Multidisciplinary System Design Optimization (MSDO)

Approximation Methods

Karen Willcox

Slides from: Theresa Robinson, Andrew March

Outline

- Introduction to approximation methods
- Data fit methods
  - Polynomial response surfaces
  - Kriging
- Model order reduction
  - Reduced-basis methods
  - Proper orthogonal decomposition
- Multifidelity methods
  - Trust-region model management

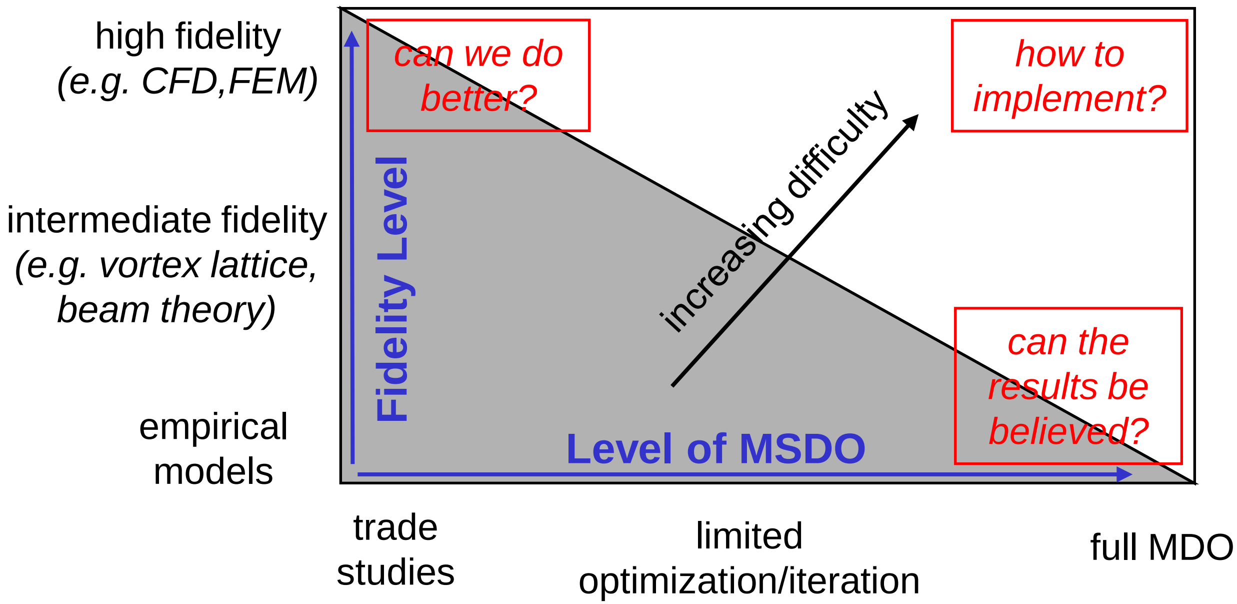
16.888
ESD.77

Approximation Methods

- Replace the simulation with an approximation or "surrogate"
- Uses some data from the initial simulation
  - Can be global or local
- Surrogate is much less computationally expensive to evaluate
- Not just optimization
  - Uncertainty Quantification (e.g. Monte Carlo simulation methods)
  - Visualization

Why Approximation Methods

We have seen throughout the course the constant trade-off between **computational cost** and **fidelity.**



from Giesing, 1998

Approximation methods provide a way to get high-fidelity model information throughout the optimization without the computational expense.

Data Fit Methods

- Sample the simulation at some number of design points
  - Use DOE methods, e.g. Latin hypercube, to select the points
- Fit a surrogate model using the sampled information
- Surrogate may be global (e.g., quadratic response surface) or local (e.g., Kriging interpolation)
- Surrogate may be updated adaptively by adding sample points based on surrogate performance (e.g., EGO)

Polynomial Response Surface Method

- Surrogate model is a local or global polynomial model
- Can be of any order
  - Most often quadratic; higher order requires many samples
- Advantages: Simple to implement, visualize, and understand, easy to find the optimum of the response surface
- Disadvantages: May be too simple, doesn't capture multimodal functions well

Global Polynomial Response Surface

- Fit objective function with a polynomial
- *e.g.* quadratic approximation:

$$J(\mathbf{x}) = \mathbf{a_0} + \sum_i b_i x_i + \sum_i c_{ii} x_i^2 + \sum_{i,j<i} c_{ij} x_i x_j$$

- Update model by including a new function evaluation then doing least squares fit to compute the new coefficients

Global Polynomial Response Surface

Estimation problem: $\mathbf{J} \approx \mathbf{X}\mathbf{c}$

$$\mathbf{J} = \begin{bmatrix} J^1 & J^2 & \cdots & J^M \end{bmatrix}^T$$

J=vector containing *M* responses

$$\mathbf{c} = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{p-1} \end{bmatrix}$$

c=vector containing *p* coefficients

$$\mathbf{X} = \begin{bmatrix} 1 & x_1^1 & x_2^1 & \cdots & x_1^1 x_1^1 & \cdots \\ 1 & x_1^2 & & & & \\ \vdots & \vdots & & & & \\ 1 & x_1^M & x_2^M & & x_1^M x_1^M & \end{bmatrix}$$

x_1^2 is the value of x_1 for the 2nd sample

X=$M \times p$ matrix

Each row corresponds to one data sample; each column corresponds to an unknown coefficient

Least squares solution: $\mathbf{c} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{J}$

Kriging

- Adopted from the geostatistics literature
- Based on Gaussian process models
- Assumes that the output function values are correlated in design space, i.e. closer points are more highly correlated
- Can have multiple extrema
- Interpolating method
  - Exact at sample points
- Gives estimate of mean squared error
  - Can use to give error bounds
  - Can use to choose new sample points

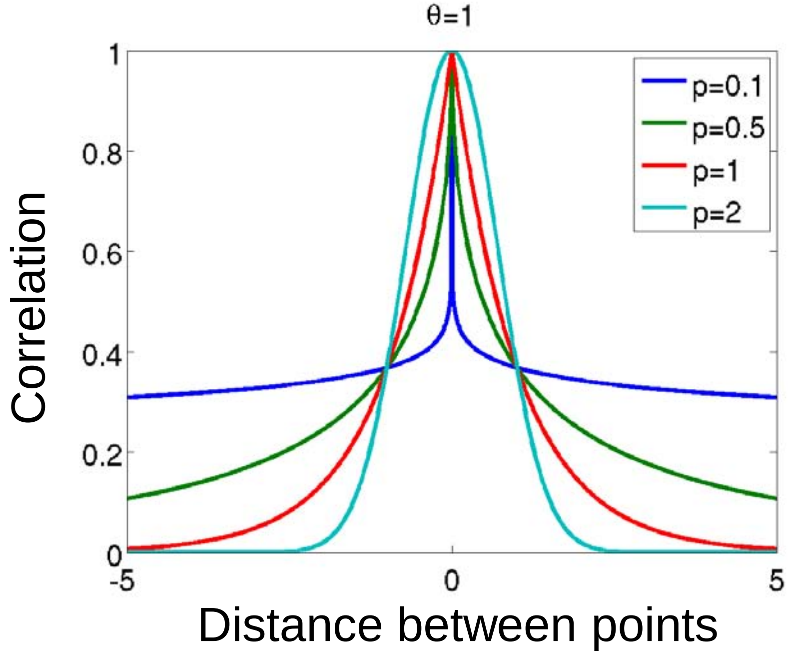
Kriging: Mathematical Background

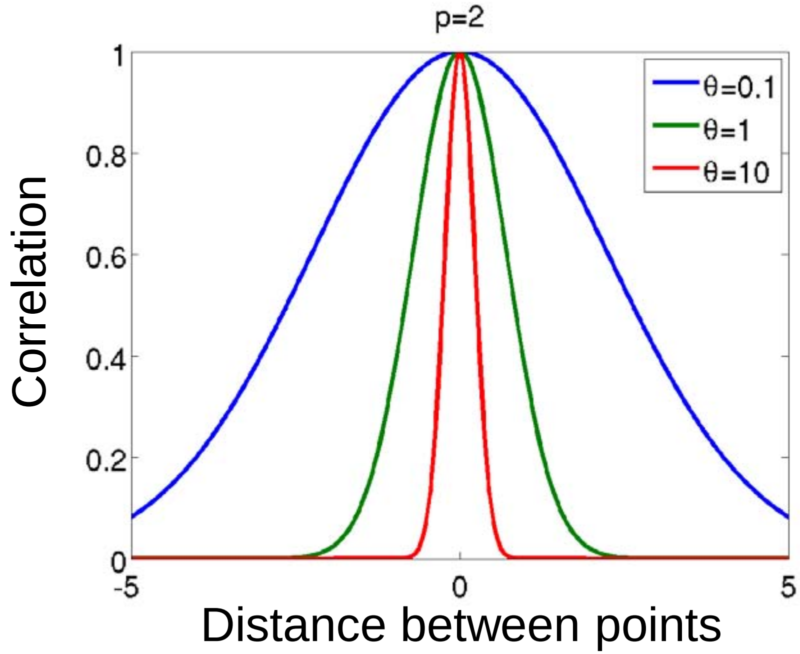
- We want to make a prediction of y at a point $\mathbf{x}$
- Uncertain of value: model as a random variable, normally distributed with mean μ and variance σ^2
- Consider two points $\mathbf{x}_i$ and $\mathbf{x}_j$
- Expect values to be close if the distance between them is small
- Formalize this idea by setting:

$$Corr[Y(x_j),Y(x_k)] = \exp\left(-\sum_{i=1}^{n} \theta_i \left| x_{ji} - x_{ki} \right|^{p_i} \right)$$

Kriging Basis Functions

$$Corr[Y(x_j),Y(x_k)] = \exp\left(-\sum_{i=1}^{n}\theta_i\left|x_{ji}-x_{ki}\right|^{p_i}\right)$$





Each p_i and θ_i is chosen to best fit the data

Kriging Mathematics Cont.

- Choose μ, p_i, and θ_i to maximize the likelihood of observing the data
  - Detailed equations in Giunta and Watson (1998), derivation in Jones (2001)

- Kriging predictor is

$$\hat{y}(\mathbf{x}^*) = \mu + \sum_{i=1}^{k} c_i \exp\left(-\sum_{j=1}^{n} \theta_j \left| \mathbf{x}^* - \mathbf{x}_i \right|^{p_j} \right)$$

mean surface

weighted sum of Gaussians, each centered at a sample point

Kriging Extensions

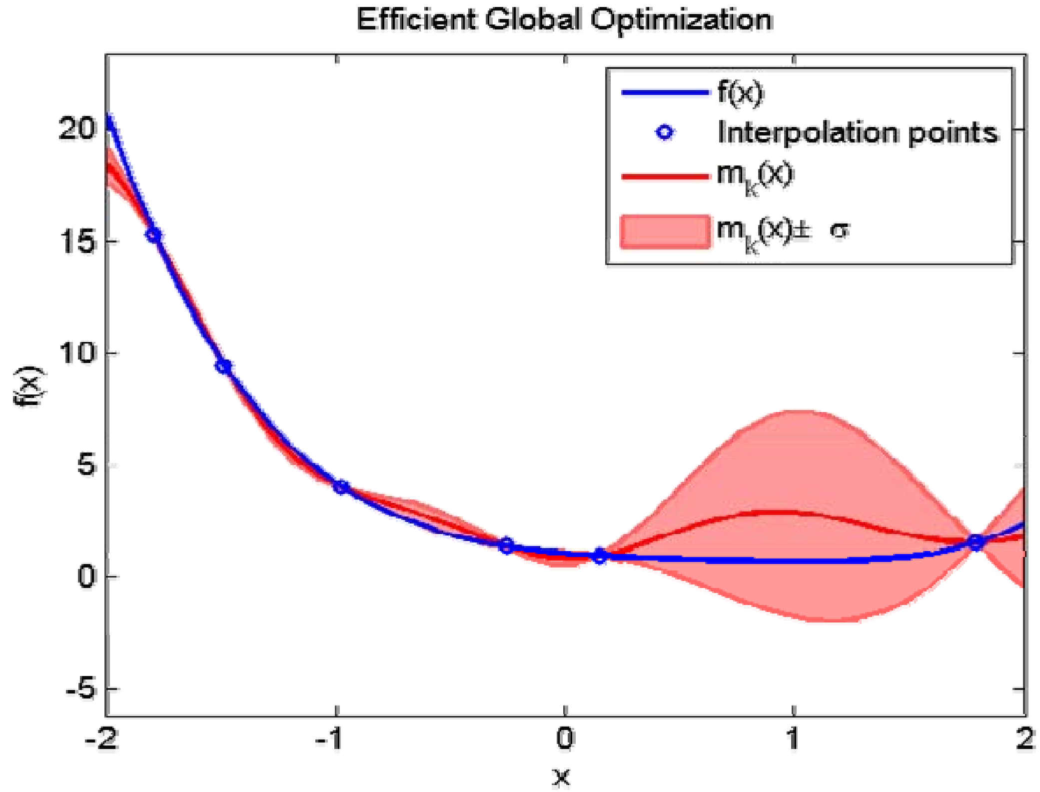
- Can combine polynomial RSM and Kriging
  - Apply Kriging to difference between sample values and polynomial approximation
- Soft Kriging allows upper and lower bounds, prior CDFs
- Efficient Global Optimization (EGO)
  - Uses Kriging to find “expected improvement”
  - Samples the point with the largest expected improvement and adds it to the sample set

Efficient Global Optimization

- Jones 1998; based on probability theory
- Assumes:

$$f(\mathbf{x}) \approx \beta^T \mathbf{x} + N\left(\mu(\mathbf{x}), \sigma^2(\mathbf{x})\right)$$

- $\beta^T \mathbf{x}$: regression term
- $N\left(\mu(\mathbf{x}), \sigma^2(\mathbf{x})\right)$: error from regression model is normally distributed, with mean $\mu(\mathbf{x})$ and variance $\sigma^2(\mathbf{x})$



Surrogate model is updated adaptively; k^th surrogate is

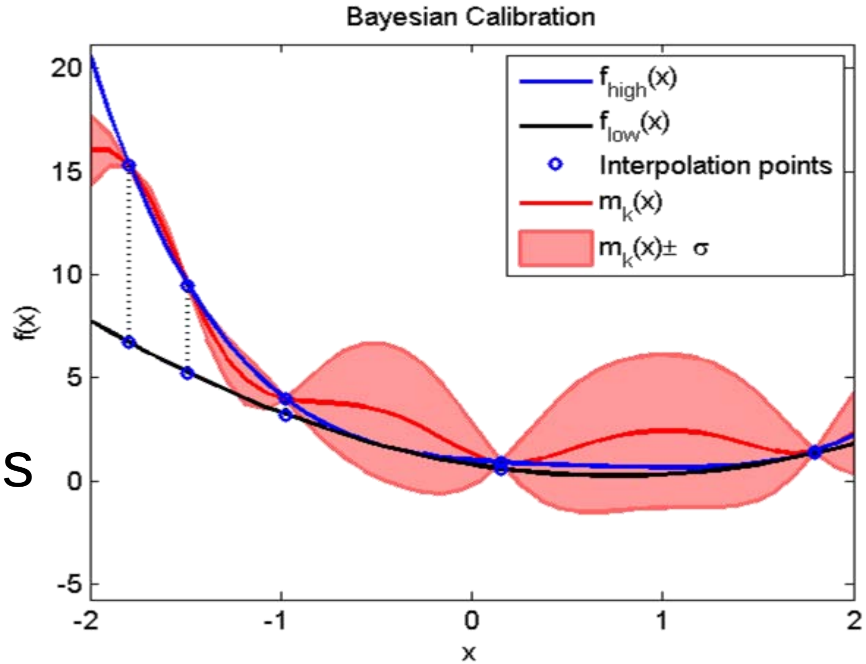
$$m_k(x) = {}^1(x) + {}^{-T}x$$

- Estimate function values with a Kriging model
  - Predicts mean and variance
- Evaluate function at "maximum expected improvement location(s)" and update model

Bayesian Model Calibration

$$f_{high}(\mathbf{x}) \approx m_k(\mathbf{x}) = f_{low}(\mathbf{x}) + \varepsilon_k(\mathbf{x})$$

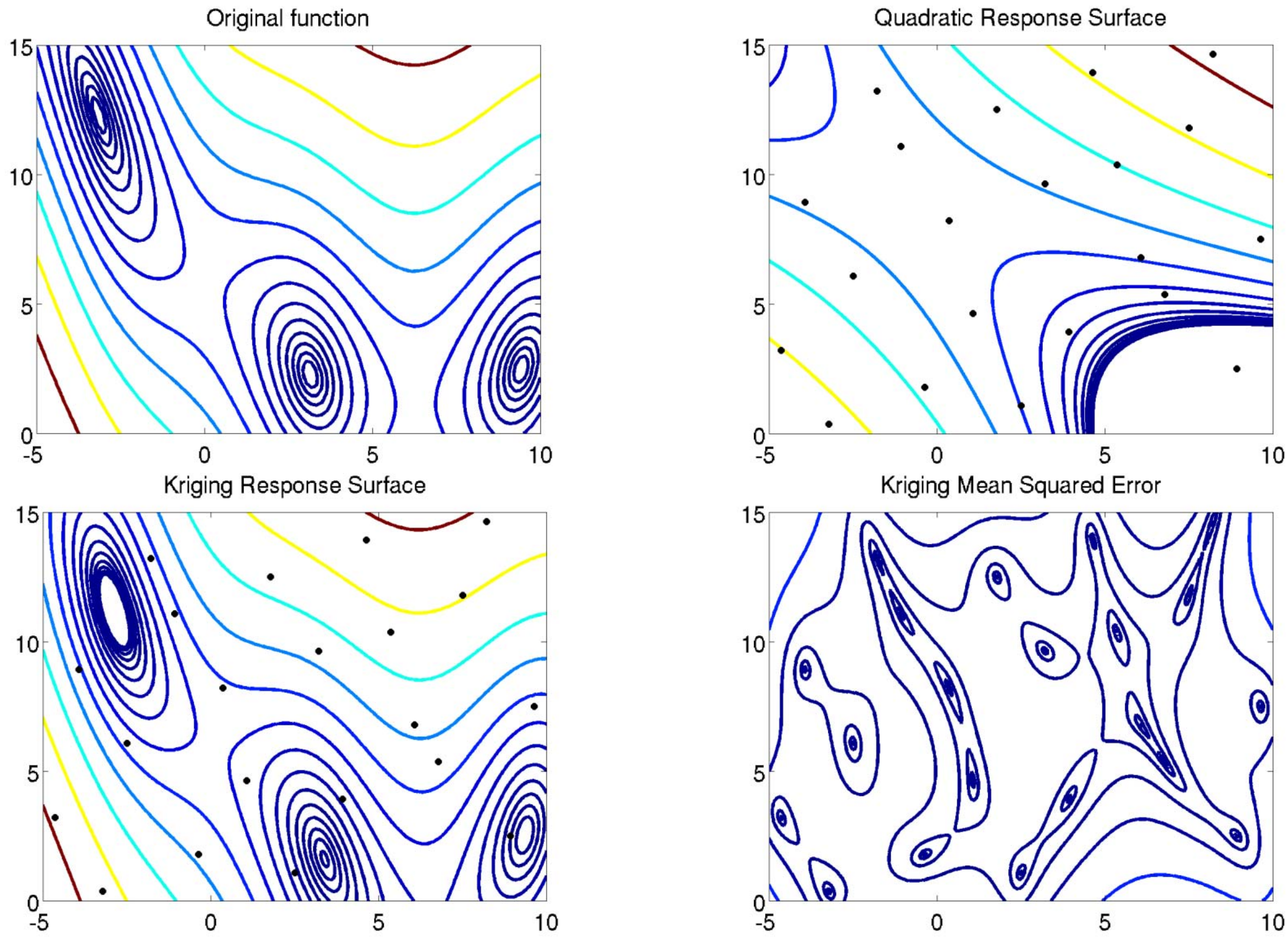
- Model the error between a high-fidelity and a low-fidelity function
  [Kennedy2000, 2001; Huang2006]



- If the low-fidelity function is "good", converges faster

- Global calibration procedure

Comparison of Data Fit Methods



Reduced-Basis Methods

Consider *r* feasible design vectors: $\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^r$

We could consider the desired design to be a linear combination of these basis vectors:

$$\mathbf{x}^* = \sum_{i=1}^{r} \alpha_i \mathbf{x}^i + \mathbf{x}^C$$

- α_i: scalar coefficient
- $\mathbf{x}^i$: basis vector
- $\mathbf{x}^C$: added for generality

Reduced-Basis Methods

We can now optimize $J(\mathbf{x})$ by finding the optimal values for the coefficients α_i.

dimension n → dimension r

- Do one full-order evaluation of resulting answer
- Approach is efficient if $r << n$
- Will give the true optimum only if $\mathbf{x}^*$ lies in the span of $\{\mathbf{x}^i\}$
- Basis vectors could be
  - previous designs
  - solutions over a particular range (DoE)
  - derived in some other way (e.g., proper orthogonal decomposition)

Reduced-Basis Example

Example using a reduced-basis approach (van der Plaats Fig 7-2): airfoil design for a unique application.

- Many airfoil shapes with known performance are available
- Design variables are (x,y) coordinates at chordwise locations (n~100)
- Use four basis airfoil shapes (low-speed airfoils) which contain the n geometry points
- Plus two basis shapes which allow trailing edge thickness to vary
- r=6 (r<<n)
- Optimize for high speed, maximum lift with a constraint on drag

Reduced-Basis Example

From Vanderplaats Figs. 7-2 and 7-3,

pg. 260

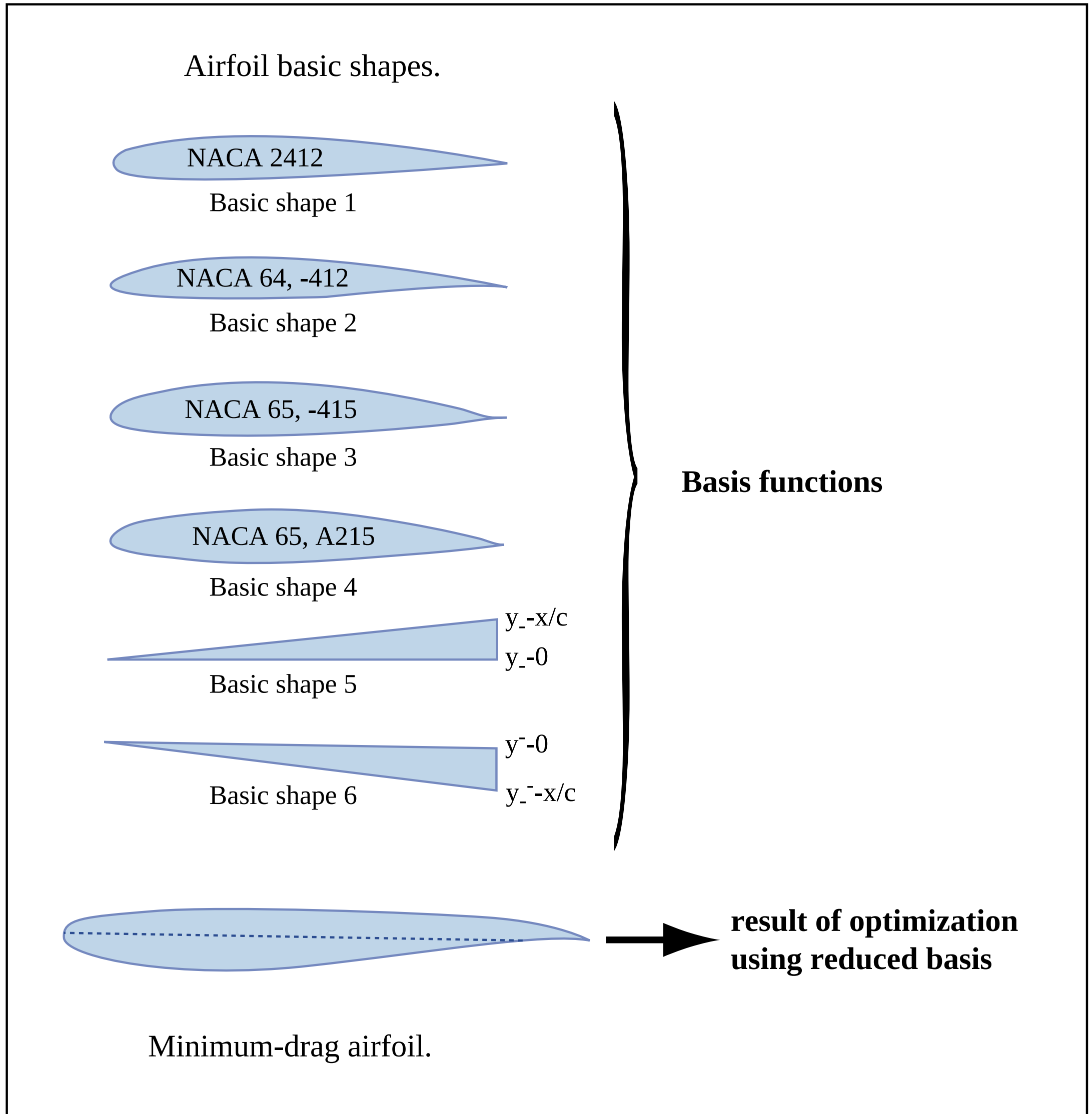


Image by MIT OpenCourseWare.

Proper Orthogonal Decomposition

(aka Karhunen-Loève expansions, Principal Components Analysis, Empirical Orthogonal Eigenfunctions, …)

Consider *K* snapshots $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_K \in \mathcal{R}^n$

(solutions at selected times or parameter values)

Form the snapshot matrix $X = [\mathbf{x}_1 \ \mathbf{x}_2 \ \ldots \ \mathbf{x}_K]$

Choose the *n* basis vectors $\mathbf{V} = [\mathbf{V}_1 \ \mathbf{V}_2 \cdots \mathbf{V}_n]$

to be left singular vectors of the snapshot matrix, with singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq \sigma_{n+1} \geq \cdots \geq \sigma_K$

This is the optimal projection in a least squares sense:

$$\min_V \sum_{i=1}^{K} ||\mathbf{x}_i - VV^T\mathbf{x}_i||_2^2 = \sum_{i=n+1}^{K} \sigma_i^2$$

Multifidelity Methods

- Sometimes there is more than one model for the same system
  - e.g. Navier Stokes and thin-airfoil theory for wing design, finite-element and beam theory for structural design
- Low-fidelity model may provide good information over a wide range, at much lower computational cost
- Would like to find optimum of high-fidelity problem, but use low-fidelity model most of the time

A Hierarchy of Models

Images of Figure 1b, 4, and 8 removed due to copyright restrictions.
Figures from: Choi, S, Alonso, JJ, Kim, S., Kroo, IM. Two-level multi-fidelity design optimization studies for supersonic jets. 43th AIAA Aerospace Sciences Meeting & Exhibit. January 2005.

Image of Low-fidelity EM and High fidelity EM models removed due to copyright restrictions.

Trust-Region Model Management

- A rigorous method for determining when to use high-fidelity function calls
- Solves a series of subproblems:

Minimize $\hat{J}^k(\mathbf{x})$

Subject to $\hat{g}^k(\mathbf{x}) \leq 0$

$$\left\| \mathbf{x} - \mathbf{x}_c^k \right\|_\infty \leq \Delta^k$$

Several methods exist to handle the approximation of constraints.

$\mathbf{x}_c^k$: center point of trust region at iteration k

Δ^k : size of trust region at iteration k

Trust-Region Model Management

- Size of trust region updated depending on how well surrogate predicts high-fidelity function value

- Merit function $\Gamma\left[J(\mathbf{x}), g(\mathbf{x})\right]$

- Ratio of actual to predicted improvement:

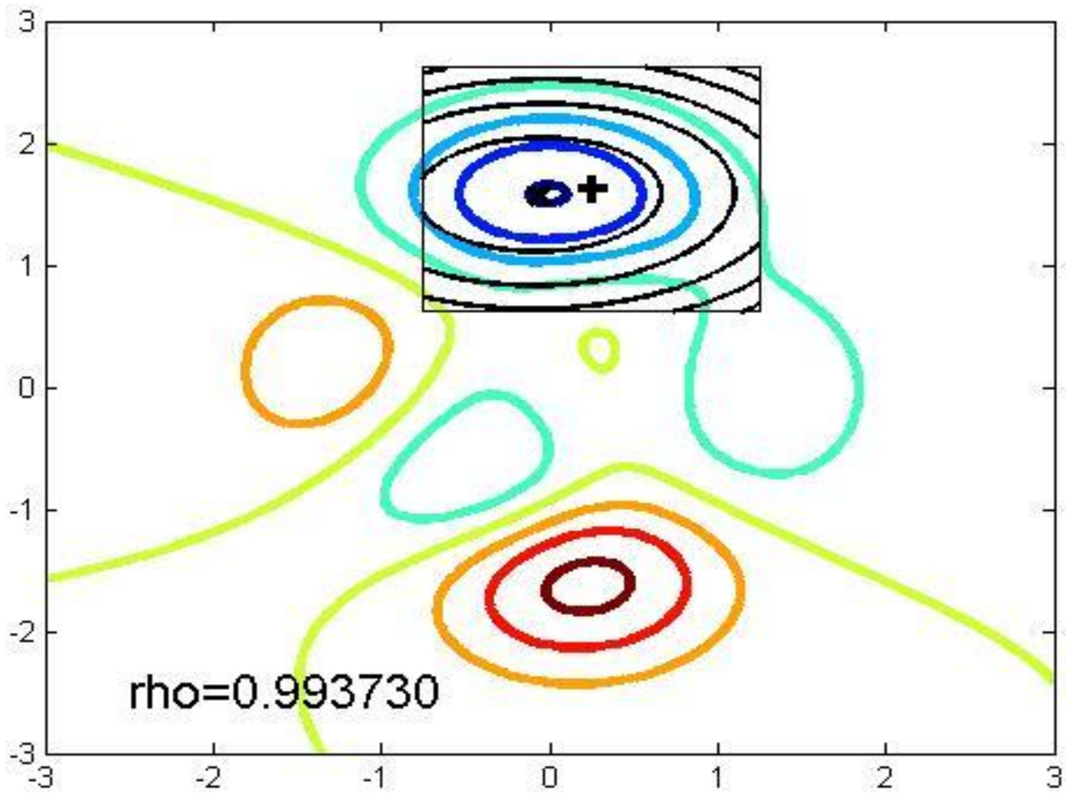
$$\rho^k = \frac{\Gamma\left(\mathbf{x}_c^k\right) - \Gamma\left(\mathbf{x}_*^k\right)}{\Gamma\left(\hat{\mathbf{x}}_c^k\right) - \hat{\Gamma}\left(\mathbf{x}_*^k\right)}$$

Trust-Region Model Management

- Trust region size update rules:

$\rho^k \le 0$	Reject step	$\Delta^{k+1} \equiv 0.5\Delta^k$
$0 < \rho^k \le 0.1$	Accept step	$\Delta^{k+1} \equiv 0.5\Delta^k$
$0.1 < \rho^k < 0.75$	Accept step	$\Delta^{k+1} \equiv \Delta^k$
$0.75 \le \rho^k$	Accept step	$\Delta^{k+1} \equiv 2\Delta^k$

Trust-Region Demonstration



Trust-Region Model Management

- Calls high-fidelity analysis once per iteration
- Calls surrogate analysis many times per iteration
- Provably convergent to local minimum of high fidelity function *if* surrogate is first-order accurate at center of trust region
- Extensions to the case of $\mathbf{x} \neq \hat{\mathbf{x}}$ in Robinson *et al.* (2008).
- Derivative-free approaches in Conn et al. (2009)

Corrections

- Include corrections in order to enforce consistency and gain provable convergence of trust-region approach

- Additive Correction:

$$\hat{J}(\mathbf{x}) = J_{lo}(\mathbf{x}) + \alpha(\mathbf{x})$$

- Multiplicative Correction:

$$\hat{J}(\mathbf{x}) = J_{lo}(\mathbf{x})\beta(\mathbf{x})$$

surrogate model

low-fidelity model

Multifidelity Optimization

- Combines several elements:
  - Trust regions
  - Bayesian model calibration
  - Adaptive sampling
  - Surrogate models (e.g., interpolation models using Kriging)
  - Estimation theory

- Active area of research

Combining Estimates of Multifidelity Models

- Use Kalman filtering approach to compute combine estimate
- Maximum likelihood estimate weights each model according to its variance (pay more attention to models in which we have more confidence)

$$1_{\mathrm{est}}(\mathbf{x}) = 1_{\mathrm{med}}(\mathbf{x}) \frac{\text{¾}^2_{\mathrm{low}}(\mathbf{x})}{\text{¾}^2_{\mathrm{low}}(\mathbf{x}) + \text{¾}^2_{\mathrm{med}}(\mathbf{x})} + 1_{\mathrm{low}}(\mathbf{x}) \frac{\text{¾}^2_{\mathrm{med}}(\mathbf{x})}{\text{¾}^2_{\mathrm{low}}(\mathbf{x}) + \text{¾}^2_{\mathrm{med}}(\mathbf{x})}$$

$$\frac{1}{\text{¾}^2_{\mathrm{est}}(\mathbf{x})} = \frac{1}{\text{¾}^2_{\mathrm{low}}(\mathbf{x})} + \frac{1}{\text{¾}^2_{\mathrm{med}}(\mathbf{x})}:$$

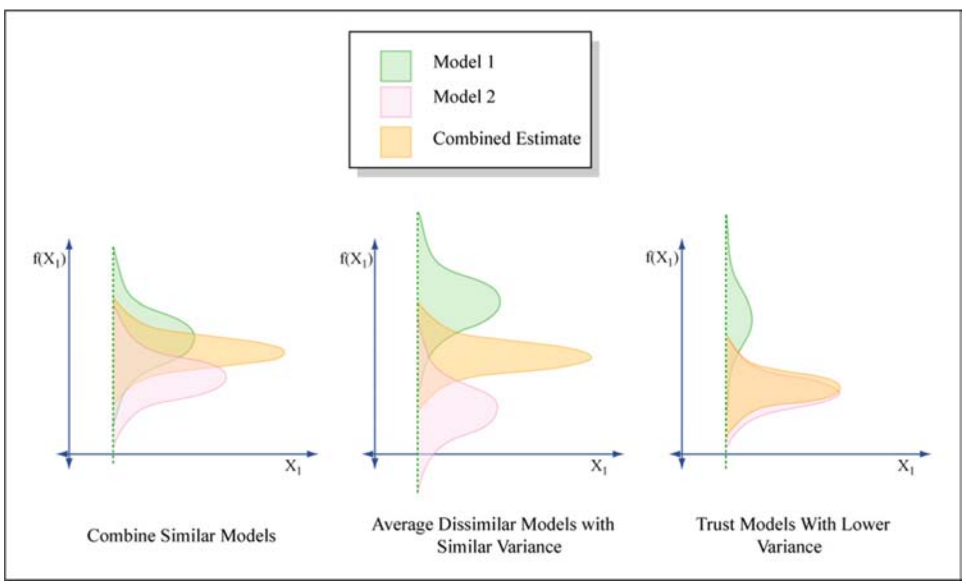


Image by MIT OpenCourseWare.

- Extends naturally to case with more than two models; much more efficient than nesting (March 2010)

Lecture Summary

- A number of ways to create approximations, or surrogates
- Each has its own area of application, advantages, and disadvantages
- Data fit surrogates
  - Polynomial response surfaces
  - Kriging
- Model order reduction
  - Reduced basis
  - Proper orthogonal decomposition
- Multifidelity methods

References

Alexandrov, N., Dennis, J.E., Lewis, R.M. and Torczon, V., “A trust region framework for managing the use of approximation models in optimization”, NASA CR-201745, ICASE Report No. 97-50, October 1997.

Barthelemy, J-F. M. and Haftka, R.T., “Approximation concepts for optimum structural design – a review”, *Structural Optimization*, 5:129-144, 1993.

Conn, A.R., Scheinberg, K. and Vicente, L., “Global Convergence of General Derivative-Free Trust-Region Algorithms to First- and Second-Order Critical Points,” *SIAM Journal of Optimization*, Vol. 20, No.1, pp. 387-415, 2009.

Gill, P.E., Murray,W. and Wright, M.H., *Practical Optimization*, Academic Press, 1986.

Giunta, A.A. and Watson, L.T.,”A comparison of approximation modeling techniques: polynomial versus interpolating models”, AIAA Paper 98-4758, 1998.

Jones, D.R., “A taxonomy of global optimization methods based on response surfaces,” Journal of Global Optimization, 21, 345-383, 2001.

LeGresley, P.A. and Alonso, J.J., “Airfoil design optimization using reduced order models based on proper orthogonal decomposition”, AIAA Paper 2000-2545, 2000.

March, A. and Willcox, K., “A Provably Convergent Multifidelity Optimization Algorithm not Requiring High-Fidelity Derivatives,” AIAA-2010-2912, presented at 3rd MDO Specialist Conference, Orlando, FL, April 12-15, 2010.

Robinson, T., Willcox, K., Eldred, M., and Haimes, R. “Multifidelity Optimization for Variable-Complexity Design,” *AIAA Journal*, Vol.46, No.11, pp. 2814-2822, 2008.

Vanderplaats, G.N., *Numerical Optimization Techniques for Engineering Design,* Vanderplaats R&D, 1999.

ESD.77 / 16.888 Multidisciplinary System Design Optimization
Spring 2010