# Outline

## *Intro to Surrogate-Based Optimization*

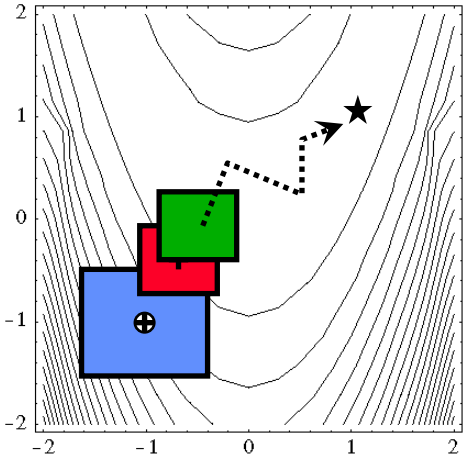
- Data fit case
- Model hierarchy case
- ROM case

## *Algorithmic Components*

- Approximate subproblem
- Iterate acceptance logic
- Merit functions
- Constraint relaxation
- Other: convergence, corrections, …

## *Computational Experiments*

- Barnes
- CUTE
- MEMS design



# Introduction to Surrogate-Based Optimization

## *Purpose:*

- **Reduce the number of expensive, high-fidelity simulations by using a succession of approximate (surrogate) models**
- **Approximations generally have a limited range of validity**
- **Trust regions adaptively manage this range based on efficacy during opt**
- **With trust region globalization and local 1$^{st}$-order consistency, SBO algorithms can be *provably-convergent***

## *Surrogate models of interest:*

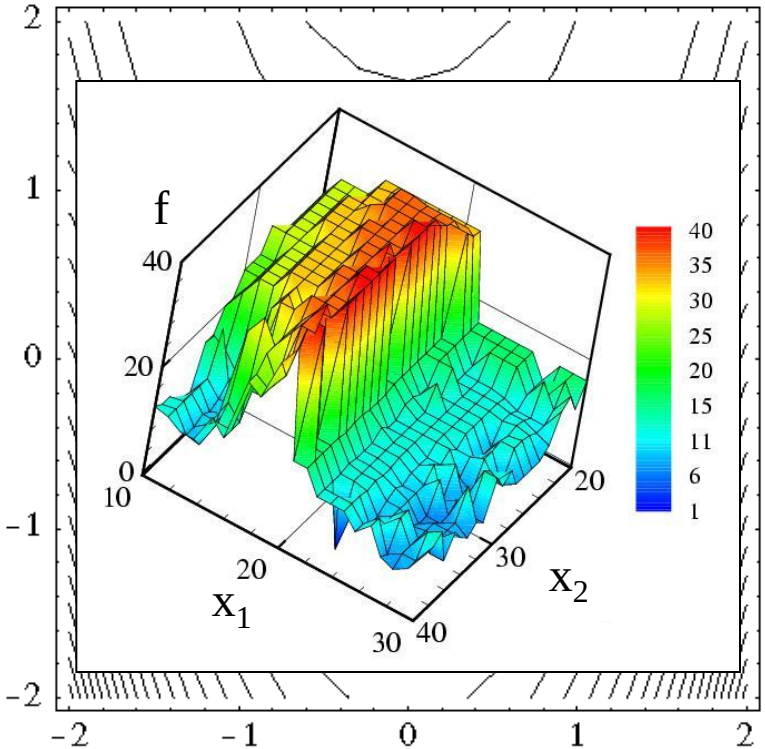
- ***Data fits*** **(local, multipoint, global)**
- ***Multifidelity*** **(special case: *multigrid* optimization)**
- ***Reduced-order models***

# Trust Region Surrogate-Based Optimization (SBO) – Data Fit Case

## *Data fit surrogates:*

- ***Global:*** **polynomial resp surf, NN, splines, kriging/GP, radial basis fn**
- ***Local:*** **1st/2nd-order Taylor series**
- ***Multipoint*****: two-point exponential approx (TPEA), two-point adaptive nonlinearity approx (TANA)**

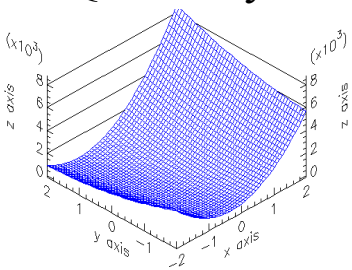
## *Sequence of trust regions*



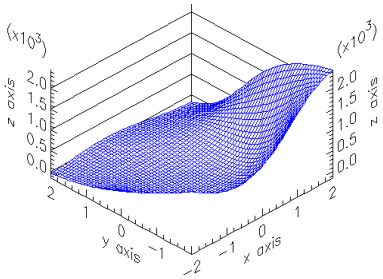

## *(Global) data fits in SBO:*

- **Smoothing: extract desired global trend from noisy data**
- **DACE: number of design variables limited to O($10^1$)**
- **Local consistency must be balanced with global accuracy**
  - **Constrained LLS**
  - **TS w/ global Hessian estimation**

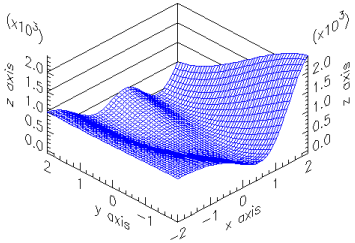
Quad Poly



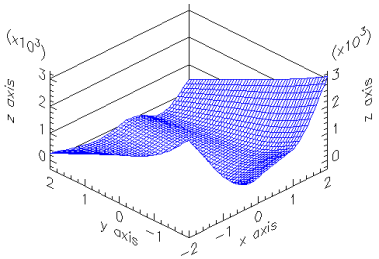

Neural Network




Kriging

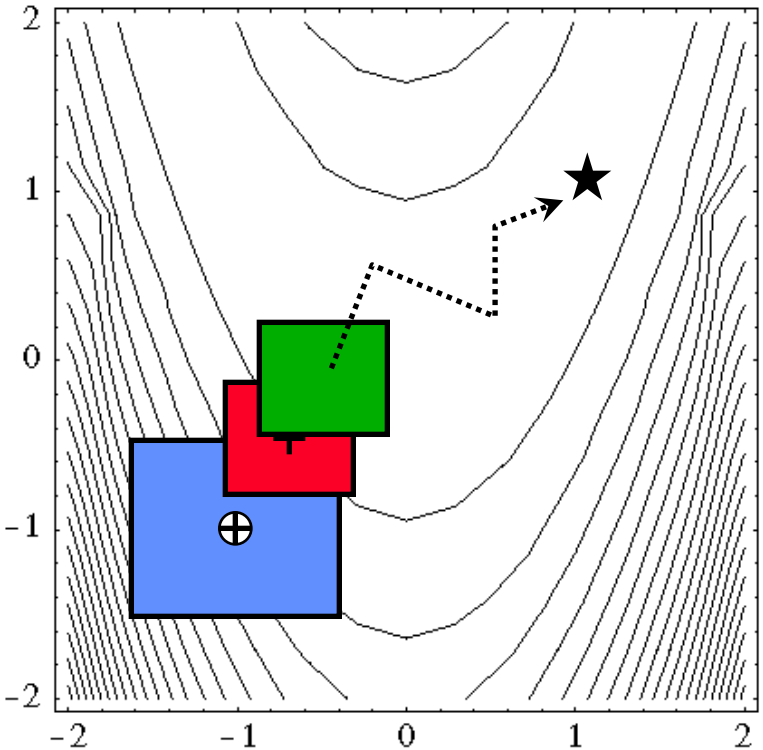



Splines



# Trust Region SBO – Multifidelity Case

*Sequence of trust regions*




***Multifidelity surrogates:***

- Coarsened discretizations (h or p)
- Loosened solver tolerances
- Omitted physics: e.g., Euler CFD, panel methods

***Multifidelity models in SBO:***

- Truth eval only at center
  - HF evals scale better with des. vars.
  - Smoothing character is lost
    → requires well-behaved LF model
- Correction quality is ***crucial***
  - Additive: $f(x)_{HF} = f(x)_{LF} + \alpha(x)$
  - Multiplicative: $f(x)_{HF} = f(x)_{LF}\, \beta(x)$
  - Combined (convex combination of add/mult)
- May require design vector mapping (Robinson)
  - Space mapping
  - Corrected space mapping
  - POD mapping
  - Hybrid POD space mapping

Sandia National Laboratories

# Trust Region SBO – ROM Case

***ROM surrogates:*** **capture critical aspects of HF simulations by projecting the system response using a reduced basis**

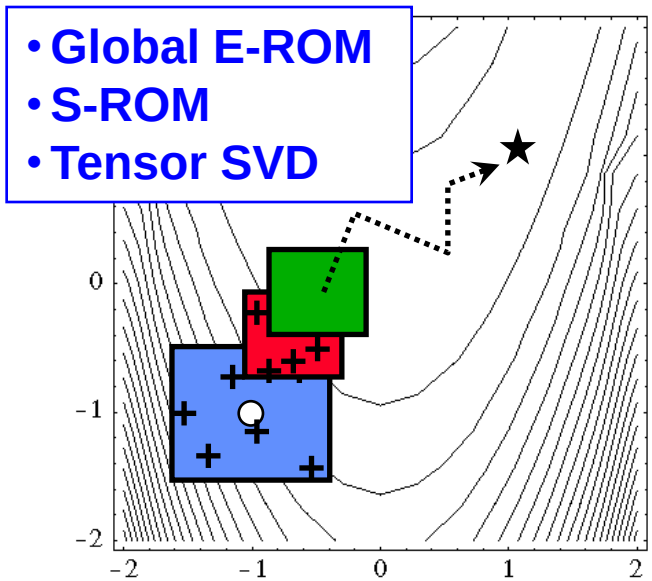
- **Spectral decomposition (str. dynamics)**
- **POD/PCA w/ SVD (CFD, image analysis)**
- **KL (random fields, stoch. proc.)**

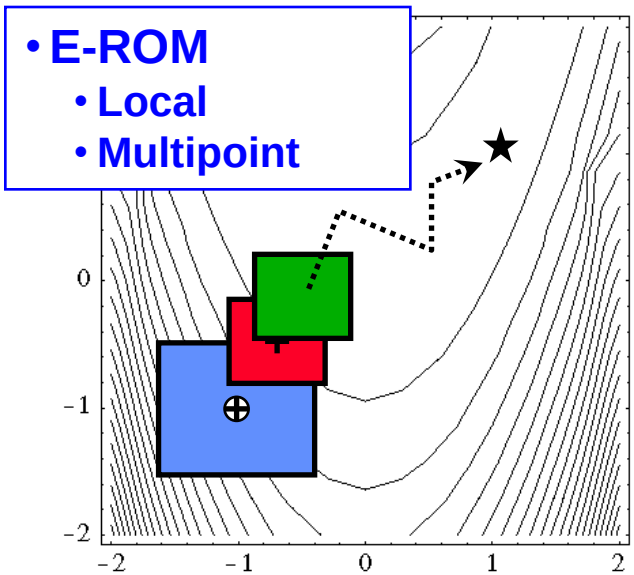
$$M\ddot{u} + C\dot{u} + Ku = f$$

$$u(t) = \sum_{j=1}^{k} \eta_j(t)\phi_{(j)} = \Phi\eta(t)$$

$$\Phi^T M\Phi\ddot{\eta}(t) + \Phi^T C\Phi\dot{\eta}(t) + \Phi^T K\Phi\eta(t) = \Phi^T f(t)$$

$$M_R\ddot{\eta} + C_R\dot{\eta} + K_R\eta = f_R$$




***ROMs in SBO:*** **Key issue is accurately capturing effect of parameter changes.**

*Many ROM-based opt. methods:* restricted; design changes = ROM inputs (e.g., Dirichlet BC control).

*General case:* design changes influence system matrices → updates to basis and projected model

→ Extended ROM (E-ROM), Spanning ROM (S-ROM), tensor SVD

Sandia National Laboratories

# Outline

*Intro to Surrogate-Based Optimization*

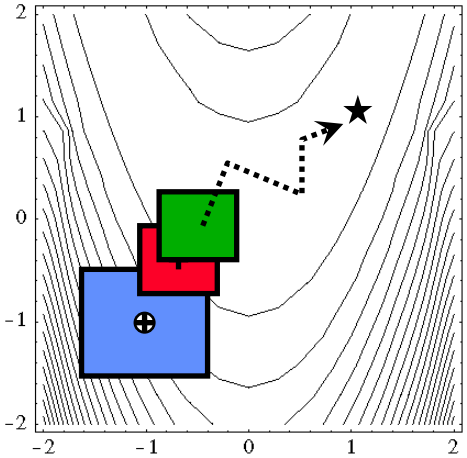
- Data fit case
- Model hierarchy case
- ROM case

***Algorithmic Components***

- **Approximate subproblem**
- **Iterate acceptance logic**
- **Merit functions**
- **Constraint relaxation**
- **Other: convergence, corrections, …**

***Computational Experiments***

- **Barnes**
- **CUTE**
- **MEMS design**




Sandia National Laboratories

# SBO Algorithm Components: Approximate Subproblem Formulations

| | Original Objective | Lagrangian | Augmented Lagrangian |
|---|---|---|---|
| No constraints | | | TRAL |
| Linearized constraints | | SQP-like | |
| Original constraints | Direct surrogate | | IPTRSAO |

$$\begin{array}{ll} \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & \mathbf{g}_l \leq \mathbf{g}(\mathbf{x}) \leq \mathbf{g}_u \\ & \mathbf{h}(\mathbf{x}) = \mathbf{h}_t \\ & \mathbf{x}_l \leq \mathbf{x} \leq \mathbf{x}_u \end{array}$$

$$\begin{array}{ll} \text{minimize} & \hat{\Phi}^k(\mathbf{x}) \\ \text{subject to} & \| \mathbf{x} - \mathbf{x}_c^k \|_\infty \leq \Delta^k \end{array}$$

**(TRAL)**

$$\begin{array}{ll} \text{minimize} & \hat{\Phi}^k(\mathbf{x}) \\ \text{subject to} & \mathbf{g}_l \leq \hat{\mathbf{g}}^k(\mathbf{x}) \leq \mathbf{g}_u \\ & \hat{\mathbf{h}}^k(\mathbf{x}) = \mathbf{h}_t \\ & \| \mathbf{x} - \mathbf{x}_c^k \|_\infty \leq \Delta^k \end{array}$$

**(IPTRSAO)**

$$\begin{array}{ll} \text{minimize} & \hat{f}^k(\mathbf{x}) \\ \text{subject to} & \mathbf{g}_l \leq \hat{\mathbf{g}}^k(\mathbf{x}) \leq \mathbf{g}_u \\ & \hat{\mathbf{h}}^k(\mathbf{x}) = \mathbf{h}_t \\ & \| \mathbf{x} - \mathbf{x}_c^k \|_\infty \leq \Delta^k \end{array}$$

**(Direct surrogate)**

$$\begin{array}{ll} \text{minimize} & \hat{\Phi}^k(\mathbf{x}) \\ \text{subject to} & \mathbf{g}_l \leq \hat{\mathbf{g}}^k(\mathbf{x}_c^k) + \nabla \hat{\mathbf{g}}^k(\mathbf{x}_c^k)^T(\mathbf{x} - \mathbf{x}_c^k) \leq \mathbf{g}_u \\ & \hat{\mathbf{h}}^k(\mathbf{x}_c^k) + \nabla \hat{\mathbf{h}}^k(\mathbf{x}_c^k)^T(\mathbf{x} - \mathbf{x}_c^k) = \mathbf{h}_t \\ & \| \mathbf{x} - \mathbf{x}_c^k \|_\infty \leq \Delta^k \, . \end{array}$$

**(SQP-like)**

# SBO Algorithm Components: Iterate Acceptance Logic

## Trust region ratio

- **Ratio of actual to predicted improvement**

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

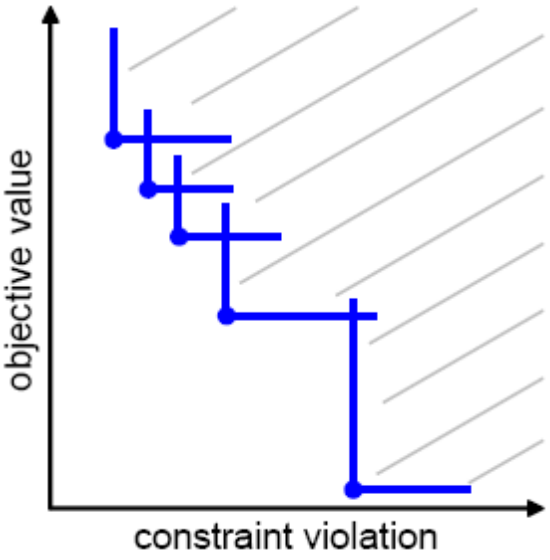
| Ratio Value | Surrogate Accuracy | Iterate Acceptance | Trust Region Sizing |
|---|---|---|---|
| $\rho^k \leq 0$ | poor | reject step | shrink |
| $0 < \rho^k \leq 0.25$ | marginal | accept step | shrink |
| $0.25 < \rho^k < 0.75$ or $\rho^k > 1.25$ | moderate | accept step | retain |
| $0.75 \leq \rho^k \leq 1.25$ | good | accept step | expand[c] |

## Filter method

- **Uses concept of Pareto optimality applied to objective and constraint viol**
- **For new iterate to be acceptable to filter:**

$$\boxed{\text{either} \quad f < f^{(i)} \quad \text{or} \quad c < c^{(i)}}$$

- **Still need logic to adapt TR size for accepted steps**



# SBO Algorithm Components: Merit function selections

## Penalty

$$\begin{aligned}\Phi(\mathbf{x}, r_p) &= f(\mathbf{x}) + r_p \mathbf{g}^+(\mathbf{x})^T \mathbf{g}^+(\mathbf{x}) + r_p \mathbf{h}(\mathbf{x})^T \mathbf{h}(\mathbf{x}) \\ r_p &= e^{(k+\text{offset})/10}\end{aligned}$$

## Adaptive penalty

**adapts $r_p$ using monotonic increases in the iteration `offset` value in order to accept any iterate that reduces the constraint violation**
**→ mimics a (monotonic) filter**

## Lagrangian

$$\Phi(\mathbf{x}, \boldsymbol{\lambda}_g, \boldsymbol{\lambda}_h) = f(\mathbf{x}) + \boldsymbol{\lambda}_g^T \mathbf{g}^+(\mathbf{x}) + \boldsymbol{\lambda}_h^T \mathbf{h}(\mathbf{x})$$

Multiplier estimation:

$$\nabla_x \Phi(\mathbf{x}, \boldsymbol{\lambda}_g, \boldsymbol{\lambda}_h) = \nabla_x f(\mathbf{x}) + \boldsymbol{\lambda}_g^T \nabla_x \mathbf{g}^+(\mathbf{x}) + \boldsymbol{\lambda}_h^T \nabla_x \mathbf{h}(\mathbf{x})$$

$$\mathbf{A}\boldsymbol{\lambda} = -\nabla_x f$$ (bound-constrained LLS: min KKT residual)

## Augmented Lagrangian

$$\begin{aligned}\Phi(\mathbf{x}, \boldsymbol{\lambda}_\psi, \boldsymbol{\lambda}_h, r_p) &= f(\mathbf{x}) + \boldsymbol{\lambda}_\psi^T \boldsymbol{\psi}(\mathbf{x}) + r_p \boldsymbol{\psi}(\mathbf{x})^T \boldsymbol{\psi}(\mathbf{x}) + \boldsymbol{\lambda}_h^T \mathbf{h}(\mathbf{x}) + r_p \mathbf{h}(\mathbf{x})^T \mathbf{h}(\mathbf{x}) \\ \psi_i &= \max\left\{ g_i, -\frac{\lambda_{\psi_i}}{2r_p} \right\}\end{aligned}$$

(derived from elimination of slacks)

Multiplier estimation:

$$\begin{aligned}\boldsymbol{\lambda}_\psi^{k+1} &= \boldsymbol{\lambda}_\psi^k + 2r_p \boldsymbol{\psi}(\mathbf{x}) \\ \boldsymbol{\lambda}_h^{k+1} &= \boldsymbol{\lambda}_h^k + 2r_p \mathbf{h}(\mathbf{x})\end{aligned}$$

(drives Lagrangian gradient to KKT)

Sandia National Laboratories

# SBO Algorithm Components: Constraint Relaxation

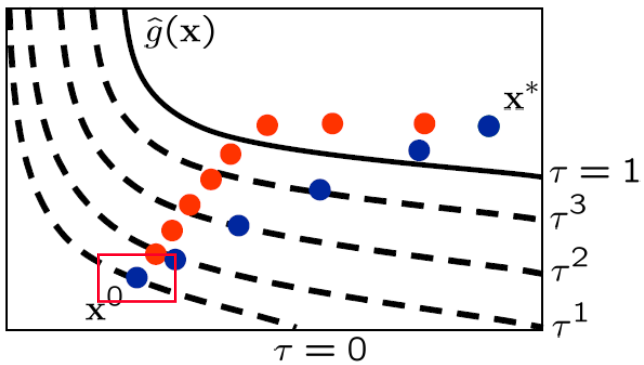
| | Original Objective | Lagrangian | Augmented Lagrangian |
|---|---|---|---|
| No constraints | | | TRAL |
| Linearized constraints | | SQP-like | |
| Original constraints | Direct surrogate | | IPTRSAO |

**Composite step**

- **Byrd-Omojokun**
- **Celis-Dennis-Tapia**
- **MAESTRO**

**Homotopy**

- ***Heuristic***
- **Probability one**




**Constrained approx subproblem**

$$\begin{array}{ll} \text{minimize} & \hat{f}^k(\mathrm{x}) \\ \text{subject to} & \mathrm{g}_l \le \hat{\mathrm{g}}^k(\mathrm{x}) \le \mathrm{g}_u \\ & \hat{\mathrm{h}}^k(\mathrm{x}) = \mathrm{h}_t \\ & \| \mathrm{x} - \mathrm{x}_c^k \|_\infty \le \Delta^k \end{array}$$

**Relax nonlinear constraints**

$$\begin{array}{lcl} \tilde{\mathrm{g}}^k(\mathrm{x},\tau) & = & \hat{\mathrm{g}}^k(\mathrm{x}) + (1-\tau)\mathrm{b}_g \\ \tilde{\mathrm{h}}^k(\mathrm{x},\tau) & = & \hat{\mathrm{h}}^k(\mathrm{x}) + (1-\tau)\mathrm{b}_h \end{array}$$

**Relaxed subproblem**

$$\begin{array}{ll} \text{minimize} & \hat{f}^k(\mathrm{x}) \\ \text{subject to} & \mathrm{g}_l \le \tilde{\mathrm{g}}^k(\mathrm{x},\tau^k) \le \mathrm{g}_u \\ & \tilde{\mathrm{h}}^k(\mathrm{x},\tau^k) = \mathrm{h}_t \\ & \| \mathrm{x} - \mathrm{x}_c^k \|_\infty \le \Delta^k \end{array}$$

**Estimate $\tau$:**

$$\begin{array}{ll} \text{maximize} & \tau^k \\ \text{subject to} & \mathrm{g}_l \le \tilde{\mathrm{g}}^k(\mathrm{x},\tau^k) \le \mathrm{g}_u \\ & \tilde{\mathrm{h}}^k(\mathrm{x},\tau^k) = \mathrm{h}_t \\ & \| \mathrm{x} - \mathrm{x}_c^k \|_\infty \le \Delta^k \\ & \tau^k \ge 0 \end{array}$$

$$\tau^k = \min\left\{1, \tau^{k-1} + \alpha\left(\tau^k_{\max} - \tau^{k-1}\right)\right\}$$
$$0 < \alpha < 1$$

**Leave portion of TR volume feasible for relaxed subproblem**

Sandia National Laboratories

# SBO Algorithm Components: Convergence, Corrections, …

- **Soft convergence assessment:** diminishing returns
- **Hard conv. assessment:** bound constrained-LLS for updating multipliers for standard Lagrangian minimizes the KKT residual

$$\nabla_x \Phi(\mathbf{x}, \boldsymbol{\lambda}_g, \boldsymbol{\lambda}_h) = \nabla_x f(\mathbf{x}) + \boldsymbol{\lambda}_g^T \nabla_x \mathbf{g}^+(\mathbf{x}) + \boldsymbol{\lambda}_h^T \nabla_x \mathbf{h}(\mathbf{x})$$
$$\mathbf{A}\boldsymbol{\lambda} = -\nabla_x f$$

If feasible and KKT residual < tolerance, then hard convergence achieved

- **Corrections (MA&O 2004)**
  - Additive $\hat{f_{hi_\alpha}}(\mathbf{x}) = f_{lo}(\mathbf{x}) + \alpha(\mathbf{x})$
  - Multiplicative $\hat{f_{hi_\beta}}(\mathbf{x}) = f_{lo}(\mathbf{x})\beta(\mathbf{x})$
  - Combined $\hat{f_{hi_\gamma}}(\mathbf{x}) = \gamma \hat{f_{hi_\alpha}}(\mathbf{x}) + (1-\gamma)\hat{f_{hi_\beta}}(\mathbf{x})$

  - Zeroth-order
  - First-order
  - Second-order (full, FD, Quasi)

# Local Correction Derivations

Exact additive/multiplicative:

$$f_{hi}(\mathbf{x}) = f_{lo}(\mathbf{x}) + A(\mathbf{x}) \qquad f_{hi}(\mathbf{x}) = f_{lo}(\mathbf{x})B(\mathbf{x})$$

Then:

$$A(\mathbf{x}) = f_{hi}(\mathbf{x}) - f_{lo}(\mathbf{x}) \qquad B(\mathbf{x}) = \frac{f_{hi}(\mathbf{x})}{f_{lo}(\mathbf{x})}$$

Approximate $A$,$B$ with 2nd-order Taylor series centered at $\mathbf{x_c}$:

$$\alpha(\mathbf{x}) = \underline{A(\mathbf{x_c})} + \underline{\nabla A(\mathbf{x_c})}^T(\mathbf{x}-\mathbf{x_c}) + \frac{1}{2}(\mathbf{x}-\mathbf{x_c})^T\underline{\nabla^2 A(\mathbf{x_c})}(\mathbf{x}-\mathbf{x_c})$$

$$\beta(\mathbf{x}) = \underline{B(\mathbf{x_c})} + \underline{\nabla B(\mathbf{x_c})}^T(\mathbf{x}-\mathbf{x_c}) + \frac{1}{2}(\mathbf{x}-\mathbf{x_c})^T\underline{\nabla^2 B(\mathbf{x_c})}(\mathbf{x}-\mathbf{x_c})$$

where:

$$A(\mathbf{x_c}) = f_{hi}(\mathbf{x_c}) - f_{lo}(\mathbf{x_c}) \qquad B(\mathbf{x_c}) = \frac{f_{hi}(\mathbf{x_c})}{f_{lo}(\mathbf{x_c})}$$

$$\nabla A(\mathbf{x_c}) = \nabla f_{hi}(\mathbf{x_c}) - \nabla f_{lo}(\mathbf{x_c}) \qquad \nabla B(\mathbf{x_c}) = \frac{1}{f_{lo}(\mathbf{x_c})}\nabla f_{hi}(\mathbf{x_c}) - \frac{f_{hi}(\mathbf{x_c})}{f_{lo}^2(\mathbf{x_c})}\nabla f_{lo}(\mathbf{x_c})$$

$$\nabla^2 A(\mathbf{x_c}) = \nabla^2 f_{hi}(\mathbf{x_c}) - \nabla^2 f_{lo}(\mathbf{x_c}) \qquad \nabla^2 B(\mathbf{x_c}) = \frac{1}{f_{lo}(\mathbf{x_c})}\nabla^2 f_{hi}(\mathbf{x_c}) - \frac{f_{hi}(\mathbf{x_c})}{f_{lo}^2(\mathbf{x_c})}\nabla^2 f_{lo}(\mathbf{x_c}) +$$

$$\frac{2f_{hi}(\mathbf{x_c})}{f_{lo}^3(\mathbf{x_c})}\nabla f_{lo}(\mathbf{x_c})\nabla f_{lo}^T(\mathbf{x_c}) -$$

$$\frac{1}{f_{lo}^2(\mathbf{x_c})}(\nabla f_{lo}(\mathbf{x_c})\nabla f_{hi}^T(\mathbf{x_c}) + \nabla f_{hi}(\mathbf{x_c})\nabla f_{lo}^T(\mathbf{x_c}))$$

# Multipoint Correction Derivation

Combined additive/multiplicative (notional):

$$f_{hi}(\mathbf{x}) = f_{lo}(\mathbf{x})B(\mathbf{x}) + A(\mathbf{x})$$

Assume a weighted sum of $\alpha$, $\beta$ corrections which preserves consistency:

$$\hat{f_{hi_\gamma}}(\mathbf{x}) = \gamma \hat{f_{hi_\alpha}}(\mathbf{x}) + (1-\gamma)\hat{f_{hi_\beta}}(\mathbf{x})$$

Enforce additional matching condition: $\hat{f_{hi_\gamma}}(\mathbf{x_p}) = f_{hi}(\mathbf{x_p})$ where $x_p$ $\begin{cases} \text{Previous } x_c \\ \text{Rejected } x^* \end{cases}$

$$\Rightarrow \quad \gamma = \frac{f_{hi}(\mathbf{x_p}) - \hat{f_{hi_\beta}}(\mathbf{x_p})}{\hat{f_{hi_\alpha}}(\mathbf{x_p}) - \hat{f_{hi_\beta}}(\mathbf{x_p})}$$

$$\hat{f_{hi_\alpha}}(\mathbf{x_p}) \rightarrow f_{hi}(\mathbf{x_p}) \quad \Rightarrow \quad \gamma \rightarrow 1$$

$$\hat{f_{hi_\beta}}(\mathbf{x_p}) \rightarrow f_{hi}(\mathbf{x_p}) \quad \Rightarrow \quad \gamma \rightarrow 0$$

Can be used to preserve global accuracy while satisfying local consistency

# Outline

*Intro to Surrogate-Based Optimization*

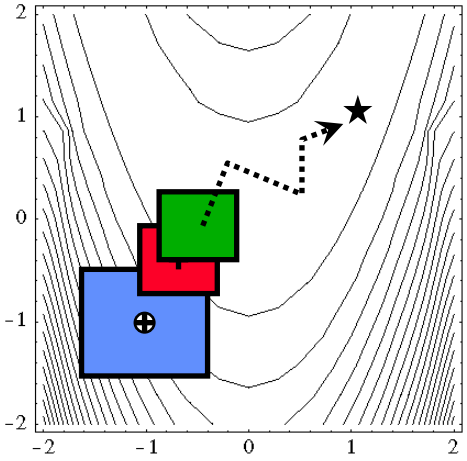
- Data fit case
- Model hierarchy case
- ROM case

*Algorithmic Components*

- Approximate subproblem
- Iterate acceptance logic
- Merit functions
- Constraint relaxation
- Other: convergence, corrections, …

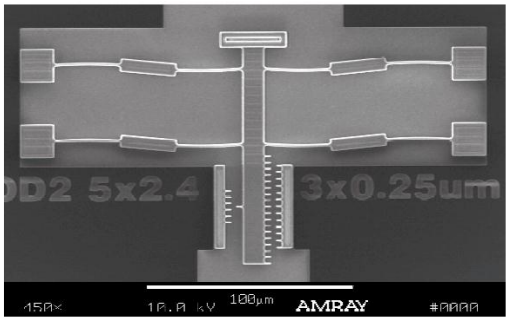
***Computational Experiments***

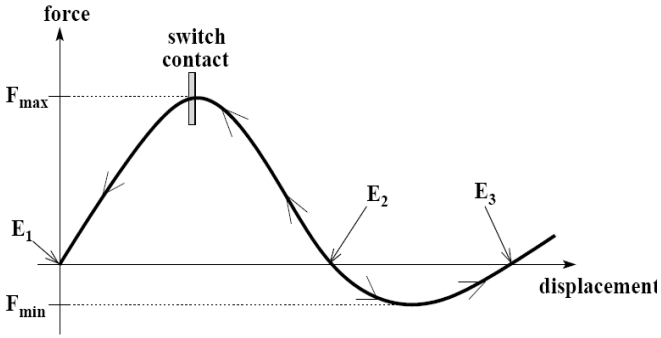
- **Barnes**
- **CUTE**
- **MEMS design**



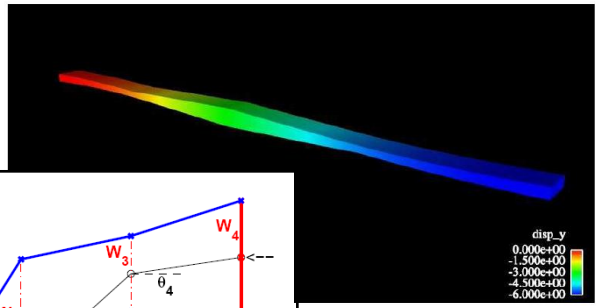

Sandia National Laboratories

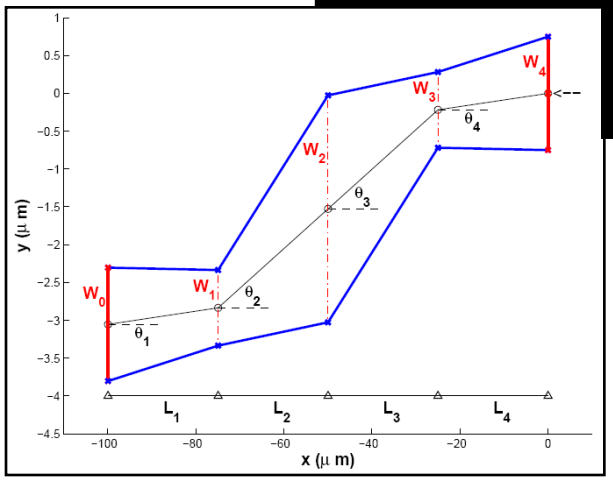
# Multifidelity SBO Testing: Bi-Stable MEMS Switch




**13 design vars:**
**$W_i$, $L_i$, $\theta_i$**




$$
\begin{array}{llcl}
\max & F_{max}(\mathrm{d}) & & \\
\text{s.t.} & E_2(\mathrm{d}) & \leq & 8 \\
 & S_{max}(\mathrm{d}) & \leq & 3000 \\
 & F_{min}(\mathrm{d}) & = & -5.
\end{array}
$$

| Element Type | Number of Elements 160 | 1200 |
|---|---|---|
| Linear | 5.0465 | 39.414 |
| Quadratic | 24.757 | 194.98 |

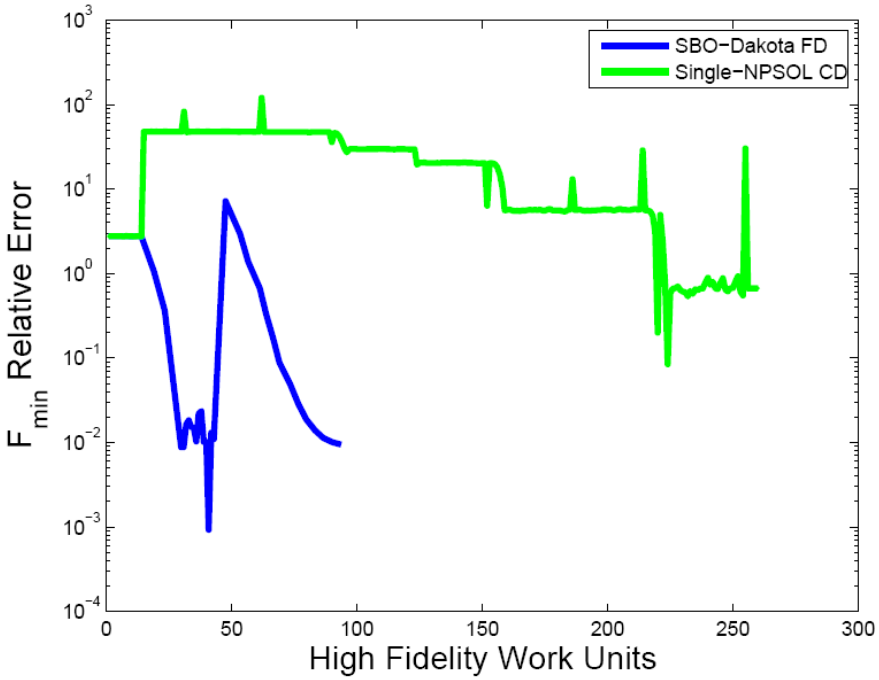
**38.6x**

Sandia National Laboratories

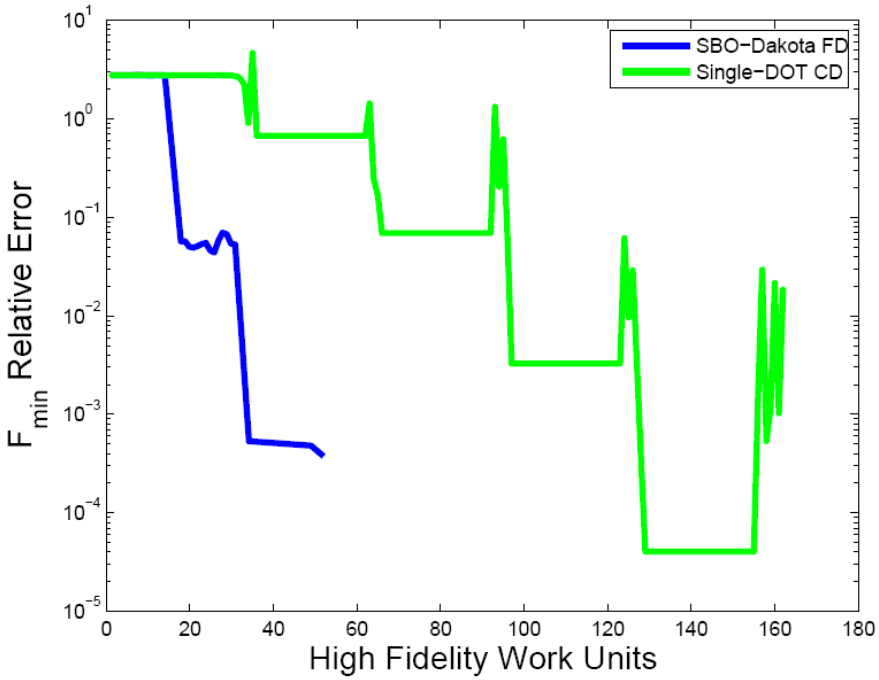
# MEMS Design Results: Single-Fidelity and Multifidelity

**Current best SBO approaches carried forward:**

- **Original, original approx subproblem**
- **Augmented Lagrangian merit function**
- **Trust region ratio acceptance**




**NPSOL**




**DOT**

Both single-fidelity fail with forward differences

Sandia National Laboratories