Convergence analysis of a multi-level kriging model
Application to UQ in CFD

Zhang, Y.; Dwight, R. P.; Han, Z. H.

Publication date
2019

Document Version
Final published version

Published in
Proceedings of the 3rd International Conference on Uncertainty Quantification in Computational Sciences and Engineering, UNCECOMP 2019

Citation (APA)
Zhang, Y., Dwight, R. P., & Han, Z. H. (2019). Convergence analysis of a multi-level kriging model: Application to UQ in CFD. In M. Papadrakakis, V. Papadopoulos, & G. Stefanou (Eds.), *Proceedings of the 3rd International Conference on Uncertainty Quantification in Computational Sciences and Engineering, UNCECOMP 2019* (pp. 400-419). (Proceedings of the 3rd International Conference on Uncertainty Quantification in Computational Sciences and Engineering, UNCECOMP 2019). National Technical University of Athens.

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Available online at www.eccomasproceedia.org
Eccomas Proceedia UNCECOMP (2019) 400-419

ECCOMAS
Proceedia

UNCECOMP 2019
3rd ECCOMAS Thematic Conference on
Uncertainty Quantification in Computational Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou (eds.)
Crete, Greece, 24-26 June 2019

CONVERGENCE ANALYSIS OF A MULTI-LEVEL KRIGING MODEL: APPLICATION TO UQ IN CFD

Y. Zhang[1,2], R.P. Dwight[2], and Z.-H. Han[1]

[1]School of Aeronautics, Northwestern Polytechnical University
Youyi west road 127, 710072 Xi'an, People's Republic of China
zhangyu91@mail.nwpu.edu.cn, hanzh@nwpu.edu.cn

[2] Aerodynamics Group, Faculty of Engineering, Delft University of Technology
Kluyverweg 1, 2629 HS Delft, The Netherlands
r.p.dwight@tudelft.nl

Abstract. *Multi-level surrogate modelling offers the promise of fast approximation to expensive simulation codes for the purposes of uncertainty quantification (UQ). The hope is that a large number of cheap samples from the simulator on coarse grids, can be corrected by a few expensive samples on a fine grid, to build an accurate surrogate. Of the various multi-level approaches, a correction-based method using Gaussian process regression (Kriging) is studied here. In particular, we examine the "additive bridge-function" method, for which – although widely applied – results on theoretical convergence rates and optimal numbers of samples per level are not present in the literature. In this paper, we perform a convergence analysis for the expectation of a quantity of interest (QoI), utilizing convergence results for single-fidelity Kriging, as well as existing multi-level analysis methodology previously applied in context of polynomial-based methods. Rigorous convergence and computational cost analyses are provided. By minimizing the total cost, optimal numbers of sampling points on each grid level are determined. Numerical tests demonstrate the theoretical results for: a 2d Genz function, Darcy flow with random coefficients, and Reynold-Averaged Navier-Stokes (RANS) for the flow over an airfoil with geometric uncertainties. The efficiency and accuracy of this method are compared with standard- and multi-level Monte Carlo. All the test cases show that using our multi-level kriging model significantly reduces cost.*

Keywords: multi-level, gaussian process, uncertainty quantification, computational fluid dynamics, Kriging.

ISSN:2623-3339
Peer-review under responsibility of the organizing committee of UNCECOMP 2019.
doi: 10.7712/120219.6349.18547

1 INTRODUCTION

With the rapid growth of computational capacity and improvements in CFD simulation techniques over the past two decades, CFD-based aerodynamic analysis and design have become standard in industry. However, in any analysis of real-world systems there exist uncertainties and errors, e.g.: discretization error, geometry uncertainty and turbulence model-form uncertainty, which can cause the prediction of performance to be poor. Thus, it is necessary to consider the effect of uncertainties on simulation predictions, i.e. which leads inexorably to solving stochastic PDEs. Statistics of the solutions, such as the expectation of a Quantity-of-Interest (QoI), $\mathbb{E}[y]$, can be straightforwardly estimated with Monte-Carlo or Taylor expansions, however, due to the high computational cost of individual CFD simulations, the lack of reliable derivatives, and strong nonlinearities, these methods are often impractical.

Surrogate-based methods on the other hand, can use a modest number of simulations to provide a fast approximation to $\mathbb{E}[y]$, provided the stochastic dimension is moderate. Some common surrogates in the literature use polynomial-chaos expansions (PCE) for interpolation or regression (often with sparsity), radial basis-function (RBF) interpolation, and Kriging. Of these, Kriging is notable for its high flexibility, thanks to its Bayesian roots – and has performed well in practical applications. Multi-fidelity and multi-level variants have been developed, which use additional low-cost simulations to assist in the estimation of $\mathbb{E}[y]$. In *multi-fidelity* methods the low-cost simulation uses a simplified model of the problem (e.g. Euler versus Navier-Stokes); in *multi-level* the low-cost simulation is the same continuous model as the high-cost, but discretized at a coarser grid resolution. In the former case the correlation between high- and low-fidelities is responsible for the cost reduction (if any); in the multi-level case we can use stronger relationships given by the rate of (grid-)convergence of the PDE solver. Although multi-fidelity and multi-level Kriging methods are widely applied in engineering, they are known to be unreliable, and do not consistently reduce costs in practice. This is at least partially a result of the lack of the convergence analysis for these methods – and as a consequence, the lack of rules for optimal sample selection.

There has been enormous work on multi-fidelity or multi-level surrogate modelling. Haftka et al. [1] [2] developed a variable-fidelity kriging model, which uses a multiplicative bridge function to correct the low-fidelity model to approximate the high-fidelity function. Gano et al. [3] developed a hybrid bridge function method, which uses a kriging model to scale the low-fidelity model. Han et. al. [4] improved variable-fidelity surrogate modeling via gradient-enhanced kriging and a generalized hybrid bridge function, to realize a more accurate and robust model. Cokriging was originally proposed in geostatistics community by Journel et al. [5] and then extended to deterministic computer experiments by Kennedy and O'Hagan, called KOH autoregressive model [6]. Han et. al. [7] proposed an improved version of cokriging, which can be built in one step, and a hierarchical kriging model [8], which avoids the cross-variance between low- and high-fidelity model thus is more robust. More recently, the multi-fidelity/level models are introduced into the field of uncertainty quantification. Palar et al. [9] developed a multi-fidelity non-intrusive polynomial chaos method based on regression, which builds two PCEs for both the low-fidelity and correction functions, and then sum it up to provide an estimation for the high-fidelity function. This method has been applied for flow around a NACA0012 airfoil and a Common Research Model wing with flow condition uncertainties, e.g. Mach number, angle of attack. Parussini et al. [10] proposed a recursive multi-fidelity cokriging model and tested it by stochastic Burgers equation and the stochastic Oberbeck-Boussinesq equations. Palar et al. [11] investigated the capability of a Hierarchical Kriging model for

uncertainty analysis and further improve it by combining with PCE method. The application to RAE2822 airfoil and CRM wing with flow condition uncertainty shows its high accuracy and robust performance. Narayan et al. [12] proposed a multi-fidelity stochastic collocation method, which leverage inexpensive low-fidelity models to generate surrogates for an expensive high-fidelity model using a parametric collocation (nonintrusive) approach. Zhu et al. [13] present a bi-fidelity algorithm for approximating the statistical moments of stochastic problems and provide a basic error analysis.

In this paper, We choose multi-level rather than multi-fidelity [14], in order to make use of stronger convergence results. The popularity of multi-level methods has increased dramatically in recent years, thanks to the success of Multi-Level Monte-Carlo (MLMC) methods [15, 16, 17]. By correctly choosing the number of Monte-Carlo samples per level, the cost of solving the stochastic PDE can be reduced to a constant multiple of the cost of a single deterministic solution, in the best case. Similar ideas were used by Teckentrup et al. [18] to devise a multi-level stochastic collocation method, dramatically improving upon MLMC for moderate stochastic dimension. This paper addresses the convergence of a particular multi-level method known as Additive Bridge-Function-based multi-level Kriging [19, 20, 21] for estimation of $\mathbb{E}[y]$. We use the Kriging mean as response surface only, the Kriging variance is not used in our analysis. Our work follows closely the outline of [18], but considering Kriging models rather than polynomial models. We employ error bounds derived for RBF interpolation [22] to estimate the interpolation error in the Kriging mean. Finally we provide expressions for optimal number of samples per level to obtain minimum computational cost.

The structure of this paper is as following: in Section 2, the additive-bridge function multi-level Kriging model is described, and in Section 3 it's convergence properties are analysed and computational cost estimates are provided. Section 4 briefly introduces the numerical test cases used in this paper, and numerical results are presented in Section 5, and compared to standard MC and MLMC.

2 METHODOLOGY

A single-fidelity ordinary kriging model is presented in Section 2.1, see also e.g. [23]; and then we describe how to construct a multi-level Gaussian process model from multiple single-level models in Section 2.2.

2.1 Single-fidelity ordinary Kriging

Consider a QoI $y \in \mathbb{R}$, which (possibly via a PDE) is a function of (deterministic) variables $\boldsymbol{\xi} \in \mathbb{R}^M$. Ordinary Kriging represents $y(\boldsymbol{\xi})$ by a Gaussian process Y of the form:

$$Y(\boldsymbol{\xi}) = \rho + Z(\boldsymbol{\xi}), \tag{1}$$

where ρ is an unknown constant and $Z(\boldsymbol{\xi})$ is a stationary Gaussian random process with zero-mean and covariance

$$\mathrm{Cov}[Z(\boldsymbol{\xi}), Z(\boldsymbol{\xi}')] = \sigma^2 R(|\boldsymbol{\xi} - \boldsymbol{\xi}'|). \tag{2}$$

Here $R : \mathbb{R}^+ \to \mathbb{R}$ is a positive-definite covariance kernel, so that the covariance of two points of the process only depends on their Euclidean distance in $\boldsymbol{\xi}$-space, and σ is the standard-deviation.

Given observations $\boldsymbol{y}_{\boldsymbol{\Xi}} \in \mathbb{R}^N$ at a number of samples $\boldsymbol{\Xi} = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_N)$, we can construct the conditional process $Y \mid \boldsymbol{y}_{\boldsymbol{\Xi}}$. Thanks to Gaussian assumptions, the mean of this process at an unoberserved location $\boldsymbol{\xi}$ can be formulated as a linear combination of the observed responses:

$$\hat{y} = \boldsymbol{\lambda}(\boldsymbol{\xi})^T \boldsymbol{y}_{\boldsymbol{\Xi}}. \tag{3}$$

In particular by minimizing mean-squared error subject to unbiasedness constraints, the predictor at any unsampled site is given by

$$\hat{y} = \rho + \boldsymbol{r}^T \boldsymbol{R}^{-1}(\boldsymbol{y}_{\boldsymbol{\Xi}} - \rho \boldsymbol{F}), \tag{4}$$

$$\rho = (\boldsymbol{F}^T \boldsymbol{R}^{-1} \boldsymbol{F})^{-1}(\boldsymbol{F}^T \boldsymbol{R}^{-1} \boldsymbol{y}_{\boldsymbol{\Xi}}). \tag{5}$$

Here $\boldsymbol{F} := \mathbf{1} \in \mathbb{R}^N$, the covariance matrix $\boldsymbol{R} := R(\boldsymbol{\Xi}, \boldsymbol{\Xi}) \in \mathbb{R}^{N \times N}$, and finally $\boldsymbol{r} := R(\boldsymbol{\Xi}, \boldsymbol{\xi}) \in \mathbb{R}^N$ is the vector consisting of the covariance of the unobserved sample with respect to all observed sample sites. Using this cheap surrogate, $\mathbb{E}_\xi[y]$ can be evaluated with e.g. Monte-Carlo to any desired accuracy.

2.2 Additive bridge function based multi-level kriging model (AMLK)

If y results from the solution of a PDE, then by varying grid resolution we can have a sequence of numerical approximations to y, denoted $y_0, \ldots, y_L$, of increasing accuracy and increasing computational cost. The heart of the additive bridge function based multi-level Kriging model (AMLK) [19, 20, 21], is then to first write y_L as the telescopic sum

$$y_L = \sum_{l=0}^{L} \delta_l, \qquad \delta_0 := y_0, \quad \delta_l := y_l - y_{l-1},\ l \in \{1, \ldots, L\}, \tag{6}$$

similarly to the MLMC method; and then approximate each δ_l with a single-level Kriging surrogate $\hat{\delta}_l$. An estimate of the expectation of the QoI can then be written:

$$\mathbb{E}_\xi[y] \simeq \mathbb{E}_\xi[\hat{y}_L] := \mathbb{E}_\xi\left[\sum_{l=0}^{L} \hat{\delta}_l\right] = \sum_{l=0}^{L} \mathbb{E}_\xi[\hat{\delta}_l], \tag{7}$$

where the expectations are then evaluated on the surrogate, independently of each other. This decomposition is worthwhile because on the finest level L, the cost of the simulation is high, but the absolute magnitude of δ_L is small, so surrogate modelling errors $(\delta_L - \hat{\delta}_L)$ contribute little to the total error in $\mathbb{E}_\xi[y]$, and therefore sufficient accuracy can be achieved with few samples. In contrast, on the coarsest level many samples are needed to reduce the surrogate modelling error there, but these samples are very cheap to obtain. Potentially then, the total cost of estimating $\mathbb{E}[y]$ at a given accuracy can be reduced compared to the single-level method. Whether or not it is, in fact, reduced is investigated in the next section.

3 CONVERGENCE ANALYSIS OF AMLK

The error of using any surrogate model to estimate $\mathbb{E}[y]$ can be bounded by the discretization error and the surrogate interpolation error separately:

$$|\mathbb{E}[y - \hat{y}_L]| \leq \underbrace{|\mathbb{E}[y - y_L]|}_{\varepsilon_{\Delta x}} + \underbrace{|\mathbb{E}[y_L - \hat{y}_L]|}_{\varepsilon_h}, \tag{8}$$

where the former is a function of grid resolution $\Delta x > 0$, and the latter depends on some sampling parameter denoted h to be specified later.

In the multi-level case, assume that the grid resolution on level l is Δx_l, and further that there exist constants $\alpha, C_d > 0$ (independent of Δx), such that for all fidelity levels $l \in \{0, \ldots, L\}$:

$$\varepsilon_{\Delta x} := |\mathbb{E}[y - y_l]| \leq C_d \Delta x_l^\alpha, \tag{9}$$

i.e. the discrete approximation of $\mathbb{E}[y]$ converges at a fixed rate, where $\alpha = 2$ implies a 2nd-order accurate discretization, etc.

Consider now the interpolation error ε_h. By analogy with radial-basis function (RBF) interpolation [22], for a single-fidelity Gaussian process, the point-wise interpolation error in the process-mean can be expressed in terms of a *fill distance* h, defined as

$$h = h_{\boldsymbol{\Xi}} := \sup_{\boldsymbol{\xi}\in\Omega} \left\{ \min_{\xi_i\in\boldsymbol{\Xi}} \|\boldsymbol{\xi} - \boldsymbol{\xi}_i\| \right\}.$$

Here Ω is the interpolation domain, and $\boldsymbol{\Xi}$ is the sample sites. Then h is the radius of the largest (hyper-)sphere, whose center is contained in Ω, and which contains none of the samples. Given approximation of an infinitely differentiable function, the convergence order is dictated by the continuity of the covariance kernel (or radial basis function in RBF interpolation). For example, when the so-called thin-plate spline $R(r) := (-1)^{k+1} r^{2k} \log r$, with $k \in \mathbb{N}$, $r = |\boldsymbol{\xi} - \boldsymbol{\xi}'| \in \mathbb{R}$, is used, the ℓ^∞-norm of the interpolation error will satisfy $\epsilon \sim O(h^{k+1})$ [22]. For infinitely differentiable covariance kernels, convergence in this norm will be spectral. In this article, all derivations and numerical tests are based on the thin-plate spline with $k = 1$, though the results can be extended to other correlation functions straightforwardly.

The interpolation error of an additive bridge based multi-level model can therefore be written

$$\begin{aligned} \varepsilon_h = |\mathbb{E}[y_L - \hat{y}_L]| &\leq \left| \mathbb{E}\left[\sum_{l=0}^{L} \delta_l - \sum_{l=0}^{L} \hat{\delta}_l \right] \right| \\ &\leq \sum_{l=0}^{L} |\mathbb{E}[\delta_l - \hat{\delta}_l]| \\ &\leq \sum_{l=0}^{L} C_I \Delta x_l^\mu h_l^\beta. \end{aligned} \tag{10}$$

where h_l is the fill distance on level l, and β is a constant depending on the correlation function used in the Gaussian process (and the smoothness of the underlying function). Here, h_l^β comes from the approximation properties of the surrogate, and Δx_l^μ describes the magnitude of the interpolation error, which is proportional to size of the function being interpolated.

To limit the total error $|\mathbb{E}[y - \hat{y}_L]|$ to less than ε, we bound both the discretization error and interpolation error by $\varepsilon/2$. First, we choose the finest level L large enough to satisfy $\varepsilon_{\Delta x} = C_d \Delta x_L^\alpha \leq \varepsilon/2$. For simplicity, we assume that $\Delta x_l = \eta^{-l} \Delta x_0$, i.e. that grid resolution is increased by a constant factor η on each level. By arbitrarily normalizing Δx_0 to 1, the discretization error constraint becomes

$$C_d \eta^{-L\alpha} \leq \varepsilon/2 \implies L = \left\lceil \frac{1}{\alpha} \log_\eta\left(\frac{2C_d}{\varepsilon}\right) \right\rceil. \tag{11}$$

Similarly, to limit the interpolation error to $\varepsilon/2$, the infill distance of the surrogate must satisfy

$$h_l^\beta \leq \frac{C_d \Delta x_L^\alpha \Delta x_l^{-\mu}}{(1+L)C_I} = \frac{C_d \eta^{l\mu - L\alpha}}{(1+L)C_I}. \tag{12}$$

Given which, the total error is bounded as

$$|\mathbb{E}[y - \hat{y}_L]| \leq 2C_d \Delta x_L^\alpha. \tag{13}$$

3.1 Cost analysis for AMLK

Having found bounds on h_l in (12), it remains to specify the optimal number of samples per level N_l. Let the cost for a single sample of y_l be T_l. Then the total computational cost is

$$T = \sum_{l=0}^{L} N_l T_l. \tag{14}$$

To choose N_l optimally, we minimize T subject to the constraint on error. Treating N_l as a continuous variables, we solve the optimization problem:

$$\min_{N_l \in \mathbb{R}^+} T, \quad \text{subject to} \quad \sum_{l=0}^{L} C_I \Delta x_l^{\mu} h_l^{\beta} = \varepsilon/2. \tag{15}$$

Further assume that there exist constants $C_c, \gamma \in \mathbb{R}$ (independent of Δx_l), such that the cost of a evaluation is

$$T_l = C_c \Delta x_l^{\gamma}, \tag{16}$$

which is approximately true for typical PDEs solvers. Finally, as it is usually difficult to estimate h (especially in high-dimensional spaces), we choose to treat the estimated error as a function of N (to which we have direct access). So, for a specific sampling method, we assume there exist constants C_s and ν, such that

$$h_l = C_s N_l^{\nu}. \tag{17}$$

Note that ν will vary with the stochastic dimension M, and depends also on the sampling method. For tensor-product samples $\nu = -\frac{1}{M}$ can be seen immediately, i.e. the curse of dimensionality. In terms of N, the convergence rate of the surrogate model method deteriorates with an increasing of number of input variables. In terms of h, it is dimension-independent.

With assumptions (16) and (17), and formulating the constrained minimization problem (15) in terms of a Lagrange multiplier λ, we obtain the equivalent problem, find N_l, λ such that:

$$\frac{\partial f}{\partial N_l} = 0, \quad \frac{\partial f}{\partial \lambda} = 0,$$

where

$$f(N_l, \lambda) = \sum_{l=0}^{L} N_l C_c \Delta x_l^{\gamma} + \lambda \left(\sum_{l=0}^{L} C_I \Delta x_l^{\mu} (C_s N_l^{\nu})^{\beta} - \varepsilon/2 \right). \tag{18}$$

Explicitly

$$\frac{\partial f}{\partial N_l} = C_c \Delta x_l^{\gamma} + \lambda C_I \Delta x_l^{\mu} C_s^{\beta} \nu \beta N_l^{\nu\beta - 1} = 0, \tag{19}$$

$$\frac{\partial f}{\partial \lambda} = \sum_{l=0}^{L} C_I \Delta x_l^{\mu} (C_s N_l^{\nu})^{\beta} - \varepsilon/2 = 0, \tag{20}$$

whereupon solving for N_l gives

$$N_l = \left\lceil \left(\frac{\varepsilon}{2 C_I C_s^{\beta} S(L)} \right)^{\frac{1}{\nu\beta}} (\Delta x_l)^{\frac{\gamma - \mu}{\nu\beta - 1}} \right\rceil, \quad S(L) = \sum_{l=0}^{L} \Delta x_l^{\frac{\mu - \gamma\nu\beta}{1 - \nu\beta}}. \tag{21}$$

With N_l determined, we now examine the complexity of the multilevel approximation:

$$\begin{aligned} T_\varepsilon &= \sum_{l=0}^{L} N_l T_l \\ &\leq \sum_{l=0}^{L} \left[\left(\frac{\varepsilon}{2C_I C_s^\beta S(L)} \right)^{\frac{1}{\nu\beta}} (\Delta x_l)^{\frac{\mu-\gamma}{1-\nu\beta}} + 1 \right] C_c \Delta x_l^\gamma \\ &= \sum_{l=0}^{L} \left(\frac{\varepsilon}{2C_I C_s^\beta S(L)} \right)^{\frac{1}{\nu\beta}} (\Delta x_l)^{\frac{\mu-\gamma}{1-\nu\beta}} C_c \Delta x_l^\gamma + \sum_{l=0}^{L} C_c \Delta x_l^\gamma \\ &= \sum_{l=0}^{L} \left(\frac{\varepsilon}{2C_I C_s^\beta S(L)} \right)^{\frac{1}{\nu\beta}} C_c (\Delta x_l)^{\frac{\mu-\gamma\nu\beta}{1-\nu\beta}} + \sum_{l=0}^{L} C_c \Delta x_l^\gamma \\ &= \left(\frac{\varepsilon}{2C_I C_s^\beta} \right)^{\frac{1}{\nu\beta}} C_c S(L)^{1-\frac{1}{\nu\beta}} + \sum_{l=0}^{L} C_c \Delta x_l^\gamma \end{aligned} \tag{22}$$

The cost analysis of the AMLK model follows that of multi-level stochastic collocation method in [18]. First, let's bound the second term on the right side. Recall (11), bounding the finest level L by $\frac{1}{\alpha}\log_\eta(\frac{2C_d}{\varepsilon}) + 1$, where η is the scaling parameter of Δx_l ($\Delta x_l = \eta^{-l}$), we have

$$\begin{aligned} \sum_{l=0}^{L} C_c \Delta x_l^\gamma &\simeq \sum_{l=0}^{L} C_c \eta^{-l\gamma} \\ &= C_c \frac{\eta^{-\gamma L} - 1}{\eta^{-\gamma} - 1} \\ &= C_c \frac{\eta^{-\gamma(L-1)} - \eta^\gamma}{1 - \eta^\gamma} \\ &\leq \frac{C_c \eta^{-\gamma(\frac{1}{\alpha}\log_\eta(\frac{2C_d}{\varepsilon}))}}{1 - \eta^\gamma} \\ &= \frac{C_c (2C_d)^{\frac{-\gamma}{\alpha}}}{1 - \eta^\gamma} (\varepsilon)^{\frac{\gamma}{\alpha}}. \end{aligned} \tag{23}$$

Then, we provide a bound on the geommetric sum $S(L)$. When $\mu \neq \gamma\nu\beta$, we have

$$\begin{aligned} S(L) = \sum_{l=0}^{L} {\Delta x_l}^{\frac{\mu-\gamma\nu\beta}{1-\nu\beta}} &\simeq \sum_{l=0}^{L} \eta^{-l\frac{\mu-\gamma\nu\beta}{1-\nu\beta}} \\ &= \frac{\eta^{-L\frac{\mu-\gamma\nu\beta}{1-\nu\beta}} - 1}{\eta^{-\frac{\mu-\gamma\nu\beta}{1-\nu\beta}} - 1} \\ &= \frac{\eta^{-(L-1)\frac{\mu-\gamma\nu\beta}{1-\nu\beta}} - \eta^{\frac{\mu-\gamma\nu\beta}{1-\nu\beta}}}{1 - \eta^{\frac{\mu-\gamma\nu\beta}{1-\nu\beta}}} \\ &= \frac{\eta^{-\frac{\mu-\gamma\nu\beta}{1-\nu\beta}(\frac{1}{\alpha}\log_\eta(\frac{2C_d}{\varepsilon}))} - \eta^{\frac{\mu-\gamma\nu\beta}{1-\nu\beta}}}{1 - \eta^{\frac{\mu-\gamma\nu\beta}{1-\nu\beta}}} \\ &= \frac{(2C_d)^{-\frac{\mu-\gamma\nu\beta}{\alpha(1-\nu\beta)}} (\varepsilon)^{\frac{\mu-\gamma\nu\beta}{\alpha(1-\nu\beta)}}}{1 - \eta^{\frac{\mu-\gamma\nu\beta}{1-\nu\beta}}} - \frac{\eta^{\frac{\mu-\gamma\nu\beta}{1-\nu\beta}}}{1 - \eta^{\frac{\mu-\gamma\nu\beta}{1-\nu\beta}}}; \end{aligned} \tag{24}$$

when $\mu = \gamma\nu\beta$, we have

$$S(L) = L + 1 = \frac{1}{\alpha}\log_\eta(\frac{2C_d}{\varepsilon}) + 2. \tag{25}$$

Finally, when $\mu \neq \gamma\nu\beta$, the computation cost versus ε can be bounded by

$$\begin{aligned} T_\varepsilon &\leq \left(\frac{\varepsilon}{2C_I C_s^\beta}\right)^{\frac{1}{\nu\beta}} C_c S(L)^{1-\frac{1}{\nu\beta}} + \sum_{l=0}^{L} C_c \Delta x_l^\gamma \\ &\leq \varepsilon^{\frac{1}{\nu\beta}} \varepsilon^{\frac{\mu-\gamma\nu\beta}{\alpha(1-\nu\beta)}(1-\frac{1}{\nu\beta})} + \varepsilon^{\frac{1}{\nu\beta}} + \varepsilon^{\frac{\gamma}{\alpha}} \\ &= \varepsilon^{\frac{1}{\nu\beta}+\frac{\gamma\nu\beta-\mu}{\alpha\nu\beta}} + \varepsilon^{\frac{1}{\nu\beta}} + \varepsilon^{\frac{\gamma}{\alpha}}; \end{aligned} \tag{26}$$

when $\mu = \gamma\nu\beta$,

$$T_\varepsilon \leq \varepsilon^{\frac{1}{\nu\beta}} |\log_\eta(\varepsilon)|^{(1-\frac{1}{\nu\beta})} + \varepsilon^{\frac{1}{\nu\beta}} + \varepsilon^{\frac{\gamma}{\alpha}}. \tag{27}$$

Consequently, we have

$$T_\varepsilon \leq \begin{cases} \varepsilon^{\frac{1}{\nu\beta}} & \text{if} \quad \mu > \gamma\nu\beta, \\ \varepsilon^{\frac{1}{\nu\beta}} |\log_\eta(\varepsilon)|^{(1-\frac{1}{\nu\beta})} & \text{if} \quad \mu = \gamma\nu\beta, \\ \varepsilon^{\frac{1}{\nu\beta}+\frac{\gamma\nu\beta-\mu}{\alpha\nu\beta}} & \text{if} \quad \mu < \gamma\nu\beta. \end{cases} \tag{28}$$

Usually, in terms of Δx, the size of the difference between two consecutive level has the same convergence rate with the discretization error, thus $\mu = \alpha$, which is also showed in the following numerical test cases. When $\mu < \gamma\nu\beta$, we have

$$\begin{aligned} T_\varepsilon &\leq \varepsilon^{\frac{1}{\nu\beta}+\frac{\gamma\nu\beta-\mu}{\alpha\nu\beta}} \\ &= \varepsilon^{\frac{1}{\nu\beta}(1-\frac{\mu}{\alpha})+\frac{\gamma}{\alpha}} \\ &\leq \varepsilon^{\frac{\gamma}{\alpha}}. \end{aligned} \tag{29}$$

3.2 Parameter estimation and practical details

A practical algorithm is given below. In the first step, a grid convergence study is needed to provide an estimation for discretization error and determine the finest level L. The second step is to estimate the constants assumed in the model of interpolation error.

Recalling (10), the interpolation error of AMLK model is assumed to be $C_I \Delta x_l^\mu h_l^\beta$. Again, h_l^β represents the convergence properties of the surrogate model, and Δx_l^μ comes from the logic that the magnitude of the interpolation error should be proportional to the size of the function being interpolated, which is the difference between two consecutive levels. Here, β and μ are both constants - which are assumed to be independent of level, and can be estimated separately. The assumption of interpolatin error should ideally be numerically verified.

As already mentioned, it is difficult to estimate h in high-dimensional spaces, so we assume that $h_l = C_s N_l^\nu$, so that

$$C_I \Delta x_l^\mu h_l^\beta = C_I \Delta x_l^\mu (C_s N_l^\nu)^\beta = C_I C_s^\beta \Delta x_l^\mu N_l^{\nu\beta},$$

and instead of estimating β directly, we treat the error as a function of N and estimate $\nu\beta$ together, as well as estimating the combined constant $C_I C_s^\beta$ together. Remember that ν varies

with the stochastic dimension M and also depends on the sampling method, so must be re-estimated for each new problem.

To estimate these parameters, firstly, by quantifying the interpolation error for the cheapest three levels using standard kriging model with a fixed, small number of sample points, an estimation for μ can be obtained. At the same time, the computational cost per sample on each level is collected, which gives us an estimation of γ. Then, using the coarsest level only, $\nu\beta$ and $C_I C_s^\beta$ can be estimated by varying the number of sampling points. With these estimated constants, the optimal number on each level is determined. Through the algorithm below, UQ with the multi-level kriging model can be conducted.

Algorithm: AMLK to estimate $\mathbb{E}[y]$ with error $< \epsilon$

Determine the finest level L from (11) interpolation error is $\varepsilon/2$.
Use the cheapest three levels $l = 0, 1, 2$ to estimate $C_I C_s^\beta$ and $\gamma, \nu\beta, \mu$.
for $l = 0 : L$ **do**
  Calculate the optimal number of samples N_l using (21);
  Generate N_l sample points $\boldsymbol{\Xi}_l$, with e.g. Latin hypercube sampling;
  Evaluate $\boldsymbol{y}_l(\boldsymbol{\Xi}_l)$ and $\boldsymbol{y}_{l-1}(\boldsymbol{\Xi}_l)$ with the PDE solver (note $y_{-1} \equiv 0$);
  Evaluate $\boldsymbol{\delta}_l(\boldsymbol{\Xi}_l) := \boldsymbol{y}_l(\boldsymbol{\Xi}_l) - \boldsymbol{y}_{l-1}(\boldsymbol{\Xi}_l)$;
  Construct a kriging model for δ_l;
  Evaluate $\mathbb{E}[\hat{\delta}_l]$ on surrogate model response surface using e.g. Monte Carlo;
end
Evaluate result $\mathbb{E}[\hat{y}_L] = \sum_{l=0}^{L} \mathbb{E}[\hat{\delta}_l]$;

4 TEST CASES

4.1 Oscillatory Genz function (M=2)

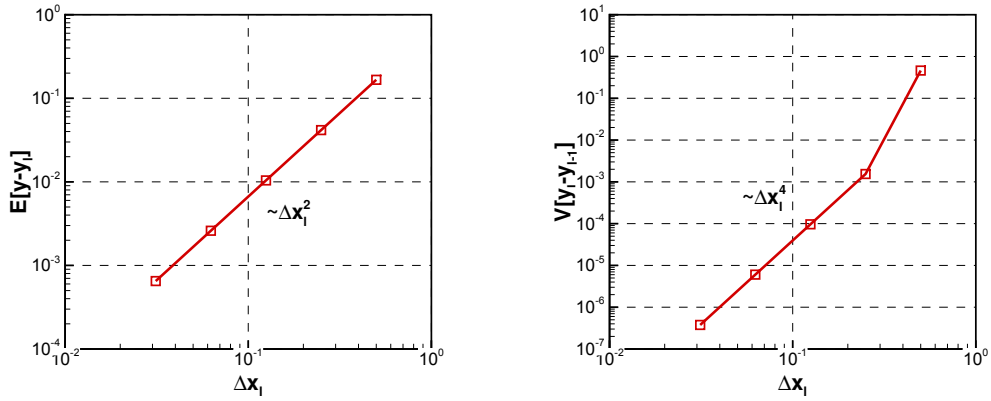
To quickly verify basic properties of AMLK, we consider an almost-trivial analytic test-case based on the osciallatory Genz function in 2d [24]. The basic function is

$$y(\boldsymbol{\xi}) := \cos\left(\pi + 5\xi_1 + 5\xi_2\right), \tag{30}$$

where $\xi_1, \xi_2 \sim \mathcal{U}(0,1)$ i.i.d. To simulate multi-level analyses we introduce an artificial mesh-dependent term:

$$y_l(\boldsymbol{\xi}) := y(\boldsymbol{\xi}) + \sin(|\boldsymbol{\xi}|)\Delta x_l^2, \tag{31}$$

where $\Delta x_l = 2^{-l}\Delta x_0$, $\Delta x_0 = \frac{1}{2}$, and $L = 4$. Note that $\delta_l(\boldsymbol{\xi})$ has the same form on every level, with only a scale difference. Two hundred random uncertainty samples are used to estimate the discretization error, shown in Figure 1a, which shows quadratic convergence as we expected, $\varepsilon_{\Delta x} = |\mathbb{E}[y - y_l]| \leq C_d \Delta x_l^\alpha \sim \Delta x_l^2$. To make a comparison with MLMC method, the variance for the difference function is also estimated, and the rate of convergence is 4, double the rate of convergence of the expectation, as expected in this case, shown in Figure 1b.



(a) Discretization error of multi-level analyses (b) Variance of difference function

Figure 1: Performance plots of multi-level analyses for Genz case

4.2 Darcy flow with random coefficients (M=21)

The first PDE-based test-case is Darcy on $D = (0,1)^d, d = 2$, with both Dirichlet and Neumann boundary conditions [17]:

$$-\nabla \cdot (k(\mathbf{x},\omega)\nabla p(\mathbf{x},\omega)) = 1, \quad \mathbf{x} \in D, \tag{32}$$

$$p|_{x_1=0} = 1, \quad p|_{x_1=1} = 0, \tag{33}$$

$$\frac{\partial p}{\partial \boldsymbol{n}}|_{x_2=0} = 0, \quad \frac{\partial p}{\partial \boldsymbol{n}}|_{x_2=1} = 0, \tag{34}$$

where k is a scalar-valued random field with

$$\log k(\mathbf{x},\omega) = Z(\mathbf{x},\omega) = \mathbb{E}[Z(\mathbf{x},\cdot)] + \sum_{n=0}^{\infty} \sqrt{\theta_n}\xi_n(\omega)b_n(\mathbf{x}), \tag{35}$$

where the Karhunen-Loeve expansion orginates from the covariance function

$$C(\mathbf{x},\mathbf{y}) := \sigma^2 \exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|_p}{\lambda}\right), \quad \lambda = 0.3, \quad \sigma^2 = 1, \quad p = 1. \tag{36}$$

where $\{\theta_n\}_{n\in\mathbb{N}}$ and $\{b_n\}_{n\in\mathbb{N}}$ are the eigenvalues and normalised eigenvectors of the covariance matrix. The uncertain variables $\{\xi_n\}_{n\in\mathbb{N}}$ are a sequence of independent, uniform random variables on $[-1,1]$. In this problem, We use 21 terms in the K-L expansion ($M = 21$), which includes 84% of the total spectral energy.

This PDE is solved with finite-volumes on a uniform Cartesian grid of $m \times m$ cells. Let $k_{i,j}$ and $p_{i,j}$ denote the values of k, and p at a cell centre $\mathbf{x}_{i,j}, (i,j = 1,\ldots,m)$. The discretization used is

$$-\bar{k}_{i,j-\frac{1}{2}}p_{i,j-1} - \bar{k}_{i-\frac{1}{2},j}p_{i-1,j} - \bar{k}_{i+\frac{1}{2},j}p_{i+1,j} - \bar{k}_{i,j+\frac{1}{2}}p_{i,j+1} + 4p_{i,j}\hat{k}_{i,j} = 0 \tag{37}$$

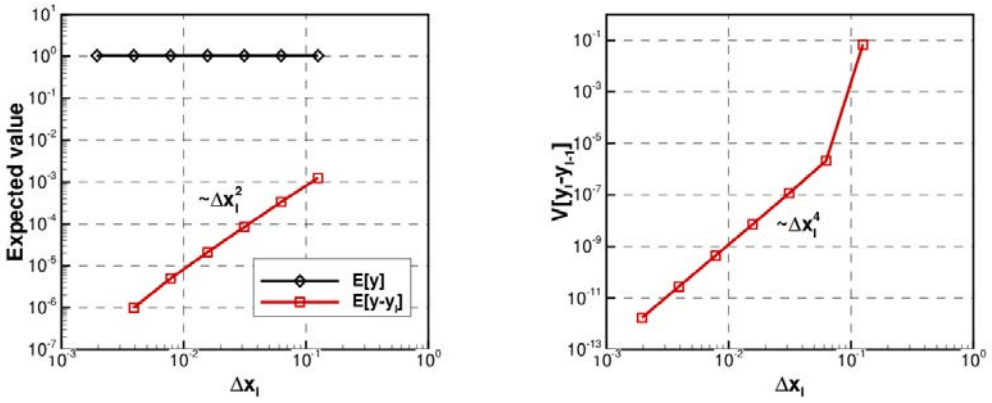
where $\hat{k}_{i,j} = (\bar{k}_{i,j-\frac{1}{2}} + \bar{k}_{i-\frac{1}{2},j} + \bar{k}_{i+\frac{1}{2},j} + \bar{k}_{i,j+\frac{1}{2}})/4$. Here $\bar{k}_{i,j+\frac{1}{2}}$ is the value at the mid-point of an edge, which is approximated by the arithmetic average of $k_{i,j+1}$ and $k_{i,j}$. At Dirichlet

boundaries, the derivative is approximated by a one-sided difference. At Neumann boundaries, the derivative is known explicitly, and k is approximated by $k_{i,j}$. The quantity of interest is

$$y := -\int_0^1 k\frac{\partial p}{\partial x_1}\bigg|_{x_1=1} \mathrm{d}x_2 \simeq \sum_{j=1}^{m} k\frac{\partial p}{\partial x_1}\bigg|_{m+\frac{1}{2},j}, \tag{38}$$

given all of which the discretization error is $O(\Delta x^2)$.

We choose a sequence of spatial grids with the cell size $\Delta x_l = 2^{-l}\Delta x_0$ and $\Delta x_0 = \frac{1}{8}$. Six levels are used, i.e. $L = 5$. The computational cost per sample is measured to be $C_l = C_c\Delta x_l^{\gamma} \propto \Delta x_l^{-2}$, thanks to an efficient multi-grid solver. Taking the grid with $\Delta x = 1/512$ as a reference, the discretization error is estimated using 200 samples per level and shown in Figure 2a. Clean 2nd-order convergence is observed, so $\alpha = 2$ in this case. The variance for the difference function is estimated as an 4th-order convergence rate in terms of Δx, shown in Figure 2b.



(a) Discretization error of multi-level analyses (b) Variance of difference function

Figure 2: Performance plots of multi-level analyses for Darcy flow case

4.3 RANS flow over an airfoil with geometric uncertainties (M=10)

Finally, a more challenging test-case is considered: Reynolds Averaged Naiver Stokes (RANS) for the RAE2822 airfoil at $M_\infty = 0.734$, $\alpha = 2.79°$, and $\mathrm{Re} = 6.5 \times 10^6$. Manufacturing variability of the aerofoil surface of approximately 0.2% of the chord length is modeled by a Gaussian random fields with the correlation

$$C(\mathbf{s_1}, \mathbf{s_2}) := \sigma^2 \exp\left(-\frac{\|\mathbf{s_1} - \mathbf{s_2}\|_2}{\lambda}\right), \quad \lambda = 0.1, \tag{39}$$

where $\mathbf{s_1}$ and $\mathbf{s_1}$ are two surface nodes, distance is the standard Euclidean norm, and the standard deviation σ is assumed to be

$$\sigma = \begin{cases} (10 - 10x)^{0.7} \times 0.001 & \text{if} \quad 0.9 \le x \le 1.0, \\ 0.001 & \text{if} \quad 0.1 \le x \le 0.9, \\ (10x)^{0.7} \times 0.001 & \text{if} \quad 0 \le x \le 0.1. \end{cases} \tag{40}$$

Similar to Darcy, the Gaussian-process is parametrized by independent standard Gaussian random variables using Karhunen-Loeve:

$$G(\mathbf{s},\omega) = \sum_{k=0}^{\infty} \sqrt{\theta_k}\xi_k(\omega)b_k(\mathbf{s}). \tag{41}$$

The eigenvalues are shown in Figure 3, and the first 10 K-L modes are used to parametrize the perturbation. Figure 4 shows three realizations of the perturbation and corresponding pressure distribution of the perturbed airfoils. The CFD solver used is finite-volume and nominally

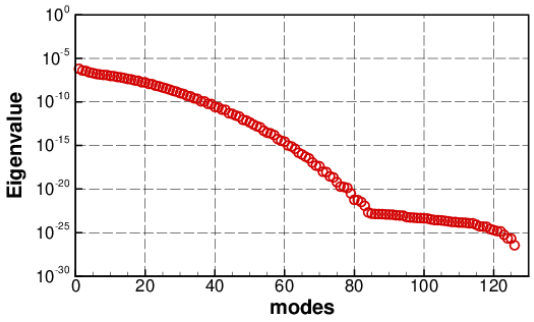
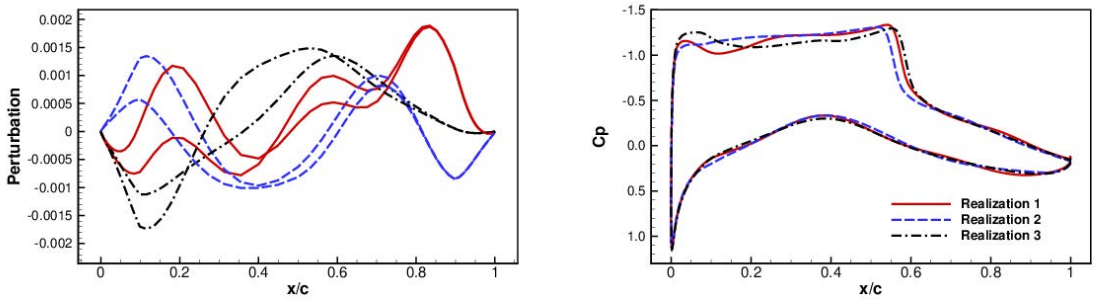


Figure 3: Eigenvalues of K-L expansion for geometric uncertainty



(a) Three realizations of K-L expansion

(b) Pressure distributions

Figure 4: Visualization of the geometric uncertainty in the airfoil surface and corresponding pressure distribution of perturbed airfoils

2nd-order accurate, though this case exhibits a shock reducing the ℓ^2-norm of the solution to 1st-order. The QoI used in this case the lift coefficient, and is observed to converge at slightly higher-order in practice.

In this final case, we define 5 grid levels, the parameters of which are given in Table 1. The parameter of the reference computational grid is also shown in the last line. Discretization error is estimated using 100 random samples, shown in Figure 6a. Note that in the Darcy case the discretization error was estimated with respect to grid-cell size, but number of cells is used in this case for simplicity. We have the discretization error $|\mathbb{E}[y - y_l]| \leq C_d K^{\alpha} \sim K^{-1.35}$. The variance for the difference function decreases at the rate of $K^{-3.5}$ in terms of K, shown in Figure 6b.

Table 1: Grid parameters for RANS flow case

Level	# cells	I-cells	J-cells	# surface cells
0	18432	192	96	128
1	41472	288	144	192
2	73728	384	192	256
3	112896	504	224	336
4	165888	576	288	384
Ref	209952	648	324	432

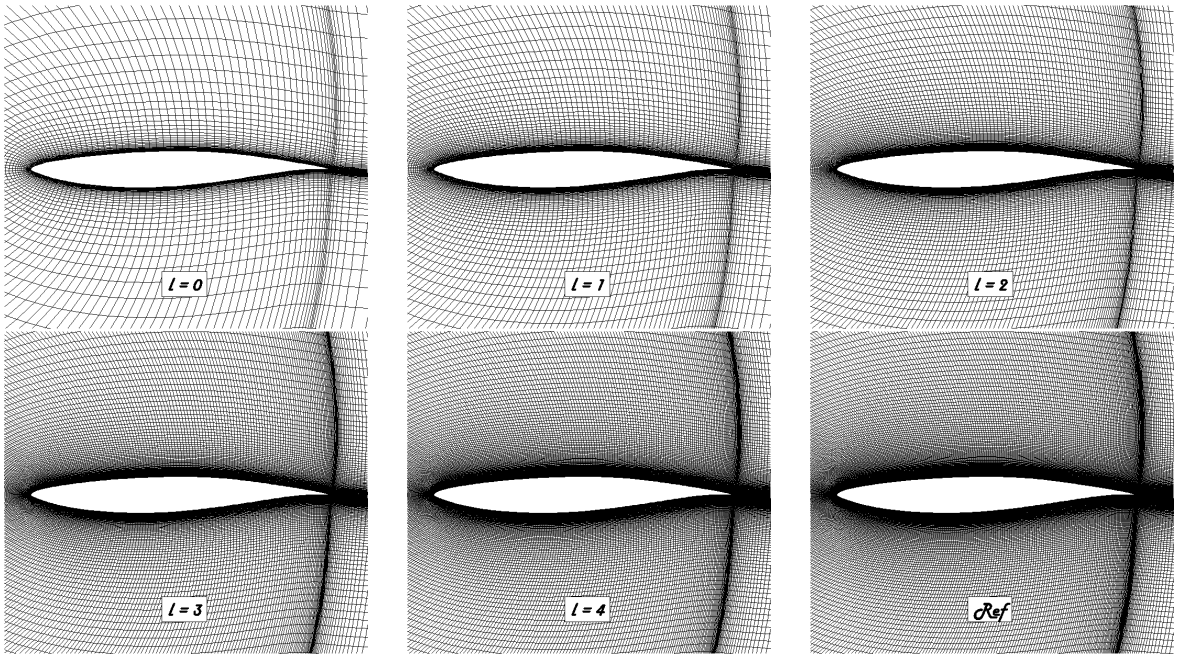
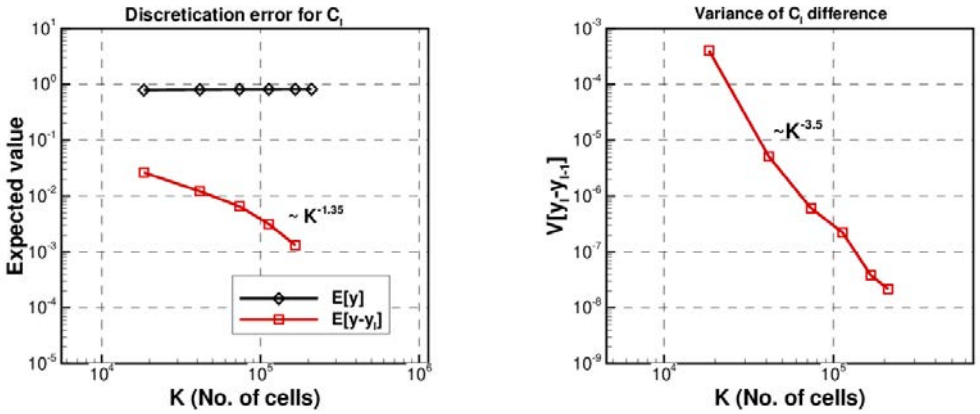


Figure 5: Multi-level computational grids for RANS flow case



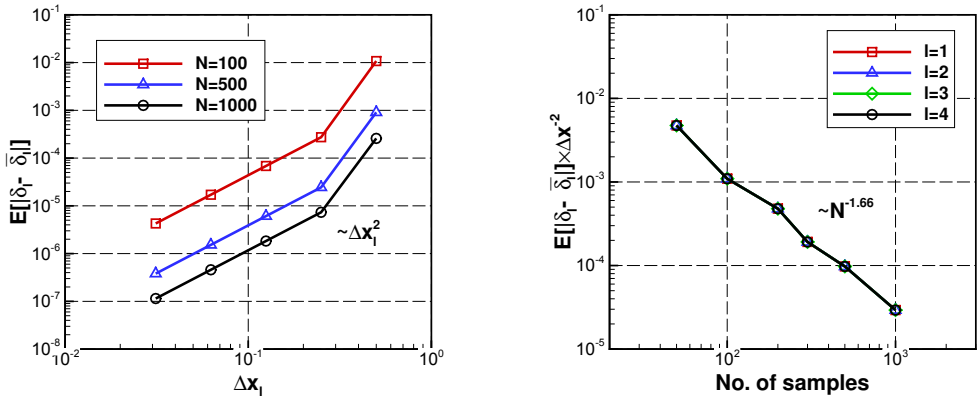
(a) Discretization error of multi-level analyses (b) Variance of difference function

Figure 6: Performance plots of multi-level analyses for RANS flow case

5 RESULTS

5.1 2D "Oscillatory" Genz funcion

First, the unknown constants in the model of interpolation error are estimated. Halton sample sequence is used to generate uniformly-distributed random points in the parameter space. The estimated interpolation error is shown in Figure 7. The left figure shows us that the magnitude of the error decreases at a second order ($\mu = 2$), and in the right figure, it can be seen that the interpolation error from different levels collapses well to a single line when scaled by Δx^2. The parameters estimated on the basis of these plots are given in Table 2. These figures confirms our assumptions in the convergence properties of interpolation error. Following Algorithm 1:



(a) Interpolation error versus grid size (b) Interpolation error versus No. of samples

Figure 7: Approximation for interpolation error of AMLK for Genz case

Table 2: Estimated parameters for Genz case

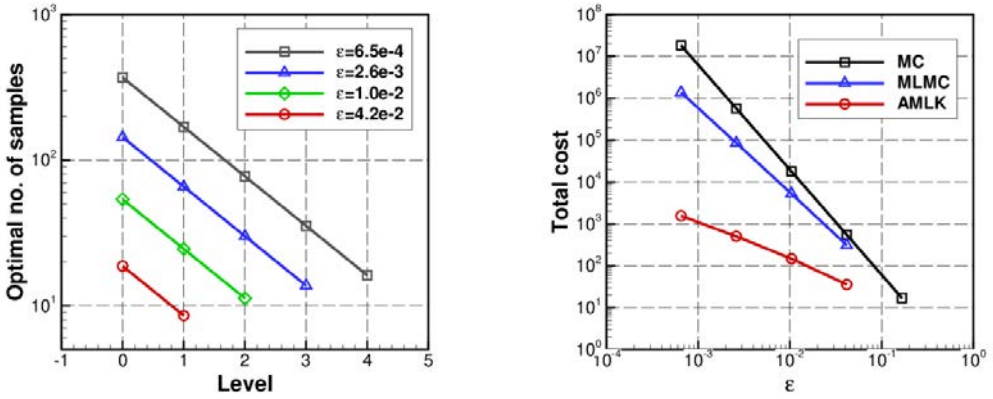
parameters	μ	$\nu\beta$	$C_I C_s^\beta$
Estimated value	2.0	-1.66	11.261

given accuracy requirement ε, L is determined by limiting the discretization error to $\varepsilon/2$. The computational cost per sample is assumed as $T_l = C_c \Delta x_l^\gamma \sim \Delta x_l^{-1}$. Then the optimal number of sampling points on each level are determined using (21), the result of which is shown in Figure 8a. For comparison, Monte-Carlo (MC) and Multilevel Monte-Carlo (MLMC) methods are also applied to this case. For MC the number of points required for a certain accuracy is simply $N = \sigma^2/\varepsilon^2$. For MLMC, the optimal points per level [17] is given by

$$N_l = \varepsilon^{-2} \left(\sum_{l=0}^{L} \sqrt{V_l C_l}\right) \sqrt{\frac{V_l}{C_l}},$$

and a comparison of total cost against MLMC and MC for this problem is shown in Figure 8b. In this paper, standardized costs are presented always, which is scaled by the cost oper sample on the coarsest level. From literature [17], the total computational costs of MC and MLMC should be proportional to $\varepsilon^{-2+\gamma/\alpha}$ and ε^{-2}, if the variance $V[y_l - y_{l-1}]$ decays faster than the increase of T_l. We observe that the computational costs of MC and MLMC method grow along with

the improvement of accuracy at the rate of $\varepsilon^{-2.5}$ and ε^{-2}, respectively, which agrees with the theoretical result. In this case, we find μ is larger than $\gamma\nu\beta$, so that the limit of the convergence of cost versus ε should be $\varepsilon^{-1/1.66}$, according to (28). In Figure 8b, the cost of AMLK model increases as $\varepsilon^{-1/1.1}$, which is much slower than other methods, but faster than the theoretical value. The reason is that the lower-order error terms in (26) is also influential. With the increase of the required accuracy, more benefit can be gained by AMLK model.



(a) Optimal No. of points of AMLK (b) Computational cost versus error

Figure 8: Performance plots of AMLK for Genz case

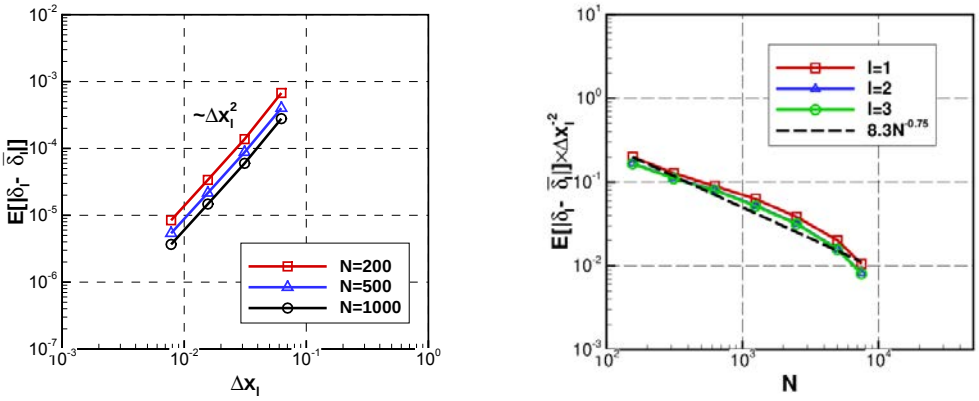
5.2 Darcy flow with random coefficients

Halton sequences are also used in this case. The estimated interpolation error is shown in Figure 9. These results confirms our assumptions again, but with different estimated parameters, shown in Table 3. Based on the estimated parameters, the optimal number of sampling points on each level are shown in Figure 10a, and the total cost are shown in Figure 10b, as well as that of MLMC and MC. The results show that the total costs of MC and MLMC achieve an error of $O(\varepsilon)$ is ε^{-3} and ε^{-2}. Same as the first case, Figure 10b also indicates the cost of AMLK method grows as $\varepsilon^{-1/0.61}$, which is a bit faster than theoretical value $\varepsilon^{-1/0.75}$, but is slower than MC and MLMC methods.

MC is seen to be completely impractical for very low ε but highly comparative for high errors (especially given its simplicity and use of a single grid). Despite MLMC achieving optimal rates, at low ε it is soundly beaten by AMLK. This result is not surprising, as similar performance was observed for polynomial surrogates in [18]. It does however rely on the regularity of the underlying response $y(\boldsymbol{\xi})$, which the MC-based techniques do not. It is therefore instructive to proceed to a case where regularity is not guaranteed, see next section.

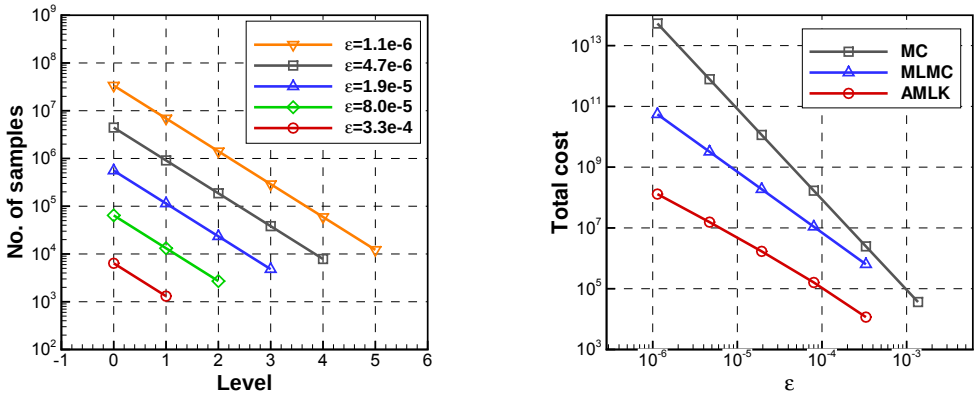
Table 3: Estimated parameters for the Darcy case

Parameter	μ	$\nu\beta$	$C_I C_s^\beta$
Estimated value	2.0	-0.75	8.3



(a) Interpolation error versus grid size (b) Interpolation error versus No. of samples

Figure 9: Approximation for interpolation error of AMLK for Darcy flow case



(a) Optimal No. of points of AMLK (b) Computational cost versus error

Figure 10: Performance plots of AMLK for Darcy flow case

5.3 RANS flow over an airfoil with geometry uncertainties

In this final case, the interpolation error of the AMLK is assumed to be $C_I K^\mu h_l^\beta$, again independent of level. Similarly to the first case, actually we estimate

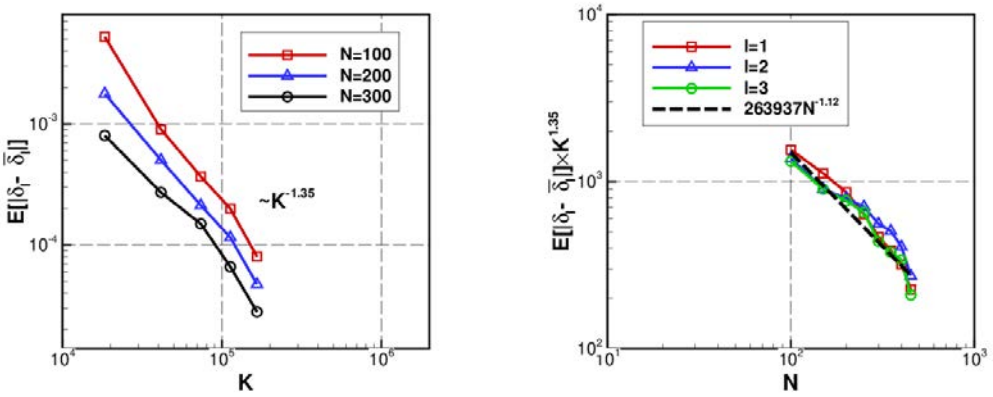
$$C_I M_l^\mu h_l^\beta = C_I K_l^\mu (C_s N_l^\nu)^\beta = C_I C_s^\beta K_l^\mu N_l^{\nu\beta}.$$

The normally-distributed random samples are obtained by transferring the Halton samples based on the probability integral property. The estimated interpolation error with respect to the number of grid-cells and sampling points are shown in Figure 11 and the estimated parameters is present in Table 4. Once more, the convergence of the interpolation is seen to be independent of the grid level (under appropriate scaling), justifying the choice of level-independent parameters in (10). The computational cost per sample on each level is estimated, shown in Figure 12, which shows that $T_l = C_c K_l^\gamma \sim K^{0.93}$.

From Figure 6a, we found that the discretization error on finest level is 0.00132, which indicates that the grid is sufficiently accurate to resolve the lift coefficient to around 0.1 count, which already meets the engineering requirement. However, the observed interpolation errors, whose magnitude ranges from $10^{-4} - 10^{-5}$, are much smaller than the discretization error. It is

Table 4: Estimated parameters for RANS flow case

parameters	μ	$\nu\beta$	$C_I C_s^\beta$
Estimated value	-1.35	-1.12	2.639×10^5



(a) Interpolation error versus grid size (b) Interpolation error versus No. of samples

Figure 11: Approximation for interpolation error of AMLK for RANS flow case

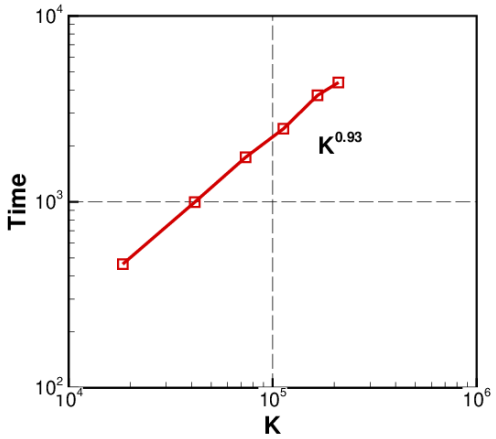
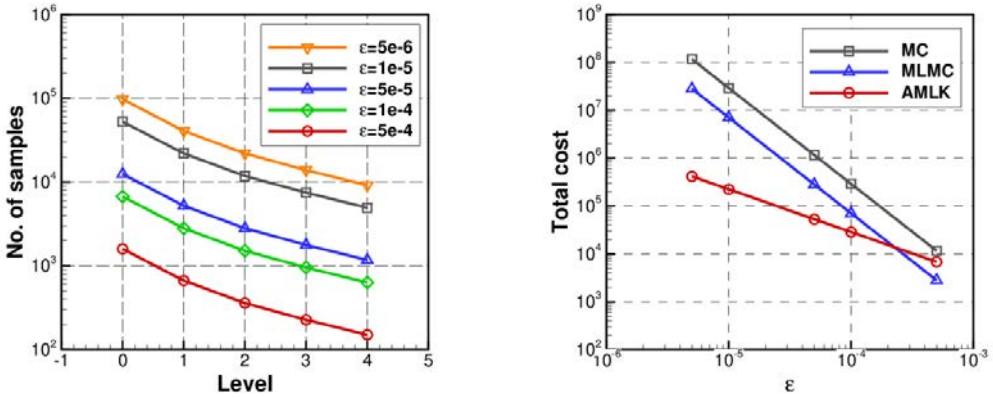


Figure 12: Computational time versus number of grid cells for RANS flow case

impractical to bound the discretization error and interpolation error equally. Therefore, in this case, we fix the discretization error on finest level - 5 levels involved, and estimate the optimal computational cost required to achieving a certain interpolation accuracy.

Based on the estimated parameters, the optimal number of sampling points on each level are shown in Figure 13a, and the total cost are shown in Figure 13b. Both the total costs of MC and MLMC grows at the rate of ε^{-2}, which is consistent with the theoretical result when the involved multi-level analyses are fixed. Meanwhile, as the multi-level analyses are fixed, the cost of AMLK method grows as $\varepsilon^{-1/1.12}$ exactly. In this practical application case, the same convergence property of AMLK model is observed, which can save much computational cost than MC and MLMC methods.



(a) No. of sampling points of AMLK (b) Computational cost versus error

Figure 13: Performance plots of AMLK for RANS case

6 CONCLUSIONS

In this work, we performed theoretical convergence and cost analyses on the AMLK model, utilizing convergence results for single-fidelity Kriging, as well as existing multi-level analysis of stochastic collocation method. Three numerical test cases with different number of uncertainty variables were utilized to demonstrate the effectiveness of proposed method. All the numerical results verified the assumptions for the mathematical form of discretization error and interpolation error. The comparisons of total computational cost showed that using multi-level kriging model for UQ can significantly reduce the cost, compared with the MLMC and standard MC method.

In this study, only the thin-plate spline was considered as the covariance kernel used in kriging model. As we mentioned before, the convergence property of a kriging model is only dependent on the smoothness of the covariance kernel. For finitely differentiable kernels, the convergence and cost analyses results are analogous to this study. However, for infinitely differentiable kernels, as it shows spectral convergence, the form of interpolation error can be assumed as $C_I \Delta x_l^{\mu} e^{-\tilde{c}/h_l}$ and the convergence study could be conducted accordingly.

On the other hand, the theoretical convergence rate of kriging model was given with respect to the fill distance h. However, it is very difficult to estimate the h, especially for high-dimensional stochastic space. We defined $h = C_s N^{\nu}$ for a specific sampling method and transferred the interpolation error in terms of N instead. In this way, the sampling method is also essential for the convergence study of a surrogate model. In this study, we used the Halton pseudo-random samples for convergence study, which are deterministic, of low discrepancy but appear randomly. To estimate the interpolation error, we generated a series of sample data set with increasing number of points and ensured that a smaller-size data set is always a subset of a larger-size data set, such that the convergence study of interpolation error is consistent and smooth. Nevertheless, it was still difficult to obtain a smooth estimation for the interpolation error in $\mathbb{E}[y]$. In fact, the error in $\mathbb{E}[y]$ is not equivalent to any norm of the point-wise error. We used the mean of the ℓ^1-norm of the point-wise error to bound the error in $\mathbb{E}[y]$. In Figure 7, 9 and 11, we can find that smooth estimations for interpolation error are obtained.

For the RANS flow case with geometric uncertainty, we chose the lift coefficient as quantity of interest and gained good estimation of discretization error. As for the drag coefficient, because of the existences of a strong shock, it was difficult to get a linear convergence of dis-

cretization error. One of the largest difficulties we met in this case is that the magnitude of the disretization error was much larger than that of the interpolation error. Thus, we could not bound the two error terms equally, as we did in other two cases. However, even the accuracy on the finest level meets the requirement of engineering application, so it is not necessary to further improve the resolution of computation grid. Therefore, we fixed the finest level, and estimated the minimal total computational cost needed in order to achieve a certain interpolation error.

Besides the additive-bridge function based multi-level kriging model, there are potential models, such as multi-level hierarchical kriging model and cokriging. In future work, the convergence analysis of these two models will be studied comparatively.

REFERENCES

[1] R. T. Haftka, Combining global and local approximations. *AIAA Journal*, **29**(9), 1523–1525, 1991.

[2] K. J. Chang, R. T. Haftka, G. L. Giles, P. J. Kao, Sensitivity-based scaling for approximation structural response. *Journal of aircraft*, **30**(2), 283–288, 1993.

[3] S. E. Gano, J. E. Renaud, B. Sanders, Hybrid variable fidelity optimization by using a kriging-based scaling function. *AIAA Journal*, **43**(11), 2422–2430, 2005.

[4] Z. H. Han, S. Göertz, R. Zimmermann, Improving variable-fidelity surrogate modeling via gradient-enhanced kriging and a generalized hybrid bridge function. *Aerospace Science and Technology*, **25**, 177–189, 2013.

[5] A. G. Journel, J. C. Huijbregts, Mining geostatistics. *Academic Press*, New York, 1978.

[6] M. C. Kennedy, A. OHagan, Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, **87**(1), 1–13, 2010.

[7] Z. H. Han, R. Zimmermann, S. Göertz, An alternative cokriging model for variable-fidelity surrogate modeling. *AIAA Journal*, **50**(5), 1205–1210, 2012.

[8] Z. H. Han, S. Göertz, Hierarchical kriging model for variable-fidelity surrogate modeling. *AIAA Journal*, **50**(5), 1285–1296, 2012.

[9] P. S. Palar, T. Tsuchiya, G. T. Parks, Multi-fidelity non-intrusive polynomial chaos based on regression. *Computer Methods in Applied Mechanics and Engineering*, **305**, 579–606, 2016.

[10] L. Parussini, D. Venturi, P. Perdikaris, G. E. Karniadakis, Multi-fidelity Gaussian process regression for prediction of random fields. *Journal of Computational Physics*, **336**, 36–50, 2017.

[11] P. S. Palar, K. Shimoyama, Multi-Fidelity Uncertainty Analysis in CFD Using Hierarchical Kriging. *35th AIAA Applied Aerodynamics Conference*, Denver, Colorado, June 5-9, 2017.

[12] A. Narayan, C. Gittelson, D. Xiu, A stochastic collocation algorithm with multi-fidelity models. *SIAM Journal On Scientific Computing*, **36**(2), 495–521, 2014.

[13] X. Zhu, E. M.Linebarger, D. Xiu, Multi-fidelity stochastic collocation method for prediction of statistical moments. *Journal of Computational Physics*, **341**, 386–396, 2017.

[14] G. Geraci, M. S. Eldred, G. Iaccarino, A multifidelity multilevel Monte Carlo method for uncertainty propagation in aerospace applications. *19th AIAA Non-Deterministic Approaches Conference*, Grapevine, Texas, 2017.

[15] M. B. Giles, Multi-level Monte Carlo path simulation. *Operations Research*, **56**(3), 607–617, 2008.

[16] S. Heinrich, Multi-level Monte Carlo Methods. *Lecture Notes in Computer Science*, **2179**, 3624–3651, Springer, Berlin, Heidelberg, 2001.

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

[18] A. L. Teckentrup, P. Jantsch, C. G. Webster, 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.

[19] S. Choi, J. J. Alonso, I. M. Kroo, M. Wintzer, Multi-fidelity design optimization of low-boom supersonic business jets. *10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Albany, New York, 2004.

[20] P. Wang ; Y. Li ; C. Li An Optimization Framework Based on Kriging Method with Additive Bridge Function for Variable-Fidelity Problem. *14th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)*, 388–392, IEEE, 2015.

[21] C. Tang, K. Gee, S. Lawrence, Generation of Aerodynamic Data using a Design Of Experiment and Data Fusion Approach. *43rd AIAA Aerospace Sciences Meeting and Exhibit*, Reno, Nevada, 2005.

[22] H. Wendland, *Scattered data approximation*, **17**, Cambridge university press, 2004.

[23] D. Zimmerman, C. Pavlik, A. Ruggles, M. P. Armstrong, An experimental comparison of ordinary and universal kriging and inverse distance weighting. *Mathematical Geology*, **31**, 375–390, 1999.

[24] A. Genz, Testing multidimensional integration routines. *Proc. of international conference on Tools, methods and languages for scientific and engineering computation*, B. Ford, J. C. Rault, and F. Thomasset (Eds.). Elsevier North-Holland, Inc., New York, 81–94, 1984.