On multilevel quadrature for elliptic stochastic partial differential equations

Helmut Harbrecht, Michael Peters,
Markus Siebenmorgen

Department of Mathematics
University of Basel
Rheinsprung 21
CH - 4051 Basel
Switzerland

Preprint No. 2011-01
July 11

www.math.unibas.ch

On multilevel quadrature for elliptic stochastic partial differential equations

Helmut Harbrecht, Michael Peters, and Markus Siebenmorgen

Abstract In the present article, we show that the multilevel Monte Carlo method for elliptic stochastic partial differential equations can be interpreted as a sparse grid approximation. By using this interpretation, the method can straightforwardly be generalized to any given quadrature rule for high dimensional integrals like the quasi Monte Carlo method or the polynomial chaos approach. Besides the multilevel quadrature for approximating the solution's expectation, a simple and efficient modification of the approach is proposed to compute the stochastic solution's variance. Numerical results are provided to demonstrate and quantify the approach.

1 Introduction

The present article is dedicated to the numerical solution of elliptic second order boundary value problems with stochastic diffusion coefficient. A principal approach to solve such stochastic boundary value problems is the Monte Carlo approach, see e.g. [15] and the references therein. However, it is extremely expensive to generate a large number of suitable samples and to solve a deterministic boundary value problem on each sample. To overcome this obstruction, the multilevel Monte Carlo method (MLMC) has been developed in [1, 7, 8, 11, 12]. From the stochastic point of view, it is a variance reduction technique which considerably decreases the complexity. The idea is to combine the Monte Carlo quadrature of the stochastic variable with a multilevel splitting of the Bochner space which contains the random solution. Then, to compute the solution's statistics, most samples can be performed on coarse spatial discretizations while only a few samples must be performed on fine spatial discretizations. As we will show, this proceeding can be interpreted as a sparse grid approximation of the expectation and variance. By replacing the Monte Carlo

Helmut Harbrecht · Michael Peters · Markus Siebenmorgen
Mathematisches Institut, Rheinsprung 21, 4051 Basel, Switzerland,
e-mail: {helmut.harbrecht,michael.peters,markus.siebenmorgen}@unibas.ch

quadrature by another quadrature rule for high dimensional integrals yields e.g. the multilevel quasi Monte Carlo method (MLQMC) or the multilevel polynomial chaos method (MLPC).

2 Sparse grids

Let

$$V_0^{(i)} \subset V_1^{(i)} \subset \cdots \subset V_j^{(i)} \subset \cdots \subset \mathscr{H}_i, \quad i = 1,2,$$

denote two sequences of nested, finite dimensional spaces with increasing approximation properties in some abstract spaces $\mathscr{H}_i$. To approximate a given object of the space $\mathscr{H}_1 \times \mathscr{H}_2$ one canonically considers the *full tensor product spaces* $U_j := V_j^{(1)} \otimes V_j^{(2)}$. However, the cost $\dim U_j = \dim V_j^{(1)} \cdot \dim V_j^{(2)}$ are often too expensive. This drawback can be avoided by considering the approximation in so-called *sparse grid* or *sparse tensor product spaces*. To this end, one defines for $j \geq 1$ the complement spaces

$$W_{j+1}^{(i)} = V_{j+1}^{(i)} \ominus V_j^{(i)}, \quad i = 1,2,$$

and gains the multilevel decompositions

$$V_j^{(i)} = \bigoplus_{i=0}^{j} W_j^{(i)}, \quad W_0^{(i)} := V_0^{(i)}, \quad i = 1,2. \tag{1}$$

Thus, the sparse grid space is defined by

$$\widehat{U}_j := \sum_{\ell+\ell' \leq j} W_\ell^{(1)} \otimes W_{\ell'}^{(2)} \tag{2}$$

It contains, up to logarithmic factors, only $\max\{\dim V_j^{(1)}, \dim V_j^{(2)}\}$[1] degrees of freedom but offers nearly the same approximation power as U_j provided that the object to be approximated offers some extra smoothness by means of mixed regularity.

Sparse grids are used for the approximation of high-dimensional data or functions [10] and for the numerical solution of partial differential equations [18]. Another application issues from the construction of quadrature formulae [5]. For all the details we refer the reader to the survey [2] and the references therein.

In the present context of stochastic partial differential equations, the sequence $\{V_j^{(1)}\}$ will correspond to a sequence of Bochner spaces which map the high dimensional unit cube $\square$ onto nested finite element spaces. Whereas, the sequence $\{V_j^{(2)}\}$ will correspond to a sequence of quadrature rules which compute the Bochner inte-

[1] This holds under the assumptions that the dimensions of $V_j^{(1)}$ and $V_j^{(2)}$ scale like geometric sequences.

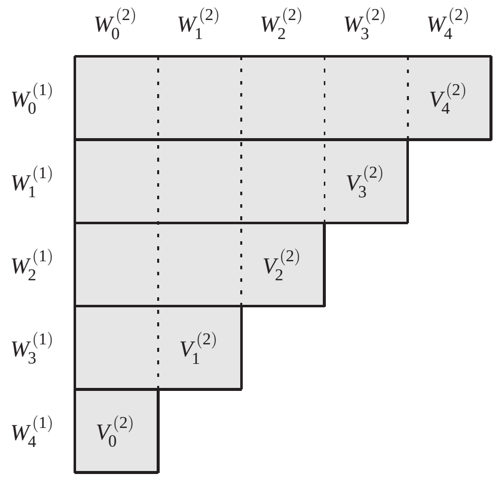

Fig. 1 Different representations of the sparse grid space $\widehat{U}_j$.

gral. To arrive at the *multilevel quadrature method* we will make use of the following equivalent representation of the sparse grid space (2). In view of (1), one can rewrite (2) according to

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

(see Fig. 1 for an illustration). This makes the specification of the complement spaces $W_j^{(2)}$ obsolete. Indeed, even the nestedness of the quadrature points can be neglected since it is sufficient that the error of quadrature decreases as the dimension of $V_j^{(2)}$ increases.

The rest of this article is organized as follows. We introduce the elliptic model problem of interest in Section 3 and transform it to a parametric diffusion problem in Section 4. The next two sections are dedicated to the discretization, namely the quadrature rule for the stochastic variable (Section 5) and the finite element method with respect to the physical domain (Section 6). The multilevel quadrature for the random solution's expectation is proposed in Section 7. The computation of its variance requires some modifications which are carried out in Section 8. Finally, in Section 9, we provide numerical results to demonstrate and quantify the approach.

3 Problem setting

In the following, let $D \subset \mathbb{R}^n$ for $n = 2,3$ be a polygonal or polyhedral domain and let (Ω, Σ, P) be a probability space with σ-field $\Sigma \subset 2^{\Omega}$ and a complete probability measure P, i.e., for all $A \subset B$ and $B \in \Sigma$ with $P[B] = 0$ it follows $A \in \Sigma$ and $P[A] = 0$. We intend to compute the expectation $\mathbb{E}_u := \mathbb{E}(u)$ and the variance $\mathbb{V}_u := \mathbb{V}(u)$ of the random function $u(\omega) \in H_0^1(D)$ which solves for almost all $\omega \in \Omega$ the stochastic diffusion problem

$$-\operatorname{div}\big(\alpha(\omega)\nabla u(\omega)\big) = f \text{ in } D. \tag{3}$$

To guarantee unique solvability of (3), we assume that $\alpha(\omega)$ is almost surely uniformly elliptic and bounded

$$0 < \alpha_{\min} \leq \alpha(\omega) \leq \alpha_{\max} < \infty \text{ in } D. \tag{4}$$

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

$$\alpha(\mathbf{x}, \omega) = \mathbb{E}_\alpha(\mathbf{x}) + \sum_{k=1}^{m} \sqrt{\lambda_k}\varphi_k(\mathbf{x})\psi_k(\omega) \tag{5}$$

with pairwise 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 (5) from the given covariance kernel

$$\operatorname{Cov}_\alpha(\mathbf{x}, \mathbf{y}) = \int_\Omega \big\{\alpha(\mathbf{x}, \omega) - \mathbb{E}_\alpha(\mathbf{x})\big\}\big\{\alpha(\mathbf{y}, \omega) - \mathbb{E}_\alpha(\mathbf{y})\big\} \,\mathrm{d}P(\omega).$$

If it contains infinitely many terms, it is appropriately truncated which will induce an additional discretization error. For all the details we refer the reader to [6, 13, 16].

4 Deterministic reformulation

The assumption that the random variables $\{\psi_k(\omega)\}$ are stochastically independent implies that the respective joint density function and the joint distribution of the random variables are given by

$$\rho(\mathbf{y}) := \prod_{k=1}^{m} \rho_k(y_k) \quad \text{and} \quad \mathrm{d}P_\rho(\mathbf{y}) := \rho(\mathbf{y})\mathrm{d}\mathbf{y}.$$

Now we are able to reformulate the stochastic problem (3) by a parametric deterministic problem by replacing the space $L_P^2(\Omega)$ by $L_\rho^2(\square)$ where $\square := [-1,1]^m$. Thus, the stochastic space Ω is identified with its image $\square$ with respect to the measurable mapping

$$\psi : \Omega \to \square, \quad \omega \mapsto \psi(\omega) := \big(\psi_1(\omega), \dots, \psi_M(\omega)\big).$$

Hence, the random variables ψ_k are substituted by coordinates $y_k \in [-1,1]$. With this construction at hand, we define the parameterized and truncated diffusion coefficient $\alpha : D \times \square \to \mathbb{R}$ via

$$\alpha(\mathbf{x},\mathbf{y}) = \mathbb{E}_\alpha(\mathbf{x}) + \sum_{k=1}^{m} \sqrt{\lambda_k}\varphi_k(\mathbf{x}) y_k$$

for all $\mathbf{x} \in D$ and $\mathbf{y} = (y_1, y_2, \dots, y_m) \in \square$. This leads to the following parametric diffusion problem:

$$\begin{aligned} &\text{find } u \in L^2_\rho(\square; H^1_0(D)) \text{ such that} \\ &-\operatorname{div}\big(\alpha(\mathbf{y})\nabla u(\mathbf{y})\big) = f \text{ in } D \text{ for all } \mathbf{y} \in \square. \end{aligned} \tag{6}$$

Here and in the sequel, for a given Banach space X, the Bochner space $L^p_\rho(\square;X)$, $1 \le p \le \infty$, consists of all functions $v : \square \to X$ whose norm

$$\|v\|_{L^p_\rho(\square;X)} := \begin{cases} \left(\int_\square \|v(\mathbf{y})\|_X^p \rho(\mathbf{y}) \,\mathrm{d}\mathbf{y}\right)^{1/p}, & p < \infty \\ \operatorname*{ess\,sup}_{\mathbf{y}\in\square} \|v(\mathbf{y})\rho(\mathbf{y})\|_X, & p = \infty \end{cases}$$

is finite. If $p = 2$ and X is a Hilbert space, the Bochner space is isomorphic to the tensor product space $L^2_\rho(\square) \otimes X$. Finally, the Bochner space $C(\square;X)$ consists of all continuous mappings $v : \square \to X$.

By the Lax-Milgram theorem, unique solvability of the parametric diffusion problem (6) in $L^2_\rho\big(\square; H^1_0(D)\big)$ follows immediately from the condition (4) on the stochastic diffusion coefficient. In [17], it has further been proven that the solution u of (6) is analytical as mapping $u : \square \to H^1_0(D)$. At least in case of uniformly distributed random variables $\{\psi_k\}$, it is even analytical as mapping $u : \square \to H^1_0(D) \cap H^2(D)$, provided that the functions $\{\varphi_k\}$ in the Karhunen-Loève expansion (5) are in $W^{1,\infty}(D)$, see [3].

5 Quadrature in the stochastic variable

Having the solution $u \in L^2_\rho\big(\square; H^1_0(D)\big)$ of (6) at hand, then its expectation and variance are given by the high dimensional integrals

$$\mathbb{E}_u(\mathbf{x}) = \int_\square u(\mathbf{x},\mathbf{y})\rho(\mathbf{y}) \,\mathrm{d}\mathbf{y} \in H^1_0(D) \tag{7}$$

and

$$\mathbb{V}_u(\mathbf{x}) = \mathbb{E}_{u^2}(\mathbf{x}) - \mathbb{E}_u^2(\mathbf{x}) = \int_\square u^2(\mathbf{x},\mathbf{y})\rho(\mathbf{y})\,\mathrm{d}\mathbf{y} - \mathbb{E}_u^2(\mathbf{x}) \in W_0^{1,1}(D). \tag{8}$$

To compute these integrals, we shall provide a sequence of quadrature formulae $\{Q_\ell\}$ for the Bochner integral

$$I : L_\rho^1(\square;X) \to X, \quad Iv = \int_\square v(\cdot,\mathbf{y})\rho(\mathbf{y})\,\mathrm{d}\mathbf{y}$$

where $X \subset L^2(D)$ denotes a Banach space. The quadrature formula

$$Q_\ell : L_\rho^1(\square;X) \to X, \quad Q_\ell v = \sum_{i=1}^{N_\ell} \omega_{\ell,i} v(\cdot,\xi_{\ell,i})\rho(\xi_{\ell,i}) \tag{9}$$

is supposed to provide the error bound

$$\|(I - Q_\ell)v\|_X \lesssim \varepsilon_\ell \|v\|_{\mathscr{H}(\square;X)} \tag{10}$$

uniformly in $\ell \in \mathbb{N}$, where $\{\varepsilon_\ell\}$ is a null sequence and $\mathscr{H}(\square;X) \subset L_\rho^2(\square,X)$ a suitable Bochner space. For our purpose, we assume that the number of points N_ℓ of the quadrature formula Q_ℓ are chosen such that

$$\varepsilon_\ell = 2^{-\ell}. \tag{11}$$

The following particular examples of quadrature rules (9) are considered in our numerical experiments:

- In the mean, the Monte Carlo method satisfies (10) with $\varepsilon_\ell = N_\ell^{-1/2}$ and $\mathscr{H}(\square;X) = L_\rho^2(\square;X)$.
- The standard quasi Monte Carlo method lead typically to $\varepsilon_\ell = N_\ell^{-1}(\log N_\ell)^m$ and the Bochner space $\mathscr{H}(\square;X) = W_{\mathrm{mix}}^{1,1}(\square;X)$ which consists of all functions $v : \square \to X$ with finite norm

$$\|v\|_{W_{\mathrm{mix}}^{1,1}(\square;X)} := \sum_{\|\mathbf{q}\|_\infty \leq 1} \int_\square \left\| \frac{\partial^{\|\mathbf{q}\|_1}}{\partial y_1^{q_1} \partial y_2^{q_2} \cdots \partial y_m^{q_m}} v(\mathbf{y}) \right\|_X \mathrm{d}\mathbf{y} < \infty, \tag{12}$$

  see e.g. [14]. Note that this estimate requires that the densities satisfy $\rho_k \in W^{1,\infty}([-1,1])$.
- If $v : \square \to X$ is analytical with respect to the variable $\mathbf{y}$, then one can apply a tensor product Gaussian quadrature rule, yielding an exponential convergence rate $\varepsilon_\ell = \exp(-bN_\ell^{1/m})$ and $\mathscr{H}(\square;X) = L^\infty(\square;X)$. In fact, the polynomial chaos approach introduced in [4] can be interpreted as a dimension weighted tensor product Gaussian quadrature rule (with weights depending on the numbers $\{\sqrt{\lambda_k}\|\varphi_k\|_{L^\infty(D)}\}$).

6 Finite element approximation in the spatial variable

In order to apply the quadrature formula (9), we shall calculate the solution $u(\mathbf{y}) \in H_0^1(D)$ of the diffusion problem (6) in certain points $\mathbf{y} \in \square$. To this end, consider a coarse grid triangulation/tetrahedralization $\mathscr{T}_0 = \{\tau_{0,k}\}$ of the domain D. Then, for $\ell \geq 1$, a uniform and shape regular triangulation/tetrahedralization $\mathscr{T}_\ell = \{\tau_{\ell,k}\}$ is recursively obtained by uniformly refining each triangle/tetrahedron $\tau_{\ell-1,k}$ into 2^n triangles/tetrahedrons with diameter $h_\ell \sim 2^{-\ell}$.

Define for $p \geq 1$ the finite element spaces

$$\mathscr{S}_\ell^p(D) := \{v \in C(D) : v|_{\partial D} = 0 \text{ and } v|_\tau \in \mathscr{P}_p \text{ for all } \tau \in \mathscr{T}_\ell\} \subset H_0^1(D),$$

where $\mathscr{P}_p$ denotes the space of all polynomials of total degree $p \geq 0$. Then, the solution $u(\mathbf{y}) \in \mathscr{S}_\ell^p(D)$ of a finite element method in the space $\mathscr{S}_\ell^p(D)$ admits the following approximation properties. [2]

Lemma 1. *Let the domain D be convex and $f \in L^2(D)$. Then, the finite element solution $u_\ell(\mathbf{y}) \in \mathscr{S}_\ell^p(D)$ of the diffusion problem* (6) *satisfies the error estimates*

$$\|u(\mathbf{y}) - u_\ell(\mathbf{y})\|_{H^1(D)} \lesssim 2^{-\ell}\|u(\mathbf{y})\|_{H^2(D)}, \tag{13}$$

and

$$\|u^2(\mathbf{y}) - u_\ell^2(\mathbf{y})\|_{W^{1,1}(D)} \lesssim 2^{-\ell}\|u(\mathbf{y})\|_{H^2(D)}^2. \tag{14}$$

Here, the constants hidden in (13) *and* (14) *depend on $\alpha_{\min}$ and $\alpha_{\max}$, but not on $\mathbf{y} \in \square$.*

Proof. The parametric diffusion problem (6) is H^2-regular since D is convex and $f \in L^2(D)$. Hence, the first error estimate immediately follows from the standard finite element theory. We further find

$$\begin{aligned}\|u^2(\mathbf{y}) - u_\ell^2(\mathbf{y})\|_{W^{1,1}(D)} &= \big\|\big(u(\mathbf{y}) - u_\ell(\mathbf{y})\big)\big(u(\mathbf{y}) + u_\ell(\mathbf{y})\big)\big\|_{W^{1,1}(D)} \\ &\lesssim \|u(\mathbf{y}) - u_\ell(\mathbf{y})\|_{H^1(D)}\|u(\mathbf{y}) + u_\ell(\mathbf{y})\|_{L^2(D)} \\ &\quad + \|u(\mathbf{y}) - u_\ell(\mathbf{y})\|_{L^2(D)}\|u(\mathbf{y}) + u_\ell(\mathbf{y})\|_{H^1(D)}.\end{aligned}$$

By using

$$\|u_\ell(\mathbf{y})\|_{H^1(D)} \leq \|u(\mathbf{y})\|_{H^1(D)} + \|u(\mathbf{y}) - u_\ell(\mathbf{y})\|_{H^1(D)} \lesssim (1 + 2^{-\ell})\|u(\mathbf{y})\|_{H^2(D)}$$

and the corresponding estimate in $L^2(D)$, we arrive at the desired estimate (14). ■

[2] Error estimates in respectively $L^2(D)$ and $L^1(D)$ are derived by straightforward modifications, yielding the convergence rate $4^{-\ell}$. Then, the error analysis of the multilevel quadrature can be performed with respect to these norms, provided that the precision of the quadrature (10) is also chosen as $\varepsilon_\ell = 4^{-\ell}$.

7 Multilevel quadrature method for the expectation

We now have to combine the quadrature method with the multilevel finite element discretization. To this end, we define the ansatz spaces

$$V_j^{(1)} := \big\{G_j(\mathbf{y})v(\mathbf{x},\mathbf{y}) : v \in C\big(\square; H_0^1(D)\big) \text{ and } \mathbf{y} \in \square\big\} \subset L_\rho^2\big(\square; \mathscr{S}_j^p(D)\big). \tag{15}$$

Herein, $G_j(\mathbf{y})$ denotes the Galerkin projection

$$G_j(\mathbf{y}) : H_0^1(D) \to \mathscr{S}_j^p(D), \quad v \mapsto v_j,$$

defined via the Galerkin orthogonality

$$\int_D \alpha(\mathbf{x},\mathbf{y})\nabla\big(v(\mathbf{x}) - v_j(\mathbf{x})\big)\nabla w_j(\mathbf{x})\,\mathrm{d}\mathbf{x} = 0 \text{ for all } w_j \in \mathscr{S}_j^p(D).$$

Note that this bilinear form issues from the weak formulation of (6) in $H_0^1(D)$.[3]

To compute the expectation (7) we shall apply the quadrature rule Q_j to the finite element solution in $\mathscr{S}_j^p(D)$ which yields

$$\mathbb{E}_u(\mathbf{x}) \approx Q_j\big(G_j(\mathbf{y})u(\mathbf{x},\mathbf{y})\big) = \sum_{i=0}^{N_j} \omega_{j,i} G_j(\xi_{j,i}) u(\mathbf{x},\xi_{j,i})\rho(\xi_{j,i}). \tag{16}$$

This can be interpreted as the *full* tensor approximation of the function $\mathbb{E}_u$ in the product space $V_j^{(1)} \otimes V_j^{(2)}$ where the quadrature rule Q_j serves as “space” $V_j^{(2)}$.

In contrast to this, setting $G_{-1}(\mathbf{y}) := 0$ for all $\mathbf{y} \in \square$, the *sparse* tensor product $\sum_{\ell=0}^{j} W_\ell^{(1)} \otimes V_{j-\ell}^{(2)}$ is performed with the help of the complement spaces

$$W_\ell^{(1)} := \big\{\big(G_\ell(\mathbf{y}) - G_{\ell-1}(\mathbf{y})\big)v(\mathbf{x},\mathbf{y}) : v \in C\big(\square; H_0^1(D)\big) \text{ and } \mathbf{y} \in \square\big\} \subset V_\ell^{(1)}.$$

Namely, we consider the sparse tensor approximation

$$\begin{aligned}\mathbb{E}_u(\mathbf{x}) &\approx \sum_{\ell=0}^{j} Q_{j-\ell}\big(G_\ell(\mathbf{y})u(\mathbf{x},\mathbf{y}) - G_{\ell-1}(\mathbf{y})u(\mathbf{x},\mathbf{y})\big) \\ &= \sum_{\ell=0}^{j}\sum_{i=0}^{N_{j-\ell}} \omega_{j-\ell,i}\big(G_\ell(\xi_{j-\ell,i})u(\mathbf{x},\xi_{j-\ell,i}) - G_{\ell-1}(\xi_{j-\ell,i})u(\mathbf{x},\xi_{j-\ell,i})\big)\rho(\xi_{j-\ell,i}).\end{aligned} \tag{17}$$

[3] There holds the identity $V_j^{(1)} = P_j\big(L_\rho^2\big(\square; H_0^1(D)\big)\big)$ where $P_j : L_\rho^2\big(\square; H_0^1(D)\big) \to L_\rho^2\big(\square; \mathscr{S}_j^p(D)\big)$ is the Galerkin projection with respect to the bilinear form $A : L_\rho^2\big(\square; H_0^1(D)\big) \times L_\rho^2\big(\square; H_0^1(D)\big) \to \mathbb{R}$ given by

$$A(v,w) := \int_\square \int_D \alpha(\mathbf{x},\mathbf{y})\nabla_\mathbf{x} v(\mathbf{x},\mathbf{y})\nabla_\mathbf{x} w(\mathbf{x},\mathbf{y})\rho(\mathbf{y})\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y}.$$

This bilinear form stems from the weak formulation of the parametric diffusion problem (6) in the Bochner space $L_\rho^2\big(\square; H_0^1(D)\big)$. The equivalence of this *weak definition* to the pointwise definition (15) follows immediately from the analyticity of the solutions to (6) in the parameter $\mathbf{y}$.

Loosely speaking, the function $u \in L^2_\rho\big(\square; H^1_0(D)\big)$ is divided into j slices in accordance with the modulus of its entity. Then, for every slice, the precision of quadrature is properly chosen. We refer to Fig. 2 for a graphical illustration.

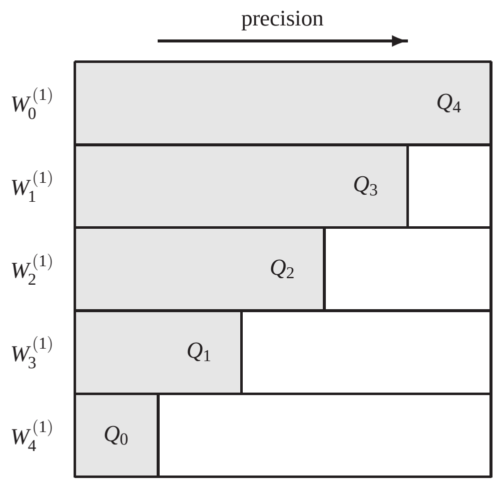

Fig. 2 Visualization of the multilevel quadrature.

Theorem 1. *Let $u \in \mathscr{H}\big(\square; H^1_0(D) \cap H^2(D)\big)$ and let $\{Q_\ell\}$ be a sequence of quadrature rules which satisfy* (10) *and* (11)*. Then, it holds*

$$\bigg\|\mathbb{E}_u(\mathbf{x}) - \sum_{\ell=0}^{j} Q_{j-\ell}\big(G_\ell(\mathbf{y})u(\mathbf{x},\mathbf{y}) - G_{\ell-1}(\mathbf{y})u(\mathbf{x},\mathbf{y})\big)\bigg\|_{H^1(D)} \lesssim 2^{-j} j \|u\|_{\mathscr{H}(\square;H_2(D))}.$$

Proof. Since $u \in \mathscr{H}\big(\square; H^1_0(D) \cap H^2(D)\big)$, the estimate (13) implies the decay

$$\|G_\ell(\mathbf{y})u(\mathbf{x},\mathbf{y}) - G_{\ell-1}(\mathbf{y})u(\mathbf{x},\mathbf{y})\|_{\mathscr{H}(\square;H^1_0(D))} \lesssim 2^{-\ell}\|u\|_{\mathscr{H}(\square;H^2(D))}.$$

This, together with (10), (11), and (13), yields

$$\begin{aligned}
&\left\|\mathbb{E}_u(\mathbf{x}) - \sum_{\ell=0}^{j} Q_{j-\ell}\big(G_\ell(\mathbf{y})u(\mathbf{x},\mathbf{y}) - G_{\ell-1}(\mathbf{y})u(\mathbf{x},\mathbf{y})\big)\right\|_{H^1(D)} \\
&\quad\lesssim \big\|\mathbb{E}_u(\mathbf{x}) - I\big(G_j(\mathbf{y})u(\mathbf{x},\mathbf{y})\big)\big\|_{H^1(D)} \\
&\qquad\qquad + \sum_{\ell=0}^{j}\left\|(I-Q_{j-\ell})\big(G_\ell(\mathbf{y})u(\mathbf{x},\mathbf{y}) - G_{\ell-1}(\mathbf{y})u(\mathbf{x},\mathbf{y})\big)\right\|_{H^1(D)} \\
&\quad\lesssim 2^{-j}\|u\|_{L^2_\rho(\square;H^2(D))} + \sum_{\ell=0}^{j} 2^{\ell-j}2^{-\ell}\|u\|_{\mathscr{H}(\square;H^2(D))} \\
&\quad\lesssim 2^{-j} j \|u\|_{\mathscr{H}(\square;H^2(D))}.
\end{aligned}$$

■

Remark 1. The convergence rate of the full tensor approximation (16) is $2^{-\ell}$. Therefore, the convergence rate of the sparse tensor approximation (17) is only the logarithmic factor j smaller. This factor can be removed, if the precision of quadrature ε_ℓ is chosen as $\ell^{-(1+\delta)}2^{-\ell}$ for some $\delta > 0$ since

$$\sum_{\ell=0}^{j} 2^{\ell-j}2^{-\ell}(j-\ell)^{-(1+\delta)} = 2^{-j}\sum_{\ell=0}^{j}\ell^{-(1+\delta)} \lesssim 2^{-j}.$$

This choice was proposed in [1].

8 Multilevel quadrature method for the second moment

The determination of the random solution's variance (8) requires the computation of its second moment $\mathbb{E}_{u^2} \in W_0^{1,1}(D)$ which can be performed similarly to the expectation. The full tensor approximation

$$\mathbb{E}_{u^2}(\mathbf{x}) \approx Q_j\big(G_j(\mathbf{y})u(\mathbf{x},\mathbf{y})\big)^2 = \sum_{i=0}^{N_j}\omega_i\big(G_j(\xi_{j,i})u(\mathbf{x},\xi_{j,i})\big)^2\rho(\xi_{j,i})$$

corresponds to the approximation in the product "space" $\widetilde{V}_j^{(1)} \otimes V_j^{(2)}$ with

$$\widetilde{V}_j^{(1)} := \big\{\big(G_j(\mathbf{y})v(\mathbf{x},\mathbf{y})\big)^2 : v \in C\big(\square;H_0^1(D)\big) \text{ and } \mathbf{y}\in\square\big\} \subset L^2_\rho\big(\square;\mathscr{S}_j^{2p}(D)\big).$$

To define the related sparse tensor approximation, we set

$$\begin{aligned}
\widetilde{W}_\ell^{(1)} := \big\{&\big(G_\ell(\mathbf{y})v(\mathbf{x},\mathbf{y})\big)^2 - \big(G_{\ell-1}(\mathbf{y})v(\mathbf{x},\mathbf{y})\big)^2 : \\
&\qquad v \in C\big(\square;H_0^1(D)\big) \text{ and } \mathbf{y}\in\square\big\} \subset \widetilde{V}_j^{(1)}.
\end{aligned}$$

Then, the approximation of $\mathbb{E}_{u^2}$ in the sparse product "space" $\sum_{\ell=0}^{j}\widetilde{W}_{j-\ell}^{(1)}\otimes V_j^{(2)}$ is given as

$$\begin{aligned}\mathbb{E}_{u^2}(\mathbf{x}) &\approx \sum_{\ell=0}^{j} Q_{j-\ell}\Big(\big(G_\ell(\mathbf{y})u(\mathbf{x},\mathbf{y})\big)^2-\big(G_{\ell-1}(\mathbf{y})u(\mathbf{x},\mathbf{y})\big)^2\Big)\\ &= \sum_{\ell=0}^{j}\sum_{i=1}^{N_{j-\ell}}\omega_{j-\ell,i}\Big(\big(G_\ell(\xi_{j-\ell,i})u(\mathbf{x},\xi_{j-\ell,i})\big)^2\\ &\qquad\qquad -\big(G_{\ell-1}(\xi_{j-\ell,i})u(\mathbf{x},\xi_{j-\ell,i})\big)^2\Big)\rho(\xi_{j-\ell,i}).\end{aligned}$$

To estimate the discretization error of this sparse tensor approximation, we will need to further estimate the expression

$$\begin{aligned}\Big\|(I-Q_{j-\ell})\Big(\big(G_\ell(\mathbf{y})u(\mathbf{x},\mathbf{y})\big)^2-\big(G_{\ell-1}(\mathbf{y})u(\mathbf{x},\mathbf{y})\big)^2\Big)\Big\|_{W^{1,1}(D))}\\ \lesssim 2^{j-\ell}\Big\|\big(G_\ell(\mathbf{y})u(\mathbf{x},\mathbf{y})\big)^2-\big(G_{\ell-1}(\mathbf{y})u(\mathbf{x},\mathbf{y})\big)^2\Big\|_{\mathscr{H}(\square;W^{1,1}(D))}.\end{aligned}\tag{18}$$

If $u\in\mathscr{H}\big(\square;H_0^1(D)\cap H^2(D)\big)$, then this is clearly possible by using (14) but requires some further specification of the Bochner space $\mathscr{H}(\square)$. For example, if $\mathscr{H}(\square)=L_\rho^2(\square)$, we arrive at

$$\big\|\big(G_\ell(\mathbf{y})u(\mathbf{x},\mathbf{y})\big)^2-\big(G_{\ell-1}(\mathbf{y})u(\mathbf{x},\mathbf{y})\big)^2\big\|_{L_\rho^2(\square;W^{1,1}(D))}\lesssim 2^{-\ell}\|u\|_{L_\rho^4(\square;H^2(D))}^2.$$

Whereas, if $\mathscr{H}(\square)=W_{\text{mix}}^{1,1}(\square)$, one gets, in view of (12), the estimate

$$\big\|\big(G_\ell(\mathbf{y})u(\mathbf{x},\mathbf{y})\big)^2-\big(G_{\ell-1}(\mathbf{y})u(\mathbf{x},\mathbf{y})\big)^2\big\|_{W_{\text{mix}}^{1,1}(\square;W^{1,1}(D))}\lesssim 2^{-\ell}\|u\|_{H_{\text{mix}}^1(\square;H^2(D))}^2,$$

where the norm in $H_{\text{mix}}^1(\square;X)$ is defined in complete analogy to (12). Finally, in case of $\mathscr{H}(\square)=L_\rho^\infty(\square)$ we obtain

$$\big\|\big(G_\ell(\mathbf{y})u(\mathbf{x},\mathbf{y})\big)^2-\big(G_{\ell-1}(\mathbf{y})u(\mathbf{x},\mathbf{y})\big)^2\big\|_{L_\rho^\infty(\square;W^{1,1}(D))}\lesssim 2^{-\ell}\|u\|_{L_\rho^\infty(\square;H^2(D))}^2.$$

These examples cover the three quadrature formulae specified in Section 5. Therefore, it is reasonable to make the assumption that a second Bochner space $\mathscr{I}\big(\square;H^2(D)\big)\subset\mathscr{H}\big(\square;H^2(D)\big)$ exists with which we can bound the right hand side in (18).

Theorem 2. *Let $u\in\mathscr{H}\big(\square;H_0^1(D)\cap H^2(D)\big)$ and let $\{Q_\ell\}$ be a sequence of quadrature rules which satisfy* (10) *and* (11)*. Moreover, assume that*

$$\big\|\big(G_\ell(\mathbf{y})u(\mathbf{x},\mathbf{y})\big)^2-\big(G_{\ell-1}(\mathbf{y})u(\mathbf{x},\mathbf{y})\big)^2\big\|_{\mathscr{H}(\square;W^{1,1}(D))}\lesssim 2^{-\ell}\|u\|_{\mathscr{I}(\square;H^2(D))}^2.\tag{19}$$

Then, there holds the following error estimate:

$$\left\|\mathbb{E}_{u^2}(\mathbf{x}) - \sum_{\ell=0}^{j} Q_{j-\ell}\Big(\big(G_\ell(\mathbf{y})u(\mathbf{x},\mathbf{y})\big)^2 - \big(G_{\ell-1}(\mathbf{y})u(\mathbf{x},\mathbf{y})\big)^2\Big)\right\|_{W^{1,1}(D)} \lesssim 2^{-j} j\big(\|u\|^2_{L^4_\rho(\square;H^2(D))} + \|u\|^2_{\mathscr{I}(\square;H^2(D))}\big).$$

Proof. Analogously to the proof of the sparse tensor approximation to the expectation, we find

$$\begin{aligned}\left\|\mathbb{E}_{u^2}(\mathbf{x}) - \sum_{\ell=0}^{j} Q_{j-\ell}\Big(\big(G_\ell(\mathbf{y})u(\mathbf{x},\mathbf{y})\big)^2 - \big(G_{\ell-1}(\mathbf{y})u(\mathbf{x},\mathbf{y})\big)^2\Big)\right\|_{W^{1,1}(D)} \\ \lesssim \big\|\mathbb{E}_{u^2} - I\big(G_j(\mathbf{y})u(\mathbf{x},\mathbf{y})\big)^2\big\|_{W^{1,1}(D)} \\ + \sum_{\ell=0}^{j}\left\|(I-Q_{j-\ell})\Big(\big(G_\ell(\mathbf{y})u(\mathbf{x},\mathbf{y})\big)^2 - \big(G_{\ell-1}(\mathbf{y})u(\mathbf{x},\mathbf{y})\big)^2\Big)\right\|_{W^{1,1}(D)}.\end{aligned}$$

Herein, the first term is estimated by

$$\big\|\mathbb{E}_{u^2} - I\big(G_j(\mathbf{y})u(\mathbf{x},\mathbf{y})\big)^2\big\|_{W^{1,1}(D)} \lesssim 2^{-j}\|u\|^2_{L^4_\rho(\square;H^2(D))}.$$

By bounding the second term by (18), (19), we derive the desired error estimate. ■

9 Numerical results

In the implementation of our numerical examples we consider the L-shaped domain $D = (-1,1)^2 \setminus (-1,0]^2 \subset \mathbb{R}^2$ and choose piecewise linear finite elements on triangles for the discretization (i.e., the spaces $\{\mathscr{S}_j^1(D)\}$). Then, the variance will be a piecewise quadratic finite element function which lives in the spaces $\{\mathscr{S}_j^2(D)\}$. Therefore, the multilevel mesh transfer can be performed by quadratic prolongations.

It is well known that an approximation of the diffusion coefficient by elementwise constant functions sustains the over-all approximation order achieved by the piecewise linear finite elements on the finest mesh. On the coarser meshes, the diffusion coefficient is successively replaced by its mean value with respect to the current triangulation. This does not affect the stiffness matrices but simplifies the computations considerably.

In the following tests, we compare the multilevel Monte Carlo method, the multilevel polynomial chaos method, and the multilevel quasi Monte Carlo method. The quasi Monte Carlo method is based on the Halton sequence [14], where the number of quadrature points is chosen equally to the number of quadrature points of the multilevel Monte Carlo method. The polynomial chaos method is implemented along the lines of [4], but the choice of the polynomial degree is coupled to the exact behaviour of the eigenvalues.

9.1 Finite dimensional stochastics

The first example treats the stochastic diffusion problem with diffusion coefficient determined by the statistics

$$\begin{aligned} \mathbb{E}_\alpha(\mathbf{x}) &= 6.5 + c_1(\mathbf{x}) + c_2(\mathbf{x}) + c_3(\mathbf{x}), \\ \mathrm{Cov}_\alpha(\mathbf{x},\mathbf{y}) &= c_1(\mathbf{x})c_1(\mathbf{y}) + c_2(\mathbf{x})c_2(\mathbf{y}) + c_3(\mathbf{x})c_3(\mathbf{y}), \end{aligned} \tag{20}$$

where the coefficient functions are given by

$$c_1(\mathbf{x}) = x_1x_2, \quad c_2(\mathbf{x}) = -x_2(1-x_1), \quad c_2(\mathbf{x}) = -x_1(1-x_2).$$

By assuming further that the emerging random variables of the Karhunen-Loève expansion are independent and uniformly distributed on $[-1,1]$, we obtain the uniqueness of the respective diffusion coefficient. Note that the covariance function is intrinsically of rank 3 which results in the stochastic dimension $m = 3$. Hence, this example fits perfectly into the considered setting of this article and is especially well suited for the use of the MLQMC method.

We consider two different load vectors for our computations, namely

$$f(\mathbf{x}) = \exp(0.5x_1 + x_2) \text{ and } f(\mathbf{x}) = 2\pi^2 \sin(\pi x_1)\sin(\pi x_2).$$

Fig. 3 shows an illustration of the mean and the standard deviation of the solution which corresponds to the first load vector. Whereas, Fig. 4 shows the mean and the standard deviation of the solution which corresponds to the second load vector.

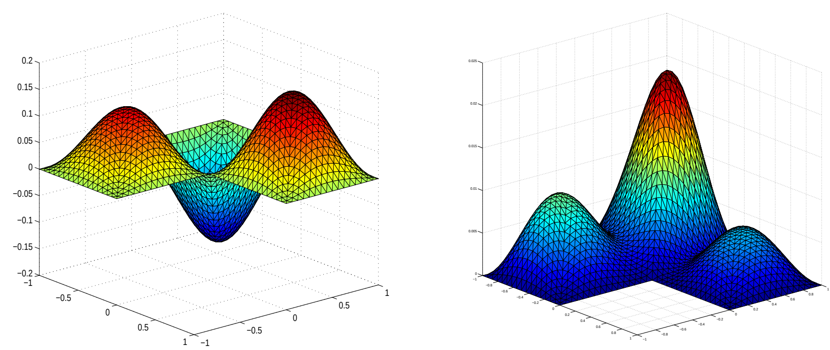

Fig. 3 The expectation and the standard deviation of the solution in case of the load vector $f(\mathbf{x}) = 2\pi^2 \sin(\pi x_1)\sin(\pi x_2)$.

We use the pivoted Cholesky decomposition to calculate the Karhunen-Loève expansion of the diffusion coefficient. The pivoted Cholesky decomposition provides

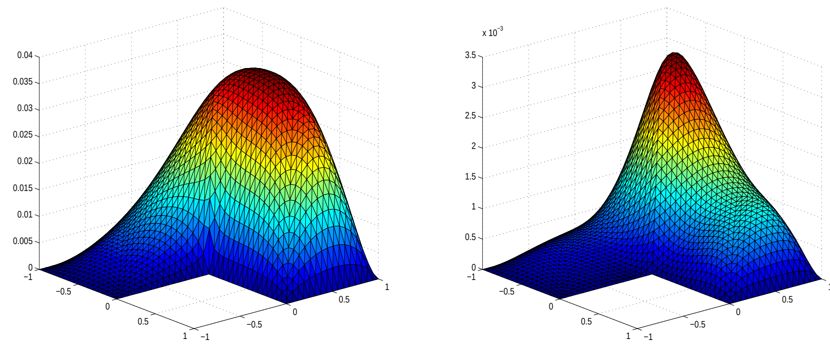

Fig. 4 The expectation and the standard deviation of the solution in case of the load vector $f(\mathbf{x}) = \exp(0.5x_1 + x_2)$.

a simple method to compute a low-rank approximation to the covariance operator. It converges optimally if the eigenvalues of the covariance operator decay sufficiently fast. In particular, if the rank of the covariance operator is finite, this is captured by the algorithm. The major advantage of this approach is that, at any time, the cut-off error of the Karhunen-Loève expansion is rigorously controlled in terms of the trace norm, cf. [9].

In the following, we denote the discretized expectation and the discretized covariance operator of the diffusion coefficient by $\mathbf{E}_\alpha \in \mathbb{R}^d$ and $\mathbf{A} \in \mathbb{R}^{d\times d}$, respectively. The arising mass matrix is neglected due to the choice of L^2-orthonormalized ansatz and test functions. Now the pivoted Cholesky decomposition yields the approximation

$$\|\mathbf{A} - \mathbf{L}\mathbf{L}^\intercal\|_{\mathrm{tr}} \leq \varepsilon$$

for some prescribed precision $\varepsilon > 0$ and a low-rank matrix $\mathbf{L} \in \mathbb{R}^{d\times m}$ with $m \ll d$. Since the eigenvalues of $\mathbf{L}\mathbf{L}^\intercal \in \mathbb{R}^{d\times d}$ coincide with those of $\mathbf{L}^\intercal\mathbf{L} \in \mathbb{R}^{m\times m}$, only a small eigenvalue problem has to be solved. Here, a reasonable speed-up can be achieved in comparison with other subspace methods for eigenvalue computation. If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ denote the orthonormal eigenvectors of the small problem, the eigenvectors of the large matrix are given by $\mathbf{L}\mathbf{v}_1, \mathbf{L}\mathbf{v}_2, \ldots, \mathbf{L}\mathbf{v}_m$. Obviously, we have

$$\mathbf{L}\mathbf{L}^\intercal(\mathbf{L}\mathbf{v}_i) = \mathbf{L}(\mathbf{L}^\intercal\mathbf{L}\mathbf{v}_i) = \lambda_i(\mathbf{L}\mathbf{v}_i)$$

for all $i = 1, 2, \ldots, m$. Due to

$$(\mathbf{L}\mathbf{v}_i)^\intercal(\mathbf{L}\mathbf{v}_j) = \mathbf{v}_i^\intercal\mathbf{L}^\intercal\mathbf{L}\mathbf{v}_j, = \lambda_i\delta_{i,j}$$

with $\delta_{i,j}$ being the Kronecker symbol, a rescaling of the eigenvectors has to be performed. Consequently, the discretized Karhunen-Loève expansion of the diffusion coefficient is simply given by

$$\alpha(\mathbf{y}) = \mathbf{E}_\alpha + \sum_{k=1}^{m} \mathbf{L}\mathbf{v}_k y_k.$$

Fig. 5 shows an illustration of the computed elementwise constant eigenfunctions of the covariance operator.

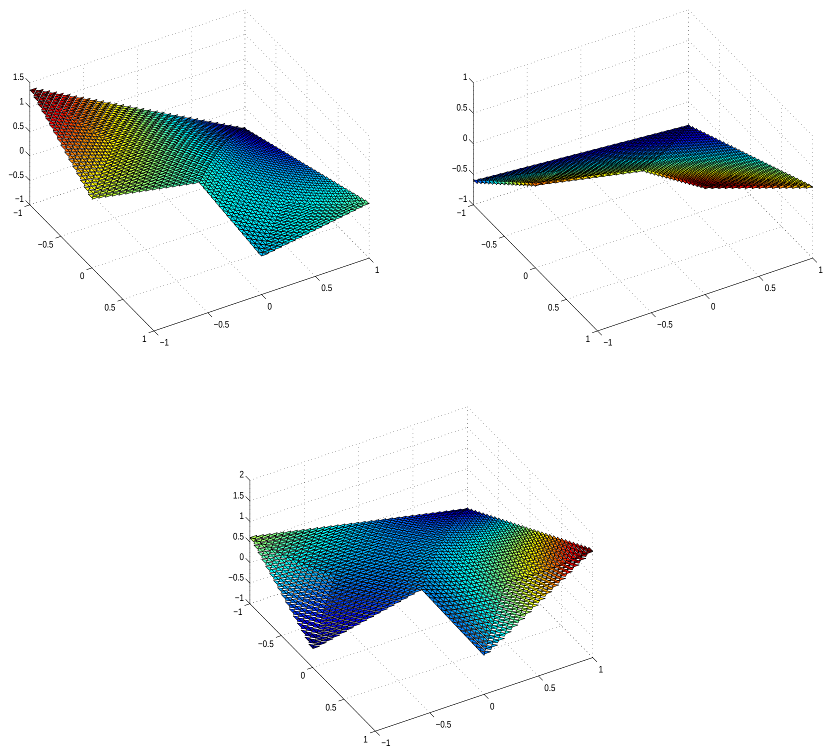

Fig. 5 The eigenfunctions of the covariance operator given by (20).

For the analysis of the convergence rates of the different quadrature rules, we average the solutions of two runs of the multilevel Monte Carlo method on level 8 with a fairly large number of samples on the related coarse spatial mesh. As we pointed out in this article, this reference solution is, up to a logarithmic factor, as accurate as the standard Monte Carlo method for the finite element discretization on level 8 with the same number of samples.

Fig. 6 shows the convergence of the solution's expectation and its second moment in case of the load vector $f(\mathbf{x}) = 2\pi^2 \sin(\pi x_1)\sin(\pi x_2)$. The dashed line corresponds to the expected convergence rate 2^{-j}. The plot implies that the convergence rate is even about 4^{-j} and thus much better than expected.

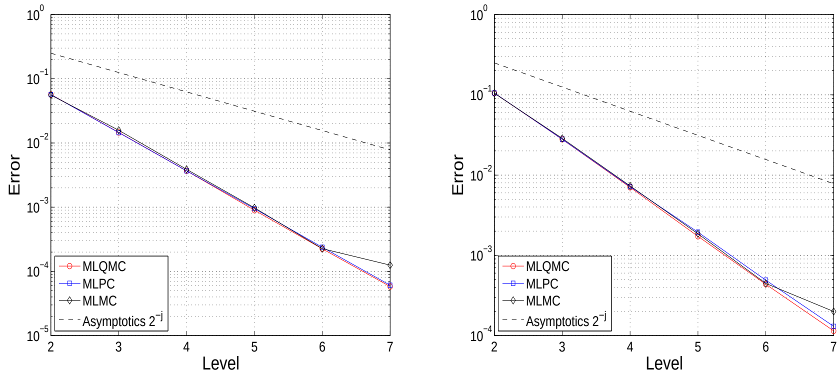

Fig. 6 Relative errors to the expectation and the second moment of the solution in case of $f(\mathbf{x}) = 2\pi^2 \sin(\pi x_1)\sin(\pi x_2)$.

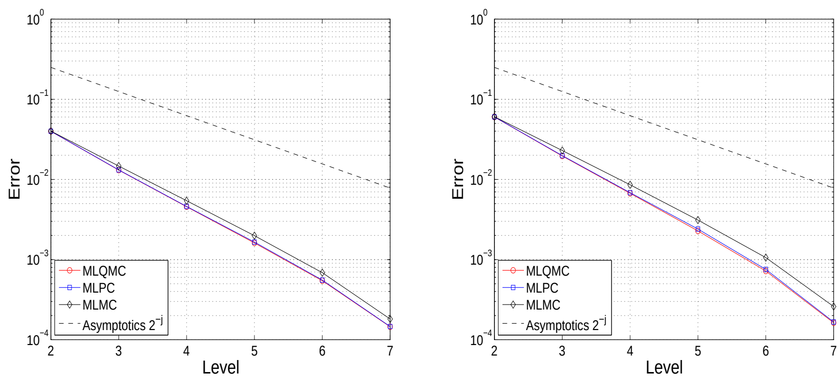

Fig. 7 Relative errors to the expectation and the second moment of the solution in case of $f(\mathbf{x}) = \exp(0.5x_1 + x_2)$.

Fig. 7 presents the convergence of the considered methods with respect to the load vector $f(\mathbf{x}) = \exp(0.5x_1 + x_2)$. Here, likewise nearly the double order of convergence can be observed.

9.2 *Infinite dimensional stochastics*

The second example involves a covariance function of infinite rank. Namely, we consider

$$\mathbb{E}_\alpha(\mathbf{x}) = 10, \quad \mathrm{Cov}_\alpha(\mathbf{x},\mathbf{y}) = \frac{10}{7}\exp\left(\frac{100}{49}\|\mathbf{x}-\mathbf{y}\|_2^2\right). \tag{21}$$

Thus, we have to compute a finite approximation to the corresponding Karhunen-Loève expansion. How the approximation error effects the solution of the respective diffusion problem is investigated in [16]. The load vector under consideration is

$$f(\mathbf{x}) = \exp(0.1x_1 + 0.2x_2).$$

The computations for this example are carried out analogously to the previous example. The only difference is that the Karhunen-Loève expansion needs to be appropriately truncated. For example, it consists of 46 terms on the finest level.

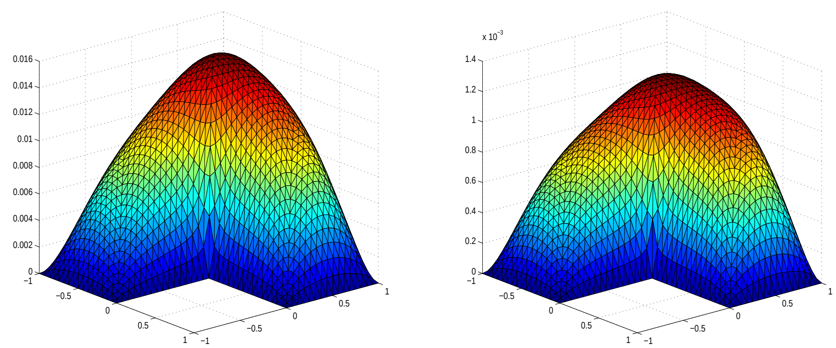

Fig. 8 The expectation and the standard deviation of the solution.

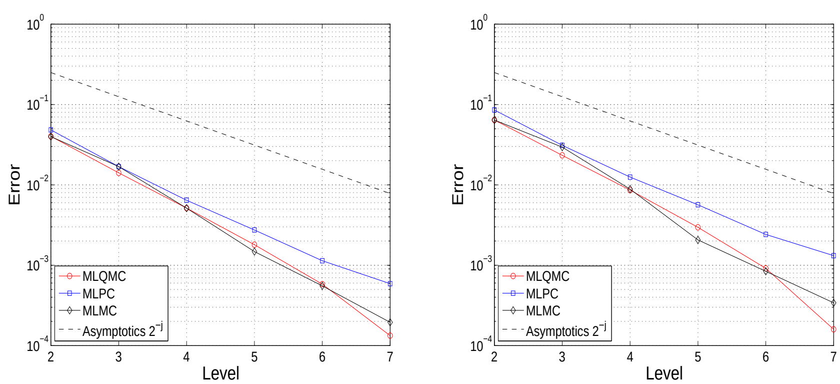

Fig. 9 Relative errors to the expectation and the second moment of the solution.

Fig. 9 shows the convergence behaviour of the different quadrature rules under consideration. Now, the convergence rate fits exactly the expected decay 2^{-j} of the error. This is even observed in case of MLQMC although the choice of the number of quadrature points is heuristically.

References

1. A. Barth, C. Schwab, and N. Zollinger. Multi-level monte carlo finite element method for elliptic PDEs with stochastic coefficients. *Numer. Math.*, 119(1):123–161, 2011.
2. H.-J. Bungartz and M. Griebel. Sparse grids. *Acta Numer.*, 13:147–269, 2004.
3. A. Cohen, R. DeVore, and C. Schwab. Convergence rates of best *N*-term Galerkin approximations for a class of elliptic sPDEs. *Found. Comput. Math.*, 10:615–646, 2010.
4. P. Frauenfelder, C. Schwab, and R. Todor. Finite elements for elliptic problems with stochastic coefficients. *Comput. Methods Appl. Mech. Engrg.*, 194(2-5):205–228, 2005.
5. T. Gerstner and M. Griebel. Numerical integration using sparse grids. *Numer. Algorithms*, 18:209–232, 1998.
6. R. Ghanem and P. Spanos. *Stochastic finite elements: a spectral approach*. Springer-Verlag, New York, 1991.
7. M. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):607–617, 2008.
8. M. Giles and B. Waterhouse. Multilevel quasi-Monte Carlo path simulation. *Radon Series Comp. Appl. Math.*, 8:1–18, 2009.
9. H. Harbrecht, M. Peters, and R. Schneider. On the low-rank approximation by the pivoted Cholesky decomposition. *Preprint 2010-32, SimTech Cluster of Excellence, Universität Stuttgart*, 2010. (submitted).
10. M. Hegland. Adaptive sparse grids. *ANZIAM J.*, 44:C335–C353, 2002.
11. S. Heinrich. The multilevel method of dependent tests. In *Advances in stochastic simulation methods (St. Petersburg, 1998)*, Stat. Ind. Technol., pages 47–61. Birkhäuser Boston, Boston, MA, 2000.
12. S. Heinrich. Multilevel Monte Carlo methods. In *Lect. Notes in Large Scale Scientific Computing*, pages 58–67, London, 2001. Springer-Verlag.
13. M. Loève. *Probability theory. I+II*, volume 45 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 4th edition, 1977.
14. H. Niederreiter. *Random number generation and quasi-Monte Carlo methods*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1992.
15. P. Protter. *Stochastic integration and differential equations: a new approach*. Springer, Berlin, 3rd edition, 1995.
16. C. Schwab and R. Todor. Karhunen-Loève approximation of random fields by generalized fast multipole methods. *J. Comput. Phys.*, 217:100–122, 2006.
17. R. Todor and C. Schwab. Convergence rates for sparse chaos approximations of elliptic problems with stochastic coefficients. *IMA J. Numer. Anal.*, 27(2):232–261, 2007.
18. C. Zenger. Sparse grids. In W. Hackbusch, editor, *Parallel Algorithms for Partial Differential Equations*, volume 31 of *Notes on Numerical Fluid Mechanics*, pages 241–251, Braunschweig/Wiesbaden, 1991. Vieweg-Verlag.