M. Griebel, H. Harbrecht and M. Multerer

Multilevel quadrature for elliptic parametric partial differential equations in case of polygonal approximations of curved domains

INS Preprint No. 1521

January 2019
Abstract. Multilevel quadrature methods for parametric operator equations such as the multilevel (quasi-) Monte Carlo method are closely related to the sparse tensor product approximation between the spatial variable and the parameter. In this article, we employ this fact and reverse the multilevel quadrature method via the sparse grid construction by applying differences of quadrature rules to finite element discretizations of increasing resolution. Besides being algorithmically more efficient if the underlying quadrature rules are nested, this way of performing the sparse tensor product approximation enables the easy use of non-nested and even adaptively refined finite element meshes. Especially, we present a rigorous error and regularity analysis of the fully discrete solution, taking into account the effect of polygonal approximations to a curved physical domain and the numerical approximation of the bilinear form. Our results facilitate the construction of efficient multilevel quadrature methods based on deterministic quadrature rules. Numerical results in three spatial dimensions are provided to illustrate the approach.

Key words. Parametric partial differential equations, multilevel quadrature, convergence

AMS subject classifications. 65N30, 65D32, 60H15, 60H35

1. Introduction. The present article is concerned with the numerical solution of elliptic parametric second order boundary value problems of the form

\[(1.1) \quad - \text{div} \left( a(y) \nabla u(y) \right) = f(y) \text{ in } D, \quad u(y) = 0 \text{ on } \partial D, \quad y \in \Gamma,\]

where \(D \subset \mathbb{R}^d\) denotes the spatial domain and \(\Gamma \subset \mathbb{R}^m\) denotes the parameter domain. Prominent representatives of such problems arise from recasting boundary value problems with random data, like random diffusion coefficients, random right hand sides and even random domains. A high-dimensional parametric boundary value problem of the form (1.1) is then derived by inserting the truncated Karhunen-Loève expansion of the random data, see e.g. \([1, 2, 12, 24, 33]\). Hence, the computation of quantities of interest amounts to a high-dimensional Bochner integration problem. The latter can be dealt with by quadrature methods. Since every quadrature method requires the repeated evaluation of the integrand for different sample or quadrature points, we have to compute the solution to (1.1) with respect to many different values of the parameter \(y \in \Gamma\).

An efficient approach to deal with the quadrature problem is the multilevel Monte Carlo method (MLMC), which has been developed in \([3, 16, 18, 27, 28]\). As first observed in \([14, 22]\), this approach mimics a certain sparse grid approximation between the physical space and the parameter space. Thus, the extension to the multilevel quasi-Monte Carlo (MLQMC) method and even more general multilevel quadrature methods is obvious. In this article, we focus on such deterministic quadrature methods, which, in particular, require extra regularity of the solution in terms of spaces.
of dominant mixed derivatives, cf. [22, 25, 31] for example. This extra regularity is available for important classes of parametric problems, see [9, 10] for the case of affine elliptic diffusion coefficients and [30] for the case of log-normally distributed diffusion coefficients. For the sake of clarity in presentation, we shall focus here on affine elliptic diffusion problems as they occur from the discretization of uniformly elliptic random diffusion coefficients. We put our emphasis on the rigorous error and regularity analysis of the fully discrete solution, taking into account the effect of polygonal approximations to a curved physical domain and the numerical approximation of the bilinear form.

In addition, we focus on a particular construction of the multilevel quadrature, which is very well suited for the use with black-box finite element solvers and adaptive mesh refinements: Taking the fact that a multilevel quadrature scheme resembles a sparse tensor product approximation between the spatial variable and the parametric variable as a starting point, we have the following abstract framework. Let

\[ V_{0}^{(i)} \subset V_{1}^{(i)} \subset \cdots \subset V_{j}^{(i)} \subset \cdots \subset \mathcal{H}_i, \quad i = 1, 2, \]

denote two sequences of finite dimensional sub-spaces with increasing approximation power in some linear spaces \( \mathcal{H}_i \). To approximate a given object of the tensor product space \( \mathcal{H}_1 \otimes \mathcal{H}_2 \), one canonically considers the full tensor product spaces \( U_j := V_{0}^{(1)} \otimes V_{j}^{(2)} \). However, the cost \( \dim U_j = \dim V_{j}^{(1)} \cdot \dim V_{j}^{(2)} \) is often too expensive. To reduce this cost, one might consider the approximation in so-called sparse grid spaces, see e.g. [7]. For \( \ell \geq 0 \), one introduces the complement spaces

\[ W_{\ell+1}^{(i)} = V_{\ell+1}^{(i)} \ominus V_{\ell}^{(i)}, \quad i = 1, 2, \]

which gives rise to the multilevel decompositions

\[ V_{j}^{(i)} = \bigoplus_{\ell=0}^{j} W_{\ell}^{(i)}, \quad W_{0}^{(i)} := V_{0}^{(i)}, \quad i = 1, 2. \]

Then, the sparse grid space is defined by

\[ \hat{U}_j := \bigoplus_{\ell + \ell' \leq j} W_{\ell}^{(1)} \otimes W_{\ell'}^{(2)}. \]

Under the assumptions that the dimensions of \( \{ V_{\ell}^{(1)} \} \) and \( \{ V_{\ell}^{(2)} \} \) form geometric series, (1.3) contains, at most up to a logarithm, only \( \mathcal{O}(\max \{ \dim V_{\ell}^{(1)}, \dim V_{\ell}^{(2)} \}) \) degrees of freedom. Nevertheless, it offers nearly the same approximation power as \( U_j \) provided that the object to be approximated has some extra smoothness by means of mixed regularity. For further details, see [19].

In view of (1.2), factoring out with respect to the first component, one can rewrite (1.3) according to

\[ \hat{U}_j = \bigoplus_{\ell=0}^{j} W_{\ell}^{(1)} \otimes \left( \bigoplus_{\ell'=0}^{j-\ell} W_{\ell'}^{(2)} \right) = \bigoplus_{\ell=0}^{j} W_{\ell}^{(1)} \otimes V_{j-\ell}^{(2)}. \]

This representation has already been proposed in [19]. Obviously, in complete analogy there holds

\[ \hat{U}_j = \bigoplus_{\ell'=0}^{j} \left( \bigoplus_{\ell=0}^{j-\ell'} W_{\ell}^{(1)} \right) \otimes W_{\ell'}^{(2)} = \bigoplus_{\ell=0}^{j} V_{j-\ell}^{(1)} \otimes W_{\ell}^{(2)}. \]
We refer to Fig. 1 for an illustration, where the left plot corresponds to the representation (1.4) and the right plot corresponds to the representation (1.5). The advantage of the representation (1.4) is that we can give up the requirement that the spaces \( \{ V^{(2)}_\ell \} \) are nested. Likewise, for the representation (1.5), the spaces \( \{ V^{(1)}_\ell \} \) need not to be nested any more.

In the context of the parametric diffusion problem (1.1), we aim at computing

\[
\int_{\Gamma} F(u(y)) \rho(y) \, dy,
\]

where \( \rho \) is the density of some measure on \( \Gamma \) and \( F \) denotes a functional or, as in the case of moment computation, it may be defined as \( F(u(y)) = u^p(y) \) for \( p = 1, 2, \ldots \). In this context, \( \{ V^{(1)}_\ell \} \) corresponds to a sequence of finite element spaces and \( \{ V^{(2)}_\ell \} \) refers to a sequence of quadrature rules. If we denote the finite element solutions of (1.1) by \( u_\ell(y) \in V^{(1)}_\ell \) and if we denote the sequence of quadrature rules by \( Q_\ell : C(\Gamma) \to \mathbb{R} \), we thus arrive with respect to (1.4) at the decomposition

\[
(1.6) \quad \int_{\Gamma} F(u(y)) \rho(y) \, dy \approx \sum_{\ell = 0}^{j} Q_{j-\ell} \Delta F_\ell(u(y)),
\]

where \( \Delta F_\ell(u(y)) := F(u_\ell(y)) - F(u_{\ell-1}(y)) \) and \( F(u_{-1}(y)) := 0 \), see [22]. On the other hand, similarly to (1.5), we obtain the decomposition

\[
(1.7) \quad \int_{\Gamma} F(u(y)) \rho(y) \, dy \approx \sum_{\ell = 0}^{j} \Delta Q_\ell F(u_{j-\ell}(y)),
\]

where \( \Delta Q_\ell := Q_\ell - Q_{\ell-1} \) and \( Q_{-1} := 0 \). Both representations are equivalent but have a different impact on its numerical implementation.

Originally, multilevel quadrature methods have been interpreted as variance reduction methods for the Monte Carlo quadrature, a view which has originally been introduced for the approximation of parametric integrals, cf. [27, 28]. Consequently, the representation (1.4), and thus the decomposition (1.6), has been used in previous articles, see, for example, [16, 17] for stochastic ordinary differential equations and [3, 22, 38, 39] for partial differential equations with random data. To this end, usually a nested sequence of approximation spaces is presumed such that the complement spaces \( \{ W^{(1)}_\ell \} \) are well-defined. In the context of partial differential equations, these
complement spaces are given via the difference of Galerkin projections onto subsequent finite element spaces. This circumstance can be avoided in the case of $F$ being a functional, cf. [20, 38]. Still, we emphasize that, particularly in the context of the Monte Carlo method, there are already results available, which allow for giving up this nestedness, see e.g. [8, 38]. A more general result addressing the resulting error in the underlying bilinear form can be found in [37].

The decomposition (1.6) is well suited if the spatial dimension is small, as it is the case for one-dimensional partial differential equations with random data or for stochastic ordinary differential equations. Nevertheless, in two or three spatial dimensions, the construction of nested approximation spaces might be difficult or even not be possible at all. Sometimes, in view of adaptive refinement strategies, it might be favourable to give up nestedness. In the article at hand, we employ the decomposition (1.7). It allows more naturally for non-nested finite element spaces which, in turn, induce different approximations of the underlying domain. Moreover, using nested quadrature formulae, a considerable speed-up is achieved in comparison to the conventional multilevel quadrature which is based on the representation (1.6).

The rest of the article is organized as follows. We start by introducing the underlying random model in Section 2 and perform the parametric reformulation that results in (1.1). Then, the next two sections are dedicated to the discretization, namely the quadrature rule for the parametric variable (Section 3) and the finite element discretization for the physical domain (Section 4). The multilevel quadrature for the model problem is discussed in Section 5. In Section 6, we present the error and regularity analysis for the multilevel quadrature taking into account polygonal approximations of curved domains. We emphasize that the key result in this section, namely Lemma 6.1, is robust with respect to the parameter dimension $m$. Afterwards, in Section 7, we consider a fully discrete approximation of the solution to (1.1) and take also quadrature errors in the bilinear form into account. Again, the main result (Theorem 7.2) is robust with respect to the parameter dimension $m$. Finally, in Section 8, we provide numerical results in three spatial dimensions to validate our approach.

Throughout this article, in order to avoid the repeated use of generic but unspecified constants, we mean by $C \lesssim D$ that $C$ can be bounded by a multiple of $D$, independently of parameters which $C$ and $D$ may depend on. Obviously, $C \gtrsim D$ is defined as $D \lesssim C$, and $C \sim D$ as $C \lesssim D$ and $C \gtrsim D$.

2. Problem setting. Let $(\Omega, \Sigma, P)$ be a complete and separable probability space with $\sigma$-field $\Sigma \subset 2^\Omega$ and probability measure $P$. We intend to compute the expectation

$$E[u] = \int_\Omega u(\omega) \, dP(\omega) \in H^1_0(D)$$

and the variance

$$\mathbb{V}[u] = \int_\Omega \{u(\omega) - E[u]\}^2 \, dP(\omega) \in W^{1,1}_0(D)$$

of the random function $u(\omega) \in H^1_0(D)$ which solves the stochastic diffusion problem

$$- \text{div}(a(\omega)\nabla u(\omega)) = f \quad \text{in } D \quad \text{for almost every } \omega \in \Omega.$$  

For sake of simplicity, we assume that the stochastic diffusion coefficient is given
by a finite Karhunen-Loève expansion

\begin{equation}
(2.2) \quad a(x, \omega) = \mathbb{E}[a](x) + \sum_{k=1}^{m} \sqrt{\lambda_{k}} \varphi_{k}(x) \psi_{k}(\omega)
\end{equation}

with pairwise $L^2$-orthonormal functions $\varphi_{k} \in L^\infty(D)$ and stochastically independent random variables $\psi_{k}(\omega) \in [-1, 1]$. Especially, it is assumed that the random variables admit continuous density functions $\rho_{k} : [-1, 1] \to \mathbb{R}$ with respect to the Lebesgue measure.

In practice, one generally has to compute the expansion (2.2) from the given covariance kernel

\[
\text{Cov}[a](x, x') = \int_{\Omega} \{a(x, \omega) - \mathbb{E}[a](x)\} \{a(x', \omega) - \mathbb{E}[a](x')\} \, dP(\omega).
\]

If the expansion contains infinitely many terms, it has to be appropriately truncated which will induce an additional discretization error. For details, we refer the reader to [15, 23, 32, 36].

The assumption that the random variables $\psi_{k}(\omega)$ are independent implies that the joint density function of the random variables is given by $\rho(y) := \prod_{k=1}^{m} \rho_{k}(y_{k})$.

Thus, we are able to reformulate the stochastic problem (2.1) as a parametric, deterministic problem in $L^2_{\rho}(\Gamma)$. To this end, the probability space $\Omega$ is identified with its image $\Gamma := [-1, 1]^{m}$ with respect to the measurable mapping

\[
\psi : \Omega \to \Gamma, \quad \omega \mapsto \psi(\omega) := (\psi_1(\omega), \ldots, \psi_m(\omega)).
\]

Hence, the random variables $\psi_{k}$ are substituted by coordinates $y_{k} \in [-1, 1]$.

We introduce the measure $\rho(y) \, dy$ on $\Gamma$, which is defined by the product density function $\rho(y) := \prod_{k=1}^{m} \rho_{k}(y_{k})$.

Next, in order to ensure $H^2$-regularity of the model problem, let $D \subset \mathbb{R}^{d}$, $d = 2, 3$, be either a convex, polygonal domain. We consider the parametric diffusion problem

\begin{equation}
\text{find } u \in L^2_{\rho}(\Gamma; H^1_{0}(D)) \text{ such that }
\begin{array}{l}
- \text{div} \left( a(y) \nabla u(y) \right) = f \text{ in } D \text{ for almost every } y \in \Gamma,
\end{array}
\end{equation}

with $f \in L^2(D)$ and $a : D \times \Gamma \to \mathbb{R}$ with

\begin{equation}
(2.4) \quad a(x, y) = \varphi_{0}(x) + \sum_{k=1}^{m} \sqrt{\lambda_{k}} \varphi_{k}(x) y_{k}, \quad \gamma_{k} := \sqrt{\lambda_{k}} \| \varphi_{k} \|_{W^{1, \infty}(D)}.
\end{equation}

Note that $u \in L^2_{\rho}(\Gamma; H^1_{0}(D))$ guarantees finite second order moments of the solution.

By the Lax-Milgram theorem, the unique solvability of the parametric diffusion problem (2.3) in $L^2_{\rho}(\Gamma; H^1_{0}(D))$ follows immediately if we impose the condition

\begin{equation}
(2.5) \quad 0 < a_{\text{min}} \leq a(y) \leq a_{\text{max}} < \infty \text{ in } D
\end{equation}

for all $y \in \Gamma$ on the diffusion coefficient. Moreover, we obtain the stability estimate

\[
\| u(y) \|_{H^1(D)} \leq \frac{1}{a_{\text{min}}} \| f \|_{H^{-1}(D)} \lesssim \frac{1}{a_{\text{min}}} \| f \|_{L^2(D)} \quad \text{for almost every } y \in \Gamma.
\]

Hence, the solution to (2.3) is essentially bounded with respect to $y \in \Gamma$. 
In e.g. [4, 9, 10, 11, 40], it has been proven that the solution \( u \) of (2.3) is analytical as mapping \( u : \Gamma \to H^1_0(D) \). Moreover, it has been shown in [9] that \( u \) is even an analytical mapping \( u : \Gamma \to W := H^1_0(D) \cap H^2(D) \) given that the \( \{ \varphi_k \} \) in (2.4) belong to \( W^{1,\infty}(D) \). This constitutes the necessary mixed regularity for a sparse tensor product discretization, see e.g. [25, 31]. A similar result for diffusion problems with coefficients of the form \( \exp(a(x, y)) \) has been shown in [30].

Since \( u \) is supposed to be in \( L^2_\rho(\Gamma; H^1_0(D)) \), we can compute its expectation

\[
\mathbb{E}[u] = \int_\Gamma u(y) \rho(y) \, dy \in H^1_0(D)
\]

and its variance

\[
\mathbb{V}[u] = \mathbb{E}[u^2] - \mathbb{E}[u]^2 = \int_\Gamma u^2(y) \rho(y) \, dy - \mathbb{E}[u]^2 \in W^{1,1}_0(D).
\]

We will focus in the sequel on the efficient numerical computation of these possibly high-dimensional integrals.

3. Quadrature in the parameter space. The expectation and the variance of the solution \( u \) to (2.3) are given by the integrals (2.6) and (2.7). To compute these integrals, we employ a sequence of quadrature formulae \( \{ Q_\ell \} \) for the Bochner integral

\[
\text{Int} : L^1_\rho(\Gamma; \mathcal{X}) \to \mathcal{X}, \quad \text{Int} v = \int_\Gamma v(y) \rho(y) \, dy
\]

where \( \mathcal{X} \subset L^2(\Gamma; D) \) denotes some Banach space. The quadrature formula

\[
Q_\ell : L^1_\rho(\Gamma; \mathcal{X}) \to \mathcal{X}, \quad (Q_\ell v)(x) = \sum_{i=1}^{N_\ell} \omega_{\ell,i} v(x, \xi_{\ell,i}) \rho(\xi_{\ell,i})
\]

is supposed to provide the error bound

\[
\| (\text{Int} - Q_\ell) v \|_{\mathcal{X}} \lesssim \varepsilon_\ell \| v \|_{\mathcal{H}(\Gamma; \mathcal{X})}
\]

uniformly in \( \ell \in \mathbb{N} \), where \( \mathcal{H}(\Gamma; \mathcal{X}) \subset L^2_\rho(\Gamma; \mathcal{X}) \) is a suitable Bochner space. Note that since the density \( \rho \) is fixed, it will be suppressed in the upcoming error estimates and will, thus, be hidden in the constants.

The following particular examples of quadrature rules (3.1) are considered in our numerical experiments:

- The Monte Carlo method satisfies (3.2) only with respect to the root mean square error. Namely, it holds

\[
\sqrt{\mathbb{E}(\| (\text{Int} - Q_\ell) v \|_{\mathcal{X}}^2)} \lesssim \varepsilon_\ell \| v \|_{\mathcal{H}(\Gamma; \mathcal{X})}
\]

with \( \varepsilon_\ell = N_{\ell,1}^{-1/2} \) and \( \mathcal{H}(\Gamma; \mathcal{X}) = L^2_\rho(\Gamma; \mathcal{X}) \).

- The quasi-Monte Carlo method leads typically to \( \varepsilon_\ell = N_{\ell,1}^{-1}(\log N_{\ell,1})^m \), where it is sufficient to consider the Bochner space \( \mathcal{H}(\Gamma; \mathcal{X}) = W^{1,1}_{m, \text{mix}}(\Gamma; \mathcal{X}) \) of all equivalence classes of functions \( v : \Gamma \to \mathcal{X} \) with finite norm

\[
\| v \|_{W^{1,1}_{m, \text{mix}}(\Gamma; \mathcal{X})} := \sum_{\| q \|_{\infty} \leq 1} \int_\Gamma \left\| \frac{\partial |q|_1}{\partial y_1} \partial y_2 \cdots \partial y_m v(y) \right\|_{\mathcal{X}} \, dy < \infty,
\]

where
see e.g. [34]. Note that, in this case, the estimate requires that the densities satisfy \( \rho_k \in W^{1,\infty}(-1,1) \). For the Halton sequence, cf. [21], it can even be shown that \( \varepsilon_\ell = N_\ell^{\delta-1} \) for arbitrary \( \delta > 0 \) given that the spatial functions in (2.4) satisfy \( \gamma_k \lesssim k^{-3-\varepsilon} \) for arbitrary \( \varepsilon > 0 \). This is a straightforward consequence from the results in [41], see e.g. [24].

- Let the densities \( \rho_k \) be in \( W^{r,\infty}(-1,1) \). If \( v : \Gamma \rightarrow \mathcal{X} \) has mixed regularity of order \( r \) with respect to the parameter \( y \), i.e.

\[
\|v\|_{W^{r,\infty}(\Gamma;\mathcal{X})} := \max_{\|\alpha\|_\infty \leq r} \left\| \partial_\alpha^\alpha v \right\|_{L^\infty(\Gamma;\mathcal{X})} < \infty,
\]

then one can apply a (sparse) tensor product Clenshaw-Curtis quadrature rule. This yields the convergence rate \( \varepsilon_\ell = 2^{-r \ell^m-1} \), where \( N_\ell \sim 2^{\ell^m-1} \) and \( \mathcal{H}(\Gamma;\mathcal{X}) = W^{r,\infty}(\Gamma;\mathcal{X}) \), see [35].

For our purposes, we shall assume that the number \( N_\ell \) of points of the quadrature formula \( Q_\ell \) is chosen such that the corresponding accuracy is given by

\[
\varepsilon_\ell = 2^{-\ell}.
\]

Then, for the respective difference quadrature \( \Delta Q_\ell := Q_\ell - Q_{\ell-1} \), we immediately obtain by combining (3.2) and (3.5) the error bound

\[
\|\Delta Q_\ell v\|_\mathcal{X} = \|(Q_\ell - Q_{\ell-1})v\|_\mathcal{X} \leq \|(\text{Int } - Q_\ell)v\|_\mathcal{X} + \|(\text{Int } - Q_{\ell-1})v\|_\mathcal{X} \lesssim 2^{-\ell}\|v\|_{\mathcal{H}(\Gamma;\mathcal{X})}.
\]

4. Finite element approximation in the spatial variable. In order to apply the quadrature formula (3.1), we shall calculate the solution \( u(y) \in H_0^1(D) \) of the diffusion problem (2.3) in certain points \( y \in \Gamma \). To this end, consider a not necessarily nested sequence of shape regular and quasi-uniform triangulations or tetrahedralizations \( \{T_\ell\} \) for \( \ell \geq 0 \) of the domain \( D \), respectively, each of which with the mesh size \( h_\ell \sim 2^{-\ell} \). If the domain is not polygonal, then we obtain a polygonal approximation \( D_\ell \) of the domain \( D \) by replacing curved edges and faces by planar ones.

In order to deal only with the fixed domain \( D \) and not with the different polygonal approximations \( D_\ell \), we follow [5] and extend functions defined on \( D_\ell \) by zero onto \( D \setminus D_\ell \). Hence, given the triangulation or the tetrahedralization \( \{T_\ell\} \), we define the spaces

\[
S_\ell(D) := \{v \in C(D) : v|_T \text{ is a linear polynomial for all } T \in T_\ell \text{ and } v(x) = 0 \text{ for all nodes } x \in \partial D\}
\]

of continuous, piecewise linear finite elements. Notice that it does hold \( S_\ell(D) \subset H^1(D) \) but in general \( S_\ell(D) \not\subset H_0^1(D) \).

We shall further introduce the finite element solution \( u_\ell(y) \in S_\ell(D) \) of (2.3) which satisfies

\[
B_y(u_\ell, v_\ell) := \int_D a(x,y)\nabla u_\ell(x,y)\nabla v_\ell(x) \, dx = \int_D f(x)v_\ell(x) \, dx
\]

for all \( v_\ell \in S_\ell(D) \). If \( D \neq D_\ell \), the bilinear form \( B_y(\cdot,\cdot) \) is also well defined for functions from \( S_\ell(D) \) since \( S_\ell(D) \subset H^1(D) \). Nevertheless, in order to maintain the

\[\text{Footnote: The Clenshaw-Curtis quadrature converges exponentially if the integrand } v : \Gamma \rightarrow \mathcal{X} \text{ and the density } \rho \text{ are analytic.}\]
ellipticity of the bilinear form, we shall assume that the mesh size $h_0$ is sufficiently small to ensure that functions in $S_U(D)$ are zero on a part of the boundary of $D$.

It is shown in e.g. [5, 6] that the finite element solution $u_\ell(y) \in S_U(D)$ of (4.1) admits the following approximation properties.

**Lemma 4.1.** Consider a convex, polygonal domain $D$ or a domain with $C^2$-smooth boundary and let $f \in L^2(D)$. Then, the finite element solution $u_\ell(y) \in S_U(D)$ of the diffusion problem (2.3) and respectively its square $u_\ell^2(y)$ satisfy the error estimate

$$
\left\|u^p(y) - u_\ell^p(y)\right\|_{\mathcal{X}} \lesssim h_\ell \|f\|_{L^2(D)},
$$

where $\mathcal{X} = H^1(D)$ for $p = 1$ and $\mathcal{X} = W^{1,1}(D)$ for $p = 2$. The constants hidden in (4.2) depend on $a_{\min}$ and $a_{\max}$, but not on $y \in \Gamma$.

**5. The multilevel quadrature method.** Based on the nomenclature from the previous sections, we now introduce the multilevel quadrature in a formal way. To that end, let $u \in \mathcal{H}(\Gamma; H^2(D))$, where the underlying Bochner space is determined by the quadrature under consideration. For the sequence $\{u_\ell(y)\}_\ell$ of finite element solutions, there obviously holds $\lim_{\ell \to \infty} u_\ell(y) = u(y)$ uniformly in $y \in \Gamma$. Thus, if $\mathcal{F}$ is continuous, we obtain

$$
\lim_{\ell \to \infty} \mathcal{F}(u_\ell(y)) = \mathcal{F}(u(y))
$$

also uniformly in $y \in \Gamma$. Moreover, we have for the sequence $\{Q_\ell\}_\ell$ of quadrature rules and for a sufficiently smooth integrand that

$$
\lim_{\ell \to \infty} Q_\ell v = \int_\Gamma v(y)\rho(y) \, dy.
$$

The combination of the relations (5.1) and (5.2) leads to

$$
\int_\Gamma \mathcal{F}(u(y))\rho(y) \, dy = \sum_{\ell=0}^{\infty} \Delta Q_\ell \mathcal{F}(u(y)) = \sum_{\ell=0}^{\infty} \Delta Q_\ell \sum_{\ell'=0}^{\infty} \Delta \mathcal{F}_{\ell'}(u(y)).
$$

Since $\Delta Q_\ell$ is linear and continuous, we end up with

$$
\int_\Gamma \mathcal{F}(u(y))\rho(y) \, dy = \sum_{\ell,\ell'=0}^{\infty} \Delta Q_\ell \Delta \mathcal{F}_{\ell'}(u(y)).
$$

Truncating this sum in accordance with $\ell + \ell' \leq j$ then yields the multilevel quadrature representation (1.6) if we recombine the operators $\Delta Q_\ell$. Analogously, we obtain the representation (1.7) if we recombine the operators $\Delta \mathcal{F}_{\ell'}$. Note that the sequence of the application of the operators $\Delta Q_\ell$ and $\Delta \mathcal{F}_{\ell'}$ is crucial here. Moreover, we have repeatedly exploited the linearity of $\Delta Q_\ell$.

Of course, the representations (1.6) and (1.7) are mathematically equivalent. More precisely, if we set $\mathcal{F}(u_{-\ell}(y)) := 0$, there holds

$$
\sum_{\ell=0}^{j} Q_{j-\ell} \Delta \mathcal{F}_{\ell}(u(y)) = \sum_{\ell=0}^{j} \Delta Q_\ell \mathcal{F}(u_{j-\ell}(y)).
$$

Thus, all available results for the representation (1.6) of the multilevel quadrature, see e.g. [22, 25] and the references therein, carry over to the representation (1.7).

 Nonetheless, the multilevel quadrature based on representation (1.7) has substantial advantages. On the one hand, it allows for an easy use of non-nested finite element meshes and even for adaptively refined finite element meshes. A further property of (1.7) is an obvious reduction of the cost if nested quadrature formulae are employed.
6. Error analysis. In the sequel, we restrict ourselves for reasons of simplicity to the situations $F(u) = u$ and $F(u) = u^2$ which yield the expectation and the second moment of the solution to (2.3). This means that we consider

$$\text{(6.1)} \quad \text{Int} \, u^p \approx \sum_{\ell=0}^{j} \Delta Q_j u^p_{\ell} = \sum_{\ell=0}^{j} Q_{j-\ell} (u^p_\ell - u^p_{\ell-1}) \quad \text{for } p = 1, 2.$$  

We derive a general approximation result for the multilevel quadrature based on the generic estimate

$$\text{(6.2)} \quad \| (\text{Int} \, Q_j) (u^p - u^p_{\ell}) \|_x \lesssim 2^{-(\ell+\ell')} \| f \|_{L^2(D)}^p \quad \text{for } p = 1, 2$$

with $f$ being the right hand side of (2.3) and $h_\ell \sim 2^{-\ell}$. In particular, any quadrature rule that satisfies this estimate gives rise to a multilevel quadrature method. In the sequel, we provide this estimate for the MLQMC as well as for the multilevel Clenshaw-Curtis quadrature (MLCC).

We remark that the derivation of the generic estimate (6.2) for the Monte Carlo quadrature is straightforward under the condition that the integrand is square integrable with respect to the parameter $y$, cf. [3, 22]. In this case, the generic estimate can be derived similarly to Strang’s lemma, see [38]. Nevertheless, since the Monte Carlo quadrature does not provide deterministic error estimates, we have to replace the norm in $X$ by the $L^p(D; \Gamma)$-norm.

The situation becomes much more involved if parametric regularity has to be taken into account. In the latter case, bounds on the derivatives of the increments have to be provided. The next lemma is a generalization of similar results from [25, 31], which provide the smoothness of the Galerkin error with respect to the parameter $y \in \Gamma$ for the non-conforming case $D \neq D_\ell$.

**Lemma 6.1.** For the error $\delta_\ell(y) := (u - u_\ell)(y)$ of the Galerkin projection, there holds the estimate

$$\text{(6.3)} \quad \| \partial_y^\alpha \delta_\ell(y) \|_{H^{1}(D)} \leq C \ell |\alpha|! \rho^{\frac{|\alpha|}{2}} \| f \|_{L^2(D)} \quad \text{for all } \alpha \in \mathbb{N}^m,$$

where $\gamma := \{\gamma_k\}_{k=1}^m$, cf. (2.4). The constants $C, \rho > 0$ are dependent on $a_{\min}$ and $a_{\max}$, but independent of the parameter dimension $m$.

**Proof.** By definition, there holds, cf. (4.1),

$$\mathcal{E}_y(u, v_\ell) = \int_D a(y) \nabla u(x) \nabla v_\ell \, dx = \int_D f v_\ell \, dx \quad \text{for all } v_\ell \in S_\ell(D).$$

On the other hand, integration by parts yields

$$\mathcal{E}_y(u, v_\ell) = \int_D a(y) \nabla u(y) \nabla v_\ell \, dx = \int f v_\ell \, dx + \int_{\partial D} \frac{\partial u}{\partial n}(y)v_\ell(x) \, d\sigma_x$$

for all $v_\ell \in S_\ell(D)$. Thus, we obtain the perturbed Galerkin orthogonality

$$\text{(6.4)} \quad \mathcal{E}_y(u - u_\ell, v_\ell) = \int_{\partial D} \frac{\partial u}{\partial n}(y)v_\ell(x) \, d\sigma_x \quad \text{for all } v_\ell \in S_\ell(D).$$

Due to the uniform ellipticity of the bilinear form, we can also define the Galerkin projection $P_\ell(y) : H^1_0(D) \rightarrow S_\ell(D)$ via

$$\mathcal{E}_y(u - P_\ell u, v_\ell) = 0 \quad \text{for all } v_\ell \in S_\ell(D).$$
It holds
\begin{equation}
\|\partial_y^\alpha (u - u_\ell)\|_{H^1(D)} \leq \|P_\ell \partial_y^\alpha (u - u_\ell)\|_{H^1(D)} + \|(I - P_\ell) \partial_y^\alpha (u - u_\ell)\|_{H^1(D)} \\
\leq \|P_\ell \partial_y^\alpha (u - u_\ell)\|_{H^1(D)} + \|(I - P_\ell) \partial_y^\alpha u\|_{H^1(D)},
\end{equation}
(6.5)
since \(\partial_y^\alpha u_\ell \in S_\ell(D)\) and hence \(P_\ell \partial_y^\alpha u_\ell = \partial_y^\alpha u_\ell\).

In order to estimate the first term, we employ the perturbed Galerkin orthogonality (6.4) and obtain
\begin{equation}
B_y (\partial_y^\alpha (u - u_\ell), v_\ell) - \partial_y^\alpha \int_{\partial D} a(y) \frac{\partial u}{\partial n} (y) v_\ell \, d\sigma_x \\
= - \sum_{\{k: \alpha_k \neq 0\}} \alpha_k \sqrt{\lambda_k} \int_D \phi_k \nabla \partial_y^{\alpha - e_k} (u - u_\ell)(y) \nabla v_\ell \, dx,
\end{equation}
(6.6)
see e.g. [9]. The derivatives of the boundary term satisfy
\begin{equation}
\partial_y^\alpha \int_{\partial D} a(y) \frac{\partial u}{\partial n} (y) v_\ell \, d\sigma_x = \sum_{\alpha' \leq \alpha} \left(\alpha \atop \alpha'\right) \int_{\partial D} \left[ \partial_y^{\alpha'} a(y) \right] \left[ \partial_y^{\alpha - \alpha'} \frac{\partial u}{\partial n} (y) \right] v_\ell \, d\sigma_x
\end{equation}
\begin{equation}
= \int_{\partial D} a(y) \frac{\partial (\partial_y^\alpha u)}{\partial n} (y) v_\ell \, d\sigma_x + \sum_{\{k: \alpha_k \neq 0\}} \alpha_k \sqrt{\lambda_k} \int_D \phi_k \nabla \partial_y^{\alpha - e_k} u(y) v_\ell \, dx.
\end{equation}

Inserting this identity into (6.6) yields
\begin{equation}
B_y (\partial_y^\alpha (u - u_\ell), v_\ell) - \int_{\partial D} a(y) \frac{\partial (\partial_y^\alpha u)}{\partial n} (y) v_\ell \, d\sigma_x
\end{equation}
(6.7)
\begin{equation}
= - \sum_{\{k: \alpha_k \neq 0\}} \alpha_k \sqrt{\lambda_k} \int_D \phi_k \nabla \partial_y^{\alpha - e_k} (u - u_\ell)(y) \nabla v_\ell \, dx
\end{equation}
\begin{equation}
- \int_{\partial D} \partial_x (\partial_y^{\alpha - e_k} u)(y) v_\ell \, d\sigma_x.
\end{equation}

In order to bound the boundary integrals, we employ the following estimate, which holds true for any \(v, w \in H^1(D)\). It holds
\begin{equation}
\left| \int_{\partial D} a(y) \frac{\partial v}{\partial n} w \, d\sigma_x \right| \leq a_{\text{max}} \left\| \frac{\partial v}{\partial n} (y) \right\|_{H^{-1/2}(\partial D)} \left\| w \right\|_{H^{1/2}(\partial D)}
\end{equation}
\begin{equation}
\leq C_{\text{inv}} a_{\text{max}} \left\| v \right\|_{H^1(D)} \left\| w \right\|_{H^{1/2}(\partial D)},
\end{equation}
where \(C_{\text{inv}}\) is the norm of the inverse Neumann trace operator. Next, we employ a discrete version of the trace theorem provided by [5, Lemma III.1.6], which reads
\begin{equation}
\|v_\ell\|_{H^{1/2}(\partial D)} \leq c_D \|v_\ell\|_{H^1(D)} \quad \text{for all } v_\ell \in S_\ell(D)
\end{equation}
(6.8)
with some constant \(c > 0\). From this, we infer
\begin{equation}
\left| \int_{\partial D} a(y) \frac{\partial (\partial_y^\alpha u)}{\partial n} (y) v_\ell \, d\sigma_x \right| \leq Ch \|\partial_y^\alpha u(y)\|_{H^1(D)} \|v_\ell\|_{H^1(D)}
\end{equation}
for all \(v_\ell \in S_\ell(D)\) and some constant \(C > 0\).
Inserting the latter estimate into (6.7) and choosing $v_\ell = \mathcal{P}_\ell \partial_y^\alpha (u - u_\ell)$ as test function, we arrive at

$$a_{\min} \| \mathcal{P}_\ell \partial_y^\alpha (u - u_\ell)(y) \|_{H^1(D)}^2 \leq \mathcal{C} h_\ell \| \mathcal{P}_\ell \partial_y^\alpha u(y) \|_{H^1(D)} \| \mathcal{P}_\ell \partial_y^\alpha (u - u_\ell)(y) \|_{H^1(D)} + C \sum_{\{k: \alpha_k \neq 0\}} \| \partial_y^\alpha (u - u_\ell)(y) \|_{H^1(D)} + C h_\ell \| \partial_y^\alpha u(y) \|_{H^1(D)}.$$  

Simplifying this expression yields

$$\| \mathcal{P}_\ell \partial_y^\alpha (u - u_\ell)(y) \|_{H^1(D)} \leq \mathcal{C} h_\ell \| \partial_y^\alpha u(y) \|_{H^1(D)} + C \sum_{\{k: \alpha_k \neq 0\}} \| \partial_y^\alpha (u - u_\ell)(y) \|_{H^1(D)}.$$  

for some other constant $C > 0$, where we employed the stability of the Galerkin projection in the first term. Next, in view of the estimate

$$\| \partial_y^\alpha u(y) \|_{H^1(D)} \leq C|\alpha|!c^{\gamma \alpha} \| f \|_{L^2(D)},$$  

for some constants $C, c > 0$, we end up with

$$\| \mathcal{P}_\ell \partial_y^\alpha (u - u_\ell)(y) \|_{H^1(D)} \leq \mathcal{C} h_\ell c^{\gamma \alpha} \| f \|_{L^2(D)} + C \sum_{\{k: \alpha_k \neq 0\}} \| \partial_y^\alpha (u - u_\ell)(y) \|_{H^1(D)},$$  

for some constants $C, c > 0$. Combining this with the initial estimate (6.5) gives then

$$\| \partial_y^\alpha (u - u_\ell)(y) \|_{H^1(D)} \leq C \sum_{\{k: \alpha_k \neq 0\}} \| \partial_y^\alpha (u - u_\ell)(y) \|_{H^1(D)} + \| (I - \mathcal{P}_\ell) \partial_y^\alpha u \|_{H^1(D)}$$  

$$\leq C \sum_{\{k: \alpha_k \neq 0\}} \| \partial_y^\alpha (u - u_\ell)(y) \|_{H^1(D)} + C h_\ell c^{\gamma \alpha} \| f \|_{L^2(D)},$$  

where we used $\| (I - \mathcal{P}_\ell) \partial_y^\alpha u \|_{H^1(D)} \leq \mathcal{C} h_\ell c^{\gamma \alpha} \| f \|_{L^2(D)}$ for some constants $C, c > 0$, which follows from the approximation property of the finite element space $S_I(D)$ and [31, Theorem 6]. The proof is now concluded similarly to the proof of [31, Theorem 7].

With this lemma, it is easy to show the following result related to the second moment, cf. [25].

**Lemma 6.2.** The derivatives of the difference $u^2 - u_\ell^2$ satisfy the estimate

$$\| \partial_y^\alpha (u^2 - u_\ell^2)(y) \|_{W^{1,1}(D)} \leq \mathcal{C} h_\ell |\alpha|!c^{\gamma \alpha} \| f \|_{L^2(D)}^2$$  

for all $\alpha \in \mathbb{N}^m$ with constants $C, c > 0$ dependent on $a_{\min}$ and $a_{\max}$.

With the aid of Lemmata 6.1 and 6.2 together with the results from [41], the generic error estimate for the MLQMC with Halton points can be derived. The next lemma is for example shown in [26, 37].
Lemma 6.3. Let \( u \in L^2_0(\Gamma; H^1_0(D)) \) be the solution to \((2.3)\) and \( u_\ell \) the associated Galerkin projection on level \( \ell \). Moreover, let \( \rho_k \in W^{1,\infty}(-1,1) \) for \( k = 1, \ldots, m \). Then, for the quasi-Monte Carlo quadrature based on Halton points, there holds
\[
\|(\text{Int} - Q_\ell)(u^p - u^p_\ell)\|_X \lesssim 2^{-(\ell + \ell')}\|f\|_{L^2(D)}^p \quad \text{for } p = 1, 2
\]
with \( N_\ell \sim 2^\ell(1 - \delta) \) for arbitrary \( \delta > 0 \).

The next lemma establishes the generic estimate for the sparse grid quadrature based on the nested Clenshaw-Curtis abscissae, cf. [13, 35]. These are given by the extrema of the Chebyshev polynomials
\[
\xi_k = \cos\left(\frac{(k - 1)\pi}{n}\right) \quad \text{for } k = 1, \ldots, n,
\]
where \( n = 2^j - 1 + 1 \) if \( j > 1 \) and \( n = 1 \) with \( \xi_1 = 0 \) if \( j = 1 \).

Lemma 6.4. Let \( u \in L^2_0(\Gamma; H^1_0(D)) \) be the solution to \((2.3)\) and let \( u_\ell \) be the associated Galerkin projection on level \( \ell \). Moreover, let \( \rho_k(y_k) \in W^{r,\infty}(-1,1) \) for \( k = 1, \ldots, m \). Then, for the sparse grid quadrature based on Clenshaw-Curtis abscissae, there holds
\[
\|(\text{Int} - Q_\ell)(u^p - u^p_\ell)\|_X \lesssim 2^{-(\ell r + \ell')m - 1}\|f\|_{L^2(D)}^p \quad \text{for } p = 1, 2
\]
provided that \( N_\ell \sim 2^\ell\ell^d - 1 \).

Proof. It is shown in [35] that the number \( N_\ell \) of quadrature points of the sparse tensor product quadrature with Clenshaw-Curtis abscissae is of the order \( O(2^\ell\ell^d - 1) \). In addition, we have for functions \( v: \Gamma \to \mathbb{R} \) with mixed regularity the following error bound:
\[
\left| \int_{\Gamma} v(y) \, dy - \sum_{i=1}^{N_\ell} w_i v(\xi_i) \right| \lesssim 2^{-\ell r \ell' (m - 1)} \max_{\|\alpha\|_{\infty} \leq r} \|\partial^\alpha_y v\|_{L^\infty(\Gamma)}.
\]
Hence, to prove the desired assertion, we have to provide estimates on the derivatives \( \partial^\alpha_y \left( (u^p(y) - u^p_\ell(y))\rho(y) \right) \). This can be accomplished by the Leibniz formula as in the proof of the previous lemma:
\[
\|
\partial^\alpha_y \left( (u^p - u^p_\ell)(y)\rho(y) \right) \|_X
\leq \sum_{\alpha' \leq \alpha} \binom{\alpha}{\alpha'} \|
\partial^{\alpha - \alpha'}_y (u^p - u^p_\ell)(y) \|_X \|
\partial^{\alpha'}_y \rho(y) \|_{L^\infty(\Gamma)}
\lesssim 2^{-\ell} \sum_{\alpha' \leq \alpha} \binom{\alpha}{\alpha'} |\alpha - \alpha'| \|
\partial^{\alpha - \alpha'}_{y} \rho \|_{L^2(D)} \|
\partial^{\alpha'}_y \rho \|_{L^\infty(\Gamma)}
\lesssim 2^{-\ell |\alpha|} (|\alpha| + 1)! \|
\rho \|_{L^\infty(\Gamma)} \|
\partial^\alpha_{y} \|_{L^2(D)} \|
\partial^\alpha_{y} \rho \|_{L^\infty(\Gamma)}.
\]
Herein, we introduced again the quantity \( \rho := \left\| \rho_1 \right\|_{W^{r,\infty}(-1,1)} \ldots \right\| \rho_m \right\|_{W^{r,\infty}(-1,1)} \) and \( \check{c} = \max_{k=1,\ldots,m} \{ e \gamma k, \rho_k \} \). We set \( C(r) := \max_{|\alpha| \leq r} (|\alpha| + 1)! \check{c}^{\alpha|} \) and obtain
\[
\|(\text{Int} - Q_\ell)(u^p - u^p_\ell)\|_X \lesssim (2^{-\ell r \ell'(m - 1)}2^{-\ell'}C(r)\|f\|_{L^2(D)}^p)^2.
\]
Then, exploiting that the bound on the derivatives of the integrand is independent of the parameter and taking square roots on both sides completes the proof. \( \square \)
Remark 6.5. As for the quasi-Monte Carlo quadrature, by slightly decreasing \( r \) in the convergence result for the sparse tensor product quadrature, we may remove the factor \( \ell^{m-1} \) since \( \ell^{m-1} \lesssim 2^\delta \) for arbitrary \( \delta > 0 \).

Estimates of the type (6.2) are crucial to show the following approximation result for the multilevel quadrature. More general, every quadrature that satisfies an estimate of type (6.2) is feasible for a related multilevel quadrature method.

Theorem 6.6. Let \( \{Q_\ell\} \) be a sequence of quadrature rules that satisfy an estimate of type (6.2), where \( u \in L^p(D) \) is the solution to (2.3) that satisfies (4.2). Then, the error of the multilevel estimator for the mean and the second moment defined in (6.1) is bounded by

\[
(6.12) \quad \left\| \text{Int} u^p - \sum_{\ell=0}^j \Delta Q_\ell u^p_{j-\ell} \right\|_X \lesssim 2^{-j} \|f\|_{L^2(D)},
\]

where \( X = H^1(D) \) if \( p = 1 \) and \( X = W^{1,1}(D) \) if \( p = 2 \).

Proof. We shall apply the following multilevel splitting of the error

\[
(6.13) \quad \left\| \text{Int} u^p - \sum_{\ell=0}^j \Delta Q_\ell u^p_{j-\ell} \right\|_X = \left\| \text{Int} u^p - Q_j u^p + \sum_{\ell=0}^j \Delta Q_\ell u^p - \sum_{\ell=0}^j \Delta Q_\ell u^p_{j-\ell} \right\|_X
\]

\[
\leq \left\| \text{Int} u^p - Q_j u^p \right\|_X + \sum_{\ell=0}^j \left\| \Delta Q_\ell (u^p - u^p_{j-\ell}) \right\|_X.
\]

The first term just reflects the quadrature error and can be bounded with similar arguments as in Lemmata 6.3 and 6.4 according to

\[
\left\| \text{Int} u^p - Q_j u^p \right\|_X \lesssim 2^{-j} \|f\|_{L^2(D)}
\]

with a constant that depends on \( m \). The term inside the sum satisfies with (6.2) that

\[
\left\| \Delta Q_\ell (u^p - u^p_{j-\ell}) \right\|_X \lesssim \left\| (\text{Int} - Q_\ell) (u^p - u^p_{j-\ell}) \right\|_X + \left\| (\text{Int} - Q_{\ell-1}) (u^p - u^p_{j-\ell}) \right\|_X
\]

\[
\lesssim 2^{-(\ell+j-\ell)} \|f\|_{L^2(D)} + 2^{-(\ell-1+j-\ell)} \|f\|_{L^2(D)} \lesssim 2^{-j} \|f\|_{L^2(D)}.
\]

Thus, we can estimate (6.13) as

\[
\left\| \text{Int} u^p - \sum_{\ell=0}^j \Delta Q_\ell u^p_{j-\ell} \right\|_X \lesssim 2^{-j} \|f\|_{L^2(D)} + \sum_{\ell=0}^j 2^{-j} \|f\|_{L^2(D)} \leq 2^{-j} (j+2) \|f\|_{L^2(D)}.
\]

This completes the proof.

Remark 6.7. Note that we can achieve in our framework also nestedness for the samples in the Monte Carlo method. This is due to the fact that independent samples have to be used only for the estimators \( Q_\ell \) for \( \ell = 0, \ldots, j \). But from the proof of the previous theorem, we see that \( Q_\ell \) has not to be sampled independently from \( Q_\ell \) for \( \ell \neq \ell' \). Thus, we may employ the same underlying set of sample points on each level.

7. Numerical approximation. The previous results guarantee that the consistency error due to the non-conformity of the finite element space is of the correct order.
In the actual implementation, instead of considering the bilinear form introduced in (4.1), we shall consider on level $\ell \geq 0$ the variational formulation
\[
\int_{D_\ell} \tilde{a}_\ell(x, y) \nabla u_\ell \nabla v_\ell \, dx = \int_{D_\ell} f v_\ell \, dx \quad \text{for all } v_\ell \in S^1_\ell(D),
\]
where $\tilde{a}_\ell(x, y)$ is a suitable piecewise constant approximation of $a(x, y)$ with respect to the triangulation $T_\ell$ on $D_\ell$. In this section, we will provide a result that also takes into account the consistency error due to numerical quadrature in the bilinear form. In particular, we account for the quadrature error that is introduced by integration with respect to $D_\ell$ instead of integration with respect to $D$.

Fig. 2. Triangle $\tilde{T}$ located at the boundary of the domain. The solid line indicates the boundary of $D$, while the dashed line indicates the boundary of $D_\ell$.

The situation is sketched in Fig. 2 for the two dimensional case: For the given triangle $T$ at the domain’s boundary, the areas of the true domain $D$ and its polygonal approximation $D_\ell$ differ by the grey shaded area. According to [5], this area is small relative to the size of the element. There holds
\[
|\tilde{T} \cap (D \triangle D_\ell)| \leq c h_\ell |\tilde{T}| \quad \text{for some constant } c > 0,
\]
where $D \triangle D_\ell := (D \setminus D_\ell) \cup (D \setminus D_\ell)$ is the symmetric difference of sets. Moreover, since we consider piecewise linear finite elements which are set to zero outside of $D_\ell$, we have
\[
\int_{\tilde{T}} a(y) \nabla u_\ell(y) \nabla v_\ell \, dx = \nabla u_\ell(y)|_T \nabla v_\ell|_T \int_{T \cap \tilde{T}} a(y) \, dx,
\]
where $T \in T_\ell$ is the polygonal approximation to $\tilde{T}$. Hence, setting
\[
a_\ell(y)|_{T \cap \tilde{T}} := \frac{1}{|T|} \int_{T \cap \tilde{T}} a(y) \, dx
\]
yields
\[
\int_{\tilde{T}} a(y) \nabla u_\ell(y) \nabla v_\ell \, dx = \int_{T} a_\ell(y) \nabla u_\ell(y) \nabla v_\ell \, dx \quad \text{for all } T \in T_\ell, v_\ell \in S^1_\ell(D)
\]
and, therefore,
\[
\int_{D} a(y) \nabla u_\ell(y) \nabla v_\ell \, dx = \int_{D_\ell} a_\ell(y) \nabla u_\ell(y) \nabla v_\ell \, dx \quad \text{for all } v_\ell \in S^1_\ell(D).
\]

Nevertheless, for numerical computations, it is more convenient to assume that $a(y) \in C^{0,1}(D \cup D_\ell)$ for all $\ell \geq 0$ and the barycenter $x_c \in T$ is also contained in $\tilde{T}$. Then, to avoid integration with respect to the curved element $\tilde{T}$, we employ a midpoint rule and consider $a_\ell(y)|_{T \cap \tilde{T}} := a(x_c, y)$ instead. We have the following
Lemma 7.1. There holds
\[ \| \partial_y^\alpha (a_\ell - \tilde{a}_\ell)(y) \|_{L^\infty(D)} \leq c h_\ell \gamma^\alpha \| a(y) \|_{W^{1,\infty}(D)} \]
for some constant \( c > 0 \) which depends on (7.1).

Proof. By Taylor’s theorem, there holds
\[ (7.2) \quad \| a(y) - a(x_c, y) \|_{L^\infty(D)} \leq c h_\ell \| a(y) \|_{W^{1,\infty}(D)}. \]
Moreover, we note that \( a_\ell \) as well as \( \tilde{a}_\ell \) differ on at most on \( |T_\ell| \) triangles, where the difference is constant for each \( T \in T_\ell \). Hence, we obtain
\[
\| \partial_y^\alpha (a_\ell - \tilde{a}_\ell)(y) \|_{L^\infty(D)} = \max_{T \in T_\ell} \frac{1}{|T|} \left| \int_{T \cap T} \partial_y^\alpha a(y) \, dx - \int_T \partial_y^\alpha a(x_c, y) \, dx \right|
\]
\[
= \max_{T \in T_\ell} \frac{1}{|T|} \left| \int_{T \cap T} \partial_y^\alpha (a - a(x_c))(y) \, dx - \int_{T \cap T} \partial_y^\alpha a(x_c, y) \, dx \right|
\]
\[
\leq \max_{T \in T_\ell} \frac{1}{|T|} \left( \int_{T \cap T} \| \partial_y^\alpha (a - a(x_c))(y) \|_W \, dx + \int_{T \cap T} \| \partial_y^\alpha a(x_c, y) \|_W \, dx \right).
\]
Obviously, since \( a(y) \) as well as \( a(x_c, y) \) are affine functions with respect to \( y \), all derivatives for \( |\alpha| > 1 \) vanish. For \( |\alpha| \leq 1 \), the first term is estimated by (7.2) together with the fact that \( |T| = |T \cap T| (1 + O(h_\ell)) \), while the second term can be bounded by \( h_\ell \gamma^\alpha \| a(y) \|_{W^{1,\infty}(D)} \) if \( \alpha_k = 1 \), due to (7.1). Consequently, we obtain
\[
\| \partial_y^\alpha (a_\ell - \tilde{a}_\ell)(y) \|_{L^\infty(D)} \leq \begin{cases} 
ch_\ell \| a(y) \|_{W^{1,\infty}(D)}, & |\alpha| = 0, \\
ch_\ell \gamma^\alpha \| a(y) \|_{W^{1,\infty}(D)}, & \alpha_k = 1, \\
0, & |\alpha| > 1, 
\end{cases}
\]
for some constant \( c > 0 \). This completes the proof. \( \Box \)

Having this lemma at our disposal, we can prove the main result of this section.

Theorem 7.2. Let \( u_\ell \in S_\ell(D) \) be the solution to
\[ \int_{D_\ell} a_\ell(y) \nabla u_\ell \nabla v_\ell \, dx = \int_{D_\ell} f v_\ell \, dx \quad \text{for all } v_\ell \in S_\ell(D), \]
while \( \tilde{u}_\ell \in S_\ell(D) \) solves
\[ \int_{D_\ell} \tilde{a}_\ell(y) \nabla \tilde{u}_\ell \nabla v_\ell \, dx = \int_{D_\ell} f v_\ell \, dx \quad \text{for all } v_\ell \in S_\ell(D). \]
Then, there holds
\[ \| \partial_y^\alpha (u_\ell - \tilde{u}_\ell)(y) \|_{H^1(D)} \leq C h_\ell |c|^{\alpha_1} |\gamma^\alpha a(y)\|_{W^{1,\infty}(D)} \| \tilde{u}(y) \|_{H^1(D)} \]
for some constants \( C, c > 0 \), which are independent of the parameter dimension \( m \).

Proof. There holds
\[ \int_{D_\ell} a_\ell(y) \nabla (u_\ell - \tilde{u}_\ell)(y) \nabla v_\ell \, dx = \int_{D_\ell} (\tilde{a}_\ell - a_\ell)(y) \nabla \tilde{u}_\ell(y) \nabla v_\ell \, dx. \]
Differentiating this equation yields via the Leibniz formula

\[
\int_{D_\ell} a_\ell(y) \nabla \partial_y^\alpha (u_\ell - \tilde{u}_\ell)(y) \nabla v_\ell \, dx
= - \sum_{\{k: \alpha_k \neq 0\}} \alpha_k \int_{D_\ell} \partial_x^\alpha a_\ell(y) \nabla \partial_y^{\alpha - \epsilon_k} (u_\ell - \tilde{u}_\ell)(y) \nabla v_\ell \, dx
+ \int_{D_\ell} (\tilde{a}_\ell - a_\ell)(y) \nabla \partial_y^\alpha \tilde{u}_\ell(y) \nabla v_\ell \, dx
+ \sum_{\{k: \alpha_k \neq 0\}} \alpha_k \int_{D_\ell} \partial_x^\alpha (\tilde{a}_\ell - a_\ell)(y) \nabla \partial_y^{\alpha - \epsilon_k} \tilde{u}_\ell(y) \nabla v_\ell \, dx.
\]

Hence, choosing \( v_\ell = \partial_y^\alpha (u_\ell - \tilde{u}_\ell)(y) \) results in

\[
a_{\ell, \min} \| \partial_y^\alpha (u_\ell - \tilde{u}_\ell)(y) \|_{H^1(D)} \leq \sum_{\{k: \alpha_k \neq 0\}} \alpha_k \gamma_k \| \partial_y^{\alpha - \epsilon_k} (u_\ell - \tilde{u}_\ell)(y) \|_{H^1(D)}
+ c h_\ell \| a(y) \|_{W^{1, \infty}(D)} \| \partial_y^\alpha \tilde{u}_\ell(y) \|_{H^1(D)}
+ \sum_{\{k: \alpha_k \neq 0\}} \alpha_k c \gamma_k h_\ell \| a(y) \|_{W^{1, \infty}(D)} \| \partial_y^{\alpha - \epsilon_k} \tilde{u}_\ell(y) \|_{H^1(D)},
\]

where \( a_{\ell, \min} > 0 \) is the constant of ellipticity associated to \( a_\ell \).

Next, we note that the standard bootstrapping argument can be employed to obtain the estimate

\[
\| \partial_y^\alpha \tilde{u}(y) \|_{H^1(D)} \leq C |c| c |\gamma| \| \tilde{u}(y) \|_{H^1(D)}
\]

for some constants \( C, c > 0 \), see e.g. [9]. Therefore, we arrive at

\[
a_{\ell, \min} \| \partial_y^\alpha (u_\ell - \tilde{u}_\ell)(y) \|_{H^1(D)} \leq \sum_{\{k: \alpha_k \neq 0\}} \alpha_k \gamma_k \| \partial_y^{\alpha - \epsilon_k} (u_\ell - \tilde{u}_\ell)(y) \|_{H^1(D)}
+ C |c| c |\gamma| \| a(y) \|_{W^{1, \infty}(D)} \| \tilde{u}(y) \|_{H^1(D)}.
\]

From the previous estimate, the claim is again obtained as in the proof of [31, Theorem 7].

The theorem directly yields to the fully discrete generic estimate

\[
\| (\text{Int} - Q_\ell)(u^p - \tilde{u}_\ell^p) \|_X \lesssim 2^{-(\ell + \ell')} \| f \|_{L^2(D)}^p \quad \text{for } p = 1, 2
\]

by using (6.2) and the triangle inequality.

8. Numerical results. The numerical examples in this section are performed in three spatial dimensions. For the finite element discretization, we employ MATLAB and the Partial Differential Equation Toolbox\(^2\). In both examples, the error is measured by interpolating the obtained solutions on a sufficiently fine grid and comparing it there to a reference solution. We consider the MLMC, the MLQMC based on the Halton sequence, and the MLCC. Moreover, we set the density to \( \rho(y) = (1/2)^m \) for our problems.

8.1. An analytical example. With our first example, we intend to validate the proposed method. To this end, we consider a simple quadrature problem on the unit ball \( \mathbf{D} = \{ \mathbf{x} \in \mathbb{R}^3 : \| \mathbf{x} \|_2 < 1 \} \). Fig. 3 depicts different tetrahedralizations for this geometry, which are in particular not nested. We aim at computing the expectation of the solution \( u \) to the parametric diffusion equation (1.1) with right hand side \( f \equiv 1 \) and random diffusion coefficient

\[
a(y) = \left( \prod_{i=1}^{6} \frac{3}{5} (2 - y_i^2) \right)^{-1}.
\]

Since the diffusion coefficient is independent of the spatial variable, we can reformulate the equation according to

\[
-\Delta u(y) = \prod_{i=1}^{6} \frac{3}{5} (2 - y_i^2) \text{ in } D, \quad u(y) = 0 \text{ on } \partial D, \quad y \in \Gamma.
\]

Thus, since the Bochner integral interchanges with closed operators, see e.g. [29], we obtain for the expectation of \( u \) the equation

\[
(8.1) \quad -\Delta \mathbb{E}[u(y)] = \mathbb{E} \left[ \prod_{i=1}^{6} \frac{3}{5} (2 - y_i^2) \right] = 1 \text{ in } D, \quad u(y) = 0 \text{ on } \partial D, \quad y \in \Gamma.
\]

Obviously, this equation is solved by \( \mathbb{E}[u](x) = (1 - \|x\|_2)^2/6. \)

![Tetrahedralizations of four different resolutions for the unit ball.](image)

In order to measure the error to the approximate solution, we interpolate the exact solution to a mesh consisting of 12,047,801 finite elements (this is level \( j = 8 \)). This involves a mesh size of \( h_8 = 0.0047 \). For the levels \( j = 0, \ldots, 7 \), the mesh sizes and corresponding degrees of freedom (DoF) are given in Table 1. Moreover, we chose \( N_0 = 10 \) for the Monte Carlo quadrature and for the quasi-Monte Carlo quadrature and set \( N_\ell = 10 \cdot 4^\ell \) and \( N_\ell = 10 \cdot 2^\ell \), respectively. For the MLMC, in order to approximate the root mean square error, we average five realizations of the related approximation error. For the Clenshaw-Curtis quadrature, the number of samples are chosen as if there would hold \( r = 1 \) in Lemma 6.4.\(^3\)

On the left side of Fig. 4, the error for the MLQMC, the MLCC and the MLMC is visualized. It is plotted against the work, which is expressed in terms of fine grid samples: In accordance with the degrees of freedom denoted in Table 1, we scale each sample on a particular level \( \ell \) with the factor \( \text{DoF}_\ell / \text{DoF}_j \), i.e. we weight a fine grid sample by \( 1 \) and scale the coarse grid samples accordingly. The work is then given by summing up the total number of samples per level times the related weight.

\(^3\)The Clenshaw-Curtis quadrature converges exponentially since the integrand is analytic. The choice \( r = 1 \) is conservative and reflects the pre-asymptotic regime.
<table>
<thead>
<tr>
<th>$\ell$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_\ell$</td>
<td>1.2</td>
<td>0.6</td>
<td>0.3</td>
<td>0.15</td>
<td>0.075</td>
<td>0.0375</td>
<td>0.0188</td>
</tr>
<tr>
<td>$\text{dof}_\ell$</td>
<td>8</td>
<td>27</td>
<td>1585</td>
<td>6042</td>
<td>29069</td>
<td>133376</td>
<td>551327</td>
</tr>
</tbody>
</table>

**Table 1**

Mesh sizes and DoF on the different levels for the unit ball.

It can be seen that MLQMC achieves the best error versus work rate. Moreover, the plot indicates that MLCC may asymptotically achieve a similar rate. MLMC seems to provide here only a halved rate compared to MLQMC. To give an insight on the number of samples spent on each particular level, we have depicted the corresponding numbers for $j = 7$ on the right hand side of Fig. 4. It turns out that the quasi-Monte Carlo quadrature requires the smallest number of quadrature points. In contrast, the number of points for the Monte Carlo quadrature and for the Clenshaw-Curtis quadrature are nearly the same. This may be caused by the conservative choice for the number of quadrature points for the latter. Nevertheless, for fixed parameter dimension $m$ and $r = 1$, we expect asymptotically similar rates for MLCC and MLQMC.

**8.2. A more complex example.** In our second example, the spatial domain is given by a model of the Zarya module of the International Space Station (ISS), which was the first module to be launched. The geometry can be imbedded into a cylinder with radius 0.52 and height 1.58.

We thank Martin Siegel (Rheinbach, Germany) who kindly provided us with this model.

---

**Fig. 5.** Tetrahedralizations of four different resolutions for the Zarya geometry.
In this example, the parametric diffusion coefficient is given by

\[
a(x, y) = 1 + \frac{\exp(||x||^2)}{20} \left( \sin(2\pi x_1) y_1 + \frac{1}{2} \sin(2\pi x_2) y_2 + \frac{1}{4} \sin(2\pi x_3) y_3 \\
+ \frac{1}{8} \sin(4\pi x_1) \sin(4\pi x_2) y_4 + \frac{1}{16} \sin(4\pi x_1) \sin(4\pi x_3) y_5 \\
+ \frac{1}{32} \sin(4\pi x_2) \sin(4\pi x_3) y_6 \right)
\]

and \( f = 10 \). For \( x \in D \) and \( y \in \Gamma \), the diffusion coefficient varies approximately in the range \([0.19, 1.81]\). Fig. 6 shows the mean (left) and the variance (right) of the reference solution. It has been computed on a mesh with 13,069,396 tetrahedrons resulting in a mesh size of \( h = 0.0039 \) by 10,000 quasi-Monte Carlo samples based on the Halton sequence. For the levels \( j = 0, \ldots, 6 \), the mesh sizes and corresponding DoF are given in Table 2.

Fig. 7 visualizes the errors of the approximate expectation and second moment for the different multilevel quadrature methods under consideration. The number of quadrature points for the presented methods are chosen as in the previous example. Again, MLQMC provides the best error versus work rate in the mean, as well as in the second moment. The rates of MLMC and MLCC are both lower here.

9. Conclusion. In the present article, we have reversed the construction of the conventional multilevel quadrature. This enables us to give up the nestedness of the spatial approximation spaces. In particular, a polygonal approximation of curved domain boundaries is sufficient for computing the finite element solution. Hence, black-box finite element solvers can be directly applied to compute the solution of the underlying boundary value problem. Note that adaptively refined finite element meshes can be easily used as well. Another aspect of our approach is that the cost is considerably reduced by the application of nested quadrature formulae. Both features have been demonstrated by numerical results for the Clenshaw-Curtis quadrature and the quasi-Monte Carlo quadrature based on Halton points. Of course, other nested quadrature formulae like the Gauss-Patterson quadrature can be used as well. The
Fig. 7. $H^1$-errors of the approximate mean (left) and $W^{1,1}$-errors of the approximate second moment (right) on the Zarya geometry for different quadrature methods.

application of quadrature formulae which are tailored to a possible anisotropy of the integrand is also straightforward. If non-nested quadrature formulae are applied, one arrives at a combination-technique-like representation of the multilevel quadrature.

REFERENCES


