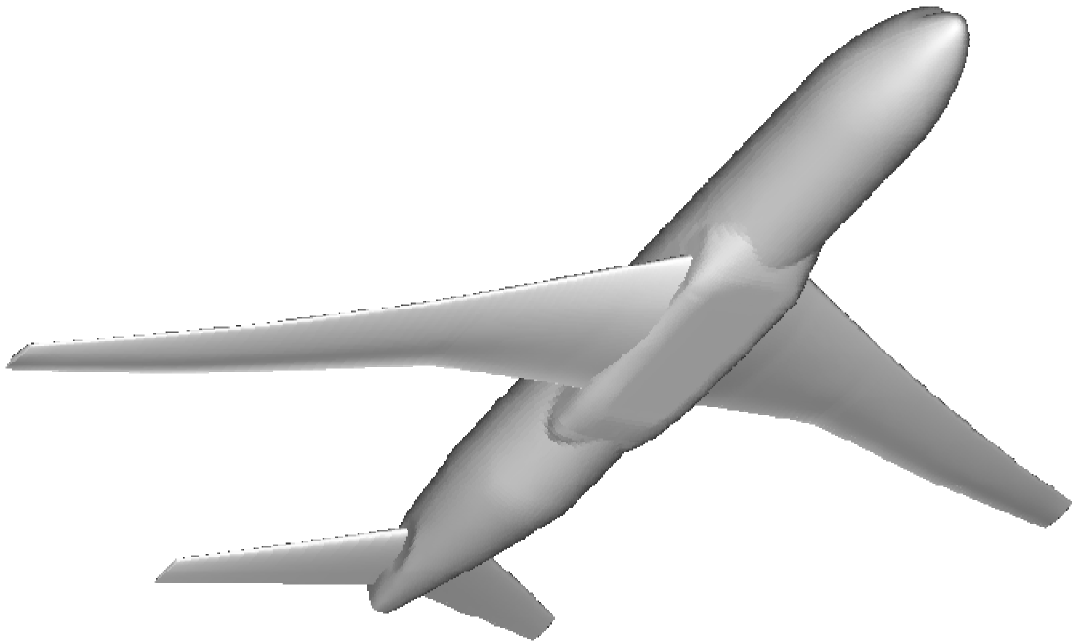
A Provably Convergent Multifidelity Optimization Algorithm not Requiring High-Fidelity Derivatives

Multidisciplinary Design Optimization
Specialist Conference

April 14, 2010

Andrew March & Karen Willcox

Motivation



- BCFD viscous solution: 2920[1] CPU hours
  - Pre/post-processing time not included
- **Can this configuration be optimized?**

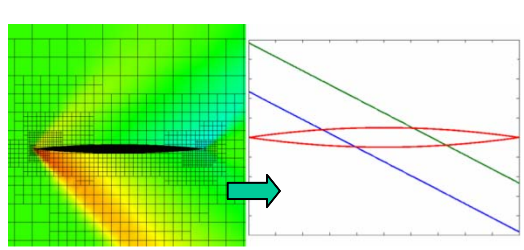
[1]Winkler et al. 4th CFD Drag Prediction Workshop, San Antonio TX, June 2009

Multifidelity Surrogates

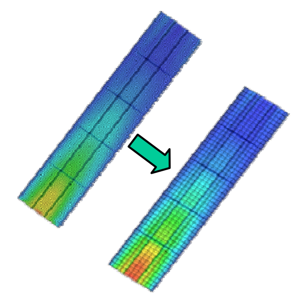
- Definition: *High-Fidelity*
  - The best model of reality that is available and affordable, the analysis that is used to validate the design.
- Definition: *Low(er)-Fidelity*
  - A method with unknown accuracy that estimates metrics of interest but requires lesser resources than the high-fidelity analysis.

Hierarchical Models

Reduced Physics

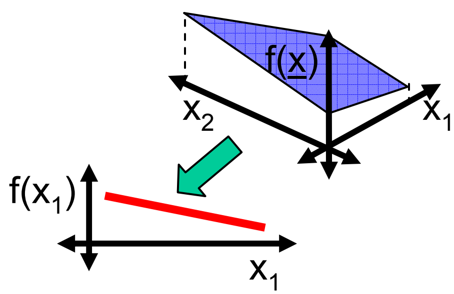


Coarsened Mesh

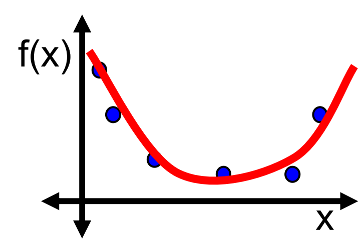


Approximation Models

Reduced Order Model



Regression Model



Main Messages

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

Motivation-Calibration Methods

- First-order trust-region methods:
  - Efficient for multifidelity optimization 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

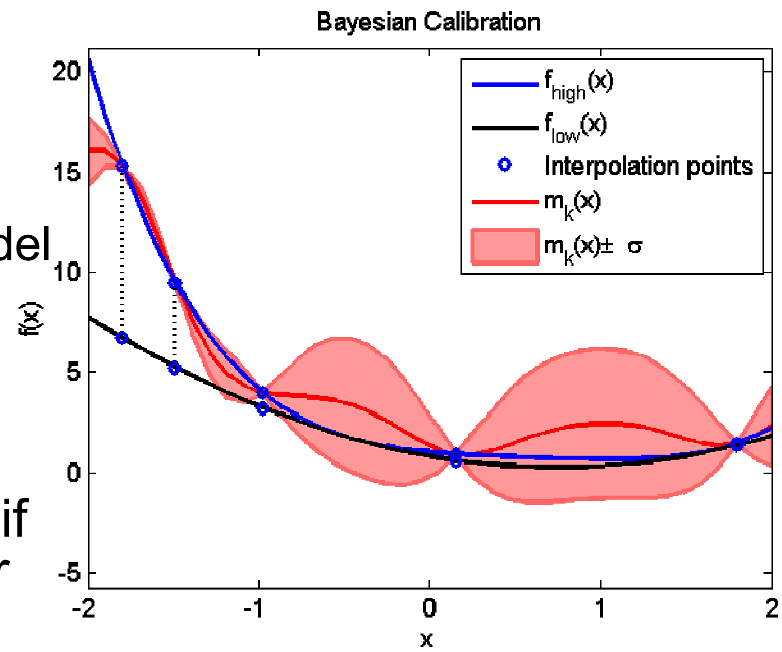
- **Goal:** Develop a multifidelity optimization algorithm that combines Bayesian calibration and reuse of high-fidelity information in a manner provably convergent to an optimum of the high-fidelity function

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



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

- Define trust region at iteration *k*:

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

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.

- Conn et al. (2009) shows that in a trust region setting, fully linear models are sufficient to prove convergence to a stationary point of $f_{high}(\mathbf{x})$.
  - Requires: $f_{high}(\mathbf{x})$ is continuously differentiable, has Lipschitz continuous first derivative, and is bounded from below
  - Multifidelity method also requires that $f_{low}(\mathbf{x})$ is continuously differentiable and has Lipschitz continuous first derivative

Fully Linear RBF Models

- Standard radial basis function model:

$$e_k(\mathbf{x}) = \sum_{i=1}^{|y|} \lambda_i \phi\left(\left\|\mathbf{x} - \mathbf{x}_k - \mathbf{y}_i\right\|; \xi\right) + \sum_{i=1}^{n+1} \nu_i \pi(\mathbf{x} - \mathbf{x}_k)$$

- Radial basis function (RBF) model requirements:
  - RBF, ϕ, is twice continuously differentiable
  - $\phi(r)$ has zero derivative at r=0
  - Polynomial basis, π, is linear

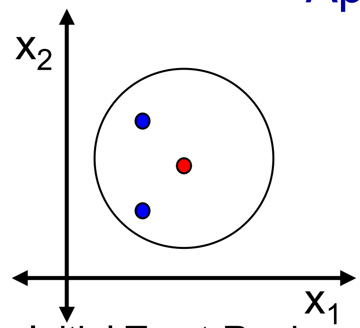
$$\phi = e^{\frac{-r^2}{\xi^2}}$$

- Wild et al. (2008) showed that an RBF model can be made fully linear by construction
  - Places conditions on the sample points used to construct the RBF model

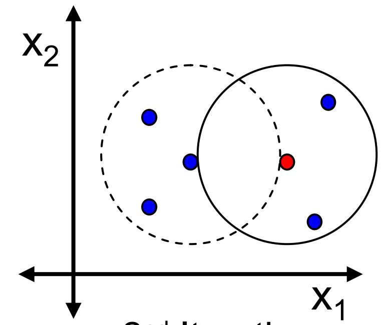
Function Evaluation Points

- RBF model has sufficient local behavior to guarantee convergence
- It also captures some global behavior
- First-order trust region approaches only look at the center of the current trust region
- RBF model will likely require fewer high-fidelity evaluations

Radial Basis Function Calibration Approach

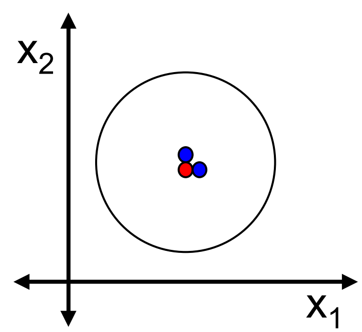


Initial Trust-Region Interpolation Points

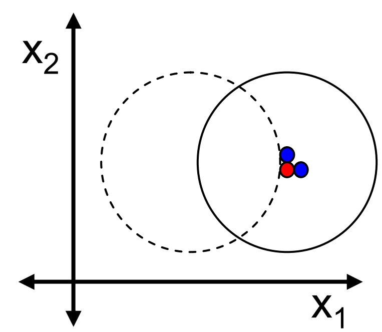


2nd Iteration Interpolation Points

First-Order Trust Region Approach



Initial Trust-Region Finite Difference Points



2nd Iteration Finite Difference Points

Unconstrained Algorithm Summary

- Solve the trust-region subproblem to determine a candidate step, $\mathbf{s}_k$:

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

- Evaluate f_{high} at the candidate point and compute the ratio of actual to predicted reduction:
$$\rho_k = \frac{f_{high}(\mathbf{x}_k) - f_{high}(\mathbf{x}_k + \mathbf{s}_k)}{m_k(\mathbf{x}_k) - m_k(\mathbf{x}_k + \mathbf{s}_k)}$$

- Accept/reject iterate: $\mathbf{x}_{k+1} = \begin{cases} \mathbf{x}_k + \mathbf{s}_k & \rho_k > 0 \\ \mathbf{x}_k & \text{otherwise} \end{cases}$

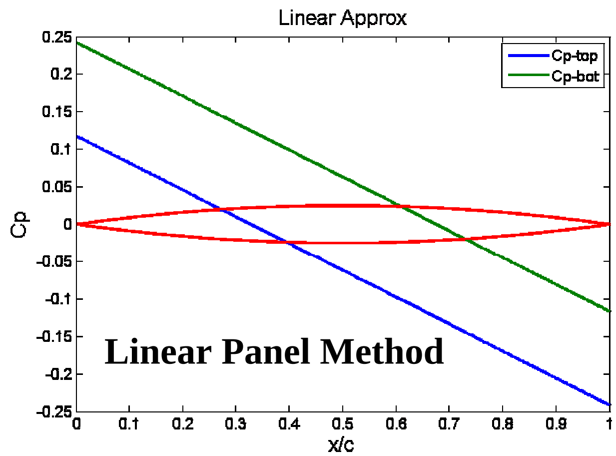
- Update trust region size: $\Delta_{k+1} = \begin{cases} \min\{2\Delta_k, \Delta_{\max}\} & \rho_k \geq \eta \\ 0.5\Delta_k & \rho_k < \eta \end{cases}$

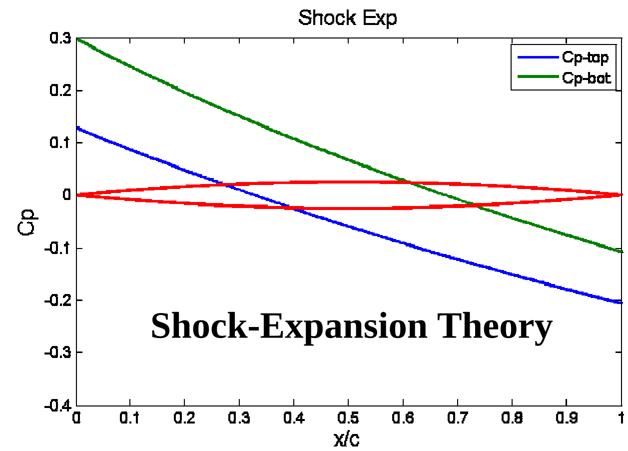
- Form new fully linear model $m_{k+1}(\mathbf{x})$, on $\{\mathbf{x} : \|\mathbf{x} - \mathbf{x}_{k+1}\| \leq \Delta_{k+1}\}$

- Perform convergence check: $\|\nabla m_k(\mathbf{x}_k)\| \leq \varepsilon_1$ and $\Delta_k \leq \varepsilon_2$

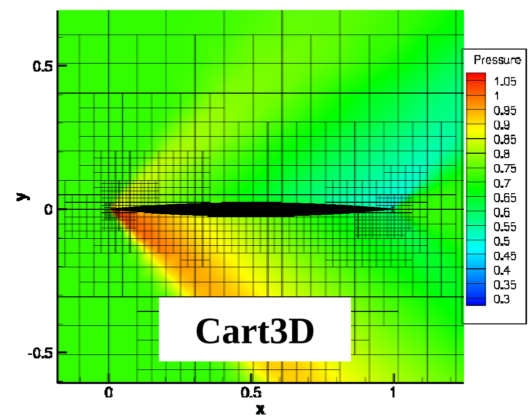
  and reduce size of trust region until convergence proved
  [called the criticality check in Conn et al. (2009)]

Supersonic Airfoil Test Problem





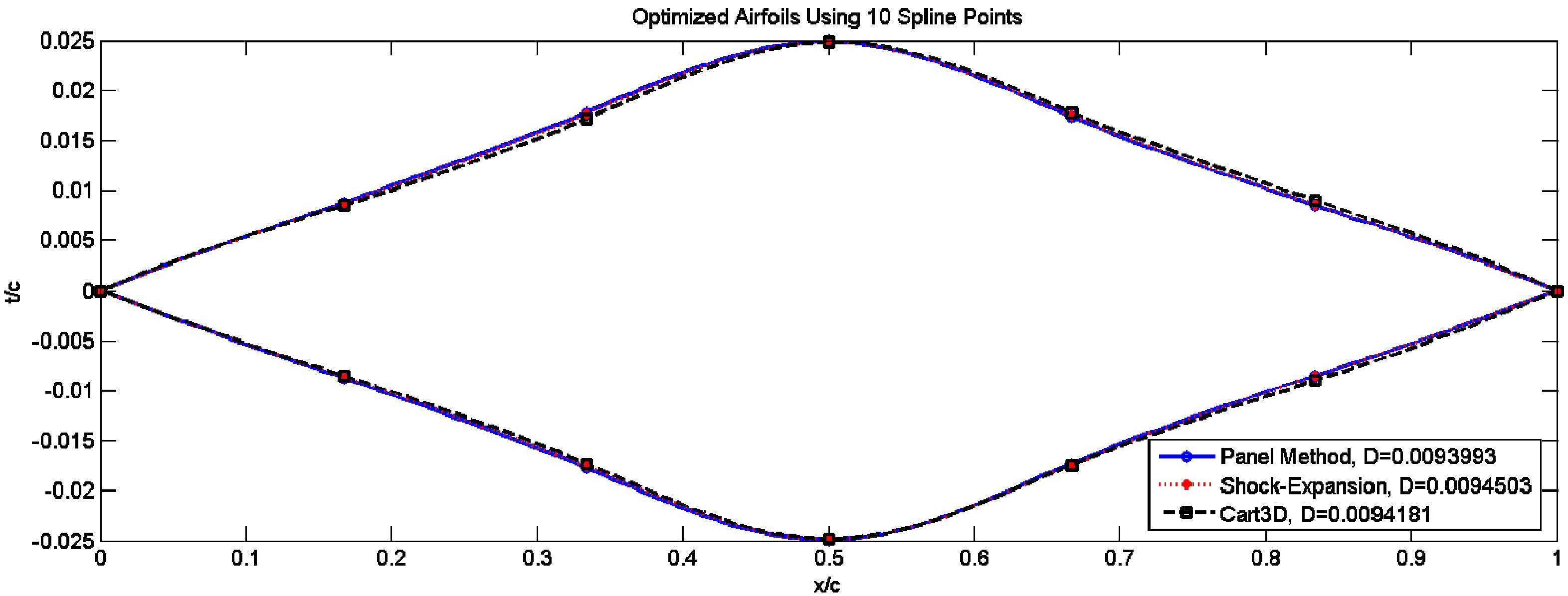
- Biconvex airfoil in supersonic flow
  - $\alpha = 2^o, M_\infty = 1.5$
  - $(t/c) = 5\%$



	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%

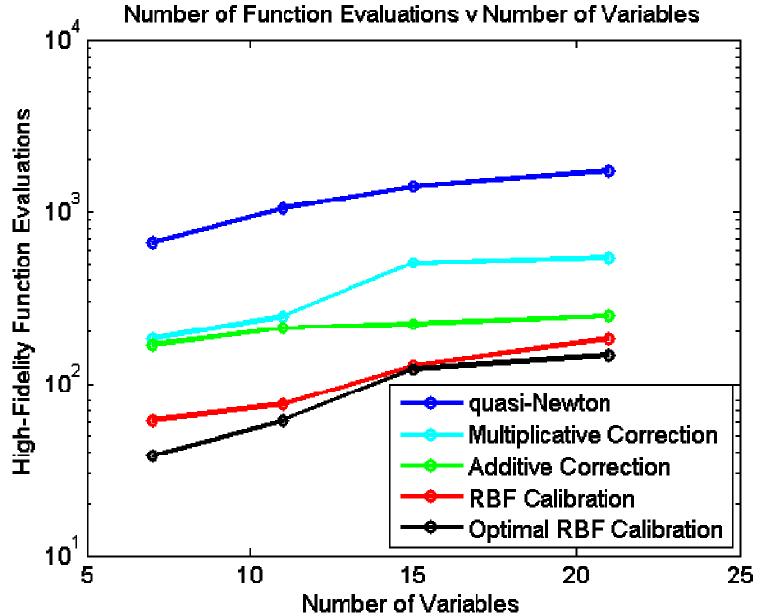
Airfoil Parameterization



- Panel method and shock-expansion theory require sharp leading and trailing edges
- Parameterization has equal number of spline points on upper and lower surface and angle of attack
- Figure shows minimum drag solutions for 11 design variables

Airfoil Optimization Results

- Models:
  - Low-Fidelity: Panel Method
  - High-Fidelity: S-E Theory
- Airfoil parameterization:
  - Angle of Attack
  - Equal # of upper/lower surface spline points
- Optimization tolerance:
  - $||step|| \leq 5x10^{-6}$ or
  - $||dm/dx|| \leq 5x10^{-4}$
- Criteria: Fewest high-fidelity function evaluations
  - Average of 5 runs with random ICs

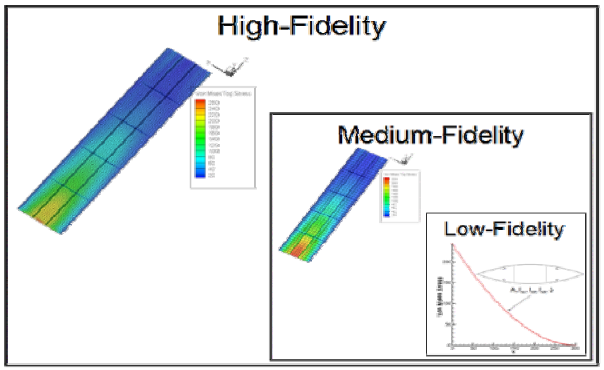


# of Variables:	7	11	15	21
Quasi-Newton	667	1048	1408	1731
β-Correlation	183	246	503	546
Add-Correction	168	211	223	249
RBF α=2	61	76	127	182
RBF α*	38	61	122	146

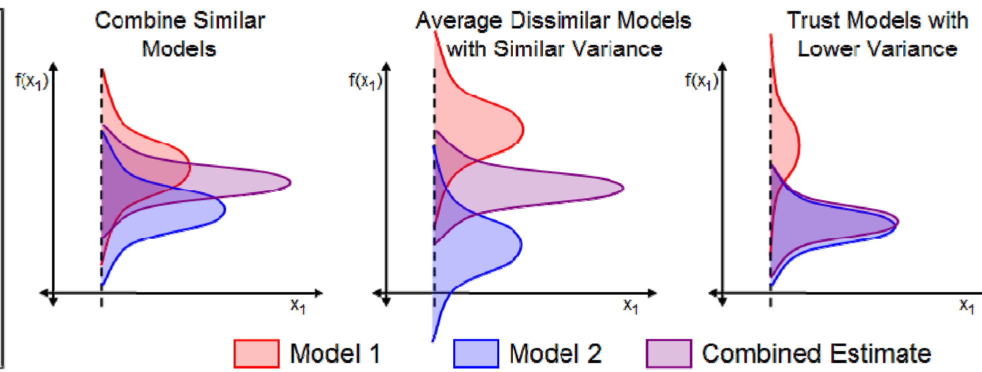
High-Fidelity Function Calls

Combining Multiple Fidelity Levels

Nesting



Maximum Likelihood



- Nested Approach
  - "Classic approach"
  - Possible exponential scaling in function evaluations,
    - e.g. 50 high-fidelity evaluations, 2500 medium-fidelity evaluations, 125,000 low-fidelity evaluations
- Maximum Likelihood Approach
  - Flexibility in selecting low/medium-fidelity function calls
  - Robust to poor models

Combining Multiple Lower-Fidelities

- The RBF error interpolation can be treated as a Kriging model that predicts the high-fidelity function with a normally distributed error:

$$f_{high}(\mathbf{x}) \approx f_{med}(\mathbf{x}) + \mathcal{N}\left(e_{med}(\mathbf{x}), \sigma^2_{med}(\mathbf{x})\right)$$

$$f_{high}(\mathbf{x}) \approx f_{low}(\mathbf{x}) + \mathcal{N}\left(e_{low}(\mathbf{x}), \sigma^2_{low}(\mathbf{x})\right)$$

- Using Kriging models for each of the lower-fidelity functions, a maximum likelihood estimate for the high-fidelity function is:

$$f_{high}(x) \approx \left(f_{low} + e_{low}\right)\left(\frac{\sigma^2_{med}}{\sigma^2_{low} + \sigma^2_{med}}\right) + \left(f_{med} + e_{med}\right)\left(\frac{\sigma^2_{low}}{\sigma^2_{low} + \sigma^2_{med}}\right)$$

$$\frac{1}{\sigma^2_{high}} = \frac{1}{\sigma^2_{low}} + \frac{1}{\sigma^2_{med}}$$

- To fit in the original multifidelity algorithm, only one of the two lower-fidelity functions needs to be sampled at the required calibration points.
  – This allows substantial flexibility in selecting when each lower-fidelity function is used.

3-Fidelity Supersonic Airfoil Results

- **Maximum likelihood approach reduced high-fidelity function calls for all cases.**
  - Results use the same calibration points for all lower-fidelity functions
  - Fancier sampling methods can be used
- Nested approach failed to converge with a non-smooth high-fidelity function (Cart3D):

	Cart3D	Shock-Expansion Theory	Panel Method
Two-Fidelities	88	0	47679
Max. Likelihood	66	23297	23297
Nested	66*	7920*	167644

Function Calls

- The maximum likelihood approach is robust to the poor information.
  - A camberline model estimates drag poorly (thickness is ignored)
  - The best result of the nested approach is shown, average result otherwise

	Shock-Expansion Theory	Panel Method	Camberline
Two-Fidelities	126	43665	0
Max. Likelihood	84	30057	30057
Nested	212**	59217**	342916**

Function Calls

Conclusion

- Explained the need for convergent high-fidelity derivative-free methods
- Demonstrated convergence of an unconstrained multifidelity optimization algorithm using Bayesian model calibration
  - Through numerical experiments, showed that the method works for nonsmooth functions
  - Has performance comparable to other state-of-the-art design methods
- Developed a maximum likelihood method to combine multiple lower fidelities into a single estimate of the high-fidelity function.
  - Showed that this technique converges faster than nesting, is robust to poor information, and allows flexible sampling.

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?