Hyperkriging: Multi-Fidelity Features for High-Fidelity Kriging Models

Atticus Rex Elizabeth Qian

Kriging/Gaussian Processes

Given:

- A kernel $\boldsymbol{\kappa}(\mathbf{x}, \mathbf{x}'|\theta)$ with parameters $\theta \in \mathbb{R}^p$
- Input data $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$
- Output data $\mathbf{y} = [y_1, \ldots, y_N]^\top$
- Test inputs $\mathbf{X}_* = \{\mathbf{x}_1^*, \ldots, \mathbf{x}_M^*\}$

We can form kernel matrices:

$$\mathbf{K} = \boldsymbol{\kappa}(\mathbf{X}, \mathbf{X}|\theta), \qquad \mathbf{K}_* = \boldsymbol{\kappa}(\mathbf{X}^*, \mathbf{X}|\theta),$$
$$\mathbf{K}_{**} = \boldsymbol{\kappa}(\mathbf{X}^*, \mathbf{X}^*|\theta)$$

which we use to obtain the normally distributed posterior of $y(\mathbf{X}_*)$:

$$\mathbb{E}\left[y(\mathbf{X}_*)|\mathbf{y};\theta\right] = \mathbf{K}_*(\mathbf{K} + \sigma^2\mathbf{I})^{-1}\mathbf{y}$$
$$\mathbb{V}\left[y(\mathbf{X}_*)|\mathbf{y};\theta\right] = \mathbf{K}_{**} - \mathbf{K}_*(\mathbf{K} + \sigma^2\mathbf{I})^{-1}\mathbf{K}_*^\top$$

Optimal parameters and noise variance maximize the log-marginal likelihood, $\log p(\mathbf{y}|\theta, \sigma)$:

$$-\frac{1}{2}\left[\mathbf{y}^\top(\mathbf{K} + \sigma^2\mathbf{I})^{-1}\mathbf{y} + \log\left|\mathbf{K} + \sigma^2\mathbf{I}\right| + N\log(2\pi)\right]$$

The offline training cost is $\mathcal{O}(N^3)$.

State of the Art Multi-Fidelity Modeling

Cokriging: (Myers, 1982; Goulard & Voltz, 1992; Karniadakis, 2016)

- Use multi-output GPs with multi-fidelity data
- Require the inversion of large kernel matrices
- Sensitive to kernel parameter selection
- Only linear mappings between fidelities

Autoregressive Estimators: (Kennedy & O'Hagan, 2000; Perdikaris et al. 2016; Cutajar et al. 2017)

- Map each level of fidelity to the next
- Require smaller kernel matrix inversion
- Induce limiting Markovian property
- Assume known accuracy hierarchy of fidelities
- May require noiseless and/or nested training data

New Hyperkriging Method

Research Question: *How can we leverage data from a set of K low-fidelity models to improve the accuracy and computational cost of training a high-fidelity surrogate model?*

Key Definitions

- **High-Fidelity model**: highly accurate and expensive simulation
- **Low-Fidelity model**: less accurate but cheaper simulation
- **Multi-Fidelity surrogate model**: a data-driven model trained on both scarce high-fidelity data and plentiful low-fidelity data.

Main idea: create a set of multi-fidelity features for the high-fidelity kriging model:

$$\phi_1(\mathbf{x}) = \left[\mathbf{x}^\top \; h_K(\mathbf{x}) \ldots \; h_2(\mathbf{x})\right]^\top$$

where h_ℓ is a data-driven surrogate model for fidelity-ℓ. Then we choose a kernel $\boldsymbol{\kappa}$ which acts on ϕ_1:

$$\text{Cov}\left[y_1(\mathbf{x}), y_1(\mathbf{x}')\right] = \boldsymbol{\kappa}(\phi_1(\mathbf{x}), \phi_1(\mathbf{x}')|\theta_1)$$

This defines a high-fidelity Kriging model. The features ϕ_1 are created recursively from surrogate models trained on $\phi_2, \ldots, \phi_K$:

$$\phi_\ell(\mathbf{x}) = \begin{bmatrix}\phi_{\ell+1}(\mathbf{x}) \\ h_{\ell+1}(\mathbf{x})\end{bmatrix}, \quad \phi_K(\mathbf{x}) = \mathbf{x} \quad \text{(base case)}$$

Hence, each model h_ℓ uses information from levels $\ell+1$ through K to make predictions about level ℓ. Further, only h_1 needs to be a Kriging model, which can alleviate computational cost to train h_2 through h_K.

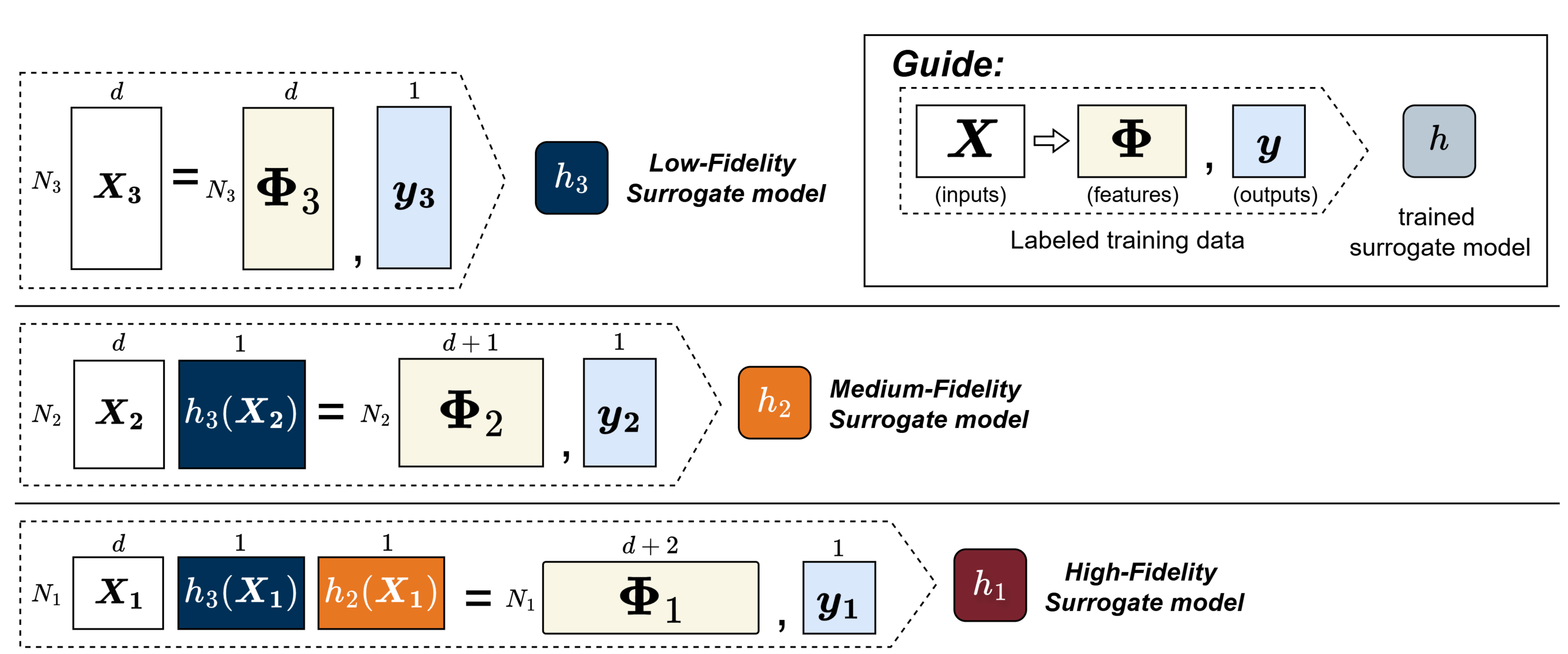

Figure 1. Hyperkriging offline training process on three levels of fidelity. The surrogate models are trained in the following order: h_3, h_2, then h_1.

Numerical Experiment

To clearly illustrate the utility of the method, we consider a simple one-dimensional test problem in which approximating the high-fidelity function requires a nonlinear combination of all low-fidelity functions:

Function	# of Data Points	$\mathbf{R}^2$
$f_1(\mathbf{x}) = \sin(2\pi\mathbf{x})\exp(-\mathbf{x})$	10	1.000
$f_2(\mathbf{x}) = \sin(2\pi\mathbf{x})$	100	0.638
$f_3(\mathbf{x}) = \exp(-\mathbf{x})$	250	0.417

Table 1. Experiment details. R^2 indicates the Pearson correlation coefficient with the high-fidelity function.

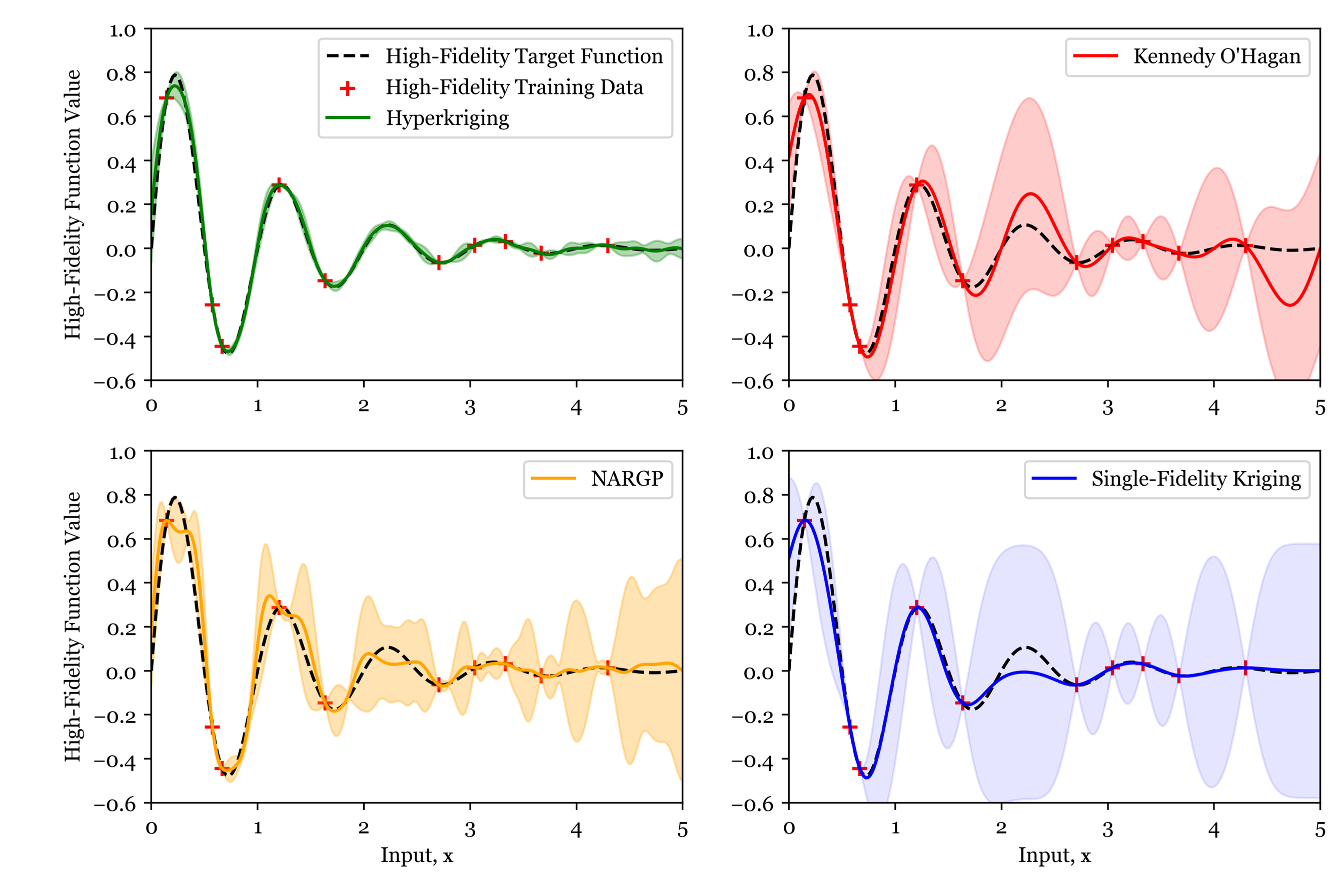

Figure 2. Hyperkriging predictions compared with KOH (Kennedy & O'Hagan, 2000), NARGP (Perdikaris et al. 2016), and single-fidelity Kriging (Rasmussen & Williams et al. 2008). Shaded regions represent 95% confidence intervals for the true function.

Method	RMSE	$\mathbf{R}^2$	MLL
Hyperkriging	2.294e-02	0.9866	10.6412
Kennedy O'Hagan	9.150e-02	0.8820	2.8686
NARGP	6.042e-02	0.9391	3.4994
Kriging	6.989e-02	0.8800	-3.4304

Table 2. Performance comparison model predictions at 250 linearly spaced test points across the input space.

Acknowledgements

AR was supported in part by the National Science Foundation under Grant No. DGE-2039655. EQ was supported in parts by the US Department of Energy Office of Science under DE-SC0024721 and by the National Science Foundation under EDSE 2442140. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National ScienceFoundation.