MULTILEVEL QUADRATURE FOR ELLIPTIC PARAMETRIC PARTIAL DIFFERENTIAL EQUATIONS IN CASE OF POLYGONAL APPROXIMATIONS OF CURVED DOMAINS*

MICHAEL GRIEBEL†, HELMUT HARBRECHT‡, AND MICHAEL D. MULTERER§

Abstract. Multilevel quadrature methods for parametric operator equations such as the multilevel (quasi–) Monte Carlo method resemble a sparse tensor product approximation between the spatial variable and the parameter. We employ this fact to reverse the multilevel quadrature method 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 nonnested and even adaptively refined finite element meshes. We moreover provide a rigorous error and regularity analysis addressing the variational crimes of using polygonal approximations of curved domains and numerical quadrature of the bilinear form. Our results facilitate the construction of efficient multilevel quadrature methods based on deterministic high order quadrature rules for the stochastic parameter. Numerical results in three spatial dimensions are provided to illustrate the approach.

Key words. parametric partial differential equations, multilevel quadrature, variational crimes

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

DOI. 10.1137/18M1236265

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

$$(1.1) \qquad -\operatorname{div}\big(a(\boldsymbol{y})\nabla u(\boldsymbol{y})\big) = f(\boldsymbol{y}) \text{ in } D, \quad u(\boldsymbol{y}) = 0 \text{ on } \partial D, \quad \boldsymbol{y} \in \Gamma,$$

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, 13, 27, 36]. 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 $\boldsymbol{y} \in \Gamma$.

An efficient approach to deal with this quadrature problem is the multilevel Monte Carlo method (MLMC), which has been developed in [3, 17, 19, 30, 31]. As first observed in [15, 25], this approach resembles the sparse tensor product approximation between the physical space and the parameter space. Thus, the extension to the

†Institut für Numerische Simulation, Universität Bonn, 53115 Bonn, Germany, and Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Schloss Birlinghoven, 53754 Sankt Augustin, Germany (griebel@ins.uni-bonn.de).
‡Department of Mathematics and Information, Universität Basel, 4051 Basel, Switzerland (helmut.harbrecht@unibas.ch).
§Institute of Computational Science, Università della Svizzera italiana, 6900 Lugano, Switzerland (michael.multerer@usi.ch).

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. [11, 22, 25, 28, 34], 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 [33] 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 analysis of variational crimes, namely the effect of polygonal approximations of curved physical domains and the numerical approximation of the bilinear form. In particular, we provide regularity results of the fully discrete solution to (1.1). If the Karhunen–Loève expansion of the coefficient $a(\boldsymbol{y})$ in this equation decays sufficiently fast, the obtained results are even robust with respect to the parameter dimension m.

The considerations in this article are based on 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 subspaces with increasing approximation power in some linear spaces $\mathcal{H}_i$. To approximate a given object of the tensor product space $\mathcal{H}_1 \otimes \mathcal{H}_2$, it is canonical to consider the full tensor product spaces

$$U_j := V_j^{(1)} \otimes V_j^{(2)}.$$

However, since $\dim U_j = \dim V_j^{(1)} \cdot \dim V_j^{(2)}$, the approximation cost in U_j is often too high. To reduce this cost, one considers the approximation in so-called sparse tensor product 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 give rise to the multilevel decompositions

$$(1.2) \qquad V_j^{(i)} = \bigoplus_{\ell=0}^{j} W_\ell^{(i)}, \quad W_0^{(i)} := V_0^{(i)}, \quad i = 1, 2.$$

Then, the sparse tensor product space is defined by

$$(1.3) \qquad \widehat{U}_j := \bigoplus_{\ell+\ell' \leq j} W_\ell^{(1)} \otimes W_{\ell'}^{(2)}.$$

Under the assumptions that the dimensions of $\{V_\ell^{(1)}\}$ and $\{V_\ell^{(2)}\}$ form geometric series, the space $\widehat{U}_j$ contains, up to possible logarithmic factors, only

$$\mathcal{O}\big(\max\big\{\dim V_j^{(1)}, \dim V_j^{(2)}\big\}\big)$$

degrees of freedom (DoF). Nevertheless, $\widehat{U}_j$ 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; see [20].

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

$$(1.4) \qquad \widehat{U}_j = \bigoplus_{\ell=0}^{j} W_\ell^{(1)} \otimes \left(\bigoplus_{\ell'=0}^{j-\ell} W_{\ell'}^{(2)}\right) = \bigoplus_{\ell=0}^{j} W_\ell^{(1)} \otimes V_{j-\ell}^{(2)}.$$
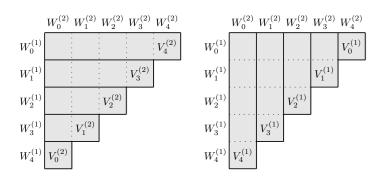
FIG. 1. *Different representations of the sparse tensor product space.*

This representation has already been proposed in [20]. Obviously, in complete analogy there holds

$$(1.5) \qquad \widehat{U}_j = \bigoplus_{\ell'=0}^{j} \left(\bigoplus_{\ell=0}^{j-\ell'} W_\ell^{(1)} \right) \otimes W_{\ell'}^{(2)} = \bigoplus_{\ell=0}^{j} V_{j-\ell}^{(1)} \otimes W_\ell^{(2)}.$$

We refer to Figure 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 $\{V_\ell^{(2)}\}$ are nested. Likewise, for the representation (1.5), the spaces $\{V_\ell^{(1)}\}$ need not be nested any more.

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

$$\int_\Gamma \mathcal{F}\big(u(\boldsymbol{y})\big)\rho(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y},$$

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

$$(1.6) \qquad \int_\Gamma \mathcal{F}\big(u(\boldsymbol{y})\big)\rho(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} \approx \sum_{\ell=0}^{j} Q_{j-\ell}\Delta\mathcal{F}_\ell\big(u(\boldsymbol{y})\big),$$

where $\Delta\mathcal{F}_\ell\big(u(\boldsymbol{y})\big) := \mathcal{F}\big(\mathfrak{u}_\ell(\boldsymbol{y})\big) - \mathcal{F}\big(\mathfrak{u}_{\ell-1}(\boldsymbol{y})\big)$ and $\mathcal{F}\big(\mathfrak{u}_{-1}(\boldsymbol{y})\big) := 0$; see [25]. On the other hand, similarly to (1.5), we obtain the decomposition

$$(1.7) \qquad \int_\Gamma \mathcal{F}\big(u(\boldsymbol{y})\big)\rho(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} \approx \sum_{\ell=0}^{j} \Delta Q_\ell \mathcal{F}\big(\mathfrak{u}_{j-\ell}(\boldsymbol{y})\big),$$

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

Originally, multilevel quadrature methods have been interpreted as variance reduction methods for the Monte Carlo quadrature, a view which was originally in-

troduced for the approximation of parametric integrals; cf. [30, 31]. Consequently, the representation (1.4), and thus the decomposition (1.6), has been used in previous articles; see, for example, [17, 18] for stochastic ordinary differential equations and [3, 25, 41, 42] for partial differential equations with random data. In this context, usually a nested sequence of approximation spaces is presumed such that the complement spaces $\{W_\ell^{(1)}\}$ 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 $\mathcal{F}$ being a functional; cf. [23, 41]. In the context of the Monte Carlo method, there are already results available, which allow for giving up this nestedness; see, e.g., [8, 41]. A more general result addressing the resulting error in the underlying bilinear form can be found in [40].

The decomposition (1.6) is well suited if the spatial dimension is small, as 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 impossible. In the article at hand, we employ the decomposition (1.7). It allows more naturally for nonnested finite element spaces which, in turn, induce different approximations of the underlying domain. Particularly, we shall focus on the error-equilibrated version of the multilevel quadrature (see [3]), rather than on the work-equilibrated one [18]. Moreover, using nested quadrature formulas, 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, i.e., 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. The key result in this section, namely Lemma 6.1, is robust with respect to the parameter dimension m. Afterward, 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 $(\Omega, \Sigma, \mathbb{P})$ be a complete and separable probability space with σ-field $\Sigma \subset 2^\Omega$ and probability measure $\mathbb{P}$. We intend to compute the expectation

$$\mathbb{E}[u] = \int_\Omega u(\omega)\,\mathrm{d}\mathbb{P}(\omega) \in H_0^1(D)$$

and the variance

$$\mathbb{V}[u] = \int_\Omega \left\{u(\omega) - \mathbb{E}[u]\right\}^2 \mathrm{d}\mathbb{P}(\omega) \in W_0^{1,1}(D)$$

of the random function $u(\omega) \in H_0^1(D)$ which solves the random diffusion problem

$$(2.1) \qquad - \operatorname{div}\big(a(\omega)\nabla u(\omega)\big) = f \text{ in } D \text{ for almost every } \omega \in \Omega.$$

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

$$(2.2) \qquad a(\boldsymbol{x}, \omega) = \mathbb{E}[a](\boldsymbol{x}) + \sum_{k=1}^{m} \sqrt{\lambda_k}\varphi_k(\boldsymbol{x})\psi_k(\omega)$$

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 $\rho_k \colon [-1, 1] \to \mathbb{R}$ with respect to the Lebesgue measure.

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

$$\operatorname{Cov}[a](\boldsymbol{x}, \boldsymbol{x}') = \int_\Omega \big\{a(\boldsymbol{x}, \omega) - \mathbb{E}[a](\boldsymbol{x})\big\}\big\{a(\boldsymbol{x}', \omega) - \mathbb{E}[a](\boldsymbol{x}')\big\} \, \mathrm{d}\mathbb{P}(\omega).$$

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 [16, 26, 35, 39].

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

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

$$\boldsymbol{\psi} \colon \Omega \to \Gamma, \quad \omega \mapsto \boldsymbol{\psi}(\omega) := \big(\psi_1(\omega), \ldots, \psi_m(\omega)\big).$$

Hence, the random variables ψ_k are substituted by coordinates $y_k \in [-1, 1]$.

We introduce the measure $\rho(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{y}$ on Γ, which is defined by the product density function $\rho(\boldsymbol{y}) := \prod_{k=1}^{m} \rho_k(y_k)$.

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 or a C^2-domain. We consider the parametric diffusion problem

$$(2.3) \qquad \begin{aligned} &\text{find } u \in L_\rho^2\big(\Gamma; H_0^1(D)\big) \text{ such that} \\ &- \operatorname{div}\big(a(\boldsymbol{y})\nabla u(\boldsymbol{y})\big) = f \text{ in } D \text{ for almost every } \boldsymbol{y} \in \Gamma, \end{aligned}$$

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

$$(2.4) \qquad a(\boldsymbol{x}, \boldsymbol{y}) = \varphi_0(\boldsymbol{x}) + \sum_{k=1}^{m} \sqrt{\lambda_k}\varphi_k(\boldsymbol{x})y_k, \quad \gamma_k := \sqrt{\lambda_k}\|\varphi_k\|_{W^{1,\infty}(D)}.$$

Note that $u \in L_\rho^2\big(\Gamma; H_0^1(D)\big)$ guarantees finite second order moments of the solution.

By the Lax–Milgram theorem, the unique solvability of the parametric diffusion problem (2.3) in $L_\rho^2\big(\Gamma; H_0^1(D)\big)$ follows immediately if we impose the condition

$$(2.5) \qquad 0 < a_{\min} \leq a(\boldsymbol{y}) \leq a_{\max} < \infty \text{ in } D$$

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

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

Hence, the solution to (2.3) is essentially bounded with respect to $\boldsymbol{y} \in \Gamma$.

In, e.g., [4, 9, 10, 12, 43], it has been proven that the solution u of (2.3) is analytic as mapping $u \colon \Gamma \to H_0^1(D)$. Moreover, it has been shown in [9] that u is even an analytic mapping $u \colon \Gamma \to \mathcal{W} := H_0^1(D) \cap H^2(D)$ given that the functions $\{\varphi_k\}$ in (2.4) belong to $W^{1,\infty}(D)$. This constitutes the necessary mixed regularity for a sparse tensor product discretization; see, e.g., [28, 34]. A similar result for diffusion problems with coefficients of the form $\exp\big(a(\boldsymbol{x}, \boldsymbol{y})\big)$ has been shown in [33].

Since u is supposed to be in $L_\rho^2\big(\Gamma; H_0^1(D)\big)$, we can compute its expectation

$$(2.6) \qquad \mathbb{E}[u] = \int_\Gamma u(\boldsymbol{y})\rho(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} \in H_0^1(D)$$

and its variance

$$(2.7) \qquad \mathbb{V}[u] = \mathbb{E}[u^2] - \mathbb{E}[u]^2 = \int_\Gamma u^2(\boldsymbol{y})\rho(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} - \mathbb{E}[u]^2 \in W_0^{1,1}(D).$$

We will focus in what follows 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 formulas $\{Q_\ell\}$ for the Bochner integral

$$\mathrm{Int} \colon L_\rho^1(\Gamma; \mathcal{X}) \to \mathcal{X}, \quad \mathrm{Int}\, v = \int_\Gamma v(\boldsymbol{y})\rho(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y},$$

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

$$(3.1) \qquad Q_\ell \colon L_\rho^1(\Gamma; \mathcal{X}) \to \mathcal{X}, \quad (Q_\ell v)(\boldsymbol{x}) = \sum_{i=1}^{N_\ell} \omega_{\ell,i} v(\boldsymbol{x}, \boldsymbol{\xi}_{\ell,i})\rho(\boldsymbol{\xi}_{\ell,i})$$

is supposed to provide the error bound

$$(3.2) \qquad \|(\mathrm{Int} - Q_\ell)v\|_{\mathcal{X}} \lesssim \varepsilon_\ell \|v\|_{\mathcal{H}(\Gamma; \mathcal{X})}$$

uniformly in $\ell \in \mathbb{N}$, where $\mathcal{H}(\Gamma; \mathcal{X}) \subset L_\rho^2(\Gamma, \mathcal{X})$ is a suitable Bochner space. Note that since the density ρ 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 that

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

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

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

$$(3.3) \qquad \|v\|_{W_{\mathrm{mix}}^{1,1}(\Gamma;\mathcal{X})} := \sum_{\|\boldsymbol{q}\|_\infty \leq 1} \int_\Gamma \left\| \frac{\partial^{\|\boldsymbol{q}\|_1}}{\partial y_1^{q_1} \partial y_2^{q_2} \cdots \partial y_m^{q_m}} v(\boldsymbol{y}) \right\|_{\mathcal{X}} \mathrm{d}\boldsymbol{y} < \infty;$$

see, e.g., [37]. 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. [24]), 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 [44]; see, e.g., [27].

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

$$(3.4) \qquad \|v\|_{W_{\mathrm{mix}}^{r,\infty}(\Gamma;\mathcal{X})} := \max_{\|\boldsymbol{\alpha}\|_\infty \leq r} \left\| \partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}} v \right\|_{L^\infty(\Gamma;\mathcal{X})} < \infty,$$

then one can apply a sparse grid Clenshaw–Curtis quadrature rule. This yields the convergence rate $\varepsilon_\ell = 2^{-\ell r} \ell^{m-1}$, where $N_\ell \sim 2^\ell \ell^{m-1}$ and $\mathcal{H}(\Gamma; \mathcal{X}) = W_{\mathrm{mix}}^{r,\infty}(\Gamma; \mathcal{X})$; see [38].[1]

For our purposes, we shall assume that the number N_ℓ of points of the quadrature formula Q_ℓ is chosen such that the corresponding accuracy is

$$(3.5) \qquad\qquad\qquad \varepsilon_\ell = 2^{-\ell}.$$

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

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

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

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

$$\mathcal{S}_\ell(D) := \{v \in C(D) : v|_T \text{ is a linear polynomial for all } T \in \mathcal{T}_\ell$$
$$\text{and } v(\boldsymbol{x}) = 0 \text{ for all nodes } \boldsymbol{x} \in \partial D\}$$

of continuous, piecewise linear finite elements. Notice that it does hold that $\mathcal{S}_\ell(D) \subset H^1(D)$ but in general $\mathcal{S}_\ell(D) \not\subset H_0^1(D)$.

---

[1] The Clenshaw–Curtis quadrature converges exponentially if the integrand $v \colon \Gamma \to \mathcal{X}$ and the density ρ are analytic.

We shall further introduce the finite element solution $\mathfrak{u}_\ell(\boldsymbol{y}) \in \mathcal{S}_\ell(D)$ of (2.3) which satisfies

$$(4.1) \qquad \mathcal{B}_{\boldsymbol{y}}(\mathfrak{u}_\ell, v_\ell) := \int_D a(\boldsymbol{y}) \nabla \mathfrak{u}_\ell(\boldsymbol{y}) \nabla v_\ell \, \mathrm{d}\boldsymbol{x} = \int_D f v_\ell \, \mathrm{d}\boldsymbol{x}$$

for all $v_\ell \in \mathcal{S}_\ell(D)$. If $D \neq D_\ell$, the bilinear form $\mathcal{B}_{\boldsymbol{y}}(\cdot, \cdot)$ is also well defined for functions from $\mathcal{S}_\ell(D)$ since $\mathcal{S}_\ell(D) \subset H^1(D)$. Nevertheless, in order to maintain the ellipticity of the bilinear form, we shall assume that the mesh size h_0 is sufficiently small to ensure that functions in $\mathcal{S}_\ell(D)$ are zero on a part of the boundary of D.

It is shown in, e.g., [5, 6] that the finite element solution $\mathfrak{u}_\ell(\boldsymbol{y}) \in \mathcal{S}_\ell(D)$ 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 $\mathfrak{u}_\ell(\boldsymbol{y}) \in \mathcal{S}_\ell(D)$ of the diffusion problem* (2.3) *and respectively its square $\mathfrak{u}_\ell^2(\boldsymbol{y})$ satisfy the error estimate*

$$(4.2) \qquad \left\| u^p(\boldsymbol{y}) - \mathfrak{u}_\ell^p(\boldsymbol{y}) \right\|_{\mathcal{X}} \lesssim h_\ell \|f\|_{L^2(D)}^p,$$

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 $\boldsymbol{y} \in \Gamma$.*

5. The multilevel quadrature method. Based on the framework introduced in 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 $\{\mathfrak{u}_\ell(\boldsymbol{y})\}_\ell$ of finite element solutions, there obviously holds $\lim_{\ell \to \infty} \mathfrak{u}_\ell(\boldsymbol{y}) = u(\boldsymbol{y})$ uniformly in $\boldsymbol{y} \in \Gamma$. Thus, if $\mathcal{F}$ is continuous, we obtain

$$(5.1) \qquad \lim_{\ell \to \infty} \mathcal{F}\big(\mathfrak{u}_\ell(\boldsymbol{y})\big) = \mathcal{F}\big(u(\boldsymbol{y})\big)$$

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

$$(5.2) \qquad \lim_{\ell \to \infty} Q_\ell v = \int_\Gamma v(\boldsymbol{y}) \rho(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{y}.$$

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

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

Since ΔQ_ℓ is linear and continuous, we end up with

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

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

Of course, the representations (1.6) and (1.7) are mathematically equivalent. More precisely, if we set $\mathcal{F}\big(\mathfrak{u}_{-1}(\boldsymbol{y})\big) := 0$, there holds

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

Thus, all available results for the representation (1.6) of the multilevel quadrature (see, e.g., [25, 28] 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 nonnested 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 formulas are employed.

6. Error analysis. In what follows, we restrict ourselves for reasons of simplicity to the situations $\mathcal{F}(u) = u$ and $\mathcal{F}(u) = u^2$ which yield the expectation and the second moment of the solution to (2.3). This means that we consider

$$(6.1) \qquad \operatorname{Int} u^p \approx \sum_{\ell=0}^{j} \Delta Q_\ell \mathfrak{u}_{j-\ell}^p = \sum_{\ell=0}^{j} Q_{j-\ell}\big(\mathfrak{u}_\ell^p - \mathfrak{u}_{\ell-1}^p\big) \quad \text{for } p = 1, 2.$$

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

$$(6.2) \qquad \big\|(\operatorname{Int}-Q_\ell)(u^p - \mathfrak{u}_{\ell'}^p)\big\|_{\mathcal{X}} \lesssim 2^{-(\ell+\ell')}\|f\|_{L^2(D)}^p \quad \text{for } p = 1, 2$$

with f being the right-hand side of (2.3) and $h_{\ell'} \sim 2^{-\ell'}$. In particular, any quadrature rule which satisfies this estimate gives rise to a multilevel quadrature method. In what follows, we provide this estimate for 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 $\boldsymbol{y}$; cf. [3, 25]. In this case, the generic estimate can be derived similarly to Strang's lemma; see [41]. Nevertheless, since the Monte Carlo quadrature does not provide deterministic error estimates, we have to replace the norm in $\mathcal{X}$ by the $L_\rho^2(\Gamma; \mathcal{X})$-norm.

The situation becomes much more involved if parametric regularity has to be taken into account. In the latter case, also bounds on the derivatives of the spatial details have to be provided. The next lemma is a generalization of similar results from [28, 34], which provide the smoothness of the Galerkin error with respect to the parameter $\boldsymbol{y} \in \Gamma$ for the nonconforming case $D \neq D_\ell$. In what follows, we denote by $\mathbb{N} := \{0, 1, 2, \ldots\}$ the set of natural numbers.

LEMMA 6.1. *For the error $\delta_\ell(\boldsymbol{y}) := (u - \mathfrak{u}_\ell)(\boldsymbol{y})$ of the Galerkin projection, there holds the estimate*

$$(6.3) \qquad \big\|\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}}\delta_\ell(\boldsymbol{y})\big\|_{H^1(D)} \leq Ch_\ell|\boldsymbol{\alpha}|!c^{|\boldsymbol{\alpha}|}\boldsymbol{\gamma}^{\boldsymbol{\alpha}}\|f\|_{L^2(D)} \quad \text{for all } \boldsymbol{\alpha} \in \mathbb{N}^m,$$

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

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

$$\mathcal{B}_{\boldsymbol{y}}(\mathfrak{u}_\ell, v_\ell) = \int_D a(\boldsymbol{y})\nabla\mathfrak{u}_\ell(\boldsymbol{y})\nabla v_\ell \, d\boldsymbol{x} = \int_D f v_\ell \, d\boldsymbol{x} \quad \text{for all } v_\ell \in \mathcal{S}_\ell(D).$$

On the other hand, integration by parts yields

$$\mathcal{B}_{\boldsymbol{y}}(u, v_\ell) = \int_D a(\boldsymbol{y}) \nabla u(\boldsymbol{y}) \nabla v_\ell \, \mathrm{d}\boldsymbol{x} = \int f v_\ell \, \mathrm{d}\boldsymbol{x} + \int_{\partial D} a(\boldsymbol{y}) \frac{\partial u}{\partial \boldsymbol{n}}(\boldsymbol{y}) v_\ell \, \mathrm{d}\sigma_{\boldsymbol{x}}$$

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

$$(6.4) \qquad \mathcal{B}_{\boldsymbol{y}}\big(u - \mathfrak{u}_\ell, v_\ell\big) = \int_{\partial D} a(\boldsymbol{y}) \frac{\partial u}{\partial \boldsymbol{n}}(\boldsymbol{y}) v_\ell \, \mathrm{d}\sigma_{\boldsymbol{x}} \quad \text{for all } v_\ell \in S_\ell(D).$$

Due to the uniform ellipticity of the bilinear form, we can also define the Galerkin projection $\mathcal{P}_\ell(\boldsymbol{y}) \colon H_0^1(D) \to S_\ell(D)$ via

$$\mathcal{B}_{\boldsymbol{y}}(u - \mathcal{P}_\ell u, v_\ell) = 0 \quad \text{for all } v_\ell \in \mathcal{S}_\ell(D).$$

It holds that

$$(6.5) \qquad \begin{aligned} \|\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}}(u - \mathfrak{u}_\ell)\|_{H^1(D)} &\leq \|\mathcal{P}_\ell \partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}}(u - \mathfrak{u}_\ell)\|_{H^1(D)} + \|(I - \mathcal{P}_\ell)\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}}(u - \mathfrak{u}_\ell)\|_{H^1(D)} \\ &\leq \|\mathcal{P}_\ell \partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}}(u - \mathfrak{u}_\ell)\|_{H^1(D)} + \|(I - \mathcal{P}_\ell)\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}} u\|_{H^1(D)}, \end{aligned}$$

since $\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}} \mathfrak{u}_\ell \in S_\ell(D)$ and hence $\mathcal{P}_\ell \partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}} \mathfrak{u}_\ell = \partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}} \mathfrak{u}_\ell$.

In order to estimate the first term, we employ the perturbed Galerkin orthogonality (6.4) and obtain

$$(6.6) \qquad \begin{aligned} \mathcal{B}_{\boldsymbol{y}}\big(\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}}(u - \mathfrak{u}_\ell), v_\ell\big) &- \partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}} \int_{\partial D} a(\boldsymbol{y}) \frac{\partial u}{\partial \boldsymbol{n}}(\boldsymbol{y}) v_\ell \, \mathrm{d}\sigma_{\boldsymbol{x}} \\ &= - \sum_{\{k : \alpha_k \neq 0\}} \alpha_k \sqrt{\lambda_k} \int_D \varphi_k \nabla \partial_{\boldsymbol{y}}^{\boldsymbol{\alpha} - \boldsymbol{e}_k}(u - \mathfrak{u}_\ell)(\boldsymbol{y}) \nabla v_\ell \, \mathrm{d}\boldsymbol{x}; \end{aligned}$$

see, e.g., [9], where $\boldsymbol{e}_k \in \mathbb{N}^m$ satisfies $\boldsymbol{e}_{k,j} = \delta_{k,j}$. The derivatives of the boundary term satisfy

$$\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}} \int_{\partial D} a(\boldsymbol{y}) \frac{\partial u}{\partial \boldsymbol{n}}(\boldsymbol{y}) v_\ell \, \mathrm{d}\sigma_{\boldsymbol{x}} = \sum_{\boldsymbol{\alpha}' \leq \boldsymbol{\alpha}} \binom{\boldsymbol{\alpha}}{\boldsymbol{\alpha}'} \int_{\partial D} \big[\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}'} a(\boldsymbol{y})\big] \left[\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha} - \boldsymbol{\alpha}'} \frac{\partial u}{\partial \boldsymbol{n}}(\boldsymbol{y})\right] v_\ell \, \mathrm{d}\sigma_{\boldsymbol{x}}$$

$$= \int_{\partial D} a(\boldsymbol{y}) \frac{\partial(\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}} u)}{\partial \boldsymbol{n}}(\boldsymbol{y}) v_\ell \, \mathrm{d}\sigma_{\boldsymbol{x}} + \sum_{\{k : \alpha_k \neq 0\}} \alpha_k \sqrt{\lambda_k} \int_{\partial D} \varphi_k \frac{\partial(\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha} - \boldsymbol{e}_k} u)}{\partial \boldsymbol{n}}(\boldsymbol{y}) v_\ell \, \mathrm{d}\sigma_{\boldsymbol{x}}.$$

Inserting this identity into (6.6) yields

$$(6.7) \qquad \begin{aligned} \mathcal{B}_{\boldsymbol{y}}\big(\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}}(u - \mathfrak{u}_\ell), v_\ell\big) &- \int_{\partial D} a(\boldsymbol{y}) \frac{\partial(\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}} u)}{\partial \boldsymbol{n}}(\boldsymbol{y}) v_\ell \, \mathrm{d}\sigma_{\boldsymbol{x}} \\ &= - \sum_{\{k : \alpha_k \neq 0\}} \alpha_k \sqrt{\lambda_k} \bigg[\int_D \varphi_k \nabla \partial_{\boldsymbol{y}}^{\boldsymbol{\alpha} - \boldsymbol{e}_k}(u - \mathfrak{u}_\ell)(\boldsymbol{y}) \nabla v_\ell \, \mathrm{d}\boldsymbol{x} \\ &\qquad\qquad - \int_{\partial D} \varphi_k \frac{\partial(\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha} - \boldsymbol{e}_k} u)}{\partial \boldsymbol{n}}(\boldsymbol{y}) v_\ell \, \mathrm{d}\sigma_{\boldsymbol{x}} \bigg]. \end{aligned}$$

In order to bound the boundary integrals, we employ the following estimate, which is valid for any $v, w \in H^1(D)$. It holds that

$$\left| \int_{\partial D} a(\boldsymbol{y}) \frac{\partial v}{\partial \boldsymbol{n}} w \, \mathrm{d}\sigma_{\boldsymbol{x}} \right| \leq a_{\max} \left\| \frac{\partial v}{\partial \boldsymbol{n}}(\boldsymbol{y}) \right\|_{H^{-1/2}(\partial D)} \|w\|_{H^{1/2}(\partial D)}$$

$$\leq C_{\mathrm{inv}} a_{\max} \|v\|_{H^1(D)} \|w\|_{H^{1/2}(\partial D)},$$

where C_{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

$$(6.8) \qquad \|v_\ell\|_{H^{1/2}(\partial D)} \le ch_\ell \|v_\ell\|_{H^1(D)} \quad \text{for all } v_\ell \in \mathcal{S}_\ell(D)$$

with some constant $c > 0$. From this, we infer

$$\left| \int_{\partial D} a(\boldsymbol{y}) \frac{\partial(\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}} u)}{\partial \boldsymbol{n}}(\boldsymbol{y}) v_\ell \, d\sigma_{\boldsymbol{x}} \right| \le Ch_\ell \|\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}} u(\boldsymbol{y})\|_{H^1(D)} \|v_\ell\|_{H^1(D)}$$

for all $v_\ell \in \mathcal{S}_\ell(D)$ and some constant $C > 0$.

Inserting the latter estimate into (6.7) and choosing $v_\ell = \mathcal{P}_\ell \partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}}(u - \mathfrak{u}_\ell)$ as a test function, we arrive at

$$a_{\min}\|\mathcal{P}_\ell \partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}}(u - \mathfrak{u}_\ell)(\boldsymbol{y})\|_{H^1(D)}^2 \le Ch_\ell \|\mathcal{P}_\ell \partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}} u(\boldsymbol{y})\|_{H^1(D)} \|\mathcal{P}_\ell \partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}}(u - \mathfrak{u}_\ell)(\boldsymbol{y})\|_{H^1(D)}$$
$$+ \sum_{\{k:\alpha_k \neq 0\}} \alpha_k \gamma_k \Big[\|\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha} - \boldsymbol{e}_k}(u - \mathfrak{u}_\ell)(\boldsymbol{y})\|_{H^1(D)} \|\mathcal{P}_\ell \partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}}(u - \mathfrak{u}_\ell)(\boldsymbol{y})\|_{H^1(D)}$$
$$+ Ch_\ell \|\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha} - \boldsymbol{e}_k} u(\boldsymbol{y})\|_{H^1(D)} \|\mathcal{P}_\ell \partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}}(u - \mathfrak{u}_\ell)(\boldsymbol{y})\|_{H^1(D)} \Big].$$

Simplifying this expression yields

$$\|\mathcal{P}_\ell \partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}}(u - \mathfrak{u}_\ell)(\boldsymbol{y})\|_{H^1(D)} \le Ch_\ell \|\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}} u(\boldsymbol{y})\|_{H^1(D)}$$
$$+ C \sum_{\{k:\alpha_k \neq 0\}} \alpha_k \gamma_k \Big[\|\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha} - \boldsymbol{e}_k}(u - \mathfrak{u}_\ell)(\boldsymbol{y})\|_{H^1(D)} + h_\ell \|\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha} - \boldsymbol{e}_k} u(\boldsymbol{y})\|_{H^1(D)} \Big]$$

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

$$\|\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}} u(\boldsymbol{y})\|_{H^1(D)} \le C|\boldsymbol{\alpha}|! c^{|\boldsymbol{\alpha}|} \boldsymbol{\gamma}^{\boldsymbol{\alpha}} \|f\|_{L^2(D)}$$

for some constants $C, c > 0$ (see [9]), we end up with

$$\|\mathcal{P}_\ell \partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}}(u - \mathfrak{u}_\ell)(\boldsymbol{y})\|_{H^1(D)}$$
$$\le Ch_\ell c^{|\boldsymbol{\alpha}|} |\boldsymbol{\alpha}|! \boldsymbol{\gamma}^{\boldsymbol{\alpha}} \|f\|_{L^2(D)} + C \sum_{\{k:\alpha_k \neq 0\}} \alpha_k \gamma_k \|\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha} - \boldsymbol{e}_k}(u - \mathfrak{u}_\ell)(\boldsymbol{y})\|_{H^1(D)}$$

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

$$\|\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}}(u - \mathfrak{u}_\ell)(\boldsymbol{y})\|_{H^1(D)} \le C \sum_{\{k:\alpha_k \neq 0\}} \alpha_k \gamma_k \|\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha} - \boldsymbol{e}_k}(u - \mathfrak{u}_\ell)(\boldsymbol{y})\|_{H^1(D)}$$
$$+ Ch_\ell c^{|\boldsymbol{\alpha}|} |\boldsymbol{\alpha}|! \boldsymbol{\gamma}^{\boldsymbol{\alpha}} \|f\|_{L^2(D)} + \|(I - \mathcal{P}_\ell) \partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}} u\|_{H^1(D)}$$
$$\le C \sum_{\{k:\alpha_k \neq 0\}} \alpha_k \gamma_k \|\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha} - \boldsymbol{e}_k}(u - \mathfrak{u}_\ell)(\boldsymbol{y})\|_{H^1(D)} + Ch_\ell c^{|\boldsymbol{\alpha}|} |\boldsymbol{\alpha}|! \boldsymbol{\gamma}^{\boldsymbol{\alpha}} \|f\|_{L^2(D)},$$

where we used $\|(I - \mathcal{P}_\ell) \partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}} u\|_{H^1(D)} \le Ch_\ell c^{|\boldsymbol{\alpha}|} |\boldsymbol{\alpha}|! \boldsymbol{\gamma}^{\boldsymbol{\alpha}} \|f\|_{L^2(D)}$ for some constants $C, c > 0$, which follows from the approximation property of the finite element space $S_\ell(D)$ and [34, Theorem 6]. The proof is now concluded similarly to the proof of [34, Theorem 7]. $\square$

With this lemma, it is easy to show the following result related to the second moment; cf. [28].

LEMMA 6.2. *The derivatives of the difference $u^2 - \mathfrak{u}_\ell^2$ satisfy the estimate*

$$(6.9) \qquad \left\|\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}}\big(u^2 - \mathfrak{u}_\ell^2\big)(\boldsymbol{y})\right\|_{W^{1,1}(D)} \leq C h_\ell |\boldsymbol{\alpha}|! c^{|\boldsymbol{\alpha}|} \boldsymbol{\gamma}^{\boldsymbol{\alpha}} \|f\|_{L^2(D)}^2 \quad \textit{for all } \boldsymbol{\alpha} \in \mathbb{N}^m$$

with constants $C, c > 0$ dependent on $a_{\min}$ and $a_{\max}$.

With the aid of Lemmas 6.1 and 6.2 together with the results from [44], the generic estimate for MLQMC with Halton points can be derived. The following lemma is, for example, shown in [29, 40].

LEMMA 6.3. *Let $u \in L_\rho^2\big(\Gamma; H_0^1(D)\big)$ be the solution to (2.3) and $\mathfrak{u}_\ell$ the associated Galerkin projection on level ℓ. 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*

$$(6.10) \qquad \left\|(\mathrm{Int} - Q_\ell)(u^p - \mathfrak{u}_{\ell'}^p)\right\|_{\mathcal{X}} \lesssim 2^{-(\ell+\ell')} \|f\|_{L^2(D)}^p \quad \textit{for } p = 1, 2$$

with $N_\ell \sim 2^{\ell/(1-\delta)}$ for arbitrary $\delta > 0$.

In what follows, we derive the generic estimate for the sparse grid quadrature based on the nested *Clenshaw–Curtis abscissae*; cf. [14, 38]. These are given by the extrema of the Chebyshev polynomials

$$\xi_k = \cos\left(\frac{(k-1)\pi}{n-1}\right) \quad \text{for } k = 1, \ldots, n,$$

where $n = 2^{j-1} + 1$ if $j > 1$ and $n = 1$ with $\xi_1 = 0$ if $j = 1$.

LEMMA 6.4. *Let $u \in L_\rho^2\big(\Gamma; H_0^1(D)\big)$ be the solution to (2.3) and let $\mathfrak{u}_\ell$ be the associated Galerkin projection on level ℓ. 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*

$$(6.11) \qquad \left\|(\mathrm{Int} - Q_\ell)(u^p - \mathfrak{u}_{\ell'}^p)\right\|_{\mathcal{X}} \lesssim 2^{-(\ell r + \ell')} \ell^{m-1} \|f\|_{L^2(D)}^p \quad \textit{for } p = 1, 2$$

provided that $N_\ell \sim 2^\ell \ell^{d-1}$.

Proof. It is shown in [38] that the number N_ℓ of quadrature points of the sparse tensor product quadrature with Clenshaw–Curtis abscissae is of the order $\mathcal{O}(2^\ell \ell^{d-1})$. In addition, we have for functions $v \colon \Gamma \to \mathbb{R}$ with mixed regularity the following error bound:

$$\left|\int_\Gamma v(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{y} - \sum_{i=1}^{N_\ell} w_i v(\boldsymbol{\xi}_i)\right| \lesssim 2^{-\ell r} \ell^{(m-1)} \max_{\|\boldsymbol{\alpha}\|_\infty \leq r} \left\|\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}} v\right\|_{L^\infty(\Gamma)}.$$

Hence, to prove the desired assertion, we have to provide estimates on the derivatives $\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}}\big[\big(u^p(\boldsymbol{y}) - \mathfrak{u}_{\ell'}^p(\boldsymbol{y})\big)\rho(\boldsymbol{y})\big]$. This can be accomplished by the Leibniz formula as in the proof of the previous lemma:

$$\left\|\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}}\big[(u^p - \mathfrak{u}_{\ell'}^p)(\boldsymbol{y})\rho(\boldsymbol{y})\big]\right\|_{\mathcal{X}}$$
$$\leq \sum_{\boldsymbol{\alpha}' \leq \boldsymbol{\alpha}} \binom{\boldsymbol{\alpha}}{\boldsymbol{\alpha}'} \left\|\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}-\boldsymbol{\alpha}'}(u^p - \mathfrak{u}_{\ell'}^p)(\boldsymbol{y})\right\|_{\mathcal{X}} \left\|\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}'}\rho(\boldsymbol{y})\right\|_{L^\infty(\Gamma)}$$
$$\lesssim 2^{-\ell'} \sum_{\boldsymbol{\alpha}' \leq \boldsymbol{\alpha}} \binom{\boldsymbol{\alpha}}{\boldsymbol{\alpha}'} |\boldsymbol{\alpha} - \boldsymbol{\alpha}'|! c^{|\boldsymbol{\alpha}-\boldsymbol{\alpha}'|} \boldsymbol{\gamma}^{\boldsymbol{\alpha}-\boldsymbol{\alpha}'} \|f\|_{L^2(D)}^p \boldsymbol{\rho}^{\boldsymbol{\alpha}'}$$
$$\lesssim 2^{-\ell'} (|\boldsymbol{\alpha}| + 1)! \|f\|_{L^2(D)}^p \tilde{c}^{|\boldsymbol{\alpha}|}.$$

Herein, we introduced again the quantity $\boldsymbol{\rho} := \big[\|\rho_1\|_{W^{r,\infty}(-1,1)}, \ldots, \|\rho_m\|_{W^{r,\infty}(-1,1)}\big]$ and $\tilde{c} = \max_{k=1,\ldots,m} \max\{c\gamma_k, \rho_k\}$. We set $C(r) := \max_{\|\boldsymbol{\alpha}\|_\infty \leq r}(|\boldsymbol{\alpha}| + 1)! \tilde{c}^{|\boldsymbol{\alpha}|}$ and obtain

$$\big\|(\text{Int} - Q_\ell)(u^p - \mathfrak{u}_{\ell'}^p)\big\|_{\mathcal{X}}^2 \lesssim \big(2^{-\ell r}\ell^{(m-1)}2^{-\ell'}C(r)\|f\|_{L^2(D)}^p\big)^2.$$

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. □

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 ℓ^{m-1} since $\ell^{m-1} \lesssim 2^{\ell\delta}$ for arbitrary $\delta > 0$.

Employing the generic estimate (6.2) it is straightforward to derive the following approximation result for the multilevel quadrature. We emphasize again that every quadrature rule which satisfies (6.2) gives rise to a corresponding multilevel quadrature method.

THEOREM 6.6. *Let $\{Q_\ell\}$ be a sequence of quadrature rules that satisfy an estimate of type* (6.2), *where $u \in L_\rho^2\big(\Gamma, H_0^1(D)\big)$ 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*

$$(6.12) \qquad \left\|\text{Int } u^p - \sum_{\ell=0}^{j} \Delta Q_\ell \mathfrak{u}_{j-\ell}^p\right\|_{\mathcal{X}} \leq c(m)2^{-j}j\|f\|_{L^2(D)}^p,$$

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

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

$$(6.13) \qquad \begin{aligned} \left\|\text{Int } u^p - \sum_{\ell=0}^{j} \Delta Q_\ell \mathfrak{u}_{j-\ell}^p\right\|_{\mathcal{X}} &= \left\|\text{Int } u^p - Q_j u^p + \sum_{\ell=0}^{j} \Delta Q_\ell u^p - \sum_{\ell=0}^{j} \Delta Q_\ell \mathfrak{u}_{j-\ell}^p\right\|_{\mathcal{X}} \\ &\leq \big\|\text{Int } u^p - Q_j u^p\big\|_{\mathcal{X}} + \sum_{\ell=0}^{j} \big\|\Delta Q_\ell\big(u^p - \mathfrak{u}_{j-\ell}^p\big)\big\|_{\mathcal{X}}. \end{aligned}$$

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

$$\big\|\text{Int } u^p - Q_j u^p\big\|_{\mathcal{X}} \leq c(m)2^{-j}\|f\|_{L^2(D)}^p.$$

The term inside the sum satisfies with (6.2) that

$$\begin{aligned} \big\|\Delta Q_\ell\big(u^p - \mathfrak{u}_{j-\ell}^p\big)\big\|_{\mathcal{X}} &\leq \big\|(\text{Int} - Q_\ell)\big(u^p - \mathfrak{u}_{j-\ell}^p\big)\big\|_{\mathcal{X}} + \big\|(\text{Int} - Q_{\ell-1})\big(u^p - \mathfrak{u}_{j-\ell}^p\big)\big\|_{\mathcal{X}} \\ &\lesssim 2^{-(\ell+j-\ell)}\|f\|_{L^2(D)}^p + 2^{-(\ell-1+j-\ell)}\|f\|_{L^2(D)}^p \lesssim 2^{-j}\|f\|_{L^2(D)}^p. \end{aligned}$$

Thus, we can estimate (6.13) as

$$\begin{aligned} \left\|\text{Int } u^p - \sum_{\ell=0}^{j} \Delta Q_\ell \mathfrak{u}_{j-\ell}^p\right\|_{\mathcal{X}} &\leq c(m)2^{-j}\|f\|_{L^2(D)}^p + c'\sum_{\ell=0}^{j} 2^{-j}\|f\|_{L^2(D)}^p \\ &\leq c(m)2^{-j}(j+2)\|f\|_{L^2(D)}^p \end{aligned}$$

for some constant c' which is independent of m. This completes the proof. □
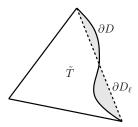
FIG. 2. *Curved element $\tilde{T}$ located at the boundary of the domain. The solid line indicates the boundary of D, while the dashed line indicates the boundary of D_ℓ.*

Remark 6.7. Note that the dependence on m only enters due to the error estimate of the quadrature for the parametric integral. Hence, if the decay of the Karhunen–Loève expansion allows for the use of a dimension-robust quadrature rule as, e.g., considered in [21, 44], then the dimension-dependence in the previous theorem can completely be removed.

We further remark 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_ℓ for $\ell = 0, \ldots, j$. But from the proof of the previous theorem, we see that Q_ℓ does not have 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 nonconformity 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

$$\int_{D_\ell} \tilde{a}_\ell(\boldsymbol{y}) \nabla \tilde{\mathfrak{u}}_\ell \nabla v_\ell \, \mathrm{d}\boldsymbol{x} = \int_{D_\ell} f v_\ell \, \mathrm{d}\boldsymbol{x} \quad \text{for all } v_\ell \in \mathcal{S}_\ell(D),$$

where $\tilde{a}_\ell(\boldsymbol{x}, \boldsymbol{y})$ is a suitable piecewise constant approximation of $a(\boldsymbol{x}, \boldsymbol{y})$ with respect to the triangulation $\mathcal{T}_\ell$ on D_ℓ. 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_ℓ instead of integration with respect to D.

The situation is sketched in Figure 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_ℓ differ by the gray shaded area. According to [5], this area is small relative to the size of the element. There holds

(7.1) $$|\tilde{T} \cap (D \triangle D_\ell)| \leq c h_\ell |\tilde{T}| \quad \text{for some constant } c > 0,$$

where $D \triangle D_\ell := (D \setminus D_\ell) \cup (D_\ell \setminus D)$ is the symmetric difference of sets. Moreover, since we consider piecewise linear finite elements which are set to zero outside of D_ℓ, we have

$$\int_{\tilde{T}} a(\boldsymbol{y}) \nabla \mathfrak{u}_\ell(\boldsymbol{y}) \nabla v_\ell \, \mathrm{d}\boldsymbol{x} = \nabla \mathfrak{u}_\ell(\boldsymbol{y})|_T \nabla v_\ell|_T \int_{\tilde{T} \cap T} a(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{x},$$

where $T \in \mathcal{T}_\ell$ is the polygonal approximation to $\tilde{T}$. Hence, setting

$$a_\ell(\boldsymbol{y})|_{T \cup \tilde{T}} := \frac{1}{|T|} \int_{\tilde{T} \cap T} a(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{x}$$

yields

$$\int_{\tilde{T}} a(\boldsymbol{y}) \nabla \mathfrak{u}_\ell(\boldsymbol{y}) \nabla v_\ell \, \mathrm{d}\boldsymbol{x} = \int_{T} a_\ell(\boldsymbol{y}) \nabla \mathfrak{u}_\ell(\boldsymbol{y}) \nabla v_\ell \, \mathrm{d}\boldsymbol{x} \quad \text{for all } T \in \mathcal{T}_\ell, v_\ell \in \mathcal{S}_\ell(D)$$

and, therefore,

$$\int_{D} a(\boldsymbol{y}) \nabla \mathfrak{u}_\ell(\boldsymbol{y}) \nabla v_\ell \, \mathrm{d}\boldsymbol{x} = \int_{D_\ell} a_\ell(\boldsymbol{y}) \nabla \mathfrak{u}_\ell(\boldsymbol{y}) \nabla v_\ell \, \mathrm{d}\boldsymbol{x} \quad \text{for all } v_\ell \in \mathcal{S}_\ell(D).$$

Nevertheless, for numerical computations, it is more convenient to assume that $a(\boldsymbol{y}) \in C^{0,1}(D \cup D_\ell)$ for all $\ell \geq 0$ and the barycenter $\boldsymbol{x}_c \in T$ is also contained in $\tilde{T}$. Then, to avoid integration with respect to the curved element $\tilde{T}$, we employ the midpoint rule, i.e., we consider the piecewise constant approximation $\tilde{a}_\ell(\boldsymbol{y})|_{T \cup \tilde{T}} := a(\boldsymbol{x}_c, \boldsymbol{y})$ of the diffusion coefficient. We have the following.

LEMMA 7.1. *There holds*

$$\left\| \partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}} (a_\ell - \tilde{a}_\ell)(\boldsymbol{y}) \right\|_{L^\infty(D)} \leq c h_\ell \boldsymbol{\gamma}^{\boldsymbol{\alpha}} \| a(\boldsymbol{y}) \|_{W^{1,\infty}(D)}$$

for some constant $c > 0$ which depends on (7.1).

Proof. By Taylor's theorem, there holds

(7.2) $$\| a(\boldsymbol{y}) - a(\boldsymbol{x}_c, \boldsymbol{y}) \|_{L^\infty(D)} \leq c h_\ell \| a(\boldsymbol{y}) \|_{W^{1,\infty}(D)}.$$

Moreover, we note that a_ℓ as well as $\tilde{a}_\ell$ differ on at most $|\mathcal{T}_\ell|$ elements, where the difference is constant for each $T \in \mathcal{T}_\ell$. Hence, we obtain

$$\left\| \partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}} (a_\ell - \tilde{a}_\ell)(\boldsymbol{y}) \right\|_{L^\infty(D)} = \max_{T \in \mathcal{T}_\ell} \frac{1}{|T|} \left| \int_{\tilde{T} \cap T} \partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}} a(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{x} - \int_{T} \partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}} a(\boldsymbol{x}_c, \boldsymbol{y}) \, \mathrm{d}\boldsymbol{x} \right|$$

$$= \max_{T \in \mathcal{T}_\ell} \frac{1}{|T|} \left| \int_{\tilde{T} \cap T} \partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}} (a - a(\boldsymbol{x}_c))(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{x} - \int_{T \setminus \tilde{T}} \partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}} a(\boldsymbol{x}_c, \boldsymbol{y}) \, \mathrm{d}\boldsymbol{x} \right|$$

$$\leq \max_{T \in \mathcal{T}_\ell} \frac{1}{|T|} \left(\left| \int_{\tilde{T} \cap T} \partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}} (a - a(\boldsymbol{x}_c))(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{x} \right| + \left| \int_{T \setminus \tilde{T}} \partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}} a(\boldsymbol{x}_c, \boldsymbol{y}) \, \mathrm{d}\boldsymbol{x} \right| \right).$$

Obviously, since $a(\boldsymbol{y})$ as well as $a(\boldsymbol{x}_c, \boldsymbol{y})$ are affine functions with respect to $\boldsymbol{y}$, all derivatives for $|\boldsymbol{\alpha}| > 1$ vanish. For $|\boldsymbol{\alpha}| \leq 1$, the first term is estimated by (7.2) together with the fact that $|T| = |T \cap \tilde{T}|(1 + \mathcal{O}(h_\ell))$, while the second term can be bounded by $h_\ell \gamma_k \| a(\boldsymbol{y}) \|_{W^{1,\infty}(D)}$ if $\alpha_k = 1$, due to (7.1). Consequently, we obtain

$$\left\| \partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}} (a_\ell - \tilde{a}_\ell)(\boldsymbol{y}) \right\|_{L^\infty(D)} \leq \begin{cases} c h_\ell \| a(\boldsymbol{y}) \|_{W^{1,\infty}(D)}, & |\boldsymbol{\alpha}| = 0, \\ c h_\ell \gamma_k \| a(\boldsymbol{y}) \|_{W^{1,\infty}(D)}, & \alpha_k = 1, \\ 0, & |\boldsymbol{\alpha}| > 1, \end{cases}$$

for some constant $c > 0$. This completes the proof. $\square$

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

THEOREM 7.2. *Let $\mathfrak{u}_\ell \in \mathcal{S}_\ell(D)$ be the solution to*

$$\int_{D_\ell} a_\ell(\boldsymbol{y}) \nabla \mathfrak{u}_\ell \nabla v_\ell \, \mathrm{d}\boldsymbol{x} = \int_{D_\ell} f v_\ell \, \mathrm{d}\boldsymbol{x} \quad \text{for all } v_\ell \in \mathcal{S}_\ell(D),$$

while $\tilde{\mathfrak{u}}_\ell \in \mathcal{S}_\ell(D)$ *solves*

$$\int_{D_\ell} \tilde{a}_\ell(\boldsymbol{y})\nabla\tilde{\mathfrak{u}}_\ell \nabla v_\ell \, \mathrm{d}\boldsymbol{x} = \int_{D_\ell} f v_\ell \, \mathrm{d}\boldsymbol{x} \quad \textit{for all } v_\ell \in \mathcal{S}_\ell(D).$$

Then, there holds

$$\|\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}}(\mathfrak{u}_\ell - \tilde{\mathfrak{u}}_\ell)(\boldsymbol{y})\|_{H^1(D)} \leq C h_\ell c^{|\boldsymbol{\alpha}|} |\boldsymbol{\alpha}|! \boldsymbol{\gamma}^{\boldsymbol{\alpha}} \|a(\boldsymbol{y})\|_{W^{1,\infty}(D)} \|\tilde{\mathfrak{u}}(\boldsymbol{y})\|_{H^1(D)}$$

for some constants $C, c > 0$, which are independent of the parameter dimension m.

Proof. There holds

$$\int_{D_\ell} a_\ell(\boldsymbol{y})\nabla\big(\mathfrak{u}_\ell - \tilde{\mathfrak{u}}_\ell\big)(\boldsymbol{y})\nabla v_\ell \, \mathrm{d}\boldsymbol{x} = \int_{D_\ell} (\tilde{a}_\ell - a_\ell)(\boldsymbol{y})\nabla\tilde{\mathfrak{u}}_\ell(\boldsymbol{y})\nabla v_\ell \, \mathrm{d}\boldsymbol{x}.$$

Differentiating this equation yields via the Leibniz formula

$$\int_{D_\ell} a_\ell(\boldsymbol{y})\nabla\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}}\big(\mathfrak{u}_\ell - \tilde{\mathfrak{u}}_\ell\big)(\boldsymbol{y})\nabla v_\ell \, \mathrm{d}\boldsymbol{x}$$

$$= -\sum_{\{k:\alpha_k \neq 0\}} \alpha_k \int_{D_\ell} \partial^{\boldsymbol{e}_k} a_\ell(\boldsymbol{y})\nabla\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}-\boldsymbol{e}_k}\big(\mathfrak{u}_\ell - \tilde{\mathfrak{u}}_\ell\big)(\boldsymbol{y})\nabla v_\ell \, \mathrm{d}\boldsymbol{x}$$

$$+ \int_{D_\ell} (\tilde{a}_\ell - a_\ell)(\boldsymbol{y})\nabla\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}}\tilde{\mathfrak{u}}_\ell(\boldsymbol{y})\nabla v_\ell \, \mathrm{d}\boldsymbol{x}$$

$$+ \sum_{\{k:\alpha_k \neq 0\}} \alpha_k \int_{D_\ell} \partial^{\boldsymbol{e}_k}(\tilde{a}_\ell - a_\ell)(\boldsymbol{y})\nabla\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}-\boldsymbol{e}_k}\tilde{\mathfrak{u}}_\ell(\boldsymbol{y})\nabla v_\ell \, \mathrm{d}\boldsymbol{x}.$$

Hence, choosing $v_\ell = \partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}}(\mathfrak{u}_\ell - \tilde{\mathfrak{u}}_\ell)(\boldsymbol{y})$ results in

$$a_{\ell,\min}\|\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}}(\mathfrak{u}_\ell - \tilde{\mathfrak{u}}_\ell)(\boldsymbol{y})\|_{H^1(D)} \leq \sum_{\{k:\alpha_k \neq 0\}} \alpha_k \gamma_k \|\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}-\boldsymbol{e}_k}(\mathfrak{u}_\ell - \tilde{\mathfrak{u}}_\ell)(\boldsymbol{y})\|_{H^1(D)}$$

$$+ c h_\ell \|a(\boldsymbol{y})\|_{W^{1,\infty}(D)}\|\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}}\tilde{\mathfrak{u}}_\ell(\boldsymbol{y})\|_{H^1(D)}$$

$$+ \sum_{\{k:\alpha_k \neq 0\}} \alpha_k c \gamma_k h_\ell \|a(\boldsymbol{y})\|_{W^{1,\infty}(D)}\|\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}-\boldsymbol{e}_k}\tilde{\mathfrak{u}}_\ell(\boldsymbol{y})\|_{H^1(D)},$$

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

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

$$\|\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}}\tilde{\mathfrak{u}}(\boldsymbol{y})\|_{H^1(D)} \leq C|\boldsymbol{\alpha}|! c^{|\boldsymbol{\alpha}|}\boldsymbol{\gamma}^{\boldsymbol{\alpha}}\|\tilde{\mathfrak{u}}(\boldsymbol{y})\|_{H^1(D)}$$

for some constants $C, c > 0$; see, e.g., [9]. Therefore, we arrive at

$$a_{\ell,\min}\|\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}}(\mathfrak{u}_\ell - \tilde{\mathfrak{u}}_\ell)(\boldsymbol{y})\|_{H^1(D)} \leq \sum_{\{k:\alpha_k \neq 0\}} \alpha_k \gamma_k \|\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}-\boldsymbol{e}_k}(\mathfrak{u}_\ell - \tilde{\mathfrak{u}}_\ell)(\boldsymbol{y})\|_{H^1(D)}$$

$$+ C|\boldsymbol{\alpha}|! h_\ell c^{|\boldsymbol{\alpha}|}\boldsymbol{\gamma}^{\boldsymbol{\alpha}}\|a(\boldsymbol{y})\|_{W^{1,\infty}(D)}\|\tilde{\mathfrak{u}}(\boldsymbol{y})\|_{H^1(D)}.$$

From the previous estimate, the claim is again obtained as in the proof of [34, Theorem 7]. □

The theorem directly yields to the fully discrete generic estimate

$$\big\|(\mathrm{Int} - Q_\ell)(u^p - \tilde{\mathfrak{u}}_{\ell'}^p)\big\|_{\mathcal{X}} \lesssim 2^{-(\ell+\ell')}\|f\|_{L^2(D)}^p \quad \text{for } p = 1, 2$$

by using (6.2) and the triangle inequality. Hence, Theorem 6.6 holds with the obvious modifications also for $\tilde{\mathfrak{u}}_{j-\ell}^p$ instead of $\mathfrak{u}_{j-\ell}^p$.
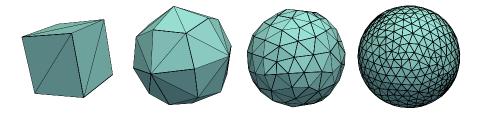
FIG. 3. *Tetrahedralizations of four different resolutions for the unit ball.*

8. Numerical results. The numerical examples in this section are performed in three spatial dimensions. For the finite element discretization, we use 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 MLMC, MLQMC based on the Halton sequence, and MLCC. Moreover, we set the density to $\rho(\boldsymbol{y}) = (1/2)^m$ 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 $D = \{\boldsymbol{x} \in \mathbb{R}^3 : \|\boldsymbol{x}\|_2 < 1\}$. Figure 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

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

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

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

Thus, since the Bochner integral interchanges with closed operators (see, e.g., [32]), we obtain for the expectation of u the equation

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

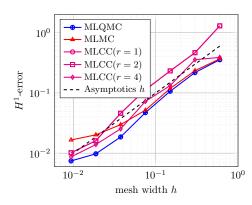Obviously, this equation is solved by $\mathbb{E}[u](\boldsymbol{x}) = (1 - \|\boldsymbol{x}\|_2)^2/6$.

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 $j = 7$). This involves a mesh size of $h_7 = 0.0047$. For the levels $j = 0, \ldots, 6$, the mesh sizes and corresponding 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 MLMC, in order to approximate the root mean square error, we average five realizations of the related approximation error. For the

---

[2]Release 2015a.

TABLE 1

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

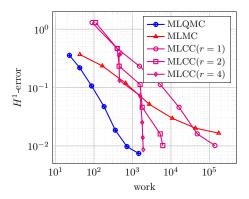| j	0	1	2	3	4	5	6
h_j	0.6	0.3	0.15	0.075	0.0375	0.0188	0.0094
dof_j	27	244	1585	6042	29069	133376	551327

FIG. 4. *H^1-errors of the mean for the different quadrature methods versus mesh width (left) and versus work (right) on the unit ball.*

Clenshaw–Curtis quadrature, we choose the sparse grid quadrature of level

$$(8.2) \qquad q = \left\lfloor \frac{j - \ell}{r} \right\rfloor + q_0,$$

on the spatial discretization level ℓ, where j is the maximum level in the finite element discretization, r is the smoothness parameter from Lemma, 6.4, and $q_0 \in \mathbb{N}$ is an offset. This results in approximately $N_\ell \approx q^5 2^q$ samples; cf. [38].

On the left-hand side of Figure 4, the error of MLQMC, MLMC, and MLCC for the approximation of the mean is depicted with respect to the H^1-norm. For MLCC, the errors for $r = 1, 2, 4$ (cf. (8.2)) are shown. In order to achieve the correct linear rate, we had to choose $q_0 = 2, 3, 4$ correspondingly. All quadrature methods exhibit the expected linear rate. The right-hand side of the same figure shows the error versus work rate, which is expressed in terms of fine grid samples: In accordance with the DoF denoted in Table 1, we scale each sample on a particular level ℓ with the factor $\mathrm{DoF}_\ell / \mathrm{DoF}_j$, i.e., we weight a fine grid sample by 1 and scale the coarse grid samples accordingly. The work is then determined by summing up the total number of samples per level times the related weight.

It can be seen that MLQMC achieves the best error versus work ratio. However, the plot indicates that MLCC for $r = 2, 4$ will outperform MLMQC for higher levels. MLMC provides only half the rate of 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, which was the first module to be launched.[3] Figure 5 shows different tetrahedralizations of this geometry with decreasing mesh size. Note that the geometry can be embedded into a cylinder with radius 0.52 and height 1.58.

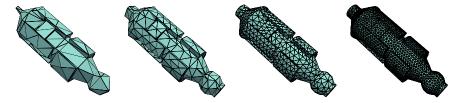[3]We thank Martin Siegel (Rheinbach, Germany), who kindly provided us with this model.

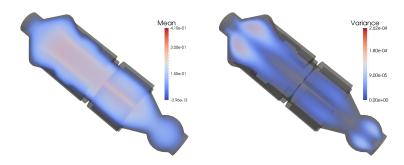FIG. 5. *Tetrahedralizations of four different resolutions for the Zarya geometry.*

FIG. 6. *Mean (left) and variance (right) of the model problem on the Zarya geometry.*

TABLE 2
Mesh sizes and DoF on the different levels for the Zarya geometry.

j	0	1	2	3	4	5	6
h_j	0.5	0.25	0.125	0.0625	0.0313	0.0156	0.0078
dof_j	174	333	1240	5846	30171	141029	617111

In this example, the parametric diffusion coefficient is given by

$$a(\boldsymbol{x}, \boldsymbol{y}) = 1 + \frac{\exp(\|\boldsymbol{x}\|_2^2)}{20}\bigg(\sin(2\pi x_1)y_1 + \frac{1}{2}\sin(2\pi x_2)y_2 + \frac{1}{4}\sin(2\pi x_3)y_3$$
$$+ \frac{1}{8}\sin(4\pi x_1)\sin(4\pi x_2)y_4 + \frac{1}{16}\sin(4\pi x_1)\sin(4\pi x_3)y_5$$
$$+ \frac{1}{32}\sin(4\pi x_2)\sin(4\pi x_3)y_6\bigg)$$

and $f = 10$. For $\boldsymbol{x} \in D$ and $\boldsymbol{y} \in \Gamma$, the diffusion coefficient varies approximately in the range $[0.19, 1.81]$. Figure 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 $j = 0, \dots, 6$, the mesh sizes and corresponding DoF are given in Table 2.

Figure 7 visualizes the approximation errors of the mean with respect to the mesh width (left) and with respect to the work (right), respectively. For MLQMC and MLCC, the number of quadrature points are chosen as in the previous example. For MLCC, the choice $q_0 = 2$ (cf. (8.2)) suffices for $r = 1, 2, 4$. Note that resulting error versus mesh width curves coincide up to the resolution of the plot and are hence represented by a single line. All quadrature methods show a similar convergence rate, which is slightly worse than the optimal rate h. In this example, MLCC for $r = 4$
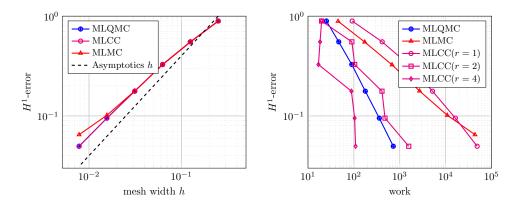
Fig. 7. H^1-errors of the mean for the different quadrature methods versus mesh width (left) and versus work (right) on the Zarya geometry.
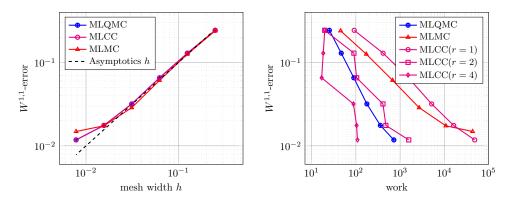

Fig. 8. $W^{1,1}$-errors of the second moment for the different quadrature methods versus mesh width (left) and versus work (right) on the Zarya geometry.

provides the best error versus work rate, while MLQMC performs similarly to MLCC for $r = 2$. Again, MLMC provides the worst error versus work rate.

Finally, Figure 8 shows the respective plots for the second moment, i.e., $\mathbb{E}[u^2]$. Here, the error slightly deteriorates, which might be caused by the interpolation procedure that is used to compute the error. In terms of error versus work, the results are similar to the approximation of the mean; cf. Figure 7.

9. Conclusion. In the present article, we have provided a rigorous analysis of the approximation of curved boundaries and of the numerical quadrature for the bilinear form in the context of multilevel quadrature methods. The obtained results are robust with respect to the parameter dimension m. Our results imply that a level-dependent polygonal approximation of curved domain boundaries is sufficient for computing the finite element solution, while maintaining the overall accuracy of the multilevel method. In addition, we have reversed the construction of the conventional multilevel quadrature. This enables us to give up the nestedness of the spatial approximation spaces. Hence, black-box finite element solvers can be directly applied to compute the solution of the underlying boundary value problem. Moreover, the cost can considerably be reduced by the application of nested quadrature formulas. The theoretical findings have been confirmed by numerical results for the Clenshaw–

Curtis quadrature and the quasi–Monte Carlo quadrature based on Halton points. Of course, other nested quadrature formulas like the Gauss–Patterson quadrature can be used as well. The application of quadrature formulas which are tailored to a possible anisotropy of the integrand is also straightforward. If nonnested quadrature formulas are applied, one arrives at a combination-technique-like representation of the multilevel quadrature.

REFERENCES

[1] I. BABUŠKA, F. NOBILE, AND R. TEMPONE, *A stochastic collocation method for elliptic partial differential equations with random input data*, SIAM J. Numer. Anal., 45 (2007), pp. 1005–1034.

[2] I. BABUŠKA, R. TEMPONE, AND G. ZOURARIS, *Galerkin finite element approximations of stochastic elliptic partial differential equations*, SIAM J. Numer. Anal., 42 (2004), pp. 800–825.

[3] A. BARTH, C. SCHWAB, AND N. ZOLLINGER, *Multi-level Monte Carlo finite element method for elliptic PDEs with stochastic coefficients*, Numer. Math., 119 (2011), pp. 123–161.

[4] J. BECK, R. TEMPONE, F. NOBILE, AND L. TAMELLINI, *On the optimal polynomial approximation of stochastic PDEs by Galerkin and collocation methods*, Math. Models Methods Appl. Sci., 22 (2012), 1250023.

[5] D. BRAESS, *Finite Elements. Theory, Fast Solvers, and Applications in Solid Mechanics*, 2nd ed., Cambridge University Press, Cambridge, UK, 2001.

[6] S. BRENNER AND L. SCOTT, *The Mathematical Theory of Finite Element Methods*, 3rd ed., Springer, New York, 2008.

[7] H.-J. BUNGARTZ AND M. GRIEBEL, *Sparse grids*, Acta Numer., 13 (2004), pp. 147–269.

[8] J. CHARRIER, R. SCHEICHL, AND A. L. TECKENTRUP, *Finite element error analysis of elliptic PDEs with random coefficients and its application to multilevel Monte Carlo methods*, SIAM J. Numer. Anal., 51 (2013), pp. 322–352.

[9] A. COHEN, R. DEVORE, AND C. SCHWAB, *Convergence rates of best N-term Galerkin approximations for a class of elliptic sPDEs*, Found. Comput. Math., 10 (2010), pp. 615–646.

[10] A. COHEN, R. DEVORE, AND C. SCHWAB, *Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDEs*, Anal. Appl., 09 (2011), pp. 11–47.

[11] J. DICK, R. N. GANTNER, Q. T. LE GIA, AND C. SCHWAB, *Multilevel higher-order quasi-Monte Carlo Bayesian estimation*, Math. Models Methods Appl. Sci., 27 (2017), pp. 953–995.

[12] O. ERNST AND B. SPRUNGK, *Stochastic collocation for elliptic PDEs with random data: The lognormal case*, in Sparse Grids and Applications—Munich 2012, J. Garcke and D. Pflüger, eds., Springer, New York, 2014, pp. 29–53.

[13] P. FRAUENFELDER, C. SCHWAB, AND R. TODOR, *Finite elements for elliptic problems with stochastic coefficients*, Comput. Methods Appl. Mech. Engrg., 194 (2005), pp. 205–228.

[14] T. GERSTNER AND M. GRIEBEL, *Numerical integration using sparse grids*, Numer. Algorithms, 18 (1998), pp. 209–232.

[15] T. GERSTNER AND S. HEINZ, *Dimension- and time-adaptive multilevel Monte Carlo methods*, in Sparse Grids and Applications, J. Garcke and M. Griebel, eds., Lect. Notes Comput. Sci. Eng. 88, Springer, New York, 2012, pp. 107–120.

[16] R. GHANEM AND P. SPANOS, *Stochastic Finite Elements. A Spectral Approach*, Springer, New York, 1991.

[17] M. GILES, *Multilevel Monte Carlo path simulation*, Oper. Res., 56 (2008), pp. 607–617.

[18] M. GILES, *Multilevel Monte Carlo methods*, Acta Numer., 24 (2015), pp. 259–328.

[19] M. GILES AND B. WATERHOUSE, *Multilevel quasi-Monte Carlo path simulation*, Radon Ser. Comput. Appl. Math., 8 (2009), pp. 1–18.

[20] M. GRIEBEL AND H. HARBRECHT, *On the construction of sparse tensor product spaces*, Math. Comp., 82 (2013), pp. 975–994.

[21] A.-L. HAJI-ALI, H. HARBRECHT, M. PETERS, AND M. SIEBENMORGEN, *Novel results for the anisotropic sparse grid quadrature*, J. Complexity, 47 (2018), pp. 62–85.

[22] A.-L. HAJI-ALI, F. NOBILE, L. TAMELLINI, AND R. TEMPONE, *Multi-index stochastic collocation convergence rates for random PDEs with parametric regularity*, Found. Comput. Math., 16 (2016), pp. 1555–1605.

[23] A.-L. HAJI-ALI, F. NOBILE, E. VON SCHWERIN, AND R. TEMPONE, *Optimization of mesh hierarchies in multilevel Monte Carlo samplers*, Stoch. Partial Differ. Equ. Anal. Comput., 4 (2016), pp. 76–112.

[24] J. HALTON, *On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals*, Numer. Math., 2 (1960), pp. 84–90.

[25] H. HARBRECHT, M. PETERS, AND M. SIEBENMORGEN, *On multilevel quadrature for elliptic stochastic partial differential equations*, in Sparse Grids and Applications, J. Garcke and M. Griebel, eds., Lect. Notes Comput. Sci. Eng. 88, Springer, New York, 2012, pp. 161–179.

[26] H. HARBRECHT, M. PETERS, AND M. SIEBENMORGEN, *Efficient approximation of random fields for numerical applications*, Numer. Linear Algebra Appl., 22 (2015), pp. 596–617.

[27] H. HARBRECHT, M. PETERS, AND M. SIEBENMORGEN, *Analysis of the domain mapping method for elliptic diffusion problems on random domains*, Numer. Math., 134 (2016), pp. 823–856.

[28] H. HARBRECHT, M. PETERS, AND M. SIEBENMORGEN, *Multilevel accelerated quadrature for PDEs with log-normally distributed diffusion coefficient*, SIAM/ASA J. Uncertain. Quantif., 4 (2016), pp. 520–551.

[29] H. HARBRECHT, M. PETERS, AND M. SIEBENMORGEN, *On the quasi-Monte Carlo method with Halton points for elliptic PDEs with log-normal diffusion*, Math. Comp., 86 (2017), pp. 771–797.

[30] S. HEINRICH, *The multilevel method of dependent tests*, in Advances in Stochastic Simulation Methods (St. Petersburg, 1998), Stat. Ind. Technol., Birkhäuser, Boston, MA, 2000, pp. 47–61.

[31] S. HEINRICH, *Multilevel Monte Carlo Methods*, in Large-Scale Scientific Computing, Lecture Notes in Comput. Sci. 2179, Springer, New York, 2001, pp. 58–67.

[32] E. HILLE AND R. PHILLIPS, *Functional Analysis and Semi-Groups*, Amer. Math. Soc. Colloq. Publ. 31, AMS, Providence, RI, 1957.

[33] V. HOANG AND C. SCHWAB, *N-term Wiener chaos approximation rate for elliptic PDEs with lognormal Gaussian random inputs*, Math. Models Methods Appl. Sci., 4 (2014), p. 797–826.

[34] F. KUO, C. SCHWAB, AND I. SLOAN, *Multi-level quasi-Monte Carlo finite element methods for a class of elliptic partial differential equations with random coefficients*, Found. Comput. Math., 15 (2015), pp. 411–449.

[35] M. LOÈVE, *Probability Theory*, 4th ed., Grad. Texts in Math. 45, Springer, New York, 1977.

[36] H. MATTHIES AND A. KEESE, *Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations*, Comput. Methods Appl. Mech. Engrg., 194 (2005), pp. 1295–1331.

[37] H. NIEDERREITER, *Random Number Generation and Quasi-Monte Carlo Methods*, CBMS-NSF Regional Conf. Ser. in Appl. Math. 63, SIAM, Philadelphia, 1992.

[38] E. NOVAK AND K. RITTER, *High dimensional integration of smooth functions over cubes*, Numer. Math., 75 (1996), pp. 79–97.

[39] C. SCHWAB AND R. TODOR, *Karhunen-Loève approximation of random fields by generalized fast multipole methods*, J. Comput. Phys., 217 (2006), pp. 100–122.

[40] M. SIEBENMORGEN, *Quadrature Methods for Elliptic PDEs with Random Diffusion*, Ph.D. thesis, Faculty of Science, University of Basel, 2015.

[41] A. TECKENTRUP, R. SCHEICHL, M. GILES, AND E. ULLMANN, *Further analysis of multilevel Monte Carlo methods for elliptic PDEs with random coefficients*, Numer. Math., 125 (2013), pp. 569–600.

[42] A. L. TECKENTRUP, P. JANTSCH, C. G. WEBSTER, AND M. GUNZBURGER, *A multilevel stochastic collocation method for partial differential equations with random input data*, SIAM/ASA J. Uncertain. Quantif., 3 (2015), pp. 1046–1074.

[43] R. TODOR AND C. SCHWAB, *Convergence rates for sparse chaos approximations of elliptic problems with stochastic coefficients*, IMA J. Numer. Anal., 27 (2007), pp. 232–261.

[44] X. WANG, *A constructive approach to strong tractability using quasi-Monte Carlo algorithms*, J. Complexity, 18 (2002), pp. 683–701.