Convergent Multifidelity Optimization using Bayesian Model Calibration

13th AIAA/ISSMO Multidisciplinary Analysis Optimization Conference
September 14, 2010

Andrew March & Karen Willcox
**Motivation**

- Engineering problems often have objective functions or constraints that are “expensive” to evaluate
- Derivatives are commonly not available and may not be easy to estimate accurately
Multifidelity Surrogates

- **Definition: High-Fidelity**
  - The best model of reality that is available and affordable, the analysis that is used to validate the design.

- **Definition: Low(er)-Fidelity**
  - A method with unknown accuracy that estimates metrics of interest but requires lesser resources than the high-fidelity analysis.
Main Messages

- Bayesian model calibration offers an efficient framework for multifidelity optimization.
- Does not require high-fidelity gradient estimates.
- Can reduce the number of high-fidelity function evaluations compared with other multifidelity methods, even those using gradients.
- Provides a flexible and robust alternative to nesting when there are multiple low-fidelity models.
Motivation-Calibration Methods

- First-order consistent trust-region methods:
  - Efficient when derivatives are available or can be approximated efficiently
  - Calibrated surrogate models are only used for one iteration

- Pattern-search methods:
  - High-fidelity information can be reused
  - Can be slow to converge

- Bayesian calibration methods (e.g., Efficient Global Optimization)
  - Reuse high-fidelity information from iteration to iteration
  - Can be quite efficient in practice
  - Heuristic, no guarantee they converge to an optimum
Bayesian Model Calibration

- Define a surrogate model of the high-fidelity function:
  \[ m_k(x) = f_{low}(x) + e_k(x) \approx f_{high}(x) \]

- The error model, \( e(x) \):
  - Is a radial basis function model
  - Interpolates \( f_{high}(x) - f_{low}(x) \) exactly at all selected calibration points
  - Based on Wild et al. 2009

- Convergence can be proven if surrogate model is fully linear within a trust region

- Define trust region at iteration \( k \):
  \[ B_k = \{ x \in \mathbb{R}^n : \| x - x_k \| \leq \Delta_k \} \]
Combining Multiple Lower-Fidelities

- Calibrate all lower-fidelity models to the high-fidelity function using radial basis function error model.

- Use a maximum likelihood estimator to predict the high-fidelity function value (Essentially a Kalman Filter).
Definition: Fully Linear Model

- Definition: For all \( x \) within a trust region of size \( \Delta_k \in (0, \Delta_{\text{max}}] \), a fully linear model, \( m_k(x) \), satisfies
  \[ \left\| \nabla f_{\text{high}}(x) - \nabla m_k(x) \right\| \leq \kappa_g \Delta_k \]
  for a Lipschitz constant \( \kappa_g \), and
  \[ \left| f_{\text{high}}(x) - m_k(x) \right| \leq \kappa_f \Delta_k^2 \]
  with a Lipschitz constant \( \kappa_f \).

- Fully linear model error bounds:
**Function Evaluation Points**

- RBF model has sufficient local behavior to guarantee convergence
- Considerable reuse of high-fidelity information
  - It captures some global behavior
- First-order trust region approaches only look at the center of the current trust region
2 Constrained Formulations

Method 1:
- High-fidelity objective without available derivatives subject to constraints with available derivatives:

\[
\min_{x \in \mathbb{R}^n} f_{\text{high}}(x) \quad \text{Derivatives unavailable}
\]

\[
s.t. \quad g(x) \leq 0
\]
\[
\quad h(x) = 0
\]

- Fully linear surrogates
  - Objective Function

- Constraints
  - Penalty method initially
  - Explicitly at termination

Method 2:
- High-fidelity objective subject to a high-fidelity constraint and constraints with available derivatives:

\[
\min_{x \in \mathbb{R}^n} f_{\text{high}}(x) \quad \text{Derivatives unavailable}
\]

\[
s.t. \quad g(x) \leq 0
\]
\[
\quad h(x) = 0
\]
\[
\quad c_{\text{high}}(x) \leq 0
\]

- Fully linear surrogates
  - Objective Function
  - High-fidelity constraint

- Constraints
  - Method 1 to find a feasible starting point
  - Interior point method to find optimum
Method 1: (Constraint Derivatives Available)

\[
\min_{x \in \mathbb{R}^n} f_{\text{high}}(x) \\
\text{s.t.} \quad g(x) \leq 0 \\
\quad h(x) = 0
\]

Derivatives available

Derivatives unavailable
**Method Summary**

- **Quadratic penalty function:**
  \[
  \hat{\phi}(\mathbf{x}, \sigma_k) = m_k(\mathbf{x}) + \frac{\sigma_k}{2} \left[ \mathbf{h}(\mathbf{x})^T \mathbf{h}(\mathbf{x}) + \mathbf{g}^+(\mathbf{x})^T \mathbf{g}^+(\mathbf{x}) \right]
  \]

- **Two trust-region subproblems:**
  \[
  \min_{\mathbf{s}_k \in \mathbb{R}^n} m_k(\mathbf{x}_k + \mathbf{s}_k) \\
  \text{s.t.} \quad g(\mathbf{x}_k + \mathbf{s}_k) \leq 0 \\
  \quad h(\mathbf{x}_k + \mathbf{s}_k) = 0 \\
  \quad \|\mathbf{s}_k\| \leq \Delta_k
  \]

- **Impose a sufficient decrease condition:**
  \[
  \lim_{k \to \infty} \Delta_k = 0
  \]

- **Termination:**
  - Fully linear model: \(\|\nabla f_{\text{high}}(\mathbf{x}) - \nabla m_k(\mathbf{x})\| \to 0; \ |f_{\text{high}}(\mathbf{x}) - m_k(\mathbf{x})| \to 0\)
  - Subproblem 1 is nearly equivalent to the original problem
**Method 2:** *(Constraint Derivatives Unavailable)*

\[
\min_{x \in \mathbb{R}^n} f_{\text{high}}(x)
\]

\[
s.t. \quad g(x) \leq 0
\]

\[
\quad h(x) = 0
\]

\[
\quad c_{\text{high}}(x) \leq 0
\]

- Finding a feasible starting point
- Finding a high-fidelity optimum
Finding a Feasible Starting Point

- Two fully linear surrogate models:
  \[ m_k(x) \approx f_{\text{high}}(x) \quad \overline{m}_k(x) \approx c_{\text{high}}(x) \]

- Find an initial feasible point:
  \[
  \begin{align*}
  \min_{x \in \mathbb{R}^n} & \quad c_{\text{high}}(x) \\
  \text{s.t.} & \quad g(x) \leq 0 \\
  & \quad h(x) = 0
  \end{align*}
  \]

- Constraint may not be bounded from below:
  \[
  \begin{align*}
  \min_{x \in \mathbb{R}^n} & \quad \max\{c_{\text{high}}(x) + d, 0\}^2 \\
  \text{s.t.} & \quad g(x) \leq 0 \\
  & \quad h(x) = 0
  \end{align*}
  \]

- Only need to iterate until \( c_{\text{high}}(x) \leq 0 \) and other constraints satisfied
Finding the Optimum

- Trust region subproblem:
  \[
  \min_{s_k \in \mathbb{R}^n} m_k(x_k + s_k) \\
  \text{s.t. } g(x_k + s_k) \leq 0 \\
  \quad h(x_k + s_k) = 0 \\
  \quad m_k(x_k + s_k) \leq 0 \\
  \quad \|s_k\| \leq \Delta_k
  \]

- Trial point acceptance:
  \[
  x_{k+1} = \begin{cases} 
  x_k + s_k & f_{\text{high}}(x_k) \geq f_{\text{high}}(x_k + s_k) \text{ and } c_{\text{high}}(x_k + s_k) \leq 0 \\
  x_k & \text{otherwise}
  \end{cases}
  \]

- Trust region size update:
  \[
  \Delta_{k+1} = \begin{cases} 
  \min\{2\Delta_k, \Delta_{\text{max}}\} & f_{\text{high}}(x_k) - f_{\text{high}}(x_k + s_k) \geq a\Delta_k \text{ and } c_{\text{high}}(x_k + s_k) \leq 0 \\
  0.5\Delta_k & \text{otherwise}
  \end{cases}
  \]

- Termination:
  - Trust region subproblem nearly equivalent to original problem when trust region is small.
Supersonic Airfoil Test Problem

- Biconvex airfoil in supersonic flow
  - $\alpha = 2^\circ, M_\infty = 1.5$
  - $t/c = 5\%$

<table>
<thead>
<tr>
<th></th>
<th>Linear Panels</th>
<th>Shock Expansion</th>
<th>Cart3D</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_L$</td>
<td>0.1244</td>
<td>0.1278</td>
<td>0.12498</td>
</tr>
<tr>
<td>% Difference</td>
<td>0.46%</td>
<td>2.26%</td>
<td>0.00%</td>
</tr>
<tr>
<td>$C_D$</td>
<td>0.0164</td>
<td>0.0167</td>
<td>0.01666</td>
</tr>
<tr>
<td>% Difference</td>
<td>1.56%</td>
<td>0.24%</td>
<td>0.00%</td>
</tr>
</tbody>
</table>
Approximate Objective Function

- 11 parameters
  - Angle of attack
  - 10 surface spline points
- Minimize drag
  - s.t. t/c ≥ 5%, all positive thickness
- Similar performance to derivative-based multifidelity methods

<table>
<thead>
<tr>
<th>High-Fidelity</th>
<th>Low-Fidelity</th>
<th>SQP</th>
<th>First-Order TR</th>
<th>RBF, ξ=2</th>
<th>RBF, ξ=ξ*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shock-Expansion</td>
<td>Panel Method</td>
<td>314 (-)</td>
<td>110 (-65%)</td>
<td>73 (-77%)</td>
<td>68 (-78%)</td>
</tr>
<tr>
<td>Cart3D</td>
<td>Panel Method</td>
<td>359' (-)</td>
<td>109 (-70%)</td>
<td>80 (-78%)</td>
<td>79 (-78%)</td>
</tr>
</tbody>
</table>

High-Fidelity Evaluations
**Multifidelity Objective and Constraint**

- Max Lift/Drag (multifidelity)
- subject to: Drag ≤ 0.01 (multifidelity)
  - t/c ≥ 5% and positive thickness

<table>
<thead>
<tr>
<th></th>
<th>High-Fidelity</th>
<th>Low-Fidelity</th>
<th>SQP</th>
<th>First-Order TR</th>
<th>RBF, (\xi=2)</th>
<th>RBF, (\xi=\xi^*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objective</td>
<td>Cart3D</td>
<td>Panel Method</td>
<td>1168</td>
<td>97 (-92%)</td>
<td>104 (-91%)</td>
<td>112 (-90%)</td>
</tr>
<tr>
<td>Constraint</td>
<td>Cart3D</td>
<td>Panel Method</td>
<td>2335</td>
<td>97 (-96%)</td>
<td>115 (-95%)</td>
<td>128 (-94%)</td>
</tr>
</tbody>
</table>

*Cart3D optimization sensitive to scaling and finite differences*
Conclusion

• Explained the need for convergent high-fidelity derivative-free methods

• Motivated the use of Bayesian model calibration methods for multifidelity optimization

• Demonstrated convergence of a constrained multifidelity optimization algorithm using Bayesian model calibration
  – Does not require high-fidelity gradient estimates
  – Has performance comparable to other gradient-based methods
  – Showed the method can be used with multiple low-fidelity models without nesting
Acknowledgements

• The authors gratefully acknowledge support from NASA Langley Research Center contract NNL07AA33C technical monitor Natalia Alexandrov.

• A National Science Foundation graduate research fellowship.

• Michael Aftosmis and Marian Nemec for support with Cart3D.
Questions?