,

$$\begin{aligned}|(w,\widetilde{\psi}_\lambda)_{L^2(D)}| &= |(w-w(x_\lambda),\widetilde{\psi}_\lambda)_{L^2(D)}|\\ &\le C2^{-|\lambda|}\|r_D^{-\delta}\|_{L^2(\mathrm{supp}(\widetilde{\psi}_\lambda))}\|r_D^\delta|\nabla w|\|_{L^\infty(\mathrm{supp}(\widetilde{\psi}_\lambda))}\|\widetilde{\psi}_\lambda\|_{L^2(D)}.\end{aligned} \tag{40}$$

We note that $\|r_D^{-\delta}\|_{L^2(\mathrm{supp}(\widetilde{\psi}_\lambda))}\le C2^{-|\lambda|(d/2-\delta)}$ for a constant $C>0$ independent of λ. For $\lambda\in\triangledown\backslash\Lambda_{L,\beta}$ and $\mathrm{supp}(\widetilde{\psi}_\lambda)\cap\mathscr{C}\neq\emptyset$, $r_D(x_\lambda)^\beta>2^{L-|\lambda|}$ and $r_D(x_\lambda)^\beta\le C2^{-|\lambda|\beta}$ for a constant independent of λ. Since $(1-\delta)/(1-\beta)>k+1$,

$$2^{-|\lambda|}\|r_D^{-\delta}\|_{L^2(\mathrm{supp}(\widetilde{\psi}_\lambda))}\le C2^{-d|\lambda|/2}2^{-L(k+1)}. \tag{41}$$

For $\lambda\in\triangledown\backslash\Lambda_{L,\beta}$ and $\mathrm{supp}(\widetilde{\psi}_\lambda)\cap\mathscr{C}=\emptyset$, $(1-\delta)/(1-\beta)>k+1$ implies that

$$2^{-|\lambda|(k+1)}\le C2^{-L(k+1)}r_D^{\delta+k}(x_\lambda). \tag{42}$$

Let $\widetilde{\Lambda}\subset\triangledown\backslash\Lambda_{L,\beta}$ be an index set such that $\overline{D}\subset\bigcup_{\lambda\in\widetilde{\Lambda}}\mathrm{supp}(\psi_\lambda)$ and for every $\lambda,\lambda'\in\widetilde{\Lambda}$, $\mathrm{supp}(\psi_\lambda)\not\subset\mathrm{supp}(\psi'_\lambda)$. For $\lambda'\in\widetilde{\Lambda}$ such that $\mathrm{supp}(\psi_\lambda)\cap\mathscr{C}=\emptyset$, by (39), the bounded support overlap property (18) of $(\psi_\lambda)_{\lambda\in\triangledown}$, and (42) there exist constants $C,C'>0$ such that

$$\begin{aligned}\|w-P_{L,\beta}w\|_{L^\infty(\mathrm{supp}(\psi_{\lambda'}))} &\le C\sum_{\ell\ge|\lambda'|}2^{-(k+1)\ell}\|w\|_{W^{k+1,\infty}(\mathrm{supp}(\widetilde{\psi}_{\lambda'}))}\\ &\le C2^{-(k+1)|\lambda'|}\sum_{\ell\ge0}2^{-\ell}\|w\|_{W^{k+1,\infty}(\mathrm{supp}(\widetilde{\psi}_{\lambda'}))}\\ &\le C'2^{-(k+1)L}\|w\|_{W_\delta^{k+1,\infty}(\mathrm{supp}(\widetilde{\psi}_{\lambda'}))}.\end{aligned}$$

Similarly, for $\lambda' \in \widetilde{\Lambda}$ such that $\mathrm{supp}(\psi_\lambda) \cap \mathscr{C} \neq \emptyset$, by (40), the bounded support overlap property of $(\psi_\lambda)_{\lambda \in \nabla}$, and (41) there exists constants $C, C' > 0$ such that

$$\begin{aligned} &\|w - P_{L,\beta} w\|_{L^\infty(\mathrm{supp}(\psi_{\lambda'}))} \\ &\quad \leq C \sum_{\ell \geq |\lambda'|} 2^{-\ell} \sum_{\lambda \in \nabla \setminus \Lambda_{L,\beta} : |\lambda| = \ell} \|r_D^{-\delta}\|_{L^2(\mathrm{supp}(\widetilde{\psi}_\lambda))} 2^{d\ell/2} \|w\|_{W^{1,\infty}_\delta(\mathrm{supp}(\widetilde{\psi}_{\lambda'}))} \\ &\quad \leq C 2^{-(k+1)L} \sum_{\ell \geq 0} 2^{-\ell} \|w\|_{W^{1,\infty}(\mathrm{supp}_\delta(\widetilde{\psi}_{\lambda'}))}. \end{aligned}$$

Since $\overline{D} \subset \bigcup_{\lambda \in \widetilde{\Lambda}} \mathrm{supp}(\psi_\lambda)$, the proof of the proposition is complete. □

The following lemma may be shown as [33, Equations (5) and (13)].

Lemma 5. *For every $L \in \mathbb{N}$ and $\beta \in [0,1)$, $|\Lambda_{L,\beta}| = \mathscr{O}(2^{dL})$.*

Proof.[of Prop. 1] We denote $w = (a_{ij}, b_j, c) \in X_{m,\delta}$ for some $m \geq 1$. We suppose that the biorthogonal wavelets $(\psi_\lambda)_\nabla$ have vanishing moments of order $m - 1 = k \geq 0$. The statement of the theorem follows applying Proposition 6 to a_{ij}, $r_D b_j$, and to $r_D^2 c$ together with Lemma 5. □

Acknowledgements This work was supported in part by the Swiss National Science Foundation (SNSF) under grant SNF 159940.

References

1. Adler, J.H., Nistor, V.: Graded mesh approximation in weighted Sobolev spaces and elliptic equations in 2D. Math. Comp. **84**(295), 2191–2220 (2015)
2. Babuška, I., Kellogg, R.B., Pitkäranta, J.: Direct and inverse error estimates for finite elements with mesh refinements. Numer. Math. **33**(4), 447–471 (1979)
3. Babuška, I., Banerjee, U., Li, H.: The effect of numerical integration on the finite element approximation of linear functionals. Numer. Math. **117**(1), 65–88 (2011)
4. Băcuţă, C., Li, H., Nistor, V.: Differential operators on domains with conical points: precise uniform regularity estimates. Rev. Roumaine Math. Pures Appl. **62**(3), 383–411 (2017)
5. Bui-Thanh, T., Girolami, M.: Solving large-scale PDE-constrained Bayesian inverse problems with Riemann manifold Hamiltonian Monte Carlo. Inverse Problems **30**(11), 114,014, 23 (2014)
6. Chen, L., Nochetto, R.H., Xu, J.: Optimal multilevel methods for graded bisection grids. Numer. Math. **120**(1), 1–34 (2012)
7. Chen, P., Villa, U., Ghattas, O.: Hessian-based adaptive sparse quadrature for infinite-dimensional Bayesian inverse problems. Comput. Methods Appl. Mech. Engrg. **327**, 147–172 (2017)
8. Chkifa, A., Cohen, A., Schwab, C.: Breaking the curse of dimensionality in sparse polynomial approximation of parametric PDEs. J. Math. Pures Appl. (9) **103**(2), 400–428 (2015)
9. Ciarlet, P.G.: The Finite Element Method for Elliptic Problems. North-Holland Publishing Co., Amsterdam-New York-Oxford (1978). Studies in Mathematics and its Applications, Vol. 4
10. Cohen, A., DeVore, R., Schwab, Ch.: Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDE's. Anal. Appl. (Singap.) **9**(1), 11–47 (2011)

11. Constantine, G.M., Savits, T.H.: A multivariate Faà di Bruno formula with applications. Trans. Amer. Math. Soc. **348**(2), 503–520 (1996)
12. Dashti, M., Stuart, A.M.: The Bayesian approach to inverse problems. pp. 1–118. Springer International Publishing (2015)
13. Davydov, O., Stevenson, R.: Hierarchical Riesz bases for $H^s(\Omega), 1 < s < \frac{5}{2}$. Constr. Approx. **22**(3), 365–394 (2005)
14. Dick, J., Gantner, R.N., Gia, Q.T.L., Schwab, C.: Higher order quasi-Monte Carlo integration for bayesian PDE inversion. Computers & Mathematics with Applications **77**(1), 144–172 (2019)
15. Dick, J., Gantner, R.N., Le Gia, Q.T., Schwab, C.: Multilevel higher-order quasi-Monte Carlo Bayesian estimation. Math. Models Methods Appl. Sci. **27**(5), 953–995 (2017)
16. Dick, J., Kuo, F.Y., Le Gia, Q.T., Nuyens, D., Schwab, Ch.: Higher order QMC Petrov-Galerkin discretization for affine parametric operator equations with random field inputs. SIAM J. Numer. Anal. **52**(6), 2676–2702 (2014)
17. Dick, J., Kuo, F.Y., Le Gia, Q.T., Schwab, Ch.: Multilevel higher order QMC Petrov-Galerkin discretization for affine parametric operator equations. SIAM J. Numer. Anal. **54**(4), 2541–2568 (2016)
18. Dick, J., Kuo, F.Y., Sloan, I.H.: High-dimensional integration: the quasi-Monte Carlo way. Acta Numer. **22**, 133–288 (2013)
19. Dick, J., Le Gia, Q.T., Schwab, C.: Higher order quasi-Monte Carlo integration for holomorphic, parametric operator equations. SIAM/ASA J. Uncertain. Quantif. **4**(1), 48–79 (2016)
20. Dodwell, T.J., Ketelsen, C., Scheichl, R., Teckentrup, A.L.: A hierarchical multilevel Markov chain Monte Carlo algorithm with applications to uncertainty quantification in subsurface flow. SIAM/ASA J. Uncertain. Quantif. **3**(1), 1075–1108 (2015)
21. Gantner, R.N., Herrmann, L., Schwab, C.: Multilevel QMC with product weights for affine-parametric, elliptic PDEs. In: J. Dick, F.Y. Kuo, H. Woźniakowski (eds.) Comtemporary Computational Mathematics - a celebration of the 80th birthday of Ian Sloan, pp. 373–405. Springer-Verlag, Cham (2018)
22. Gantner, R.N., Herrmann, L., Schwab, C.: Quasi–Monte Carlo integration for affine-parametric, elliptic PDEs: local supports and product weights. SIAM J. Numer. Anal. **56**(1), 111–135 (2018)
23. Gaspoz, F.D., Morin, P.: Convergence rates for adaptive finite elements. IMA J. Numer. Anal. **29**(4), 917–936 (2009)
24. Giles, M.B.: Multilevel Monte Carlo methods. Acta Numer. **24**, 259–328 (2015)
25. Herrmann, L.: Ph.D. thesis. In preparation
26. Herrmann, L., Schwab, C.: Multilevel quasi-Monte Carlo integration with product weights for elliptic PDEs with lognormal coefficients. Tech. Rep. 2017-19, Seminar for Applied Mathematics, ETH Zürich, Switzerland (2017). URL https://www.sam.math.ethz.ch/sam_reports/reports_final/reports2017/2017-19_rev2.pdf
27. Hoang, V.H., Schwab, C., Stuart, A.M.: Complexity analysis of accelerated MCMC methods for Bayesian inversion. Inverse Problems **29**(8), 085,010, 37 (2013)
28. Houston, P., Schwab, C., Süli, E.: Discontinuous hp-finite element methods for advection-diffusion-reaction problems. SIAM J. Numer. Anal. **39**(6), 2133–2163 (2002)
29. Kuo, F.Y., Scheichl, R., Schwab, C., Sloan, I.H., Ullmann, E.: Multilevel quasi-Monte Carlo methods for lognormal diffusion problems. Math. Comp. **86**(308), 2827–2860 (2017)
30. Kuo, F.Y., Schwab, Ch., Sloan, I.H.: Quasi-Monte Carlo methods for high-dimensional integration: the standard (weighted Hilbert space) setting and beyond. ANZIAM J. **53**(1), 1–37 (2011)
31. Kuo, F.Y., Schwab, Ch., Sloan, I.H.: Multi-level quasi-Monte Carlo finite element methods for a class of elliptic PDEs with random coefficients. Found. Comput. Math. **15**(2), 411–449 (2015)
32. Müller, F., Schötzau, D., Schwab, C.: Symmetric interior penalty discontinuous galerkin methods for elliptic problems in polygons. SIAM J. Numer. Anal. **55**(5), 2490–2521 (2017)
33. Nitsche, P.A.: Sparse approximation of singularity functions. Constr. Approx. **21**(1), 63–81 (2005)

34. Nochetto, R.H., Veeser, A.: Primer of adaptive finite element methods. In: Multiscale and adaptivity: modeling, numerics and applications, *Lecture Notes in Math.*, vol. 2040, pp. 125–225. Springer, Heidelberg (2012)
35. Nuyens, D., Cools, R.: Fast algorithms for component-by-component construction of rank-1 lattice rules in shift-invariant reproducing kernel Hilbert spaces. Math. Comp. **75**(254), 903–920 (electronic) (2006)
36. Nuyens, D., Cools, R.: Fast component-by-component construction of rank-1 lattice rules with a non-prime number of points. J. Complexity **22**(1), 4–28 (2006)
37. Opic, B., Kufner, A.: Hardy-type inequalities, *Pitman Research Notes in Mathematics Series*, vol. 219. Longman Scientific Technical, Harlow (1990)
38. Rekatsinas, N., Stevenson, R.: A quadratic finite element wavelet Riesz basis. Int. J. Wavelets Multiresolut. Inf. Process. **16**(4), 1850,033, 17 (2018)
39. Scheichl, R., Stuart, A.M., Teckentrup, A.L.: Quasi-Monte Carlo and multilevel Monte Carlo methods for computing posterior expectations in elliptic inverse problems. SIAM/ASA J. Uncertain. Quantif. **5**(1), 493–518 (2017)
40. Schillings, C., Schwab, C.: Sparse, adaptive Smolyak quadratures for Bayesian inverse problems. Inverse Problems **29**(6), 065,011, 28 (2013)
41. Schillings, C., Schwab, C.: Sparsity in Bayesian inversion of parametric operator equations. Inverse Problems **30**(6), 065,007, 30 (2014)
42. Schneider, R.: Optimal convergence rates of adaptive algorithms for finite element multiscale methods. In: Boundary value problems and integral equations in nonsmooth domains (Luminy, 1993), *Lecture Notes in Pure and Appl. Math.*, vol. 167, pp. 269–284. Dekker, New York (1995)
43. Schwab, C.: *p*- and *hp*-Finite Element Methods. Numerical Mathematics and Scientific Computation. The Clarendon Press, Oxford University Press, New York (1998). Theory and Applications in Solid and Fluid Mechanics
44. Stevenson, R.: Stable three-point wavelet bases on general meshes. Numer. Math. **80**(1), 131–158 (1998)
45. Stevenson, R.: Locally supported, piecewise polynomial biorthogonal wavelets on nonuniform meshes. Constr. Approx. **19**(4), 477–508 (2003)
46. Yoshiki, T.: Bounds on Walsh coefficients by dyadic difference and a new Koksma–Hlawka type inequality for Quasi-Monte Carlo integration. Hiroshima Math. J. **47**(2), 155–179 (2017)