Minimax Approach to Variable Fidelity Data Interpolation

A.Zaytsev
Skoltech, IITP RAS

E.Burnaev
Skoltech, IITP RAS, HSE

Abstract

Engineering problems often involve data sources of variable fidelity with different costs of obtaining an observation. In particular, one can use both a cheap low fidelity function (e.g. a computational experiment with a CFD code) and an expensive high fidelity function (e.g. a wind tunnel experiment) to generate a data sample in order to construct a regression model of a high fidelity function. The key question in this setting is how the sizes of the high and low fidelity data samples should be selected in order to stay within a given computational budget and maximize accuracy of the regression model prior to committing resources on data acquisition.

In this paper we obtain minimax interpolation errors for single and variable fidelity scenarios for a multivariate Gaussian process regression. Evaluation of the minimax errors allows us to identify cases when the variable fidelity data provides better interpolation accuracy than the exclusively high fidelity data for the same computational budget. These results allow us to calculate the optimal shares of variable fidelity data samples under the given computational budget constraint. Real and synthetic data experiments suggest that using the obtained optimal shares often outperforms natural heuristics in terms of the regression accuracy.

1 INTRODUCTION

In some cases sample data for regression modeling has variable fidelity: some data comes from a high fidelity source, some – from a low fidelity source [Forrester et al., 2007]. While there are many approaches to handle variable fidelity data including transfer learning [Pan and Yang, 2010] and space mapping [Bandler et al., 2004] techniques, engineers often use cokriging approach [Kennedy and O'Hagan, 2000] based on the Gaussian process framework [Zaytsev et al., 2014, Rasmussen and Williams, 2006]. Numerous applications of cokriging include geostatistics [Xu et al., 1992], aerospace [Han et al., 2013], and engineering [Koziel et al., 2014]. In this paper we also consider this approach for modeling data, obtained from high and low fidelity data sources.

The interest in accuracy of Gaussian process models for single fidelity data dates back to Wiener and Kolmogorov [Wiener, 1949, Kolmogorov, 1941]. They obtained an error at a specified point in the univariate case. Further progress in refining this estimate is available in the book by Stein [Stein, 2012], inspired by Ibragimov results [Ibragimov and Rozanov, 2012]. Recent results expand this setting by considering a more general interpolation error, equal to the integral of the squared difference between the true function and an its interpolation over the domain of interest, see [van der Vaart and van Zanten, 2008] for finite sample results in the multivariate case, and [Golubev and Krymova, 2013] for results about the minimax error of interpolation over an infinite regular univariate sample. References [Zhang et al., 2015b, Suzuki, 2012, Bhattacharya et al., 2014], to name a few, report similar results.

While in case of single fidelity data results are quite well established, there is only one paper [Zhang et al., 2015a], to our knowledge, that investigates the interpolation error for the variable fidelity data case from a theoretical point of view. For a squared exponential covariance function and a squared error at a single point, authors identify cases when regression modeling based on variable fidelity data is superior to using only the high fidelity data. Other papers dealing with variable fidelity regression modeling, [Velandia et al., 2016, Bevilacqua et al., 2015,

Proceedings of the 20^{th} International Conference on Artificial Intelligence and Statistics (AISTATS) 2017, Fort Lauderdale, Florida, USA. JMLR: W&CP volume 54.

Pascual and Zhang, 2006], focus on statistical properties of regression parameters estimates, but provide little insight into understanding how and why the variable fidelity modeling works.

Due to current apparent scarcity of theoretical foundations practitioners usually adopt heuristic rules in determining sizes of data samples of different fidelity and quantify when to use the variable fidelity data [Alexandrov et al., 1999, Simpson et al., 2008, Kuya et al., 2011]; or they use adaptive design of experiments approaches and surrogate based optimization directly, see [Ranjan et al., 2011, Kandasamy et al., 2016, Burnaev and Panov, 2015, Le Gratiet and Cannamela, 2015] and references therein.

The main contributions of this paper are the following:

- **Minimax interpolation error for the multivariate case.** We start with obtaining the interpolation error for the Gaussian process regression with a known covariance function. Then we derive the minimax interpolation error for functions from a general smoothness class in the multivariate case. This error is a nontrivial generalization of the univariate results obtained in [Golubev and Krymova, 2013].

- **The optimal ratio of sizes of variable fidelity data samples.** We obtain the interpolation error for the specified covariance function in the variable fidelity case, and then derive the minimax interpolation error in the general additive setting (cokriging) [Kennedy and O'Hagan, 2000]. With the derived minimax interpolation error we identify when and to which extent the variable fidelity regression modeling is beneficial compared to the regression modeling using only a high fidelity data under the same computational budget. We calculate the optimal ratio of sizes of variable fidelity data samples given the budget constraint. There is a certain gap between the theoretical setup we consider and the real world: we consider a setting that uses an infinite grid as a design of experiments and requires knowledge of relative complexities of high and low fidelity functions to calculate the optimal ratio of sample sizes. Nevertheless these theoretical results are sufficient to provide justification for the corresponding applied algorithm we develop.

- **The technique to select the ratio of sizes of variable fidelity data samples.** We elaborate on a method to choose the ratio inspired by our theoretical investigations. While the existing approaches usually work in adaptive design of experiments setting and pick points using sufficiently accurate regression models constructed beforehand [Ranjan et al., 2011], we offer a method to select sizes of high and low fidelity data samples to fit into a given computational budget and maximize accuracy of a resulting regression model prior to spending any significant resources on data generation. Our estimate depends only on the computational cost of variable fidelity data generation and on a correlation between high and low fidelity functions. As after an application of our approach we hope to get a good enough model, our approach can be used at a step that anticipates further adaptive design of experiments. We investigate the applicability of the proposed technique by comparing it to a number of natural baselines on synthetic and real datasets.

We provide proofs of all theorems in the supplementary materials.

2 MINIMAX INTERPOLATION ERROR FOR GAUSSIAN PROCESS REGRESSION

In case of Gaussian process regression there is a gap between theoretically tractable problems and practice. Namely, since the main tool for theoretical investigation is the Fourier transform, it is a common approach to consider the design of experiments based on an infinite grid [Golubev and Krymova, 2013, Stein, 2012], though in many cases the theoretical results are transferable to practical solutions. In this section we consider a design of experiments, belonging to some infinite grid, and later in the experimental section we show that our conclusions remain valid under finite sample random designs.

2.1 Interpolation Error

Let $f(\mathbf{x})$ be a stationary Gaussian process on $\mathbb{R}^d$ with a covariance function $R(\mathbf{x}) = \mathbb{E}(f(\mathbf{x}_0 + \mathbf{x}) - \mathbb{E}f(\mathbf{x}_0 + \mathbf{x}))(f(\mathbf{x}_0) - \mathbb{E}f(\mathbf{x}_0))$ and a spectral density $F(\boldsymbol{\omega})$

$$F(\boldsymbol{\omega}) = \int_{\mathbb{R}^d} e^{2\pi \mathrm{i} \boldsymbol{\omega}^{\mathrm{T}} \mathbf{x}} R(\mathbf{x}) d\mathbf{x}\,.$$

Suppose that we know values of realizations of $f(\cdot)$ at the infinite rectangular grid $D_H = \{\mathbf{x}_\mathbf{k} : \mathbf{x}_\mathbf{k} = H\mathbf{k}, \mathbf{k} \in \mathbb{Z}^d\}$, where H is a diagonal matrix with elements $h_1, \ldots, h_d$. An example of such design in the case of the input dimension $d = 2$ is provided in Figure 1.

We measure the interpolation error over the domain of

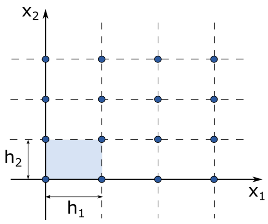


Figure 1: Design D_H for $d = 2$.

interest $\Omega_H = [0, h_1] \times \ldots \times [0, h_d]$ as follows:

$$\sigma_H^2(\tilde{f}, F) \stackrel{\text{def}}{=} \frac{1}{\mu(\Omega_H)} \int_{\Omega_H} \mathbb{E}\left[\tilde{f}(\mathbf{x}) - f(\mathbf{x})\right]^2 d\mathbf{x}\,, \quad (1)$$

where $\mu(\Omega_H) = \prod_{i=1}^d h_i$ is the Lebesgue measure of Ω_H, and $\tilde{f}(\mathbf{x})$ is an interpolation of $f(\mathbf{x})$. Here we consider $\tilde{f}(\mathbf{x})$ of the form

$$\tilde{f}(\mathbf{x}) = \mu(\Omega_H) \sum_{\mathbf{x}' \in D_H} K(\mathbf{x} - \mathbf{x}') f(\mathbf{x_k})\,, \quad (2)$$

where $K(\cdot)$ is a symmetric kernel.

Theorem 1. *The error of interpolation with $\tilde{f}(\mathbf{x})$ from (2), based on observations at points from D_H of a stationary Gaussian process $f(\mathbf{x})$ with spectral density $F(\boldsymbol{\omega})$, is equal to*

$$\sigma_H^2(\tilde{f}, F) = \int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \left[\left(1 - \hat{K}(\boldsymbol{\omega})\right)^2 + \right.$$
$$\left. + \sum_{\mathbf{x} \in D_{H^{-1}} \setminus \{\mathbf{0}\}} \hat{K}^2(\boldsymbol{\omega} + \mathbf{x})\right] d\boldsymbol{\omega}\,,$$

where $\hat{K}(\boldsymbol{\omega})$ is the Fourier transform of $K(\boldsymbol{\omega})$. Furthermore, the optimal $\hat{K}(\boldsymbol{\omega})$, minimizing the interpolation error, has the form

$$\hat{K}(\boldsymbol{\omega}) = \frac{F(\boldsymbol{\omega})}{\sum_{\mathbf{x} \in D_{H^{-1}}} F(\boldsymbol{\omega} + \mathbf{x})}\,.$$

Remark 1. *The function $\tilde{f}(\mathbf{x})$ that minimizes the squared error $\mathbb{E}(\tilde{f}(\mathbf{x}) - f(\mathbf{x}))^2$ has the form (2), where $K(\cdot)$ is a symmetric kernel. This motivates us to use $\tilde{f}(\mathbf{x})$ from (2) for interpolation.*

Remark 2. *It is easy to see that for $\tilde{f}(\mathbf{x})$ from (2) it holds that*

$$\sigma_H^2(\tilde{f}, F) = \sigma_{SH}^2(\tilde{f}, F)\,,$$

where $S = \mathrm{diag}(s_1, \ldots, s_d)$, with $s_i \in \mathbb{Z}^+, i = 1, \ldots, d$.

Using Theorem 1 one can estimate interpolation errors for various covariance functions. For example,

Corollary 1. *For a Gaussian process on $\mathbb{R}$ with exponential spectral density $F_\theta(\omega) = \frac{\theta}{\theta^2 + \omega^2}$ the interpolation error (1) for the best interpolation has the form:*

$$\sigma_h^2(\tilde{f}, F_\theta) \approx \frac{2}{3}\pi^2 \theta h + O((\theta h)^2),\ \theta h \to 0\,.$$

Corollary 2. *For a Gaussian process on $\mathbb{R}$ with squared exponential spectral density $F_\theta(\omega) = \frac{1}{\sqrt{\theta}} \exp\left(-\frac{\omega^2}{2\theta}\right)$ the interpolation error (1) for the best interpolation is bounded by:*

$$\frac{4}{3} h\sqrt{\theta} \exp\left(-\frac{1}{8h^2\theta}\right) \leq \sigma_h^2(\tilde{f}, F_\theta) \leq$$
$$\leq 7h\sqrt{\theta} \exp\left(-\frac{1}{8h^2\theta}\right),\ \theta h^2 \to 0\,.$$

2.2 Minimax Interpolation Error

For many covariance functions direct evaluation of the interpolation error can be technically cumbersome, especially for $d > 1$. Furthermore, in many cases the true covariance function is not known exactly, and calculating the interpolation error in such misspecified cases is even a harder task.

Instead we consider a minimax interpolation error that provides an answer in the worst case scenario. We define a set $\mathcal{F}(L, \boldsymbol{\lambda})$ of spectral densities $F(\boldsymbol{\omega})$ for a given $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d) \in \mathbb{R}^d$ and $L > 0$ as

$$\mathcal{F}(L, \boldsymbol{\lambda}) \stackrel{\text{def}}{=} \left\{F \,:\, \mathbb{E} \sum_{i=1}^d \lambda_i^2 \left(\frac{\partial f_F(\mathbf{x})}{\partial x_i}\right)^2 \leq L\,, \mathbf{x} \in \mathbb{R}^d\right\}, \quad (3)$$

where $f(\mathbf{x}) = f_F(\mathbf{x})$ is a realization of a Gaussian process with the spectral density $F(\boldsymbol{\omega})$ at the point $\mathbf{x} \in \mathbb{R}^d$. Sample realizations of Gaussian processes for different L in the case of $d = 1$ and the Matérn covariance function [Rasmussen and Williams, 2006] are shown in Figure 2.

The minimax interpolation error that describes how large the interpolation error is for the worst case scenario is defined as follows:

$$R^H(L, \boldsymbol{\lambda}) \stackrel{\text{def}}{=} \inf_{\tilde{f}} \sup_{F \in \mathcal{F}(L, \boldsymbol{\lambda})} \sigma_H^2(\tilde{f}, F)\,.$$

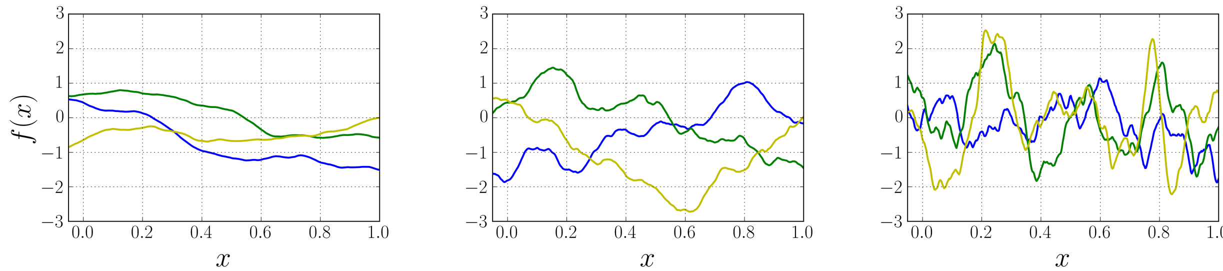
Then

Theorem 2. *For a Gaussian process $f(\mathbf{x})$, defined on $\mathbb{R}^d$ and evaluated on the design D_H, with the spectral density from the set $\mathcal{F}(L, \boldsymbol{\lambda})$, the minimax interpolation error has the form*

$$R^H(L, \boldsymbol{\lambda}) = \frac{L}{2\pi^2} \max_{i \in \{1, \ldots, d\}} \left(\frac{h_i}{\lambda_i}\right)^2.$$

Moreover, the minimax optimal interpolation $\tilde{f}(\mathbf{x})$ has the form

$$\tilde{f}(\mathbf{x}) = \mu(\Omega_H) \sum_{\mathbf{x}' \in D_H} K(\mathbf{x} - \mathbf{x}') f(\mathbf{x}')\,,$$



(a) $\mathcal{F}(10, 1)$ (b) $\mathcal{F}(100, 1)$ (c) $\mathcal{F}(1000, 1)$

Figure 2: Realizations of Gaussian processes with the Matérn covariance function $R(x) = (1 + \sqrt{3}\theta|x|)\exp(-\sqrt{3}\theta|x|)$ ($\nu = \frac{3}{2}$) for different values of L in $\mathcal{F}(L, 1)$ and $d = 1$.

where $K(\mathbf{x})$ *is a symmetric kernel with the Fourier transform* $\hat{K}(\boldsymbol{\omega})$ *defined as*

$$\hat{K}(\boldsymbol{\omega}) = \begin{cases} 1 - \sqrt{\sum_{i=1}^{d} \omega_i^2 \cdot h_i^2} & \textit{if } \sum_{i=1}^{d} \omega_i^2 \cdot h_i^2 \leq 1\,, \\ 0, & \textit{otherwise}\,. \end{cases}$$

While there is no explicit dependence of the minimax interpolation error on the input dimension d, growth of d leads to an exponential growth of the number of points in an unit hypercube. Thus, there is an exponential dependence of the minimax interpolation error on d if the density of observations is constant.

Note, that we can minimize the minimax interpolation error w.r.t. the diagonal matrix H in such a way as to keep fixed the average number of points belonging to a unit hypercube: $\prod_{i=1}^{d} \frac{1}{h_i} = n$. The diagonal elements h_i^* of the corresponding optimal matrix $H^* = \mathrm{diag}(h_1^*, \ldots, h_d^*)$ have the form $h_i^* = \sqrt[d]{\frac{n\lambda_i^d}{\prod_{j=1}^{d} \lambda_j}}$. The minimal minimax interpolation error is then equal to $R^{H^*}(L, \boldsymbol{\lambda}) = \frac{L}{2\pi^2} \sqrt[d/2]{\frac{n}{\prod_{i=1}^{d} \lambda_i}}$.

3 MINIMAX INTERPOLATION ERROR FOR A VARIABLE FIDELITY MODEL

3.1 Variable Fidelity Data Model

Suppose that the true function is modelled as

$$u(\mathbf{x}) = \rho f(\mathbf{x}) + g(\mathbf{x})\,, \tag{4}$$

where ρ is a fixed constant, and $f(\mathbf{x})$ and $g(\mathbf{x})$ are stationary independent Gaussian processes, defined on $\mathbb{R}^d$. This is the state-of-the-art cokriging approach used to model a variable fidelity data [Kennedy and O'Hagan, 2000].

We refer to a realization of $u(\mathbf{x})$ as a high fidelity function, and to a realization of $f(\mathbf{x})$ as a low fidelity function. Therefore $g(\mathbf{x})$ is a correction of $f(\mathbf{x})$ that appears due to a low fidelity nature of $f(\mathbf{x})$. The parameter ρ provides information on a strength of the relation between $f(\mathbf{x})$ and $u(\mathbf{x})$.

We observe values of $u(\mathbf{x})$ and $f(\mathbf{x})$ and we want to construct an interpolation $\tilde{u}(\mathbf{x})$ of the high fidelity function $u(\mathbf{x})$ on the basis of these variable fidelity observations.

3.2 Interpolation Error

It is natural to assume that we observe the cheap low fidelity function $f(\mathbf{x})$ on denser grid than the expensive high fidelity function $u(\mathbf{x})$. We observe $u(\mathbf{x})$ at points from $D_u = D_H$, and $f(\mathbf{x})$ at points from $D_f = D_{\frac{H}{m}}$ with a grid size ratio $m \in \mathbb{Z}^+$.

Using these observations we attempt to interpolate $u(\mathbf{x})$ within the hypercube Ω_H using a function $\tilde{u}(\mathbf{x})$ in order to minimize the interpolation error:

$$\sigma_{H,m}^2(\tilde{u}, F, G, \rho) \stackrel{\text{def}}{=} \frac{1}{\mu(\Omega_H)} \int_{\Omega_H} \mathbb{E}\left[\tilde{u}(\mathbf{x}) - u(\mathbf{x})\right]^2 d\mathbf{x}\,. \tag{5}$$

Theorem 3. *The minimum of interpolation error (5) of the variable fidelity data model* $u(\mathbf{x})$ *from (4), based on observations of* $u(\mathbf{x})$ *at points from* D_H *and observations of* $f(\mathbf{x})$ *at points from* $D_{\frac{H}{m}}$*, has the form:*

$$\sigma_{H,m}^2(\tilde{u}, F, G, \rho) = \sigma_H^2(\tilde{g}, G) + \rho^2 \sigma_{\frac{H}{m}}^2(\tilde{f}, F)\,, \tag{6}$$

where $\tilde{g}(\mathbf{x})$ *and* $\tilde{f}(\mathbf{x})$ *minimize* $\sigma_H^2(\tilde{g}, G)$ *and* $\sigma_{\frac{H}{m}}^2(\tilde{f}, F)$ *respectively.*

The explicit formula for optimal $\tilde{u}(\mathbf{x})$ is similar to the formula for $\tilde{f}(\mathbf{x})$ in Theorem 1, while as it is more cumbersome, we provide it in supplementary materials in the proof of the above theorem.

3.3 Minimax Interpolation Error

We obtain the minimax interpolation error for the variable fidelity case in the manner similar to the single fidelity case. Let us assume that the true spectral densities of the processes $f(\cdot)$ and $g(\cdot)$ are unknown, but sufficiently smooth, i.e. they belong to classes $\mathcal{F}(L_f) = \mathcal{F}(L_f, \mathbf{1})$ and $\mathcal{F}(L_g) = \mathcal{F}(L_g, \mathbf{1})$ respectively. Here for clarity of the presentation we limit ourselves to the case $\boldsymbol{\lambda} = \mathbf{1} \in \mathbb{R}^d$ and $H = h\mathrm{I}$ for some $h > 0$, where I is an identity matrix. In fact, results below hold in a more general setting, described in section 2 and defined by general values of $\boldsymbol{\lambda} \in \mathbb{R}^d$ and H. However, this additional sophistication blurs the main conclusions and provides little additional insight.

The goal is to obtain the minimax interpolation error for $u(\mathbf{x})$. In particular we want to get the minimax interpolation error for the variable fidelity data

$$R^{h,m}(L_f, L_g) \stackrel{\text{def}}{=} \inf_{\tilde{u}} \sup_{\substack{F \in \mathcal{F}(L_f),\\ G \in \mathcal{F}(L_g)}} \sigma^2_{h\mathrm{I},m}(\tilde{u}, F, G, \rho)\,. \quad (7)$$

Theorem 4. *Minimax interpolation error (7) of model (4), based on observations of $u(\mathbf{x})$ at points from D_H and observations of $f(\mathbf{x})$ at points from $D_{\frac{H}{m}}$, has the form*

$$R^{h,m}(L_f, L_g) = \rho^2 \frac{L_f}{2} \left(\frac{h}{m\pi}\right)^2 + \frac{L_g}{2}\left(\frac{h}{\pi}\right)^2. \quad (8)$$

4 OPTIMAL RATIO OF SIZES OF VARIABLE FIDELITY DATA SAMPLES

Obtained results allow us to get the optimal ratio m of sizes of variable fidelity data samples. We consider the following setting: one evaluation of $u(\mathbf{x})$ costs c, whereas one evaluation of $f(\mathbf{x})$ is 1; the total evaluation cost is equal to the number of points in a unit hypercube $\frac{1}{h^d}$ multiplied by the evaluation price; and the computational budget is set to B.

For such setup the total budget is equal to $c\frac{1}{h^d} + \delta\frac{1}{h^d}$, where $\delta = m^d$ is the ratio of sizes of variable fidelity data samples.

Using Theorem 4 we prove

Theorem 5. *The minimum of the minimax interpolation error (8) given the computational budget B has the form*

$$\min_{\substack{h,\delta:\\ B = \frac{c+\delta}{h^d}}} R^{h,m}(L_f, L_g) = \rho^2 \frac{L_f}{2}\left(\frac{c+\delta^*}{\pi B \delta^*}\right)^{\frac{2}{d}} + \frac{L_g}{2}\left(\frac{c+\delta^*}{\pi B}\right)^{\frac{2}{d}},$$

and the optimal ratio is $\delta^* = \left(\frac{L_f}{L_g} c \rho^2\right)^{\frac{d}{d+2}}$.

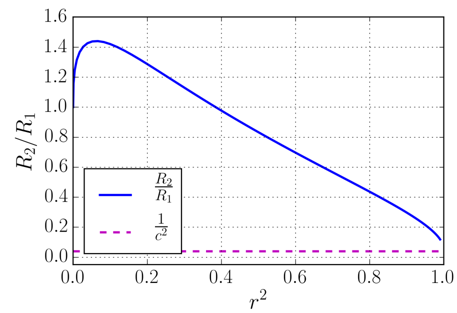


Figure 3: Dependence of the ratio $\frac{R_2}{R_1}$ of the minimax interpolation errors on the correlation coefficient r for $L_f = 3, L_g = 1, c = 5,\ d = 1$.

The optimal ratio δ^* depends on the relative cost c of the high fidelity function evaluation, the coefficient ρ and the smoothnesses L_f and L_g of $f(\mathbf{x})$ and $g(\mathbf{x})$ respectively and input dimension d.

If we evaluate exclusively $u(\mathbf{x})$, then we get the following minimax interpolation error given the budget B:

$$\min_{h:\, Bh^d = c} R^h(L_f, L_g) = \rho^2 \frac{L_f}{2}\left(\frac{c}{\pi B}\right)^{\frac{2}{d}} + \frac{L_g}{2}\left(\frac{c}{\pi B}\right)^{\frac{2}{d}}.$$

Note, that we can get similar results for a specific covariance function using Theorem 3 and Corollaries 1 and 2.

4.1 Comparison of Minimax Interpolation Errors Under Different Scenarios

Let us now investigate under what conditions and to what extent the usage of the variable fidelity data can decrease the interpolation error compared to using single fidelity data within the same computational budget. We denote by $R_2 = R^{h,\delta^*}(L_f, L_g, \rho)$ the minimax interpolation error, obtained when using the variable fidelity data, and by $R_1 = R^h(L_f, L_g, \rho)$ the minimax interpolation error, obtained when using only the high fidelity data. The ratio $\frac{R_2}{R_1}$ characterizes benefits of the variable fidelity data over single fidelity data: $\frac{R_2}{R_1} \geq 1$ means there is no advantage to using the variable fidelity data, while $\frac{R_2}{R_1} < 1$ implies that the variable fidelity data improves the accuracy of the interpolation.

The ratio $\frac{R_2}{R_1}$ has the form:

$$\frac{R_2}{R_1} = \frac{\left(1 + \left(\frac{L_f^d \rho^{2d}}{L_g^d c^2}\right)^{\frac{1}{d+2}}\right)^{\frac{d+2}{d}}}{1 + \rho^2 \frac{L_f}{L_g}}.$$

If we put $V_f = \mathbb{E} f^2(\mathbf{x})$ and $V_g = \mathbb{E} g^2(\mathbf{x})$, then the correlation coefficient r between $u(\mathbf{x})$ and $f(\mathbf{x})$ is $r =$

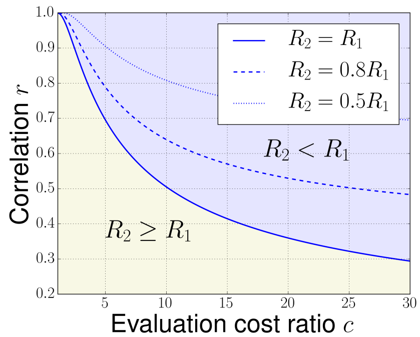


Figure 4: Curves $R_2 = kR_1$ for $L_f = 2$, $L_g = 1$, $d = 1$.

$\frac{1}{\sqrt{1+\frac{V_g}{V_f}\frac{1}{\rho^2}}}$. Thus for $r \to 0$ or $r \to 1$ it holds that

$$r \to 0 : \frac{R_2}{R_1} \approx 1 + \frac{d+2}{d}\left(\frac{L_f V_f}{L_g V_g}\right)^{\frac{d}{d+2}} \frac{r^{\frac{2d}{d+2}}}{c^{\frac{2}{d+2}}},$$

$$r \to 1 : \frac{R_2}{R_1} \approx \frac{1}{c^{\frac{2}{d}}} + \frac{2+d}{d}\left(\frac{L_g V_f}{L_f V_g}\right)^{\frac{d}{d+2}} \frac{(1-r^2)^{\frac{d}{d+2}}}{c^{\frac{4}{d(d+2)}}}.$$

If $r \to 0$ then the variable fidelity data is unable to improve the accuracy of the interpolation, while when $r \to 1$ the ratio $\frac{R_2}{R_1}$ approaches $\frac{1}{c^{\frac{2}{d}}}$, where usually $c \gg 1$.

In Figure 3 we show how the ratio $\frac{R_2}{R_1}$ depends on $r = r(\rho)$ in case of $d = 1$. For small r it holds that $R_2 > R_1$ no matter how large c is, while for high enough r the value of $\frac{R_2}{R_1}$ tends to $\frac{1}{c^2}$, $c \gg 1$.

Figure 4 depicts the smallest values of r in the case of $d = 1$, which for the fixed $c > 1$ provides $R_2 \leq R_1$.

For $d > 1$ and $c \gg 1$ the minimal value of r that provides $R_2 \leq R_1$ is of the order $\frac{1}{\sqrt{c}}$:

$$r \approx \frac{1}{\sqrt{c}}\left(\frac{V_f}{V_g}\right)^{\frac{1}{2}}\left(\frac{L_g}{L_f}\right)^{\frac{1}{2}}.$$

4.2 Optimal Ratio of Sample Sizes for Variable Fidelity Data

If we know the true covariance function it is easy to estimate the parameters L_f and L_g with the second derivatives of the covariance function $\frac{\partial^2 R(\mathbf{x})}{\partial x_i \partial x_j}$ at the point $\mathbf{x} = \mathbf{0}$. However, in the small sample case it is difficult to estimate the parameters of the covariance function [Zaytsev et al., 2014] or the sum of partial derivatives [Kucherenko et al., 2009] accurately. Also, in many practical cases it is often the case that L_f and L_g are close enough.

Therefore, assuming $L_f = L_g$ and using Theorem 4, we propose **Technique 1**, that can be used to estimate the optimal ratio of sample sizes δ^* and produce a design of experiments for the case of variable fidelity data. The advantage of the proposed technique is that it can be used even for a variable fidelity modeling approach different from the Gaussian process regression framework; and it requires little prior knowledge about the dependence structure between the high and the low fidelity functions, in particular, we only have to estimate the correlation coefficient r.

Technique 1 Generation of designs of experiments D_f and D_u for evaluations of the low fidelity function and the high fidelity function respectively.

Input: Correlation r between the variable fidelity observations, budget B, cost c of one high fidelity function evaluation (the cost of evaluating the low fidelity function is fixed at 1)

$\rho^2 \leftarrow 1/(\frac{1}{r^2} - 1)$
$\delta^* \leftarrow (c\rho^2)^{\frac{d}{d+2}}$
$n_f \leftarrow \frac{B\delta^*}{c+\delta^*}$, $n_u \leftarrow \frac{B}{c+\delta^*}$
Generate random nested designs of experiments D_f, D_u, $D_u \subseteq D_f$, with $|D_f| = n_f$, $|D_u| = n_u$.
return D_f, D_u

5 EXPERIMENTS

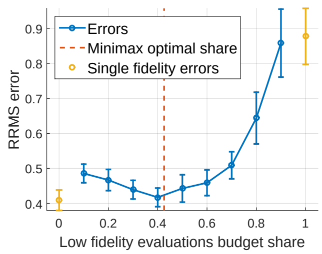
We evaluate the performance of the proposed algorithm for estimation of the optimal ratio of sample sizes for variable fidelity data in two steps: we start with synthetic data generated as realizations of Gaussian processes, and then consider real data problems that mostly originate from engineering applications.

We use the Matérn covariance function $R_{\boldsymbol{\theta}}(\mathbf{x}-\mathbf{x}')$ with $\nu = \frac{3}{2}$ that provides differentiable realizations of Gaussian processes [Rasmussen and Williams, 2006]:

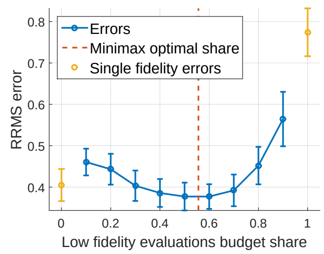
$$R_{\boldsymbol{\theta}}(\mathbf{x}-\mathbf{x}') = (1+\sqrt{3}\mathrm{d}_{\boldsymbol{\theta}}(\mathbf{x}-\mathbf{x}'))\exp(-\sqrt{3}\mathrm{d}_{\boldsymbol{\theta}}(\mathbf{x}-\mathbf{x}')),$$

where $\mathrm{d}_{\boldsymbol{\theta}}(\mathbf{x}-\mathbf{x}') = \sqrt{\sum_{i=1}^d \theta_i (x_i - x_i')^2}$. To construct a Gaussian process regression model we use Bayesian estimates of the covariance function parameters [Burnaev et al., 2016] obtained in a way similar to [Burnaev and Zaytsev, 2015], as open source software alternatives require manual tuning for each particular problem [Le Gratiet and Garnier, 2014].

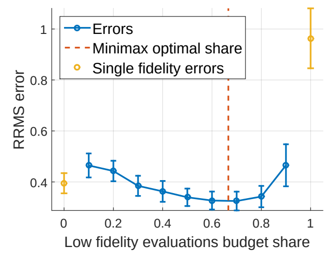
To assess model accuracy we use the Relative Root Mean Squared Error (RRMS) estimated using a dedicated test sample in case of a synthetic data and the cross-validation procedure in case of a real data. For a model $\tilde{u}(\mathbf{x})$ and a test sample $S_* = \{\mathbf{x}_i^*, u_i^* = u(\mathbf{x}_i^*)\}_{i=1}^{n_t}$ the RRMS error is given by RRMS $= \sqrt{\frac{\sum_{i=1}^{n_t}(u_i^* - \tilde{u}(\mathbf{x}_i^*))^2}{\sum_{i=1}^{n_t}(u_i^* - \overline{u})^2}}$, where $\overline{u} = \frac{1}{n_t}\sum_{i=1}^{n_t} u_i^*$.



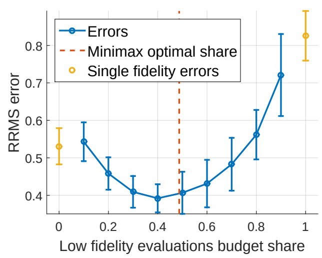
(a) $c = 5, r = 0.8$



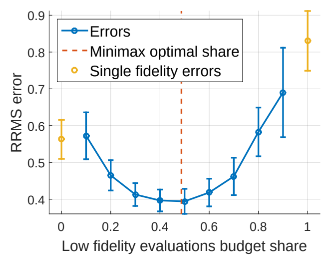
(b) $c = 5, r = 0.9$



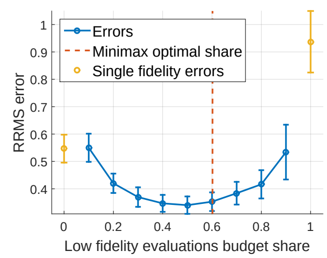
(c) $c = 5, r = 0.95$



(d) $c = 10, r = 0.8$



(e) $c = 10, r = 0.9$



(f) $c = 10, r = 0.95$

Figure 5: Synthetic data. Dependence of the RRMS error on the share of the budget, allocated for the low fidelity function evaluations. We consider the case of $d = 3$, different correlations r between the high and low fidelity functions and different cost of evaluating the high fidelity function c. Yellow points correspond to the case when we use either exclusively high or exclusively low fidelity data. The results are averaged across 20 runs.

Data and scripts, used to run the experiments, are at `gitlab.com/JohnDoe1989/VariableFidelityData`.

5.1 Synthetic Data Experiments

In this section we generate synthetic data as a realization of a Gaussian process with a specified covariance function. We follow the model $u(\mathbf{x}) = \rho f(\mathbf{x}) + g(\mathbf{x})$, with nested designs, i.e. $D_u \subseteq D_f$, and design points picked uniformly at random from $[0, 1]^d$. The total computational budget is 300, and the cost of evaluating $u(\mathbf{x})$ is either 5 or 10. Since the exact values of ρ and r are known, we use them in our experiments to get δ^*.

Figures 5 depict the dependence of the RRMS error on the proportion of the computational budget allocated for the low fidelity function evaluations. It can be seen that our estimate of the optimal ratio δ^* is close to the true optimal ratio despite the fact that estimates of the unknown parameters of the Gaussian Process regression model were used, and the design of experiments was not a grid.

5.2 Baseline Techniques

We compare our technique for estimation of the optimal ratio of sample sizes, which we call **MinMinimax**, to four baseline heuristics:

- **High** — only the high fidelity data is used,
- **Low** — we use only the low fidelity data,
- **EqualSize** — the sizes of low and high fidelity data samples are equal,
- **EqualBudget** — the budget is devoted equally to low and high fidelity function evaluations.

Baseline technique	n_{u}	n_{f}
High	B/c	0
EqualSize	$B/(c+1)$	$B/(c+1)$
EqualBudget	$B/(2c)$	$B/2$
Low	0	B
MinMinimax	$B/(c+\delta^*)$	$\delta^* B/(c+\delta^*)$

Table 1: Sizes of the high fidelity sample n_u and the low fidelity sample n_f in case of the budget B.

Relative sizes of samples for these techniques are given in Table 1.

5.3 Real Data Experiments

We consider the following real data problems. The first three of them (**Euler**, **Airfoil** [Bernstein et al., 2011], **MachAngle**) are concerned with calculation of lift and drag coefficients of an airfoil depending on flight conditions and airfoil geometry. To evaluate these outputs we use different solvers for the high and the low fidelity data sources. The next

Problem	High	EqualSize	EqualBudget	MinMinimax	Low
Euler-1	0.767 ± 0.550	0.892 ± 0.552	0.846 ± 0.177	$\mathbf{0.742 \pm 0.227}$	0.913 ± 0.0226
Euler-2	$\mathbf{0.066 \pm 0.022}$	0.077 ± 0.029	0.269 ± 0.189	0.380 ± 0.184	0.397 ± 0.102
Airfoil-1	0.546 ± 0.040	0.594 ± 0.085	0.539 ± 0.072	0.522 ± 0.050	$\mathbf{0.485 \pm 0.022}$
Airfoil-2	$\mathbf{0.120 \pm 0.009}$	0.142 ± 0.030	0.130 ± 0.031	0.138 ± 0.040	0.296 ± 0.020
MachAngle-1	$\mathbf{0.088 \pm 0.017}$	0.106 ± 0.025	0.195 ± 0.063	0.195 ± 0.063	0.405 ± 0.007
MachAngle-2	$\mathbf{0.093 \pm 0.005}$	0.114 ± 0.005	0.171 ± 0.009	0.179 ± 0.008	0.365 ± 0.004
Press12-1	0.559 ± 0.071	0.601 ± 0.072	0.3580 ± 0.022	$\mathbf{0.2779 \pm 0.016}$	0.2843 ± 0.013
Press12-2	0.443 ± 0.075	0.491 ± 0.079	0.2715 ± 0.037	$\mathbf{0.1768 \pm 0.016}$	$\mathbf{0.1768 \pm 0.016}$
Press13-1	0.559 ± 0.027	0.575 ± 0.025	0.386 ± 0.046	$\mathbf{0.348 \pm 0.046}$	0.5435 ± 0.011
Press13-2	0.449 ± 0.073	0.485 ± 0.066	0.278 ± 0.024	$\mathbf{0.1798 \pm 0.017}$	$\mathbf{0.1798 \pm 0.017}$
Disk-1	0.299 ± 0.066	0.3400 ± 0.079	0.192 ± 0.030	0.193 ± 0.029	$\mathbf{0.1638 \pm 0.010}$
Disk-2	0.446 ± 0.136	0.457 ± 0.125	0.299 ± 0.038	0.299 ± 0.038	$\mathbf{0.2723 \pm 0.032}$
SVM-1	$\mathbf{0.148 \pm 0.022}$	0.149 ± 0.026	0.184 ± 0.061	0.1642 ± 0.072	0.6081 ± 0.015
Supernova-1	0.0367 ± 0.0132	0.0439 ± 0.0145	$\mathbf{0.0153 \pm 0.0051}$	0.0574 ± 0.0003	0.0574 ± 0.0003

Table 2: RRMS errors averaged over 20 runs of the cross-validation procedure for the real data problems. Numbers after hyphen denote output number in a problem.

two problems (**Press** [Burnaev and Zaytsev, 2015], **Disk** [Zaytsev, 2016]) investigate dependence of maximum stress and maximum displacement on geometry of the equipment considered. Although three data fidelities are available in the **Press** problem, in each experiment we use only two. The last two problems ([Kandasamy et al., 2016], **SVM**, **Supernova**) are related to modeling dependence of the goodness-of-fit on model parameters. Input dimensions for these problems vary from two to eleven. More details on the problems are in the supplementary materials.

The budget B is equal to 300 for all problems except **Euler**, as in this problem the sample size is small. For consistency of the comparison the cost ratio c is 5 for all given problems. If the **MinMinimax** technique returns the sample size $n_u < 1$, then only the low fidelity data is used. For the **MinMinimax** technique we use the correlation coefficient r estimated using the whole available data sample. In addition, to keep the comparison meaningful we normalize all the data before constructing regression models to get variables with zero mean and unit variance.

We provide errors in Table 2, which show that the best results are typically obtained using the proposed **MinMinimax** approach. However, there are two drawbacks: sometimes it is impossible to improve the model accuracy using variable fidelity data; or too small sample size is selected making it impossible to construct a reliable regression model. For example, for the **Supernova** dataset the **MinMinimax** method works poorly because it suggests to use the high fidelity sample size equal to four, which is insufficient for the cokriging to work efficiently. Thus, we suggest to impose a lower bound for the size of the high fidelity data sample.

6 CONCLUSIONS

We prove the minimax interpolation error for the Gaussian process regression in the multivariate case. The obtained results are used to estimate the interpolation error for the regression modeling with the variable fidelity data. This allows us to identify settings in which the accuracy of the regression model can be improved with the variable fidelity data.

Moreover, we estimate the optimal ratio of sizes of the variable fidelity data samples. Using both synthetic and real problems, we demonstrate that this ratio can be used when producing a design of experiments.

However, there is still room for improvement of the proposed approach: it requires an accurate estimate of the correlation coefficient, and it doesn't take into account inaccuracies of estimates of the regression model parameters. Furthermore, in this paper we consider the case of two fidelity levels only, whereas in practice multiple fidelity levels can be accessible.

Acknowledgments

The research, presented in Section 5 of this paper was supported solely by the Russian Science Foundation grant (project 14-50-00150).

The authors would like to thank Yu. Golubev for his helpful comments, I. Nazarov and M. Panov for their help with proofreading.

References

[Alexandrov et al., 1999] Alexandrov, N., Lewis, R., Gumbert, C., Green, L., and Newman, P. (1999).

Optimization with variable-fidelity models applied to wing design. Technical report, DTIC Document.

[Bandler et al., 2004] Bandler, J., Cheng, Q., Dakroury, S., Mohamed, A., Bakr, M., Madsen, K., and Søndergaard, J. (2004). Space mapping: the state of the art. *Microwave Theory and Techniques, IEEE Transactions on*, 52(1):337–361.

[Bernstein et al., 2011] Bernstein, A., Burnaev, E., Chernova, S., Zhu, F., and Qin, N. (2011). Comparison of three geometric parameterization methods and their effect on aerodynamic optimization. In *Proceedings of International Conference on Evolutionary and Deterministic Methods for Design, Optimization and Control with Applications to Industrial and Societal Problems, Eurogen*, pages 14–16.

[Bevilacqua et al., 2015] Bevilacqua, M., Fassò, A., Gaetan, C., Porcu, E., and Velandia, D. (2015). Covariance tapering for multivariate Gaussian random fields estimation. *Statistical Methods & Applications*, pages 1–17.

[Bhattacharya et al., 2014] Bhattacharya, A., Pati, D., and Dunson, D. (2014). Anisotropic function estimation using multi-bandwidth gaussian processes. *Annals of statistics*, 42(1):352.

[Burnaev and Panov, 2015] Burnaev, E. and Panov, M. (2015). Adaptive design of experiments based on Gaussian processes. In *Statistical Learning and Data Sciences*, pages 116–125. Springer.

[Burnaev et al., 2016] Burnaev, E., Panov, M., and Zaytsev, A. (2016). Regression on the basis of nonstationary gaussian processes with bayesian regularization. *Journal of Communications Technology and Electronics*, 61(6):661–671.

[Burnaev and Zaytsev, 2015] Burnaev, E. and Zaytsev, A. (2015). Surrogate modeling of multifidelity data for large samples. *Journal of Communications Technology and Electronics*, 60(12):1348–1355.

[Forrester et al., 2007] Forrester, A., Sóbester, A., and Keane, A. (2007). Multi-fidelity optimization via surrogate modelling. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 463(2088):3251–3269.

[Golubev and Krymova, 2013] Golubev, G. and Krymova, E. (2013). On interpolation of smooth processes and functions. *Problems of Information Transmission*, 49(2):127–148.

[Han et al., 2013] Han, Z., Görtz, S., and Zimmermann, R. (2013). Improving variable-fidelity surrogate modeling via gradient-enhanced kriging and a generalized hybrid bridge function. *Aerospace Science and Technology*, 25(1):177–189.

[Ibragimov and Rozanov, 2012] Ibragimov, I. and Rozanov, Y. (2012). *Gaussian random processes*, volume 9. Springer Science & Business Media.

[Kandasamy et al., 2016] Kandasamy, K., Dasarathy, G., Oliva, J., Schneider, J., and Póczos, B. (2016). Gaussian process bandit optimisation with multi-fidelity evaluations. In *Advances in Neural Information Processing Systems*, pages 992–1000.

[Kennedy and O'Hagan, 2000] Kennedy, M. and O'Hagan, A. (2000). Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, 87(1):1–13.

[Kolmogorov, 1941] Kolmogorov, A. (1941). Interpolation and extrapolation of stationary random sequences. *Izv. Akad. Nauk SSSR, Ser. Mat.*, 5(1):3–14.

[Koziel et al., 2014] Koziel, S., Bekasiewicz, A., Couckuyt, I., and Dhaene, T. (2014). Efficient multi-objective simulation-driven antenna design using co-kriging. *Antennas and Propagation, IEEE Transactions on*, 62(11):5900–5905.

[Kucherenko et al., 2009] Kucherenko, S. et al. (2009). Derivative based global sensitivity measures and their link with global sensitivity indices. *Mathematics and Computers in Simulation*, 79(10):3009–3017.

[Kuya et al., 2011] Kuya, Y., Takeda, K., Zhang, X., and Forrester, A. (2011). Multifidelity surrogate modeling of experimental and computational aerodynamic data sets. *AIAA journal*, 49(2):289–298.

[Le Gratiet and Cannamela, 2015] Le Gratiet, L. and Cannamela, C. (2015). Cokriging-based sequential design strategies using fast cross-validation techniques for multi-fidelity computer codes. *Technometrics*, 57(3):418–427.

[Le Gratiet and Garnier, 2014] Le Gratiet, L. and Garnier, J. (2014). Recursive co-kriging model for design of computer experiments with multiple levels of fidelity. *International Journal for Uncertainty Quantification*, 4(5).

[Pan and Yang, 2010] Pan, S. and Yang, Q. (2010). A survey on transfer learning. *Knowledge and Data Engineering, IEEE Transactions on*, 22(10):1345–1359.

[Pascual and Zhang, 2006] Pascual, F. and Zhang, H. (2006). Estimation of linear correlation coefficient of two correlated spatial processes. *Sankhyā: The Indian Journal of Statistics*, pages 307–325.

[Ranjan et al., 2011] Ranjan, P., Lu, W., Bingham, D., Reese, S., Williams, B. J., Chou, C.-C., Doss, F., Grosskopf, M., and Holloway, J. P. (2011). Follow-up experimental designs for computer models and physical processes. *Journal of Statistical Theory and Practice*, 5(1):119–136.

[Rasmussen and Williams, 2006] Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian processes for machine learning*. The MIT Press.

[Simpson et al., 2008] Simpson, T., Toropov, V., Balabanov, V., and Viana, F. (2008). Design and analysis of computer experiments in multidisciplinary design optimization: a review of how far we have come or not. In *12th AIAA/ISSMO multidisciplinary analysis and optimization conference*, volume 5, pages 10–12.

[Stein, 2012] Stein, M. (2012). *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media.

[Suzuki, 2012] Suzuki, T. (2012). PAC-Bayesian bound for gaussian process regression and multiple kernel additive model. In *Proceedings of the 25th annual conference on Computational Learning Theory*, volume 22, pages 39–60.

[van der Vaart and van Zanten, 2008] van der Vaart, A. and van Zanten, J. (2008). Rates of contraction of posterior distributions based on Gaussian process priors. *The Annals of Statistics*, pages 1435–1463.

[Velandia et al., 2016] Velandia, D., Bachoc, F., Bevilacqua, M., Gendre, X., and Loubes, J. (2016). Maximum likelihood estimation for a bivariate Gaussian process under fixed domain asymptotics. *arXiv preprint arXiv:1603.09059*.

[Wiener, 1949] Wiener, N. (1949). *Extrapolation, interpolation, and smoothing of stationary time series*, volume 2. MIT press Cambridge, MA.

[Xu et al., 1992] Xu, W., Tran, T., Srivastava, R., and Journel, A. (1992). Integrating seismic data in reservoir modeling: the collocated cokriging alternative. In *SPE annual technical conference and exhibition*. Society of Petroleum Engineers.

[Zaytsev, 2016] Zaytsev, A. (2016). Variable fidelity regression using low fidelity function blackbox and sparsification. In *Conformal and Probabilistic Prediction with Applications*, pages 147–164. Springer.

[Zaytsev et al., 2014] Zaytsev, A., Burnaev, E., and Spokoiny, V. (2014). Properties of the Bayesian parameter estimation of a regression based on Gaussian processes. *Journal of Mathematical Sciences*, 203(6):789–798.

[Zhang et al., 2015a] Zhang, H., Cai, W., et al. (2015a). When doesnt cokriging outperform kriging? *Statistical Science*, 30(2):176–180.

[Zhang et al., 2015b] Zhang, Y., Duchi, J., and Wainwright, M. (2015b). Divide and conquer kernel ridge regression: A distributed algorithm with minimax optimal rates. *Journal of Machine Learning Research*, 16:3299–3340.