Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Identification of Uncertain, Spatially Varying Parameters through Multilevel Sampling

Hans-Werner van Wyk *

* *Department of Scientific Computing, Florida State University, FL 32306 USA (e-mail: hvanwyk@fsu.edu)*

Abstract: While stochastic sampling methods remain widely used to simulate solutions of partial differential equations with random parameters, due to ease of implementation on a parallel computing architecture, they can nevertheless carry a considerable computational cost, especially if reliable statistics are sought at high levels of spatial fidelity. Multilevel sampling methods improve upon the efficiency of traditional sampling schemes without compromising on accuracy and parallelizability, by dynamically incorporating the model's spatial discretization into the sampling procedure. Moreover, the diagnostic estimates that are computed to monitor the solution's convergence behaviour during the course of the algorithm, can be used to inform adaptive spatial refinement strategies. In this paper we leverage these properties to develop an adaptive, multilevel algorithm for identifying the statistical distribution of a spatially varying parameter from the statistical description of a related model output, based on experimental measurements.

Keywords: Parameter identification, uncertainty, statistical inference, multilevel sampling, adaptive algorithm, distributed-parameter systems.

1. INTRODUCTION

As numerical simulation plays an increasingly central role in modern decision making, it becomes critical to assess the reliability of numerical models by quantifying the effect of uncertainty in the underlying model parameters on model outputs.

Suppose q is a random, spatially varying parameter whose statistical distribution is known. To be specific, $q(x, y)$ is a parameter that depends on both a spatial variable $x \in D \subset \mathbb{R}^d$, where D denotes an underlying spatial domain, and on a random vector $y \in \Gamma \subset \mathbb{R}^N$, whose statistical variation is determined by its density function $\rho : \Gamma \to [0, \infty)$. Consequently, the model response u of the system also takes the form of a random field that depends on the same underlying random vector, i.e. $u = u(x, y)$.

As a model problem, we consider the second order elliptic equation with Dirichlet boundary conditions and deterministic forcing term f, i.e.

$$\begin{aligned} -\nabla \cdot (q(x,y)\nabla u(x,y)) &= f(x) \text{ in } D \times \Gamma \\ u(x,y) &= 0 \quad \text{in } \partial D \times \Gamma. \end{aligned} \tag{1}$$

This equation is can be viewed as a collection of deterministic elliptic equations parameterized by the random vector y an may be used to model the temperature distribution in a material with random conductivity, or subsurface flow through a medium with random diffusion coefficient. Under the assumption that $f \in L^2(D)$ and $q \in L^\infty(\Gamma, C^1(\bar{D}))$ so that $0 < q_{\min} < q(x, y)$ a.s. on $\Gamma \times D$ for constant $q_{\min}$, the solution u of (1) can be shown to exist and be unique, see Babŭska (2007).

A statistical description of the model output u, or some related physical quantity, such as a local spatial average or a boundary flux, usually takes the form of a set of statistical quantities of interest Q, such as measures of central tendency, spread, and correlation, as well as exceedence probabilities and confidence intervals. A large class of such statistical quantities of interest Q can be written in the form of a stochastic integral,

$$Q = E[v] := \int_\Gamma v(x, y)\rho(y)\, dy, \tag{2}$$

where $v = G(u)$ for some mapping G. The probability that the local spatial average $\bar{u}(y) = \int_{D_1} u(x, y)\, dx$ of u over the region $D_1 \subset D$ exceeds some upper threshold $u_{\max}$ could for instance be written as the integral

$$\mathbb{P}(\bar{u} \geq u_{\max}) = \int_\Gamma \chi_{\{\bar{u} \geq u_{\max}\}}(y)\rho(y)\, dy,$$

where χ_A represents the indicator function over a set $A \subset \Gamma$.

Stochastic sampling methods can be thought of as numerical quadrature approximations $I_\eta[v]$ of Q, computed as a weighted sum of of the form $I_\eta[v] := \sum_{i=1}^{\eta} w_i v^{(i)}$, where $\{v^{(i)}\}_{i=1}^{\eta}$ are sample paths corresponding to different values of the random input vector y. These may be generated randomly from the underlying density function ρ, such as in Monte Carlo sampling, or may be based on a low discrepancy sequence in the stochastic domain Γ, such as in quasi Monte Carlo methods (see Kuo (2012a)), or chosen according to a fixed set of predetermined abscissas defined on Γ, such as in stochastic collocation (Babŭska

(2007)). In all cases, sample paths $v^{(i)}$ can be computed independently, allowing for straightforward parallelization.

In practice, the model parameter q is rarely known and must be estimated from measurements of related observable quantities, most often measurements $\hat{u}$ of the model response u. The task of inferring information about q from $\hat{u}$ constitutes an inverse problem, which is typically ill-posed. Additionally, in the presence of limited, noisy measurements, q can only be known in a probabilistic sense, i.e. as a random field. Suppose, we have a statistical description of the measurements $\hat{u}$, as a random field, i.e. $\hat{u}(x, y)$, where $x \in D$ is a spatial variable and $y \in \Gamma$ a random vector with joint density function $\rho : \Gamma \to [0, \infty)$. The corresponding statistical inverse problem then amounts to obtaining a statistical description of q from that of $\hat{u}$.

In Borggaard (2012), we consider a least squares type formulation of the statistical inverse problem related to (1), in which we seek to identify the random parameter q so as to minimize of the discrepancy between the statistical distribution of the corresponding model output and that of $\hat{u}$, measured in the appropriate stochastic Sobolev norm. Furthermore, we develop an augmented Lagrangian optimization algorithm (see Ito (1990)) to simultaneously approximate the spatial-, as well as stochastic components of q.

In this paper we consider an alternative, sampling-based algorithm, in which sample paths $q^{(i)}$ of q are generated from sample paths $\hat{u}^{(i)}$ of $\hat{u}$, by solving a deterministic inverse problem. It can readily be shown that parameters with such paths minimize the cost functional in the aforementioned least squares formulation, while being more amenable to parallel implementation. These paths can then be statistically aggregated to yield descriptive statistical quantities of interest Q. In this case, the integrand in (2) is of the form $v(x, y) = G(q)$. We are interested in approximating integrals of this form, using sampling.

As an illustration of how naturally this method arises in an experimental or industrial setting, consider the problem of determining the statistical distribution of the conductivity coefficient q in a collection of plates that are manufactured from an inhomogeneous material. An intuitively appealing procedure to obtain a statistical estimate of q is to subject a random sample of plates to a heat source under similar conditions and record the temperature distribution for each plate. Upon estimating the conductivity coefficient deterministically for each plate in the sample, we can statistically aggregate the results. It therefore seems like a natural idea to treat the inverse problem itself as a sampling problem.

Certainly, accurate estimates of statistics Q relating to q may require a large sample, depending on the statistical complexity of the measurements. Moreover, since deterministic parameter identification methods usually take the form of an optimization routine that computes $q^{(i)}$ iteratively from an initial guess $q_0^{(i)}$, each sample path of q may be costly to compute, requiring multiple forward- and adjoint solves.

In the following sections, we show how these shortcomings can be mitigated through the use of multilevel sampling methods. Not only do these improve the allocation of computational resources through coordination of the sample sizes with spatial refinement levels, but also provide good initial guesses $q_0^{(i)}$ for the deterministic parameter identification algorithms, based on coarse approximations, provided the underlying stochastic quadrature grid is nested. Lastly, in situations where statistical fluctuations are localized, the statistical variation of the corrections between successive spatial refinement levels can be used to inform adaptive mesh refinement strategies.

The remainder of this paper is structured as follows. Section 2, In section 3 we outline how multilevel methods can be modified to reduce the overall computational cost of sampling based parameter estimation, without compromising on the accuracy. In Section 4, we then demonstrate the workings of the algorithm by means of a numerical example.

2. STOCHASTIC SAMPLING METHODS

The type of sampling scheme used is usually determined by factors such as the stochastic dimension N, the complexity of the underlying parameter space, and the regularity of v as a function of y.

Monte Carlo methods, well-known for their robustness with respect to stochastic dimension and regularity, nevertheless have the notoriously slow convergence rate of $O(\eta^{-\frac{1}{2}})$. Quasi Monte Carlo sampling schemes exhibit the stronger convergence rate of $O(\eta^{-1}\log(\eta)^N)$ (see Kuo (2012a)), which deteriorates, however, as the stochastic dimension grows. For integrands that depend relatively smoothly on the underlying random vector y, interpolation-based quadrature methods, such as sparse grid stochastic collocation, have shown to yield considerably higher convergence rates than the aforementioned schemes (see Nobile (2008)). For problem (1), such smoothness can be determined *a priori*, based on that of the input parameter q. These smoothness requirements can, however also be partially circumvented in moderate stochastic dimensions N, by using interpolants with local support, instead of polynomials.

Throughout this paper, we make use of sparse grid stochastic collocation, based on Clenshaw-Curtis grids. These grids have the additional benefit of being nested, so that high order stochastic interpolants can be constructed hierarchically from lower order ones.

For distributed parameter systems such as (1), the computation of the statistical quantity of interest Q requires not only the approximation of the integral in (2), but often also the spatial estimation $v_h(y)$ of sample paths v. Here, we make use of the standard finite element space of piecewise polynomials based on a regular triangulation $\mathcal{T}_h$ of D with maximum mesh spacing parameter $h = \max_{\tau\in\mathcal{T}_h} \mathrm{diam}(\tau)$.

The two predominant sources of error in the estimation of Q are therefore the spatial error, that results from the numerical approximation $v_h^{(i)}$ of each sample path $v^{(i)}$ and depends on the mesh spacing parameter h, and the sampling error, that results from the quadrature approximation of the integral (2) and depends on the sample size η. In fact, denoting the resulting estimate of Q by

$\widehat{Q}_{\eta,h} := I_\eta[v_h]$, we can decompose the total approximation error into two parts, i.e.

$$\begin{aligned}\|Q - \widehat{Q}_{\eta,h}\| =& \|E[v] - E[v_h] + E[v_h] - I_\eta[u_h]\| \\ \leq& \|E[v - v_h]\| + \|E[v_h] - I_\eta[v_h]\|, \end{aligned} \quad (3)$$

where $\|\cdot\|$ is the norm with which the discrepancy in the statistical quantity of interest is measured.

2.1 Improving the Efficiency of Sampling Methods through Multilevel Sampling

A convenient measure of the efficiency of a sampling method is given by its ε-cost $\mathcal{C}_\varepsilon$, the computational cost needed to obtain an overall error of order ε. Let $\mathcal{C}(v_h^{(i)})$ be the computational cost of generating each sample path $v_h^{(i)}$. Under the assumption that it is more or less the same for all paths, depending only on the underlying mesh spacing parameter, i.e. $\mathcal{C}(v_h^{(i)}) = \mathcal{C}_h$ for all $i = 1, ..., \eta$, we can determine the total computational cost through $\mathcal{C}(\widehat{Q}_{\eta,h}) = \eta C_h$. This expression, together with the overall error estimate (3), imply that an accurate approximation $\widehat{Q}_{\eta,h}$ of Q requires a large number of samples, each of which should be generated on a fine mesh, leading to a considerable computational overhead, depending on the extent to which $\mathcal{C}_h$ increases as $h \to 0^+$.

Multilevel sampling schemes aim to reduce the ε-cost through the use of a hierarchy of spatial refinement levels, by generating the bulk of the sample paths $v_h^{(i)}$ cheaply on coarser meshes while sampling more sparingly at higher refinement levels. Originally developed as a variance reduction algorithm for Monte Carlo sampling (Cliffe (2011)), it has since been shown for both quasi Monte Carlo- (Kuo (2012b)) and stochastic collocation methods (van Wyk (2013)), that it is possible to coordinate the sample size with the spatial refinement so as to considerably reduce the ε-cost, while maintaining the same level of accuracy.

3. MULTILEVEL SAMPLING METHODS FOR PARAMETER IDENTIFICATION

Let $v(x, y) = G(q)$ be a random field whose stochastic integral yields a desired statistical quantity of interest Q related to the distribution of q. We are then interested in approximating an integral of the form (2).

Let $\{h_\ell\}_{\ell=0}^L$ be a sequence of spatial discretization parameters giving an increasing level of accuracy, where h_L is chosen such that

$$\|E[v - v_{h_L}]\|_W \leq \frac{\varepsilon}{2}.$$

We can now write sample paths of this fine-scale approximation v_{h_L} of v as the sum

$$v_{h_L} = v_{h_0} + \sum_{\ell=1}^{L}(v_{h_\ell} - v_{h_{\ell-1}}) =: \sum_{\ell=0}^{L} \triangle v_\ell, \text{ where}$$

$$\triangle v_\ell := \begin{cases} v_{h_0} & \text{if } \ell = 0 \\ v_{h_\ell} - v_{h_{\ell-1}} & \text{if } \ell > 0 \end{cases}$$

The expected value, $E[v_{h_L}]$ can then be expanded to

$$E[v_{h_L}] = \sum_{\ell=0}^{L} E[\triangle v_\ell],$$

a sum which in turn may be estimated by a sum of numerical integrals of the form

$$\widehat{Q}_{\{\eta_\ell\},\{h_\ell\}} := \sum_{\ell=0}^{L} I_{\eta_\ell}[\triangle v_\ell]$$

where the sample sizes η_ℓ may be chosen separately for each spatial refinement level ℓ. If it is possible to choose $\eta_0, ..., \eta_L$ so that the bulk of the sample paths are computed cheaply at the lower refinement levels whereas those at higher refinement levels are sampled sparingly, while maintaining an overall error within tolerance,

As before, the total error splits into a spatial error and a multilevel sampling error. Indeed,

$$\begin{aligned} &\|Q - \widehat{Q}_{\{\eta_\ell\},\{h_\ell\}}\| \\ =& \left\| E[v] - E[v_{h_L}] + \sum_{\ell=0}^{L}(E[\triangle v_\ell] - I_{\eta_\ell}[\triangle v_\ell]) \right\| \\ \leq& \|E[v - v_{h_L}]\| + \sum_{\ell=0}^{L} \|E[\triangle v_\ell] - I_{\eta_\ell}[\triangle v_\ell]\|. \end{aligned} \quad (4)$$

Moreover, under the assumptions made earlier concerning the computational cost, the overall computational cost $\mathcal{C}(\widehat{Q}_{\{\eta_\ell\},\{h_\ell\}})$ of generating a multilevel estimate takes the form of the sum

$$\mathcal{C}(\widehat{Q}_{\{\eta_\ell\},\{h_\ell\}}) = \sum_{\ell=0}^{L} \eta_\ell \mathcal{C}_{h_\ell}. \quad (5)$$

3.1 The Optimal Allocation Sub-problem

In light of (4) and (5), the optimal choice of sample sizes $\eta_0, ..., \eta_L$ can now be formulated as the discrete constrained optimization problem

$$\begin{aligned} &\min_{\eta_0,\eta_1,...,\eta_L} \sum_{\ell=0}^{L} \eta_\ell \mathcal{C}_{h_\ell} \\ &\text{subject to } \sum_{\ell=0}^{L} \|E[\triangle v_\ell] - I_{\eta_\ell}[\triangle v_\ell]\| \leq \frac{\varepsilon}{2} \end{aligned} \quad (6)$$

For Monte Carlo sampling, the multilevel sampling error can readily be seen to satisfy

$$\|Q - \widehat{Q}_{\{\eta_\ell\},\{h_\ell\}}\|^2 = \sum_{\ell=0}^{L} \eta_\ell^{-1} \text{var}(\triangle v_\ell),$$

where $\text{var}(\cdot)$ denotes the variance. Consequently, explicit formulae can be derived for the sample sizes $\{\eta_\ell\}_{\ell=0}^L$.

For stochastic collocation methods such explicit upper bounds are, however, more difficult to obtain. Even for the stochastic forward problem for which the model output depends smoothly on y and the error can be shown to have the form

$$\|\mathbb{E}[v_h] - I_\eta[v_h]\| \leq c\eta^{-\mu_2} \log(\eta)^{\mu_1} \varphi(v_h), \quad (7)$$

where the constants $c, \mu_1, \mu_2 > 0$ and $\varphi(\cdot)$ is a non-negative functional loosely related to the 'size' of the integrand, the precise convergence rates must be estimated.

If stochastic collocation is used and the smoothness of v cannot be verified *a priori*, as in the case of inverse sampling, or if the sampling strategy has no useful *a*

priori error bound, such as adaptive sampling methods, no explicit formulae exist and the convergence rate, as well as the solution to the optimal allocation sub-problem, must be approximated numerically. In our numerical examples, we calculate the relative sampling error of estimates computed on successive stochastic grids, and approximate the corresponding log-log plot of sample size versus relative sampling error by a piecewise linear interpolant. We then estimate the minimizer of problem (6) by means of an active set method.

For sampling schemes based on interpolatory numerical quadrature, there are additional constraints as to the precise values of the sample sizes that are admissible. In this case additional binning of the sample sizes may be necessary, which potentially decreases the efficiency of the algorithm. This can be partially mitigated by sorting the levels according to the computational work required for each unit of error and refining until the sample error is below tolerance.

3.2 The Multilevel Algorithm

A practical implementation of the multilevel sampling method is outlined in Algorithm 3.2. Given an initial coarse mesh $\mathcal{T}_0$ and error tolerance $\varepsilon > 0$, we compute a single-level estimate $\widehat{Q}_{\eta,h}$ of Q and use it to estimate the relative sampling error. We then proceed to refine the mesh until the spatial error, estimated here by comparing approximations on successive meshes, is within tolerance. After each mesh refinement, we re-compute the optimal sample sizes $\eta_0, ..., \eta_L$ as minimizers of (6) and adjust the multilevel estimate $\widehat{Q}_{\{\eta_\ell\},\{h_\ell\}}$ when necessary. Note that for Monte Carlo sampling, each new sample path of the correction term $\triangle v_\ell^{(i)}$ must be independent. We therefore need to generate two new sample paths, one at the present- and one at the previous spatial refinement level. For sparse grid stochastic collocation methods based on nested grids on the other hand, sample paths can be re-used.

Input: ε, h_0.
Output: Multilevel estimate $\widehat{Q}_{\{\eta_\ell\},\{h_\ell\}}$ of Q
while $e_L^{\text{space}} > \frac{\varepsilon}{2}$ **do**
  $L \leftarrow L + 1$;
  Refine grid at new discretization level h_L;
  Determine optimal sample sizes $\{\eta_0, ..., \eta_L\}$;
  Generate samples $\left\{\triangle v_\ell^{(i)}\right\}_{i=1}^{\eta_\ell}$ for $\ell = 0, ..., L$;
  Update the multilevel estimate $\widehat{Q}_{\{\eta_\ell\},\{h_\ell\}}$
end while

Algorithm 1. The basic multilevel sampling algorithm

Note that both the spatial- and sampling errors are not available and must be estimated, using successive relative differences for instance.

3.3 Spatial Adaptivity

In cases where the statistical quantity of interest Q is spatially varying, the statistical distribution of the correction terms $\triangle v_\ell$ can be used to assess the spatial distribution of the sampling error. Refinement should thus occur in regions where the variation in $\triangle v_\ell$, i.e. the discrepancy between successive estimates, is large. In our numerical example, we compute spatial averages of the variance of the correction terms over each element and compare these values to the overall spatial average. In an effort to equilibrate the variance over the entire region, we then subdivide elements in which the variance exceeds a predetermined fraction of the overall average, all the while ensuring that there are no hanging nodes or skinny triangles.

4. NUMERICAL EXAMPLE

In this section, we demonstrate the use of multilevel sampling in a parameter identification problem in which the parameter's stochastic variability is restricted to a fixed region of the domain. Consider the second order elliptic equation (1) defined on the region $D = [0, 1]^2$, with discontinuous conductivity coefficient q

$$q(x, r) = \begin{cases} 2, \text{ if } |x| \leq r \\ 1, \text{ if } |x| > r, \end{cases}$$

where the radius r varies randomly with $5(r - 0.75) \sim \text{Beta}(2, 5)$ and $|x|$ denotes the euclidean norm. The statistical distribution of the transition boundary is given in Figure 1. We generate samples of $\hat{u}$ by solving the stochastic forward problem on a coarse mesh an adding random noise whose magnitude is about 5% of the norm $\|\mathbb{E}[\hat{u}]\|$. We are interested in estimating the expected value of q. Other statistical quantities can also be estimated.

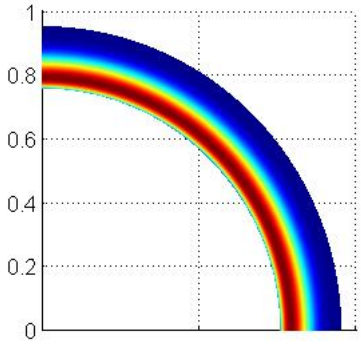


Fig. 1. Distribution of the transition boundary

Throughout the multilevel sampling procedure, we refine adaptively by equilibrating the variance of the correction terms over the entire spatial domain, as discussed in Subsection 3.3. The resulting meshes are shown in Figure 2. As expected, the mesh is fine over the subregion in which the discontinuity of diffusion coefficient q can lie.

Figure 3 shows the improvement in the estimates of the expected value $\mathbb{E}[q]$ of q for successive levels of spatial refinement. The top row represents the estimates themselves, while the bottom row depicts the correction terms, or improvement of the current estimate over the previous one. Not only do the magnitudes of successive correction terms decrease, but they also become more concentrated in certain parts of the domain.

Figure 4 shows the estimated spatial- as well as sampling errors. For this problem, the spatial error seems much harder to resolve than the sampling error. Indeed, a sample size of only 17 is sufficient to achieve an approximate sampling error within tolerance, while 5 levels of spatial refinement are necessary for the spatial error. As a

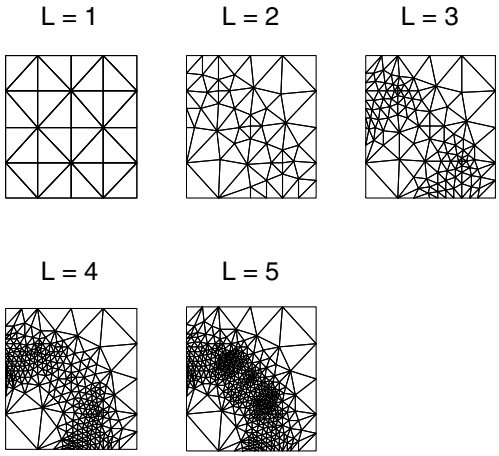


Fig. 2. Mesh at successive levels of spatial refinement

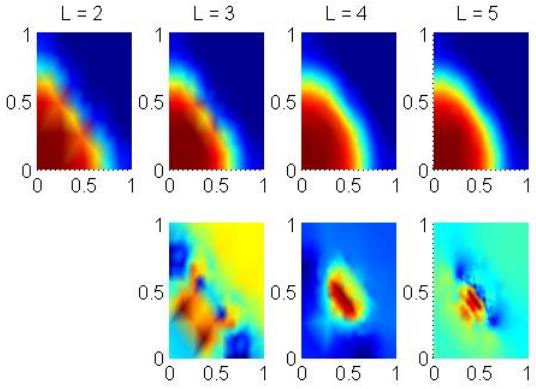


Fig. 3. Estimates of the mean of q, for increasing levels spatial refinement (top row) and correction terms (bottom row).

consequence, most computational effort is expended on the finest refinement levels, where the cost per solve is considerable, as shown in Figure 5. Nevertheless, the total computational effort spent by the multilevel algorithm amounts to 968.7573 while the single level sample with the same accuracy has a total computational cost of 2046.4.

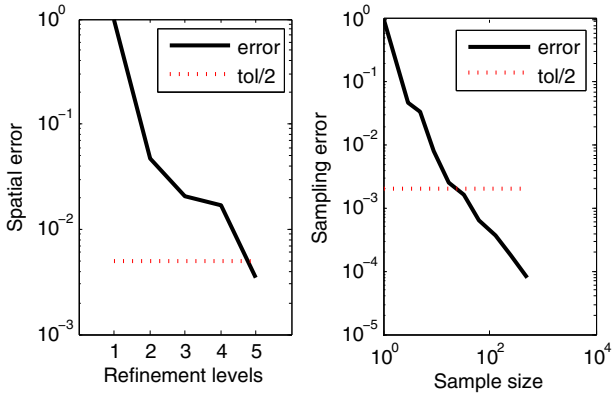
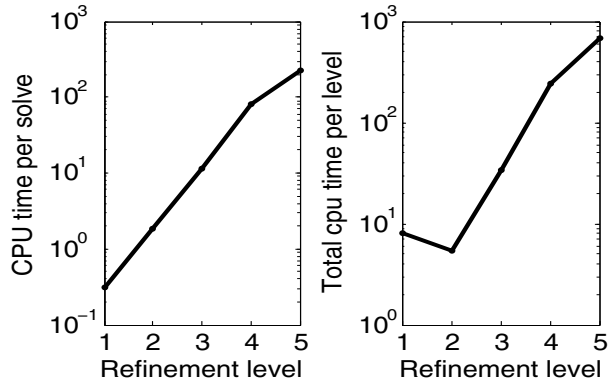


Fig. 4. Estimates of the spatial- and sampling errors, together with the requisite tolerance levels



s

Fig. 5. The cpu time required to generate a single sample path at each refinement level (left) and the total computational time spent on each refinement level (right).

5. CONCLUSION

In this paper, we have shown how multilevel sampling can be used to estimate the statistical variation of uncertain, spatially distributed parameters, based on a statistical description of the related model output. Since such statistical inverse problems are generally ill-posed, the regularity of q as a function of y is not easily determined and adaptive sampling methods may be necessary.

REFERENCES

I. Babŭska, F. Nobile, and R. Tempone. A stochastic collocation method for elliptic partial differential equations with random input data *SIAM Journal on Numerical Analysis*, SIAM, 2007 volume 45, pages 1005–1034

K.A. Cliffe, M.B. Giles, R. Scheichl, A.L. Teckentrup. Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients, *Computing and Visualization in Science*, 2011, volume 14, number 1, pages 3–15

K. Ito, K. Kunisch. The augmented Lagrangian method for parameter estimation in elliptic systems *SIAM Journal on Control and Optimization*, 1990, volume 28, pages 113–136

F. Y. Kuo, C. Schwab, I.H. Sloan. Quasi-Monte Carlo finite element methods for a class of elliptic partial differential equations with random coefficients *SIAM Journal on Numerical Analysis*, SIAM, 2012, volume 50, pages 3351–3374

F. Y. Kuo, C. Schwab, I.H. Sloan. Multi-level quasi-Monte Carlo finite element methods for a class of elliptic partial differential equations with random coefficients *Unpublished*, arXiv preprint arXiv:1208.6349, 2012

H.-W. van Wyk, M. Gunzburger. Multilevel sparse grid methods for partial differential equations with random coefficients *Computers & Mathematics with Applications, submitted*

J. Borggaard, H.-W. van Wyk. Optimization-based estimation of random distributed parameters in elliptic partial differential equations *Proc. 51st IEEE Conference on Decision and Control*, 2012

F. Nobile, R. Tempone, C.G. Webster. A sparse grid stochastic collocation method for partial differential equations with random input data *SIAM Journal*

on Numerical Analysis, SIAM, 2008, volume 46, pages 2309–2345