

6. Multifidelity uncertainty quantification (MIT, FSU, LANL)

Research thrusts: *Validation, adaptation and management of models; Optimization under uncertainty*

Research sub-thrusts: *Goal-driven model adaptivity; Multifidelity uncertainty quantification*

Evaluating uncertainty in predictions of interest is a critical element in supporting effective decision making in optimal control and design. For example, characterizing—with quantified uncertainty—the properties of a subsurface medium hosting a contaminant plume is essential to calculating the risk of future aquifer contamination, and thus managing the size and extent of costly remediation efforts. Monte Carlo methods are the most general way to estimate statistics of high-fidelity model outputs, but they typically require a large number of model evaluations. In situations of interest where the high-fidelity model is a discretized PDE and a single evaluation is already computationally demanding, uncertainty quantification via Monte Carlo sampling quickly becomes prohibitively expensive.

In many fields, multiple computational models are available describing a system of interest. These models have differing evaluation costs and differing fidelities. Expensive high-fidelity models describe the system with the accuracy required by the application at hand, while lower-fidelity models are less accurate but cheaper than the high-fidelity model. Often lower-fidelity models are simplified-physics representations of the problem; in other cases, they may be mathematically derived reduced-order models. Multifidelity methods seek to accelerate the solution of outer-loop applications (uncertainty quantification, optimization, inference) by combining high-fidelity and low-fidelity model evaluations [113]. The overall premise of these multifidelity methods is that low-fidelity models are leveraged for speedup while the high-fidelity model is kept in the loop to establish accuracy and/or convergence guarantees.

Our research has developed an optimal model management strategy that exploits multifidelity surrogate models to accelerate Monte Carlo estimation of statistics of outputs of computationally expensive high-fidelity models [112]. Our approach uses a control variate estimator, leveraging cheap surrogate models for speedup together with occasional recourse to the high-fidelity model to guarantee an unbiased estimator of statistics of interest. This guarantee on the estimator holds even in the absence of error bounds and error estimators for the surrogate models themselves. Solution of an optimization problem balances the number of model evaluations across the high-fidelity and surrogate models with respect to error and costs. The distribution of the number of model evaluations is optimal in the sense that our multifidelity Monte Carlo (MFMC) method yields an estimator with minimal mean-squared error for a given computational budget. Our MFMC method is applicable to general collections of surrogate models, which may include projection-based reduced models, data-fit models, support vector machines, and simplified-physics models.

In ongoing research, the asymptotic convergence behavior of multifidelity estimators is investigated for special cases of lower-fidelity models with known error and cost rates. A special case of interest is multilevel Monte Carlo, where the high-fidelity model is a discretized PDE, and lower-fidelity models are constructed by coarsening the high-fidelity mesh. Under certain assumptions on the error and cost rates of the coarse-grid approximations, multilevel Monte Carlo yields an estimator with mean-squared error in $\mathcal{O}(\varepsilon)$ and with costs bounded by $\mathcal{O}(\varepsilon^{-1})$. First results suggest that the multifidelity estimators with the optimal model management strategy achieve the same error/cost behavior if applied in the same setting as multilevel Monte Carlo. In that sense, multilevel Monte Carlo methods can be seen as a special class of the multifidelity methods. On the other hand, other instances of the multifidelity framework for which a variety of approximate models, e.g., reduced order, data fit, etc., are introduced can result in even greater cost saving when compared to multilevel Monte Carlo.

As an example, we have considered LANL subsurface flow models. Upscaling permeability is a fundamental problem: the heterogeneity represented by our models (often tens of meters) is much

larger than the heterogeneity in the subsurface (often centimeters or smaller). Here, we upscale permeability by solving the flow equations for a heterogeneous medium, computing the ratio between the flux out of the domain and the pressure gradient across the domain, then applying Darcy's law to estimate the effective permeability of the large scale system. The MFMC simulation uses three models. The high-fidelity model discretizes the domain into a 16×16 grid, using a permeability field defined by permeabilities at pilot points on a 4×4 grid. The medium-fidelity model is an 8×8 discretization using the permeability field defined by the same pilot points. The lowest-fidelity model is a support vector machine that takes the permeabilities at the pilot points as inputs. Fig. 10 shows that to achieve the same level of accuracy in the mean estimate of the hydraulic conductivity, the multifidelity Monte Carlo results in over an order of magnitude savings in computational time.

This research crosscuts core areas of applied mathematics in multifidelity methods, model reduction, optimal design and control, and uncertainty quantification. The research contributions include mathematical analysis of the convergence behavior of multifidelity estimators under various assumptions on the surrogate models; new scalable algorithms for multifidelity estimators of mean, variance, and sensitivity indices, with code for MFMC mean estimators contributed to an open-source Github repo; and application of the methods to a variety of problems, including LANL subsurface simulation tools. Across these applications, we consistently see that our multifidelity approach achieves speedups of orders of magnitude. This translates into reduction from many hours to just a few minutes for UQ in complex PDE models.

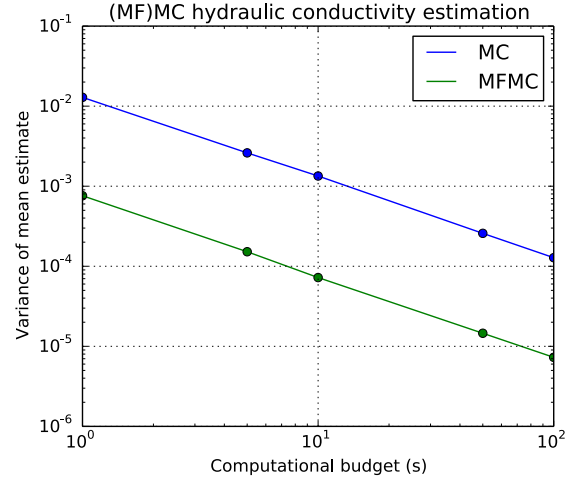


Figure 10: Multifidelity Monte Carlo (MFMC) leads to over 10X speedups in estimating mean hydraulic conductivity compared to standard Monte Carlo (MC).