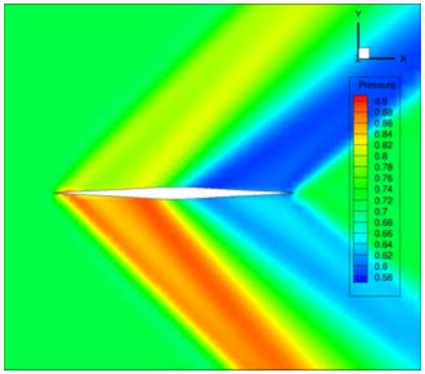
Convergent Multifidelity Optimization using Bayesian Model Calibration

13th AIAA/ISSMO Multidisciplinary Analysis Optimization Conference

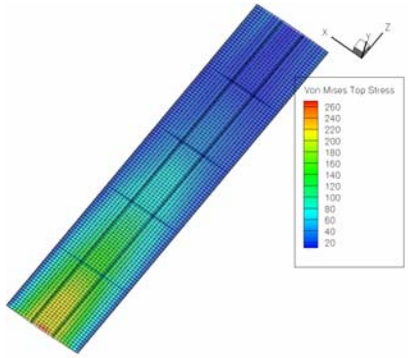
September 14, 2010

Andrew March & Karen Willcox

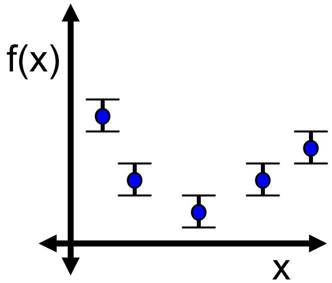
Motivation



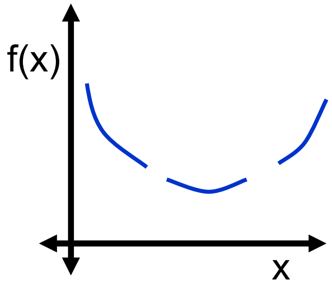
Finite Convergence Tolerance



Non-smooth Objective (Max Stress)



Experimental Result



Function Fails to Exist/Converge

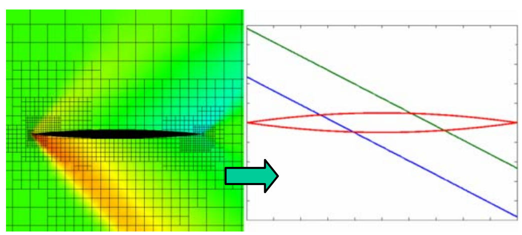
- Engineering problems often have objective functions or constraints that are "expensive" to evaluate
- Derivatives are commonly not available and may not be easy to estimate accurately

Multifidelity Surrogates

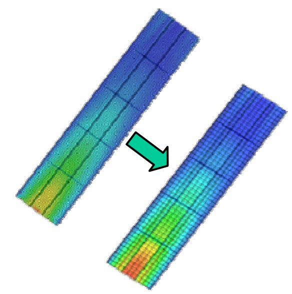
- Definition: *High-Fidelity*
  - The best model of reality that is available and affordable, the analysis that is used to validate the design.
- Definition: *Low(er)-Fidelity*
  - A method with unknown accuracy that estimates metrics of interest but requires lesser resources than the high-fidelity analysis.

Hierarchical Models

Reduced Physics

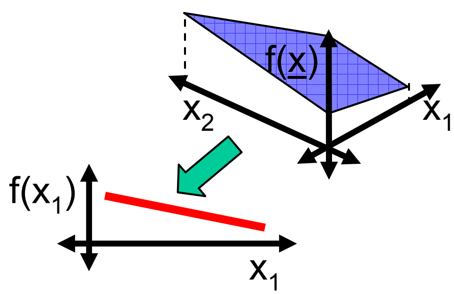


Coarsened Mesh

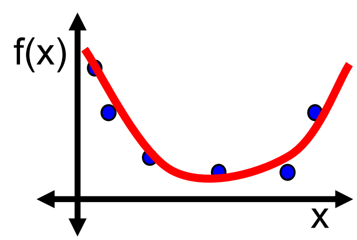


Approximation Models

Reduced Order Model



Regression Model

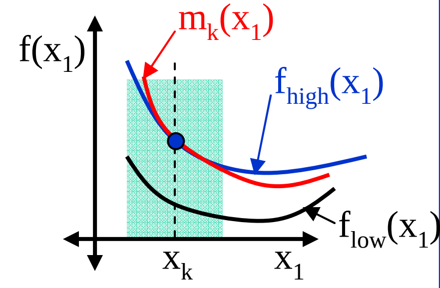


Main Messages

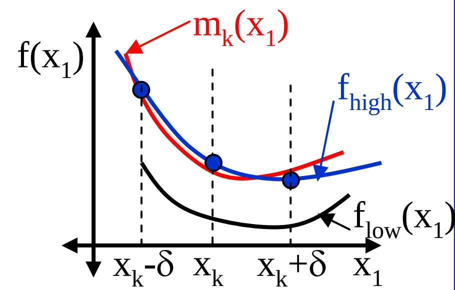
- Bayesian model calibration offers an efficient framework for multifidelity optimization.
- Does not require high-fidelity gradient estimates.
- Can reduce the number of high-fidelity function evaluations compared with other multifidelity methods, even those using gradients.
- Provides a flexible and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

Motivation-Calibration Methods

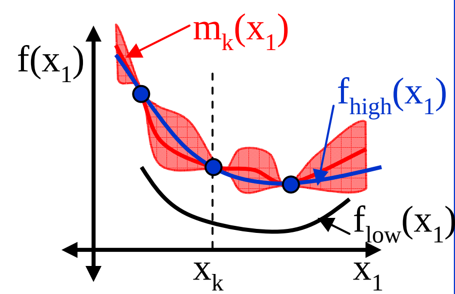
- First-order consistent trust-region methods:
  - Efficient when derivatives are available or can be approximated efficiently
  - Calibrated surrogate models are only used for one iteration



- 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

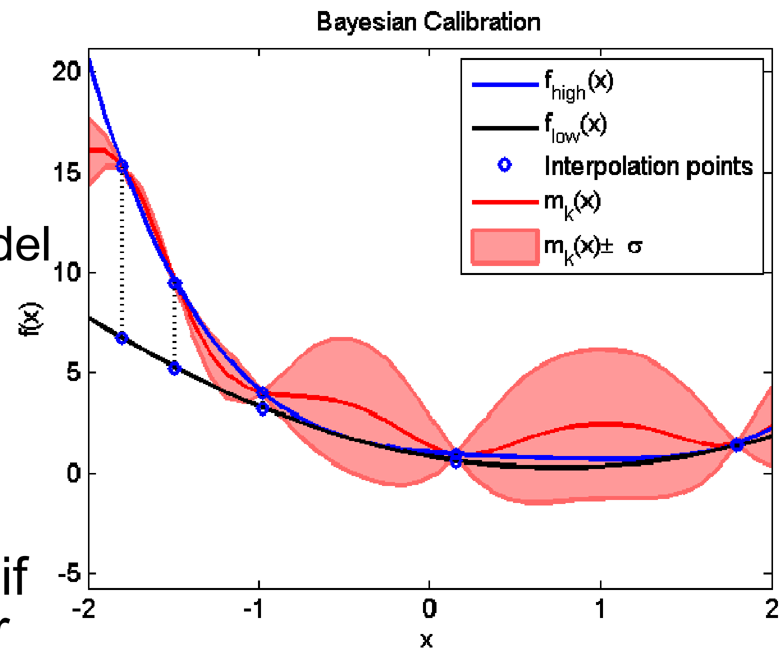


Bayesian Model Calibration

- Define a surrogate model of the high-fidelity function:

$$m_k(\mathbf{x}) \equiv f_{low}(\mathbf{x}) + e_k(\mathbf{x}) \approx f_{high}(\mathbf{x})$$

- The error model, $e(x)$:
  - Is a radial basis function model
  - Interpolates $f_{\text{high}}(\mathbf{x})$- $f_{\text{low}}(\mathbf{x})$ 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

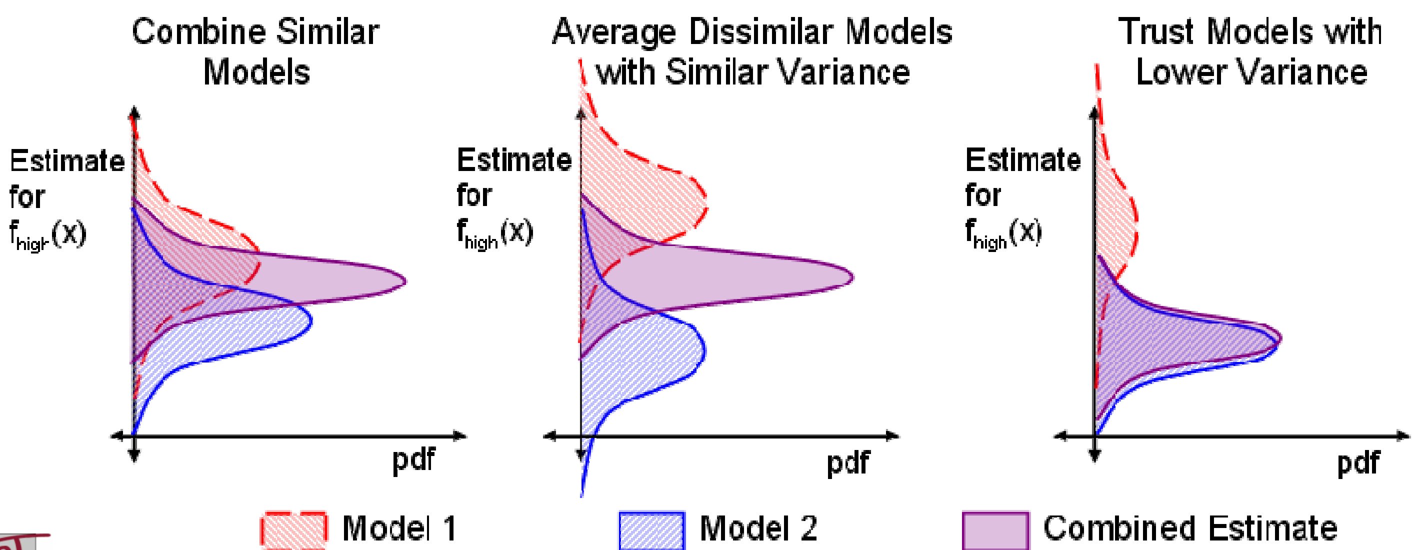
- Define trust region at iteration *k*:

$$B_k = \left\{ \mathbf{x} \in \Re^n : \left\| \mathbf{x} - \mathbf{x}_k \right\| \le \Delta_k \right\}$$

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 $\Delta_k \in (0, \Delta_{max}]$, a *fully linear* model, $m_k(\mathbf{x})$, satisfies

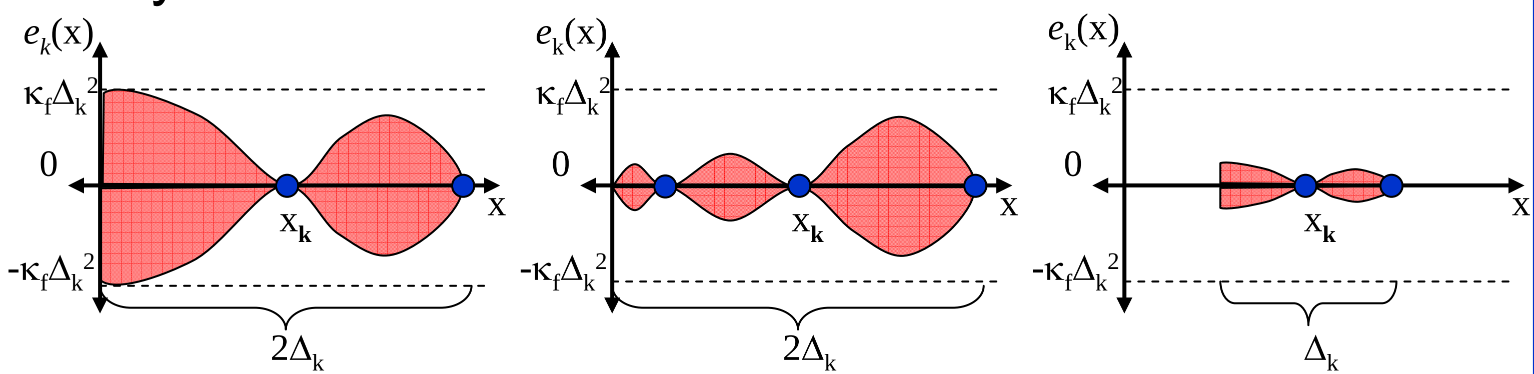
$$\left\| \nabla f_{high}(\mathbf{x}) - \nabla m_k(\mathbf{x}) \right\| \le \kappa_g \Delta_k$$

for a Lipschitz constant κ_g, and

$$\left| f_{high}(\mathbf{x}) - m_k(\mathbf{x}) \right| \le \kappa_f \Delta_k^2$$

with a Lipschitz constant κ_f.

- Fully linear model error bounds:

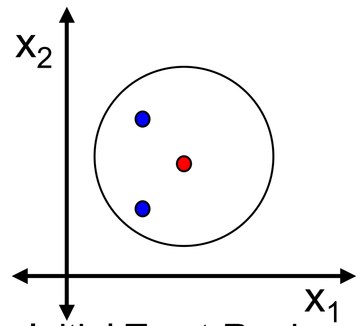


Initial Error Bound | Added Calibration Point | Reduced Trust Region Size

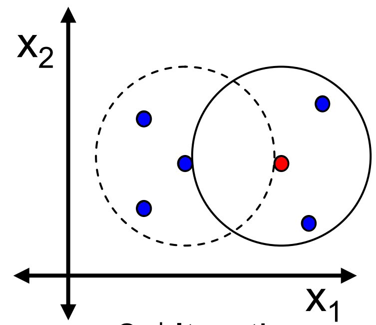
Function Evaluation Points

- 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

Bayesian Model Calibration Approach

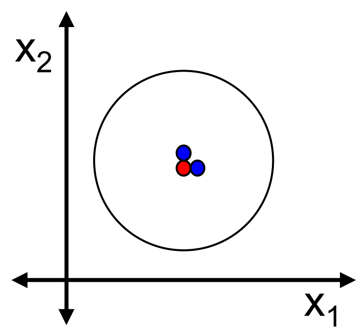


Initial Trust-Region Interpolation Points

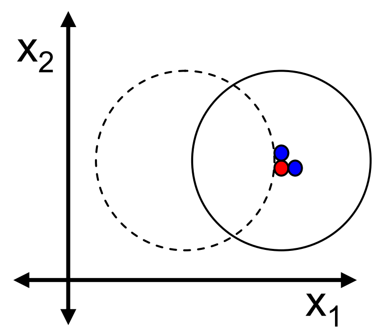


2nd Iteration Interpolation Points

First-Order Trust Region Approach



Initial Trust-Region Finite Difference Points



2nd Iteration Finite Difference Points

2 Constrained Formulations

Method 1:

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

$$\min_{\mathbf{x}\in\Re^n} f_{high}(\mathbf{x}) \quad \text{Derivatives unavailable}$$

$$s.t. \quad g(\mathbf{x}) \le 0, \quad h(\mathbf{x}) = 0 \quad \text{Derivatives available}$$

- 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:

$$\min_{\mathbf{x}\in\Re^n} f_{high}(\mathbf{x}) \quad \text{Derivatives unavailable}$$

$$s.t. \quad g(\mathbf{x}) \le 0, \quad h(\mathbf{x}) = 0 \quad \text{Derivatives available}$$

$$c_{high}(\mathbf{x}) \le 0 \quad \text{Derivatives unavailable}$$

- 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)*

$$\min_{\mathbf{x}\in\Re^n} f_{high}(\mathbf{x})$$

← Derivatives unavailable

$$s.t. \quad g(\mathbf{x}) \le 0$$

$$h(\mathbf{x}) = 0$$

} Derivatives available

Method Summary

$$\min_{\mathbf{x}\in\Re^n} f_{high}(\mathbf{x})$$ ← Derivatives unavailable

$$s.t.\quad g(\mathbf{x}) \le 0,\quad h(\mathbf{x}) = 0$$ ← Derivatives available

- 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:

$$\min_{\mathbf{s}_k\in\Re^n} m_k(\mathbf{x}_k+\mathbf{s}_k)$$
$$s.t.\quad g(\mathbf{x}_k+\mathbf{s}_k)\le 0,\quad h(\mathbf{x}_k+\mathbf{s}_k)=0,\quad \|\mathbf{s}_k\|\le\Delta_k$$

Possibly incompatible

or

$$\min_{\mathbf{s}_k\in\Re^n} \hat{\phi}_k(\mathbf{x}_k+\mathbf{s}_k)$$
$$s.t.\quad \|\mathbf{s}_k\|\le\Delta_k$$

Hessian norm unbounded

- Impose a sufficient decrease condition: $\lim_{k\to\infty}\Delta_k = 0$
- Termination:
  - Fully linear model: $\left\|\nabla f_{high}(\mathbf{x}) - \nabla m_k(\mathbf{x})\right\| \to 0;\quad \left|f_{high}(\mathbf{x}) - m_k(\mathbf{x})\right| \to 0$
  - Subproblem 1 is nearly equivalent to the original problem

Method 2: *(Constraint Derivatives Unavailable)*

$$\min_{\mathbf{x}\in\Re^n} f_{high}(\mathbf{x})$$ ← Derivatives unavailable

$$s.t. \quad g(\mathbf{x}) \leq 0$$
$$h(\mathbf{x}) = 0$$
} Derivatives available

$$c_{high}(\mathbf{x}) \leq 0$$ ← Derivatives unavailable

- Finding a feasible staring point
- Finding a high-fidelity optimum

Finding a Feasible Starting Point

$$\min_{\mathbf{x}\in\Re^n} f_{high}(\mathbf{x})$$ ← Derivatives unavailable

$$s.t.\quad g(\mathbf{x}) \le 0,\quad h(\mathbf{x}) = 0$$ ← Derivatives available

$$c_{high}(\mathbf{x}) \le 0$$ ← Derivatives unavailable

- Two fully linear surrogate models:

$$m_k(\mathbf{x}) \approx f_{high}(\mathbf{x}) \qquad \overline{m}_k(\mathbf{x}) \approx c_{high}(\mathbf{x})$$

- Find an initial feasible point:

$$\begin{aligned} \min_{\mathbf{x}\in\Re^n} \quad & c_{high}(\mathbf{x}) \\ s.t. \quad & g(\mathbf{x}) \le 0 \\ & h(\mathbf{x}) = 0 \end{aligned}$$

- Constraint may not be bounded from below:

$$\begin{aligned} \min_{\mathbf{x}\in\Re^n} \quad & \max\{c_{high}(\mathbf{x}) + d, 0\}^2 \\ s.t. \quad & g(\mathbf{x}) \le 0 \\ & h(\mathbf{x}) = 0 \end{aligned}$$

- Only need to iterate until $c_{high}(\mathbf{x}) \le 0$ and other constraints satisfied

Finding the Optimum

- Trust region subproblem:

$$\begin{aligned}
&\min_{\mathbf{s}_k \in \Re^n} \quad m_k(\mathbf{x}_k + \mathbf{s}_k) \\
&s.t. \quad g(\mathbf{x}_k + \mathbf{s}_k) \le 0 \\
&\qquad\; h(\mathbf{x}_k + \mathbf{s}_k) = 0 \\
&\qquad\; \overline{m}_k(\mathbf{x}_k + \mathbf{s}_k) \le 0 \\
&\qquad\; \|\mathbf{s}_k\| \le \Delta_k
\end{aligned}$$

$$\begin{aligned}
&\min_{\mathbf{x} \in \Re^n} f_{high}(\mathbf{x}) \quad \leftarrow \text{Derivatives unavailable} \\
&s.t. \quad g(\mathbf{x}) \le 0 \\
&\qquad\; h(\mathbf{x}) = 0 \quad \Big\} \text{ Derivatives available} \\
&\qquad\; c_{high}(\mathbf{x}) \le 0 \quad \leftarrow \text{Derivatives unavailable}
\end{aligned}$$

- Trial point acceptance:

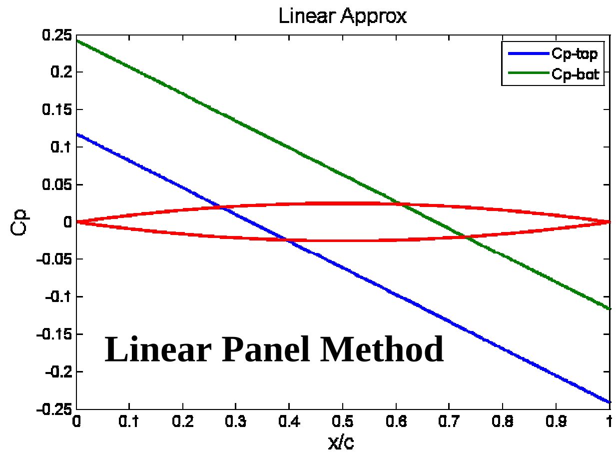
$$\mathbf{x}_{k+1} = \begin{cases} \mathbf{x}_k + \mathbf{s}_k & f_{high}(\mathbf{x}_k) \ge f_{high}(\mathbf{x}_k + \mathbf{s}_k) \text{ and } c_{high}(\mathbf{x}_k + \mathbf{s}_k) \le 0 \\ \mathbf{x}_k & \text{otherwise} \end{cases}$$

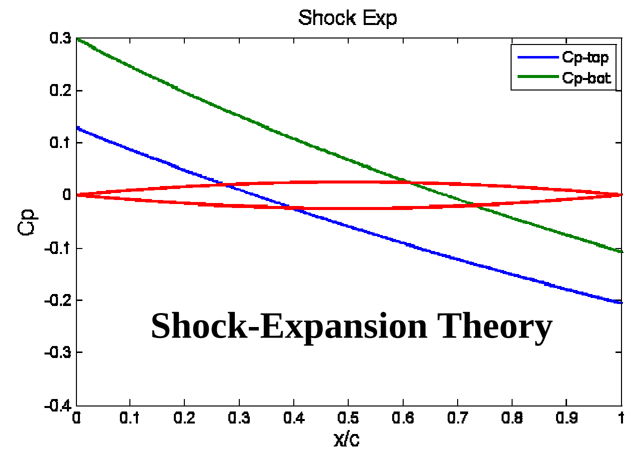
- Trust region size update:

$$\Delta_{k+1} = \begin{cases} \min\{2\Delta_k, \Delta_{\max}\} & f_{high}(\mathbf{x}_k) - f_{high}(\mathbf{x}_k + \mathbf{s}_k) \ge a\Delta_k \text{ and } c_{high}(\mathbf{x}_k + \mathbf{s}_k) \le 0 \\ 0.5\Delta_k & \text{otherwise} \end{cases}$$

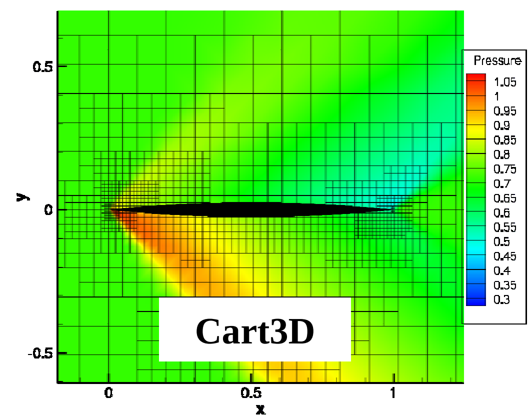
- Termination:
  - Trust region subproblem nearly equivalent to original problem when trust region is small.

Supersonic Airfoil Test Problem



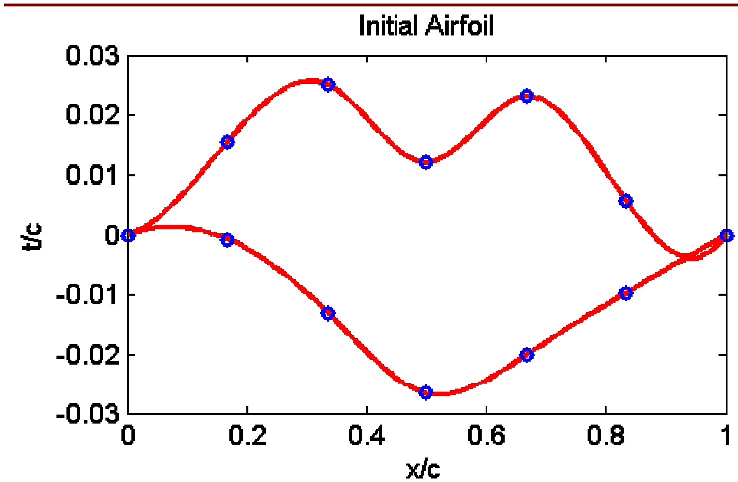


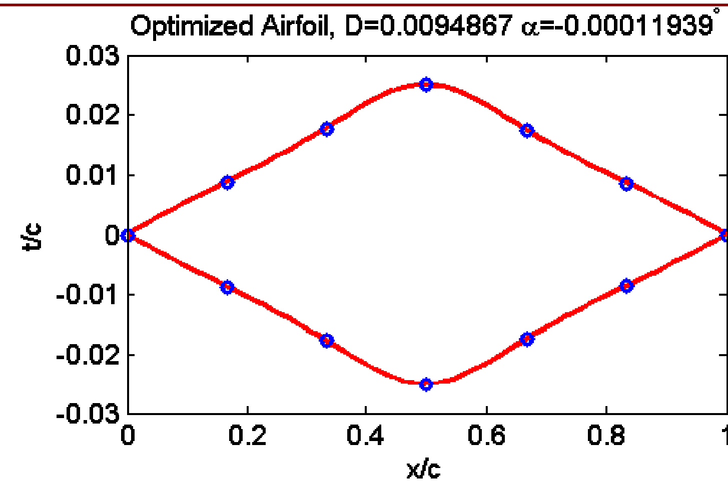
- Biconvex airfoil in supersonic flow
  - $\alpha = 2^o, M_\infty = 1.5$
  - (t/c) = 5%



	Linear Panels	Shock Expansion	Cart3D
C_L	0.1244	0.1278	0.12498
% Difference	0.46%	2.26%	0.00%
C_D	0.0164	0.0167	0.01666
% Difference	1.56%	0.24%	0.00%

Approximate Objective Function

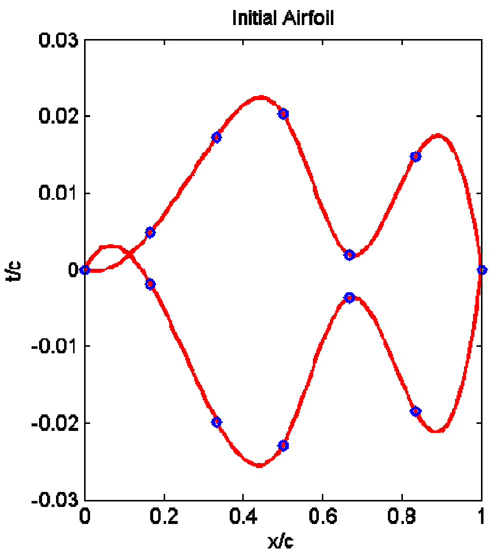


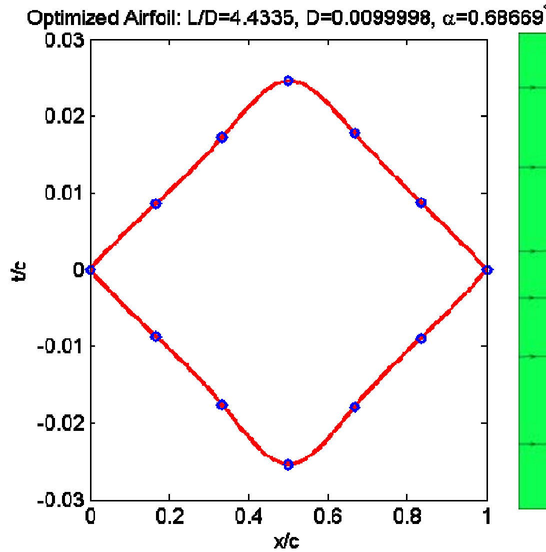


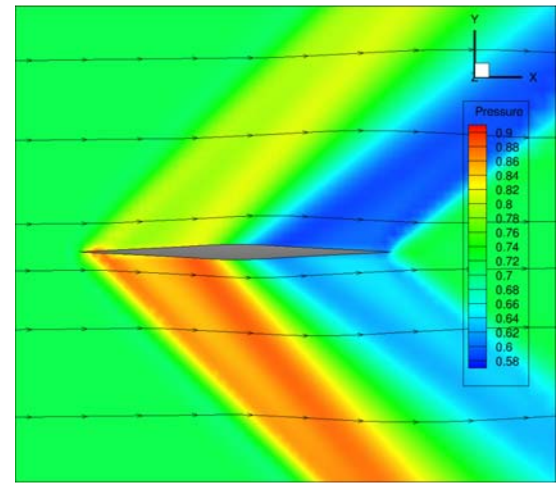
- 11 parameters
  - Angle of attack
  - 10 surface spline points
- Minimize drag
  - s.t. t/c≥5%, all positive thickness
- Similar performance to derivative-based multifidelity methods

High-Fidelity	Low-Fidelity	SQP	First-Order TR	RBF, ξ=2	RBF, $\xi=\xi^*$
Shock-Expansion	Panel Method	314 (-)	110 (-65%)	73 (-77%)	68 (-78%)
Cart3D	Panel Method	359* (-)	109 (-70%)	80 (-78%)	79 (-78%)

Multifidelity Objective and Constraint







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

	High-Fidelity	Low-Fidelity	SQP	First-Order TR	RBF, ξ=2	RBF, ξ=ξ*
Objective	Cart3D	Panel Method	1168* (-)	97 (-92%)	104 (-91%)	112 (-90%)
Constraint	Cart3D	Panel Method	2335* (-)	97 (-96%)	115 (-95%)	128 (-94%)

- *Cart3D optimization sensitive to scaling and finite differences

	High-Fidelity	Low-Fidelity	SQP	First-Order TR	RBF, ξ=2	RBF, ξ=ξ*
Objective	Shock-Exp	Panel Method	773 (-)	132 (-83%)	93 (-88%)	90 (-88%)
Constraint	Shock-Exp	Panel Method	773 (-)	132 (-83%)	97 (-87%)	96 (-88%)

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?