Hybrid multi-level Monte Carlo polynomial chaos method for global sensitivity analysis

Mike Merritt

Gianluca Geraci, Mike Eldred

Sandia National Labs, Department of Optimization and Uncertainty Quantification

September 3, 2020

Sensitivity analysis for expensive models

- We consider high-dimensional and high-fidelity models:
  - Large number of uncertain parameters
  - Model represents the system of interest with high accuracy
  - High accuracy models are often prohibitively expensive
- Sensitivity analysis allows one to characterize the uncertainty in such models
- Performing sensitivity analysis can be prohibitively expensive due to the larger number of function evaluations required
- We consider a hierarchy of related models, which are organized by fidelity/accuracy and the corresponding cost
- Our goal is to perform sensitivity analysis efficiently on a high-fidelity model by leveraging the information provided by the cheaper, lower-fidelity models

Polynomial chaos expansions - NISP

- Given a scalar-valued function $Q(\boldsymbol{\xi})$ with random vector $\boldsymbol{\xi}$, the polynomial chaos expansion (PCE) of Q is given as

$$Q_{PC}(\boldsymbol{\xi}) = \sum_{k=0}^{P} \beta_k \Psi_k(\boldsymbol{\xi}), \quad \beta_k = \frac{\mathbb{E}[Q(\boldsymbol{\xi})\Psi(\boldsymbol{\xi})]}{\mathbb{E}[\Psi_k^2(\boldsymbol{\xi})]}$$

  where $\{\Psi_k\}_{k\geq 1}$ is a family of orthogonal polynomials, the β_k's are the corresponding PCE coefficients, and P is the truncation level
- For $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_d)$, the polynomial basis is a tensor product of 1D orthogonal polynomials

$$\Psi_k(\xi_1, \ldots, \xi_d) = \prod_{i=1}^{d} \psi_{a_i^k}(\xi_i), \quad a^k = (a_1^k, \ldots, a_d^k),$$

  where a_i^k is a multi-index, denoting the degree of the ith 1D polynomial for the kth multivariate polynomial

Polynomial chaos expansions

- The choice of polynomial basis is made to guarantee orthogonality with respect to the distribution of $\boldsymbol{\xi}$
(e.g. Normal and Hermite, Uniform and Legendre, etc.)[1]
- Computing a PCE can become prohibitively expensive for a high-dimensional Q due to the number of terms involved,

$$P + 1 = \frac{(p+d)!}{p!d!}, \quad \text{where } p = \text{total polynomial order}$$

- PCE is also well-suited for functions with some underlying smoothness

[1] Le Maıtre and Knio, *Spectral methods for uncertainty quantification: with applications to computational fluid dynamics.*

PCE for GSA

- One notable benefit of PCE is the ability to compute global sensitivity analysis (GSA) indices as a post-process
- Given a square-integrable, scalar-valued $f(\boldsymbol{\xi})$ with $\boldsymbol{\xi} = (\xi_1, \dots, \xi_d)$, the ANOVA decomposition of f is defined as:

$$f(\boldsymbol{\xi}) = f_0 + \sum_{i=1}^{d} f_i(\xi_i) + \sum_{i<j}^{d} f_{i,j}(\xi_i, \xi_j) + \cdots + f_{1,\dots,d}(\boldsymbol{\xi}_{1,\dots,d})$$

where

$$\begin{aligned} f_0 &= \mathbb{E}[f(\boldsymbol{\xi})] \\ f_i(\xi_i) &= \mathbb{E}[f(\boldsymbol{\xi})|\xi_i] - f_0 \\ f_{i,j}(\xi_i, \xi_j) &= \mathbb{E}[f(\boldsymbol{\xi})|\xi_i, \xi_j] - f_i - f_j - f_0 \ \ \dots \end{aligned}$$

- For a PC expansion, computing the ANOVA decomposition simply involves summing the proper terms

PCE for GSA

- Given the ANOVA decomposition of f, one may define the Sobol' indices w.r.t. $u \subseteq \{1, \ldots, d\}$ as:

$$S_u(f) = \frac{\mathbb{V}ar[\mathbb{E}[f(\boldsymbol{\xi}) \mid \boldsymbol{\xi}_u]]}{\mathbb{V}ar[f(\boldsymbol{\xi})]} \quad \text{and} \quad T_u(f) = \sum_{v \cap u \neq \emptyset} S_v(f),$$

  where the order of S_u is $|u|$ and T_u is a total index
- In order to compute the Sobol' indices of Q_{PC}, we have

$$\mathbb{V}ar[Q_{PC}] = \mathbb{E}[Q_{PC}^2] - \mathbb{E}[Q_{PC}]^2 = \sum_{k=1}^{P} \beta_k^2 \, \mathbb{E}[\Psi_k^2]$$

$$S_u(Q_{PC}) = \frac{\sum_{k \in K_u} \beta_k^2 \, \mathbb{E}[\Psi_k^2]}{\sum_{k=1}^{P} \beta_k^2 \, \mathbb{E}[\Psi_k^2]}$$

- Here, K_u denotes the indices of the PCE terms that only depend on the parameter subset $\boldsymbol{\xi}_u$

PCE for GSA

- The limiting factors in computing GSA indices with a PCE are the error in each $\hat{\beta}_k$ and the truncation level P
- For many polynomial families, the norms $\mathbb{E}[\Psi_k^2]$ are known analytically, so the real cost in building a PCE is in computing the spectral projection, $\mathbb{E}[Q\Psi_k]$
- A variety of methods exist for this task, including quadrature methods, Galerkin projection, least squares approximations[2]
- We will estimate this expectation using Monte Carlo integration
- In the case of a high-dimensional Q, we will attempt to accelerate the estimation of $\mathbb{E}[Q\Psi_k]$ using a hierarchy of related models

[2] Crestaux, Le Maıtre, and Martinez, "Polynomial chaos expansion for sensitivity analysis".

Monte Carlo sampling

- Let $Q(\boldsymbol{\xi})$ be a scalar quantity of interest from a high-fidelity model, where $\boldsymbol{\xi}$ is a vector of uncertain parameters
- We want to compute $\mathbb{E}[Q(\boldsymbol{\xi})]$ and we define the estimator

$$\hat{Q} = \frac{1}{N}\sum_{i=1}^{N} Q(\boldsymbol{\xi}^i)$$

- If $\boldsymbol{\xi}^1, \ldots, \boldsymbol{\xi}^N$ are i.i.d., then $\hat{Q}$ is an unbiased estimator (i.e. $\mathbb{E}[\hat{Q}] = \mathbb{E}[Q]$) and the mean-squared error (MSE) is given by

$$\mathbb{E}[(\hat{Q} - \mathbb{E}[Q])^2] = \frac{\mathbb{V}ar[Q]}{N} + (\mathbb{E}[\hat{Q} - Q])^2 = \frac{\mathbb{V}ar[Q]}{N}$$

- Reducing the MSE through sampling alone can be expensive because convergence will be slow: rate $\mathcal{O}(N^{-1/2})$
- Another approach is to decrease $\mathbb{V}ar[Q]$, without changing $\mathbb{E}[Q]$
- In general, estimator bias is not guaranteed to be zero

Multi-level Monte Carlo

- Let $Q_0, Q_1, \dots, Q_L$ denote a hierarchy of models parameterized by a scalar ℓ with associated costs $C_0 \leq C_1 \leq \cdots \leq C_L$ for each "level"
  - A natural example of this is solving a differential equation on a mesh where the number of points is controlled by the index ℓ for Q_ℓ
- If we want to estimate $\mathbb{E}[Q_L]$, we can use

$$\mathbb{E}[Q_L] = \sum_{\ell=0}^{L} \mathbb{E}[Q_\ell - Q_{\ell-1}], \quad Q_{-1} = 0$$

- This defines a multi-level Monte Carlo (MLMC) estimator

$$\hat{Q}_L^{ML} = \sum_{\ell=0}^{L} Q_\ell \widehat{- Q}_{\ell-1} = \sum_{\ell=0}^{L} \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} Q_\ell^i - Q_{\ell-1}^i,$$

with the associated cost $C_{tot} = \sum_{\ell=0}^{L} N_\ell (C_\ell + C_{\ell-1})$

Multi-level Monte Carlo

- The MSE of an unbiased $\hat{Q}_L^{ML}$ can be expressed as

$$\mathbb{E}\left[(\hat{Q}_L^{ML} - \mathbb{E}[Q_L])^2\right] = \mathbb{V}ar[\hat{Q}_L^{ML}] = \sum_{\ell=0}^{L} \frac{\mathbb{V}ar\left[Q_\ell - Q_{\ell-1}\right]}{N_\ell},$$

  where independent sampling among levels removes any covariance
- The goal then is to minimize $\mathbb{V}ar[\hat{Q}_L^{ML}]$ by appropriately allocating N_ℓ samples to each level
- In the case that $\mathbb{V}ar[Q_\ell - Q_{\ell-1}]$ is decreasing for $\ell \to L$, one is able to evaluate the majority of samples at the cheaper levels
- The optimization problem:

$$\min_{N_0,\dots,N_L} C_{tot} \quad \text{s.t.} \quad \mathbb{V}ar[\hat{Q}_L^{ML}] \leq \varepsilon^2$$

  can be solved in closed form for the optimal sample allocation[3]

[3] Giles, "Multilevel monte carlo methods".

MLMC for PCE formulation

- Returning to the estimation of PCE coefficients, we decompose the spectral projection as

$$\mathbb{E}[Q\Psi_k] = \sum_{\ell=0}^{L} \mathbb{E}[(Q_\ell - Q_{\ell-1})\Psi_k],$$

which leads to the multi-level estimator

$$\hat{\beta}_k = \frac{1}{\mathbb{E}[\Psi_k^2]} \sum_{\ell=0}^{L} (Q_\ell \widehat{- Q_{\ell-1}})\Psi_k = \frac{1}{\mathbb{E}[\Psi_k^2]} \sum_{\ell=0}^{L} \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} (Q_\ell^i - Q_{\ell-1}^i)\Psi_k^i$$

- In this case, the variance of the estimator can be expressed as

$$\mathbb{V}ar[\hat{\beta}_k] = \frac{1}{\mathbb{E}[\Psi_k^2]^2} \sum_{\ell=0}^{L} \frac{\mathbb{V}ar[(Q_\ell - Q_{\ell-1})\Psi_k]}{N_\ell}$$

MLMC for PCE formulation

- We want to find an optimal sampling allocation scheme that balances the estimator variance and the cost of sampling the QoIs
- We have the optimization problem:

$$\min_{N_0,\dots,N_L} \sum_{\ell=0}^{L} N_\ell C_\ell + \mu^2 \left(\mathbb{V}ar[\hat{\beta}_k] - \varepsilon^2 \right),$$

where μ^2 is a Lagrange multipler and ε^2 is the target variance.
- The optimal sample allocation[4] can be shown to be

$$N_\ell = \mu \sqrt{\frac{\mathbb{V}ar[(Q_\ell - Q_{\ell-1})\Psi_k]}{\mathbb{E}[\Psi_k^2]\ C_\ell}} \quad \text{where} \tag{1}$$

$$\mu = \varepsilon^{-2} \sum_{\ell=0}^{L} \frac{\sqrt{\mathbb{V}ar[(Q_\ell - Q_{\ell-1})\Psi_k]\ C_\ell}}{\mathbb{E}[\Psi_k^2]} \tag{2}$$

[4] Giles, "Multilevel monte carlo methods".

Computing ensembles of PCE coefficients

- For the purposes of GSA, we need a set of PCE coefficients. Thus for L levels and P coefficients, we estimate

$$\mathbb{E}[Q\Psi_k] = \sum_{\ell=0}^{L} \mathbb{E}[(Q_\ell - Q_{\ell-1})\Psi_k], \quad k = 0, 1, \ldots, P$$

- Ideally, we share N_ℓ samples when estimating $\mathbb{E}[(Q_\ell - Q_{\ell-1})\Psi_k]$ for $k = 1, \ldots, P$
- We have two initial sample allocation schemes:
    1. Individual: estimate each β_k separately, each $\hat{\beta}_k$ has the target variance, with no sample sharing this is expensive
    2. Worst case: same samples at each level where N_ℓ is computed to minimize $\max_k \mathbb{V}ar[(Q_\ell - Q_{\ell-1})\Psi_k]$ - the worst case coefficient
- These schemes work for more general sets of PCE coefficients

Example algorithm - estimating a single PCE coefficient

Algorithm 1 Estimate kth PCE coefficient

Input: Multi-level model, target accuracy ε^2, number of pilot samples N, coefficient to estimate k, max iterations

Output: coefficients $\hat{\beta}_k$, estimator variance $\mathbb{V}[\hat{\beta}_k]$, function evaluations

Draw N pilot samples: $\boldsymbol{\xi}_{pilot}$ {Distribution included with the model}

Evaluate Ψ_k and $Q_\ell - Q_{\ell-1}$ for $\ell = 0, \ldots, L$ at $\boldsymbol{\xi}_{pilot}$

while $\mathbb{V}ar[\hat{\beta}_k] > \varepsilon^2$ and iteration $<$ max iterations **do**

    for $\ell = 0, \ldots, L$ **do**

        Estimate $\mathbb{V}ar[(Q_\ell - Q_{\ell-1})\Psi_k]$

        Compute N_ℓ in order to minimize $\mathbb{V}ar[\hat{\beta}_k]$ {See (1) and (2)}

        Draw additional samples of $\boldsymbol{\xi}$

        Evaluate functions Ψ_k and $Q_\ell - Q_{\ell-1}$ for $\ell = 0, \ldots, L$

    end for

    Estimate $\mathbb{E}[(Q_\ell - Q_{\ell-1})\Psi_k]$

    Compute $\mathbb{V}ar[\hat{\beta}_k]$

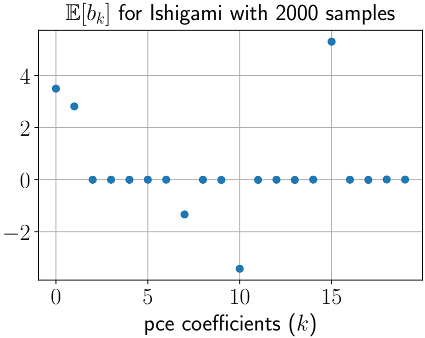
end while

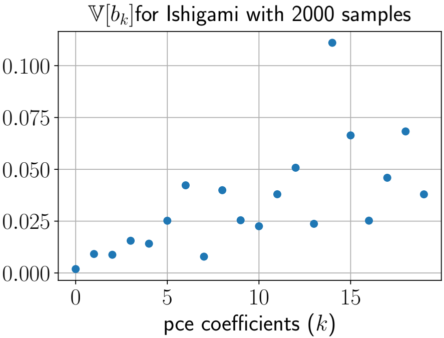
Compute final estimate of $\hat{\beta}_k$

Example with Ishigami function

$$f(\boldsymbol{\xi}) = \sin(\xi_1) + a\sin^2(\xi_2) + b\xi_3\sin(\xi_1)$$
$$\xi_i \sim \mathcal{U}[-\pi,\pi],\ i=1,2,3 \quad a=7,\ b=0.1$$

- We take the first 20 PCE modes and compute 2000 realizations of the coefficients, looking at the mean and variance of $\hat{\beta}_k$

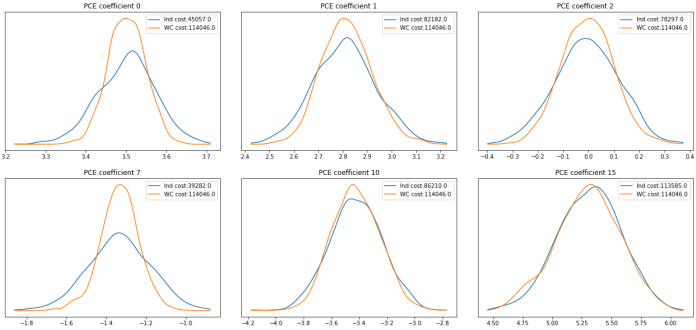




Example with multi-level Ishigami

$$
\begin{aligned}
Q &= \sin(\xi_1) + a\sin^2(\xi_2) + b\xi_3^4\sin(\xi_1) \\
Q_0 &: a = (0.6)7.0,\ \ b = (0.6)0.1,\ \ C_0 = 1 \\
Q_1 &: a = (0.8)7.0,\ \ b = (0.8)0.1,\ \ C_1 = 10 \\
Q_2 &: a = (1.0)7.0,\ \ b = (1.0)0.1,\ \ C_2 = 100
\end{aligned} \tag{3}
$$

- We compare the individual and worst-case methods for



The worst case coefficient is β_{15}. Cost is provided in the legend.

Globally optimal sample allocation scheme

- We want a sample allocation method that balances the total cost and variance for each of the estimated coefficients
- Moreover, the real goal is in targeting the Sobol' indices, minimizing the variance of $\hat{S}_u$. Thus given a PCE

$$Q_{PC} = \sum_{k=1}^{P} \hat{\beta}_k \Psi_k \quad \text{we consider} \quad \mathbb{V}ar\left[\frac{\sum_{k \in K_u} (\hat{\beta}_k)^2 \ \mathbb{E}[\Psi_k^2]}{\sum_{k=1}^{P} (\hat{\beta}_k)^2 \ \mathbb{E}[\Psi_k^2]}\right]$$

- We instead can target particular indices by subdividing

$$\mathbb{V}ar\left[\sum_{k=1}^{P} \hat{\beta}_k^2 \ \mathbb{E}[\Psi_k^2]\right] = \sum_{k=1}^{P} \mathbb{V}ar[\hat{\beta}_k^2] \ \mathbb{E}[\Psi_k^2]^2 + \sum_{k \neq z} \mathbb{E}[\Psi_k^2]\mathbb{E}[\Psi_z^2]\mathbb{C}ov\left[\hat{\beta}_k^2, \hat{\beta}_z^2\right]$$

- Expressions are needed for $\mathbb{V}ar[\hat{\beta}_k^2]$ and $\mathbb{C}ov\left[\hat{\beta}_k^2, \hat{\beta}_z^2\right]$ in terms of moments of $Q_\ell - Q_{\ell-1}$, Ψ_k, and Ψ_z

Other concerns

- Given an unbiased estimator for β_k (i.e. $\mathbb{E}[\hat{\beta}_k] = \beta_k$), the quantity $(\hat{\beta}_k)^2$ will be a biased estimator for β_k^2, resulting in error
- This will require a bias correction to be incorporated into the estimator for each β_k^2
- This, in turn, will require unbiased estimators for the moments of the relevant $Q_\ell - Q_{\ell-1}$ and Ψ_k terms
- If $\mathbb{V}ar[\hat{\beta}_k]$ increases with k (as expected), how does one determine the appropriate truncation level?

Ongoing work

- Completed derivation of multi-level variances and covariances
- Derivation of sample allocation strategy, targeting a given accuracy for a set of GSA indices
- Rigorous evaluation of the efficiency of PCE and MLMC-PCE hybrid for GSA, considering the effects of dimension and regularity
- Extension of this hybrid MLMC-PCE method to multi-fidelity models and Approximate Control Variates[5]

[5] Gorodetsky et al., "A generalized approximate control variate framework for multifidelity uncertainty quantification".

References

Crestaux, Thierry, Olivier Le Maıtre, and Jean-Marc Martinez. "Polynomial chaos expansion for sensitivity analysis". *Reliability Engineering & System Safety* 94.7 (2009), pp. 1161–1172.

Giles, Michael B. "Multilevel monte carlo methods". *Acta Numerica* 24 (2015), p. 259.

Gorodetsky, Alex A et al. "A generalized approximate control variate framework for multifidelity uncertainty quantification". *Journal of Computational Physics* 408 (2020), p. 109257.

Le Maıtre, Olivier and Omar M Knio. *Spectral methods for uncertainty quantification: with applications to computational fluid dynamics.* Springer Science & Business Media, 2010.