Uncertainty Quantification in Random Heterogeneous Media

Nicholas Zabaras

Materials Process Design and Control Laboratory
Sibley School of Mechanical and Aerospace Engineering
101 Frank H. T. Rhodes Hall
Cornell University
Ithaca, NY 14853-3801

Email: zabaras@cornell.edu
URL: http://mpdc.mae.cornell.edu/
Motivation

All physical systems have inherent associated randomness

SOURCES OF UNCERTAINTIES
- Multiscale nature inherently statistical
- Uncertainties in process conditions
- Material heterogeneity
- Model formulation – approximations, assumptions

Why uncertainty modeling?
Predictive Modeling
Assess product and process reliability
Estimate confidence level in model predictions
Identify relative sources of randomness
Provide robust design solutions
Motivation of HDMR

- Conventional and adaptive collocation methods are not suitable for high-dimensional problems due to their weakly dependence on the dimensionality (logarithmic) in the error estimate.

- Although ASGC can alleviate this problem to some extent, its performance depends on the regularity of the problem and the method is only effective when some random dimensions are more important than others.

- In e.g. random heterogeneous media we often deal with a very small correlation length and this results in a rather high-dimensional stochastic space with nearly the same weights along each dimension. In this case, all the previously developed stochastic methods are obviously not applicable.

- These modeling issues for high-dimensional stochastic problems motivate the use of the High Dimensional Model Representation (HDMR) technique.
The finite-dimensional noise assumption

- The solution of the problem of interest is described by a set of random variables, i.e.
\[ u(x, \omega) = u(x, Y_1(\omega), \ldots, Y_N(\omega)) \]

- The SPDE problem is stated as: Find the stochastic function \( u : \Gamma \times D \rightarrow \mathbb{R} \) such that
\[
L(x, Y; u) = f(x, Y), \quad \forall x \in D \times \Gamma \\
B(x, Y; u) = g(x, Y), \quad \forall x \in \partial D \times \Gamma
\]

- In this work, we assume that \( \{Y_i(\omega)\}_{i=1}^{N} \) are independent random variables with probability density function \( \rho_i \). Let \( \Gamma_i \) be the image of \( Y_i \). Then
\[
\rho(Y) = \prod_{i=1}^{N} \rho_i(Y_i), \quad \forall Y \in \Gamma
\]

is the joint probability density of \( Y = (Y_1, \ldots, Y_N) \) with support
\[
\Gamma \equiv \prod_{i=1}^{N} \Gamma_i \in \mathbb{R}^N
\]
Stochastic collocation based framework

$$f : \mathbb{R}^N \rightarrow \mathbb{R}$$

Need to represent this function

Sample the function at a finite set of points $$\Omega_M = \{Y^{(i)}\}_{i=1}^M$$

Use polynomials (Lagrange polynomials) to get an approximate representation

$$\tilde{f}(Y) = \sum_{i=1}^M f(Y^{(i)})a_i(Y^{(i)})$$

Function value at any point is simply $$\tilde{f}(\xi)$$

Stochastic function in 2 dimensions

Spatial domain is approximated using a FEM discretization.

Stochastic domain is approximated using multidimensional interpolating functions
In the context of incorporating adaptivity, we use the Newton-Cotes grid with equidistant support nodes and the linear hat function as the univariate nodal basis.

\[ m_i = \begin{cases} 
1, & \text{if } i = 1 \\
2^{i-1} + 1, & \text{if } i > 1 
\end{cases} \]

\[ Y_j^i = \begin{cases} 
\frac{j - 1}{m_i - 1}, & \text{for } j = 1, \ldots, m_i, \text{ if } m_i > 1, \\
0.5, & \text{for } j = 1, \text{ if } m_i = 1. 
\end{cases} \]

In this manner, one ensures a local support and that discontinuities in the stochastic space can be resolved. The piecewise linear basis functions is defined as

\[ a_1^i = 1, \text{ for } i = 1, \text{ and } \]

\[ a_j^i = \begin{cases} 
1 - (m_i - 1) \cdot |Y - Y_j^i|, & \text{if } |Y - Y_j^i| < 1/(m_i - 1), \\
0, & \text{otherwise,} 
\end{cases} \]

for \( i > 1 \) and \( j = 1, \ldots, m_i \).
Conventional sparse grid collocation (CSGC)

- Denote the one dimensional interpolation formula as
  \[ U^i(f) = \sum_{j=1}^{k_i} f(Y^i_j) \cdot a^i_j \]

- In higher dimensions, a simple case is the tensor product formula
  \[ (U^{i_1} \otimes \cdots \otimes U^{i_N})(f) = \sum_{j_1=1}^{k_{i_1}} \cdots \sum_{j_N=1}^{k_{i_N}} f(Y^{i_1}_{j_1}, \ldots, Y^{i_N}_{j_N}) \cdot (a^{i_1}_{j_1} \otimes \cdots \otimes a^{i_N}_{j_N}) \]

- Using the 1D formula, the sparse interpolant \( A_{q,N} \), where \( q \) is the depth of sparse grid interpolation \( (q \geq 0, q \in \mathbb{N}_0) \) and \( N \) is the number of stochastic dimensions, is given by the Smolyak algorithm as
  \[ A_{q,N}(f) = A_{q-1,N}(f) + \Delta A_{q,N}(f) \quad \Delta^i = U^i - U^{i-1} \quad A_{-1,N} = 0 \]
  \[ U^0 = 0 \quad ||i|| = i_1 + \cdots + i_N \]

  \[ A_{q,N}(f) = \sum_{||i|| \leq N+q} \left( \Delta^{i_1} \otimes \cdots \otimes \Delta^{i_N} \right) A_{q,N}(f) = \sum_{||i|| \leq N+q} \sum_{j_i \in B_i} \left( a_{i_1}^{i_1} \otimes \cdots \otimes a_{i_N}^{i_N} \right) \]

  \[ \cdot \left( f(Y^{i_1}_{j_1}, \ldots, Y^{i_N}_{j_N}) - A_{q-1,N}(f)(Y^{i_1}_{j_1}, \ldots, Y^{i_N}_{j_N}) \right) \]

- Here, we define the **hierarchical surplus** as:
  \[ w^i_j = f(Y^{i_1}_{j_1}, \ldots, Y^{i_N}_{j_N}) - A_{||i||-1,N}(f)(Y^{i_1}_{j_1}, \ldots, Y^{i_N}_{j_N}) \]
Nodal basis versus hierarchical basis

Nodal basis

\[ f = f(Y_1^3)a_1^3 + f(Y_2^3)a_2^3 + f(Y_3^3)a_3^3 + f(Y_4^3)a_4^3 + f(Y_5^3)a_5^3 \]

Hierarchical basis

\[ f = w_1^1 a_1^1 + w_1^2 a_1^2 + w_2^2 a_2^2 + w_1^3 a_1^3 + w_2^3 a_2^3 \]
The mean of the random solution can be evaluated as follows:

\[
\mathbb{E}[f(x)] = \sum_{\|i\| \leq N+q} \sum_{j \in B_i} w^i_j(x) \cdot \int_{\Gamma} a^i_j(Y) dY 
\]

\[
\int_{0}^{1} a^i_j(Y) dY = \begin{cases} 
1, & \text{if } i = 1, \\
\frac{1}{4}, & \text{if } i = 2, \\
2^{1-i}, & \text{otherwise.}
\end{cases}
\]

Denoting \( \int_{\Gamma} a^i_j(Y) dY = I^i_j \), we rewrite the mean as

\[
\mathbb{E}_q[f(x)] = \sum_{\|i\| \leq N+q} \sum_{j \in B_i} w^i_j(x) \cdot I^i_j
\]

To obtain the variance of the solution, we need to first obtain an approximate expression for \( u^2 \)

\[
u^2(x, Y) = \sum_{\|i\| \leq N+q} \sum_{j \in B_i} v^i_j(x) \cdot a^i_j(Y)
\]

\[
\text{Var}[u(x)] = \mathbb{E}[u^2(x)] - (\mathbb{E}[u(x)])^2
\]

\[
= \sum_{\|i\| \leq N+q} \sum_{j \in B_i} v^i_j(x) \cdot I^i_j - \left( \sum_{\|i\| \leq N+q} \sum_{j \in B_i} w^i_j(x) \cdot I^i_j \right)^2
\]
Let us first revisit the 1D hierarchical interpolation

- For smooth functions, the hierarchical surpluses tend to zero as the interpolation level increases.

- Finite discontinuities are indicated by the magnitude of the hierarchical surplus.

- The bigger the magnitude is, the stronger the underlying discontinuity is.

- Therefore, the hierarchical surplus is a natural candidate for error control and adaptivity. If the hierarchical surplus is \textit{larger} than a pre-defined value (threshold), we simply add the $2N$ neighboring points to the current point.

---

Definition of the error indicator

- The mean of the random solution can be evaluated as follows:

\[
\mathbb{E}[f(x)] = \sum_{\|i\| \leq N+q} \sum_{j \in B_i} w_{ij}(x) \cdot \int_{\Gamma} a_j^i(Y) dY = \begin{cases} 
1, & \text{if } i = 1, \\
\frac{1}{4}, & \text{if } i = 2, \\
2^{1-i}, & \text{otherwise.}
\end{cases}
\]

- Denoting \( \int_{\Gamma} a_j^i(Y) dY = I_j^i \), we rewrite the mean as

\[
\mathbb{E}_q[f(x)] = \sum_{\|i\| \leq N+q} \sum_{j \in B_i} w_{ij}(x) \cdot I_j^i
\]

- We now define the error indicator as follows:

\[
\gamma_j^i = \frac{\|w_{ij}(x) \cdot I_j^i\|_{L^2}}{\|\mathbb{E}_{\|i\|-N-1}\|_{L^2}}
\]

- In addition to the surpluses, this error indicator incorporates information from the basis functions. This forces the error to decrease to a sufficient small value for a large interpolation level. This error indicator guarantees that the refinement would stop at a certain interpolation level.
Adaptive sparse grid interpolation: Algorithm

**Algorithm 1** Adaptive sparse grid interpolation

1. Set level of Smolyak construction $q = 0$.
2. Construct the first level adaptive sparse grid $\mathcal{G}_{N,N}$.
3. Calculate the function value at the point $(0.5, \ldots, 0.5)$.
4. Generate the $2N$ neighbor points and add them to the active index set.
5. Set $q = q + 1$.

   **while** $q \leq q_{max}$ and the active index set is not empty **do**

   1. Copy the points in the active index set to an old index set and clear the active index set.
   2. Calculate in parallel the hierarchical surplus of each point in the old index set according to

      $$w^i_j = f(Y^{i_1}_{j1}, \ldots, Y^{i_N}_{jN}) - A_{q-1,N}(f)(Y^{i_1}_{j1}, \ldots, Y^{i_N}_{jN}).$$

   Here, we use all of the existing collocation points in the current adaptive sparse grid $\mathcal{G}_{N+q-1,N}$. This allows us to evaluate the surplus for each point from the old index set in parallel.

   3. For each point in the old index set, if $\gamma^i_j \geq \varepsilon$
      
      - Generate $2N$ neighbor points of the current active point.
      - Add them to the active index set.
   4. Add the points in the old index set to the existing adaptive sparse grid $\mathcal{G}_{q-1,N}$. Now the adaptive sparse grid becomes $\mathcal{G}_{q,N}$.
   5. $q = q + 1$.

   **end while**
**Adaptive sparse grid interpolation**

Ability to detect and reconstruct steep gradients

\[ f(x, y) = \frac{1}{|10^{-3} - x^2 - y^2| + 10^{-3}} \]

\[ f(x, y) = \frac{1}{|0.3 - x^2 - y^2| + 10^{-1}} \]
Let \( f(Y) \) be a real-value multivariate stochastic function: \( \mathbb{R}^N \rightarrow \mathbb{R} \), which depends on a \( N \)-dimensional random vector \( Y = (Y_1, Y_2, \ldots, Y_N) \in [0,1]^N \). A HDMR of \( f(Y) \) can be described by

\[
f(Y) = f_0 + \sum_{s=1}^{N} \sum_{i_1 < \cdots < i_s} f_{i_1 \cdots i_s}(Y_{i_1}, \ldots, Y_{i_s})
\]

where the interior sum is over all sets of \( S \) integers \( i_1, \ldots, i_s \), that satisfy \( 1 \leq i_1 < \cdots < i_s \leq N \). This relation means that

\[
f(Y) = f_0 + \sum_{i=1}^{N} f_i(Y_i) + \sum_{i_1 < i_2} f_{i_1 i_2}(Y_{i_1}, Y_{i_2}) + \sum_{i_1 < i_2 < i_3} f_{i_1 i_2 i_3}(Y_{i_1}, Y_{i_2}, Y_{i_3}) + \cdots + \sum_{i_1 < \cdots < i_s} f_{i_1 \cdots i_s}(Y_{i_1}, \ldots, Y_{i_s}) + \cdots + f_{12\cdots N}(Y_1, \ldots, Y_N)
\]

It can be viewed as a finite hierarchical correlated function expansion in terms of the input random variables with increasing dimensions.

For most physical systems, the first- and second-order expansion terms are expected to have most of the impact upon the output.

This equation is often written in a more compact notation:

\[ f(Y) = \sum_{u \subseteq D} f_u(Y_u) \]

for a given set \( u \subseteq D \) where \( D := \{1, \ldots, N\} \) denotes the set of coordinate indices and \( f_0(Y_0) = f_0 \). Here, \( Y_u \) denotes the \(|u|\) - dimensional vector containing those components of \( Y \) whose indices belong to the set \( u \), where \(|u|\) is the cardinality of the corresponding set \( u \), i.e. \( Y_u = (Y_i)_{i \in u} \).

For example, if \( u = \{1, 3, 5\} \), then \(|u| = 3\) and \( f_u(Y_u) \) implies \( f_{135}(Y_1, Y_3, Y_5) \).

The component functions \( f_u(Y_u) \) can be derived by minimizing the error functional

\[
\int_{\Gamma} \left[ f(Y) - f_0 - \sum_{i=1}^{N} f_i(Y_i) - \cdots - \sum_{i_1 < \cdots < i_s} f_{i_1 \cdots i_s}(Y_{i_1}, \ldots, Y_{i_s}) \right]^2 d\mu(Y) \quad 0 \leq s \leq N
\]

subject to the orthogonal constraint

\[ \int_{\Gamma^N} f_u(Y_u) f_v(Y_v) d\mu(Y) = 0, \text{ for } u \neq v \]
HDMR: Component functions

- The measure $d\mu$ determines the particular form of the error functional and of the component functions.

- By the variational principle, the component functions $f_u(Y_u)$ can be explicitly given as

$$f_u(Y_u) := \sum_{v \subseteq u} (-1)^{|u|-|v|} P_v f(Y_v)$$

where the measure $\mu$ induces the projection operator $P_u : \Gamma^N \rightarrow \Gamma^{|u|}$

$$P_u f(Y_u) := \int_{\Gamma^{N-|u|}} f(Y) d\mu_{D\setminus u}(Y)$$

where

$$d\mu_{D\setminus u}(Y) := \prod_{i \notin u} d\mu_i(Y_i)$$

- There are two different forms of HDMR induced by different measure: ANOVA-HDMR and CUT-HDMR.
**ANOVIA-HDMR versus CUT-HDMR**

HDMR

ANOVA-HDMR

CUT-HDMR

Lebesgue measure $d\mu(Y) = d(Y) = \prod_{i=1}^{N} Y_i$

$P_u f(Y_u) := \int_{\mathcal{Y}_{N-|u|}} f(Y) dY_{\mathcal{D}\setminus u}$

$N - |u|$ dimensional integration

$f_0 = \int_{\mathcal{Y}} f(Y) dY$, $f_i(Y_i) = \int_{\mathcal{Y}_{N-1}} f(Y) \prod_{j\neq i} dY_j - f_0$

$f_{ij}(Y_i, Y_j) = \int_{\mathcal{Y}_{N-2}} f(Y) \prod_{k\neq i, j} dY_k - f_i(Y_i) - f_j(Y_j) - f_0$, ...

Computational expensive -- requires a N-dimensional integral for the constant terms

Dirac measure $d\mu(Y) = \prod_{i=1}^{N} \delta(Y_i - \bar{Y}_i) dY_i$

at a reference point $Y = (\bar{Y}_1, \bar{Y}_2, ..., \bar{Y}_N)$

$P_u f(Y_u) := f(Y) |_{Y=\bar{Y}\setminus Y_u}$

$|u|$ dimensional function

$f_0 = f(\bar{Y})$, $f_i(Y_i) = f(Y) |_{Y=\bar{Y}\setminus Y_i} - f_0$

$f_{ij}(Y_i, Y_j) = f(Y) |_{Y=\bar{Y}\setminus (Y_i, Y_j)} - f_i(Y_i) - f_j(Y_j) - f_0$, ...

Computational efficient -- requires function evaluations at sample points

---

Cornell University
Materials Process Design and Control Laboratory
Within the framework of CUT-HDMR, we can write

\[
f(Y) = \sum_{\mathbf{u} \subseteq \mathcal{D}} f_{\mathbf{u}}(Y_{\mathbf{u}}) = \sum_{\mathbf{u} \subseteq \mathcal{D}} \sum_{\mathbf{v} \subseteq \mathbf{u}} (-1)^{|\mathbf{u}| - |\mathbf{v}|} f(Y_{\mathbf{v}})_{Y=\bar{Y} \setminus Y_{\mathbf{v}}}
\]

where the notation \(Y=\bar{Y} \setminus Y_{\mathbf{v}}\) means that the components of \(Y\) other than those indices that belong to the set \(\mathbf{u}\) equal to those of the reference point.

If the HDMR is a converged expansion, the choice of this point does not affect the approximation. In this work, the mean of the random input vector is chosen as the reference point.

Therefore, the \(N\)-dimensional stochastic problem is transformed to several lower-order \(|\mathbf{v}|\)-dimensional problems \(f(Y_{\mathbf{v}})_{Y=\bar{Y} \setminus Y_{\mathbf{v}}}\) which can easily solved by ASGC:

\[
f(Y) = \sum_{\mathbf{u} \subseteq \mathcal{D}} \sum_{\mathbf{v} \subseteq \mathbf{u}} (-1)^{|\mathbf{u}| - |\mathbf{v}|} \sum_{||i|| \leq N+q} \sum_{j \in B_i} w_{ij}^{\mathbf{v}}(x) \cdot a_j^{i}(Y_{\mathbf{v}})
\]

where \(w_{ij}^{\mathbf{v}}(x)\) are the hierarchical surpluses for different sub-problems indexed by \(\mathbf{v}\) and \(a_j^{i}(Y_{\mathbf{v}})\) is only a function of the coordinates belonging to \(\mathbf{v}\).
Let us denote

\[ J_u = \sum_{v \subseteq u} (-1)^{|u|-|v|} \sum_{\|i\| \leq N+q} \sum_{j \in B_i} w_{ij}(x) \cdot I_j^i \]

as the mean of the component function \( f_u \). Then the mean of the HDMR expansion is simply \( \mathbb{E}[f(Y)] = \sum_{u \subseteq D} J_u \).

The basic conjecture underlying HDMR is that the component functions arising in typical physical systems will not likely exhibit high-order cooperativity among the input variables such that the 1st- and 2nd-order expansion terms are expected to have most of the impact upon the output and the contribution of higher-order terms would be insignificant.

In other words, instead of solving the \( N \)-dimensional problem directly using ASGC, which is impractical for extremely high dimensional problems, we only need to solve several one- or two-dimensional problems, which can be solved efficiently via ASGC.
Effective dimension of a stochastic function

- Let $\hat{f} := \sum_{\mathbf{u} \subseteq \mathcal{D}} |J_{\mathbf{u}}|$ be the sum of all contributions to the mean value. Here, $|\cdot|$ denotes the absolute value.

- Then, for the proportion $\alpha \in (0, 1]$, the truncation dimension is defined as the smallest integer $N_t$, such that
  $$\sum_{\mathbf{u} \subseteq \{1, \ldots, N_t\}} |J_{\mathbf{u}}| \geq \alpha \hat{f},$$
  whereas, the superposition dimension is defined as the smallest integer $N_s$, such that
  $$\sum_{|\mathbf{u}| \leq N_s} |J_{\mathbf{u}}| \geq \alpha \hat{f}$$

- The superposition dimension is also called the order of the HDMR expansion.

- With the definition of effective dimensions, we can thus truncate the expansion and take only a subset $\mathcal{S}$ of all indices $\mathbf{u} \subseteq \mathcal{D}$. Here we assume that the set $\mathcal{S}$ satisfies the following admissibility condition:
  $$\mathbf{u} \in \mathcal{S} \text{ and } \mathbf{v} \subset \mathbf{u} \Rightarrow \mathbf{v} \in \mathcal{S}$$
  This is to guarantee that all the terms can be calculated via the recursive expression for computing the component functions.
In practice, we always truncate the expansion by taking only a subset of all indices \( u \subseteq D \). We can define an interpolation formula \( A_s f \) for the \( S \) approximation of \( f \) as

\[
A_s f := \sum_{u \in S} A(f_u)
\]

It is common to refer to the terms \( \{ f_u : |u| = l \} \) collectively as the \( l \)-“order- terms”. Then the expansion order is the maximum of \( l \). The number of collocation points in this expansion is defined as the sum of the number of points for each sub-problem, i.e. \( M = \sum_{v \in S} M_v \)

However, the number of order- \( l \) component functions is \( \sum_{i=1}^{l} \frac{N!}{i!(N-i)!} \), which increases quickly with the number of dimensions. Therefore, we developed an adaptive version of HDMR.
We fix \( \alpha \in (0, 1] \) and assume that \( N_s \) and \( N_t \), the corresponding superposition and truncation dimensions, are known. With the definition of the index set \( S_{N_t,N_s} := \{u \subseteq \{1, \ldots, N_t\}, |u| \leq N_s\} \), we have the following theorem:

**Theorem 1.** Let \( S = S_{N_t,N_s} \), and let \( A \) be the ASGC interpolant with the same error threshold \( \varepsilon \) for all the sub-problems. Then:

\[
|f - A_S f| \leq c(N_s, N_t) \varepsilon + \varepsilon_t,
\]

for all \( f \in F_N \). Here, the constant \( c(N_t, N_s) \) depends on the effective dimensions, but does not depend on the nominal dimension \( N \). \( \varepsilon_t \) is the truncation error of according to the definition of effective dimensions.

Therefore, it is expected that the expansion converges to the true value with decreasing error threshold \( \varepsilon \) and increasing number of component functions.
Adaptive HDMR

- For extremely high dimensional problems, even a 2\textsuperscript{nd} order expansion is impractical due to the increase of the number of component functions. Therefore, we would like to develop an adaptive version of HDMR for automatically and simultaneously detecting the truncation and superposition dimensions.

- We assume each component function $f_u$ is associated with a weight $\eta_u \geq 0$ which describes the contribution of the term $f_u$ to the HDMR.

- First, we try to find the important dimensions. To this end, we always construct the 0\textsuperscript{th} - and 1\textsuperscript{st}-order HDMR expansion. We define a weight:

\[
\eta_i = \frac{\| J \{i\} \|_{L_2}}{\| f_0(\mathbf{Y}) \|_{L_2}} \quad J \{i\} = \int f_i(Y_i) dY_i
\]

Then we define the important dimensions as those whose weights are larger than a predefined error threshold $\theta_1$. Only higher-order terms which consist of only these important dimensions are considered.

Here, the $L_2$ norm is defined in the spatial domain.
For example, if the important dimensions are 1, 3 and 5, then only the higher-order terms \{13\}, \{15\}, \{35\} and \{135\} are considered.

However not all the possible terms are computed. For higher-order term, a weight is also defined as

\[
\eta_u = \frac{||J_u||_{L_2}}{||\sum_{v \in S, |v| \leq |u| - 1} J_v||_{L_2}}
\]

We also define the important terms in a similar way. We put all the important dimensions and higher-order terms in to a set \(T\). When adaptively constructing HDMR for each new order, we only calculate the term \(f_u\) whose indices satisfy the admissibility relation

\[
u \in D \text{ and } v \subseteq u \Rightarrow v \in T\]
Continued with the example, now if we want to construct the 2\textsuperscript{nd} order expansion, only \{13\}, \{15\} and \{35\} are calculated.

Then we compute the weights for each term. Assume \{13\} is the important term, the important index set \( T = \{0,1,3,5,13\} \).

Now, we go to 3\textsuperscript{rd} order expansion. The only possible term is \{135\}. Its subsets \{35\} and \{15\} do not belong to the important index set \( T \), i.e. \{1,3,5\} does not satisfy the admissibility condition. Therefore, the construction stops.

In other words, among all possible indices, we only find the terms which can be computed using the previous known important component functions and have significant contributions to the overall expansion.
Let us denote the order of expansion as $p$. Furthermore, we also define a relative error $\rho$ of the integral value between two consecutive expansion orders $p$ and $p - 1$ as

$$
\rho = \frac{\left\| \sum_{\lvert u \rvert \leq p} J_u - \sum_{\lvert u \rvert \leq p-1} J_u \right\|_{L_2}}{\left\| \sum_{\lvert u \rvert \leq p-1} J_u \right\|_{L_2}}
$$

If $\rho$ is smaller than another predefined error threshold $\theta_2$, the HDMR is regarded as converged and the construction stops.

In this way, the construction will automatically stop and then the obtained HDMR expansion can be used as a stochastic surrogate model (response surface) for the solution. Any statistics can be easily computed through this expansion.
**Algorithm 2** Adaptive construction of the index set $S$

Initialize: Let $S = \{\emptyset\}$, $R = \{\emptyset\}$ and $T = \{\emptyset\}$. Set $p = 1$. Construct the zeroth and first-order component functions:

- Solve each sub-problem using the ASGC method with error threshold $\varepsilon$ and add all the indices to $S$.
- Compute the weights of each first-order term. Add those dimensions which satisfy $\eta \geq \theta_1$ to set $T$.

repeat

- $p \leftarrow p + 1$. Construct the set $R$ whose indices satisfy the admissibility relation for $|u| = p$.
- If $R \neq \{\emptyset\}$, for each index $u \in R$, solve the corresponding sub-problem using ASGC with error threshold $\varepsilon$ and add all the indices to $S$.
- Compute the weight of component functions. Add those indices which satisfy $\eta \geq \theta_1$ to set $T$ and clear set $R$.
- Compute the relative error $\rho$.

until $R = \{\emptyset\}$ or $\rho < \theta_2$;
Numerical example: Flow through random media

Basic equation for pressure and velocity in a domain

\[ \nabla \mathbf{u} = f \quad \text{in} \quad \mathcal{D} \]
\[ \mathbf{u} = -k \nabla p \quad \text{in} \quad \mathcal{D} \]

where \( f(x) \) denotes the deterministic source/sink term. Homogeneous boundary condition is applied. Mixed finite element method is used to solve the deterministic problem at the collocation points.

To impose the non-negativity of the permeability, we will treat the permeability as a log random field obtained from the K-L expansion

\[ Y(\omega) = \log(K(\omega)) = \sum_{i=1}^{N} \sqrt{\lambda_i} \phi_i(x) Y_i \]

where \( Y \) is a zero mean Gaussian random field with covariance function

\[ \text{Cov}(x, y) = \sigma^2 \exp \left( -\frac{|x_1 - y_1|}{L} - \frac{|x_2 - y_2|}{L} \right) \]

where \( L \) is the correlation length and \( \sigma \) is the standard deviation.
Numerical example: K-L Expansion

The eigenvalues and their corresponding eigenfunctions can be determined analytically. The $Y_i$ are assumed as i.i.d. uniform random variables on $[-1,1]$.

According to the decay rate of eigenvalues, the number of stochastic dimensions is $N = 33,108$ and $500$, respectively for $L = 1.0$, $0.5$ and $0.25$.

Monte Carlo simulations are conducted for the purpose of comparison. For each case, the reference solution is taken from $10^6$ samples and all errors are defined as normalized $L_2$ errors. In all cases, $\theta_2 = 10^{-4}$.
Standard deviation for different correlation lengths

\[ \sigma^2 = 1.0, \varepsilon = 10^{-6} \]

\[ N = 33 \]

error = 1.46 \times 10^{-3}

\[ N = 500 \]

error = 1.39 \times 10^{-3}

\[ N = 108 \]

error = 1.19 \times 10^{-3}

Number of component functions is 2271 while for the full 2\textsuperscript{nd}-order expansion it is 125251. The advantage of using adaptive HDMR is obvious.

Standard deviation of the \( v \) velocity-component along the cross section \( y = 0.5 \) for different correlation lengths
PDF at (0,0.5) for different correlation lengths

PDF of the $v$ velocity-component at point (0,0.5) for different correlation lengths

$N = 33$

$N = 108$

$N = 500$

$\sigma^2 = 1.0, \varepsilon = 10^{-6}$
Convergence of the normalized errors of the standard deviation of the \( v \) velocity-component for different correlation lengths

Convergence of the normalized errors

\[ N = 33 \]

\[ N = 108 \]

\[ N = 500 \]

Algebraic convergence rate better than MC

Nearly the same for three cases, it does not depend on the smoothness of the random space

\( \sigma^2 = 1.0, \theta_1 = 10^{-4} \)
Standard deviations for different $\sigma^2$ with $N = 500$

$e = 8.08 \times 10^{-4}$

$\sigma^2 = 0.01$

$e = 7.37 \times 10^{-4}$

$\sigma^2 = 0.25$

Standard deviation of the $\nu$ velocity-component along the cross section $y = 0.5$ for different $\sigma^2$. 
PDF at (0,0.5) for different $\sigma^2$ with $N = 500$

For low input variability, even 1st-order expansion is accurate.

For high input variability, the 1st-order expansion deviates from MC. More component terms are needed to improve accuracy.

For moderate input variability, 1st order does not deviate significantly from MC. However, a few 2nd-order terms are still needed.

$L = 0.25, N = 500$
Convergence of the normalized errors with $N = 500$

Direct solution of the 500 dimensional problem using ASGC is impractical due to the huge computational cost.

Convergence rate deteriorates with increasing input variability. However, it is still better than that of MC.
Effect of choices of the reference point

From our test studies, the mean vector always gives satisfactory results with much less computational cost.

It is also interesting to note that when $\mathbf{Y} = (0.6, \ldots, 0.6)$, in order to achieve an error of order $O(10^{-3})$, the number of component functions is 880 while the number is 6018 when using conventional HDMR.

$N_1$ is the number of important dimension
$p$ is the highest expansion order

<table>
<thead>
<tr>
<th>Reference Point</th>
<th>$\mathbf{Y}$</th>
<th># Terms</th>
<th># Points</th>
<th>$N_1$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.5, ..., 0.5)</td>
<td>(0.5, ..., 0.5)</td>
<td>287</td>
<td>27968</td>
<td>23</td>
<td>2</td>
</tr>
<tr>
<td>(0.6, ..., 0.6)</td>
<td>(0.6, ..., 0.6)</td>
<td>880</td>
<td>170992</td>
<td>33</td>
<td>3</td>
</tr>
<tr>
<td>(0.1, ..., 0.1)</td>
<td>(0.1, ..., 0.1)</td>
<td>5526</td>
<td>185460</td>
<td>33</td>
<td>4</td>
</tr>
</tbody>
</table>
Natural Convection: $L = 0.1$ and $\sigma=1.0$

$u$ velocity

$N = 10, \ Y_i \sim [-1,1]$

Mean

Std

HDMR

10^5 MC
Natural Convection: $L = 0.1$ and $\sigma = 1.0$

- We choose $\varepsilon = 10^{-5}$ and $\theta_1 = 10^{-4}$. The construction stopped after order 2. However, not all terms are needed. Only 17 component functions are in the final HDMR.

- Only 1085 collocation points are needed to achieve an error $1.78 \times 10^{-3}$. 
- Thermal and fluid transport in heterogeneous media are ubiquitous
- Range from large scale systems (geothermal systems) to the small scale
- Most critical devices/applications utilize heterogeneous/polycrystalline/functionally graded materials
- Properties depend on the distribution of material/microstructure
- But only possess limited information about the microstructure/property distribution (e.g. 2D images)

Incorporate limited information into stochastic analysis:
- worst case scenarios
- variations on physical properties
1. Property extraction
Extract properties P1, P2, .. Pn, that the structure satisfies.
These properties are usually statistical: Volume fraction, 2 Point correlation, auto correlation

2. Microstructure/property reconstruction
Reconstruct realizations of the structure satisfying the correlations.

3. Reduced model
Construct a reduced stochastic model of property variations from the data. This model must be able to approximate the class of structures.

4. Stochastic analysis
Solve the heterogeneous property problem in the reduced stochastic space for computing property variations.

Methodology for creating linear models from data

- Given some limited information (either in terms of statistical correlation functions or sample microstructure/property variations)

- Utilize some reconstruction methodology to create a finite set of realizations of the property/microstructure.

- Utilize this data set to construct a model of this variability

Use Proper Orthogonal Decomposition (POD), Principal Component Analysis (PCA) to construct a reduced order model of the data.

\[ = a_1 + a_2 + \ldots + a_n \]

Convert variability of property/microstructure to variability of coefficients.

Not all combinations allowed. Developed subspace reducing methodology\(^1\) to find the space of allowable coefficients that reconstruct plausible microstructures

LINEAR APPROACH TO MODEL GENERATION

Successfully applied to investigate effect of heterogeneous media in diffusion phenomena in two-phase microstructures.

PCA based methods are easy to implement and are well understood.

But PCA based methods are linear projection methods

*Only guaranteed to discover the true structure of data lying on a linear subspace of the high dimensional input space*

PCA works very well when the input space is linear

What about when the input space is curved/non-linear?

PCA based techniques tend to overestimate the dimensionality of the model

Further related issues:

- How to generalize it to other properties/structures? Can PCA be applied to other classes of microstructures, say, polycrystals?
- How does convergence change as the amount of information increases? Computationally?
- Find structure of data lying on a possibly non-linear subspace of the high-dimensional input data.

- PCA finds a low-dimensional embedding of the data points that best preserves their variance as measured in the high-dimensional input space. Variance is measured based on Euclidian distance. This results in a linear subspace approximation of the data.

- But what if the data lie on nonlinear curve in high-dimensional space?

- Have to unfold/unravel the curve

- Need non-linear approaches
Set of images. Each image = 64x64 = 4096 pixels
Each image is a point in 4096 dimensional space.
But each and every image is related (they are pictures of the same object). *Same object but different poses.*
That is, all these images lie on a unique curve (manifold) in $\mathbb{R}^{4096}$.
Can we get a parametric representation of this curve?

**Problem:** Can the parameters that define this manifold be extracted, ONLY given these images (points in $\mathbb{R}^{4096}$)

**Solution:** Each image can be uniquely represented as a point in 2D space (UD, LR).

**Strategy:** based on the ‘manifold learning’ problem
A FORMAL DEFINITION OF THE PROBLEM

State the problem as a parameterization problem (also called the manifold learning problem)

**Given a set of N unordered points belonging to a manifold \( \mathcal{M} \) embedded in a high-dimensional space \( \mathbb{R}^n \), find a low-dimensional region \( \mathcal{A} \subset \mathbb{R}^d \) that parameterizes \( \mathcal{M} \), where \( d \ll n \)**

Classical methods in manifold learning have been methods like the Principle Component Analysis (PCA) and multidimensional scaling (MDS).

These methods have been shown to extract optimal mappings when the manifold is embedded linearly or almost linearly in the input space.

In most cases of interest, the manifold is non-linearly embedded in the input space, making the classical methods of dimension reduction highly approximate.

Two approaches developed that can extract non-linear structures while maintaining the computational advantage offered by PCA\(^1,2\).

1) Geometry can be preserved if the distances between the points are preserved – Isometric mapping.

2) The geometry of the manifold is reflected in the geodesic distance between points.

3) First step towards reduced representation is to construct the geodesic distances between all the sample points.
PROPERTIES OF THE CURVE $\mathcal{M}$

Given a set of $N$ sample points $\{x_i\}$ lying on $\mathcal{M}$
Must have a notion of distance between these sample points.
Define an appropriate function

$$\mathcal{D}: \mathcal{M} \times \mathcal{M} \rightarrow [0, \infty)$$

that determines the difference between any two points
Ensure that it satisfies the properties of positive-definiteness, symmetry and the triangle inequality.
Lemma 1: $(\mathcal{M}, \mathcal{D})$ is a metric space

Lemma 2a: $(\mathcal{M}, \mathcal{D})$ is a bounded metric space

Lemma 2b: $(\mathcal{M}, \mathcal{D})$ is dense

Lemma 2c: $(\mathcal{M}, \mathcal{D})$ is complete

Theorem: $(\mathcal{M}, \mathcal{D})$ is compact\(^1\).

A compact manifold embedded in a high-dimensional space can be isometrically mapped to a region in a low-dimensional space\(^2\).


A compact manifold embedded in a high-dimensional space can be isometrically mapped to a region in a low-dimensional space.

Isometry encoded into the geodesic distances

Have no notion of the geometry of the manifold to start with. Hence cannot construct true geodesic distances!

$$\mathcal{D}_M(i, j) = \inf_{\gamma} \{\text{length}(\gamma)\}$$

Approximate the geodesic distance using the concept of graph distance $\mathcal{D}_G(i,j)$: the distance of points far away is computed as a sequence of small hops.

This approximation, $\mathcal{D}_G$, asymptotically matches the actual geodesic distance $\mathcal{D}_M$. In the limit of large number of samples$^{1,2}$. (Theorem 4.5 in Ref. 1)

$$(1 - \lambda_1)\mathcal{D}_M(i, j) \leq \mathcal{D}_G(i, j) \leq (1 + \lambda_2)\mathcal{D}_M(i, j)$$

2. M.Bernstein, V. deSilva, J.C.Langford, J.B.Tenenbaum, Graph approximations to geodesics on embedded manifolds, Dec 2000
The length functional of the minimal spanning tree of the geodesic matrix, \( M \), is related to the intrinsic dimensionality of the low-dimensional representation of the manifold \(^1,^2\).

**Theorem 4.6:** Let \( \mathcal{M}_{S^2} \) be a smooth \( d \) dimensional manifold embedded in \( \mathbb{R}^n \) through a conformal map \( \varphi : \mathbb{R}^d \rightarrow \mathcal{M}_{S^2} \). Let \( 2 \leq d \leq n \). Suppose that \( \{x_i\} \), \( i = 1, \ldots, N \) are random vectors in \( \mathcal{M}_{S^2} \). Assume that each of the edge lengths \( |e_{ij}|_M \) in the edge matrix (or the matrix of manifold distances) \( M \) converge to \( |\varphi^{-1}(x_i) - \varphi^{-1}(x_j)|_2 \) as \( N \rightarrow \infty \) (i.e the graph distance converges to the true manifold distance). This is guaranteed by Theorem 4.5. Then the length functional, \( L(\{x\}) \) of the GMST satisfies:

\[
\lim_{N \rightarrow \infty} \frac{L(\{x\})}{N^{(d'-1)/d'}} = \begin{cases} 
\infty & \text{if } d' < d; \\
\beta_m C & \text{if } d' = d; \\
0 & \text{if } d' > d.
\end{cases}
\] (17)


THE NONLINEAR MODEL REDUCTION FRAMEWORK

Given N unordered samples

N points in a low dimensional space

The procedure results in N points in a low-dimensional space. The geodesic distance + MDS step (Isomap algorithm\(^1\)) results in a low-dimensional convex, connected space\(^2\), \(\mathcal{A} \subset \mathbb{R}^d\).

Using the N samples, the reduced space is given as \(\mathcal{A}\) serves as the surrogate space for \(\mathcal{M}\).

Access variability in \(\mathcal{M}\) by sampling over \(\mathcal{A}\).

BUT have only come up with \(\mathcal{M} \rightarrow \mathcal{A}\) map …. Need \(\mathcal{A} \rightarrow \mathcal{M}\) map too

The Reduced Order Stochastic Model

Only have N pairs to construct $\mathcal{A} \rightarrow \mathcal{M}$ map. Various possibilities based on specific problem at hand. But have to be conscious about computational effort and efficiency. Illustrate 3 such possibilities below. Error bounds can be computed.¹

1. Nearest neighbor map
2. Local linear interpolation
3. Local linear interpolation with projection

Algorithm consists of two parts.

1) Compute the low-dimensional representation of a set of N unordered sample points belonging to a high-dimensional space.

Given N unordered samples → Compute pairwise geodesic distance → Perform MDS on this distance matrix → N points in a low dimensional space

For using this model in a stochastic collocation framework, must sample points in $\mathcal{A} \rightarrow \mathcal{M}$

2) For an arbitrary point $\xi \in \mathcal{A}$ must find the corresponding point $x \in \mathcal{M}$. Compute the mapping from $\mathcal{A} \rightarrow \mathcal{M}$
Seamlessly couple stochastic analysis with multiscale analysis.

Multiscale framework (large deformation/thermal evolution) + Adaptive stochastic collocation framework

Provides roadmap to efficiently link any validated multiscale framework

Coupled with a data-driven input model strategy to analyze realistic stochastic multiscale problems.

\[ r(\mu) \]
\[ g(r) \]

Limited data  \( T = -0.5 \)  \( T = 0.5 \)

Stochastic multiscale framework

Statistics extraction + model generation

Mean statistics

Higher-order statistics

Statistics extraction + model generation

Limited data

Mean statistics

Higher-order statistics
OTHER APPLICATIONS OF THIS FRAMEWORK

From unordered samples in high-dimensional space to a convex space representing the parameterization

- This methodology has significant applications to problems where working in high-dimensional spaces is computationally intractable.
- Can pose the problem in a low-dimensional space

visualizing property evolution, process-property maps, searching and contouring, representing input uncertainty, data mining …
Some input parameters, boundary conditions, initial conditions, or operational sequences have a significant impact on the stochastic solution.

Improved input representation (i.e., lesser uncertainty) significantly reduces the uncertainty in the final solution.

The stochastic solution is **sensitive to which parameters or input terms?** Need a robust, seamlessly interfacing methodology to estimate this.

- Which input parameters to consider as random?
- Quantify the effect of uncertainty on each parameter on system response
- Decide the length scale of experimentation and the needed fidelity of each measurement.
Variability of Material Properties of Polycrystals

Problem definition

Given:
- Grain size snapshots constrained by moments (mean size, standard deviation, higher-order moments).
- Texture snapshots from random process.

Goal:
- The variability in material properties and response

Methodologies

- Model reduction to reduce the complexity of stochastic input
  - Nonlinear Model Reduction (manifold learning) to reduce grain size space
  - Karhunen-Loeve Expansion to reduce texture space
- Adaptive sparse grid collocation to solve stochastic partial differential equations

Zheng Li, Bin Wen and N. Zabaras, "Data-driven stochastic models for the representation of polycrystalline microstructures", Computational Materials Science, in press.
Nonlinear Model Reduction on Grain Size Feature

Given a set of \( N \) unordered points belonging to a manifold \( \mathcal{M} \) embedded in a high-dimensional space \( \mathbb{R}^n \), find a low-dimensional region \( \mathcal{A} \subset \mathbb{R}^d \) that parameterizes \( \mathcal{M} \), where \( d \ll n \).

The properties of a polycrystalline microstructure are highly dependent on its texture: orientation distribution of grains.

**Orientation representation:** Rodrigues parameters

\[ r_1 = w_1 \tan \frac{\phi}{2}, \quad r_2 = w_2 \tan \frac{\phi}{2}, \quad r_3 = w_3 \tan \frac{\phi}{2} \]

**Texture representation:** Orientation Distribution Function (ODF)

A discrete form: \( \tau(\mathbf{r}) = \left\{ r_1, r_2, r_3, \ldots, r_1^n, r_2^n, r_3^n \right\} \)

Orientation dependence of slip system (anisotropy in crystalline materials)

\[ \mathbf{m}^{j,\alpha} = R^j \mathbf{m}_{local}^{\alpha} \]
\[ \mathbf{n}^{j,\alpha} = R^j \mathbf{n}_{local}^{\alpha} \]

where

\[ R^j = \frac{1}{1 + \mathbf{r}^j \cdot \mathbf{r}^j} \left( \mathbf{I} (1 - \mathbf{r}^j \cdot \mathbf{r}^j) + 2 (\mathbf{r}^j \otimes \mathbf{r}^j - \mathbf{I} \times \mathbf{r}^j) \right) \]

12 slip systems in FCC

- \( \mathbf{m}_{local}^{\alpha} : \langle 111 \rangle \)
- \( \mathbf{n}_{local}^{\alpha} : [110] \)
Karhunen-Loeve Expansion on Texture Samples

Deterministic texture

Random process controlled by \( \{\omega_1, \omega_2, \ldots, \omega_8\} \)

Initial texture samples

Given N texture examples, construct covariance matrix of these samples

\[
\tilde{C} = \frac{1}{N-1} \sum_{i=1}^{N} (\tau_i - \bar{\tau})^T (\tau_i - \bar{\tau}),
\]

\[
\bar{\tau} = \frac{1}{N} \sum_{i=1}^{N} \tau_i
\]

The truncated Karhunen-Loeve Expansion of a random vector \( \tau \) is

\[
\tau(r, \omega) = \bar{\tau}(r, \omega) + \sum_{i=1}^{d} \sqrt{\lambda_i} \phi_i(r) \eta_i(\omega)
\]

where \( \phi_i, \lambda_i \) are the \( i \)th eigenvector and eigenvalue of \( \tilde{C} \), respectively. \( \{\eta_i(\omega)\} \) are a set of uncorrelated random variables satisfying

\[
E(\eta_i(\omega)) = 0, \quad E(\eta_i(\omega) \eta_j(\omega)) = \delta_{ij}, \quad i, j = 1, \ldots, d
\]

Texture random field thus transformed to low-dimensional space \( \eta \in \mathbb{R}^d \)
Use sparse grid collocation to obtain the stochastic characteristic of macroscale properties.

Mean of A at the end of deformation process

Variance of A at the end of deformation process

\[ \bar{E}(\text{MPa}) \quad \text{Var}(E) \quad (\text{MPa})^2 \]

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.41e05</td>
<td>4.42e08</td>
<td>Adaptive Sparse grid (level 8)</td>
</tr>
<tr>
<td>1.41e05</td>
<td>4.39e08</td>
<td>MC 10,000 runs</td>
</tr>
</tbody>
</table>

Variation of stress-strain response

\[ \times 10^5 \text{ MPa} \]

FCC copper

B. Kouchmeshky and N. Zabaras, Computational Materials Science 47 (2009) 342--352
Kernel Principal Component Analysis on input model reduction

- Let \( \{x_1, \ldots, x_N\} \) be a set of realizations of a random field in the input stochastic space \( \Gamma = \mathbb{R}^d \).

- Kernel PCA performs the traditional PCA in the feature space \( F \) corresponding to the kernel \( k = \langle \cdot, \cdot \rangle \). The kernel gives the inner product between two points in the feature space, i.e., \( k(x, y) = \langle \phi(x), \phi(y) \rangle \).

- This fact can be used to obtain the eigenvectors in the feature space \( F \) even though the non-linear map \( \phi \) is unknown. This is the so called kernel trick.
Kernel Principal Component Analysis

- Analogous to linear PCA, it involves the following eigen decomposition:
  \[ HKH = U \Sigma U^T \]
  where, \( K \) is the kernel matrix with entries \( K_{ij} = k(x_i, x_j) \), \( H \) is the centering matrix given by \( H = I - \frac{1}{N} 11^T \).

- \( I \) is the \( N \times N \) identity matrix, \( 1 = [1 \ldots 1]^T \) is an \( N \times 1 \) vector, \( U = [a_1, \ldots, a_N] \) with \( a_i = [a_{i1}, \ldots, a_{iN}]^T \) is the matrix containing the eigenvectors and \( \Sigma = \text{diag}(\lambda_1, \ldots, \lambda_N) \) contains the corresponding eigenvalues.

- The \( k \)th orthonormal eigenvector of the covariance matrix in the feature space can then be shown to be \( V_k = \sum_{i=1}^{N} a_{ki} \tilde{\phi}(x_i) \).

- Denote the projection of the \( \phi \)-image of an arbitrary realization onto the \( k \)th component by \( \beta_k \). Then, \( \beta_k = \frac{1}{\sqrt{\lambda_k}} \sum_{i=1}^{N} a_{ki} \tilde{k}(x, x_i) \).

- The projection of \( \phi(x) \) onto the subspace spanned by the first \( n \) eigenvectors is given by \( P\phi(x) = \sum_{k=1}^{n} \beta_k V_k \). This is where we achieve the dimensionality reduction.

- Finally, we need to transform the image back into the image space.
Example

- We first generate 200 sample permeabilities, where each permeability follows a mixture Gaussian distribution.

- We then perform KPCA on these samples and use the first 30 eigenvectors for reconstruction. We also keep the same number of eigenvectors in linear PCA for reconstruction.

- We then use an additional permeability realization for testing purposes.
Example: Kernel PCA

The marginal PDF of the original and reconstructed samples. It is seen that the KPCA successfully captures the multimodal feature of the PDF.
Example: Linear PCA

The marginal PDF of the original and reconstructed samples. It is seen that the Linear PCA failed to capture the multimodal feature of the PDF as expected.
At each macro scale point, due to the randomness of the microstructure, we will have a set of (microstructure) samples (realizations).

One can construct a reduced order stochastic model for each point (e.g. linear or non-linear POD). However, in this case, the reduced-order model cannot see the correlation between the set of random variables at different macropoints.

This will result in a huge dimension of the random space and thus the problem becomes computationally intractable.
Define a microstructure using its corresponding texture (ODF) as \( A(x, s, \omega) \), which is a function of macroscopic coordinate \( x \), microscopic coordinate \( s \) and the random event \( \omega \).

We will use the bi-orthogonal decomposition to construct the reduced order model.

\[
\bar{a}(x, s, \omega) = \log(A(x, s, \omega) - A_{\text{min}})
\]

\[
\bar{a}(x, s, \omega) = \bar{a}(x, s, \omega) + \hat{a}(x, s, \omega)
\]

where \( \hat{a}(x, s, \omega) = \sum_{i=1}^{\infty} \sqrt{\rho_i} \psi_i(s) \Phi_i(x, \omega) \) (1)

where \( \psi_i(s) \) are modes strongly orthogonal in microscale space and \( \Phi_i(x, \omega) \) are spatial random modes weakly orthogonal in macroscale space [1].

\[
(\psi_i, \psi_j) := \int_{\mathcal{R}} \psi_i(s) \psi_j(s) ds = \delta_{ij} \quad (2)
\]

\[
(\Phi_i, \Phi_j)^\# = \delta_{ij} \quad (3)
\]

where \( (f, g)^\# := \int_{D} <f, g> dx \quad <f, g> := \int_{\Omega} f(\omega) g(\omega) P(\omega) d\omega \)

\( D \) macroscale domain, \( \mathcal{R} \) microscale domain, \( \Omega \) random sample space

B. Kouchmeshky and N. Zabaras, Computational Materias Science 48 (2010) 213--227
From (1) and (3) we can have

$$\sqrt{\rho_i \psi_i(s)} = \left(\hat{a}(x, s, \omega), \Phi_i(x, \omega)\right)^\dagger = \int_D \langle \hat{a}(x, s, \omega), \Phi_i(x, \omega) \rangle \, dx$$  

$$= \int_D \int_G \hat{a}(x, s, \omega) \Phi_i(x, \omega) \, dP(\omega) \, dx$$  

From (1) and (2),

$$\sqrt{\rho_i \Phi_i(x, \omega)} = \int_{\mathbb{R}} \hat{a}(x, s, \omega) \psi_i(s) \, ds$$  

Now From (4) and (5) we can derive

$$\rho_i \psi_i(s) = \int_{\mathbb{R}} \int_D \int_G \hat{a}(x, s, \omega) \hat{a}(x, s', \omega) \, dP(\omega) \psi_i(s') \, ds' \, dx$$

$$= \int_{\mathbb{R}} \int_D \frac{1}{n_r} \sum_{j=1}^{n_r} \hat{a}(x, s, \xi_j) \hat{a}(x, s', \xi_j) \, dx \psi_i(s') \, ds'$$

$$\Rightarrow \rho_i \psi_i(s) = \int_{\mathbb{R}} \left( \frac{1}{n_r} \sum_{j=1}^{n_r} \sum_{i_n=1}^{n_{el}} \sum_{i_m=1}^{n_{im}} \hat{a}(x_{i_m}, s, \xi_j) \hat{a}(x_{i_m}, s', \xi_j) \hat{\eta}_{i_n} | J_{i_n} | \right) \psi_i(s') \, ds'$$

$$\Rightarrow \rho_i \psi_i(s) = \int_{\mathbb{R}} C(s, s') \psi_i(s') \, ds'$$  

An eigenvalue problem in microscale space
Comparison between the original microstructure and the reduced order one

Original

Mean(B) Mpa

Mean(G) Mpa

Mean(E) Mpa

Reconstructed

Mean(B) Mpa

Mean(G) Mpa

Mean(E) Mpa
Comparison between the original microstructure and the reduced order one

Original

Reconstructed
Numerical Example

Deterministic Solver: Multiscale forging FE simulation.

Macro

\[ \int_{B_n} \left( \langle P_r \rangle \cdot \nabla_{\tilde{u}} \right) dV_n = \int_{\Gamma_n} t \cdot \tilde{u} dA_n \]

- Formulation for macro scale
- Update macro displacements
- Macro-deformation gradient
- Homogenized stress

Meso

\[ \frac{\partial A(s, t)}{\partial t} + A(s, t) \nabla \cdot v(s, t) = 0 \]

- Texture evolution update
- Polycrystal averaging for macro-quantities
- Integration of single crystal slip and twinning laws

\[ \tau^\alpha = T \cdot S^\alpha \]
After the interpolation in the stochastic space for the texture have been obtained one can use them to obtain the realizations of the texture. Using these realizations statistics of the macro-scale properties can be obtained.
Define a complete probability space $\Omega, \mathcal{F}, \mathbb{P}$. We are interested to find stochastic functions $u : \Omega \times D \to \mathbb{R}$, $p : \Omega \times D \to \mathbb{R}$ and $S : \Omega \times [0,T] \times D \to \mathbb{R}$ such that for $\mathbb{P}$-almost everywhere (a.e.) $\omega \in \Omega$, the following holds:

\[
\nabla \cdot u(x, \omega) = 0, \quad u(x, \omega) = -K(x, \omega)\nabla p(x, \omega) \quad \forall x \in D
\]

\[
\frac{\partial S(x, t, \omega)}{\partial t} + u(x, t, \omega) \cdot \nabla S(x, t, \omega) = 0, \quad \forall x \in D, t \in [0, T]
\]

with the boundary condition

\[
p = \bar{p} \quad \text{on} \quad \partial D_p, \quad u \cdot n = 0 \quad \text{on} \quad \partial D_u.
\]

Length scale of the system,

Length scale of permeability variation,

**Multiscale paradigm**: Construct a strategy that solves for the coarse-scale features while accounting for the fine-scale features.

But exact permeability unknown. Some statistics or limited data.

Permeability is a realization from corresponding probability space $K(\omega, .) \in \Omega$.
For each collocation point, \( Y^m, m = 1, \ldots, M \), we have

\[
\nabla \cdot \mathbf{u}(x, Y^m) = 0
\]

\[
\mathbf{u}(x, Y^m) = -k(x, Y^m)\nabla p(x, Y^m)
\]

\[
\frac{\partial S(x, Y^m)}{\partial t} + \mathbf{u}(x, Y^m) \cdot \nabla S(x, Y^m) = 0
\]

For each collocation point, \( y, 1, \ldots, m \), we have

\[
\nabla \mathbf{u}_i(x) = 0, \quad \mathbf{u}_i(x) = -K \nabla \mathbf{p}_i(x), \quad \forall x \in E
\]

\[
\mathbf{u}_i \cdot \mathbf{n} = \frac{T_{\nu_a}}{\sum_{\nu_b \subset \Lambda} T_{\nu_b} |\nu_b|} \mathbf{q}_i, \quad \text{on } \Lambda \subset \partial E,
\]

Modified bilinear form

\[
A_h(K^{-1} \mathbf{u}_c, \mathbf{v}_c) := \sum_{E \in T_c} \sum_{k=1}^{4} \frac{T_k}{|E|} \int_E \mathbf{U}_k(x) \cdot K^{-1} \mathbf{V}_k(x) \, dx.
\]

Reconstruct fine scale velocity:

\[
\nabla \cdot \mathbf{u}_h = 0, \quad \mathbf{u}_h = -K \nabla \mathbf{p}_h, \quad \forall x \in E,
\]

\[
\mathbf{u}_h \cdot \mathbf{n} = \mathbf{q}_c, \quad \text{on } \Lambda \subset \partial E
\]
Solution Methodology

- Generate collocation point
- Generate the permeability sample given the collocation point, set coarse discretization
- Compute the stiffness matrix for each coarse element
- Compute the stochastic coarse-scale fluxes
- Solve the subgrid problems for each basis function at quadrature points
- Reconstruct the fine-scale velocity
- Solve the subgrid problems with coarse-scale flux
- Solve the transport problem
- POSTPROCESSING: Compute the statistics of the solution
- Solve stochastic multiscale problem with HDMR
- Return function value at collocation point
Example

Fine-scale grid: 80×80

Coarse-scale scale grid: 20×20

\[ \mathbf{Y}(\omega) = \log (K(\omega)) = \sum_{i=1}^{N} \sqrt{\lambda_i} \phi_i(\mathbf{x}) Y_i \]

\[ \text{Cov}(\mathbf{x}, \mathbf{y}) = \sigma^2 \exp \left( -\frac{|x_1 - y_1|}{L} - \frac{|x_2 - y_2|}{L} \right) \]

where \( \mathbf{Y} \) is a zero mean random field with covariance function

\( L \) is the correlation length and \( \sigma \) is the standard deviation.

\[ L = 0.1, \quad \sigma = 1.0 \Rightarrow N = 500 \]

We truncate the stochastic dimension after 500 terms and we assume the random variables to be uniform in [-1,1].
It is interesting to note that the mean is nearly the same as the homogeneous solution with the mean permeability. This is called “heterogeneity-induced dispersion”, where the heterogeneity smoothes the water saturation problem in the ensemble sense, although individual realizations show heterogeneity.
The results indicate that higher water saturation variance are concentrated near displacement fronts, which are areas of steep saturation gradients.

The number of component functions is 1281 and the total number of collocation points is 36087, while there are $10^5$ MC samples.
MC method with fine scale solver

HDMR method with multiscale solver

➢ The saturation front has the largest variation occurred.
Nonlinear POD for SPDES: KPCA on output model

- The basic idea is similar to input model reduction.
- We perform traditional POD in the feature space through the kernel transformation.
- Then we solve the reduced model in the feature space. However, in general, the governing equation is only known in the input space and is not known in the feature space!
- We learn the governing equations in the feature space through the known feature points.
- As mentioned, the feature space is assumed to have better linear variation than the original input space. Thus, we can assume the feature points follow a stochastic linear kernel dynamical model:

\[ \Phi(u_{t+1}) = A\Phi(u_t) \]

where A is the transition matrix.
KPCA on output model reduction

- If only keeping the first largest $m$ eigenvalues, any nonlinear mapping of the random snapshot $u$ can be expanded in the feature space as:

$$\Phi(u) = \sum_{i=1}^{m} \zeta_m V_m$$

- We plug the above equation into the linear dynamical model and take the kernel projection, we will get a linear $m$-dimensional dynamical model with the unknown coefficients $\zeta := (\zeta_1, \ldots, \zeta_m)$:

$$\zeta_{t+1} k(V_{t+1}, V_j) = A \zeta_t k(V_t, V_j), j = 1, \ldots, m$$

- The only thing left is to determine the matrix $A$ given the available random snapshots. We can solve it through:

$$k(u_i, u_{t+1}) := \int_{\Omega} \langle \Phi(u_i) \cdot \Phi(u_{t+1}) \rangle \, dx = A \int_{\Omega} \langle \Phi(u_i) \cdot \Phi(u_t) \rangle \, dx$$

- So for any initial condition, we can find its projection onto the random basis in the feature space and then use this initial value to solve the time evolution problem in the feature space.
Related Research in Progress

- Curse of Dimensionality for stochastic PDEs
  - Higher-order hierarchical basis functions
  - Sparse kernel machines (very promising)
  - ....

- Non-linear stochastic input models:
  - IsoTop
  - Kernel PCA methods
  - ....

- Non-linear data-driven model reduction for the solution of stochastic PDEs:
  - Kernel PCA
  - Manifold Learning
  - Multiscale, biorthogonal KLE decomposition coupled with KPCA
  - ....