Adaptive Multilevel Stochastic Collocation Method for Randomized Elliptic PDEs

Jens Lang*
Technische Universität Darmstadt, Department of Mathematics
Dolivostraße 15, 64293 Darmstadt, Germany

Robert Scheichl
University of Bath, Department of Mathematical Sciences
Bath BA2 7AY, United Kingdom

November 12, 2017

Abstract

In this paper, we propose and analyse a new adaptive multilevel stochastic collocation method for randomized elliptic PDEs. A hierarchical sequence of adaptive mesh refinements for the spatial approximation is combined with adaptive anisotropic sparse Smolyak grids in the stochastic space in such a way as to minimize computational cost. We provide a rigorous analysis for the convergence and computational complexity of the adaptive multilevel algorithm.

Keywords. multilevel methods, hierarchical methods, adaptivity, stochastic collocation, PDEs with random input data, sparse grids, uncertainty quantification, high-dimensional approximation

AMS subject classifications. 65C20, 65C30, 65N35, 65M75

1 Introduction

Multilevel methods have made their way into Monte Carlo and stochastic collocation methods [2, 11]. They were originally proposed by Giles [6]. Recently, also combinations with spatial adaptivity have been investigated [1, 4, 10]. Here, we formulate first ideas to include rigorous error control in both the solution of the physical PDE and the stochastic collocation method. We will follow the paper by

*corresponding author

Teckentrup, Jantsch, Webster, and Gunzburger [11] and concentrate on elliptic partial differential equations with random input data in the standard setting: Find $u(x,y) : D \times \Gamma \to \mathbb{R}$ such that almost surely

$$\begin{aligned} A(x,y)\,u(x,y) &= f(x,y), \qquad (x,y) \in D \times \Gamma, & (1)\\ u(x,y) &= g(x,y), \qquad (x,y) \in \partial D \times \Gamma, & (2) \end{aligned}$$

where $\Gamma = \Pi_{n=1,\ldots,N}\Gamma_n$ with bounded Γ_n (boundedness assumption) is a stochastic parameter space of finite dimension N (finite noise assumption) and $D \subset \mathbb{R}^d$, $d = 1,2,3$, is the deterministic (physical) space. The random variables $y_1, \ldots, y_n$ have a joint probability density function $\rho(y) = \Pi_{n=1,\ldots,N}\hat{\rho}_n \in L^\infty(\Gamma)$, $\hat{\rho}_n : [-1,1] \to \mathbb{R}$. We assume that problem (1)-(2) admits a unique solution $u(x,y) \in L^2_\rho(\Gamma; H^1(D))$ in the weighted Bochner space

$$L^2_\rho(\Gamma; H^1(D)) = \{v : \Gamma \to H^1(D) \text{ measurable: } \int_\Gamma \|v(y,\cdot)\|^2_{H^1(D)}\rho(y)dy < \infty\} \quad (3)$$

with corresponding norm

$$\|v\|^2_{L^2_\rho(\Gamma;H^1(D))} = \int_\Gamma \|v(y,\cdot)\|^2_{H^1(D)}\rho(y)dy\,. \quad (4)$$

The standard, single-level stochastic collocation method uses a set of sampling points $\{y_m\}_{m=1,\ldots,M}$ in Γ and finite-dimensional spatial approximations $u_h(y) \in V_h \subset H^1(D)$ to construct an interpolant

$$u^{(SL)}_{M,h}(x,y) = \mathcal{I}[u_h](x,y) = \sum_{m=1}^{M} u_m(x)\phi_m(y), \quad (5)$$

in the polynomial space $\mathcal{P}_M = \text{span}\{\phi_m\}_{m=1,\ldots,M} \subset L^2_\rho(\Gamma)$ with basis functions $\{\phi_m\}_{m=1,\ldots,M}$. The coefficients $u_m(x)$ are determined by the interpolating condition $\mathcal{I}[u_h](x,y_m) = u_h(x,y_m)$ for $m = 1,\ldots,M$. The quality of the interpolation process depends on the accuracy of the spatial approximations $u_h(y)$ and the number of collocation points M, which in practice can quickly grow with increasing stochastic dimension N. Multilevel methods aim at reducing the overall computational cost through exploiting hierarchies of different resolutions in both spatial and stochastic approximations. Examples are uniform mesh refinement and and generalized sparse grids as used in [11]. In what follows, we will combine an adaptive mesh refinement for the approximation of the spatial approximations $u_h(y)$ with an adaptive (anisotropic) sparse Smolyak grid in order to improve the multilevel method and to reach a user-prescribed tolerance for the accuracy of the multilevel interpolant.

2 Adaptive Spatial Approximation

Let $\{TolX_k\}_{k=0,\ldots,K}$ be a decreasing sequence of tolerances with

$$1 \geq TolX_0 > TolX_1 > \cdots > TolX_k > \cdots > TolX_K > 0. \tag{6}$$

Then, for each fixed $y \in \Gamma$, we adaptively compute approximate spatial solutions $u_k(y) \in V_k(y)$ on nested subspaces

$$V_0(y) \subset V_1(y) \subset \cdots \subset V_K(y) \subset H^1(D) \tag{7}$$

with the pointwise error estimates

$$\|u(\cdot, y) - u_k(y)\|_{H^1(D)} \leq C_x \cdot TolX_k, \quad k = 0, \ldots, K. \tag{8}$$

We assume that the positive constant C_x does not depend on k and y. Supposing measurability of the discrete spaces $V_k(y)$, we directly get

$$\|u - u_k\|_{L^2_\rho(\Gamma, H^1(D))} \leq C_x \cdot TolX_k, \quad k = 0, \ldots, K. \tag{9}$$

Adaptive algorithms as proposed in [3, 7] converge for fixed $y \in \Gamma$ and $TolX_k \to 0$. The constant C_x depends on the quality of the a posteriori error estimator. Values close to one can be obtained by hierarchical error estimators and gradient recovery techniques in the asymptotic regime.

3 Adaptive Stochastic Interpolation

Let us assume $u \in C^0(\Gamma; H^1(D))$ and denote by $\{\mathcal{I}_{M_k}\}_{k=0,1,\ldots}$ a sequence of interpolation operators

$$\mathcal{I}_{M_k} : C^0(\Gamma) \to L^2_\rho(\Gamma) \tag{10}$$

with M_k points from the N-dimensional space $\Gamma = \Pi_{n=1,\ldots,N}\Gamma_n$. We construct each of these operators by a hierarchical sequence of one-dimensional Lagrange interpolation operators with the anisotropic Smolyak algorithm, which was introduced by Gerstner and Griebel [5]. The method is dimension-adaptive, using the individual surplus spaces in the multi-dimensional hierarchy as natural error indicators.

Let $\{TolY_k\}_{k=0,\ldots,K}$ be a second sequence of tolerances. Under suitable regularity assumptions for the uncertain data (see e.g. [11, Lemma 5.7]), we can assume that there exists numbers M_k, $k = 0, 1, \ldots, K$, and a positive constant C_s not depending on k such that

$$\|(u_k - u_{k-1}) - \mathcal{I}_{M_{K-k}}[u_k - u_{k-1}]\|_{L^2_\rho(\Gamma; H^1(D))} \leq C_s \cdot TolY_{K-k}, \quad k = 0, \ldots, K, \tag{11}$$

where, for simplicity, we set $u_{-1} = 0$. Since $\|u_k - u_{k-1}\|_{L^2_\rho(\Gamma; H^1(D))} \leq C \cdot TolX_{k-1}$ and therefore decreasing as $k \to \infty$, we can expect that less accurate interpolation

operators, i.e. smaller numbers M_{K-k}, are needed for higher k to achieve a required accuracy. Indeed this is the main motivation to set up a multilevel interpolation approximation. Suitable values for the tolerances $TolY_k$ will be given in the next Section.

4 Adaptive Multilevel Method

Given the sequences $\{u_k\}$ and $\{\mathcal{I}_{M_k}\}$, we define the multilevel interpolation approximation in the usual way through

$$u_K^{(ML)} = \sum_{k=0}^{K} \mathcal{I}_{M_{K-k}}[u_k - u_{k-1}] = \sum_{k=0}^{K} \left(u^{(SL)}_{M_{K-k},k} - u^{(SL)}_{M_{K-k},k-1} \right). \tag{12}$$

Observe that the most accurate interpolation operator $\mathcal{I}_{M_K}$ is used on the coarsest spatial approximation u_0 whereas the least accurate interpolation operator $\mathcal{I}_{M_0}$ is applied to the difference of the finest spatial approximations $u_K - u_{K-1}$. From (12), the close relations of the single index k for the spatial and stochastic approximations are clearly visible.

Convergence analysis
To show the convergence of the multilevel approximation $u_K^{(ML)}$ to the true solution u, we split the error into the sum of a spatial discretization error and a stochastic interpolation error. This yields with the triangle inequality

$$\|u - u_K^{(ML)}\|_{L^2_\rho(\Gamma;H^1(D))} \leq \|u - u_K\|_{L^2_\rho(\Gamma;H^1(D))} + \|u_K - u_K^{(ML)}\|_{L^2_\rho(\Gamma;H^1(D))}. \tag{13}$$

Due to (9) the spatial discretization error is bounded by $C_x \cdot TolX_K$. The aim is now to choose the tolerances $TolY_k$ in an appropriate way to reach the same accuracy. From (11), we estimate the stochastic interpolation error as follows:

$$\begin{aligned} \|u_K - u_K^{(ML)}\|_{L^2_\rho(\Gamma;H^1(D))} &= \left\| \sum_{k=0}^{K} (u_k - u_{k-1}) - \mathcal{I}_{M_{K-k}}[u_k - u_{k-1}] \right\|_{L^2_\rho(\Gamma;H^1(D))} \\ &\leq \sum_{k=0}^{K} \left\| (u_k - u_{k-1}) - \mathcal{I}_{M_{K-k}}[u_k - u_{k-1}] \right\|_{L^2_\rho(\Gamma;H^1(D))} \leq \sum_{k=0}^{K} C_s \cdot TolY_{K-k}. \end{aligned} \tag{14}$$

To obtain an accuracy of the same size as for the spatial discretization error, we simply require $TolY_k \leq C_x \cdot TolX_K/((K+1)C_s)$ for $k = 0, \ldots, K$. It follows

$$\|u - u_K^{(ML)}\|_{L^2_\rho(\Gamma;H^1(D))} \leq 2\, C_x \cdot TolX_K, \tag{15}$$

and thus convergence of the adaptive multilevel method for $TolX_K \to 0$. The values for $TolY_k$ can be optimized by minimizing the computational costs while keeping

the desired accuracy. This is considered next.

Cost analysis
We will analyse the computational costs, $C_\epsilon^{(ML)}$, of the multilevel stochastic collocation estimator $u_K^{(ML)}$, required to achieve an accuracy ϵ. In order to quantify the contributions from the spatial discretization and the stochastic collocation, we make two assumptions to link the costs with the accuracies in (9) and (11). Let A_k denote the cost for solving the deterministic PDE for one sample point from Γ with accuracy $TolX_k$. Then, we assume

$$\begin{aligned} (A1) \qquad A_k &\leq C_c \cdot TolX_k^{-s}, \\ (A2) \qquad C_s \cdot TolY_{K-k} &= C_I(N)\, M_{K-k}^{-\mu}\, TolX_{k-1} \end{aligned}$$

for all $k = 0, \ldots, K$, and the special case $\|u_0\|_{H^1(D)} \leq TolX_{-1} := const$. Here, $C_I(N) > 0$, $C_c > 0$ does not depend on k, and $s, \mu > 0$ are two real numbers. Assumption $(A1)$ usually holds for adaptive spatial discretization methods with $s = d$, when they are coupled with optimal linear solvers such as multigrid. The factors on the right-hand side in $(A2)$ reflect best the convergence of the sparse grid approximations in (11) with respect to the total number M_{K-k} of collocation points, see [8, 9] or [11, Theorem 5.4]. To estimate the difference $u_k - u_{k-1}$, we have used the fact that $\|u_k - u_{k-1}\|_{H^1(D)} \leq C \cdot TolX_{k-1}$ with a constant $C > 0$ close to one. It follows from $\|u - u_{k-1}\|_{H^1(D)} \approx \|u_k - u_{k-1}\|_{H^1(D)}$, which is the basis for the very good performance of hierarchical error estimators. We absorb C in C_I. The factor μ strongly depends on the dimension N. Several examples, including also the anisotropic classical Smolyak, are given in [11, Table 5.2].

The total computational cost of the approximation $u_K^{(ML)}$ can be defined as

$$C^{(ML)} = \sum_{k=0}^{K} M_{K-k}\,(A_k + A_{k-1}). \tag{16}$$

In a first step, we will consider a general sequence $\{TolX_k\}$ without defining a decay rate a priori. We have the following result for the ϵ-cost $C_\epsilon^{(ML)}$ and the optimal choice of the tolerances $TolY_k$ in (11).

Theorem 4.1. *Let a decreasing sequence of spatial tolerances $\{TolX_k\}_{k=0,1,\ldots,K}$ in (6) with unfixed K are given. Suppose assumptions $(A1)$ and $(A2)$ hold. Then, for any epsilon, there exist an integer $K(\epsilon)$ and a sequence of tolerances in (11), $\{TolY_k\}_{k=0,\ldots,K}$, such that*

$$\|u - u_K^{(ML)}\|_{L^2_\rho(\Gamma;H^1(D))} \leq \epsilon \tag{17}$$

and

$$C_\epsilon^{(ML)} \leq C \cdot G_K(\mu)^{\frac{\mu+1}{\mu}}\, \epsilon^{-\frac{1}{\mu}} + C_c \sum_{k=0}^{K} \left(TolX_k^{-s} + TolX_{k-1}^{-s}\right) \tag{18}$$

with $C = C_c\,(2\,C_I)^{\frac{1}{\mu}}$ *and*

$$G_K(\mu) = \sum_{k=0}^{K} \left(TolX_k^{-s} + TolX_{k-1}^{-s}\right)^{\frac{\mu}{\mu+1}} \left(TolX_{k-1}\right)^{\frac{1}{\mu+1}}. \tag{19}$$

The optimal tolerances $TolY_k$ *are given through*

$$TolY_{K-k} = (2\,C_s\,G_K(\mu))^{-1}\,\left(TolX_k^{-s} + TolX_{k-1}^{-s}\right)^{\frac{\mu}{\mu+1}}\,\left(TolX_{k-1}\right)^{\frac{1}{\mu+1}}\,\epsilon. \tag{20}$$

Proof: As in the convergence analysis, we split the error and make sure that both the spatial discretization error and stochastic interpolation error are bounded by $\epsilon/2$. First, we choose an appropriate $K \geq 0$ and $TolX_K$ such that $TolX_K < \epsilon/2$. This is, of course, always possible and fixes the number K. Next we determine the set $\{M_k\}_{k=0,\ldots,K}$ so that the computational cost in (16) are minimized subject to the requirement that the stochastic interpolation error is bounded by $\epsilon/2$. Using assumptions $(A1)$ and $(A2)$, this reads

$$\begin{aligned} &\min_{M_0,\ldots,M_K} \sum_{k=0}^{K} C_c \cdot M_{K-k} \left(TolX_k^{-s} + TolX_{k-1}^{-s}\right) \\ &s.t. \quad \sum_{k=0}^{K} C_I \cdot M_{K-k}^{-\mu}\, TolX_{k-1} = \frac{\epsilon}{2}. \end{aligned} \tag{21}$$

Applying the Lagrange multiplier method with all M_k treated as continuous variables, gives the optimal choice for the number of samples

$$M_{K-k} = (2\,C_I\,G_K(\mu))^{\frac{1}{\mu}}\,F_k^{-\frac{1}{\mu+1}}\,\epsilon^{-\frac{1}{\mu}} \tag{22}$$

with

$$F_k = \left(TolX_k^{-s} + TolX_{k-1}^{-s}\right)\,\left(TolX_{k-1}\right)^{-1}, \quad k = 0, \ldots, K, \tag{23}$$

$$G_K = \sum_{k=0}^{K} F_k^{\frac{\mu}{\mu+1}}\,TolX_{k-1}. \tag{24}$$

To make sure that M_{K-k} is an integer, we round up to next integer. The complexity

of the multilevel approximation can now be estimated:

$$\begin{aligned} C_\epsilon^{(ML)} &\leq \sum_{k=0}^{K} C_c \cdot (M_{K-k}+1)\left(TolX_k^{-s}+TolX_{k-1}^{-s}\right) \\ &= \sum_{k=0}^{K} C_c \cdot \left((2\,C_I\,G_K(\mu))^{\frac{1}{\mu}} F_k^{-\frac{1}{\mu+1}}\,\epsilon^{-\frac{1}{\mu}}+1\right)\left(TolX_k^{-s}+TolX_{k-1}^{-s}\right) \\ &\leq C\cdot G_K(\mu)^{\frac{1}{\mu}}\epsilon^{-\frac{1}{\mu}} \sum_{k=0}^{K} F_k^{\frac{\mu}{\mu+1}}\,TolX_{k-1} + C_c \sum_{k=0}^{K}\left(TolX_k^{-s}+TolX_{k-1}^{-s}\right) \\ &= C\cdot G_K(\mu)^{\frac{\mu+1}{\mu}}\epsilon^{-\frac{1}{\mu}} + C_c \sum_{k=0}^{K}\left(TolX_k^{-s}+TolX_{k-1}^{-s}\right) \end{aligned}$$

with $C = C_c\,(2\,C_I)^{\frac{1}{\mu}}$.

The optimal tolerances $TolY_k$ can be directly determined from assumption (A2). We get with the above defined M_k

$$TolY_{K-k} = (2\,C_s\,G_K(\mu))^{-1}\,F_k^{\frac{\mu}{\mu+1}}\,TolX_{k-1}\,\epsilon. \tag{25}$$

Note that with these values $\sum_{k=0,\ldots,K} C_s \cdot TolY_k = \epsilon/2$, which gives the desired accuracy in (14). □

Observe that the function $G_K(\mu)$ and the sum over powers of all spatial tolerances still depend on ϵ, because K is a function of ϵ. In this way, the rate $(-1/\mu)$ is influenced by the choice of the tolerances $TolX_k$, which could be also an issue for further optimization. A typical design is a geometric sequence $TolX_k = q^k\,TolX_0,\ k = 1, 2, \ldots$, with a positive reduction factor $q < 1$. This will be treated next. In the following, we use the relation $a \lesssim b \Leftrightarrow a < Cb$ with a generic constant C that does not depend on the dimension N and the number of samples M_k, and $a \eqsim b \Leftrightarrow (a \lesssim b$ and $b \lesssim a)$.

Theorem 4.2. *Let the sequence of spatial tolerances $\{TolX_k\}_{k=0,1,\ldots,K}$ in (6) are defined by $TolX_k = q^k\,TolX_0$ with a reduction factor $q < 1$. Suppose assumptions (A1) and (A2) hold. Then, for any $\epsilon < 1$, there exists an integer $K(\epsilon)$ such that*

$$\|u - u_K^{(ML)}\|_{L^2_\rho(\Gamma;H^1(D))} \leq \epsilon \tag{26}$$

and

$$C_\epsilon^{(ML)} \lesssim \begin{cases} \epsilon^{-\frac{1}{\mu}} & \text{if } s\mu < 1 \\ \epsilon^{-\frac{1}{\mu}}\,|\log\epsilon|^{1+\frac{1}{\mu}} & \text{if } s\mu = 1 \\ \epsilon^{-s} & \text{if } s\mu > 1. \end{cases} \tag{27}$$

Proof: We start with identifying the number K. From the accuracy requirement $TolX_K = q^K\, TolX_0 < \epsilon/2$ we deduce

$$K = \left\lceil \log_q \left(\frac{\epsilon}{2\, TolX_0} \right) \right\rceil . \tag{28}$$

This gives $K \leq \log_q(\epsilon/(2\, TolX_0)) + 1$ and the estimate

$$C_c \sum_{k=0}^{K} \left(TolX_k^{-s} + TolX_{k-1}^{-s} \right) \lesssim \sum_{k=0}^{K} q^{-ks} \lesssim \frac{q^{-sK}}{1-q^s} \lesssim \epsilon^{-s}. \tag{29}$$

To estimate the term $G_K(\mu)$ in (18), we first state that

$$G_K(\mu) \lesssim \sum_{k=0}^{K} q^{k(1-s\mu)/(\mu+1)}. \tag{30}$$

The behaviour of this geometric sum depends on the sign of $1-s\mu$. When $1-s\mu > 0$, the sum converges to a limit independent of K. Since in this case $\epsilon^{-s} < \epsilon^{-1/\mu}$ for $\epsilon < 1$, we end up with $C_\epsilon^{(ML)} \lesssim \epsilon^{-1/\mu}$. When $1 - s\mu = 0$, we have $G_K(\mu) \lesssim K + 1$, which gives together with K as defined in (28) the additional logarithmic term, i.e., $C_\epsilon^{(ML)} \lesssim \epsilon^{-1/\mu} |\log \epsilon|^{1+1/\mu}$. Eventually, when $1 - s\mu < 0$, we estimate

$$G_K(\mu) \lesssim q^{\frac{K(1-s\mu)}{\mu+1}} \sum_{k=0}^{K} \left(q^{-\frac{1-s\mu}{\mu+1}} \right)^{K-k} \lesssim q^{\frac{K(1-s\mu)}{\mu+1}} \lesssim \epsilon^{\frac{1-s\mu}{\mu+1}} \tag{31}$$

and find

$$C_\epsilon^{(ML)} \lesssim \epsilon^{-\frac{1}{\mu}} \epsilon^{\frac{1-s\mu}{\mu+1} \cdot \frac{\mu+1}{\mu}} + \epsilon^{-s} \eqsim \epsilon^{-s}. \tag{32}$$

This completes the proof. □

For a comparison with the standard single-level stochastic collocation method, we set $K = 0$ and use the estimate

$$\|u - u_0^{(SL)}\|_{L^2_\rho(\Gamma; H^1(D))} \leq C_x \cdot TolX_0 + C_I \, \|u_0\|_{H^1(D)} \, M_0^{-\mu}. \tag{33}$$

Balancing both contributions with $\epsilon/2$, requests $TolX_0 \eqsim \epsilon$ and $M_0 \eqsim \epsilon^{-1/\mu}$. The computational ϵ-cost is then bounded by

$$C_\epsilon^{(SL)} \eqsim M_0 \cdot TolX_0^{-s} \eqsim \epsilon^{-s-\frac{1}{\mu}}. \tag{34}$$

In terms of savings, we find a reduction factor $\Theta := C_\epsilon^{(ML)} / C_\epsilon^{(SL)} \eqsim \epsilon^s$ for the case $s\mu < 1$, which also holds, up to a log factor, for $s\mu = 1$. When $s\mu > 1$, we have $\Theta \eqsim \epsilon^{1/\mu}$. The advantage of the multilevel method is obvious.

5 A few remarks

We would like to give a few remarks.

Remark 5.1. *The proposed method has a self-adaptive nature. Once the tolerances* $\{TolX_k\}_{k=0,\dots,K}$ *and* $\{TolY_k\}_{k=0,\dots,K}$ *are set, the algorithm delivers a solution* $u_K^{(ML)}$ *with an accuracy close to* ϵ*, provided that the constants* C_x *and* C_s*, which describe the reliability of the estimation for the adaptive spatial discretization and the adaptive Smolyak algorithm, are close to one. While the spatial tolerances* $TolX_k$ *can be freely chosen, the optimal choice of the tolerances* $TolY_k$ *in (20) requests the knowledge of the parameters* s *and* μ*. They have to be determined in advance through an appropriate number of samples. Usually, it holds* $s = d$*. Note that the adaptive anisotropic Smolyak algorithm will automatically detect the importance of various directions in the parameter space* $\Gamma \subset \mathbb{R}^N$*.*

Remark 5.2. *A crucial point already mentioned in [11] is that the optimal rounded values for the number of samples,* M_k*, will not be used by the algorithm, because they do not necessarily correspond to an adaptive sparse grid level. However, for each level* k*, the tolerance* $TolY_k$ *will be ensured with a perturbed number* $\tilde{M}_k \geq M_k$*, resulting in a slight inefficiency of the sparse grid approximation. Note that, in practice, the same behaviour is observed for adaptive spatial discretizations. In any case, there is no restart necessary as used in [11, Section 6.3].*

6 Acknowledgement

The first author is supported by the German Research Foundation within the collaborative research center TRR154 "Mathematical Modeling, Simulation and Optimisation Using the Example of Gas Networks" (DFG-SFB TRR154/1-2014, TP B01), the Graduate School of Excellence Computational Engineering (DFG GSC233), and the Graduate School of Excellence Energy Science and Engineering (DFG GSC1070). The second author is supported by the UK Engineering and Physical Sciences Research Council (EPSRC) under grant number EP/K031368/1.

References

[1] E. Babushkina and R. Kornhuber. Adaptive multilevel Monte Carlo methods for stochastic variational inequalities. Technical Report arXiv:1611.06012v1, 2016.

[2] K.A. Cliffe, M.B. Giles, R. Scheichl, and A.L. Teckentrup. Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients. *Computing and Visualization in Science*, 14:3–15, 2011.

[3] W. Dörfler. A convergent adaptive algortihm for Poisson's equation. *SIAM J. Numer. Anal.*, 33:1106–1124, 1996.

[4] M. Eigel, Ch. Merdon, and J. Neumann. An adaptive Monte Carlo method with stochastic bounds for quantities of interest with uncertain data. *SIAM/ASA Journal on Uncertainty Quantification*, 4:1219–1245, 2016.

[5] T. Gerstner and M. Griebel. Dimension-adaptive tensor-product quadrature. *Computing*, 71:65–87, 2003.

[6] M. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56:607–617, 2008.

[7] C. Kreuzer and K.G. Siebert. Decay rates of adaptive finite elements with Dörfler marking. *Numerische Mathematik*, 1147:679–716, 2011.

[8] F. Nobile, R. Tempone, and C. Webster. An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J. Numer. Anal.*, 46:2411–2442, 2008.

[9] F. Nobile, R. Tempone, and C. Webster. A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J. Numer. Anal.*, 46:2309–2345, 2008.

[10] B. Schieche and J. Lang. Adjoint error estimation for stochastic collocation methods. In J. Garcke and D. Pflüger, editors, *Sparse Grids and Applications - Munich 2012*, volume 97 of *Lecture Notes on Computational Science and Engineering*, pages 271–294. Springer, 2014.

[11] 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 Journal on Uncertainty Quantification*, 3:1046–1074, 2015.