Model management strategy for hierarchical Kriging

Motivation

Many-query analysis

- Requires repeated runs of simulations
- Becomes computationally intractable when using expensive high fidelity simulations

Can we introduce cheaper low fidelity data into model training?

“How” should we allocate the samples across fidelities?

Vision

Explore budget allocation strategy for hieararchical Kriging

Building blocks

1. **Hierarchical Kriging** ⟵ **Lack of budget allocation strategy**
2. **Multifidelity Monte Carlo (MFMC)** ⟵ **Budget allocation strategy for multifidelity data**

Goal

Apply MFMC budget allocation for hierarchical Kriging

Contents

Gaussian process: Notation

- $z_i \in \mathbb{R}^d$: high fidelity input
- $z^* \in \mathbb{R}^d$: test input
- $f^{(1)}: \mathbb{R}^d \to \mathbb{R}$: high fidelity model
- $k: \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$: kernel covariance function

Gaussian process: Assumptions

Models input-output relationship by assuming a Gaussian process prior

$$f^{(1)} \sim \mathcal{GP}(0, k(\cdot,\cdot))$$

$y_i^{(1)} \in \mathbb{R}$: high fidelity observation

ϵ_i: noise

$\sigma_e^2 \in \mathbb{R}$: noise variance

$$y_i^{(1)} = f^{(1)}(z_i) + \epsilon_i, \qquad \epsilon_i \sim \mathcal{N}(0, \sigma_e^2 I)$$

Gaussian process: Definition

Given training data $\mathcal{D} = \{\boldsymbol{z}, \boldsymbol{y}^{(1)}\}$, GP posterior distribution

$$f^{(1)}(z^*|\mathcal{D}) \sim \mathcal{N}\left(\mathbb{E}\left[f^{(1)}(z^*|\mathcal{D})\right], \mathbb{V}\mathrm{ar}\left[f^{(1)}(z^*|\mathcal{D})\right]\right)$$

$$\mathbb{E}\left[f^{(1)}(z^*|\mathcal{D})\right] = k(\boldsymbol{z}, z^*; \theta)^\top \left(k(\boldsymbol{z}, \boldsymbol{z}; \theta) + \sigma_e^2 I\right)^{-1} \boldsymbol{y}^{(1)} = \hat{f}^{(1)}(z^*)$$

Squared exponential kernel

$$k(z, z'; \theta) = \theta_1 \exp\left(-\frac{\|z - z'\|_2^2}{2\theta_2^2}\right)$$

θ_1, θ_2: kernel hyperparameters

How to find θ, σ_e^2?

Gaussian process: Training

Find θ, σ_e^2 that maximize log-likelihood by gradient-based optimization methods

$$\max_{\theta,\sigma_e^2} \log p(y^{(1)}|\theta, \sigma_e^2)$$

$$= -\frac{1}{2}\left[{y^{(1)}}^{\top} (k(z,z;\theta) + \sigma_e^2 I)^{-1} y^{(1)} + \log|k(z,z;\theta) + \sigma_e^2 I| + n \log 2\pi \right]$$

summary

$$\mathbb{E}\left[f^{(1)}(z^*|\mathcal{D})\right] = k(\boldsymbol{z}, z^*;\theta)^{\top}(k(\boldsymbol{z},\boldsymbol{z};\theta) + \sigma_e^2 I)^{-1}\boldsymbol{y}^{(1)} = \hat{f}^{(1)}(z^*)$$

Hierarchical Kriging: Notation

- $z_i^{(2)} \in \mathbb{R}^d$: low fidelity input
- $f^{(2)}: \mathbb{R}^d \to \mathbb{R}$: low fidelity model
- $y_i^{(2)} \in \mathbb{R}$: low fidelity observation

$\boldsymbol{y}^{(1)} \in \mathbb{R}^n$: high fidelity output vector

$\boldsymbol{y}^{(2)} \in \mathbb{R}^m$: low fidelity output vector

$$n < m$$

Hierarchical Kriging: Definition

- Based on **Kennedy O'Hagan** approach

$$f^{(1)}(z^*) = \alpha f^{(2)}(z^*) + \delta(z^*)$$

$\alpha \in \mathbb{R}$: scaling factor

- Assumes $f^{(2)}$ is a low fidelity GP model
- $\hat{\delta}$ is a discrepancy GP model
- **Predictor (posterior mean)**

$$\hat{f}^{(1)}(z^*) = \alpha \underbrace{\hat{f}^{(2)}(z^*)}_{\text{Posterior mean of } f^{(2)}} + \underbrace{\hat{\delta}(z^*)}_{\text{Posterior mean of } \delta^{(2)}}$$

Hierarchical Kriging: Definition

- **Predictor (posterior mean)**

Trained with low fidelity data $\boldsymbol{y}^{(2)} \in \mathbb{R}^m$

$$\hat{f}^{(1)}(z^*) = \alpha \underline{\hat{f}^{(2)}(z^*)} + \underline{\hat{\delta}(z^*)}$$

Trained with discrepancy data

$$\boldsymbol{y}_d = \boldsymbol{y}^{(1)} - \alpha \hat{f}^{(2)}(\boldsymbol{z}),$$

$$\boldsymbol{y}^{(1)} \in \mathbb{R}^n, \boldsymbol{y}_d \in \mathbb{R}^n$$

summary

$$\alpha\left(k(\boldsymbol{z}, z^*)^\top (k(\boldsymbol{z}, \boldsymbol{z}; \theta) + \sigma_e^2 I)^{-1} \boldsymbol{y}^{(2)}\right) + k(\boldsymbol{z}, z^*; \theta)^\top (k(\boldsymbol{z}, \boldsymbol{z}; \theta) + \sigma_e^2 I)^{-1} \boldsymbol{y}_d$$

How to find θ, σ_e^2 for δ?

Gaussian process: Training

Find θ, σ_e^2 that maximize log-likelihood by gradient-based optimization methods

$$\max_{\theta,\sigma_e^2} \log p(y_d|\theta,\sigma_e^2)$$

$$= -\frac{1}{2}\left[y_d^\top (k(z,z;\theta)+\sigma_e^2 I)^{-1} y_d + \log|k(z,z;\theta)+\sigma_e^2 I| + n\log 2\pi\right]$$

$$\boldsymbol{y}_d = \boldsymbol{y}^{(1)} - \alpha \hat{f}^{(2)}(\boldsymbol{z})$$

How to find α?

Hierarchical Kriging: Scaling factor

- By solving generalized least squares

$$\left\|\boldsymbol{y}^{(1)} - \alpha f^{(2)}(\boldsymbol{z})\right\|^2_{k(\boldsymbol{z},\boldsymbol{z})+\sigma_e^2 I}$$

$$\alpha = \left(\hat{f}^{(2)}(z)^\top (k(z,z) + \sigma_e^2 I)^{-1}\,\hat{f}^{(2)}(\boldsymbol{z})\right)^{-1} (\hat{f}^{(2)}(z)^\top (k(z,z) + \sigma_e^2 I)^{-1} y^{(1)})$$

Predictor

$$\hat{f}^{(1)}(z^*) = \alpha\left(k(\boldsymbol{z}, z^*)^\top (k(\boldsymbol{z},\boldsymbol{z};\theta) + \sigma_e^2 I)^{-1} \boldsymbol{y}^{(2)}\right) + k(\boldsymbol{z}, z^*;\theta)^\top (k(\boldsymbol{z},\boldsymbol{z}) + \sigma_e^2 I)^{-1} \boldsymbol{y}_d$$

$$\boldsymbol{y}^{(2)} \in \mathbb{R}^m, \boldsymbol{y}_d \in \mathbb{R}^n$$

How to allocate n, m?

MFMC budget allocation: MFMC estimator

More robust way to estimate mean

- Monte Carlo estimator

$$\mathbb{E}[f^{(1)}(Z)] \approx \bar{y}^{(1)} = \frac{1}{n}\sum_{i=1}^{n} y_i^{(1)}$$

- MFMC estimator adds low fidelity data

Assumes nested samples $z \subset z^{(2)}$

$$\mathbb{E}[f^{(1)}(Z)] \approx \frac{1}{n}\sum_{i=1}^{n} y_i^{(1)} + \alpha\left(\frac{1}{m}\sum_{i=1}^{m} y_i^{(2)} - \frac{1}{n}\sum_{i=1}^{n} y_i^{(2)}\right)$$

How to allocate n, m?

MFMC budget allocation: Variance

Obtain optimal n and m that **minimizes variance of the estimator**

$$\frac{\sigma_1^2}{n} + \left(\frac{1}{n} - \frac{1}{m}\right)\left(\alpha^2\sigma_2^2 - 2\alpha\rho_{1,2}\sigma_1\sigma_2\right)$$

- σ_1: standard deviation of high fidelity data
- σ_2: standard deviation of low fidelity data
- $\rho_{1,2}$: correlation coefficient of high and low fidelity data

MFMC budget allocation

c: computational budget

w_1: high fidelity model evaluation cost

w_2: low fidelity model evaluation cost

τ: ratio of number of low fidelity samples to high fidelity samples

$$n = \frac{c}{w_1 + w_2 \tau}, \qquad m = \tau n, \qquad \tau = \sqrt{\frac{w_1 \rho_{1,2}^2}{w_2 (1 - \rho_{1,2}^2)}}$$

Code structure

- Class MFMC
- Class Kriging
- Class DiscKriging
- Class MFKriging

$$f^{(1)}(z^*) = \alpha f^{(2)}(z^*) + \delta(z^*)$$

Code structure

Class MFMC

- def stats: Compute statistics $\sigma_1, \sigma_2, \rho_{1,2}$
- def alloc: Compute sample allocations n, m

$$n = \frac{c}{w_1 + w_2 \tau}, \qquad m = \tau n, \qquad \tau = \sqrt{\frac{w_1 \rho_{1,2}^2}{w_2(1-\rho_{1,2}^2)}}$$

Code structure

Class Kriging, DiscKriging

- Def neg_loglikeli: Evaluate negative log likelihood

$$\max_{\theta,\sigma_e^2} \log p(y^{(1)}|\theta,\sigma_e^2)$$

$$= -\frac{1}{2}\left[{y^{(1)}}^{\top}(k(z,z;\theta)+\sigma_e^2 I)^{-1}y^{(1)} + \log|k(z,z;\theta)+\sigma_e^2 I| + n\log 2\pi\right]$$

Cholesky decomposition

$$(k(\mathbf{z},\mathbf{z};\theta)+\sigma_e^2 I) = LL^{\top}$$

- Def train: Train single fidelity Kriging model

- Def predict: Compute posterior mean and variance at unseen points

Code structure

Class MFKriging

- Def train: Train low fidelity and discrepancy Kriging models
- Def predict: Compute posterior mean and variance at unseen points

Ishigami function example: Set up

- Input: $\boldsymbol{x} = (x_1, x_2, x_3)$, $x_i \sim \mathcal{U}(-\pi, \pi)$
- High fidelity model

$$a = 5, b = 0.1$$

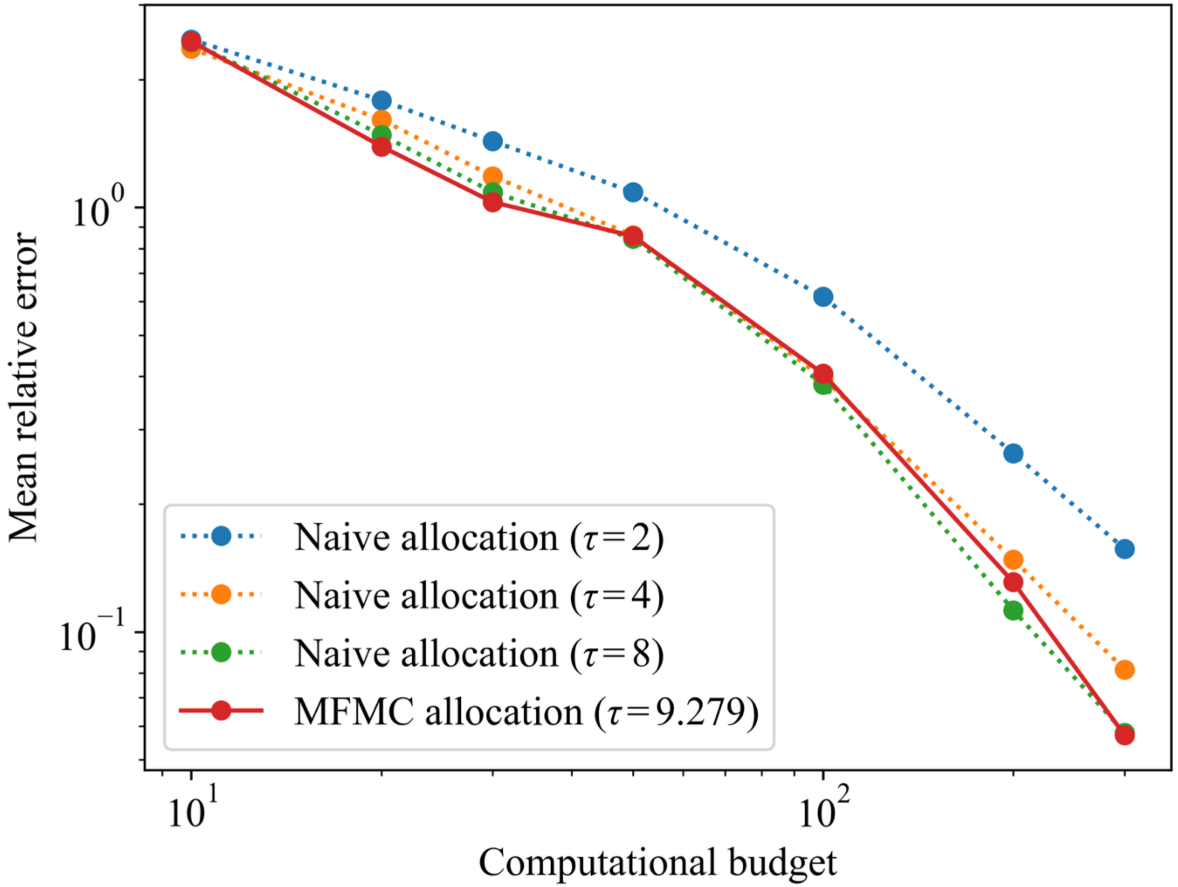
$$f^{(1)}(\boldsymbol{x}) = \sin x_1 + a \sin^2 x_2 \ + \ b x_3^4 \sin x_1 \ ,$$

- Low fidelity model

$$f^{(2)}(\boldsymbol{x}) = \sin x_1 + 0.6 a \sin^2 x_2 \ + 9 b x_3^2 \sin x_1 \ ,$$

- Cost = [1, 0.1]
- Statistics: $\sigma_1 = 3.29$,$\sigma_2 = 3.53, \rho_{1,2} = 0.9465$

Ishigami function example: Results



$$\tau = \frac{m}{n}$$

At computational budget 100

Allocation	τ	n	m
	2	83	166
Naive	4	71	285
	8	55	444
MFMC	9.279	51	481

Error comparable with τ =8

Wing structural analysis problem: Set up

- Input: 4 wing geometry parameters

  Wing span, dihedral, twist, sweep angles

- Output: maximum von Mises stress

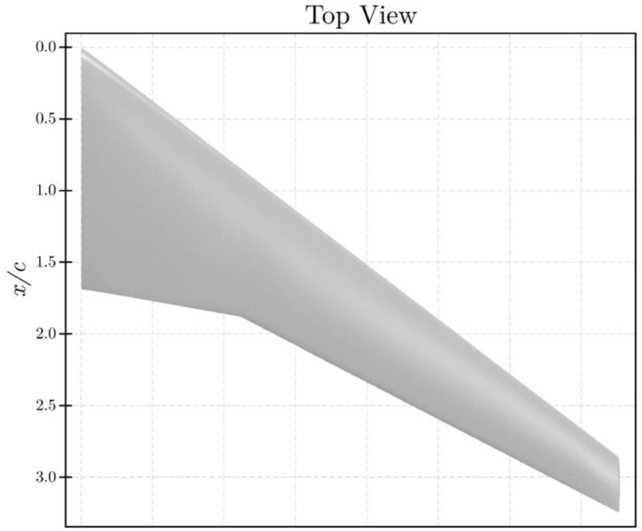
- High fidelity: higher rib count

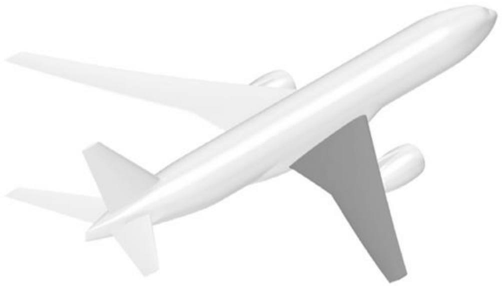
- Low fidelity: lower rib count

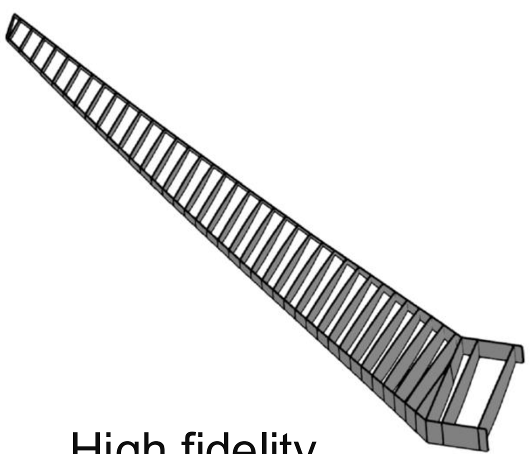
- Cost = [5.4, 4.7] CPU s

- Statistics:

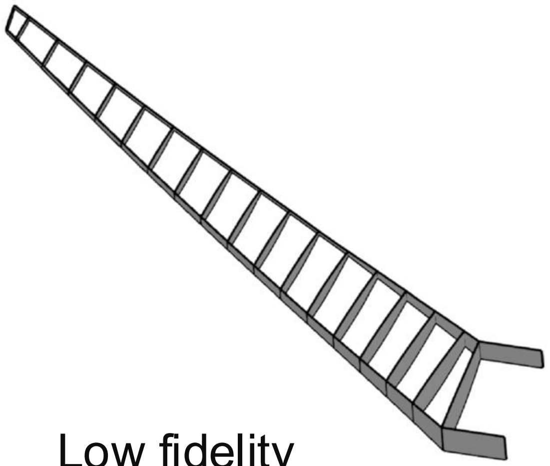
  $\sigma_1 = 9131.61, \sigma_2 = 8838.04, \rho_{1,2} = 0.9732$







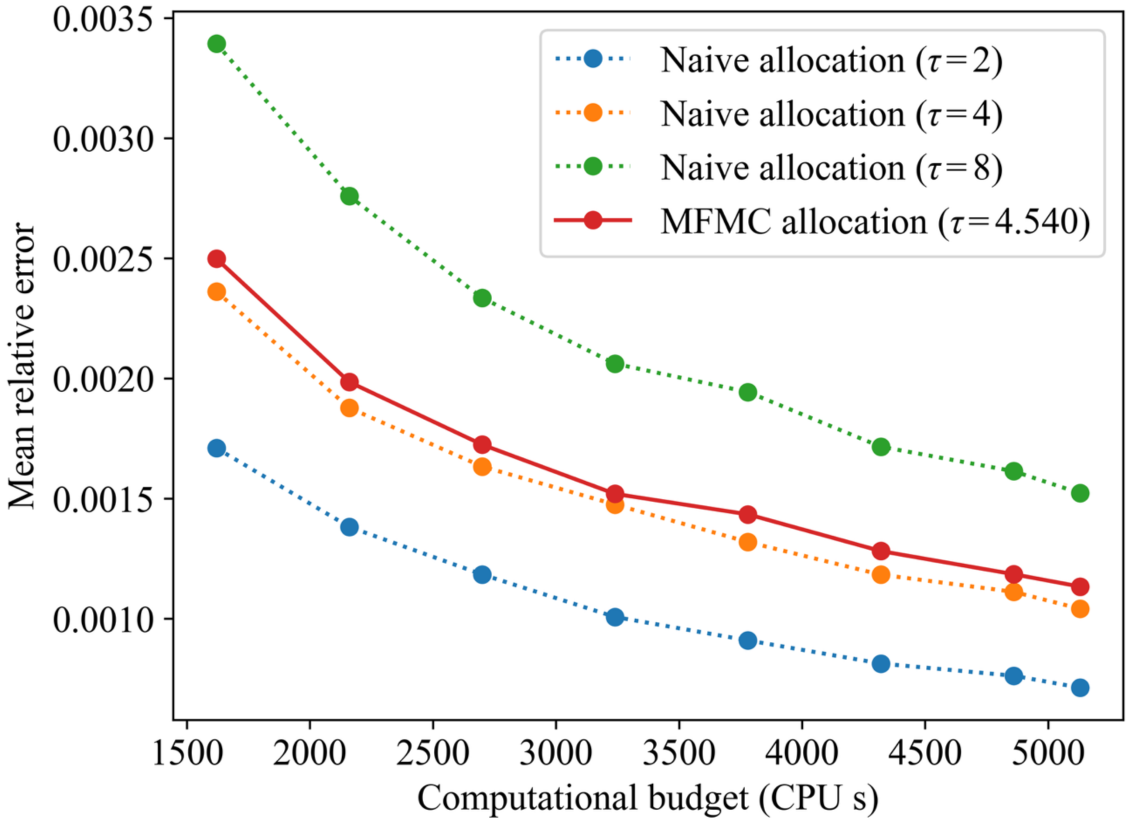
High fidelity



Low fidelity

Source: Perron, C., Rajaram, D., & Mavris, D. N. (2021). Multi-fidelity non-intrusive reduced-order modelling based on manifold alignment. Proceedings of the Royal Society A, 477(2253), 20210495.

Wing structural analysis problem: Results



$$\tau = \frac{m}{n}$$

At computational budget 1,620 CPU s

Allocation	τ	n	m
	2	109	218
Naive	4	66	267
	8	37	301
MFMC	4.54	60	275

Error comparable with $\tau = 4$

Summary and conclusion

- Proposed MFMC budget allocation strategy for hierarchical Kriging
- Hierarchical Kriging with MFMC allocation achieves comparable accuracy
- MFMC allocation functions as a practical guideline for sample allocation