Stacking designs: designing multi-fidelity computer experiments with target predictive accuracy

Chih-Li Sung

Department of Statistics and Probability
Michigan State University

Joint work with Wenjia Wang (HKUST(GZ)), Yi Ji, Tao Tang, and Dr. Simon Mak (Duke)

UMass Amherst, April 17, 2024

Outline

Multi-Fidelity Simulations

- Computer models have been widely adopted to understand a real-world feature, phenomenon or event.
- Computer simulations are used to solve these models (e.g., finite element / finite difference) .

Multi-Fidelity Simulations

- Computer models have been widely adopted to understand a real-world feature, phenomenon or event.
- Computer simulations are used to solve these models (e.g., finite element / finite difference) .
- The simulation can be either
  - High-fidelity simulation: costly but close to the truth

Multi-Fidelity Simulations

- Computer models have been widely adopted to understand a real-world feature, phenomenon or event.
- Computer simulations are used to solve these models (e.g., finite element / finite difference) .
- The simulation can be either
  - High-fidelity simulation: costly but close to the truth
  - Low-fidelity simulation: cheaper but less accurate

Multi-Fidelity Simulations

- Computer models have been widely adopted to understand a real-world feature, phenomenon or event.
- Computer simulations are used to solve these models (e.g., finite element / finite difference) .
- The simulation can be either
  - High-fidelity simulation: costly but close to the truth
  - Low-fidelity simulation: cheaper but less accurate
  - (intermediate-fidelity simulation)

Motivated Example: Finite Element Simulations

- Thermal stress of jet engine turbine blade can be analyzed through a static structural computer model.
- The model can be *numerically* solved via finite element method.

Motivated Example: Finite Element Simulations

- Thermal stress of jet engine turbine blade can be analyzed through a static structural computer model.
- The model can be *numerically* solved via finite element method.
- **Input**: $\mathbf{x} = (x_1, x_2) = (\text{pressure}, \text{suction})$

Motivated Example: Finite Element Simulations

- Thermal stress of jet engine turbine blade can be analyzed through a static structural computer model.
- The model can be *numerically* solved via finite element method.
- **Input**: $\mathbf{x} = (x_1, x_2) = (\text{pressure}, \text{suction})$
- **Output**: $f(\mathbf{x})$: average of thermal stress
- e.g., $\mathbf{x} = (0.23, 0.71)$

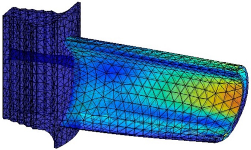


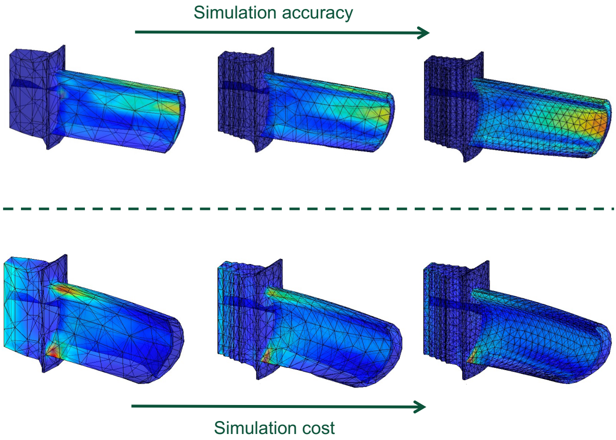
Figure: *average of thermal stress* $f(0.23, 0.71) = 10.5$

Multi-Fidelity Simulations via Mesh Configuration

less accurate but cheaper — accurate but expensive

$\mathbf{x} = (0.50, 0.50)$

$\mathbf{x} = (0.23, 0.71)$



Statistical Emulation

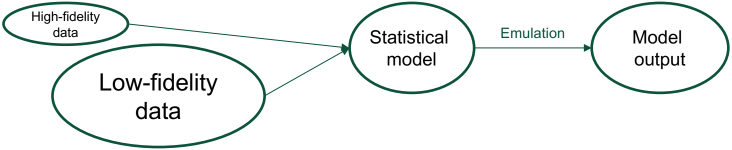
- Can we leverage both low- and high-fidelity simulations in order to

Statistical Emulation

- Can we leverage both low- and high-fidelity simulations in order to
  - maximize the accuracy of model predictions,
  - while minimizing the cost associated with the simulations?

Statistical Emulation

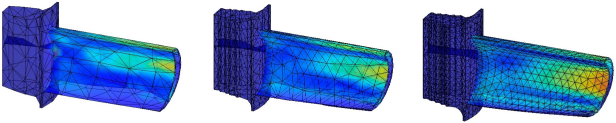
- Can we leverage both low- and high-fidelity simulations in order to
  - maximize the accuracy of model predictions,
  - while minimizing the cost associated with the simulations?
  - A cheaper statistical model emulating the model output based on the simulations with multiple fidelities
  - Often called emulator or surrogate model



Notation

fidelity level	1		2		3
output	$f_1(\mathbf{x})$		$f_2(\mathbf{x})$		$f_3(\mathbf{x})$
mesh size	h_1	$>$	h_2	$>$	h_3
cost	C_1	$<$	C_2	$<$	C_3

$\mathbf{x} = (0.50, 0.50)$



Existing Methods

- Modeling:
  - Co-kriging (Kennedy and O'Hagan, 2000, and many others)

$$f_l(\mathbf{x}) = \rho_{l-1} f_{l-1}(\mathbf{x}) + Z_{l-1}(\mathbf{x}), \quad l = 2, \ldots, L$$

where both $f_{l-1}(\mathbf{x})$ and $Z_{l-1}(\mathbf{x})$ have Gaussian Process (GP) priors.

Existing Methods

- Modeling:
  - Co-kriging (Kennedy and O'Hagan, 2000, and many others)

    $$f_l(\mathbf{x}) = \rho_{l-1} f_{l-1}(\mathbf{x}) + Z_{l-1}(\mathbf{x}), \quad l = 2, \ldots, L$$

    where both $f_{l-1}(\mathbf{x})$ and $Z_{l-1}(\mathbf{x})$ have Gaussian Process (GP) priors.
  - Non-stationary GP (Tuo, Wu and Yu, 2014): emulate $f_\infty(\mathbf{x})$ as $h_\infty \to 0$

Existing Methods

- Modeling:
  - Co-kriging (Kennedy and O'Hagan, 2000, and many others)

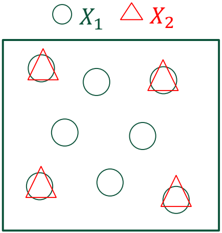
$$f_l(\mathbf{x}) = \rho_{l-1} f_{l-1}(\mathbf{x}) + Z_{l-1}(\mathbf{x}), \quad l = 2, \ldots, L$$

    where both $f_{l-1}(\mathbf{x})$ and $Z_{l-1}(\mathbf{x})$ have Gaussian Process (GP) priors.

  - Non-stationary GP (Tuo, Wu and Yu, 2014): emulate $f_\infty(\mathbf{x})$ as $h_\infty \to 0$

- Experimental Design: Nested space-filling design (Qian, Ai, and Wu, 2009, and many others)

$$X_L \subseteq X_{L-1} \subseteq \cdots \subseteq X_1$$



Questions

- Q1: Sample size of each level?

Questions

- Q1: Sample size of each level?
- Q2: How many fidelity levels?

Questions

- Q1: Sample size of each level?
- Q2: How many fidelity levels?
- Q3: Mesh size/density specification?

Questions

- Q1: Sample size of each level?
- Q2: How many fidelity levels?
- Q3: Mesh size/density specification?
- Q4: Is it better than single-fidelity simulation?

Multi-Level Interpolator

- Goal: Emulate $f_\infty(\mathbf{x})$

Multi-Level Interpolator

- Goal: Emulate $f_\infty(\mathbf{x})$
- Idea: with $f_0(\mathbf{x}) = 0$

$$f_L(\mathbf{x}) = (f_1(\mathbf{x}) - f_0(\mathbf{x})) + (f_2(\mathbf{x}) - f_1(\mathbf{x})) + \cdots + (f_L(\mathbf{x}) - f_{L-1}(\mathbf{x}))$$

- Assume the data is nested $X_L \subseteq X_{L-1} \subseteq \cdots \subseteq X_1$

Multi-Level Interpolator

- Denote the difference $Z(X_l) = f_l(X_l) - f_{l-1}(X_l)$ at nested sites X_l

Multi-Level Interpolator

- Denote the difference $Z(X_l) = f_l(X_l) - f_{l-1}(X_l)$ at nested sites X_l
- The reproducing kernel Hilbert space (RKHS) interpolator for each $(f_l(\mathbf{x}) - f_{l-1}(\mathbf{x}))$ is

$$\hat{Z}_l(\mathbf{x}) = \Phi_l(\mathbf{x}, X_l)\Phi_l(X_l, X_l)^{-1}Z(X_l),$$

where Φ_l is a positive definite kernel function.

Multi-Level Interpolator

- Denote the difference $Z(X_l) = f_l(X_l) - f_{l-1}(X_l)$ at nested sites X_l
- The reproducing kernel Hilbert space (RKHS) interpolator for each $(f_l(\mathbf{x}) - f_{l-1}(\mathbf{x}))$ is

$$\hat{Z}_l(\mathbf{x}) = \Phi_l(\mathbf{x}, X_l)\Phi_l(X_l, X_l)^{-1}Z(X_l),$$

  where Φ_l is a positive definite kernel function.

- ML Interpolator:

$$\hat{f}_L(\mathbf{x}) = \hat{Z}_1(\mathbf{x}) + \hat{Z}_2(\mathbf{x}) + \cdots + \hat{Z}_L(\mathbf{x}).$$

Matérn kernel

Assumption: Matérn kernel Φ

$$\Phi_l(\mathbf{x}, \mathbf{x}') = \phi_l(\|\theta_l \odot (\mathbf{x} - \mathbf{x}')\|_2)$$

with

$$\phi_l(r) = \frac{\sigma_l^2}{\Gamma(\nu_l)2^{\nu_l - 1}}(2\sqrt{\nu_l}r)_l^{\nu} B_{\nu_l}(2\sqrt{\nu_l}r),$$

- ν_l: smoothness parameter
- θ_l: lengthscale parameter
- σ_l^2: scalar parameter
- B_ν: the modified Bessel function of the second kind

Matérn kernel

Assumption: Matérn kernel Φ

$$\Phi_l(\mathbf{x}, \mathbf{x}') = \phi_l(\|\theta_l \odot (\mathbf{x} - \mathbf{x}')\|_2)$$

with

$$\phi_l(r) = \frac{\sigma_l^2}{\Gamma(\nu_l) 2^{\nu_l - 1}} (2\sqrt{\nu_l} r)_l^{\nu} B_{\nu_l}(2\sqrt{\nu_l} r),$$

- ν_l: smoothness parameter
- θ_l: lengthscale parameter
- σ_l^2: scalar parameter
- B_ν: the modified Bessel function of the second kind
- Parameters can be estimated via either CV or MLE (by a GP assumption)

Note of ML Interpolator

- Alternatively, one can assume $Z_l(\mathbf{x})$ follows a Gaussian process (GP) prior.

Note of ML Interpolator

- Alternatively, one can assume $Z_l(\mathbf{x})$ follows a Gaussian process (GP) prior.
- The posterior mean is equivalent to the ML Interpolator $\hat{f}_L(\mathbf{x})$.
- Can be viewed as a special case of Kennedy and O'Hagan (2000) model ($\rho_l = 1$)

Error Analysis of ML Interpolator

- ML Interpolator $\hat{f}_L(\mathbf{x}) = \hat{Z}_1(\mathbf{x}) + \hat{Z}_2(\mathbf{x}) + \cdots + \hat{Z}_L(\mathbf{x})$
- Recall our goal is to emulate $f_\infty(\mathbf{x})$

Error Analysis of ML Interpolator

- ML Interpolator $\hat{f}_L(\mathbf{x}) = \hat{Z}_1(\mathbf{x}) + \hat{Z}_2(\mathbf{x}) + \cdots + \hat{Z}_L(\mathbf{x})$
- Recall our goal is to emulate $f_\infty(\mathbf{x})$

$$|f_\infty(\mathbf{x}) - \hat{f}_L(\mathbf{x})| \leq \underbrace{|f_\infty(\mathbf{x}) - f_L(\mathbf{x})|}_{\text{simulation error}} + \underbrace{|f_L(\mathbf{x}) - \hat{f}_L(\mathbf{x})|}_{\text{emulation error}}.$$

Idea of Stacking Design

- Given a desired accuracy $\epsilon > 0$

Idea of Stacking Design

- Given a desired accuracy $\epsilon > 0$
- We wish $\|f_\infty - \hat{f}_L\| < \epsilon$ (i.e., with target predictive accuracy!)

Idea of Stacking Design

- Given a desired accuracy $\epsilon > 0$
- We wish $\|f_\infty - \hat{f}_L\| < \epsilon$ (i.e., with target predictive accuracy!)

$\|f_\infty - f_L\| < \frac{\epsilon}{2}$ → determine L

$\|f_L - \hat{f}_L\| < \frac{\epsilon}{2}$ → determine sample sizes n_l

→ Stacking Design

Control emulation error $\|f_L - \hat{f}_L\|$

Proposition 1: Emulation error

Suppose that

- the input space is d-dimensional and is bounded and convex,
- X_l is **quasi-uniform** with sample size n_l,

Then,

$$|f_L(\mathbf{x}) - \hat{f}_L(\mathbf{x})| \leq c \sum_{l=1}^{L} \|\theta_l\|_2^{\nu_l} n_l^{-\nu_l/d} \|f_l - f_{l-1}\|_{\mathcal{N}_{\Phi_l}(\Omega)},$$

where $\| \cdot |||_{\mathcal{N}_{\Phi_l}(\Omega)}$ is the RKHS norm.

Control emulation error $\|f_L - \hat{f}_L\|$

Proposition 1: Emulation error

Suppose that

- the input space is d-dimensional and is bounded and convex,
- X_l is **quasi-uniform** with sample size n_l,

Then,

$$|f_L(\mathbf{x}) - \hat{f}_L(\mathbf{x})| \leq c \sum_{l=1}^{L} \|\theta_l\|_2^{\nu_l} n_l^{-\nu_l/d} \|f_l - f_{l-1}\|_{\mathcal{N}_{\Phi_l}(\Omega)},$$

where $\|\cdot\||\|_{\mathcal{N}_{\Phi_l}(\Omega)}$ is the RKHS norm.

- Denote $q_X = \min_{1\leq j\neq k\leq n} \|\mathbf{x}_j - \mathbf{x}_k\|/2$ and $h_{X,\Omega}$ as the fill distance.
- A design $\mathbf{X}_n$ satisfying $h_{X,\Omega}/q_X \leq C$ for some constant C is called a **quasi-uniform** design.

Sample size determination n_l

- Sample size n_l can be determined by minimizing the error bound and the total cost by the method of Lagrange multipliers

$$\sum_{l=1}^{L} \|\theta_l\|_2^{\nu_{\min}} n_l^{-\nu_{\min}/d} \|f_l - f_{l-1}\|_{\mathcal{N}_{\Phi_l}(\Omega)} + \lambda \sum_{l=1}^{L} n_l C_l,$$

where $\nu_{\min} = \min_{l=1,\ldots,L} \nu_l$, which gives

Sample size determination n_l

- Sample size n_l can be determined by minimizing the error bound and the total cost by the method of Lagrange multipliers

$$\sum_{l=1}^{L} \|\theta_l\|_2^{\nu_{\min}} n_l{}^{-\nu_{\min}/d} \|f_l - f_{l-1}\|_{\mathcal{N}_{\Phi_l}(\Omega)} + \lambda \sum_{l=1}^{L} n_l C_l,$$

where $\nu_{\min} = \min_{l=1,\ldots,L} \nu_l$, which gives

$$n_l = \mu \left(\frac{\|\theta_l\|^{\nu_{\min}}}{C_l} \|f_l - f_{l-1}\|_{\mathcal{N}_{\Phi_l}(\Omega)} \right)^{d/(\nu_{\min}+d)}$$

for some constant $\mu > 0$.

Sample size determination n_l

- Sample size n_l can be determined by minimizing the error bound and the total cost by the method of Lagrange multipliers

$$\sum_{l=1}^{L} \|\theta_l\|_2^{\nu_{\min}} n_l^{-\nu_{\min}/d} \|f_l - f_{l-1}\|_{\mathcal{N}_{\Phi_l}(\Omega)} + \lambda \sum_{l=1}^{L} n_l C_l,$$

where $\nu_{\min} = \min_{l=1,\ldots,L} \nu_l$, which gives

$$n_l = \mu \left(\frac{\|\theta_l\|^{\nu_{\min}}}{C_l} \|f_l - f_{l-1}\|_{\mathcal{N}_{\Phi_l}(\Omega)} \right)^{d/(\nu_{\min}+d)}$$

for some constant $\mu > 0$.

- Find μ such that $\|f_L - \hat{f}_L\| < \epsilon/2$

Sample size determination n_l

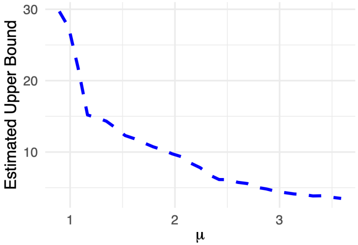
$$\|f_L - \hat{f}_L\| < \sum_{l=1}^{L} \|P_l\| \|f_l - f_{l-1}\|_{\mathcal{N}_{\Phi_l(\Omega)}} < \epsilon/2$$

- $P_l(\mathbf{x})$ is a *power function*
- $\|f_l - f_{l-1}\|_{\mathcal{N}_{\Phi_l(\Omega)}}$ can be estimated by $\|\hat{Z}_l\|_{\mathcal{N}_{\Phi_l(\Omega)}}$

Sample size determination n_l

$$\|f_L - \hat{f}_L\| < \sum_{l=1}^{L} \|P_l\| \|f_l - f_{l-1}\|_{\mathcal{N}_{\Phi_l(\Omega)}} < \epsilon/2$$

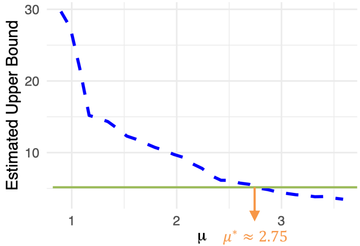
- $P_l(\mathbf{x})$ is a *power function*
- $\|f_l - f_{l-1}\|_{\mathcal{N}_{\Phi_l(\Omega)}}$ can be estimated by $\|\hat{Z}_l\|_{\mathcal{N}_{\Phi_l(\Omega)}}$



Sample size determination n_l

$$\|f_L - \hat{f}_L\| < \sum_{l=1}^{L} \|P_l\| \|f_l - f_{l-1}\|_{\mathcal{N}_{\Phi_l}(\Omega)} < \epsilon/2$$

- $P_l(\mathbf{x})$ is a *power function*
- $\|f_l - f_{l-1}\|_{\mathcal{N}_{\Phi_l}(\Omega)}$ can be estimated by $\|\hat{Z}_l\|_{\mathcal{N}_{\Phi_l}(\Omega)}$



Questions

- Q1: Sample size of each level? n_l
- Q2: How many fidelity levels?
- Q3: Mesh size/density specification?
- Q4: Is it better than single-fidelity simulation?

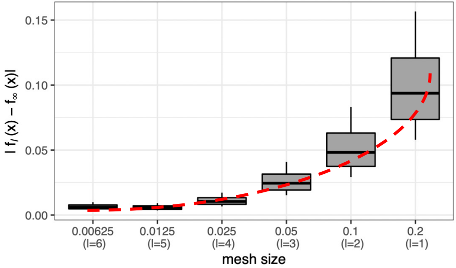
Control simulation error $\|f_\infty - f_L\|$

Error Rate of finite element simulations

(Brenner and Scott, 2007, Tuo, Wu and Yu, 2014) Under some regularity conditions, for a constant $\alpha \in \mathbb{N}$,

$$|f_\infty(\mathbf{x}) - f_L(\mathbf{x})| < c(\mathbf{x})h_L^\alpha.$$

Recall h_L is the mesh size.



Determine the fidelity level L

- Let $h_l = h_0 2^{-l}$ where $h_0/2$ is the mesh size of the lowest fidelity simulator $f_1(\mathbf{x})$.

Determine the fidelity level L

- Let $h_l = h_0 2^{-l}$ where $h_0/2$ is the mesh size of the lowest fidelity simulator $f_1(\mathbf{x})$.
- Suppose $|f_\infty(\mathbf{x}) - f_L(\mathbf{x})| = c_1(\mathbf{x}) h_L^\alpha + O(h_L^{\alpha+1})$

Determine the fidelity level L

- Let $h_l = h_0 2^{-l}$ where $h_0/2$ is the mesh size of the lowest fidelity simulator $f_1(\mathbf{x})$.
- Suppose $|f_\infty(\mathbf{x}) - f_L(\mathbf{x})| = c_1(\mathbf{x}) h_L^\alpha + O(h_L^{\alpha+1})$
- One can show that

$$\|f_\infty - f_L\| = \frac{\|f_L - f_{L-1}\|}{2^\alpha - 1},$$

assuming that the terms of order $h_L^{\alpha+1}$ and higher can be neglected.

- $\|f_L - f_{L-1}\|$ can be approximated by $\|\hat{Z}_L\|$.

Determine the fidelity level L

- Let $h_l = h_0 2^{-l}$ where $h_0/2$ is the mesh size of the lowest fidelity simulator $f_1(\mathbf{x})$.
- Suppose $|f_\infty(\mathbf{x}) - f_L(\mathbf{x})| = c_1(\mathbf{x}) h_L^\alpha + O(h_L^{\alpha+1})$
- One can show that

$$\|f_\infty - f_L\| = \frac{\|f_L - f_{L-1}\|}{2^\alpha - 1},$$

  assuming that the terms of order $h_L^{\alpha+1}$ and higher can be neglected.
- $\|f_L - f_{L-1}\|$ can be approximated by $\|\hat{Z}_L\|$.
- Find L that ensures $\frac{\|\hat{Z}_L\|}{2^\alpha - 1} \leq \epsilon/2$

Determination of α

- Tuo, Wu and Yu (2014) determines α according to the quantity of interest

Determination of α

- Tuo, Wu and Yu (2014) determines α according to the quantity of interest
- Alternatively, it can be determined by collected data (can be done only when $L \geq 3$) (details omitted)

$$\hat{\alpha} = \frac{1}{L-2} \sum_{l=3}^{L} \sum_{\mathbf{x} \in X_l} \frac{\log\left(\left|\frac{f_{l-1}(\mathbf{x}) - f_{l-2}(\mathbf{x})}{f_l(\mathbf{x}) - f_{l-1}(\mathbf{x})}\right|\right)}{n_l \log 2}.$$

Questions

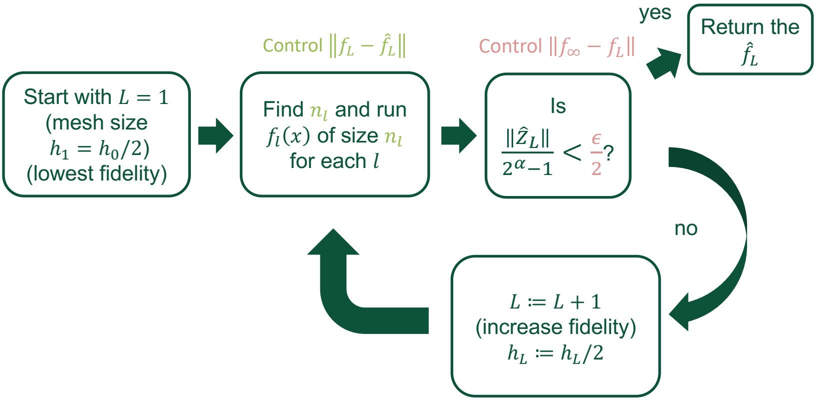
- Q1: Sample size of each level? n_l
- Q2: How many fidelity levels? L
- Q3: Mesh size/density specification? $h_l = h_0 2^{-l}$
- Q4: Is it better than single-fidelity simulation?

Stacking design with error upper bound ϵ

- **Idea:** Start with low-fidelity simulations and sequentially increase the fidelity level until $\frac{\|\hat{Z}_L\|}{2^\alpha - 1} \leq \epsilon/2$.

Stacking design with error upper bound ϵ

- **Idea:** Start with low-fidelity simulations and sequentially increase the fidelity level until $\frac{\|\hat{Z}_L\|}{2^{\alpha}-1} \leq \epsilon/2$.

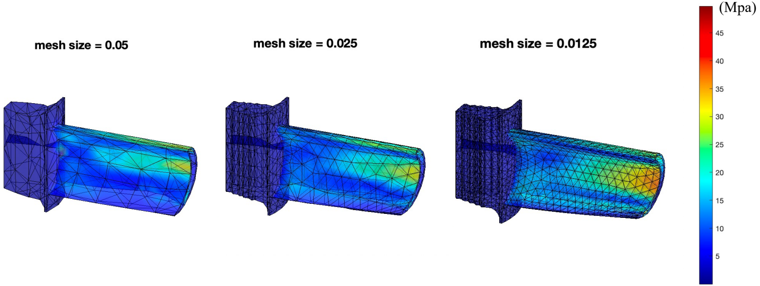


Prediction Uncertainty

- An approximated pointwise error interval of $f_\infty(\mathbf{x})$ can be constructed as

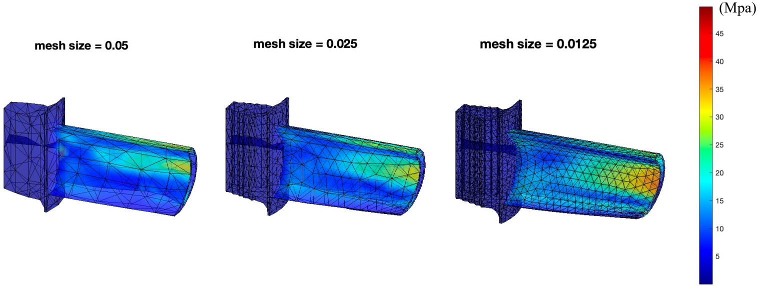
$$\hat{f}_L(\mathbf{x}) \pm \left(\frac{|\hat{Z}_L(\mathbf{x})|}{2^\alpha - 1} + \sum_{l=1}^{L} P_l(\mathbf{x})(Z_l(X_l)^T \Phi_l(X_l, X_l)^{-1} Z_l(X_l))^{1/2}\right).$$

Revisit motivated example



- **Input**: $\mathbf{x} = (x_1, x_2) = (\text{pressure}, \text{suction})$
- **Output**: $f(\mathbf{x})$: average of thermal stress

Revisit motivated example



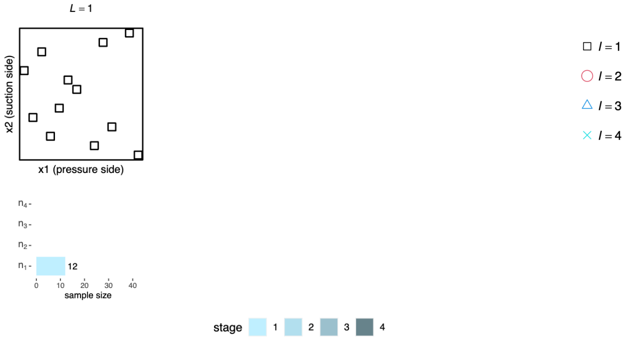
- **Input**: $\mathbf{x} = (x_1, x_2) = (\text{pressure}, \text{suction})$
- **Output**: $f(\mathbf{x})$: average of thermal stress
- **Test data**: Simulations with mesh size $h \approx 0$ at 20 uniform test input locations are conducted to examine the performance

Revisit motivated example

- We wish $\|f_\infty - \hat{f}_L\|_{L_2(\Omega)} < \epsilon = 5$

Revisit motivated example

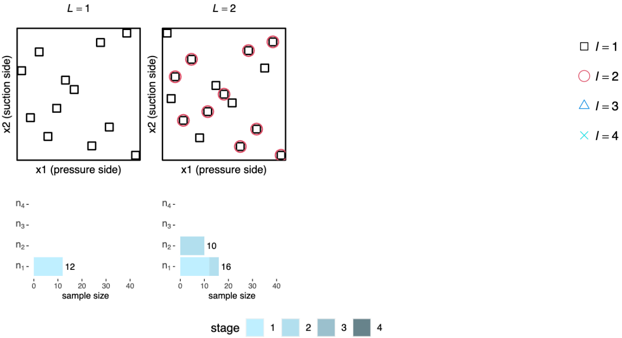
- We wish $\|f_\infty - \hat{f}_L\|_{L_2(\Omega)} < \epsilon = 5$



	$L=1$
h_L	0.05
C_L (sec.)	0.75
$\|f_L - \hat{f}_L\|_{L_2(\Omega)}$	**2.324**
$\|f_\infty - f_L\|_{L_2(\Omega)}$	NA

Revisit motivated example

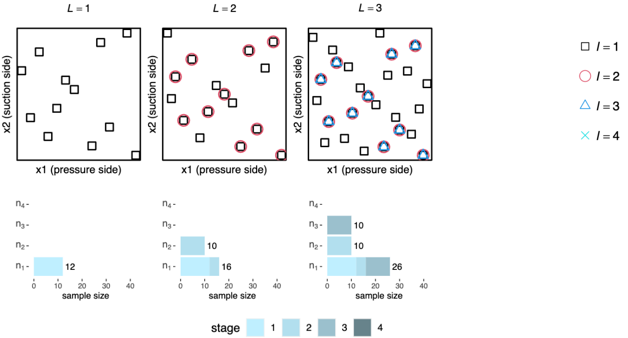
- We wish $\|f_\infty - \hat{f}_L\|_{L_2(\Omega)} < \epsilon = 5$



	$L=1$	$L=2$
h_L	0.05	0.025
C_L (sec.)	0.75	1.07
$\|f_L - \hat{f}_L\|_{L_2(\Omega)}$	2.324	2.408
$\|f_\infty - f_L\|_{L_2(\Omega)}$	NA	NA

Revisit motivated example

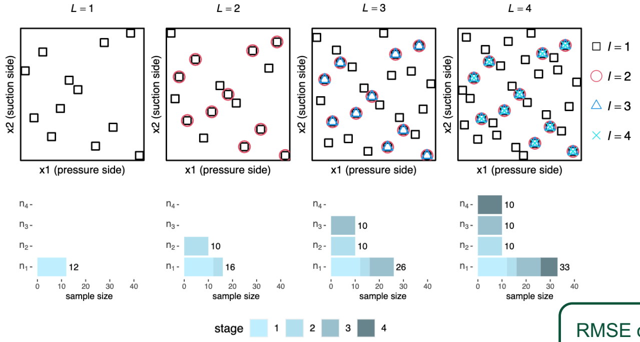
- We wish $\|f_\infty - \hat{f}_L\|_{L_2(\Omega)} < \epsilon = 5$



	$L=1$	$L=2$	$L=3$
h_L	0.05	0.025	0.0125
C_L (sec.)	0.75	1.07	2.13
$\|f_L - \hat{f}_L\|_{L_2(\Omega)}$	**2.324**	**2.408**	**2.481**
$\|f_\infty - f_L\|_{L_2(\Omega)}$	NA	NA	2.969

Revisit motivated example

- We wish $\|f_\infty - \hat{f}_L\|_{L_2(\Omega)} < \epsilon = 5$



	$L=1$	$L=2$	$L=3$	$L=4$	
h_L	0.05	0.025	0.0125	0.00625	
C_L (sec.)	0.75	1.07	2.13	11.51	
$\|f_L - \hat{f}_L\|_{L_2(\Omega)}$	**2.324**	**2.408**	**2.481**	**2.491**	$< \frac{\epsilon}{2}$
$\|f_\infty - f_L\|_{L_2(\Omega)}$	NA	NA	2.969	**0.956**	$< \frac{\epsilon}{2}$

RMSE of $\hat{f}_4$ $= 1.60$

Visualize $\hat{f}_L(\mathbf{x})$

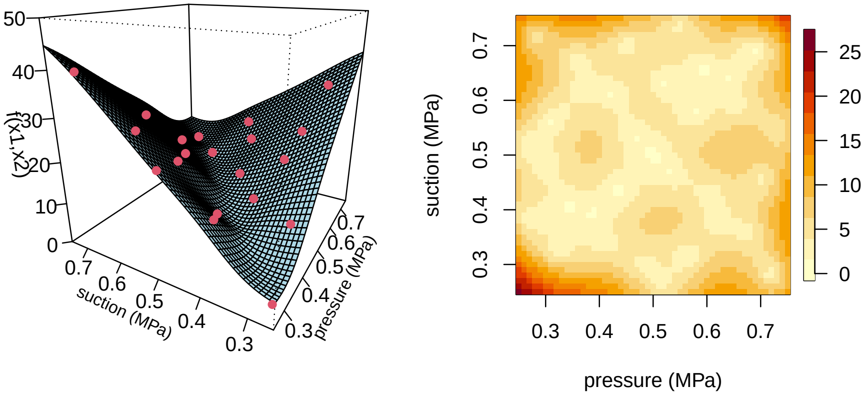


Figure: *(left)* $\hat{f}_4(\mathbf{x})$ *and true test points (red dots); (right) pointwise error bounds*

Cost complexity theorem

Theorem

Suppose that

- $\nu := \nu_1 = \cdots = \nu_L$
- $|f_\infty(\mathbf{x}) - f_l(\mathbf{x})| < c_1 2^{-\alpha l}$
- $C_l < c_2 2^{\beta l}$

Cost complexity theorem

Theorem

Suppose that

- $\nu := \nu_1 = \cdots = \nu_L$
- $|f_\infty(\mathbf{x}) - f_l(\mathbf{x})| < c_1 2^{-\alpha l}$
- $C_l < c_2 2^{\beta l}$

Under some regularity conditions, it follows that

$$|f_\infty(\mathbf{x}) - \hat{f}_L(\mathbf{x})| < \epsilon,$$

with a **total computational cost** C_{tot} bounded by

$$C_{\text{tot}} \leq \begin{cases} c_3 \epsilon^{-\frac{d}{\nu}}, & \frac{\alpha}{\beta} > \frac{2\nu}{d}, \\ c_3 \epsilon^{-\frac{d}{\nu}} |\log \epsilon|^{1+\frac{d}{\nu}}, & \frac{\alpha}{\beta} = \frac{2\nu}{d}, \\ c_3 \epsilon^{-\frac{d}{\nu} - \frac{2\beta\nu - \alpha d}{2\alpha(\nu+d)}}, & \frac{\alpha}{\beta} < \frac{2\nu}{d}. \end{cases}$$

Insight on budget allocation

- When $\frac{\alpha}{\beta} > \frac{2\nu}{d}$, the budget C_{tot} is expended on coarser mesh

Insight on budget allocation

- When $\frac{\alpha}{\beta} > \frac{2\nu}{d}$, the budget C_{tot} is expended on coarser mesh
- When $\frac{\alpha}{\beta} < \frac{2\nu}{d}$, the budget C_{tot} is expended on denser mesh

Insight on budget allocation

- When $\frac{\alpha}{\beta} > \frac{2\nu}{d}$, the budget C_{tot} is expended on coarser mesh
- When $\frac{\alpha}{\beta} < \frac{2\nu}{d}$, the budget C_{tot} is expended on denser mesh
- When $\frac{\alpha}{\beta} = \frac{2\nu}{d}$, evenly spread

Insight on budget allocation

- When $\frac{\alpha}{\beta} > \frac{2\nu}{d}$, the budget C_{tot} is expended on coarser mesh
- When $\frac{\alpha}{\beta} < \frac{2\nu}{d}$, the budget C_{tot} is expended on denser mesh
- When $\frac{\alpha}{\beta} = \frac{2\nu}{d}$, evenly spread

$\frac{\alpha}{\beta}$	$\frac{2\nu}{d}$
simulation error reduction over the rate computational cost as fidelity increases	the rate of convergence of RKHS interpolator as sample size increases

Single-fidelity vs Multi-fidelity

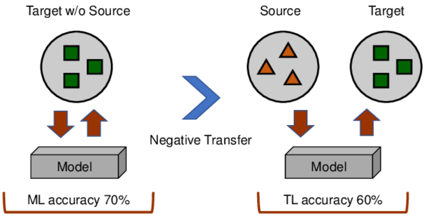
- What if all the budget is expanded on single fidelity simulations?

Single-fidelity vs Multi-fidelity

- What if all the budget is expanded on single fidelity simulations?
- Independent kriging vs co-kriging?

Single-fidelity vs Multi-fidelity

- What if all the budget is expanded on single fidelity simulations?
- Independent kriging vs co-kriging?
- Negative transfer?



Zhang et al. (2021) A Survey on Negative Transfer. *IEEE Transactions on Neural Networks and Learning Systems*

Complexity of single-fidelity interpolator

Corollary

- Let $\hat{f}_H(x)$ be the RKHS interpolator based on single-fidelity data $(X_H, f_H(X_H))$

Complexity of single-fidelity interpolator

Corollary

- Let $\hat{f}_H(x)$ be the RKHS interpolator based on single-fidelity data $(X_H, f_H(X_H))$
- Suppose that $(\epsilon/2)^{1+\frac{\alpha d}{2\nu\beta}} \leq c_1 h_H^{\alpha} \leq \epsilon/2$, where $c_1 = \sup_{\mathbf{x}\in\Omega} c_1(\mathbf{x})$

Complexity of single-fidelity interpolator

Corollary

- Let $\hat{f}_H(x)$ be the RKHS interpolator based on single-fidelity data $(X_H, f_H(X_H))$
- Suppose that $(\epsilon/2)^{1+\frac{\alpha d}{2\nu\beta}} \leq c_1 h_H^{\alpha} \leq \epsilon/2$, where $c_1 = \sup_{\mathbf{x}\in\Omega} c_1(\mathbf{x})$

Under some regularity conditions, it follows that

$$|f_\infty(\mathbf{x}) - \hat{f}_H(\mathbf{x})| < \epsilon,$$

with a **total computational cost** C_H bounded by

$$C_H \leq c_h \epsilon^{-\frac{\beta}{\alpha}-\frac{d}{2\nu}}.$$

Single-fidelity vs Multi-fidelity

- When $\frac{\alpha}{\beta} < \frac{2\nu}{d} \Rightarrow$ ML interpolator has a slower cost rate

Single-fidelity vs Multi-fidelity

- When $\frac{\alpha}{\beta} < \frac{2\nu}{d} \Rightarrow$ ML interpolator has a slower cost rate
- When $\frac{\alpha}{\beta} \geq \frac{2\nu}{d} \Rightarrow$ Single-fidelity RKHS interpolator has a slower cost rate

Single-fidelity vs Multi-fidelity

- When $\frac{\alpha}{\beta} < \frac{2\nu}{d} \Rightarrow$ ML interpolator has a slower cost rate
- When $\frac{\alpha}{\beta} \geq \frac{2\nu}{d} \Rightarrow$ Single-fidelity RKHS interpolator has a slower cost rate
- Example 1: β is small and α is large

Single-fidelity vs Multi-fidelity

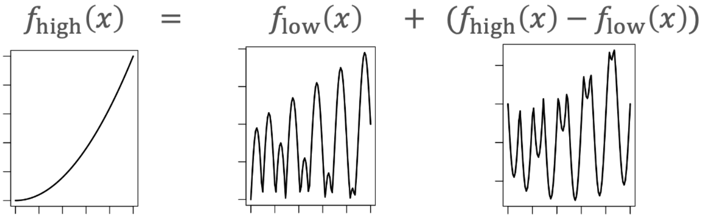
- When $\frac{\alpha}{\beta} < \frac{2\nu}{d} \Rightarrow$ ML interpolator has a slower cost rate
- When $\frac{\alpha}{\beta} \geq \frac{2\nu}{d} \Rightarrow$ Single-fidelity RKHS interpolator has a slower cost rate
- Example 1: β is small and α is large
  - $C_1 = 2.9$ and $C_5 = 3$
  - $|f_\infty(\mathbf{x}) - f_1(\mathbf{x})| = 10$, and $|f_\infty(\mathbf{x}) - f_5(\mathbf{x})| = 0.001$

Single-fidelity vs Multi-fidelity

- When $\frac{\alpha}{\beta} < \frac{2\nu}{d} \Rightarrow$ ML interpolator has a slower cost rate
- When $\frac{\alpha}{\beta} \geq \frac{2\nu}{d} \Rightarrow$ Single-fidelity RKHS interpolator has a slower cost rate
- Example 1: β is small and α is large
  - $C_1 = 2.9$ and $C_5 = 3$
  - $|f_\infty(\mathbf{x}) - f_1(\mathbf{x})| = 10$, and $|f_\infty(\mathbf{x}) - f_5(\mathbf{x})| = 0.001$
- Example 2: ν is very small (i.e., nonsmooth $(f_l - f_{l-1})$)

Single-fidelity vs Multi-fidelity

- When $\frac{\alpha}{\beta} < \frac{2\nu}{d} \Rightarrow$ ML interpolator has a slower cost rate
- When $\frac{\alpha}{\beta} \geq \frac{2\nu}{d} \Rightarrow$ Single-fidelity RKHS interpolator has a slower cost rate
- Example 1: β is small and α is large
  - $C_1 = 2.9$ and $C_5 = 3$
  - $|f_\infty(\mathbf{x}) - f_1(\mathbf{x})| = 10$, and $|f_\infty(\mathbf{x}) - f_5(\mathbf{x})| = 0.001$
- Example 2: ν is very small (i.e., nonsmooth $(f_l - f_{l-1})$)



Questions

- Q1: Sample size of each level? n_l
- Q2: How many fidelity levels? L
- Q3: Mesh size/density specification? $h_l = h_0 2^{-l}$
- Q4: Is it better than single-fidelity simulation? In some cases, yes

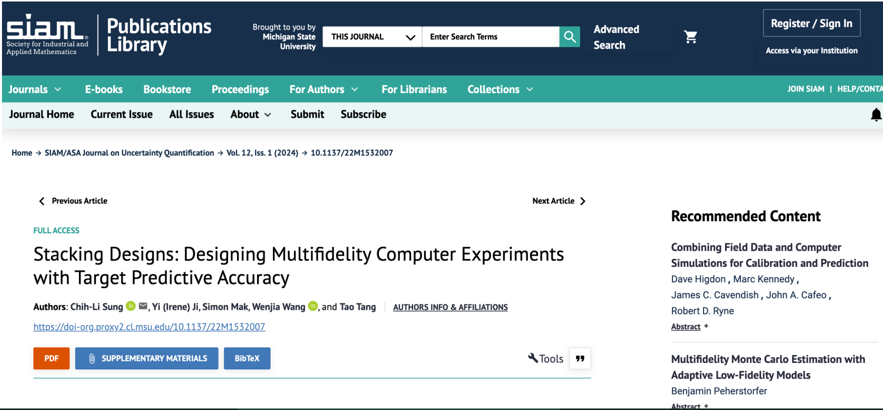
Conclusion

- Stacking design for multi-fidelity simulations with desired accuracy
  - Sample determination
  - Mesh size determination

- Cost complexity
  - Budget allocation
  - Comparison with single fidelity simulation

Reference

- **Sung, C.-L.**, Ji, Y., Mak, S., Wang, W., & Tang, T. (2024). Stacking Designs: Designing Multifidelity Computer Experiments with Target Predictive Accuracy. *SIAM/ASA Journal on Uncertainty Quantification*, 12(1), 157-181.



Thank You!

Thank NSF DMS 2113407 for supporting this work.

MICHIGAN STATE
UNIVERSITY