# Learning context-aware surrogate models for multifidelity importance sampling and Bayesian inverse problems

NYU | COURANT

Terrence Alsup, Benjamin Peherstorfer
Courant Institute of Mathematical Sciences

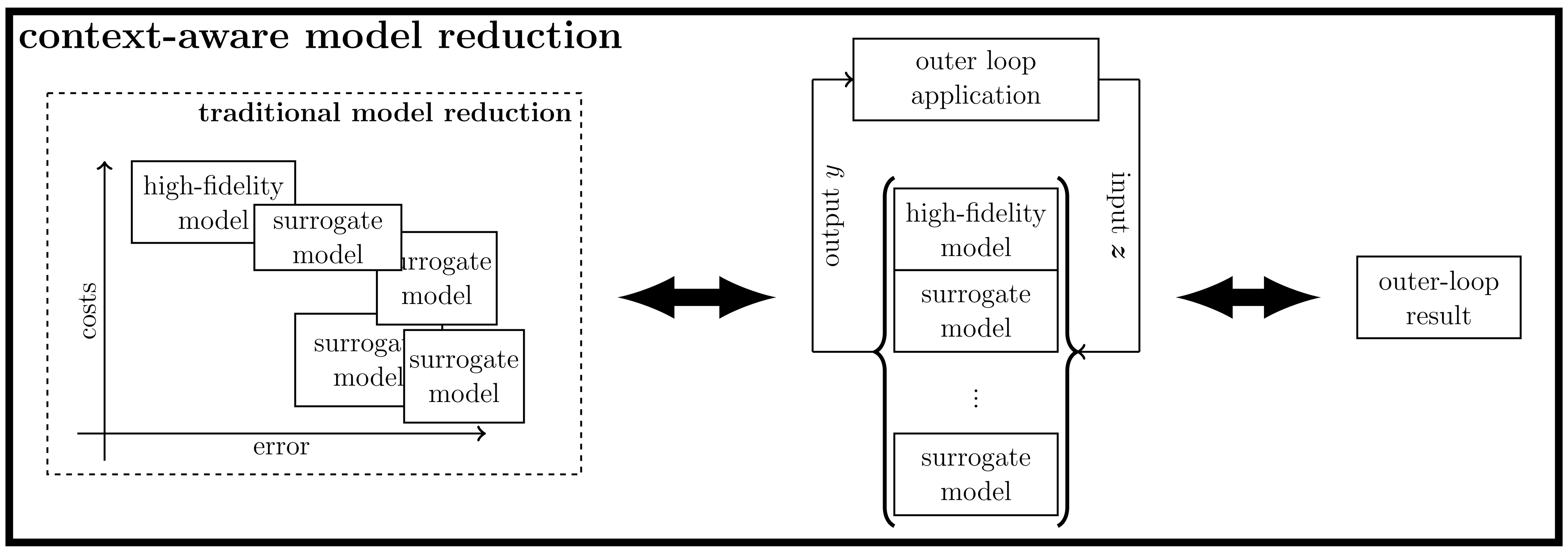
## Learning context-aware surrogate models for multifidelity computations

**Traditional model reduction is separate from multifidelity computations**

- Measures error w.r.t. high-fidelity output while outer-loop result is goal.
- Ignores that surrogates are used together with other information sources.
- Approximating high-fidelity model hard, but supporting it may be easy.

**Our approach: Context-aware model reduction**

- Learn w.r.t. outer-loop result, rather than high-fidelity model output.
- Learn surrogate explicitly for multifidelity computations.
- Optimally trade off surrogate model accuracy and high-fidelity recourse.




## Context-aware model reduction for Bayesian inverse problems

**Posterior of Bayesian inverse problem**

- For Bayesian inverse problems, posterior takes the form

$$p_L(\boldsymbol{\theta}|\boldsymbol{y}) \propto \exp\left(-\frac{1}{2}\|\boldsymbol{y} - F^{(L)}(\boldsymbol{\theta})\|^2_{\Gamma^{-1}}\right)\pi(\boldsymbol{\theta}),$$

  with parameter $\boldsymbol{\theta}$, high-fidelity parameter-to-observation map $F^{(L)}$, observed data $\boldsymbol{y}$, noise covariance matrix $\Gamma$, and Gaussian prior $\pi$.

- Estimating expectations $\mathbb{E}_{p_L}[f(\boldsymbol{\theta})]$ of test functions $f$ challenging if $F^{(L)}$ is computationally expensive to evaluate.

**Multifidelity importance sampling (MFIS)** [P., Cui, Marzouk, Willcox, 2015]

- Fidelity of surrogate models determined by $\ell$ with $F^{(\ell)} \to F^{(L)}$ as $\ell \to \infty$.
- Use surrogate models $F^{(\ell)}$ and posteriors $p_\ell$ to build biasing density $q_\ell$ for importance sampling.
- MFIS estimator

$$\hat{f}^{\text{MFIS}}_{\ell,N} = \frac{\sum_{i=1}^N f(\boldsymbol{\theta}_i)\frac{p_L(\boldsymbol{\theta}_i)}{q_\ell(\boldsymbol{\theta}_i)}}{\sum_{i=1}^N \frac{p_L(\boldsymbol{\theta}_i)}{q_\ell(\boldsymbol{\theta}_i)}}, \quad \{\boldsymbol{\theta}_i\}_{i=1}^N \overset{\text{i.i.d.}}{\sim} q_\ell(\boldsymbol{\theta}).$$

## 1. Trade-off: Quality of biasing density vs. number of high-fidelity samples

**Laplace approximation as a biasing density**

Laplace approximation $q_\ell$ of the posterior $p_\ell$

$$q_\ell \sim N\left(\boldsymbol{\theta}^*_\ell, \boldsymbol{H}^{-1}_\ell(\boldsymbol{\theta}^*_\ell)\right)$$

where $\boldsymbol{\theta}^*_\ell$ is MAP point of $p_\ell$ and $\boldsymbol{H}^{-1}_\ell(\boldsymbol{\theta}^*_\ell)$ is inverse Hessian of $-\log p_\ell$ at $\boldsymbol{\theta}^*_\ell$.

**Biasing density should be close in the $\chi^2$-divergence** [Agapious et. al, 2017]

- $\chi^2$-divergence

$$\chi^2(p_L||q_\ell) = \text{Var}_{q_\ell}\left[\frac{p_L}{q_\ell}\right] = \int \left(\frac{p_L(\boldsymbol{x})}{q_\ell(\boldsymbol{x})} - 1\right)^2 q_\ell(\boldsymbol{x})d\boldsymbol{x}$$

- If $p_L \ll q_\ell$, $\mathbb{E}_{q_\ell}[(p_L/q_\ell)^2] < \infty$, and $|f| \leq 1$, then

$$\text{MSE}(\hat{f}^{\text{MFIS}}_{\ell,N}) = \mathbb{E}\left[|\hat{f}^{\text{MFIS}}_{\ell,N} - \mathbb{E}_{p_L}[f]|^2\right] \leq \frac{4}{N}\left(\chi^2(p_L||q_\ell) + 1\right).$$

- If $\chi^2(p_L||q_\ell) \approx 0$, then $q_\ell$ is "good" biasing density and fewer samples $N$ are needed.

## 2. Optimal level for context-aware surrogate model

**Fidelity of surrogate controls closeness of biasing density in $\chi^2$-divergence**

If $\|F^{(L)}(\boldsymbol{x}) - F^{(\ell)}(\boldsymbol{x})\| \leq \delta(\ell)\|\boldsymbol{x}\|^2$ and $q_\ell$ is Laplace approximation of $p_\ell$, then

$$\chi^2(p_L||q_\ell) + 1 \leq K_0 \exp\left(K_1\delta(\ell)\right),$$

with $K_0, K_1$ constants independent of $\ell$.

**Find level $\ell$ and number of samples $N$ that minimize cost and achieve MSE within tolerance $\epsilon$**

$$\underset{N,\ell\geq 0}{\text{minimize}} \quad \underbrace{c(L)N}_{\text{online cost}} + \underbrace{Mc(\ell)}_{\text{offline cost}}$$

$$\text{subject to} \quad \text{MSE}(\hat{f}^{\text{MFIS}}_{\ell,N}) \leq \frac{K_0}{N}\exp\left(K_1\delta(\ell)\right) \leq \epsilon$$

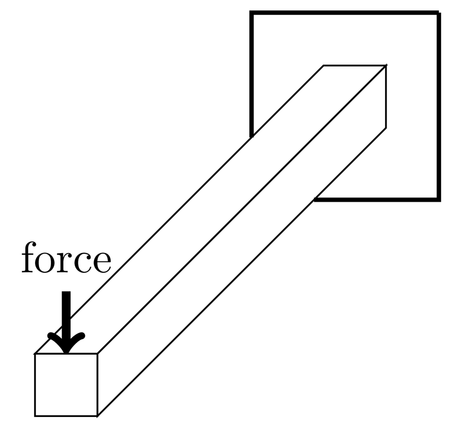
- Costs $c(\ell), c(L)$ of surrogate and high-fidelity models, respectively, fixed number of offline evaluations $M$ to construct biasing density.
- Unique solution $(\ell^*, N^*)$ exists if $c''(\ell), \delta''(\ell) \geq 0$.
- Optimal context-aware importance sampling estimator is $\hat{f}^{\text{MFIS}}_{\ell^*,N^*}$.

## Numerical example: Inferring effective stiffness of Euler-Bernoulli beam

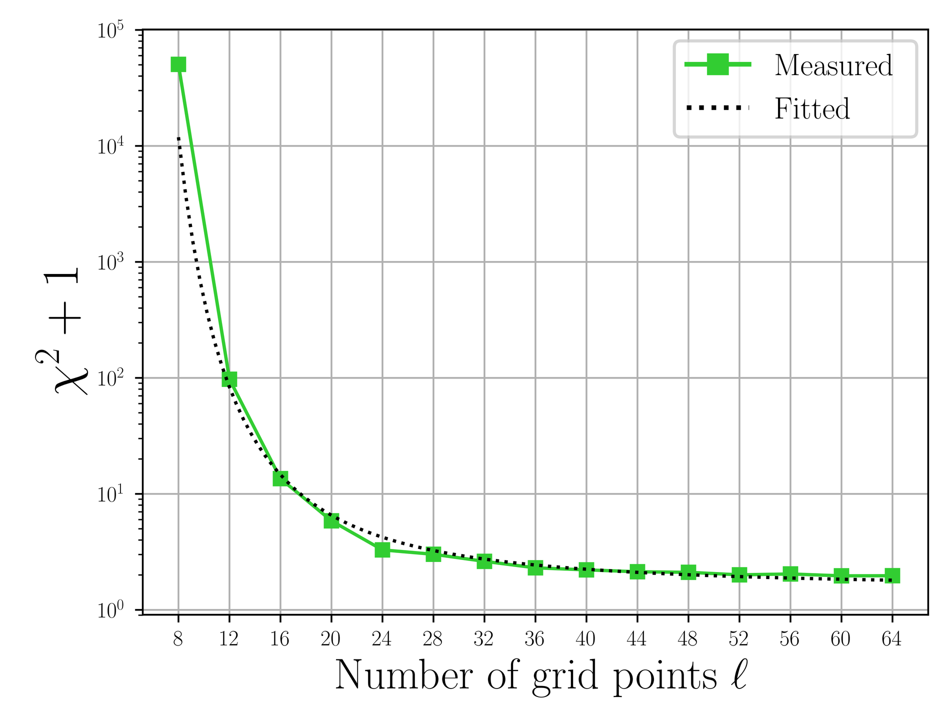
- Euler-Bernoulli problem governed by

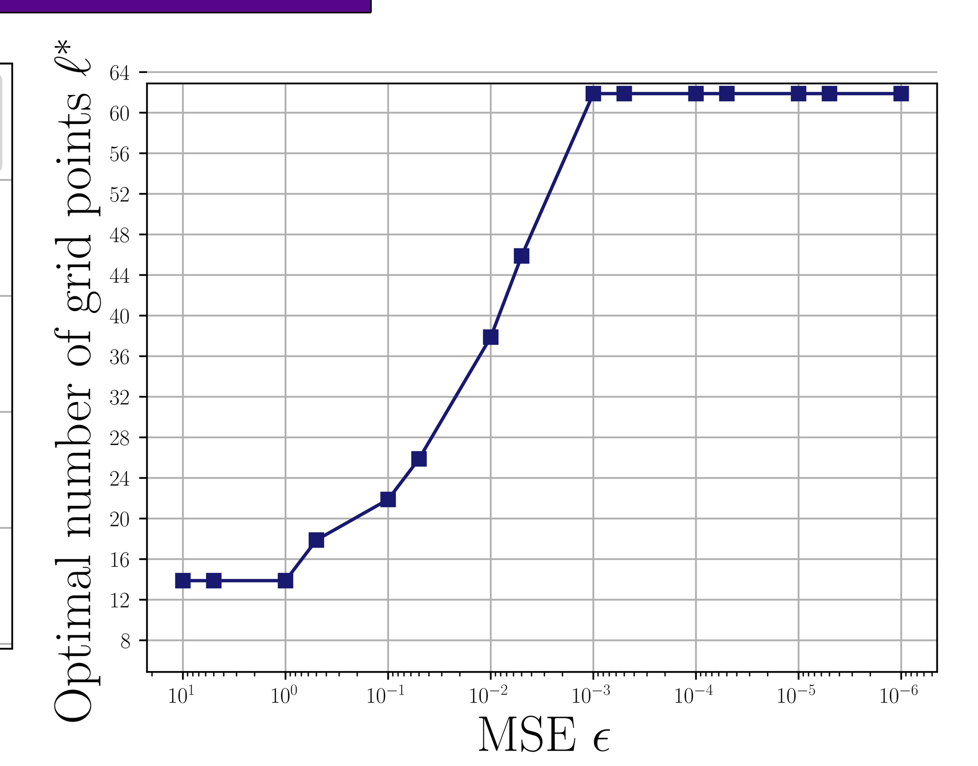
$$\frac{\partial^2}{\partial x^2}\left(E(x;\boldsymbol{\theta})\frac{\partial^2}{\partial x^2}u(x)\right) = 1, \quad x \in [0,1]$$

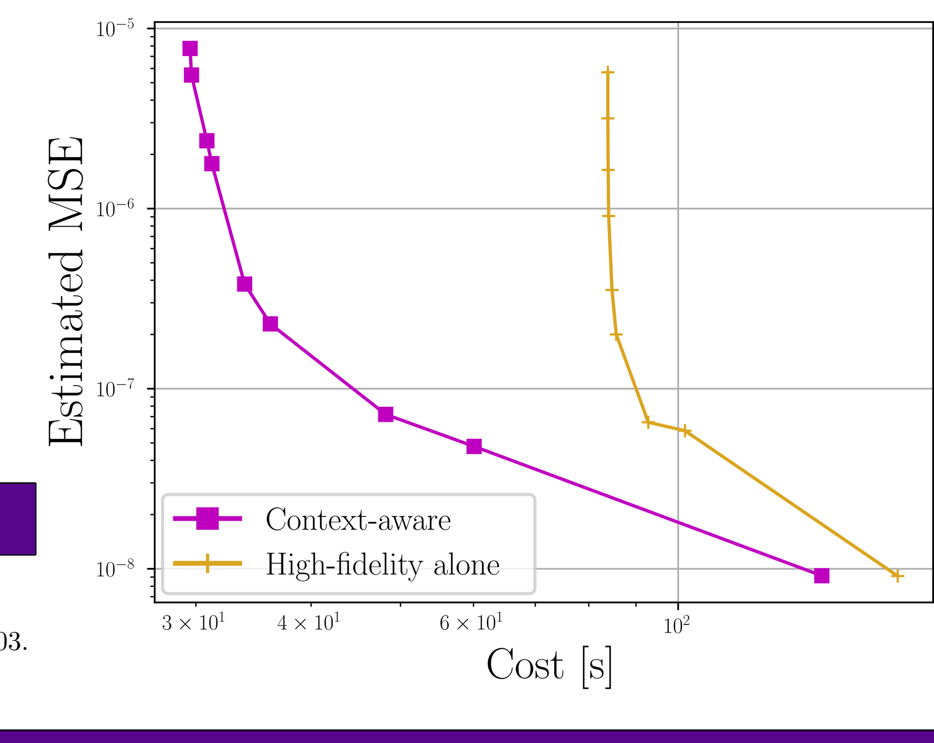
  with boundary conditions $u(0) = u_x(0) = u_{xx}(1) = u_{xxx}(1) = 0$.




- Parameter dependence through effective stiffness $E(x;\boldsymbol{\theta})$ with $\boldsymbol{\theta} \in \mathbb{R}^6$.
- High-fidelity parameters-to-observation map: $F^{(L)}: \mathbb{R}^6 \to \mathbb{R}^{10}$. Compute the finite difference solution using $L = 256$ points and then interpolate at 10 linearly spaced points over $[0,1]$.
- Surrogates $F^{(\ell)}$ use $\ell = 8, 12, \ldots, 64$ grid points.
- Costs are $c(\ell) \propto \ell$ and finite difference accuracy is $\delta(\ell) \propto \ell^{-2}$.
- Noise covariance $\Gamma = \text{diag}(10^{-4}, \ldots, 10^{-4})$ corresponds to $\sim 3\%$ error.
- Gaussian prior with mean $\boldsymbol{\mu} = (1, \ldots, 1)^T$, covariance $\boldsymbol{\Sigma} = \text{diag}(10^{-5}, \ldots, 10^{-5})$.




**Top-left** Estimated $\chi^2(p_L||q_\ell)$ shows close fit to theoretical curve $K_0\exp(K_1\delta(\ell))$.

**Top-right** Optimal $\ell^*$ shows cheap surrogates are sufficient for outer-loop application.

**Bottom** Multifidelity importance sampling with context-aware surrogate model achieves 3× speedup compared to using high-fidelity model alone.

## References


- S. Agapious, O. Papaspiliopoulos, D. Sanz-Alonso, and A. M. Stuart, *Importance Sampling: Intrinsic Dimension and Computational Cost.* (2017).
- B. Peherstorfer, *Multifidelity Monte Carlo estimation with adaptive low-fidelity models.* SIAM/ASA Journal on Uncertainty Quantification, 7 (2019), pp. 579-603.
- A. Stuart, *Inverse problems: A Bayesian perspective.* Acta Numerica, 19 (2010), pp. 451-559.


Terrence Alsup (alsup@cims.nyu.edu), https://cims.nyu.edu/~alsup/