Recent advancements in Multilevel-Multifidelity techniques for forward UQ in the DARPA Sequoia project

Gianluca Geraci* and Michael S. Eldred†

Sandia National Laboratories, Albuquerque, NM 87185, USA

Alex A. Gorodetsky‡

University of Michigan, Ann Arbor, MI 48109, USA

John D. Jakeman§

Sandia National Laboratories, Albuquerque, NM 87185, USA

In the context of the DARPA funded project SEQUOIA we are interested in the design under uncertainty of a jet engine nozzle subject to the performance requirements of a reconnaissance mission for a small unmanned military aircraft. This design task involves complex and expensive aero-thermo-structural computational analyses where it is of a paramount importance to also include the effect of the uncertain variables to obtain reliable predictions of the device’s performance. In this work we focus on the forward propagation analysis which is a key part of the design under uncertainty workflow. This task cannot be tackled directly by means of single fidelity approaches due to the prohibitive computational cost associated to each realization. We report here a summary of our latest advancement regarding several multilevel and multifidelity strategies designed to alleviate these challenges. The overall goal of these techniques is to reduce the computational cost of analyzing a high-fidelity model by resorting to less accurate, but less computationally demanding, lower fidelity models. The features of these multifidelity UQ approaches are initially illustrated and demonstrated on several model problems and afterward for the aero-thermo-structural analysis of the jet engine nozzle.

I. Introduction

In the context of the DARPA funded project SEQUOIA [1] we are interested in the design under uncertainties of devices that involve aero-structural-thermal interactions in advanced aerospace vehicles. One problem of interest is the analysis and design of a jet engine nozzle for small unmanned military aircrafts as, for instance, the Northrop Grumman UCAS X-47B. As in many other complex engineering systems, many sources of uncertainty are present in this problem definition, for instance the material properties, the atmospheric and flight conditions might not be known a priori. All these source of uncertainty need to be characterized and propagated through the numerical code in order to understand their impact on the final performance of the device and, ultimately, provide the basis for a design optimization under uncertainty.

The accurate prediction of the aerodynamic, thermal and mechanical loadings for the nozzle requires sophisticated and computationally expensive numerical simulations, therefore both the Uncertainty Quantification (UQ) and the Optimization Under Uncertainty (OUU) activities cannot be performed with classical single fidelity approaches. In the context of the SEQUOIA project we explored several algorithmic solutions to accelerate the UQ analysis with a lower numerical cost and without sacrificing the overall accuracy. We contributed in designing and analyzing both sampling based approaches and surrogate based methods. Moreover, we also proposed a latent variable network approach that might potentially bridge these two areas.

In Fig. 1 we report a flowchart of the different activities that constitute our effort in advancing multifidelity UQ algorithm for complex engineering analysis.

*Limited Term Employee, Optimization and Uncertainty Estimation, Sandia National Laboratories, 1515 Eubank SE, MS1318, Albuquerque, NM 87123, AIAA Member.
†Distinguished Member of Technical Staff, Optimization and Uncertainty Estimation, Sandia National Laboratories, 1515 Eubank SE, MS1318, Albuquerque, NM 87123, AIAA Associate Fellow.
‡Assistant Professor, Department of Aerospace Engineering, University of Michigan, AIAA Member
§Senior Member of Technical Staff, Optimization and Uncertainty Estimation, Sandia National Laboratories, 1515 Eubank SE, MS1318, Albuquerque, NM 87123.
Table: Algorithms and Contributions

<table>
<thead>
<tr>
<th></th>
<th>Recursive</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sampling</strong></td>
<td>Multilevel Monte Carlo</td>
<td>Approximate control variates</td>
</tr>
<tr>
<td></td>
<td>Multifidelity Monte Carlo</td>
<td></td>
</tr>
<tr>
<td><strong>Surrogates</strong></td>
<td>Multi-index collocation</td>
<td>Bayesian latent variable networks</td>
</tr>
<tr>
<td></td>
<td>Adaptive multi-index collocation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Greedy adaptive MLMC</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 1  The relationships between algorithms proposed in this work. Bold-faced text indicates our contributions.

The remainder of the paper is structured as follows. Algorithmic highlights are presented for both the sampling-based approaches and the surrogate-based approaches in Sec. III and Sec. IV respectively. In these sections we also present numerical results on model problems in order to briefly illustrate the features of the different algorithms. Numerical results for the nozzle problems are instead reported in Sec. VI. The paper ends with conclusions in Sec. VII.

II. Aero-thermo-structural model of a jet engine nozzle

Our aero-thermo-elastic problem is inspired by the Pratt & Whitney F100-PW-220 turbofan engine without after-burning capabilities, that powers the Northrop Grumman UCAS X-47B military unmanned aircraft, for which we want to design its nozzle. During normal operative conditions the nozzle structure is exposed to both thermal and mechanical loading. Therefore it needs to include a two-layers array, namely a thermal and a loading layer, placed internally and externally, respectively.

Evaluating the performance of this device requires solving a multi-physics problem, and this solution consists of the following steps. An engine model simulator [9–11] provides the inlet conditions for the nozzle. These conditions are subsequently used as input for the computational fluid dynamics (CFD) solver SU2 [4]. This solver computes the solution for the internal flow and the temperature and pressure profile along the nozzle internal walls. The temperature and pressure loading are used by a Finite Element (FEM) solver AERO-S [8] which computes several quantities regarding the thermal and mechanical stresses inside the structure. More details on the computational approach adopted are provided in [13].

The intrinsic multiphysics nature of this problem provides a natural stage for the development and test of several model alternatives. For instance, a 1D ideal solver can be adopted to replace the SU2 CFD solver and consequently decrease the overall computational cost. In a previous work we demonstrated that a similar strategy might be exploited in a multifidelity forward UQ analysis [16]. In the same spirit, in this contribution we deploy an array of model fidelities that ranges from the 1D ideal solver to a more accurate 3D Reynolds-Averaged Navier Stokes (RANS) model for the CFD. For the FEM solvers, we relied on several meshes with increasing spatial resolution and even axi-symmetric mesh configurations to avoid the cost associated to the cost of a fully three-dimensional solution.

1. Nozzle structure

The structure of nozzle’s walls is comprised of the following elements. An inner thermal layer and an outer structural layer which are separated by an air gap. The material for the thermal layer is a ceramic matrix composite that insulates the structural layer from high temperatures. The structural layer is constituted by a composite sandwich material composed of a titanium honeycomb between two layers of graphite-bismaleimide (Gr/BMI) composite material. The air gap between the layers adds additional insulation from the thermal loading for the structural layer. Additional supports between these two layers are not modeled, however the nozzle structure also includes six composite sandwich baffles to attach the nozzle to the surrounding structure and stingers that provide additional support in the axial direction. For a schematic view of the nozzle structure the reader should refer to [12].
2. Operative conditions

The operative conditions for the nozzle are inspired by a typical reconnaissance mission for a small unmanned high-subsonic military aircraft. In particular, we decided to focus on the most critical conditions which occurs for the top-of-climb condition when the required thrust is 21,500 N at an altitude of approximately 40,000 ft and Mach 0.511 [12].

3. Uncertain parameters

Following the numerical setup in [12] we use 40 uncertain parameters that are divided in 35 material properties variables, 2 atmospheric conditions variables, 2 inlet conditions and 1 heat transfer. All the variables are considered to be independent. More details on the bounds and distributions of the 40 uncertain parameters are reported in [12] however, we mention here that for this problem we deal with a set of uniformly and log-normally distributed variables.

III. Sampling-based approaches: highlights

A. Approximate Control Variate

The Monte Carlo algorithm is popular due to its simplicity, flexibility, and provably convergent behavior. Let \((\Omega, \mathcal{F}, P)\) a probability space and \(Z \in \mathbb{R}^d\) a vector valued random variable defined on this space. We consider a generic QoI \(Q\) as the mapping from \(Z\) to a scalar valued output. We desire to compute some statistics of \(Q\), for instance its expected value \(\mathbb{E}[Q]\). The MC estimator for the expected value of \(Q\) based on \(N\) realizations is

\[
\hat{Q} = \frac{1}{N} \sum_{i=1}^{N} Q^{(i)}.
\]

A classical results for MC is derived from the Central Limit Theorem and states that the mean square error for this value.

\[
\text{var} \{Q^{(i)}\} \rightarrow \frac{\text{var} \{Q\}}{N}, \quad \text{for} \ N \rightarrow \infty.
\]

The Monte Carlo algorithm is popular due to its simplicity, flexibility, and provably convergent behavior. Let \((\Omega, \mathcal{F}, P)\) a probability space and \(Z \in \mathbb{R}^d\) a vector valued random variable defined on this space. We consider a generic QoI \(Q\) as the mapping from \(Z\) to a scalar valued output. We desire to compute some statistics of \(Q\), for instance its expected value \(\mathbb{E}[Q]\). The MC estimator for the expected value of \(Q\) based on \(N\) realizations is

\[
\hat{Q} = \frac{1}{N} \sum_{i=1}^{N} Q^{(i)}.
\]

Moreover the unbiased terms are weighted by a sets of weights \(\alpha = [\alpha_1, \ldots, \alpha_M]^T\). The optimal set of weights can be obtained by maximizing the variance reduction of the estimator \[21\]. If we denote \(C \in \mathbb{R}^{M \times M}\) as the covariance matrix among \(Q_i\) and \(c \in \mathbb{R}^M\) as the vector of covariances between \(Q\) and each \(Q_i\), the highest variance reduction is obtained by choosing \(\alpha^* = -C^{-1}c\). If we further define \(\bar{c} = c/\text{var}[Q] = [\rho_1 \sqrt{\text{var}[Q]}, \ldots, \rho_M \sqrt{\text{var}[Q]}]^{T}\), where \(\rho_i\) is the Pearson’s correlation coefficient between \(Q\) and \(Q_i\), then the variance reduction corresponding to optimal weights becomes

\[
\gamma(CV, \alpha^*) = \frac{\text{var}[\hat{Q}_{CV}]}{\text{var}[\hat{Q}]} = 1 - R_{OCV}^2 = 1 - \bar{c}^T C^{-1} \bar{c}, \quad 0 \leq R_{OCV}^2 \leq 1.
\]

For one single low-fidelity model we obtain the classical result \(R_{OCV}^2 = \rho_1^2\) and therefore the variance reduction is equal to the correlation of the low-fidelity model \(\gamma(CV, \alpha^*) = 1 - \rho_1^2\) where \(\alpha_1 = \rho_1 \sqrt{\text{var}[Q]} / \sqrt{\text{var}[Q]}\).
2. Approximate Control Variate

Traditional control variate assumes that the expected value of each low-fidelity model \( \mu_i \) is known \textit{a priori}. In the presence of complex engineering system simulations we instead have cheaper to evaluate low-fidelity models whose means we also need to estimate. To do so we consider \( z \) a set of samples used to evaluate the high-fidelity model \( Q \) and \( z_i \) an ordered set of \( N_i \) evaluations for \( Q_i \). Furthermore, we partition \( z_i \) in two ordered subsets according to \( z^1_i \cup z^2_i = z_i \) (note that in general \( z^1_i \cap z^2_i \neq \emptyset \)).

Combining these subsets, we obtain the following for the \textit{approximate control variate}

\[
\hat{Q}(\alpha, z) = \hat{Q}(z) + \sum_{i=1}^{M} \alpha_i \left( \hat{Q}_i(z^1_i) - \hat{\mu}_i(z^2_i) \right) = \hat{Q}(z) + \sum_{i=1}^{M} \alpha_i \Delta_i(z_i) = \hat{Q} + \alpha^T \Delta
\]

where \( \Delta = [\Delta_1(z_1), \ldots, \Delta_M(z_M)]^T \), \( \Delta_i(z_i) = \hat{Q}_i(z^1_i) - \hat{\mu}_i \), and the input values for the high- and low-fidelity models are denoted as \( z = (z_1, z_2, \ldots, z_M) \). Note that all the estimators in literature (for instance MLMC [19, 20] and MFMC [38]) can be obtained from the approximate control variate estimator (ACV) \( \hat{Q} \) by using different partitioning schemes for \( z^1_i, z^2_i \) of each low-fidelity model. To facilitate easier analysis, we assume that the number of evaluations available for each low-fidelity model \( N_i \) is proportional to the number of evaluations of the high-fidelity model \( N \), as \( N_i = [r_i N] \) where \( r_i \in \mathbb{R}_+ \). In [21] we derived the expressions for both the optimal weights \( \alpha^{ACV} \) and the variance of \( \hat{Q} \). These results are reported here for convenience but the reader must refer to [21] for their derivation

\[
\hat{Q}^{ACV} = -\text{Cov} \left[ \Delta \Delta^{-1} \text{Cov} \left[ \Delta \hat{Q} \right] \right] \\
\text{Var} \left[ \hat{Q} \right] = \text{Var} \left[ \hat{Q} \right] \left( 1 - \text{Cov} \left[ \Delta \hat{Q} \right] \text{Cov} \left[ \Delta \hat{Q} \right] \right) \\
= \text{Var} \left[ \hat{Q} \right] \left( 1 - R_{ACV}^2 \right) .
\]

For reference, if only a single low-fidelity model is used, the classical results from [34, 37] is recovered

\[
R_{ACV-1}^2 = \frac{r_i - 1}{r_i} \rho_i^2 .
\]

To interpret this result, consider the term \( \frac{r_i - 1}{r_i} < 1 \) as a penalization on the efficiency with respect to the traditional CV with a single low-fidelity model, \textit{i.e.} OCV-1 Eq. (2), where a full reduction of \( \rho_i^2 \) is achieved.

3. Multilevel Monte Carlo

In [21] we demonstrated that the MLMC framework [20] can be derived directly from the ACV estimator by using the following samples partitioning: \( z^1_i = z \) and \( z^2_i = z_{i+1}^2 \) for \( i = 1, \ldots, M - 1 \) and \( \alpha_i = -1 \) for all \( i \). Moreover, for \( i \neq j \) the set \( z^2_i \) and \( z^2_j \) are independent, \textit{i.e.} \( z^2_i \cap z^2_j = \emptyset \).

One important remark is that the MLMC implies a recursive sampling strategy, \textit{i.e.} in order to estimate the expected value of each low-fidelity model \( \hat{\mu}_i \) a model \( Q_{i+1} \) is introduced as control variate. Following these definitions, the MLMC estimator can be written as

\[
\hat{Q}^{MLMC} (z) = \hat{Q} + \sum_{i=1}^{M} (-1) \left( \hat{Q}_i(z^1_i) - \hat{\mu}_i(z^2_i) \right) .
\]

Given the partitioning of \( z_i \), the cardinality of \( z^1_i \) and \( z^2_i \) is \( \bar{r}_i N \) and \( \bar{r}_i N \), respectively. Moreover if \( \bar{r}_i = 1 \), it holds that \( r_i = \bar{r}_i + \bar{r}_{i-1} \). In this case, the estimator variance is obtained as

\[
\text{Var} \left[ \hat{Q}^{MLMC} \right] = \text{Var} \left[ \hat{Q} \right] \left( 1 - R_{MLMC}^2 \right) \\
R_{MLMC}^2 = \sum_{i=1}^{M} \frac{\bar{r}_i + \bar{r}_{i-1}^2}{\bar{r}_i \bar{r}_{i-1}} + 2 \sum_{i=1}^{M-1} \frac{\rho_{i+1} \tau_i \tau_{i+1}}{\bar{r}_i} - 2 \rho_1 \tau_1 ,
\]

where \( \tau_i = \sqrt{\text{Var}[\hat{Q}_i]/\text{Var}[\hat{Q}]} \) and \( \rho_{i+1} \) is the Pearson correlation coefficient between \( Q_i \) and \( Q_{i+1} \). The recursive sampling strategy on which the MLMC estimator is built on indeed limits the maximum attainable variance reduction
to the one of a traditional control variate Eq.(7) based on a single low fidelity model, i.e. $R^2_{\text{MLMC}} < \rho_1^2$. However, it is important to note that the MLMC estimator is not in general limited by the variance reduction obtained by an approximate control variate with only one low-fidelity model, i.e. $R^2_{\text{MLMC}}$, might be greater than $R^2_{\text{ACV}} = \frac{\rho_1 - 1}{\rho_1^2}$.

We derived the previous estimator by assuming $\alpha_i = -1$ for all $i$. This assumption can be relaxed leading virtually in all the cases to an estimator (Weighted-MLMC or W-MLMC) that achieves a greatest variance reduction for the same sample set.

4. Multifidelity Monte Carlo

The Multifidelity Monte Carlo estimator has been recently proposed in [37][38]. We demonstrated in [21] that it can be obtained as a particular realization of the ACV estimator. Therefore, in the following we summarize some of the results that stems from its analysis in the ACV context. The MFMC is obtained from the ACV estimator (see Eq. (4)) by choosing the following sampling scheme $z_i = z_{i-1}$ and $z_i \notin z_i$ for $i = 2, \ldots, M$ and $z_1 = z$ and $z_1 = z_1$. The optimal weights for this scheme are obtained as

$$a_i^{\text{MFMC}} = -\frac{\text{Cov}[\hat{Q}, Q_i]}{\text{Var}[Q_i]}, \quad \text{for } i = 1, \ldots, M,$$

and the variance of the estimator is

$$\text{Var}[\hat{Q}^{\text{MFMC}}] = \text{Var}[\hat{Q} (1 - R^2_{\text{MFMC}})],$$

$$R^2_{\text{MFMC}} = \sum_{i=1}^{M} \frac{r_i - r_{i-1}}{r_i r_{i-1}} \rho_i^2 = \rho_1 \left( \frac{r_1 - 1}{r_1} + \sum_{i=2}^{M} \frac{r_i - r_{i-1}}{r_i r_{i-1}} \rho_i^2 \right).$$

As for MLMC, it is evident from the expression of $R^2_{\text{MFMC}}$ that the variance reduction for MFMC is greater than the one obtained by ACV with only one fidelity model, $R^2_{\text{ACV}} = (r_1 - 1)/r_1 \rho_1^2$, however being based on a recursive sampling approach even MFMC is limited by the variance reduction of a traditional control variate with a single model, i.e. $R^2_{\text{OCV}} = \rho_1^2$.

5. Examples of convergent Approximate Control Variates estimators

In [21] we proposed two algorithms that enabled us to overcome the limitations of the recursive schemes like MLMC and MFMC by enabling the ACV estimator to converge to the OCV variance reduction whenever additional low-fidelity evaluations are available.

The first estimator we proposed is obtained by using the following sampling strategy. The approximation of each term $\hat{Q}_i$ is obtained by using the same set of samples $z$ used for the evaluation of the high-fidelity model. Moreover all the samples $z_i$ are used to evaluate the low-fidelity control mean $\hat{\mu}_i$, however the additional set of samples used for the evaluation of this latter term are unique to each low-fidelity model. More formally, these assumptions translate in the following definitions $z_i^1 = z$, $z_i^2 = z_i$ and $(z_i \setminus z_i^1) \cap (z_i \setminus z_i^2) = 0$. Since the estimator for $\hat{\mu}_i$ is obtained with independent samples for each low-fidelity model, we called this algorithm ACV-IS. The optimal weights and the variance for the ACV-IS estimator are

$$\hat{Q}^{\text{ACV-IS}} = -\left[ C \cdot F^{(\text{IS})} \right]^{-1} \left[ \text{diag}(F^{(\text{IS})}) \cdot c \right],$$

$$\text{Var}[\hat{Q}^{\text{ACV-IS}}] = \text{Var}[\hat{Q} (1 - R^2_{\text{ACV}})].$$

where

$$R^2_{\text{ACV-IS}} = \left[ \text{diag}(F^{(\text{IS})}) \cdot c \right]^T \left[ C \cdot \text{diag}(F^{(\text{IS})}) \right]^{-1} \left[ \text{diag}(F^{(\text{IS})}) \cdot c \right]^T.$$

We indicated with $\circ$ the Hadamard, i.e. element-by-element product between two matrices, and $F^{(\text{IS})} \in \mathbb{R}^{M \times M}$ is the matrix that encodes the sampling structure of the particular ACV estimator. For the ACV-IS this matrix is defined as (see [21] for the full derivation of this result)

$$F^{(\text{IS})}_{ij} = \left\{ \begin{array}{cc} \frac{r_i - 1}{r_i} & \text{if } i \neq j \\ \frac{r_i - 1}{r_i} & \text{otherwise} \end{array} \right..$$

5
A second convergent ACV estimator is obtained as a slight modified version of MFMC. If the same sampling strategy of MFMC is used but the term \( \hat{Q}_i \) is evaluated by only using the first \( N \) samples (the ones for which the high-fidelity model is evaluated), it is possible to show that variance reduction of the estimator tends to the one of the OCV estimator for infinite number of low-fidelity evaluations. The formal definition of this estimator, that we named ACV-MF, is obtained from the OCV estimator Eq. 4 by using \( z_i^1 = z, z_i^2 = z \) and \( z_i^{(k)} = z_i^{(k)} \) for \( j > i \) and \( k \leq \min(r_i,r_j)N \). For ACV-MF the optimal weights and the variance are obtained as

\[
\alpha_{\text{ACV-MF}} = - \left[ C \circ F^{\text{MF}} \right]^{-1} \left[ \text{diag} \left( F^{\text{MF}} \right) \circ \epsilon \right]
\]

where

\[
\forall ar \left[ Q^{\text{ACV-MF}} (\alpha^{\text{ACV-MF}}) \right] = \forall ar \left[ \hat{Q} \right] \left( 1 - R^2_{\text{ACV-MF}} \right),
\]

Even in this case the matrix \( F^{\text{MF}} \in \mathbb{R}^{M \times M} \) encodes the particular sampling strategy and is defined as

\[
F^{\text{MF}}_{ij} = \begin{cases} \frac{\min(r_i,r_j)-1}{\min(r_i,r_j)} & \text{if } i \neq j \\ \frac{1}{r_i} & \text{otherwise} \end{cases}
\]

Interestingly, the form of the optimal estimators for ACV-IS and ACV-MF only differ in terms for \( F^{\text{MF}} \) and \( F^{\text{IS}} \). The way in which these matrices enter is algebraically identical. For both matrices it is possible to show that their limit for \( r_i \to \infty \) and \( i = 1, \ldots, M \) is \( R^2_{\text{OCV}} \). This feature enable both estimators to overcome the limitation of both MLMC and MFMC for which the variance reduction is always smaller than \( R^2_{\text{OCV}-1} \).

6. Summary and relation between the sampling schemes

We report in Table 1 a summary of all the estimators we presented so far, their sampling strategies and variance reduction.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Relation between ( z ) and ( z_i )</th>
<th>( z_i^1 )</th>
<th>( z_i^2 )</th>
<th>Reduction ratio ( \gamma )</th>
</tr>
</thead>
<tbody>
<tr>
<td>OCV</td>
<td>( z_i = z )</td>
<td>( z )</td>
<td>( 0 )</td>
<td>( 1 - R^2_{\text{OCV}} )</td>
</tr>
<tr>
<td>OCV-1</td>
<td>( z_i \supseteq z )</td>
<td>( z )</td>
<td>( 0 )</td>
<td>( 1 - R^2_{\text{OCV}} )</td>
</tr>
<tr>
<td>MLMC [19]</td>
<td>( z_i^1 = z, z_i \cap z_i = \emptyset ) for ( i &gt; 1 )</td>
<td>( z_i^{(k)} )</td>
<td>( z_i^{(k)} )</td>
<td>( 1 - R^2_{\text{MLMC}} ), ( R^2_{\text{MLMC}} &lt; \rho^2_1 )</td>
</tr>
<tr>
<td>MFMC [38]</td>
<td>( z_i \supseteq z ) for all ( i )</td>
<td>( z_i^{(k)} )</td>
<td>( z_i^{(k)} )</td>
<td>( 1 - R^2_{\text{MFMC}} ), ( R^2_{\text{MFMC}} &lt; \rho^2_1 )</td>
</tr>
<tr>
<td>ACV-1</td>
<td>( z_i \supseteq z )</td>
<td>( z )</td>
<td>( z )</td>
<td>( 1 - R^2_{\text{ACV-1}} )</td>
</tr>
<tr>
<td>ACV-IS</td>
<td>( z_i = z^1, z_i^1 \cap z_i^2 = \emptyset ) for ( 1 \leq i \neq j )</td>
<td>( z_i )</td>
<td>( z_i )</td>
<td>( 1 - R^2_{\text{ACV-IS}} )</td>
</tr>
<tr>
<td>ACV-MF</td>
<td>( z_i \cap (z_i \setminus z_i^j) = \emptyset )</td>
<td>( z )</td>
<td>( z )</td>
<td>( 1 - R^2_{\text{ACV-MF}} )</td>
</tr>
</tbody>
</table>

Table 1: Summary of the sampling scheme for several multifidelity sampling estimator from [21].

7. Accelerating the Approximate Control Variate

We demonstrated in [21] that a recursive sampling strategy can be exploited to accelerate the estimator to a target control variate level in case there is not enough data from the higher-fidelity models. The reason for that is that a recursive scheme can achieve the faster convergence to a given target control variate (for instance the OCV-1 or OCV-2). One possible approach is to partition the control variates in two groups: the first \( K \) variables form a \( K \)-level approximate control variate, and the last \( M - K \) variables are used to reduce the variance of estimating \( \mu_L \) for some \( L \leq K \). The resulting estimator accelerates quickly to OCV-\( K \), and \( L \) provides a degree of freedom for targeting a control variate level that contributes the greatest to the estimator variance. We report in the following only the final results for one estimator we derived, for the entire derivation the interested reader should refer to [21]. Let \( K, L \leq M \) and \( K \in \mathbb{Z}_+ \) with \( 0 \leq L \leq K \). The ACV-KL estimator is given by the following expression and sampling scheme

\[
\hat{Q}^{\text{ACV-KL}}(\alpha, z) = \hat{Q}(z) + \sum_{i=1}^{K} \alpha_i \left( \hat{Q}_i(z) - \hat{\mu}_i(z_i) \right) + \sum_{i=K+1}^{M} \alpha_i \left( \hat{Q}_i(z_i) - \hat{\mu}_i(z_i) \right),
\]
where we used that \( z_i^1 = z \) for \( i \leq K \) and \( z_i^j = z_i \) for \( i > K \). Furthermore \( z_i^2 = z_j \) for all \( i \). The sets \( z_i \setminus z_i^1 \) can be chosen in several ways. Here we choose the same sampling strategy as ACV-MF: \( z_i^1 = z, z_i^2 = z_i \) and \( z_i^{(k)} = z_i^{(k)} \) for \( j > i \) and \( k \leq \min(r_i, r_j)N \).

Assume \( r_i > r_L \) for \( i > L \), then the optimal weights for the ACV-KL control variate are

\[
e_{ACV-KL}(K, L) = -\left[ C \circ F(K, L) \right]^{-1} \left[ \text{diag} \left( F(K, L) \right) \circ c \right],
\]

and the estimator variance is

\[
\mathbb{V} \text{ar}\left[ \hat{Q}^2_{ACV-KL}(e_{ACV-KL}(K, L)) \right] = \mathbb{V} \text{ar}\left[ \hat{Q} \right] \left( 1 - R^2_{ACV-KL}(K, L) \right),
\]

where

\[
R^2_{ACV-KL}(K, L) = \left[ \text{diag} \left( F(K, L) \right) \circ \check{c} \right]^T \left[ C \circ F(K, L) \right]^{-1} \left[ \text{diag} \left( F(K, L) \right) \circ \check{c} \right],
\]

and \( F(K, L) \in \mathbb{R}^{M \times M} \) such that

\[
F_{ij}^{(K, L)} = \begin{cases} 
\min(r_i, r_j)^{-1} & \text{if } i, j \leq K \\
\min(r_i, r_j)^{-1} & \text{if } i, j > K \\
\frac{r_i - r_l}{r_i - r_L} & \text{if } L < i \leq K, \ j > K, \ i \neq j \\
\frac{r_j - r_L}{r_i - r_L} & \text{if } L < j \leq K, \ i > K \\
0 & \text{otherwise}
\end{cases}
\]

The diagonal elements are \( F_{ii}^{(K, L)} = \frac{r_i^{-1}}{r_i} \) if \( i \leq K \) and \( F_{ii}^{(K, L)} = \frac{r_i - r_L}{r_i} \) otherwise.

8. Numerical optimization procedure

For the algorithms presented so far, the total computational cost can be expressed as a function of the cost of obtaining an evaluation of \( \hat{Q} \) and \( Q_i \), \( w \) and \( w_i \) respectively. An optimal sample allocation strategy can be obtained for MLMC and MFMC in closed form. This is not the case for the ACV estimators. In this latter case we resort to a numerical optimization that seeks to maximize the variance reduction subject to a constraint on the total cost \( N \left( w + \sum_{i=1}^{M} w_i r_i \right) \). For ACV-KL we use a mixed-integer formulation where \( (K, L) \) are treated as categorical integer variable (addressed through enumeration). We mention here that the best choice of \( K \) and \( L \) is problem dependent; however, they can be estimated at negligible cost. Additional details on the optimization are reported in [21].

9. Parametric model problem

In order to illustrate the features of the ACV estimators compared to the other multifidelity estimators in literature we present a parametric model problem that enables us to explore different scenarios in term of correlation among models and numerical cost. We consider a two-dimensional problem (two random variables) and two low-fidelity models that serve as control variate

\[
Q = A \left( \cos \theta x^5 + \sin \theta y^5 \right), \\
Q_1 = A_1 \left( \cos \theta_1 x^3 + \sin \theta_1 y^3 \right), \\
Q_2 = A_2 \left( \cos \theta_2 x + \sin \theta_2 y \right),
\]

where \( x, y \sim \mathcal{U}(-1, 1) \) and all \( A_i \) and \( \theta_i \) coefficients are real. For this particular example we choose to set \( A = \sqrt{11}, A_1 = \sqrt{7} \) and \( A_2 = \sqrt{3} \) to obtain unitary variance for each model. This choice further reduce the degree-of-freedom in the problem parameterization since the correlation and covariance matrices are identical. Specifically, the analytic correlation/covariance matrix is given in Table [2]. To create enough variability in the scenarios we want to analyze, we fix the parameter for the high-fidelity and lowest fidelity model to \( \theta = \pi/2 \) and \( \theta_2 = \pi/6 \), respectively, and we let the middle fidelity model to vary, i.e. \( \theta_1 \) varies uniformly in the bounds \( \theta_2 < \theta_1 < \theta \). Each particular value of \( \theta_1 \) induces a
different correlation \( \rho_1 \) between \( Q_1 \) and \( Q \) and a different correlation \( \rho_{12} \) between \( Q_1 \) and \( Q_2 \), whereas the correlation \( \rho_2 \) between \( Q_2 \) and \( Q \) remains fixed. These correlations are reported for different settings of \( \theta_1 \) in Fig. 2a.

First we demonstrate that the variance reduction ratio of the control variate that considers both models, i.e., OCV, is larger than the one obtained using only a single control variate, i.e., OCV-1. This ratio of the estimator variance of OCV-1 and OCV is reported in Fig. 2b. In Fig. 2b we also report the variance reduction obtained by OCV and OCV-1 with respect to MC. The greatest gap between OCV and OCV-1 occurs when \( \theta_1 \) is approximately \( \pi/3 \) (the middle of its range).

The presence of a gap between OCV and OCV-1 indicates that the ACV estimators might outperform the recursive-based approaches as MLMC and MFMC because these can attain at most the variance reduction of OCV-1. Efficiently exploiting this gap ultimately depends on the possibility to obtain enough samples for the low-fidelity models and, in turn, it depends on the cost ratios between the different models. We explore several cost scenarios by assigning a relative cost of \( Q_1 \) to \( Q_2 \) for different \( \theta_1 \) and \( \theta_2 \). However, in this particular example ACV-KL is nearly identical to ACV-MF since large deviations in \( K, L \) cannot occur for only two low-fidelity models. In most of the scenarios, these two estimators also outperform OCV-1 and the relative benefits with respect to OCV-1 increases whenever the low-fidelity models become less expensive. For very inexpensive low-fidelity models \( (w = 1000) \) all the ACV estimators almost converge to the OCV estimator, whereas the recursive estimators only converge to OCV-1 as suggested by the theory.

### B. Leveraging Active Subspaces for multifidelity sampling

In the previous section we demonstrated that the performance of all the multifidelity sampling based algorithms are ultimately related to the correlation matrix among the models. Moreover, in many circumstances, in order to increase the efficiency of the lower fidelity models, the physics and the numerical approximations need to be much simplified. This might potentially lead to a lack of consistency with respect to the original model parametrization. We recently proposed \cite{17} to employ a dimension reduction strategy before deploying a multifidelity sampling estimator to seek a

---

**Table 2** Correlation/covariance matrix for problem given by Eq. 22 with \( A = \sqrt{11} \), \( A_1 = \sqrt{7} \), and \( A_2 = \sqrt{3} \).

<table>
<thead>
<tr>
<th>( Q )</th>
<th>( Q_1 )</th>
<th>( Q_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Q )</td>
<td>1</td>
<td>( AA_1/9 (\sin \theta \sin \theta_1 + \cos \theta \cos \theta_1) )</td>
</tr>
<tr>
<td>( Q_1 )</td>
<td>sym</td>
<td>1</td>
</tr>
<tr>
<td>( Q_2 )</td>
<td>sym</td>
<td>sym</td>
</tr>
</tbody>
</table>

---

Fig. 2 Correlation (left) and variance reduction gap (right) for the parametric problem of Eq. 22 with \( \theta = \pi/2 \), \( \theta_2 = \pi/6 \) and \( \theta_2 < \theta_1 < \theta \). Results from \cite{21}.
greater correlation and overcome the inconsistent models’ parametrization. As an example of this strategy we proposed to adopt an Active Subspace dimension reduction step to accelerate a control variate with a single low-fidelity model. The extension to several control variates and ACV estimators is straightforward therefore here we focus on a single control variate for illustration purpose. The idea is very simple and in order to present it we consider two models $Q$ and $Q_1$, the high- and low-fidelity model respectively. We assume $z$ and $z_1$, the input vector of random parameters for $Q$ and $Q_1$ respectively. Moreover we consider that the model can have a different dimensionality, i.e. $z \in \mathbb{R}^d$ and $z_1 \in \mathbb{R}^{d_1}$. For simplicity of exposure we consider the input parameters to be a vector of i.i.d. standard normal Gaussian variables, but we will relax this assumption later on. For each model, we perform an Active Subspace analysis and we identify the active directions independently. This step is accomplished by using only the pilot samples (which are usually generated in any case for a control variate estimator) and does not strictly require the knowledge of the gradient of a function because a low-order regression might be adopted (we tested several problems with only a linear regression and obtained satisfactory results). This step produces a vector of active variables $z_A$ and $z_{1,A}$ for both models

$$z_A = (W_A)^T z \quad z_{1,A} = (W_{1A})^T z_1,$$  \hfill (23)

where $W_A$ and $W_{1A}$ indicates the orthogonal matrices whose columns are the normalized eigenvectors for the high- and low-fidelity model respectively. These matrices are obtained through the standard Active Subspaces analysis, i.e. the solution of the eigenvalue problem for the expected value of the matrix generated by the outer product for the gradient (see [3] for more details). Given the linearity of the transformation, the orthonormality of the eigenvectors and the normal distribution of both $z$ and $z_1$, both variables $z_A$ and $z_{1,A}$ are standard normal variables. Therefore it is possible to identify a shared space between the models and a vector of standard normal variables $\mathbb{R}^{d_A} \ni z^* \sim N(0, I_{d_A})$ where $d_A^* = \max(d_A, d_{1A})$.

In a simple version of the algorithm the approximate control variate estimator can be built by evaluating a set of $N_{pilot}$ samples for $\{z^*[i]\}_{i=1}^{N_{pilot}}$ for which the models are evaluated after the rotation to the original coordinates by using Eq. (23). At this point the approximate control variate approach (ACV-1) can be deployed as usual just sampling the variable $z^*$ to create a new relationship between the models through the matrices $W_A$ and $W_{1A}$. As we demonstrated in [17] if the original distribution of the input variables of the low- and high-fidelity model is not a standard Gaussian, a
non-linear transformation can be adopted and the product between this function and the original model becomes the subject of both the Active Subspace analysis and the control variate.

In order to demonstrate the idea we consider a simple algebraic problem where the high- and low-fidelity model are defined by

\[
\begin{align*}
  f(x, y) &= \exp(0.7x + 0.3y) + 0.15 \sin(2\pi x), \\
  g(x, y) &= \exp(0.01x + 0.99y) + 0.15 \sin(3\pi y),
\end{align*}
\]

respectively. The two functions are reported in Fig. 4 for reference.

**Fig. 4** Test model functions.

We consider the input variable \(x, y \sim N(0, 1/3)\), therefore in this case only a scaling is applied to obtain standard normal variables. The correlation between the two models obtained by running 1000 realizations is approximately 0.05. This low correlation prevents to adopt a control variate if the problem is handled in its standard form. However, if we resort to the active variables (in this case both models have a single active variable, \(d^*_A = 1\)) the correlation can be much improved. In Fig. 5 we report both the functions \(f\) and \(g\) obtained by sampling the common active variable \(t\) and the resulting scatter plot obtained by mapping back the samples independently for each model fidelity. The correlation measured for this set of samples is approximately 0.9, therefore we can attain an increase in the performance for the OCV-1 estimator. In particular, as described in the previous section, if we assume the mean of the low-fidelity to be known the variance reduction ratio is equal to \(1 - \rho^2_1\). On the contrary, if the expected value for the low-fidelity model is unknown, the cost of a single fidelity realization needs to be considered in order to obtain the performance of the ACV-1 estimator.

**Fig. 5** Test model functions \(f\) and \(g\) represented along the active coordinate \(t\) with \(t = (W_A)^T z_1 = (W_{1A})^T z_1\) (Left). Scatter plot obtained by sampling \(t\) and mapping back independently each model (Right).

As an illustrative example, we assume a cost for the low-fidelity model \(g\) to be \(C^1 = 0.01C\), which is a representative cost ratio we often observe for practical applications. We arbitrary decided a computational budget of 300 HF simulations and compared four sampling estimators with equivalent cost, namely the MC estimator; the MC estimator obtained by
sampling along $t$ (MC-AS); a standard CV estimator with 100 HF realizations and 20000 LF simulations (MC-MF); a CV estimator obtained by sampling along $t$ and rotating the samples back to the original coordinates independently for each model (MC-MFAS). We note that for the two estimators adopting the AS, the active directions are estimated every time in order to make explicit the variance of this process in the overall estimator variance. We perform 1000 repetitions of each estimator and we report in Fig. 6 the normalized histograms for the estimated expect values (for this problem the exact solution is also known).

![Normalized histograms for 1000 realizations of several expected value estimators. The exact solution is reported as a black bold vertical line.](image)

As expected all the estimators are unbiased and close to a Gaussian distribution. The MC and MC-AS are almost identical because the variance is captured by the AS technique and there is no advantage between MC-AS and MC for the same number of samples (this is due to the independence of the performance of MC from the number of random dimensions for a problem with same variance). The MC-MF method has been used forcing it to compute enough low-fidelity models to obtain the same equivalent cost of the other methods, i.e. 300 high-fidelity computations. However, if the optimal samples allocation had been used, the algorithm would have allocated zero samples for the low-fidelity due to the low correlation between the models. The MC-MFAS instead recovers much sharply the expected value for this problem since the correlation is very high when the sampling is carried out with respect to the active variables of each model fidelity independently.

**IV. Surrogate-based approaches: highlights**

In this section, we describe two approaches for greedy adaptive refinement of stochastic expansions utilizing a multilevel, multifidelity, or multi-index construction. In the first approach, we describe a general construction that supports polynomial chaos expansion (PCE) methods, based on either projection or regression approaches, and stochastic collocation (SC) methods, built using either nodal or hierarchical interpolants. In the second section, we describe a fully integrated approach that employs adaptive sparse grid refinement across a union of random variable and model resolution dimensions.

**A. Greedy multilevel/multifidelity adaptation**

In the area of multilevel PCE approaches, we have previously explored bi-fidelity approaches [7,33], followed by approaches that shaped a sample profile based on convergence rate estimation for statistical estimator variance, on satisfying the restricted isometry property for compressed sensing, and on greedy multilevel refinement [6]. Of these, the greedy multilevel approach has been the most successful, and we describe the latest advancements here.

In a greedy approach, one competes a set of refinement candidates in order to select one or more best candidates, as indicated by maximal change in statistical quantities of interest (QoI) per unit of computational cost. This supports goal orientation in that the statistical QoI can involve expected values, variances, probabilities of failure, or other statistical goals. Each selected refinement is then used to generate additional candidates that continue the evolution of the approximation in the most fruitful directions.
Multilevel-multifidelity schemes add an additional dimension to greedy refinement, in that now we compete one or more candidates per level/fidelity and normalize their impact based both on the number of evaluations and on the relative cost of the evaluations for a particular level or fidelity. Level candidate generators of interest include:

- **Uniform refinement**: coarse-grained refinement with one expansion order / grid level candidate per model level
  - Tensor / sparse grids: projection PCE and nodal/hierarchical SC
  - Regression PCE: least squares / compressed sensing using a fixed sample ratio

- **Anisotropic refinement**: coarse-grained refinement with one expansion order / grid level candidate per model level
  - Tensor / sparse grids: projection PCE and nodal/hierarchical SC

- **Index-set-based refinement**: fine-grained refinement with multiple index set candidates per model level; exponential growth in size of candidate set with dimension.
  - Generalized sparse grids: projection PCE and nodal/hierarchical SC

- **Basis selection**: coarse-grained refinement with a few expansion order frontier advancements per model level
  - Regression PCE

1. **Model problem definition**

We present preliminary results for the steady state diffusion problem defined as follows:

\[- \frac{d}{dx} \left[ a(x, \xi) \frac{du}{dx}(x, \xi) \right] = 10, \quad (x, \xi) \in (0, 1) \times I_{\xi}, \quad (25)\]

where \(x\) is the spatial coordinate, \(\xi\) a vector of independent random input parameters and \(a(x, \xi)\) denotes the (random) diffusivity field. The following Dirichlet boundary conditions are also assumed

\[u(0, \xi) = 0, \quad u(1, \xi) = 0. \quad (26)\]

We are interested in quantifying the uncertainty in the solution \(u\) at specified spatial locations: \(\bar{x} = 0.05, 0.5, 0.95\). We represent the random diffusivity field \(a\) using the following expansion

\[a(x, \xi) = 1 + \sigma \sum_{k=1}^{d} \frac{1}{k^2 \pi^2} \cos(2\pi k x) \xi_k \quad (27)\]

where \(\xi_k \in [-1, 1], \ k = 1, \ldots, d\) are bounded random variables. The form of Eq. (27) is similar to that obtained from a Karhunen-Loève expansion. If the variables are i.i.d. uniform in \([-1,1]\) the the diffusivity satisfies satisfies the auxiliary properties

\[\mathbb{E}[a(x, \xi)] = 1 \quad \text{and} \quad 1 - \frac{\sigma}{6} < a(x, \xi) < 1 + \frac{\sigma}{6}. \quad (28)\]

For the particular test case in this section \(d = 9\), therefore \(I_{\xi} = [-1,1]^9\) and five uniform spatial discretizations of 4, 8, 16, 32 and 64 elements are selected.

2. **Greedy multilevel PCE with compressed sensing**

In Figure 7, we present results for greedy multilevel PCE with uniform refinement of candidate expansion orders using compressed sensing. The efficacy of the greedy approach is demonstrated by comparing its convergence envelope to several distinct sample allocations for the same number of levels (previous rate estimation results using compressed sensing). This envelope encompasses all of the other sample allocations. To make this more evident in Figure 8 we report the magnification of the standard deviation envelope.

In Table 3, we present the final sample profiles for this greedy multilevel compressed sensing approach. Here, the minimum evaluation requirement per level is 9 samples, resulting from initialization using a first-order basis with 10 terms and an enforced collocation ratio of 0.9 (note: a second-order candidate in nine dimensions contains 55 terms such that the collocation ratio requires 50 terms).

3. **Greedy multilevel PCE with generalized sparse grids**

The greedy approach can be also based on sparse grid refinement built using uniform refinement candidates (level-based) or generalized sparse grid candidates (index set-based). For these two cases, we report in Figure 9 the
Fig. 7 Convergence for greedy multilevel PCE based on compressed sensing. Test problem is steady state diffusion with nine random variables and one, two, or five discretization levels.

Fig. 8 Zoom in for the standard deviation error of Figure 7b)

<table>
<thead>
<tr>
<th>Conv Tol</th>
<th>(N_1)</th>
<th>(N_2)</th>
<th>(N_3)</th>
<th>(N_4)</th>
<th>(N_5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.e-1</td>
<td>198</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>1.e-2</td>
<td>644</td>
<td>198</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>1.e-3</td>
<td>1802</td>
<td>644</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>1.e-4</td>
<td>4505</td>
<td>1802</td>
<td>50</td>
<td>9</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 3 Final sample profiles for greedy multilevel compressed sensing applied to steady state diffusion (9 random variables, 5 discretization levels).
Fig. 9 Convergence for greedy multilevel PCE based on (generalized) sparse grids. Test problem is steady state diffusion with nine random variables and one or five discretization levels (solid and dashed lines, respectively).

<table>
<thead>
<tr>
<th>Conv Tol</th>
<th>$N_1$</th>
<th>$N_2$</th>
<th>$N_3$</th>
<th>$N_4$</th>
<th>$N_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.e-2</td>
<td>43</td>
<td>23</td>
<td>19</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>1.e-4</td>
<td>211</td>
<td>83</td>
<td>19</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>1.e-6</td>
<td>391</td>
<td>271</td>
<td>156</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>1.e-8</td>
<td>1359</td>
<td>743</td>
<td>327</td>
<td>59</td>
<td>19</td>
</tr>
<tr>
<td>1.e-10</td>
<td>3535</td>
<td>2311</td>
<td>1039</td>
<td>391</td>
<td>19</td>
</tr>
<tr>
<td>1.e-12</td>
<td>10319</td>
<td>5783</td>
<td>2783</td>
<td>1343</td>
<td>43</td>
</tr>
<tr>
<td>1.e-14</td>
<td>26655</td>
<td>14991</td>
<td>8063</td>
<td>3703</td>
<td>1535</td>
</tr>
</tbody>
</table>

Table 4 Final sample profiles for greedy multilevel refinement applied to steady state diffusion (9 random variables, 5 discretization levels).

convergence for mean and standard deviation, where it is evident that the greedy multilevel polynomial chaos (ML PCE) and multilevel hierarchical stochastic collocation (ML HSC) variants outperform their corresponding single-level benchmarks. Moreover, the ML HSC case shows initial promise relative to ML PCE since it can directly calculate hierarchical increments in statistics and avoid loss of precision due to subtractive cancellation. Finally, the greedy uniform refinement cases appear promising early on, but then stagnate for the highest grid levels that are currently affordable.

In Table 4 we present the final sample profiles for greedy multilevel PCE using generalized sparse grid refinement candidates. In nine dimensions for generalized sparse grids initiated from level-zero, all levels incur a minimum $2n + 1 = 19$ evaluation cost due to the initial set of level-one candidate index sets. The table shows how this minimum evaluation requirement is eventually replaced with an optimized multilevel sample profile, which also sheds light on the relative efficiency for low sample levels.

In Figure 10 we present all the results together in order to allow an easier comparison of the different methods. It is evident that the explicit nature of the sparse grid approaches allows for more precise convergence. The compressed sensing approaches, while supporting sample profiles at the lower end of the cost spectrum, are currently hampered in accuracy by solution of the large implicit systems that are allocated at the coarse level.
B. Adaptive Multi-index collocation

The multi-level surrogate based approaches we described in the previous section are useful when the level of fidelity is controlled by a single hyper-parameter specifying the level of discretization. In many applications, however, multiple hyper-parameters may control the model discretization, such as the mesh and time step sizes. In these situations multi-index stochastic collocation (MISC), can be used to deal with multivariate hierarchies with multiple refinement hyper-parameters [22].

In our recent work [25] we developed an adaptive extension of MISC and use this algorithm to reduce the cost of quantifying uncertainty in the performance of a supersonic jet engine nozzle. Our work utilizes a sparse grid approximation that spans both random dimensions and model resolution dimensions and uses generalized sparse grid refinement to select model evaluations from multiple candidates based on maximizing benefit per unit cost. In the following we will describe our algorithm and summarize our existing numerical results.

Given a sequence of models \( \{ \alpha \mid \alpha \leq (l_1^\alpha, \ldots, l_n^\alpha) \} \) of varying physical fidelity, which can be used to approximate the QoI \( f \), MISC approximates the QoI as a linear combination of low-resolution tensor-product interpolants

\[
\hat{f}_J(z) = \sum_{[\alpha, \beta] \in J} c_{\alpha, \beta} \hat{f}_{\alpha, \beta}(z). 
\]  

(29)

The tensor-product interpolants are themselves a weighted linear combination of tensor-product of univariate Lagrange polynomials

\[
\phi_{i,j}(z_i) = \prod_{k=1, k \neq j}^{m_{\beta_i}} \frac{z_i - z_i^{(k)}}{z_i^{(j)} - z_i^{(k)}}, \quad i \in [d],
\]  

(30)

defined on a set of univariate points \( z_i^{(j)}, \ j \in [m_{\beta_i}] \) (we use Clenshaw-Curtis points). Specifically the multivariate interpolant is given by

\[
\hat{f}_{\alpha, \beta}(z) = \sum_{j \leq \beta} \hat{f}_{\alpha}(z^{(j)}) \prod_{i \in [d]} \phi_{i,j_i}(z_i). 
\]  

(31)

The partial ordering \( j \leq \beta \) is true if all the component wise conditions are true.
The accuracy and efficiency of the MISC approximation is dependent on the index set $\mathcal{F}$. In Figure 11, we plot the tensor-product grids and corresponding interpolants that make up an isotropic MISC approximation of the function

$$f(z) = \cos\left(\frac{\pi}{4}(z_1 + \frac{1}{2})\right),$$

with $\mathcal{F}(l) = \{[\alpha, \beta] \mid (\max(0,l-1) \leq \|\alpha + \beta\|_1 \leq l + n_\alpha + n_\beta - 2\}$. The MISC approximation uses a sequence of models $\hat{f}_{\alpha l} = \cos(\frac{\pi}{4}(z_1 + \frac{1}{2} + \epsilon_{\alpha}))$ where $\epsilon_l > \epsilon_{l+1} \geq 0$ to approximate $f$. For each of these models MISC builds an interpolant $\hat{f}_{\alpha l, \beta}$ that converges to $\hat{f}_{\alpha l}$ as $\beta$ increases. Refinement of both $\alpha$ and $\beta$ is needed for the interpolants to converge to $f$. MISC achieves its efficiency by allocating resources to balance physical prediction bias and error of the stochastic interpolant. For the isotropic MISC, the number of evaluations allocated to an approximate model $f_{\alpha l}$ decreases as the level of fidelity $\alpha_l$ increases. For example, 5 samples are allocated to evaluating $\hat{f}_1$, 3 to $\hat{f}_2$ and 1 to $\hat{f}_3$. Despite the fact only one sample is allocated to evaluating $\hat{f}_3$, the MISC approximation, shown in the right plot of

![Figure 11](image)

Fig. 11 The tensor-product grids and interpolants of the level $l = 2$ isotropic MISC approximation $\hat{f}_{\alpha l}(z)$. The grids plotted with gray background are included in the MISC approximation, whereas the remaining grids are ignored. The numbers in the top-right of each plot are the coefficients $c_{\alpha, \beta} = (-1)^{l-\|\alpha + \beta\|_1} l!^{d-1} |\alpha + \beta|! l^{-|\alpha + \beta|}$ in (29).

The dash-dotted black lines represent the tensor-product interpolants $\hat{f}_{\alpha l, \beta}$. These interpolants converge to the lower fidelity models $\hat{f}_\alpha$ (dashed blue lines) as $\beta$ increases, and to the high-fidelity model $f$ (solid red line) as both $\alpha$ and $\beta$ are increased.

Figure 12 is more accurate than any of the tensor-product interpolants that make up the MISC approximation (gray). Understanding of this observation can be gained by viewing the discrepancies between models $\hat{f}_{\alpha l, \beta} - \hat{f}_l$ depicted in
the middle plot of Figure 12. The magnitude of these discrepancies decreases substantially as \(\alpha_1\) increases. Thus a constant approximation \(\hat{f}_{2,0}\) of \(f_2\) is sufficient to balance the deterministic and stochastic errors. Specifically the constant approximation is useful because it reduces the bias in the MISC interpolant by decreasing the physical discretization error. In contrast the approximation \(\hat{f}_{0,2}\) is useful as it captures the variation of the function \(f\) with respect to the variable \(z_1\) thereby reducing the stochastic error.

\[
\begin{align*}
\Delta E_{\alpha, \beta} &= \|\hat{f}_J \setminus \|\alpha, \beta\| \setminus \hat{f}_J\| \quad \Delta W_{\alpha, \beta} = |\text{Work}[\hat{f}_J \setminus \|\alpha, \beta\|] - \text{Work}[\hat{f}_J]| \quad \text{(33)}
\end{align*}
\]

respectively denote the difference between two successive MISC approximations and the work needed to update the MISC approximation, we choose grids using the following indicator to guide refinement

\[
\gamma_{\alpha, \beta} = \frac{1}{\Delta W_{\alpha, \beta}} \left( \kappa \Delta E_{\alpha, \beta}^\mu + (1 - \kappa) \Delta E_{\alpha, \beta}^{\sigma^2} \right), \quad \kappa \in [0, 1]. \quad \text{(34)}
\]

where

\[
\begin{align*}
\Delta E_{\alpha, \beta}^\mu &= \frac{1}{\|f_0\|} \mathbb{E}[\hat{f}_J \setminus \|\alpha, \beta\|](z) - \mathbb{E}[\hat{f}_J(z)] \quad \Delta E_{\alpha, \beta}^{\sigma^2} &= \frac{1}{\|f_0\|^2} \|\text{var}[\hat{f}_J \setminus \|\alpha, \beta\|](z) - \text{var}[\hat{f}_J(z)]\| \quad \text{(35)}
\end{align*}
\]

The index \([\alpha, \beta]\) with the largest refinement indicator is refined by adding all indices \([\alpha, \beta] + e_k, k \in [n_\alpha + n_\beta]\) that satisfy the following (downward-closed) admissibility criterion

\[
[\alpha, \beta] + e_k - e_j \in \mathcal{J} \text{ for } j \in [n_\alpha + n_\beta], l_k > 1. \quad \text{(36)}
\]

V. Multifidelity modeling with Bayesian networks

The surrogate methodologies described above require explicit ordering and hierarchical assumptions. However, the approximate control variate methodology has shown that enforcing such assumptions, which lead to MLMC and MFMC in the sampling case, creates inefficiency in the multifidelity models. In contrast to the pure sampling based approaches – the surrogate methodology can be much more brittle to inaccuracies in the modeling. For instance, it might require significantly more resources to accurately approximate discrepancies between two models which were
ordered hierarchically, but in practice are actually not so. For example, it is common that a computational simulation of a high-fidelity model may not be fully converged over the entire input space and therefore exhibit some noisy behavior in a quantity of interest or its derivative.

We propose a new methodology for leveraging multiple information sources that first specifies hidden variables for each model and then learns the structure between the latent variable models. These latent variables serve to explain any perceived or measured relationships between the information sources or models. Once such a network structure is learned, any additional analysis including uncertainty quantification, experimental design, and optimization can be made more efficient by exploiting the structure.

A. Networks of surrogates

Specifically, our new model for the relationships between physical models consists of a network of \( M \) latent variables \( \theta = \{ \theta_1, \ldots, \theta_M : \theta_i \in \Theta_i \subseteq \mathbb{R}^{p_i} \} \), each of which can be multivariate with \( p_i \) univariate parameters. These latent variables correspond to each of the models we seek to manage. For example, suppose we assume a linear relationship between models and latent parameters, then we define a set of basis functions for each model \( \phi = \{ \phi_1, \ldots, \phi_M : \phi_i : \mathcal{X}_i \rightarrow \mathbb{R}^{p_i \times d_i} \} \). Each model output becomes a random variable

\[
q_i = \tilde{Q}_i(x_i, \theta_i) = \phi_i^T(x_i)\theta_i = \sum_{k=1}^{p_i} \phi_{ik}(x_i)\theta_{ik}, \quad x_i \in \mathcal{X}_i.
\]  

Essentially, we introduce a new source of uncertainty due to incomplete knowledge of the latent variables. While we cannot reduce the original source of uncertainty (the input variables), we can reduce the uncertainty due to the latent variables as more data is obtained. The primary purpose of introducing the latent variables in the UQ context is to provide explanatory variables for observed correlations in model outputs. Furthermore, they allow us to consider relationships between models with different input/output spaces by developing relationships between their respective random variables. This notion of explanatory behavior is especially important in managing physical models where engineering and science intuition can often qualitatively specify the relationships between models. As an example, these latent variables could be the rotated variables obtained from active subspace approach described above and the basis functions could be the rotation matrices.

Next introduce a network relationship between each of the parameters \( \theta_i \) to obtain a network-informed high-fidelity model. A graph \( G = (V, E) \) is a tuple of vertices (or nodes) and edges. A Bayesian network (BN) structure is a directed acyclic graph whose nodes are random variables and edges are conditional probability distributions. The network encodes the set of conditional independence assumptions, called local Markov assumptions. The parents of random variable \( X_i \) in a graph are denoted as \( X_{pa(i)} \). The set of conditional probability distributions (CPDs), given the graph, is denoted as \( \mathbf{p} = \{ P(X_i | X_{pa(i)}) : X_i \in V \} \).

For simplicity, we consider linear-Gaussian CPDs so that

\[
\theta_i | \theta_{pa(i)} \sim \mathcal{N} (A_i | \theta_{pa(i)} \theta_{pa(i)} + b_i | \theta_{pa(i)}, \Gamma_i | \theta_{pa(i)})
\]

where \( c \sim \mathcal{N}(b, \Gamma) \) denotes that \( c \) is normally distributed with mean \( b \) and covariance \( \Gamma \). Then, the expectation of the high-fidelity model over the input uncertainty is a random variable depending on the latent parameters

\[
\mathbb{E}_X \left[ \tilde{Q}(x_1, \ldots, x_M, \theta_1, \ldots, \theta_M) \right] = \mathbb{E}_X \left[ \tilde{Q}(x_1, \theta_1) \right] = \mathbb{E}_X [\phi_1(x_1)^T] \theta_1,
\]

where the first equality uses the Markov property (implicitely indicated by Equation (37)) that indicates that given \( \theta_1 \) and \( x_1 \), \( f \) is independent of the other latent parameters. Together this procedure is shown in Flgure 13.

Learning the hyperparameters of a fixed graph is an important aspect before using it for prediction. Let \( \tau = \{ \tau_1, \ldots, \tau_M \} \) denote the set of hyperparameters defining the distribution of \( \theta_j | \theta_{pa(j)} \). For example for a network of \( M - 1 \) low-fidelity peer models we factor the joint distribution, given the hyperparameters, according to

\[
\theta_j | \theta_j \sim \mathcal{N}(b_j, \Gamma_j) \quad \text{for} \quad j = 2, \ldots, M
\]

\[
\theta_1 | \theta_2, \ldots, \theta_M, \tau \sim \mathcal{N} (b_{1|pa(1)} + \sum_{j=2}^{M} A_{1|j} \theta_j, \Gamma_{1|pa(1)})
\]
when the additional information is not known, but must be estimated from data. where the prior mean of the high fidelity model is \( \mu_2 \) and variance \( \sigma^2_2 \). Let the CPD between the second and first model be \( \theta_1 = \theta_2 + b + \xi \) where the disturbance \( \xi \) has zero mean and variance \( \sigma^2_2 \).

We can define three non-dimensional quantities \( \eta_1 \equiv \sigma^2_1 m_1 / \sigma^2_1, \eta_2 \equiv \sigma^2_2 m_2 / \sigma^2_2 \), and \( \eta = \sigma^2_2 / \sigma^2_2 \) that denote the ratio of uncertainty between HF prior and HF data; LF prior and LF data; and LF prior and HF prior, respectively. If we now denote \( \beta = \eta_1 (\eta^2 + \eta_2 + 1) \), then we can define a new weight \( \nu^2 = \frac{1}{\beta + 2 \eta_1 \eta_2} \). Using these non-dimensional quantities we can write the posterior mean as

\[
\hat{\mu}_1 = \nu_2 \bar{y} + (1 - \nu_2) (a \mu_2 + b) + \frac{\eta_2}{\beta} \nu_2 \left( \bar{y}^2 - \mu_2 \right)
\]

where the prior mean of the high fidelity model is \( \mu_1 = \mu_2 + b \). Note that the first two terms are an average of the likelihood and the marginal prior on the high fidelity model; however, the weights are now balanced by a combination of uncertainties of the LF/HF data/prior. The realizations of the LF data enter through a correction of this averaged quantity (third term). The strength of this correction depends on the strength of the correlation, which is provided by the scaling factor \( a \). Note the similarity of Equation (39) to the control variate formula (2). If we specify \( \mu_1 = \bar{y} \) by setting the hyperparameter \( b = \bar{y} - \mu_2 \), then we get a control-variate-like formula

\[
\hat{\mu}_1 = \bar{y} + \frac{\eta_2}{\beta} \nu_2 \left( \bar{y}^2 - \mu_2 \right)
\]

where instead of \( \mu_2 \) representing a known mean value, it now represents the prior belief of the mean value. Note that if the prior belief of the low-fidelity model is also set to the mean of the data, then, as expected, the posterior mean becomes the mean of the high-fidelity data. In some respect, this analysis justifies the use of control variates in the case when the additional information is not known, but must be estimated from data.

The posterior marginal variance can be decomposed into the two sources of prior uncertainty: the uncertainty of the second model and the uncertainty in the CPD

\[
\bar{c}_1 = \sigma^2_1 (1 - \nu_2) + \frac{\eta_2^2}{\beta} \nu_2
\]

B. Connections to sampling methods

We can show that sampling methods can be recovered by making some assumptions on the surrogate techniques. Consider if we use only a single constant basis function then we recover the surrogate \( \hat{f}(x) = \theta_1 \). Furthermore, consider the simple case of two models: a “high-fidelity” model \( \theta_1 \) and a “low-fidelity” model \( \theta_2 \). This case contains only one non-trivial graph \( \theta_2 \rightarrow \theta_1 \). The prior for the second model has mean \( \mu_2 \) and variance \( \sigma^2_2 \). Let the CPD between the second and first model be \( \theta_1 = \theta_2 + b + \xi \) where the disturbance \( \xi \) has zero mean and variance \( \sigma^2_2 \).

Statistics can then be extracted from this high-fidelity model. For instance here, we have \( \bar{h} = E [Q] \). Furthermore, note that the predictions are now probabilistic over the latent variable space.
where we the posterior variance is a weighted average of two terms. The first term is the posterior variance of a single, high-fidelity model, which decays as $\nu^2 \to 1$ (uncertainty in data decreases). The second term reflects the uncertainty in the low-fidelity model that goes to zero as $\beta \to \infty$ denoting that the data is extremely informative. Note that the posterior variance decay does not actually depend on the realization of the data, a common characteristic of linear-Gaussian models.

We present initial results using this approach for the SEQUOIA problem in Section VI.

VI. Numerical results

In this section we present numerical results regarding the aero-thermo-structural analysis of the SEQUOIA jet engine nozzle presented in Sec. II. For each algorithm, we tested scenarios that are based on a different selections of models among the ones available.

A. Sampling-based approaches

1. Approximate Control Variate

We demonstrated in Section III.A that all the recursive sampling based schemes cannot achieve a variance reduction higher that the one corresponding to OCV-1. We proposed the ACV framework as a possible approach to overcome this issue and obtain converging estimators, i.e. multifidelity sampling estimators that converge to OCV in the limit of an infinite number of low-fidelity evaluations. As we demonstrated by means of the parametric model problem in Sec. III.A the efficiency of the ACV estimators ultimately depends on the presence of a separation between the performance, i.e. variance reduction, of OCV-1 and OCV. If this gap exists and the cost ratio is adequate we demonstrated that this framework outperforms the recursive schemes. In this section we analyze the nozzle problem to understand if a similar gap between OCV and OCV-1 exists for several QoIs, namely the thrust of the nozzle, a temperature failure criterion in the inner load layer, and a strain failure criterion in the thermal layer. We focus on an array of seven models ranging from the 1D ideal fluid solver up to a RANS 3D model (each with a different spatial resolution for the FEM analyses). The full range of models and their costs are reported in Table 5. We note that the meshes for both the CFD solver and the FEM solvers are denoted as COARSE, MEDIUM and FINE, however, they vary for each fidelity, for instance meshes COARSE for Euler 2D, Euler 3D and RANS 3D are all distinct.

<table>
<thead>
<tr>
<th>CFD</th>
<th>FEM (Thermal/Structural)</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>1D</td>
<td>COARSE</td>
<td>2.63e-04</td>
</tr>
<tr>
<td>Euler 2D COARSE</td>
<td>COARSE (axisymmetric)</td>
<td>9.69e-04</td>
</tr>
<tr>
<td>Euler 2D MEDIUM</td>
<td>MEDIUM (axisymmetric)</td>
<td>3.18e-03</td>
</tr>
<tr>
<td>Euler 2D FINE</td>
<td>FINE (axisymmetric)</td>
<td>9.05e-03</td>
</tr>
<tr>
<td>Euler 3D COARSE</td>
<td>COARSE</td>
<td>1.16e-02</td>
</tr>
<tr>
<td>Euler 3D MEDIUM</td>
<td>MEDIUM</td>
<td>3.58e-02</td>
</tr>
<tr>
<td>RANS 3D COARSE</td>
<td>COARSE</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 5 Relative computational cost for several model fidelities for the nozzle problem. All the cost are normalized with respect to the 3D RANS solver.

The performance of all the estimators, given the ACV framework, can be estimated by knowing the covariance matrix $C$ (whose first column is also the vector $c$) and the correlation matrix which enable us to compute the vector $\tilde{c}$. The correlation matrix is simpler to interpret, therefore we report the matrix for Thrust (Table 6), the temperature failure criterion in the load layer (Table 7) and the strain failure criterion in the thermal layer (Table 8). All the values here have been computed by means of 1000 random samples.

A general observation is that the correlations among all the models are very high for all the three QoIs we considered. We report in Table 6 the variance reduction that could be potentially achieved by the OCV estimator (or potentially a convergent ACV estimator). The variance reduction is expressed as the term $1 - R^2$, i.e. for a generic estimator the absolute variance is given by the Monte Carlo estimator variance with the same number of high-fidelity evaluations times the variance reduction term $1 - R^2$ (see Section III.A for more details). We also compare the ratio of the variance reduction between OCV-1 and OCV to make explicit the performance increase that could be attained by using the ACV
<table>
<thead>
<tr>
<th></th>
<th>RANS 3D C</th>
<th>Eul 3D M</th>
<th>Eul 3D C</th>
<th>Eul 2D F</th>
<th>Eul 2D M</th>
<th>Eul 2D C</th>
<th>1D</th>
</tr>
</thead>
<tbody>
<tr>
<td>RANS 3D C</td>
<td>1.00000</td>
<td>0.97446</td>
<td>0.96878</td>
<td>0.88782</td>
<td>0.88841</td>
<td>0.89041</td>
<td>0.97419</td>
</tr>
<tr>
<td>Eul 3D M</td>
<td>0.97446</td>
<td>1.00000</td>
<td>0.99950</td>
<td>0.83390</td>
<td>0.83438</td>
<td>0.83684</td>
<td>0.99426</td>
</tr>
<tr>
<td>Eul 3D C</td>
<td>0.96878</td>
<td>0.99950</td>
<td>1.00000</td>
<td>0.81976</td>
<td>0.82027</td>
<td>0.82283</td>
<td>0.99376</td>
</tr>
<tr>
<td>Eul 2D F</td>
<td>0.88782</td>
<td>0.83390</td>
<td>0.81976</td>
<td>1.00000</td>
<td>0.99990</td>
<td>0.99988</td>
<td>0.85284</td>
</tr>
<tr>
<td>Eul 2D M</td>
<td>0.88841</td>
<td>0.83438</td>
<td>0.82027</td>
<td>0.99990</td>
<td>1.00000</td>
<td>0.99997</td>
<td>0.85351</td>
</tr>
<tr>
<td>Eul 2D C</td>
<td>0.89041</td>
<td>0.83684</td>
<td>0.82283</td>
<td>0.99990</td>
<td>0.99997</td>
<td>1.00000</td>
<td>0.85595</td>
</tr>
<tr>
<td>1D</td>
<td>0.97419</td>
<td>0.99426</td>
<td>0.99376</td>
<td>0.85284</td>
<td>0.85351</td>
<td>0.85595</td>
<td>1.00000</td>
</tr>
</tbody>
</table>

Table 6  Correlation matrix for the nozzle problem: Thrust.

<table>
<thead>
<tr>
<th></th>
<th>RANS 3D C</th>
<th>Eul 3D M</th>
<th>Eul 3D C</th>
<th>Eul 2D F</th>
<th>Eul 2D M</th>
<th>Eul 2D C</th>
<th>1D</th>
</tr>
</thead>
<tbody>
<tr>
<td>RANS 3D C</td>
<td>1.00000</td>
<td>0.99621</td>
<td>0.99622</td>
<td>0.99642</td>
<td>0.99641</td>
<td>0.99642</td>
<td>0.99779</td>
</tr>
<tr>
<td>Eul 3D M</td>
<td>0.99621</td>
<td>1.00000</td>
<td>0.99999</td>
<td>0.99992</td>
<td>0.99992</td>
<td>0.99992</td>
<td>0.99854</td>
</tr>
<tr>
<td>Eul 3D C</td>
<td>0.99622</td>
<td>0.99999</td>
<td>1.00000</td>
<td>0.99992</td>
<td>0.99992</td>
<td>0.99992</td>
<td>0.99855</td>
</tr>
<tr>
<td>Eul 2D F</td>
<td>0.99642</td>
<td>0.99991</td>
<td>0.99991</td>
<td>1.00000</td>
<td>0.99999</td>
<td>0.99999</td>
<td>0.99874</td>
</tr>
<tr>
<td>Eul 2D M</td>
<td>0.99641</td>
<td>0.99992</td>
<td>0.99992</td>
<td>0.99999</td>
<td>1.00000</td>
<td>0.99999</td>
<td>0.99874</td>
</tr>
<tr>
<td>Eul 2D C</td>
<td>0.99642</td>
<td>0.99992</td>
<td>0.99992</td>
<td>0.99999</td>
<td>1.00000</td>
<td>0.99999</td>
<td>0.99875</td>
</tr>
<tr>
<td>1D</td>
<td>0.99779</td>
<td>0.99854</td>
<td>0.99855</td>
<td>0.99874</td>
<td>0.99874</td>
<td>0.99875</td>
<td>1.00000</td>
</tr>
</tbody>
</table>

Table 7  Correlation matrix for the nozzle problem: Temperature load layer failure.

<table>
<thead>
<tr>
<th></th>
<th>RANS 3D C</th>
<th>Eul 3D M</th>
<th>Eul 3D C</th>
<th>Eul 2D F</th>
<th>Eul 2D M</th>
<th>Eul 2D C</th>
<th>1D</th>
</tr>
</thead>
<tbody>
<tr>
<td>RANS 3D C</td>
<td>1.00000</td>
<td>0.99412</td>
<td>0.99628</td>
<td>0.99254</td>
<td>0.99382</td>
<td>0.99599</td>
<td>0.99931</td>
</tr>
<tr>
<td>Eul 3D M</td>
<td>0.99412</td>
<td>1.00000</td>
<td>0.99911</td>
<td>0.99965</td>
<td>0.99994</td>
<td>0.99911</td>
<td>0.99211</td>
</tr>
<tr>
<td>Eul 3D C</td>
<td>0.99628</td>
<td>0.99911</td>
<td>1.00000</td>
<td>0.99794</td>
<td>0.99898</td>
<td>0.99990</td>
<td>0.99436</td>
</tr>
<tr>
<td>Eul 2D F</td>
<td>0.99254</td>
<td>0.99965</td>
<td>0.99794</td>
<td>1.00000</td>
<td>0.99978</td>
<td>0.99817</td>
<td>0.99058</td>
</tr>
<tr>
<td>Eul 2D M</td>
<td>0.99382</td>
<td>0.99994</td>
<td>0.99898</td>
<td>0.99978</td>
<td>1.00000</td>
<td>0.99913</td>
<td>0.99186</td>
</tr>
<tr>
<td>Eul 2D C</td>
<td>0.99599</td>
<td>0.99911</td>
<td>0.99990</td>
<td>0.99817</td>
<td>0.99913</td>
<td>1.00000</td>
<td>0.99417</td>
</tr>
<tr>
<td>1D</td>
<td>0.99931</td>
<td>0.99211</td>
<td>0.99436</td>
<td>0.99058</td>
<td>0.99186</td>
<td>0.99417</td>
<td>1.00000</td>
</tr>
</tbody>
</table>

Table 8  Correlation matrix for the nozzle problem: Strain thermal layer failure.
Variance reduction

<table>
<thead>
<tr>
<th>QoI</th>
<th>OCV</th>
<th>OCV-1</th>
<th>Ratio OCV/OCV-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thrust</td>
<td>0.020595</td>
<td>0.050432</td>
<td>0.41</td>
</tr>
<tr>
<td>Temperature failure criterion</td>
<td>0.0043612</td>
<td>0.0075662</td>
<td>0.58</td>
</tr>
<tr>
<td>Strain failure criterion</td>
<td>6.2981e-04</td>
<td>0.011720</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Table 9  Comparison between the performance of OCV and OCV-1 for the nozzle problem and three different QoIs.

framework (in the limit of infinite low-fidelity data).

We note that for all the QoIs considered a separation exists between OCV-1 and OCV. The variance reduction of OCV-1 with respect to the plain MC estimator is greater than one order of magnitude for all QoIs and an additional performance increase of approximately 40% could be obtained for Thrust and the temperature failure criterion by using the OCV estimator. Moreover, for the strain failure criterion the performance increase between OCV and OCV-1 accounts for approximately 95%. We note that these values correspond to the maximum attainable variance reduction under the assumption that the statistics of the low-fidelity models are known (or equivalently the cost of the low-fidelity model is zero). Therefore, the effective performance of both ACV and ACV-1 with respect to a plain Monte Carlo estimator are inferior compared to their OCV and OCV-1 counterparts.

2. Active Subspaces for Control Variate

In this section we present some preliminary results regarding the use of the Active Subspaces to accelerate the convergence of an ACV-1 estimator for the nozzle problem. We consider only the ACV-1 estimator for simplicity, but the extension to multiple model is straightforward. The high-fidelity model in this case uses 3D Euler for the CFD solver and a COARSE mesh for the FEM thermal and structural solver. Two low-fidelity models are separately considered; both models are based on the same numerical approximation, namely 2D Euler solutions for the CFD and axi-symmetric FEM mesh. However, albeit the nominal parametrization of the two low-fidelity model is the same, one model uses a consistent parametrization with the 3D geometry, i.e. the preservation of the nozzle internal area, section-by-section, is enforced whereas for the other one the axi-symmetric mesh is defined starting from the mid-plane section of the 3D geometry. Henceforth, we refer to these models as consistent and inconsistent parametrization, respectively. The consistency of the parameterization is enforced underneath by the aero-thermo-structural solver, therefore we consider it as a distinct model with a slightly higher computational cost due to this pre-processing stage. The computational cost for the three models used in this section is reported in Table 10.

<table>
<thead>
<tr>
<th>CFD</th>
<th>FEM (Thermal/Structural)</th>
<th>Parameterization</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D Euler COARSE</td>
<td>COARSE</td>
<td>Consistent</td>
<td>1.00</td>
</tr>
<tr>
<td>2D Euler COARSE</td>
<td>COARSE (axisymmetric)</td>
<td>Consistent</td>
<td>0.201</td>
</tr>
<tr>
<td>2D Euler COARSE</td>
<td>COARSE (axisymmetric)</td>
<td>Inconsistent</td>
<td>0.135</td>
</tr>
</tbody>
</table>

Table 10  Relative computational cost for the models used for the Active Subspace tests for the nozzle problem. All the costs are normalized with respect to the 3D Euler COARSE solver.

The numerical test has been configured to include 40 uncertain parameters that describe operative conditions and material properties as described in [12] and 96 design parameters that describe control points coordinates for the nozzle geometry. Moreover, the set of random parameters is constituted by a combination of uniform and log-normal distributed variables. In order to explore an array of possible scenarios we designed three distinct numerical tests to assess the potentiality of the combination of Active Subspaces with ACV-1. All the scenarios are based on datasets that we generated according to the following situation.

1. High- and low-fidelity model with inconsistent parametrization evaluated for the same set of samples. The inconsistency between the model is hidden in the pre-processing step of the deterministic solver;
2. High- and low-fidelity model with consistent parametrization evaluated at an independent set of samples. The inconsistency between the models arises from the lack of common evaluations between the high- and low-fidelity datasets (this might occur for instance if multifidelity estimators need to be built starting from legacy data);
Fig. 14  QoIs w.r.t. the active variable for the nozzle problem in the case of inconsistent parameterization for both the original data and the PCE regression with respect to the active variable (Scenario 1).

3 High- and low-fidelity model with inconsistent parametrization evaluated for the same set of nominal samples. However, in this case, we also consider the design parameters of the high-fidelity model to be uniform variable with a range of ±2.5% with respect to their nominal values. The inconsistency in this case stems from the different dimensionality of the two input spaces (136 versus 40 for the high- and low-fidelity, respectively).

We note here that for all the tests in this section we used a linear regression for the evaluation of the gradients required in the Active Subspaces analysis since they were not readily available from the multiphysics solvers. Also, an ordinary least square method is used to compute the coefficients of the polynomial chaos expansion (PCE) with respect to the active variables (in this case a single active variable is considered).

For the first scenario we verified that the active subspaces exist and that the polynomial regression is able to reduce the noise introduced by the inactive variables after the projection. This test is important to verify that the introduction of the linear mapping between the uniform and log-normal variables toward the standard Gaussian variables used in our approach is not hampering the correlation. The four QoIs reported as a function of the active variable are reported in Fig. [14] We note that all the QoIs are well approximated by a monodimensional Active Subspace.

For the second scenario we perform the same Active Subspace analysis to demonstrate that the performance of the consistent parametrization model are slightly improved with respect to the model with inconsistent parameterization, Fig.[15] The improved capabilities of the model with consistent parametrization are evident especially for the thrust and the thermal stresses. Moreover, even in this case a single active variable is enough to obtain a regular PCE regression.

For the third scenario in addition to the inconsistent parameterization between the models we also added a variability of ±2.5% around each design variable. The high-fidelity model has now a total of 136 uncertain parameters. We omit for brevity the representation of the active subspaces for the four QoIs, but we report their scatter plots in Fig. [16] We can note that the additional 96 uncertainty parameters in the design variable add only a limited amount of variability to the scatter plots especially for the thermal and mechanical stresses. However, their impact is more evident for the mass and thrust when considering the original variables. By using the PCE regression along the active variable we are able to
Fig. 15 QoIs w.r.t. the active variable for the nozzle problem in the case of consistent parameterization for both the original data and the PCE regression with respect to the active variable (Scenario 2).
compute, *a posteriori* and without additional realizations of the low-fidelity model, low-fidelity responses for a set of active variables consistent with the high-fidelity samples. Therefore, after this projection, the scatter plots reveal more clearly the correlation between the models.

In Table 11 we report the variation of the correlation squared between the samples taken in the physical variables and its estimated value from the PCE regression along the active variable. This value enable us to estimate the reduction in the estimator standard deviation w.r.t. to plain MC approach for both the standard OCV-1 with the sampling performed in the original/physical variables and OCV-1 (AS) with the sampling carried out in the active variables.

<table>
<thead>
<tr>
<th>QoIs</th>
<th>$\rho^2$</th>
<th>$\rho_{AS}^2$</th>
<th>MC</th>
<th>ACV-1</th>
<th>ACV-1 (AS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass</td>
<td>0.822</td>
<td>0.999</td>
<td>1</td>
<td>0.178</td>
<td>0.001</td>
</tr>
<tr>
<td>Thrust</td>
<td>0.956</td>
<td>0.998</td>
<td>1</td>
<td>0.044</td>
<td>0.002</td>
</tr>
<tr>
<td>Thermal Stress</td>
<td>0.982</td>
<td>0.998</td>
<td>1</td>
<td>0.018</td>
<td>0.002</td>
</tr>
<tr>
<td>Mechanical Stress</td>
<td>0.985</td>
<td>0.986</td>
<td>1</td>
<td>0.015</td>
<td>0.014</td>
</tr>
</tbody>
</table>

Table 11 (Estimated) Standard Deviation for OCV-1 and OCV-1 (AS) (normalized w.r.t. MC) for the Sequoia application problem in the case of inconsistent parameterization and uncertain design input in HF (Scenario 3).

The results reported in Table 11 are promising and predict an increase of performance of at least one order of magnitude for all the QoIs but the mechanical stresses. However, we remark that these results are obtained by resorting to a PCE regression along the active direction which is used to evaluate the correlation between the models. Albeit we can visually verify that the regression accurately capture the model trend, more rigorous tests need to be performed to confirm these results for this specific application.
B. Surrogate-based approaches

1. Multifidelity Multi-Index Collocation

We report here some results regarding the performance of the adaptive multi-index Stochastic Collocation (AMISC) algorithm for four QoIs, namely the mass, thrust, the thermal and mechanical stress failure criteria.

For this problem, we are interested in approximating four QoIs, so we must adjust slightly the refinement indicator introduced in Section IV.B. More precisely, we use a worst case error indicator that refines the multi-index set in the AMISC approximation according to the largest error across the set of QoIs. Letting \( \gamma_{\alpha, \beta}^k \) denote the error indicator Eqn. (34), we use

\[
\gamma_{\alpha, \beta} = \max_{k \in \{4\}} \gamma_{\alpha, \beta}^k
\]

across the set of \( k \) QoIs. Moreover we select \( k = 1/2 \) to obtain a balanced evaluation of both mean and variance.

We first compare the performance of the AMISC algorithm with a single-level adaptive sparse grid in Figure 17 (only thrust and temperature layer failure are reported for brevity). Work is measured relative to the cost of the highest fidelity. We terminate refinement of each approximation after the cost of 3400 high-fidelity samples is reached. We observe that AMISC can approximate each QoI with the same accuracy as a single fidelity approximation at only a fraction of the cost. The AMISC approximation at some points on the convergence curves is an order of magnitude less expensive. This approaches the limit of potential acceleration, as there is only a factor of 20 difference in cost between the cheapest and most expensive models.

![Comparison of the convergence of single-fidelity and multi-fidelity approximations of thrust and temperature layer failure. The \([0, 1, 2], [0, 1, 2, 3, 4] \) specify the physical discretization indices used for the AMISC approximation and \([2], [4] \) the physical discretization indices used for the single high-fidelity sparse grid approximation.](image)

Note that the convergence curves of the approximation of thrust (top right) is noisy in the right tail. This is related to the tolerance we use for the non-linear CFD solver; in particular, the error is saturating at the error dictated by this tolerance.

For the thermal-layer failure criteria we observe that AMISC is cheaper to construct for accuracies up to \( 10^{-4} \), however the single-fidelity approximation is cheaper to construct for higher accuracy. This situation is caused by the error of the two coarsest structural meshes to be smaller than the deterministic prediction error of those models. Therefore, only the high-fidelity structural model can reduce the error. The downward closed assumption for the sparse grid index set implies that the coarsest structural meshes are always evaluated before the highest fidelity structural mesh even if they are not effective. Therefore an overhead cost is associated to these low-fidelity evaluations and the single-level sparse grid approximation of the load-layer failure criteria will eventually become more efficient than the AMISC approximation.

In Figure 18 we plot the number of model evaluations allocated to each model discretization. The AMISC algorithm determined that the CFD and structural model are both large sources of deterministic prediction error. This is reflected...
by the large number of samples allocated to the most refined CFD and structural models. However, it is not necessary to refine both the CFD and structural meshes simultaneously. This is indicated by the fact only one sample is assigned to any model indexed by $(\alpha_1, \alpha_2)$ with both $\alpha_1$ and $\alpha_2$ non zero. The error in the CFD model output, which is used as input to the structural model, does not significantly affect the prediction of the structural quantities of interest – the load layer and temperature layer failure criteria.

Fig. 18  The number of simulation and fraction of total work assigned to each nozzle model.

The advantage of surrogate-based approach with respect to sampling-based methods is that once the surrogate has been constructed it can be used to compute additional metrics of interest. For instance, for this problem we used AMISC to compute the sensitivity indices, namely the Sobol indices, that evaluate the contribution of each parameter or group of parameters to the overall variance. In Figure 19 we plot the dominant sources of uncertainty in the nozzle predictions. We only report the parameter combinations that contribute to a portion of the first 99.9% of variance of at least one nozzle QoI. It is interesting to note that, for each QoI, one parameter acting independently of all other parameters contributes over 90% of the total variance. Moreover almost all contributions to 99.9% of the variance of each QoI are caused by individual parameters. Only two second-order interactions are significant. However, the convergence curves suggest that resolving multivariate interactions becomes important as we drive the surrogate error to tight tolerances.

Fig. 19  Sobol indices that contribute a portion of the first 99.9% of the variance of at least one nozzle QoI. The labels of the x-axis represent the parameter of the Sobol index. The maximum value of any Sobol index is one. We zoom in on the y-scale to better highlight the smaller Sobol index values, and report values greater than the 0.1 plot scale in the gray boxes.

The most sensitive parameter identified for each QoI matches physical intuition. The density of the CMC heat layer $\rho_{\text{CMC}}$ is the largest contribution to the uncertainty in the mass of the nozzle, consistent with the observations that the
minimum density of this layer is an order of magnitude larger the the maximum density of the other layers and the volume of this layer is greater than any other layer. The inlet stagnation pressure $P_i$ has the greatest influence on thrust. The heat transfer coefficient $\beta$ which transfers heat from the nozzle to the environment has the greatest impact on the Load layer temperature failure ratio, which is the ratio of the temperature in the load layer to the maximum allowable temperature. When $\beta$ is decreased, less heat is able to escape the nozzle and thus the temperature in the nozzle increases. Finally, the failure strain in the CMC heat layer $\epsilon_{CMC}^f$ has the greatest impact on the failure of the thermal failure.

2. Latent variables

Next we apply the Bayesian network approach to a SEQUOIA model consisting with 40 uncertain inputs and three mesh resolutions. Our goal here is to demonstrate the difference in posteriors due to differing graphs. The first graph is the peer graph where we have $\theta_1 = a_1 \theta_2 + a_2 \theta_3 + \xi$. The second graph is a recursive graph $\theta_1 = a_1 \theta_2 + \xi_1$ and $\theta_2 = a_2 \theta_3 + \xi_2$. The peer model indicates that the high fidelity model is a combination of the two low fidelity models. The recursive model is a standard recursive co-kriging (see e.g., [29]) approach in parametric form with a single, constant basis function. For each graph we choose the statics of $\xi$, $\xi_1$, and $\xi_2$ to ensure that the marginal priors of all of the parameters $\theta_1$, $\theta_2$, $\theta_3$ are identical. Only the joint marginals differ amongst the graph. The posteriors over the mean of two quantities of interest, recall this is a Bayesian procedure and we have more than a point estimate, are shown in Figure 20.

VII. Concluding remarks

We have provided an overview of recent activities in the area of multilevel-multifidelity strategies for forward propagation of uncertainties. Our algorithms span both sampling-based and surrogate-based approaches, and are evolving from methods that assume hierarchical (recursive) modeling relationships to ones that are not bound by these assumptions and can learn model topologies from the simulation data.

For numerical investigations of these algorithms, we have described results both from model problems and from the deployment of these approaches to a realistic engineering application, namely an aero-thermo-structural analysis of a jet engine nozzle. The nozzle case exhibits significant computational cost for each realization, making the use of traditional single-fidelity UQ techniques extremely computationally demanding. Through the exploitation of multiple fidelities and resolutions, accurate resolution of the statistical quantities of interest becomes feasible.

Acknowledgments

The authors thank Dr. Rick Fenrich, Dr. Victorien Menier and Prof. Juan Alonso at Stanford University for their help in setting the nozzle computations. The authors are also greatly indebted with Dr. Jeff Hokanson and Prof. Paul Constantine for sharing the dataset used to generate the results reported in Sec. VI.A.1. Support for this research was provided by the DARPA program EQUiPS. Sandia National Laboratories is a multimission
laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy’s National Nuclear Security Administration under contract DE-NA0003525. The views expressed in the article do not necessarily represent the views of the U.S. Department Of Energy or the United States Government.

References


