RAVEN Workshop

Multi-Step Input Reduction

Nuclear Engineering Methods Development Department
Idaho National Laboratory

Discussion Points

Overview

Overview: Uncertainty Quantification

Benefits:

- Quantity of Interest Variance
- Failure Probabilities
- Limit Surface Construction
- Design of Experiment

Overview: Session Goal

Reduce high-dimension input spaces

- Use PCA to eliminate correlated inputs
- Use sensitivity to eliminate low-impact inputs
- Accurate UQ on a reduced input space

Overview: Assumptions

- Simulation codes are expensive to run
- All inputs are initially perturbable
- Saving time and money is good

Motivations

Motivations: Sampling Strategies

Two classes of forward sampling strategies in RAVEN:

- Structured (Orthogonal Grids, Limit Surface, Latin Hypercube)
- Unstructured (Monte Carlo)

Motivations: Unstructured Sampling

Traditional Monte Carlo sampling

- Agnostic of Dimensionality
- Consistent, but slow convergence

Motivations: Structured Sampling

`RAVEN` has several structured solvers

- Orthogonal Grids
- Sparse Grid
- Stratified
- Limit Surface Search

All of these suffer from the Curse of Dimensionality

Motivations: Proposed Solution

Remove unnecessary input dimensions

- Correlated inputs have redundancy
- Low-impact inputs aren't useful to perturb

Perform UQ on reduced space

Methods

Methods: Procedure Overview

- (optional) Benchmark original problem
- Perform PCA on original input space
- Truncate to essential latent variables
- Perform global sensitivity analysis
- Truncate latent variables to exclude non-essential

Methods: Principal Component Analysis

Methods: Principal Component Analysis

Used to orthogonalize input space
Represent input space as sum of "latent" variables $\boldsymbol{L}$

$$\boldsymbol{M} = \boldsymbol{Q} \cdot \boldsymbol{L}, \tag{1}$$

where

- $\boldsymbol{M}$ is the vector of original variables, $|M| \times 1$,
- $\boldsymbol{Q}$ is the PCA transformation matrix, $|M| \times |L|$,
- $\boldsymbol{L}$ is the vector of latent variables, $|L| \times 1$

All $\boldsymbol{L}$ are distributed as standard normal distributions

Methods: Principal Component Analysis

$$\boldsymbol{M} = \boldsymbol{Q} \cdot \boldsymbol{L}, \tag{2}$$

Perform eigenvalue decomposition of covariance matrix and rank eigenvalues
Truncate latent variables at desirable eigenvalue

Methods: PCA in RAVEN

Add transformation to `Sampler`:

```
<Samplers>
  <MonteCarlo name='mc'>
    ... TODO add variables with dimensions
    <variablesTransformation distribution='MultivariateNormalReduction'>
      <latentVariables>y1,y2,y3,y4,y5,y6,y7</latentVariables>
      <manifestVariables>x1,x2,x3,x4,x5,x6,x7</manifestVariables>
      <method>pca</method>
  </MonteCarlo>
</Samplers>
```

Methods: PCA in RAVEN

Use a PostProcessor to extract the rankings
PostProcessor

```
<Models>
  <PostProcessor TODO>
  </PostProcessor>
</Models>
```

Step

```
<Steps>
  <PostProcessor name='TODO'>
  </PostProcessor>
</Steps>
```

Use results from file TODO to pick PCA truncation

Methods: PCA in RAVEN

Once reduction level is decided, change sampler truncation

```
<Samplers>
  <MonteCarlo name='mc'>
    ... TODO add variables with dimensions
    <variablesTransformation distribution='MultivariateNormalReduction'>
      <latentVariables>y1,y2,y3,y4</latentVariables>
      <manifestVariables>x1,x2,x3,x4,x5,x6,x7</manifestVariables>
      <method>pca</method>
  </MonteCarlo>
</Samplers>
```

Methods: Global Sensitivity Analysis

Now that we've reduced the input, we can do global sensitivity analysis

- Pearson Correlation Coefficients
- Spearman Rank Coefficients
- Sobol Sensitivity Coefficients

Pearson and Spearman can be sampled using forward samplers
Linear Sobol expansion often more efficient (1 run per latent variable)

Methods: Global Sensitivity Analysis

Pearson Coefficients

- Global correlation between dimensions
- Can be input-input, input-output, or output-output
- Can be calculated using most RAVEN Samplers

Methods: Global Sensitivity Analysis

More on Sobol sampler and HDMRRom in Collocation workshop!
Sobol sampler

```
<Samplers>
  <Sobol name='sobol'>
    ... TODO add variables with dimensions
    <ROM class='Models' type='HDMRRom'>rom</ROM>
    <variablesTransformation distribution='MultivariateNormalReduction'>
      <latentVariables>y1,y2,y3,y4</latentVariables>
      <manifestVariables>x1,x2,x3,x4,x5,x6,x7</manifestVariables>
      <method>pca</method>
  </Sobol>
</Samplers>
```

Methods: Global Sensitivity Analysis

HDMRRom

```
<Models>
  <ROM name='rom' subType='HDMRRom'>
    <SobolOrder>1</SobolOrder>
    <PolynomialOrder>1</PolynomialOrder>
    <IndexSet>TotalDegree</IndexSet>
    <Features>y1,y2,y3</Features>
    <Targets>QoI</Targets>
    ... TODO
  </ROM>
</Models>
```

ROM Output

```
<OutStreams>
  <Print name='print_rom'>
    <type>xml</type>
    <source>rom</source>
    <what>indices</what>
  </Print>
</OutStreams>
```

Methods: Global Sensitivity Analysis

Steps

```
<Steps>
  <MultiRun name='sample'>
    <Input class='Files' type=''>infile</Input>
    <Sampler class='Sampelrs' type='Sobol'>sobol</Sampler>
    <Model class='Models' type='ROM'>rom</Models>
    <Output class='DataObjects' type='PointSet'>training_data</Output>
  </MultiRun>
  <RomTrainer name='train'>
    <Input class='DataObjects' type='PointSet'>training_data</Input>
    <Output class='Models' type='ROM'>rom</Input>
  </RomTrainer>
</Steps>
```

Methods: Global Sensitivity Analysis

Output from ROM

```
<ROM>
  <QoI>
    <indices>
      <tot\_variance>TODO</tot\_variance>
      <variables>TODO
        <partial\_variance>TODO</partial\_variance>
        <Sobol\_index>TODO</Sobol\_index>
      </variables>
    </indices>
  </QoI>
</ROM>
```

Methods: Twice-Reduced Analysis

Change dimensions in Sampler to prioritize

```
TODO sampler
```

Methods: Twice-Reduced Analysis

Now what?

- Make a ROM
- Perform surrogate statistics sampling
- Compute limit surface on surrogate

Full Example

Example case: Cross Section Model

- 308 Correlated Cross Sections (SCALE)
- Simulation model is polynomial combinations
- PCA Reduction: 308 to 50
- Sensitivity Reduction: 50 to 9

See included examples:

- Original Monte Carlo benchmark: `run_mc_orig.xml`
- PCA reduction: `run_mc_pca.xml`
- PCA output: `first/TODO`
- Sensitivity reduction: `run_mc2_[#].xml`
- Sensitivity output: `first/sobol_dump.xml`
- ROM with twice-reduced space: `sc_td[#].xml`
- Twice-reduced output: `td_[#]_dump.xml`

Full Example: Results

TODO RESULTS PLOTS