Binned Multilevel Monte Carlo for Bayesian Inverse Problems with Large Data

Robert N. Gantner[1], Claudia Schillings[1], and Christoph Schwab[1]

Abstract We consider Bayesian inversion of parametric operator equations for the case of a large number of measurements. Increased computational efficiency over standard averaging approaches, per measurement, is obtained by binning the data and applying a multilevel Monte Carlo method, specifying optimal forward solution tolerances per level. Based on recent bounds of convergence rates of adaptive Smolyak quadratures in Bayesian inversion [7] for single observation data, the bin sizes in large sets of measured data are optimized and a rate of convergence of the error vs. work is derived analytically and confirmed by numerical experiments.

1 Introduction

In recent years, various methods have been developed for solving parametric operator equations, mainly focusing on the *estimation of parameters given measurements* of the parametric solution, subject to a stochastic observation error model. A second objective is *prediction of a "most likely" response of the parametric system given noisy measurements*. The *Bayesian approach* to such inverse problems for partial differential equations (PDEs for short) has been the focus of numerous papers [10, 9, 7, 8] and will be considered here. For multiple data points, averaging is often done with a standard Monte Carlo approach. We consider here the case where computational resources are limited and develop a multilevel Monte Carlo method (MLMC) achieving an error of the same order while requiring less work [6, 5, 2, 1].

[1]Seminar for Applied Mathematics, ETH Zürich, CH-8092 Zürich, Switzerland

2 Bayesian Inversion of Parametric Operator Equations

We assume an operator equation depending on a distributed, uncertain "parameter" u with values in a separable Banach space X. It takes the form of the operator equation

$$\text{Given } u \in \tilde{X} \subseteq X, \text{ find } q \in \mathscr{X}: \quad A(u;q) = F(u) \quad \text{in} \quad \mathscr{Y}', \tag{1}$$

where we denote by $\mathscr{X}$ and $\mathscr{Y}$ two reflexive Banach spaces over $\mathbb{R}$ with (topological) duals $\mathscr{X}'$ and $\mathscr{Y}'$, respectively and $A(u;\cdot) \in \mathscr{L}(\mathscr{X},\mathscr{Y}')$. Assuming that the forcing function $F : \tilde{X} \mapsto \mathscr{Y}'$ is known, and the uncertain operator $A(u;\cdot) : \mathscr{X} \mapsto \mathscr{Y}'$ is locally boundedly invertible for uncertain input u in a sufficiently small neighborhood $\tilde{X}$, let the *uncertainty-to-observation map* $\mathscr{G} : \tilde{X} \mapsto \mathbb{R}^K$ have the structure

$$X \supseteq \tilde{X} \ni u \mapsto \mathscr{G}(u) := \mathscr{O}(G(u;F)) \in \mathbb{R}^K . \tag{2}$$

Here, $\tilde{X} \ni u \mapsto q(u) = G(u;F) \in \mathscr{X}$ denotes the (noise-free) response of the forward problem for a given instance of $u \in \tilde{X}$ and $\mathscr{O}$ a *bounded linear observation operator* $\mathscr{O} \in \mathscr{L}(\mathscr{X},\mathbb{R}^K),\ K < \infty$. The goal of computation is assumed to be the low-order statistics of a *quantity of interest* (QoI) given noisy observational data δ of the form

$$\delta = \mathscr{G}(u) + \eta , \tag{3}$$

where δ represents the observation of $\mathscr{G}(u)$ perturbed by the noise η, a random variable with given statistical properties. We restrict ourselves to the case where the measurement error is Gaussian and the covariance matrix symmetric positive definite, i.e. $\eta \sim \mathscr{N}(0,\Gamma)$ with $\Gamma \in \mathbb{R}^{K\times K}_{\text{spd}}$.

We work in the following under the assumption that the uncertainty u admits a parametric representation of the form

$$u = u(\mathbf{y}) := \langle u \rangle + \sum_{j\in\mathbb{J}} y_j \psi_j \in X$$

for some "nominal" value $\langle u \rangle \in X$ of the uncertain datum u, a countable sequence $(\psi_j)_{j\in\mathbb{J}}$ of X with $\mathbb{J} := \{1,\dots,J\}, J < \infty$ or $\mathbb{J} = \mathbb{N}$ and for some coefficient sequence $\mathbf{y} = (y_j)_{j\in\mathbb{J}}$ (after possibly rescaling the fluctuations) in the reference domain $U = [-1,1]^{\mathbb{J}} = \bigotimes_{j\in\mathbb{J}}[-1,1]$ with unconditional convergence. We assume $\mathbf{y}$ to be a random variable on the countable product probability space $(U,\mathscr{B}(U),\mu_0)$ with U as above and with $\mu_0(\mathrm{d}\mathbf{y}) = \prod_{j\in\mathbb{J}} \frac{1}{2}\lambda^1(\mathrm{d}y_j)$. This also makes δ a random variable; for a fixed value of $\mathbf{y}$, (3) gives an expression for $\delta(\mathbf{y})$, denoted by $\delta|\mathbf{y}$.

In general, our aim is to compute the "most likely" value of a QoI over all realizations of u, with the QoI defined as a function $\phi : U \to \mathscr{S}$ mapping from the parameter space U to some Banach space $\mathscr{S}$. Bayes' theorem characterizes this value as the mathematical expectation with respect to a probability measure μ_0 (the "Bayesian prior") on U which we choose as a countable product of uniform measures. In particular, we are interested in $\phi = G$, the response of the system. To this end, we use Bayes' Theorem to obtain an expression for $\mathbf{y}|\delta$, as in [9, 10].

Theorem 1 (Bayes' Theorem). *Let* $\mathscr{G}\Big|_{u=\langle u\rangle+\sum_{j\in\mathbb{J}} y_j\psi_j} : U \to \mathbb{R}^K$ *be bounded and continuous. Then,* $\mu^\delta(d\mathbf{y})$*, the distribution of* $\mathbf{y}|\delta$*, is absolutely continuous with respect to* $\mu_0(d\mathbf{y})$*, and*

$$\frac{\mathrm{d}\mu^\delta(\mathbf{y})}{\mathrm{d}\mu_0(\mathbf{y})} = \frac{1}{Z_\delta}\exp\left(-\frac{1}{2}||\delta - \mathscr{G}(\mathbf{y})||_\Gamma^2\right)$$

with $Z_\delta := \int_U \exp(-\Phi(\mathbf{y};\delta))\,\mu_0(\mathrm{d}\mathbf{y}) > 0$.

In the Bayesian setting, the distribution $\mathrm{d}\mu_0(\mathbf{y})$ is called the *prior distribution* and is assumed to be known and easily computable. Thus, we can write our desired expectation as an integral over the prior measure μ_0:

$$\mathbb{E}^{\mu^\delta}[\phi] = \int_U \phi(\mathbf{y})\,\mu^\delta(\mathrm{d}\mathbf{y}) = \frac{1}{Z_\delta}\int_U \phi(\mathbf{y})\exp\left(-\frac{1}{2}||\delta - \mathscr{G}(\mathbf{y})||_\Gamma^2\right)\mu_0(\mathrm{d}\mathbf{y}) =: \frac{Z'_\delta}{Z_\delta}\,. \quad (4)$$

This formulation of the expectation $\mathbb{E}^{\mu^\delta}[\cdot]$ is based on just one measurement δ. For a given model for the measurement errors η, we would like to additionally compute the expectation over all errors. Assuming that the perturbations η are normally distributed as above, this can be written as an expectation with respect to the measure $\gamma_\Gamma^K(\eta)$, the K-variate Gaussian measure with covariance Γ. *Here, and throughout, we assume the observation noise* η *to be statistically independent from the uncertain parameter* u *in* (1). This yields the total expectation of the QoI ϕ in terms of Z'_δ and Z_δ as

$$\mathbb{E}^{\gamma_\Gamma^K}\left[\mathbb{E}^{\mu^\delta}[\phi]\right] = \int_{\mathbb{R}^K} \frac{Z'_\delta}{Z_\delta}\Big|_{\delta=\mathscr{G}(\mathbf{y}_0)+\eta} \gamma_\Gamma^K(\mathrm{d}\eta)\,, \quad (5)$$

where $\mathscr{G}(\mathbf{y}_0)$ denotes the observation at the unknown, exact parameter $\mathbf{y}_0$.

In practice, we are given a set of measurements $\Delta := \{\delta_i, i = 1,\dots,M\}$ with which this outer expectation should be approximated. The measurements can be taken at different positions, i.e. with respect to different observation maps $\mathscr{O}_i$ in (2). In the derivations below, we consider the notationally more convenient case where the measurements are all obtained using the same observation map. We do, however, impose the restriction that the measurements are homoscedastic, i.e. δ_i is Gaussian with the same covariance Γ for all $i = 1,\dots,M$. In Section 4, we will approximate the outer expectation in (5) by a multilevel Monte Carlo averaging approach.

3 Approximation of Posterior Expectation

A first simplification of (5) is achieved by replacing the inner expectation over the posterior distribution μ^δ by an approximation $E_{\tau_L}^{\mu^\delta}[\phi]$ with tolerance parameter $\tau_L > 0$. We assume that the following bound holds for the considered QoI ϕ:

$$\left\|\mathbb{E}^{\mu^\delta}[\phi] - E_{\tau_L}^{\mu^\delta}[\phi]\right\|_{\mathscr{X}} \le \tau_L\,. \quad (6)$$

Our method of choice is the adaptive Smolyak algorithm developed in [7], which adaptively constructs a sparse tensor quadrature rule that approximates Z_δ and Z'_δ. More precisely, the results in [7, 8] ensure existence of a monotone index set Λ with

$$\left\| \mathbb{E}^{\mu^\delta}[\phi] - E^{\mu^\delta}_{\tau_L}[\phi] \right\|_{\mathscr{X}} \le C_\Gamma^{\mathrm{SM}} N_L^{-(\frac{1}{p}-1)} , \tag{7}$$

where N_L is the cardinality of the index set Λ assuming that the forward solution map $U \ni \mathbf{y} \mapsto q(\mathbf{y})$ is $(\mathbf{b}, p, \varepsilon)$-analytic for some $0 < p < 1$ and $\varepsilon > 0$, i.e.

instance well-posedness of the forward problem:

for each instance $\mathbf{y} \in U$, there exists a unique realization $u(\mathbf{y}) \in \tilde{X} \subseteq X$ of the uncertainty and a unique solution $q(\mathbf{y}) \in \mathscr{X}$ of the forward problem (1) satisfying $\|q(\mathbf{y})\|_{\mathscr{X}} \le C_0$ for all $\mathbf{y} \in U$.

analyticity:

There exists a $0 \le p \le 1$ and a positive sequence $\mathbf{b} = (b_j)_{j\in\mathbb{J}} \in \ell^p(\mathbb{J})$ such that for every sequence $\rho = (\rho_j)_{j\in\mathbb{J}}$ of poly-radii $\rho_j > 1$ with $\sum_{j\in\mathbb{J}}(\rho_j - 1)b_j \le 1-\varepsilon$, the solution map $U \ni \mathbf{y} \mapsto q(\mathbf{y}) \in \mathscr{X}$ admits an analytic continuation to the open poly-ellipse $\mathscr{E}_\rho := \bigotimes_{j\in\mathbb{J}} \mathscr{E}_{\rho_j} \subset \mathbb{C}^{\mathbb{J}}$ and satisfies $\|q(z)\|_{\mathscr{X}} \le C_\varepsilon(\rho)$, $\forall z \in \mathscr{E}_\rho$.

The concept of $(\mathbf{b}, p, \varepsilon)$-analyticity allows to analyze the regions of analyticity $\mathscr{E}_\rho$ of the solution in each parameter and exploit the anisotropic smoothness of the problem reflected by the poly-radii ρ. Sufficient conditions on the $(\mathbf{b}, p, \varepsilon)$-analyticity of the forward problem (1) are given in [4, 8]. The results presented in [7, 8] suggest dimension robust convergence rates of the form (7) for adaptive Smolyak-based quadrature algorithms using a greedy-type approach to construct the monotone index set. The underlying quadrature points are symmetrized Léja sequences (see [3] and the references therein for more details), which allow us to relate the number of quadrature points to the prescribed tolerance τ_L as follows.

Proposition 1. *The work required for the evaluation of the adaptive Smolyak approximation up to the tolerance $\tau > 0$ based on symmetrized Léja quadrature is bounded by $C(\Gamma)\tau^{-\log_2 3\cdot(\frac{1}{p}-1)^{-1}}$ with $C(\Gamma) > 0$ independent of τ.*

Proof. For a multiindex ν in a monotone index set Λ_N with $\#\Lambda_N \le N$, the bound $\#\{i \in \mathbb{J} : \nu \ne 0\} \le \lfloor \log_2 N \rfloor$ holds as argued in the proof of Lemma 5.4 in [7]. A worst case bound can be derived by considering an isotropic refinement in the first $\lfloor \log_2 N \rfloor$ dimensions, i.e. it holds for the maximal number of quadrature points $M \le 3^{\log_2 N} = N^{\log_2 3}$. Equating (7) to τ, solving for N and inserting into the above yields the claimed bound on the number of quadrature points.

Remark 1. Note that the result derived in Proposition 1 is based on a worst case bound on the number of quadrature points arising in the case of isotropic refinement.

4 Binned Multilevel Monte Carlo

In this section, we formulate our method for combining M realizations of δ, $\Delta = \{\delta_i : i = 1, \ldots, M\}$ to compute an approximation to

$$\mathbb{E}^{\gamma_\Gamma^K}\left[\mathbb{E}^{\mu^\delta}[\phi]\right] = \int_{\mathbb{R}^K} \frac{1}{Z_\delta} \int_U \phi(\mathbf{y}) \exp\left(-\frac{1}{2}||\delta - \mathscr{G}(\mathbf{y})||_\Gamma^2\right)\bigg|_{\delta=\mathscr{G}(\mathbf{y}_0)+\eta} \mu_0(\mathrm{d}\mathbf{y})\, \gamma_\Gamma^K(\mathrm{d}\eta)\,. \tag{8}$$

Our approach is based on the multilevel Monte Carlo method originally applied by Heinrich [6] and Giles [5] and, in the current form, by Barth et al. [2].

Formulation of the Binned MLMC Algorithm. We interpret the approximation $E_{\tau_L}^{\mu^\delta}[\phi]$ obtained by the method explained in Section 3 as corresponding to a discretization level L and write it as a telescopic sum over the levels $\ell = 0,\dots,L$. Using the convention $E_{\tau_{-1}}^{\mu^\delta} = 0$, we obtain the exact reformulation

$$E_{\tau_L}^{\mu^\delta}[\phi] = \sum_{\ell=0}^{L}\left(E_{\tau_\ell}^{\mu^\delta}[\phi] - E_{\tau_{\ell-1}}^{\mu^\delta}[\phi]\right)\,. \tag{9}$$

Inserting this back into (8) and applying the linearity of the expectation yields

$$\mathbb{E}^{\gamma_\Gamma^K}\left[\sum_{\ell=0}^{L}\left(E_{\tau_\ell}^{\mu^\delta}[\phi] - E_{\tau_{\ell-1}}^{\mu^\delta}[\phi]\right)\right] = \sum_{\ell=0}^{L}\mathbb{E}^{\gamma_\Gamma^K}\left[E_{\tau_\ell}^{\mu^\delta}[\phi] - E_{\tau_{\ell-1}}^{\mu^\delta}[\phi]\right]\,.$$

Replacing the expectations on each level by a sample mean over a level-dependent number of samples M_ℓ yields a full approximation to (8),

$$E_{\mathrm{ML},L}^{\gamma_\Gamma^K}[E_{\tau_L}^{\mu^\delta}[\phi]] := \sum_{\ell=0}^{L} E_{M_\ell}^{\gamma_\Gamma^K}\left[E_{\tau_\ell}^{\mu^\delta}[\phi] - E_{\tau_{\ell-1}}^{\mu^\delta}[\phi]\right]\,, \tag{10}$$

where we denote by $E_M[Y]$ the standard Monte Carlo estimator for realizations $\hat{Y}_i$ of a random variable $Y : \Omega \to S$, given by $E_M[Y] = \frac{1}{M}\sum_{i=1}^{M}\hat{Y}_i$.

A crucial aspect of this formulation is the choice of the number of samples per level $(M_\ell)_{\ell=0}^L$ and the tolerances per level $(\tau_\ell)_{\ell=0}^L$. Since the total number of samples is fixed, a natural approach is to make an ansatz for M_ℓ and then choose τ_ℓ optimally.

Number of Samples per Level. Thinking of the levels $\ell = 0,\dots,L$ as "bins" containing measurements over which we wish to average, we distribute the samples according to the ansatz $M_\ell = b^{L-\ell+1}$ with $b \in \mathbb{N}, b > 1$. The analysis presented can also be generalized to the case $b \in \mathbb{R}, b > 1$. The total number of samples is $\sum_{\ell=0}^{L} b^{L-\ell-1}$, which we assume to be the given number of measurements M.

Error Bounds. For the computation of the error, we consider the Gaussian probability space $(\Omega, \mathscr{B}(\Omega), \gamma_\Gamma^K)$ and the random variable η. The approximation of the inner expectation is an $\mathscr{X}$-valued random variable whereas the full expectation is in $\mathscr{X}$. Clearly, $E_{\tau_\ell}^{\mu^\delta}[\phi] \in L^2(\Omega;\mathscr{X})$ and the error of (10) in the $L^2(\Omega;\mathscr{X})$ norm can be bounded by

$$\left\| \mathbb{E}^{\gamma_\Gamma^K}\left[\mathbb{E}^{\mu^\delta}[\phi]\right] - E^{\gamma_\Gamma^K}_{\mathrm{ML},L}[E^{\mu^\delta}_{\tau_L}[\phi]] \right\|_{L^2(\Omega;\mathscr{X})} \leq \left\| \mathbb{E}^{\gamma_\Gamma^K}\left[\mathbb{E}^{\mu^\delta}[\phi]\right] - \mathbb{E}^{\gamma_\Gamma^K}\left[E^{\mu^\delta}_{\tau_L}[\phi]\right] \right\|_{L^2(\Omega;\mathscr{X})}$$
$$+ \left\| \mathbb{E}^{\gamma_\Gamma^K}\left[E^{\mu^\delta}_{\tau_L}[\phi]\right] - \sum_{\ell=0}^{L} E^{\gamma_\Gamma^K}_{M_\ell}\left[E^{\mu^\delta}_{\tau_\ell}[\phi] - E^{\mu^\delta}_{\tau_{\ell-1}}[\phi]\right] \right\|_{L^2(\Omega;\mathscr{X})} . \qquad (11)$$

Since the first term on the right in (11) already contains the expectation with respect to γ_Γ^K, we simply obtain the discretization error from (6), $\|\mathbb{E}^{\mu^\delta}[\phi] - E^{\mu^\delta}_{\tau_L}[\phi]\|_{\mathscr{X}} \leq \tau_L$. Inserting an expansion analagous to (9) into the second term of (11) yields

$$\left\| \sum_{\ell=0}^{L} \left(\mathbb{E}^{\gamma_\Gamma^K}\left[E^{\mu^\delta}_{\tau_\ell}[\phi] - E^{\mu^\delta}_{\tau_{\ell-1}}[\phi]\right] - E^{\gamma_\Gamma^K}_{M_\ell}\left[E^{\mu^\delta}_{\tau_\ell}[\phi] - E^{\mu^\delta}_{\tau_{\ell-1}}[\phi]\right] \right) \right\|_{L^2(\Omega;\mathscr{X})}$$
$$\leq \sum_{\ell=0}^{L} \left\| \mathbb{E}^{\gamma_\Gamma^K}\left[E^{\mu^\delta}_{\tau_\ell}[\phi] - E^{\mu^\delta}_{\tau_{\ell-1}}[\phi]\right] - E^{\gamma_\Gamma^K}_{M_\ell}\left[E^{\mu^\delta}_{\tau_\ell}[\phi] - E^{\mu^\delta}_{\tau_{\ell-1}}[\phi]\right] \right\|_{L^2(\Omega;\mathscr{X})} .$$

For each summand above, we use the standard Monte Carlo error bound that holds for any $M \in \mathbb{N}$, $Y \in L^2(\Omega;\mathscr{X})$, i.e. $\|\mathbb{E}[Y] - E^{\gamma_\Gamma^K}_M[Y]\|_{L^2(\Omega;\mathscr{X})} \leq \frac{1}{\sqrt{M}}\|Y\|_{L^2(\Omega;\mathscr{X})}$. Combining this with the given bound (7) as follows

$$\left\| E^{\mu^\delta}_{\tau_\ell}[\phi] - E^{\mu^\delta}_{\tau_{\ell-1}}[\phi] \right\|_{L^2(\Omega;\mathscr{X})} = \mathbb{E}^{\gamma_\Gamma^K}\left[\left\| E^{\mu^\delta}_{\tau_\ell}[\phi] - E^{\mu^\delta}_{\tau_{\ell-1}}[\phi] \right\|^2_{\mathscr{X}} \right]^{\frac{1}{2}} \leq$$
$$\mathbb{E}^{\gamma_\Gamma^K}\left[\left(\left\| E^{\mu^\delta}_{\tau_\ell}[\phi] - \mathbb{E}^{\mu^\delta}[\phi] \right\|_{\mathscr{X}} + \left\| \mathbb{E}^{\mu^\delta}[\phi] - E^{\mu^\delta}_{\tau_{\ell-1}}[\phi] \right\|_{\mathscr{X}} \right)^2 \right]^{\frac{1}{2}} \leq C_\ell \tau_\ell + C_{\ell-1}\tau_{\ell-1} ,$$

and using $\tau_{-1} = 0$, we obtain a total sampling error bound of

$$\left\| \mathbb{E}^{\gamma_\Gamma^K}\left[E^{\mu^\delta}_{\tau_L}[\phi]\right] - E^{\gamma_\Gamma^K}_{\mathrm{ML},L}[E^{\mu^\delta}_{\tau_L}[\phi]] \right\|_{L^2(\Omega;\mathscr{X})} \leq \sum_{\ell=1}^{L} M_\ell^{-\frac{1}{2}}(C_\ell\tau_\ell + C_{\ell-1}\tau_{\ell-1}) + C_0 M_0^{-\frac{1}{2}} .$$

Combined with the discretization error, the total error is then bounded by

$$\mathrm{e_{tot}} \leq \tau_L + \sum_{\ell=1}^{L} M_\ell^{-\frac{1}{2}}(C_\ell\tau_\ell + C_{\ell-1}\tau_{\ell-1}) + C_0 M_0^{-\frac{1}{2}} . \qquad (12)$$

Theorem 2 (Optimal Tolerances). *Given the sample distribution $M_\ell = b^{L-\ell+1}$, the optimal tolerances for the inner expectation that minimize the total work bound at given error are*

$$\tau_\ell = \frac{M_0^{-\frac{1}{2}}}{C(s,b,L)} \left(\frac{M_\ell}{D_\ell}\right)^{\frac{1}{s+1}} , \quad 0 \leq \ell \leq L ,$$

for a constant $C(s,b,L)$ and $M_{-1} = D_{-1} = 0$, $D_0 = C_0 M_1^{-1/2}$, $D_L = C_L(1 + M_L^{-1/2})$ and for $0 < \ell < L$, $D_\ell = C_\ell(M_\ell^{-1/2} + M_{\ell+1}^{-1/2})$.

Proof. The optimization problem we consider is the minimization of the total work subject to the constraint that the discretization and sampling errors are equilibrated,

$$\min \sum_{\ell=0}^{L} M_\ell w_\ell \quad \text{s.t.} \quad \tau_L + \sum_{\ell=1}^{L} \frac{C_\ell \tau_\ell + C_{\ell-1}\tau_{\ell-1}}{\sqrt{M_\ell}} = \frac{C_0}{\sqrt{M_0}} \ ,$$

where $w_\ell \sim \tau_\ell^{-s}$, $s > 0$ denotes the work on level ℓ (for the Smolyak approach mentioned above, we use $s = (\frac{1}{p} - 1)^{-1}$). Using Lagrange multipliers, one can impose the necessary condition that the partial derivatives of the Lagrange function $\mathscr{L}(\tau_0, \dots, \tau_L, \lambda) = \sum_{\ell=1}^{L} M_\ell \frac{1}{\tau_\ell^s} + \lambda \left(\tau_L + \sum_{\ell=0}^{L} \frac{\tau_\ell + \tau_{\ell-1}}{\sqrt{M_\ell}} - \frac{C_0}{\sqrt{M_0}} \right)$ vanish in the optimum.

A straightforward calculation reveals that the total work when using M samples and the tolerances from above is bounded for $0 < s < 2$ and a constant $C(s,b)$ by $W_{\text{tot}}^L \le C(s,b) M^\gamma w_L$, $\gamma = \frac{2-s}{2(s+1)} \in (0,1)$. A slightly more involved computation yields an optimal error versus work relationship with exponent $-1/2$, independent of s, given by $e_{\text{tot}} \sim C(s,b) \left(W_{\text{tot}}^L \right)^{-1/2}$.

5 Numerical Experiment

As a model problem of the abstract, $(\mathbf{b}, p, \varepsilon)$-analytic operator equation described in Section 2 we consider the diffusion equation $-\nabla \cdot (u \nabla p) = 100x$ in $D := [0,1]$, $p = 0$ on ∂D with stochastic diffusion coefficient u modeled as a random field described by $u = u(\mathbf{y}) := \langle u \rangle + \sum_{j=1}^{64} y_j \psi_j \in X$ with constant mean $\langle u \rangle = 1$, parameters $\mathbf{y} = (y_j)_{j=1}^{64} \in U := [-1,1]^{64}$ and basis functions $\psi_j = \frac{0.9}{j^3} \chi_{D_j}$, $D_j = [\frac{j-1}{64}, \frac{j}{64})$ describing the fluctuations and $X = \cup_{j=1}^{64} C^0(\overline{D_j})$. The problem is solved by a finite element approach with piecewise linear basis functions on a uniform mesh. The meshwidth is $h = 2^{-14}$ to avoid discretization error effects. Given a noisy measurement with $\eta \sim \mathscr{N}(0,1)$, our goal is to evaluate the conditional expectation $\mathbb{E}^{\mu^\delta}[\phi]$ of the QoI $\phi(u) = \mathscr{G}(u)$. The observation operator $\mathscr{O}$ consists of a system response at $x_1 = 0.5$.

For MLMC, the maximal level was chosen by numerically observing that (12) is convex in L and increasing the value of L until the error bound stops decreasing. For each L, b is computed such that $\sum_{\ell=0}^{L} M_\ell = M$ is satisfied. The reference solution is computed to high accuracy using 96-point Gauss-Hermite quadrature.

6 Conclusion

Assuming a given set of measured responses of a forward problem, a multilevel Monte Carlo averaging method was derived by computing optimal values of forward map evaluation tolerances on each level. Numerical results based on Bayesian

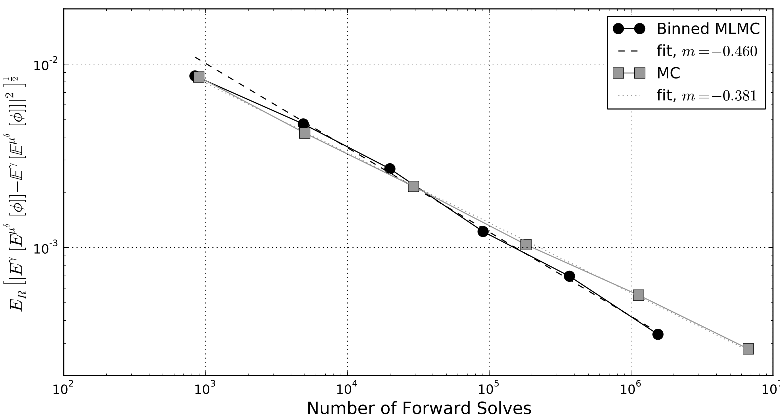

Fig. 1 Convergence of L^2 error approximation $E_R[|E^{\gamma_l^1}_{\mathrm{ML},L}[E^{\mu^\delta}_{\tau_\ell}[\phi]] - \mathbb{E}^{\gamma_l^1}[\mathbb{E}^{\mu^\delta}[\phi]]|^2]^{\frac{1}{2}}$ with $R = 200$ vs. the work, which is assumed proportional to the number of forward evaluations. The theoretically computed rates are $-1/3$ for Monte Carlo (MC) and $-1/2$ for multilevel Monte Carlo (MLMC). The number of measurements were $M = 16, 64, 256, 1024, 4096, 16384$ and all tolerances were scaled with $C = 0.1$. For MLMC, the first point is not used in computing the slope, as $L = 0$ which corresponds to a MC simulation.

inversion of a parametric diffusion equation confirm the analytically derived optimal convergence rate of the error with respect to the work.

Acknowledgments: This work is supported by the Swiss National Science Foundation (SNF) and the European Research Council (ERC) under FP7 Grant AdG247277.

References

1. A. Barth, A. Lang, and Ch. Schwab, *Multilevel Monte Carlo method for parabolic stochastic partial differential equations*, BIT Numerical Mathematics (2011), 3–27.
2. A. Barth, Ch. Schwab, and N. Zollinger, *Multi-level Monte Carlo finite element method for elliptic PDEs with stochastic coefficients*, Numer. Math. **119** (2011), no. 1, 123–161.
3. A. Chkifa, *On the Lebesgue constant of Leja sequences for the complex unit disk and of their real projection*, J. Approx. Theory **166** (2013), 176–200.
4. A. Cohen, A. Chkifa, and Ch. Schwab, *Breaking the curse of dimensionality in sparse polynomial approximation of parametric pdes*, Journ. Math. Pures et Appliquees **103** (2014), no. 2, 400–428.
5. M. B. Giles, *Multilevel Monte Carlo path simulation*, Operations Research (2008), 1–25.
6. S. Heinrich, *Multilevel Monte Carlo methods*, Large-scale scientific computing (2001), 58–67.
7. C. Schillings and Ch. Schwab, *Sparse, adaptive Smolyak quadratures for Bayesian inverse problems*, Inverse Problems **29** (2013), no. 6, 065011.
8. ———, *Sparsity in bayesian inversion of parametric operator equations*, Inverse Problems **30** (2014), no. 6, 065007.
9. Ch. Schwab and A. M. Stuart, *Sparse deterministic approximation of Bayesian inverse problems*, Inverse Problems **28** (2012), no. 4, 045003.
10. A. M. Stuart, *Inverse problems: a Bayesian perspective*, Acta Numer. **19** (2010), 451–559.