Accelerating Forward Uncertainty Propagation

Roy H. Stogner
Vikram Garg

Institute for Computational Engineering and Sciences
The University of Texas at Austin

October 18, 2012
Outline

1. Notation

2. Enhanced Model Evaluation
   - Goal-Oriented Adaptivity

3. Enhanced Monte Carlo
   - Enhanced Sampling Strategies
   - Enhanced Estimators
   - Results
<table>
<thead>
<tr>
<th>Primal/Forward Problem</th>
</tr>
</thead>
</table>
| **Parameter value(s)** | $\xi \in \Xi^\mathbb{R} \subset \Xi \equiv \mathbb{R}^{N_p}$  
| **System solution**    | $\tilde{u}(x, t; \xi) \in U^\mathbb{R} \subset U$  
| **Primal weighted residual** | $\mathcal{R}(\tilde{u}(\xi), v; \xi) \equiv 0 \ \forall v \in V$  
| **Quantity of interest functional** | $Q : U^\mathbb{R} \times \Xi^\mathbb{R} \rightarrow \mathbb{R}$  
| **Quantity of interest output** | $q = Q(\tilde{u}(\xi), \xi)$ |
### Notation

#### Primal/Forward Problem

<table>
<thead>
<tr>
<th>Parameter value(s)</th>
<th>( \xi \in \Xi^R \subset \Xi \equiv \mathbb{R}^{N_p} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>System solution</td>
<td>( \tilde{u}(x, t; \xi) \in U^R \subset U )</td>
</tr>
<tr>
<td>Primal weighted residual</td>
<td>( \mathcal{R}(\tilde{u}(\xi), v; \xi) \equiv 0 \quad \forall v \in V )</td>
</tr>
<tr>
<td>Quantity of interest functional</td>
<td>( Q : U^R \times \Xi^R \rightarrow \mathbb{R} )</td>
</tr>
<tr>
<td>Quantity of interest output</td>
<td>( q = Q(\tilde{u}(\xi), \xi) )</td>
</tr>
</tbody>
</table>

- Highly nonlinear forward solves
- Large \( N_p \)-dimensional parameter space
- Few quantities of interest
## Notation

### Adjoint Problem

<table>
<thead>
<tr>
<th>Adjoint solution</th>
<th>( \tilde{z}(x, t; \xi) \in V )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjoint equation</td>
<td>( \mathcal{R}_u(\tilde{u}, \tilde{z}; \xi)(u) \equiv Q_u(\tilde{u}; \xi)(u) \ \forall u \in U )</td>
</tr>
<tr>
<td>Adjoint-based sensitivities</td>
<td>( \frac{dq}{d\xi} = Q_\xi(\tilde{u}; \xi) - \mathcal{R}_\xi(\tilde{u}, \tilde{z}; \xi) )</td>
</tr>
<tr>
<td>Adjoint-based error</td>
<td>( Q(\tilde{u}^h) - Q(\tilde{u}) = \mathcal{R}(\tilde{u}^h, \tilde{z}; \xi) + H.O.T. )</td>
</tr>
</tbody>
</table>
## Notation

### Adjoint Problem

<table>
<thead>
<tr>
<th>Adjoint solution</th>
<th>$\tilde{z}(x, t; \xi) \in V$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjoint equation</td>
<td>$\mathcal{R}_u(\tilde{u}, \tilde{z}; \xi)(u) \equiv Q_u(\tilde{u}; \xi)(u) \quad \forall u \in U$</td>
</tr>
<tr>
<td>Adjoint-based sensitivities</td>
<td>$\frac{dq}{d\xi} = Q_\xi(\tilde{u}; \xi) - \mathcal{R}_\xi(\tilde{u}, \tilde{z}; \xi)$</td>
</tr>
<tr>
<td>Adjoint-based error</td>
<td>$Q(\tilde{u}^h) - Q(\tilde{u}) = \mathcal{R}(\tilde{u}^h, \tilde{z}; \xi) + H.O.T.$</td>
</tr>
</tbody>
</table>

- Linear, relatively cheap adjoint solves
- One linear solve per QoI
- Sensitivity: One dot product per parameter
- Error: One weighted residual evaluation per QoI
Enhanced Model Evaluation
Goal-Oriented Adaptivity

Adjoint Residual Error Indicator

Error Indicators

- Efficiently bounding $e_Q$ via per-element terms
- From our error estimator,

$$R(u^h, z; \xi) = \sum_{E} R^E(u^h|_E, z|_E; \xi)$$

- $z^h$ is cheaper than higher order approximation of $z$;
- $z^h$ is already needed for sensitivity calculations

Ignoring higher order terms:

$$q - q^h = -R_u(u, z - z^h; \xi)(\tilde{u} - u^h)$$

$$\left| q - q^h \right| \leq \sum_{E} \left| R_u^E \right|_{B(U^E, V^{E\times})} \left| \tilde{u}|_E - u^h |_E \right|_{U^E} \left| z|_E - z^h |_E \right|_{V^E}$$
Adjoint Residual Error Indicator

\[
\eta_E \equiv \left\| \tilde{u} |_E - \tilde{u}^h |_E \right\|_{\mathbf{U}_E} \left\| \tilde{z} |_E - \tilde{z}^h |_E \right\|_{\mathbf{V}_E}
\]
(neglecting \(\left\| \mathcal{R}_u E \right\|\))
Norm vs QoI Based Refinement

- Global error indicator ignores QoI sensitivity, plateaus
- Rapid convergence from adjoint-based refinement
Multiphysics and Adjoint Weighting

Where does $\| R^E_u \|_{B(U^E, V^{E*})}$ become non-negligible?

- Spatially varying constitutive parameters, nonlinearities
- Variable scaling in dimensionalized multiphysics problems
- Interaction pattern between multiphysics variables
Enhanced Model Evaluation | Goal-Oriented Adaptivity

Multiphysics and Adjoint Weighting

Where does \( \| R^E_u \|_{B(U^E, V^{E*})} \) become non-negligible?

- Spatially varying constitutive parameters, nonlinearities
- Variable scaling in dimensionalized multiphysics problems
- Interaction pattern between multiphysics variables

Generalized Adjoint Residual Error Indicator

\[
\left| Q(\tilde{u}^h) - Q(\tilde{u}) \right| \leq \left| R(\tilde{u}^h, \tilde{z} - \tilde{z}^h)(\tilde{u} - \tilde{u}^h) \right| \\
\leq \sum_{E} \left| \tilde{z}_i - \tilde{z}_i^h \right|_i M_{ij} \left| \tilde{u}_j - \tilde{u}_j^h \right|_j
\]

Choose \( M, ||\cdot||_i, ||\cdot||_j \) to match physics
Multiphysics Adjoint Weighting Example

Stokes Flow

\[
Q(\tilde{u}) - Q(\tilde{u}^h) = \int_{\Omega} \nabla (\tilde{u} - \tilde{u}^h) \cdot \nabla (\tilde{u}^* - \tilde{u}^{*h}) \, dx \\
- \int_{\Omega} \nabla \cdot (\tilde{u} - \tilde{u}^h) (p^* - p^{*h}) \, dx - \int_{\Omega} \nabla \cdot (\tilde{u}^* - \tilde{u}^{*h}) (p - p^h) \, dx
\]

Matrix Form

\[
Q(\tilde{u}) - Q(\tilde{u}^h) \leq \begin{bmatrix}
    e(u_1) \\
    e(u_2) \\
    e(p)
\end{bmatrix}^T \begin{bmatrix}
    1 & 0 \\
    0 & 1 \\
    1 & 1
\end{bmatrix} \begin{bmatrix}
    e(u_1^*) \\
    e(u_2^*) \\
    e(p^*)
\end{bmatrix}
\]

+ \begin{bmatrix}
    e(u_{1,1}) \\
    e(u_{2,2}) \\
    e(p)
\end{bmatrix}^T \begin{bmatrix}
    0 & 0 & 1 \\
    0 & 0 & 1 \\
    1 & 1 & 0
\end{bmatrix} \begin{bmatrix}
    e(u_{1,1}^*) \\
    e(u_{2,2}^*) \\
    e(p^*)
\end{bmatrix}
\]
Multiphysics Adjoint Weighting Example

- Stokes flow + transport, corner singularities
- Naive weighting “staggars”
- Matrix weighting converges smoothly to penalty BC precision
## Notation

### Uncertainty Propagation Problems

<table>
<thead>
<tr>
<th>Input PDF</th>
<th>$\xi \sim p^{\xi}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expected (mean) output</td>
<td>$\bar{q} \equiv \mathbb{E}[q] \equiv \int_{\Xi} q(\xi) dp^{\xi}$</td>
</tr>
<tr>
<td>Output standard deviation</td>
<td>$\sigma_q \equiv \sqrt{\mathbb{E}[(q - \bar{q})^2]}$</td>
</tr>
<tr>
<td>Output risk</td>
<td>$P_C \equiv \mathbb{E}[q \in C_q]$</td>
</tr>
</tbody>
</table>

*Integrals in high dimensional spaces $R^N$*

*"Curse of dimensionality"*

*Deterministic integration cost exponential in $N$*

---

Stogner & Garg  
Uncertainty Propagation  
October 18, 2012  
11 / 31
Notation

Uncertainty Propagation Problems

<table>
<thead>
<tr>
<th>Input PDF</th>
<th>$\xi \sim p^\xi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expected (mean) output</td>
<td>$\bar{q} \equiv \mathbb{E}[q] \equiv \int_{\Xi} q(\xi) d\xi , p^\xi \xi$</td>
</tr>
<tr>
<td>Output standard deviation</td>
<td>$\sigma_q \equiv \sqrt{\mathbb{E}[(q - \bar{q})^2]}$</td>
</tr>
<tr>
<td>Output risk</td>
<td>$P_C \equiv \mathbb{E}[q \in C_q]$</td>
</tr>
</tbody>
</table>

- Integrals in high dimensional spaces $\mathbb{R}^N_P$
- “Curse of dimensionality”
  - Deterministic integration cost exponential in $N_P$
- Monte Carlo: dimension independent
Monte Carlo Integration

Errors

- \( q_{MC}^M - \bar{q} \) variance: \( \sigma^2_{e[\bar{q}]} = \frac{\sigma^2}{N_s} \)
- \( (\sigma^2_q) - \sigma_q^2 \) variance: \( \sigma^2_{e[\sigma^2]} = \sigma^4 \left( \frac{2}{N_s-1} + \frac{\kappa}{N_s} \right) \)
- \( P_{MC}^M - P_C \) variance: \( \sigma^2_{e[P_C]} = \frac{P_C - P_C^2}{N_s} \)
Monte Carlo Integration

Errors

- $\bar{q}^{MC} - \bar{q}$ variance: $\sigma_{e[\bar{q}]}^2 = \frac{\sigma^2}{N_s}$
- $(\sigma_{q}^{MC})^2 - \sigma_q^2$ variance: $\sigma_{e[\sigma^2]}^2 = \sigma^4 \left( \frac{2}{N_s-1} + \frac{\kappa}{N_s} \right)$
- $P_C^{MC} - P_C$ variance: $\sigma_{e[P_C]}^2 = \frac{P_C - P_C^2}{N_s}$

- Sampling-based scheme errors are PDFs
- Errors based on variance $\sigma^2$ of sampled entity
- Error “bound” width: $\sigma_e \propto N_s^{-1/2}$
- Importance sampling improves convergence constant, not rate
Latin Hypercube Sampling

Algorithm
- Quantiles in each parameter
- Samples permuted to bins
  - Reducing correlations?
- Random sample placement within bins

Uses
- Reduces variance from additive components
- Higher-order convergence for separable functions
Hierarchical Latin Hypercube Sequences

- LHS improves convergence but not naturally incremental
- HLHS based methods offer more flexible incrementation

(a) Hierarchical Latin Hypercubes

(b) Tree Structure

i = 5, 25 samples each
i = 4, 24 samples each
i = 3, 23 samples each
i = 2, 22 samples each
Numerical Experiments

- Correlation reduction used for both ILHS and HLHS
- HILHS points show finer grained increments

\[ Q(\xi) = \text{round} \left( \sum_{i=1}^{N_P=16} \xi_i \right) \]

\[ Q(\xi) = e^{\sum_{i=1}^{N_P=16} \xi_i} \]
Control Variate

Known Surrogate

- Quantity of interest $q$ has correlated surrogate statistic $q_s$
- “Known” mean $\bar{q}_s \equiv \mathbb{E}[q_s]$

$$c_{qq_s} \equiv \frac{\mathbb{E}[(q - \bar{q})(q_s - \bar{q}_s)]}{\sigma_q \sigma_{q_s}}$$

Unbiased, Variance-reduced estimator

$$\mathbb{E}[q] = \mathbb{E}[q - \alpha q_s] + \mathbb{E}[\alpha q_s] = \mathbb{E}[s]$$

$$s \equiv q - \alpha q_s + \alpha \bar{q}_s$$

$$\sigma_s^2 = \sigma_q^2 - 2\alpha c_{qq_s} \sigma_q \sigma_{q_s} + \alpha^2 \sigma_{q_s}^2$$

Integrating $s$ via MC sampling, $\alpha \equiv 1$, $q_s \rightarrow q$ gives $\sigma_{e[q]} \rightarrow 0$. 
Sensitivity Derivative Enhancement

SDEMC Surrogate

- Justification: Adjoint are cheap
- One linearization at input mean
  - Adds one forward, one sensitivity solve
- \( q_s(\xi) \equiv q(\bar{\xi}) + q_\xi(\bar{\xi})(\xi - \bar{\xi}) \)
Local Sensitivity Derivative Enhancement

LSDEMC Surrogate

- Take any input sample set
- Evaluate forward and adjoint at each
  - Adds one sensitivity solve per sample
- Linearize around each nearby sample
Bias

Problem

- Use every sample to construct surrogate?
- \[ \mathbb{E} \left[ \sum_{i=1}^{N_s} q(\xi_i) - q_s(\xi_i) \right] = 0 \]
- \[ \mathbb{E} \left[ \bar{q}^L \right] = \bar{q}_s^L \neq \bar{q} \]
- May have systemic bias for any problem
- Huge error for high-dimensional benchmark problems
Bias

Problem

- Use every sample to construct surrogate?
- \[ E \left[ \sum_{i=1}^{N_s} q(\xi_i) - q_s(\xi_i) \right] = 0 \]
- \[ E \left[ \bar{q}^L \right] = \bar{q}_s \neq \bar{q} \]
- May have systemic bias for any problem
- Huge error for high-dimensional benchmark problems

Solution

- Subdivide sample set (e.g. 2, 4, 8 subsets; HLHS applicable)
- Use one subset to construct surrogate, remainder to integrate bias
- Repeat for all subsets; average.
Bias

Unbiased LSDEMC Estimator

\[
\frac{1}{N_R} \sum_{r=1}^{N_R} \left( \frac{1}{N_{ss}} \sum_{l=1}^{N_{ss}} Q_{r,1}(\xi_l) + \frac{1}{N_s - \frac{N_s}{N_R}} \sum_{c=1}^{N_s - \frac{N_s}{N_R}} (Q(\xi_c) - Q_{r,1}(\xi_c)) \right)
\]
Convergence

Proofs
- Holding $N_s/N_R$ constant
- Assuming sufficiently smooth $Q$
- LSDEMC converges asymptotically faster than Monte Carlo

Issues
- How many subsets is optimal?
  - More subsets == larger $N$ integrating bias term
  - Fewer subsets == smaller $\sigma$ in bias term
  - Numerical results show dimension dependence
- What convergence rate to expect?
## Visualizing 250 Thousand Trials

### Methodology
- Dependent variable: mean(abs(error)) from 256 trials per case
- 32, 128, 512 true sample evaluations per trial
- 32, 128, 512 surrogate samples per true sample
- Statistics: mean, standard deviation
- Sampling: SRS, HLHS
- Methods: MC, SDEMC, LSDEMC2/4/8
Visualizing 250 Thousand Trials

Methodology

- Dependent variable: mean(abs(error)) from 256 trials per case
- 32, 128, 512 true sample evaluations per trial
- 32, 128, 512 surrogate samples per true sample
- Statistics: mean, standard deviation
- Sampling: SRS, HLHS
- Methods: MC, SDEMC, LSDEMC2/4/8

Graph Key

- Purple: MC, Red: SDEMC
- Blue/Cyan/Green: LSDEMC2/4/8
- X: SRS, O: LHS
Benchmark Problem: Smooth

Normal → Lognormal Distribution

\[ \xi \equiv \left( \xi_1, \ldots, \xi_{N_p} \right) \]
\[ q(\xi) \equiv e^{\sum_i \xi_i} \]
\[ \xi_i \sim \mathcal{N} \left( \frac{\mu}{N_p}, \frac{\sigma}{\sqrt{N_p}} \right) \]
\[ q \sim \text{Log-N} \left( \mu, \sigma \right) \]

- \( \mu \equiv 1, \ \sigma \equiv 1 \)
- Arbitrary dimensionality \( N_p \)
- All parameters equally significant
- Simple analytic exact solution moments
- Variance, MC error independent of \( N_p \)
1 Parameter

Forward UQ Convergence: Exp Benchmark, Mean

Error in Approximated Output
Number of Forward Evaluations

Forward UQ Convergence: Exp Benchmark, Mean

MC+SRS
SDEMC+SRS
LSDEMC2+SRS
LSDEMC4+SRS
LSDEMC8+SRS
MC+LHS
SDEMC+LHS
LSDEMC2+LHS
LSDEMC4+LHS
LSDEMC8+LHS

Stogner & Garg
Uncertainty Propagation
October 18, 2012 24 / 31
4 Parameters

Forward UQ Convergence: Exp Benchmark, Mean

Error in Approximated Output vs. Number of Forward Evaluations for different methods:
- MC+SRS
- SDEMC+SRS
- LSDEMC2+SRS
- LSDEMC4+SRS
- LSDEMC8+SRS
- MC+LHS
- SDEMC+LHS
- LSDEMC2+LHS
- LSDEMC4+LHS
- LSDEMC8+LHS

Stogner & Garg Uncertainty Propagation
October 18, 2012
16 Parameters

Forward UQ Convergence: Exp Benchmark, Mean

- MC+SRS
- SDEMC+SRS
- LSDEMC2+SRS
- LSDEMC4+SRS
- LSDEMC8+SRS
- MC+LHS
- SDEMC+LHS
- LSDEMC2+LHS
- LSDEMC4+LHS
- LSDEMC8+LHS
64 Parameters

Forward UQ Convergence: Exp Benchmark, Mean

Error in Approximated Output vs. Number of Forward Evaluations

Stogner & Garg
Uncertainty Propagation
October 18, 2012 27 / 31
Benchmark Problem: $C^0$ abs

Normal $\rightarrow$ Folded Normal Distribution

$$\xi \equiv (\xi_1, \ldots, \xi_{N_p})$$

$$q(\xi) \equiv \left| \sum_i \xi_i \right|$$

$$\xi_i \sim \mathcal{N} \left( \frac{\mu}{N_p}, \frac{\sigma}{\sqrt{N_p}} \right)$$

$$q \sim \mathcal{F} - \mathcal{N} (\mu, \sigma)$$

- $\mu \equiv 1, \sigma \equiv 1$
- Same benefits as previous benchmark
- Response function now piecewise linear
- Differentiable except on one hyperplane
- Derivative defined as $\bar{0}$ there
4 Parameters

Forward UQ Convergence: Abs Benchmark, Mean

- MC+SRS
- SDEMC+SRS
- LSDEMC2+SRS
- LSDEMC4+SRS
- LSDEMC8+SRS
- MC+LHS
- SDEMC+LHS
- LSDEMC2+LHS
- LSDEMC4+LHS
- LSDEMC8+LHS
Enhanced Monte Carlo Results

64 Parameters

Forward UQ Convergence: Abs Benchmark, Mean

Stogner & Garg
October 18, 2012
Future Improvements

Adaptivity
- Regularization of “peak value” QoIs

HILHS
- Correlation Reduction improvements

LSDEMC
- Stochastic Voronoi moments lemma
- Anisotropic Voronoi metric
- Smooth (multi-sample-based) surrogates

Questions?