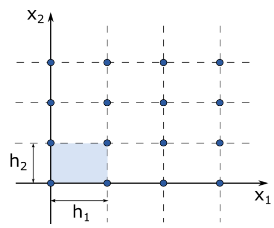


Figure 1: Design D_H for $d = 2$.

Supplementary to the paper "Minimax Approach to Variable Fidelity Data Interpolation"

A.Zaytsev
a.zaytsev@skoltech.ru
Skoltech, IITP RAS

E.Burnaev
e.burnaev@skoltech.ru
Skoltech, IITP RAS, HSE

1 INTRODUCTION

This supplementary provides proofs of the main statements from the paper. We start with description of all statements similar to that in the paper, and then give proofs for them in separate sections. At the end there is also Section 9 devoted to description of the real problems considered in the paper.

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

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

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

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

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*

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

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

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

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*

$$\sigma^2_H(\tilde{f}, F) = \sigma^2_{SH}(\tilde{f}, F) \,,$$

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

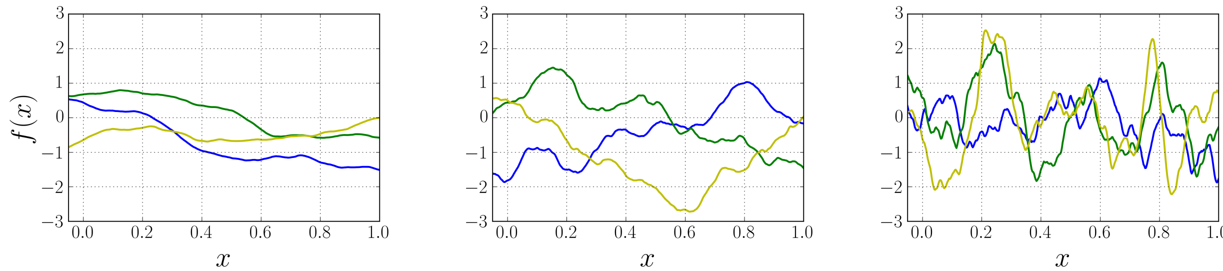
Using Theorem 1 one can estimate interpolation errors for various covariance functions. For example,

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:*

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

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:*

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



(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 $d = 1$.

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 $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_d) \in \mathbb{R}^d$ and $L > 0$ as

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

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*

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

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

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

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

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

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

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

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*

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

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

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.

5 PROOFS FOR SUBSECTION 2.1

Proof of Theorem 1. It is easy to see that

$$\mathbb{E}[f(\mathbf{x}) - \tilde{f}(\mathbf{x})]^2 = \int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \left| 1 - |H| \sum_{\mathbf{k}\in\mathbb{Z}^d} K(\mathbf{x} - \mathbf{x_k}) \exp(-2\pi i \boldsymbol{\omega}^T (\mathbf{x_k} - \mathbf{x})) \right|^2 d\boldsymbol{\omega} =$$

$$= \int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \left| 1 - |H| \sum_{\mathbf{k}\in\mathbb{Z}^d} \left(\int_{\mathbb{R}^d} \hat{K}(\mathbf{u}) \exp(-2\pi i \mathbf{u}^T (\mathbf{x} - \mathbf{x_k})) d\mathbf{u} \right) \exp(-2\pi i \boldsymbol{\omega}^T (\mathbf{x_k} - \mathbf{x})) \right|^2 d\boldsymbol{\omega},$$

where $\hat{K}(\mathbf{u})$ is the Fourier transform of $K(\mathbf{x})$. As Poisson summation formula suggests:

$$\sum_{\mathbf{k}\in\mathbb{Z}^d} \exp(2\pi i \mathbf{k}^T \boldsymbol{\omega}) = \sum_{\mathbf{k}\in\mathbb{Z}^d} \delta(\boldsymbol{\omega} + \mathbf{k}),$$

where $\delta(\boldsymbol{\omega})$ is the Dirac delta function, then

$$\mathbb{E}[f(\mathbf{x}) - \tilde{f}(\mathbf{x})]^2 = \int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \left| 1 - |H| \sum_{\mathbf{k}\in\mathbb{Z}^d} \int_{\mathbb{R}^d} \hat{K}(\mathbf{u}) \exp(2\pi i (\boldsymbol{\omega} - \mathbf{u})^T \mathbf{x}) \delta(\mathbf{u} - \boldsymbol{\omega} + H^{-1}\mathbf{k}) d\mathbf{u} \right|^2 d\boldsymbol{\omega} =$$

$$= \int_{\mathbb{R}^d} \left| 1 - \sum_{\mathbf{k}\in\mathbb{R}^d} \hat{K}(\boldsymbol{\omega} - H^{-1}\mathbf{k}) \exp(2\pi i H^{-1} \mathbf{x}^T \mathbf{k}) \right|^2 d\boldsymbol{\omega}.$$

Taking into account orthogonality of the system of functions $\exp(2\pi i H^{-1}\mathbf{x}^T\mathbf{k})$ on $\mathbf{x} \in [0, h_1] \times \ldots \times [0, h_d]$ we integrate the equality to get the interpolation error

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

To get $\hat{K}(\boldsymbol{\omega})$ that minimizes the interpolation error we rewrite it as

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

To minimize this error we solve this quadratic optimization problem for each $\boldsymbol{\omega}$ and get:

$$\hat{K}(\boldsymbol{\omega}) = \frac{\hat{F}(\boldsymbol{\omega})}{\sum_{\mathbf{k} \in \mathbb{Z}^d} \hat{F}(\boldsymbol{\omega} + H^{-1}\mathbf{k})}.$$

Then

$$\sigma_H^2(\tilde{f}, F) = \int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \frac{\sum_{\mathbf{k} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}} \hat{F}(\boldsymbol{\omega} + H^{-1}\mathbf{k})}{\sum_{\mathbf{k} \in \mathbb{Z}^d} \hat{F}(\boldsymbol{\omega} + H^{-1}\mathbf{k})} d\boldsymbol{\omega}. \tag{9}$$

□

Proof of Remark 1. It holds that the best approximation has the form

$$\tilde{f}(\mathbf{x}) = \mu(\Omega_H) \sum_{\mathbf{k} \in \mathbb{Z}^d} \phi(\mathbf{x}, \mathbf{x_k}) f(\mathbf{x_k})$$

for some $\phi(\mathbf{x}, \mathbf{x}')$. As Wiener-Hopf equations for the covariance function $R(\mathbf{x})$ hold, then

$$\sum_{\mathbf{k} \in \mathbb{Z}^d} \phi(\mathbf{x}, \mathbf{x_k}) R(\mathbf{x_k} - \mathbf{x_m}) = R(\mathbf{x} - \mathbf{x_m}) \tag{10}$$

for all $\mathbf{m} \in \mathbb{Z}^d$. Let us prove that $\phi(\mathbf{x}, \mathbf{x_k}) = \phi(\mathbf{x} - \mathbf{x_k})$.

Let us consider two sums from (10):

$$\sum_{\mathbf{k} \in \mathbb{Z}^d} \phi(\mathbf{x}, \mathbf{x_k}) R(\mathbf{x_k} - \mathbf{x_m}) = R(\mathbf{x} - \mathbf{x_m}),$$

$$\sum_{\mathbf{k} \in \mathbb{Z}^d} \phi(\mathbf{x} - \mathbf{x_s}, \mathbf{x_k}) R(\mathbf{x_k} - \mathbf{x_{m-s}}) = R(\mathbf{x} - \mathbf{x_s} - \mathbf{x_{m-s}}).$$

As $\mathbf{x_{m-s}} = H\mathbf{m} - H\mathbf{s} = \mathbf{x_m} - \mathbf{x_s}$, then

$$\sum_{\mathbf{k} \in \mathbb{Z}^d} \phi(\mathbf{x}, \mathbf{x_k}) R(\mathbf{x_k} - \mathbf{x_m}) = R(\mathbf{x} - \mathbf{x_m}) = R(\mathbf{x} - \mathbf{x_s} - \mathbf{x_{m-s}}) = \sum_{\mathbf{k} \in \mathbb{Z}^d} \phi(\mathbf{x} - \mathbf{x_s}, \mathbf{x_k}) R(\mathbf{x_k} - \mathbf{x_{m-s}}).$$

Consequently,

$$\sum_{\mathbf{k} \in \mathbb{Z}^d} [\phi(\mathbf{x}, \mathbf{x_k}) - \phi(\mathbf{x} - \mathbf{x_s}, \mathbf{x_k} - \mathbf{x_s})] \, R(\mathbf{x_k} - \mathbf{x_m}) = 0.$$

Positive definiteness of the covariance function $R(\mathbf{x})$ implies that

$$\phi(\mathbf{x}, \mathbf{x_k}) = \phi(\mathbf{x} - \mathbf{x_s}, \mathbf{x_k} - \mathbf{x_s}).$$

For $\mathbf{x_s} = \mathbf{x_k}$ we get

$$\phi(\mathbf{x}, \mathbf{x_k}) = \phi(\mathbf{x} - \mathbf{x_k}, \mathbf{0}) = K(\mathbf{x} - \mathbf{x_k}).$$

Due to Poisson summation formula it holds that

$$\frac{1}{\mu(\Omega_H)}\Phi(\boldsymbol{\omega})\sum_{\mathbf{k}\in\mathbb{Z}^d}F(\boldsymbol{\omega}-\boldsymbol{\omega}_{\mathbf{k}}) = F(\boldsymbol{\omega}),$$

where $\Phi(\boldsymbol{\omega})$ is the Fourier transform of $\phi(\mathbf{x})$. Then

$$\Phi(\boldsymbol{\omega}) = \frac{\mu(\Omega_H)F(\boldsymbol{\omega})}{\sum_{\mathbf{k}\in\mathbb{Z}^d}F(\boldsymbol{\omega}-\boldsymbol{\omega}_{\mathbf{k}})}.$$

So, optimal interpolation has the form:

$$\tilde{f}(\mathbf{x}) = \mu(\Omega_H)\sum_{\mathbf{k}\in\mathbb{Z}^d}K(\mathbf{x}-\mathbf{x}_{\mathbf{k}})f(\mathbf{x}_{\mathbf{k}}).$$

Also

$$\hat{K}(\boldsymbol{\omega}) = \frac{\Phi(\boldsymbol{\omega})}{\mu(\Omega_H)}.$$

□

Proof of Corollary 1. We get the interpolation error for an exponential covariance function of the form $R(x) = \sqrt{\frac{\pi}{2}}\exp\left(-\theta|x|\right)$ for $x \in \mathbb{R}$. The spectral density for this covariance function is $F(\omega) = \frac{\theta}{\theta^2+\omega^2}$.

We want to evaluate the interpolation error

$$\sigma_h^2\left(\tilde{f}, F\right) = \int_{-\infty}^{\infty}F(\omega)\frac{\sum_{k\neq 0}F(\omega+\frac{k}{h})}{\sum_k F(\omega+\frac{k}{h})}d\omega.$$

It holds that

$$\sum_{k=-\infty}^{\infty}F(\omega+\frac{k}{h}) = \sum_{k=-\infty}^{\infty}\frac{\theta}{(\omega+\frac{k}{h})^2+\theta^2} = h\sum_{k=-\infty}^{\infty}\frac{h\theta}{(h\omega+k)^2+h^2\theta^2} =$$
$$= \pi h\coth(\pi\theta h)\frac{1}{1+\sin^2(\pi h\omega)(\coth^2(\pi\theta h)-1)}.$$

Then

$$\int_{-\infty}^{\infty}F(\omega)\frac{\sum_{k\neq 0}F(\omega+\frac{k}{h})}{\sum_k F(\omega+\frac{k}{h})}d\omega = \int_{-\infty}^{\infty}\frac{\theta}{\theta^2+\omega^2}\left(1-\frac{\theta}{\theta^2+\omega^2}\frac{1+\sin^2(\pi h\omega)(\coth^2(\pi\theta h)-1)}{\pi h\coth(\pi\theta h)}\right)d\omega.$$

We can integrate three terms inside the integral analytically. Namely,

$$\int_{-\infty}^{\infty}\frac{\theta}{\theta^2+\omega^2}d\omega = \pi.$$

Also

$$\int_{-\infty}^{\infty}\frac{\theta^2}{(\theta^2+\omega^2)^2}d\omega = \frac{\pi}{2\theta}.$$

Finally

$$\int_{-\infty}^{\infty}\frac{\theta^2}{(\theta^2+\omega^2)^2}\sin^2(\pi\omega h)d\omega = -\frac{\pi^2 h}{2}\left(\cosh(\pi\theta h)-\sinh(\pi\theta h)\right)\left(\cosh(\pi\theta h)-\left(\frac{1}{\pi\theta h}+1\right)\sinh(\pi\theta h)\right).$$

Consequently,

$$\int_{-\infty}^{\infty}\frac{\theta}{\theta^2+\omega^2}\left(1-\frac{\theta}{\theta^2+\omega^2}\frac{1+\sin^2(\pi h\omega)(\coth^2(\pi\theta h)-1)}{\pi h\coth(\pi\theta h)}\right)d\omega =$$
$$\pi - \frac{\pi}{2\pi\theta h\coth(\pi\theta h)}+$$
$$+\frac{\pi^2}{2}\exp(-\pi\theta h)\left(\exp(-\pi\theta h)-\frac{1}{\pi\theta h}\sinh(\pi\theta h)\right)\frac{\coth^2(\pi\theta h)-1}{\coth(\pi\theta h)}.$$

For $h \to 0$ we get Taylor series for the obtained interpolation error:

$$\sigma_h^2(\tilde{f}, F) = \frac{2\pi^2}{3}\theta h + O((\theta h)^2).$$

□

Proof of Corollary 2. Note, that the interpolation error has the form

$$\sigma_h^2(\tilde{f}, F) = \int_{-\infty}^{\infty} F(\omega) \frac{\sum_{k \neq 0} F(\omega + \frac{k}{h})}{\sum_s F(\omega + \frac{s}{h})} d\omega.$$

We get lower and upper bounds for this expression. We denote $v = \frac{1}{h}$.

We get upper bound for the interpolation error by splitting integration region $(-\infty, \infty)$ to three intervals $(-\infty, -v/2]$, $(-v/2, v/2]$, $(v/2, +\infty)$ and obtaining an upper bound for each of them.

Note that

$$0 \leq \frac{\sum_{k \neq 0} F(\omega + kv)}{\sum_s F(\omega + sv)} \leq 1.$$

Consequently, using Chernov type bounds [Chang et al., 2011] we get

$$\begin{aligned} &\int_{v/2}^{\infty} F(\omega) \frac{\sum_{k \neq 0} F(\omega + kv)}{\sum_s F(\omega + sv)} d\omega \leq \int_{v/2}^{\infty} F(\omega) d\omega = \\ &= \int_{v/2}^{\infty} \frac{1}{\sqrt{\theta}} \exp\left(-\frac{\omega^2}{2\theta}\right) d\omega \leq \sqrt{2} \exp\left(-\frac{v^2}{8\theta}\right). \end{aligned} \tag{11}$$

In a similar way get an upper bound for the interval $(-\infty, -v/2)$.

Now we get an estimate for the interval $(-v/2, v/2)$. We start with an upper bound and a lower bound for series $\sum_{s \neq 0} F(\omega + sv)$. Spectral density for squared exponential covariance function decreases at $[0, +\infty)$ with respect to ω. Thus,

$$\int_{\Delta + u}^{+\infty} F(x) dx \leq \sum_{s=1}^{\infty} \Delta F(\Delta s + u) \leq \Delta F(s + u) + \int_{\Delta + u}^{+\infty} F(x) dx.$$

Using [Abramowitz and Stegun, 1964], Formula 7.1.13, we get for ω such that $|\omega| \leq \frac{v}{2}$:

$$\frac{4\sqrt{\theta}}{v + \sqrt{v^2 + 16\theta}} \exp\left(-\frac{v^2}{8\theta}\right) \leq \int_{\frac{v}{2}}^{\infty} F(\omega) d\omega \leq \frac{4\sqrt{\theta}}{v + \sqrt{v^2 + \frac{32}{\pi}\theta}} \exp\left(-\frac{v^2}{8\theta}\right).$$

And

$$v \sum_{k \in \mathbb{Z}^+} F(\omega + kv) \leq vF(\omega + v) + \int_{\frac{v}{2}}^{\infty} F(\omega) d\omega \leq \frac{v}{\sqrt{\theta}} \exp\left(-\frac{v^2}{8\theta}\right) + \frac{4\sqrt{\theta}}{v + \sqrt{v^2 + \frac{32}{\pi}\theta}} \exp\left(-\frac{v^2}{8\theta}\right).$$

Now we are ready to get an upper bound for the integral over the interval $(-v/2, v/2)$ for big enough v:

$$\int_{-v/2}^{v/2} F(\omega)\frac{\sum_{k\neq 0} F(\omega+kv)}{\sum_s F(\omega+sv)} d\omega \leq$$
$$\leq \int_{-v/2}^{v/2} F(\omega)\frac{F(\omega+v)+F(\omega-v)+\frac{4\sqrt{\theta}}{v\left(v+\sqrt{v^2+\frac{32}{\pi}\theta}\right)}\exp\left(-\frac{v^2}{8\theta}\right)}{F(\omega)+F(\omega+v)+F(\omega-v)+\frac{4\sqrt{\theta}}{v\left(v+\sqrt{v^2+\frac{32}{\pi}\theta}\right)}\exp\left(-\frac{v^2}{8\theta}\right)} d\omega \leq$$
$$\leq \int_{-v/2}^{v/2} F(\omega)\frac{F(\omega+v)+F(\omega-v)}{F(\omega)+F(\omega+v)+F(\omega-v)} d\omega + \int_{-v/2}^{v/2} F(\omega)\frac{\frac{4\sqrt{\theta}}{v\left(v+\sqrt{v^2+\frac{32}{\pi}\theta}\right)}\exp\left(-\frac{v^2}{8\theta}\right)}{F(\omega)+\frac{4\sqrt{\theta}}{v\left(v+\sqrt{v^2+\frac{32}{\pi}\theta}\right)}\exp\left(-\frac{v^2}{8\theta}\right)} d\omega \leq$$
$$\leq \int_{-v/2}^{v/2} F(\omega+v)+F(\omega-v) d\omega + \frac{4\sqrt{\theta}}{v+\sqrt{v^2+\frac{32}{\pi}\theta}}\exp\left(-\frac{v^2}{8\theta}\right) \leq$$
$$\leq \frac{12\sqrt{\theta}}{v+\sqrt{v^2+\frac{32}{\pi}\theta}}\exp\left(-\frac{v^2}{8\theta}\right) \leq \frac{7\sqrt{\theta}}{v}\exp\left(-\frac{v^2}{8\theta}\right).$$

It holds that
$$\frac{\sum_{k\neq 0} F(\omega+kv)}{\sum_s F(\omega+sv)} \geq \frac{F(\omega+v)+F(\omega-v)}{F(\omega)+F(\omega+v)+F(\omega-v)}.$$
For ω such that $|\omega| \leq \frac{v}{2}$ we get that:
$$1+\frac{F(\omega+v)}{F(\omega)}+\frac{F(\omega-v)}{F(\omega)} \leq 3.$$

Then for sufficiently large v the following lower bound holds:

$$\int_{-\infty}^{\infty} F(\omega)\frac{\sum_{k\neq 0} F(\omega+kv)}{\sum_s F(\omega+sv)} d\omega \geq \int_{-v/2}^{v/2} F(\omega)\frac{\sum_{k\neq 0} F(\omega+kv)}{\sum_s F(\omega+sv)} d\omega \geq$$
$$\geq \int_{-\frac{v}{2}}^{\frac{v}{2}} F(\omega)\frac{F(\omega+v)+F(\omega-v)}{F(\omega)+F(\omega+v)+F(\omega-v)} d\omega \geq \int_{-\frac{v}{2}}^{\frac{v}{2}} \frac{F(\omega+v)+F(\omega-v)}{3} d\omega =$$
$$= \frac{2}{3}\int_{\frac{v}{2}}^{\frac{3v}{2}} F(\omega) d\omega \geq \frac{2}{3}\left(\frac{4\sqrt{\theta}}{v+\sqrt{v^2+16\theta}}\exp\left(-\frac{v^2}{8\theta}\right) - \frac{4\sqrt{\theta}}{3v+\sqrt{9v^2+\frac{32}{\pi}\theta}}\exp\left(-\frac{9v^2}{8\theta}\right)\right) \geq$$
$$\geq \frac{4}{3}\frac{\sqrt{\theta}}{v}\exp\left(-\frac{v^2}{8\theta}\right).$$

□

6 PROOFS FOR SUBSECTION 2.2

We need the following lemma to complete the proof of the main result

Lemma 1. *Let $c \geq 0$ and $\omega \geq 0$ be such that $c^2+\omega^2 \leq 1$, $c^2+(1-\omega^2) \leq 1$. Then*

$$\left(1-\sqrt{c^2+\omega^2}\right)^2 + \left(1-\sqrt{c^2+(1-\omega)^2}\right)^2 \leq \left(1-\sqrt{c^2}\right)^2 = (1-c)^2. \quad (12)$$

Proof. We start with a scheme of the proof. We prove that for ω, that maximizes left hand side of the inequality (12), this inequality holds. To prove this we show that for admissible $\omega \in [1-\sqrt{1-c^2}, \frac{1}{2}]$ derivative of the left hand side with respect to ω is smaller than zero for all admissible c, so $\omega = 1-\sqrt{1-c^2}$ provides maximum of the left hand side, and for such ω inequality holds.

Partial derivative of the left hand side with respect to ω is equal to

$$g(\omega, c) = \frac{\partial}{\partial \omega}\left(\left(1-\sqrt{c^2+\omega^2}\right)^2 + \left(1-\sqrt{c^2+(1-\omega)^2}\right)^2\right) =$$
$$= -2\frac{\left(1-\sqrt{c^2+\omega^2}\right)\omega}{\sqrt{c^2+\omega^2}} + 2\frac{\left(1-\sqrt{c^2+(1-\omega)^2}\right)(1-\omega)}{\sqrt{c^2+(1-\omega)^2}} =$$
$$= -2\left(\frac{1}{\sqrt{c^2+\omega^2}}-1\right)\omega + 2\left(\frac{1}{\sqrt{c^2+(1-\omega)^2}}-1\right)(1-\omega).$$

If $\omega = \frac{1}{2}$, then the partial derivative is zero. We show that for such $\omega < \frac{1}{2}$ that $c^2+\omega^2 < 1$, $c^2+(1-\omega)^2 < 1$, it holds that $g(\omega, c) < 0$. This fact means that the initial function decreases for $\omega \in [1-\sqrt{1-c^2}, \frac{1}{2}]$.

We start with maximization of $g(\omega, c)$ with respect to c. The function $g(\omega, c)$ attains maximum at the edge of admissibility region or in a local optimum with respect to c. To find local optima we search for c, such that the partial derivative $g(\omega, c)$ with respect to c is equal to zero:

$$\frac{c(1-\omega)}{((1-\omega)^2+c^2)^{\frac{3}{2}}} - \frac{c\omega}{(\omega^2+c^2)^{\frac{3}{2}}} = 0.$$

Consequently,

$$\frac{1-\omega}{\omega} = \frac{((1-\omega)^2+c^2)^{\frac{3}{2}}}{(\omega^2+c^2)^{\frac{3}{2}}}. \tag{13}$$

So,

$$c^2 = \omega^{\frac{2}{3}}(1-\omega)^{\frac{2}{3}}(\omega^{\frac{2}{3}}+(1-\omega)^{\frac{2}{3}}).$$

We show that this is a local maximum. Namely, we prove that the second partial derivative of $g(\omega, c)$ with respect to c is smaller than 0:

$$-\frac{(1-\omega)}{((1-\omega)^2+c^2)^{\frac{3}{2}}} + \frac{\omega}{(\omega^2+c^2)^{\frac{3}{2}}} + \frac{3c^2(1-\omega)}{((1-\omega)^2+c^2)^{\frac{5}{2}}} - \frac{3c^2\omega}{(\omega^2+c^2)^{\frac{5}{2}}} \leq 0.$$

Or:

$$\frac{\omega}{(\omega^2+c^2)^{\frac{5}{2}}}((\omega^2+c^2)-3c^2) - \frac{(1-\omega)}{((1-\omega)^2+c^2)^{\frac{5}{2}}}(((1-\omega)^2+c^2)-3c^2) \leq 0.$$

In a local optimum (13) holds, and we can rewrite inequality as:

$$\frac{(1-\omega)}{(\omega^2+c^2)((1-\omega)^2+c^2)^{\frac{3}{2}}}(\omega^2-2c^2) - \frac{(1-\omega)}{((1-\omega)^2+c^2)^{\frac{5}{2}}}((1-\omega)^2-2c^2) \leq 0.$$

Then,

$$\frac{(1-\omega)}{((1-\omega)^2+c^2)^{\frac{5}{2}}(\omega^2+c^2)}\left(((1-\omega)^2+c^2)(\omega^2-2c^2) - (\omega^2+c^2)((1-\omega)^2-2c^2)\right) \leq 0.$$

Due to constraints on values of ω this is equivalent to:

$$((1-\omega)^2+c^2)(\omega^2-2c^2) - (\omega^2+c^2)((1-\omega)^2-2c^2) \leq 0,$$

or

$$2c^2\omega^2 - c^2(1-\omega)^2 - 2c^2(1-\omega)^2 + c^2\omega^2 \leq 0.$$

This inequality holds, as $\omega \leq \frac{1}{2}$ and $(1-\omega)^2 \geq \omega^2$.

So, the extremum is a local minimum, and the function attains maximum values at the edges of the admissibility region. Namely, $c^2 = 1 - (1-\omega)^2$ or $c^2 = 0$ provides maximum values.

For such values of c the derivative is smaller than zero. Using $c^2 = 1 - (1-\omega)^2$ we get —

$$-2\left(\frac{1}{\sqrt{1-(1-\omega)^2+\omega^2}} - 1\right)\omega \leq 0.$$

In a similar way for $c^2 = 0$

$$-2(1-\omega) + 2\omega = 4\omega - 2 \leq 0.$$

Consequently, the target function decreases with respect to ω at $[1-\sqrt{1-c^2}, \frac{1}{2}]$, and $\omega = \frac{1}{2}$ provides a local minimum. So, the local maximum for left hand side is at $\omega = 1 - \sqrt{1-c^2}$. It is easy to see that in this case the left side of (12) is not larger than the right side. □

Let us now prove the main theorem.

Proof of Theorem 2. We provide upper and lower bounds for $R^H(L, \boldsymbol{\lambda})$ that are equal to $\frac{L}{2\pi^2}\max_{i\in\{1,\ldots,d\}}\left(\frac{h_i}{\lambda_i}\right)^2$. We start with a lower bound, and then continue with an upper bound.

We consider a functional

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

that is equal to the interpolation error $\sigma_H^2(\tilde{f}, F)$ for

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

such that $\hat{K}(\boldsymbol{\omega})$ is the Fourier transform of $K(\mathbf{x})$.

The functional is linear in $F(\boldsymbol{\omega})$ and quadratic in $\hat{K}(\boldsymbol{\omega})$, and we search for a saddle point of the functional $R^H(L, \boldsymbol{\lambda})$ such that:

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

It holds that (9)

$$\min_{\hat{K}} \Phi(F, \hat{K}) = \int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \frac{\sum_{\mathbf{x}\in D_{H-1}\setminus\{\mathbf{0}\}} F(\boldsymbol{\omega}+\mathbf{x})}{\sum_{\mathbf{x}\in D_{H-1}} F(\boldsymbol{\omega}+\mathbf{x})} d\boldsymbol{\omega}.$$

Let us consider a class of spectral densities $F_\varepsilon(\boldsymbol{\omega})$:

$$F_\varepsilon(\boldsymbol{\omega}) = \begin{cases} \frac{A_\varepsilon}{(2\varepsilon)^d}, & \exists \mathbf{s}\in U_h : \|\boldsymbol{\omega}-\mathbf{s}\|_\infty \leq \varepsilon, \\ 0, & \text{otherwise,} \end{cases}$$

here $U_h = \left\{\left(0,0,\ldots,\frac{1}{2h_j},\ldots,0\right), \left(0,0,\ldots,-\frac{1}{2h_j},\ldots,0\right)\right\}$, and an index j is such that

$$j = \arg\max_{i\in\{1,\ldots,d\}} \left(\frac{h_i}{\lambda_i}\right)^2.$$

Due to (3)

$$(2\pi)^2 \int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \sum_{i=1}^d \lambda_i^2 \omega_i^2 d\boldsymbol{\omega} \leq L,$$

and for $\varepsilon \to 0$

$$A_\varepsilon \to \frac{L}{2\pi^2}\left(\frac{h_j}{\lambda_j}\right)^2.$$

Really, for $\varepsilon \to 0$:

$$(2\pi)^2 \int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \sum_{i=1}^d \lambda_i^2 \omega_i^2 d\boldsymbol{\omega} \to 2(2\pi)^2 \frac{A_\varepsilon}{(2\varepsilon)^d}(2\varepsilon)^d \left(\frac{\lambda_j}{h_j}\right)^2 = 2A_\varepsilon \left(\frac{\pi\lambda_j}{h_j}\right)^2 = L.$$

Now for $\varepsilon \to 0$ it holds that

$$\min_{\hat{K}} \Phi(F_\varepsilon, \hat{K}) \to 2\frac{1}{2}\frac{A_\varepsilon}{(2\varepsilon)^d}(2\varepsilon)^d = A_\varepsilon = \frac{L}{2\pi^2}\left(\frac{h_j}{\lambda_j}\right)^2.$$

Consequently, we get a lower bound that equals $\frac{L}{2\pi^2}\left(\frac{h_j}{w_j}\right)^2$. Now we continue with a proof of the upper bound.

For any $\hat{K}(\boldsymbol{\omega})$ it holds that

$$\begin{aligned} R^H(L, \boldsymbol{\lambda}) &\leq \max_{F \in \mathcal{F}(L,\boldsymbol{\lambda})} \Phi(F, \hat{K}) \leq \\ &\leq L\left(\frac{1}{2\pi}\right)^2 \max_{\boldsymbol{\omega}} \left\{ \frac{1}{\sum_{i=1}^d \lambda_i^2 \omega_i^2} \left[(1 - \hat{K}(\boldsymbol{\omega}))^2 + \sum_{\mathbf{x} \in D_{H-1} \setminus \{\mathbf{0}\}} \hat{K}^2(\boldsymbol{\omega} + \mathbf{x}) \right] \right\}. \end{aligned}$$

Now let us consider

$$\hat{K}(\boldsymbol{\omega}) = \begin{cases} 1 - \|\boldsymbol{\omega}\|, \|\boldsymbol{\omega}\|^2 \leq 1, \\ 0, \text{otherwise}. \end{cases}$$

Now we prove that for such $\hat{K}(\boldsymbol{\omega})$ it holds that

$$\left[(1 - \hat{K}(\boldsymbol{\omega}))^2 + \sum_{\mathbf{x} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}} \hat{K}^2(\boldsymbol{\omega} + \mathbf{x}) \right] \leq 2\|\boldsymbol{\omega}\|^2. \tag{14}$$

It holds that $(1 - \hat{K}(\boldsymbol{\omega}))^2 \leq \|\boldsymbol{\omega}\|^2$. Now let us prove that

$$\sum_{\mathbf{x} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}} \hat{K}^2(\boldsymbol{\omega} + \mathbf{x}) \leq \|\boldsymbol{\omega}\|^2. \tag{15}$$

We use mathematical induction by d for $\boldsymbol{\omega}$ such that $\|\boldsymbol{\omega}\|_\infty < 1$. We prove that for $\|\boldsymbol{\omega}\|_\infty < 1$ and $c^2 \geq 0$:

$$\sum_{\substack{\mathbf{x} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}, \\ \|\boldsymbol{\omega}+\mathbf{x}\|^2 + c^2 \leq 1}} \left(1 - \sqrt{c^2 + \sum_{i=1}^d (\omega_i + x_i)^2} \right)^2 \leq \sum_{\substack{i \in \{1, \cdots, d\}, \\ c^2 + (1-\omega_i)^2 \leq 1}} \left(1 - \sqrt{c^2 + (1 - \omega_i)^2}\right)^2.$$

For $d = 1$ the induction statement is trivial, as the right hand side and the left hand side coincide. Suppose that for $(d-1)$ the induction statement holds. Now let us prove that the induction statement holds for d.

For $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_d)$ such that $\|\boldsymbol{\omega}\|_\infty < 1$, i-th component of the vector $\boldsymbol{\omega} + \mathbf{x}, \mathbf{x} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}$ such that $\|\boldsymbol{\omega} + \mathbf{x}\| \leq 1$ is either ω_i or $(1 - \omega_i)$. Consequently, all such $\boldsymbol{\omega} + \mathbf{x}$ has the form $\mathbf{s}\boldsymbol{\omega} + (\mathbf{1} - \mathbf{s})(\mathbf{1} - \boldsymbol{\omega})$, where $\mathbf{s}$ is a vector with all components of it belong to $\{0, 1\}$.

It holds that $(1 - \sqrt{c^2 + (1 - \omega_1)^2 + \omega_2^2 + \ldots + \omega_d^2})^2 \leq (1 - \sqrt{c^2 + (1 - \omega_1)^2})^2$, if $c^2 + (1 - \omega_1)^2 + \omega_2^2 + \ldots + \omega_d^2 \leq 1$.

Now let us consider all terms of the form $(1 - \sqrt{c^2 + (1-\omega_1)^2 + \ldots})^2$ for which there exists $j \neq 1$, such that $(1-\omega_j)^2$ is in the sum inside the squared root. Due to the induction statement sum of these terms is bounded by:

$$\sum_{\substack{i \in \{2, \ldots, d\}, \\ c^2 + (1-\omega_1)^2 + (1-\omega_i)^2 \leq 1}} (1 - \sqrt{c^2 + (1-\omega_1)^2 + (1-\omega_i)^2})^2.$$

In the same way we prove that the sum of terms $(1 - \sqrt{c^2 + \omega_1^2 + \ldots})^2$ with a term $(1-\omega_j)^2$ inside the root is upper bounded by:

$$\sum_{\substack{i \in \{2, \ldots, d\}, \\ c^2 + \omega_1^2 + (1-\omega_i)^2 \leq 1}} (1 - \sqrt{c^2 + \omega_1^2 + (1-\omega_i)^2})^2.$$

Using Lemma 1 for a pair of terms $(1 - \sqrt{c^2 + (1-\omega_1)^2 + (1-\omega_i)^2})^2 + (1 - \sqrt{c^2 + \omega_1^2 + (1-\omega_i)^2})^2$ we get:

$$(1 - \sqrt{c^2 + (1-\omega_1)^2 + (1-\omega_i)^2})^2 + (1 - \sqrt{c^2 + \omega_1^2 + (1-\omega_i)^2})^2 \leq$$
$$\leq (1 - \sqrt{c^2 + (1-\omega_i)^2})^2.$$

This upper bound also holds if there are no or only one term for i-th index. Consequently, the induction statement holds: the target sum is bounded by $\sum_{i=1}^d (1 - \sqrt{c^2 + (1-\omega_i)^2})^2$.

Using $c^2 = 0$ we get (15).

Now let us consider the case $\|\boldsymbol{\omega}\|_\infty \geq 1$. We look at the case $\boldsymbol{\omega} = \{\hat{\omega}_1 + 1, \omega_2, \ldots, \omega_d\}$, moreover $\|(\hat{\omega}_1, \omega_2, \ldots, \omega_d)\|_\infty < 1$, and $\hat{\omega}_1 \geq 0$, $\omega_i \geq 0, i = \overline{2, d}$. Then $\|\boldsymbol{\omega}\|^2 = 1 + 2\hat{\omega}_1 + \hat{\omega}_1^2 + \sum_{i=2}^d \omega_i^2$. For vector $(\hat{\omega}_1, \omega_2, \ldots, \omega_d)$ we have the induction statement above (15). For the initial vector $\boldsymbol{\omega}$ we have an additional term $\hat{K}^2((\hat{\omega}_1, \omega_2, \ldots, \omega_d))$ if Euclidian norm of such a vector is below or equal 1 — but this new term is smaller or equal to 1, as otherwise this term is not in the sum. So, the target estimate for $\boldsymbol{\omega}$ holds. Other cases for $\|\boldsymbol{\omega}\|_\infty > 1$ are similar. Consequently for all $\boldsymbol{\omega}$ the estimate (14) holds.

It holds that

$$\max_{\boldsymbol{\omega}} \frac{\sum_{i=1}^d \omega_i^2}{\sum_{i=1}^d \left(\frac{\lambda_i}{h_i}\right)^2 \omega_i^2} = \max_{i \in \{1, \ldots, d\}} \left(\frac{h_i}{\lambda_i}\right)^2.$$

Consequently, an upper bound for minimax interpolation error holds

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

The upper bound coincides with the lower bound. The theorem holds. □

7 PROOFS FOR SUBSECTION 3.2

Proof of Theorem 3. For convenience we redefine all points that belong to D_H as $D_H = \{\mathbf{x}_i\}$ and all points that belong to $D_{\frac{H}{m}}$ as $D_{\frac{H}{m}} = \{\tilde{\mathbf{x}}_j\}$. Then for Gaussian process regression the best unbiased estimator is linear in known values:

$$\tilde{u}(\mathbf{x}) = \sum_i k_i u(\mathbf{x}_i) + \sum_j \tilde{k}_j f(\tilde{\mathbf{x}}_j).$$

for some $k_i, \tilde{k}_j$. Our problem is then to find coefficients k_i, $\tilde{k}_j$ that minimize $\mathbb{E}(u(\mathbf{x}) - \tilde{u}(\mathbf{x}))^2$. Using independence of random processes $f(\mathbf{x})$ and $g(\mathbf{x})$ we get:

$$\mathbb{E}(u(\mathbf{x}) - \tilde{u}(\mathbf{x}))^2 = \mathbb{E}\left[\rho f(\mathbf{x}) + g(\mathbf{x}) - \sum_i k_i(\rho f(\mathbf{x}_i) + g(\mathbf{x}_i)) - \sum_j \tilde{k}_j f(\tilde{\mathbf{x}}_j)\right]^2 =$$
$$= \mathbb{E}\left[\rho f(\mathbf{x}) - \sum_i \rho k_i f(\mathbf{x}_i) - \sum_j \tilde{k}_j f(\tilde{\mathbf{x}}_j)\right]^2 + \mathbb{E}\left[g(\mathbf{x}) - \sum_i k_i g(\mathbf{x}_i)\right]^2.$$

For each i there exists an index j such $\mathbf{x}_i = \tilde{\mathbf{x}}_j$. Denote

$$\tilde{k}'_j = \begin{cases} \frac{1}{\rho}\tilde{k}_j, & \forall i, \tilde{\mathbf{x}}_j \neq \mathbf{x}_i, \\ \frac{1}{\rho}\tilde{k}_j + k_i, & \exists i, \tilde{\mathbf{x}}_j = \mathbf{x}_i. \end{cases}$$

There is a one-to-one correspondence between $(\{k_i\}, \{\tilde{k}_j\})$ and $(\{k_i\}, \{\tilde{k}'_j\})$, so minimization of $\mathbb{E}(u(\mathbf{x}) - \tilde{u}(\mathbf{x}))^2$ with respect to $k_i, \tilde{k}_j$ is equivalent to minimization of this function with respect to $k_i, \tilde{k}'_j$. Then

$$\mathbb{E}\left[\rho f(\mathbf{x}) - \sum_i k_i \rho f(\mathbf{x}_i) - \sum_j \tilde{k}_j f(\tilde{\mathbf{x}}_j)\right]^2 + \mathbb{E}\left[g(\mathbf{x}) - \sum_i k_i g(\mathbf{x}_i)\right]^2 =$$
$$= \rho^2 \mathbb{E}\left[f(\mathbf{x}) - \sum_j \tilde{k}'_j f(\tilde{\mathbf{x}}_j)\right]^2 + \mathbb{E}\left[g(\mathbf{x}) - \sum_i k_i g(\mathbf{x}_i)\right]^2.$$

For terms $\mathbb{E}\left[f(\mathbf{x}) - \sum_j \tilde{k}'_j f(\tilde{\mathbf{x}}_j)\right]^2$ and $\mathbb{E}\left[g(\mathbf{x}) - \sum_i k_i g(\mathbf{x}_i)\right]^2$ minimization problems are equivalent to that of single fidelity data — and the first term contains only coefficients $\tilde{k}'_j$, the second term contains only coefficients k_i.

For k_i and $\tilde{k}'_j$ that minimize interpolation error at point for the single fidelity scenario it holds that $\tilde{k}'_j = K_f(\mathbf{x} - \mathbf{x}_j)$, $k_i = K_g(\mathbf{x} - \mathbf{x}_i)$ for some symmetric kernels $K_f(\mathbf{x} - \mathbf{x}_j)$, $K_g(\mathbf{x} - \mathbf{x}_i)$.

Now we continue proof for $f(\mathbf{x})$ and $g(\mathbf{x})$ in a way similar to the single fidelity case. For $\mathbb{E}\left[g(\mathbf{x}) - \sum_i K_g(\mathbf{x} - \mathbf{x}_i) g(\mathbf{x}_i)\right]^2$ it holds that

$$\frac{1}{|H|}\int_{\substack{x_i \in [0, h_i], \\ i = \overline{1,d}}} \mathbb{E}\left[g(\mathbf{x}) - \sum_i K_g(\mathbf{x} - \mathbf{x}_i) g(\mathbf{x}_i)\right]^2 d\mathbf{x} =$$
$$= \int_{\mathbb{R}^d} G(\boldsymbol{\omega}) \left[\left[1 - \hat{K}_g(\boldsymbol{\omega})\right]^2 + \sum_{\mathbf{k} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}} \hat{K}_g^2\left(\boldsymbol{\omega} + H^{-1}\mathbf{k}\right)\right] d\boldsymbol{\omega}$$

In a similar way we get for the interval $[0, \frac{h_1}{m}] \cdot \ldots \cdot [0, \frac{h_d}{m}]$ for $\mathbb{E}\left[f(\mathbf{x}) - \sum_j K_f(\mathbf{x} - \tilde{\mathbf{x}}_j) f(\tilde{\mathbf{x}}_j)\right]^2$:

$$\frac{m^d}{|H|}\int_{\substack{x_i \in [0, \frac{h_i}{m}], \\ i = \overline{1,d}}} \mathbb{E}\left[f(\mathbf{x}) - \sum_j K_f(\mathbf{x} - \tilde{\mathbf{x}}_j) f(\tilde{\mathbf{x}}_j)\right]^2 d\mathbf{x} =$$
$$= \int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \left[\left[1 - \hat{K}_f(\boldsymbol{\omega})\right]^2 + \sum_{\mathbf{k} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}} \hat{K}_f^2\left(\boldsymbol{\omega} + H^{-1}\mathbf{k}\right)\right] d\boldsymbol{\omega}.$$

Consequently,

$$\frac{1}{|H|}\int_{\substack{x_i \in [0, h_i], \\ i = \overline{1,d}}} \mathbb{E}\left[f(\mathbf{x}) - \sum_j K_f(\mathbf{x} - \tilde{\mathbf{x}}_j) f(\tilde{\mathbf{x}}_j)\right]^2 d\mathbf{x} =$$
$$= \int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \left[\left[1 - \hat{K}_f(\boldsymbol{\omega})\right]^2 + \sum_{\mathbf{k} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}} \hat{K}_f^2\left(\boldsymbol{\omega} + mH^{-1}\mathbf{k}\right)\right] d\boldsymbol{\omega}.$$

So, the target interpolation error (5) has the form:

$$\sigma^2_{H,m}(\tilde{u}, F, G, \rho) = \int_{\mathbb{R}^d} G(\boldsymbol{\omega}) \left[\left[1 - \hat{K}_g(\boldsymbol{\omega})\right]^2 + \sum_{\mathbf{k} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}} \hat{K}_g^2 \left(\boldsymbol{\omega} + H^{-1}\mathbf{k}\right) \right] d\boldsymbol{\omega} +$$

$$+ \rho^2 \int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \left[\left[1 - \hat{K}_f(\boldsymbol{\omega})\right]^2 + \sum_{\mathbf{k} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}} \hat{K}_f^2 \left(\boldsymbol{\omega} + mH^{-1}\mathbf{k}\right) \right] d\boldsymbol{\omega}.$$

Finally,

$$\sigma^2_{H,m}(\tilde{u}, F, G, \rho) = \sigma^2_H(\tilde{g}, G) + \rho^2 \sigma^2_{\frac{H}{m}}(\tilde{f}, F).$$

□

8 PROOFS FOR SECTION 4

In this section we provide a proof of Theorem 5.

Proof of Theorem 5. Minimax interpolation error has the form:

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

Denote $\delta = (m^*)^d$. Then we need to minimize with respect to a the following expression

$$\frac{L_g}{2} (c + \delta)^{\frac{2}{d}} + \rho^2 \frac{L_f}{2} \left(\frac{c + \delta}{\delta} \right)^{\frac{2}{d}}.$$

Partial derivative with respect to δ should equal 0:

$$\frac{L_g}{2} (c + \delta)^{\frac{2}{d} - 1} \frac{2}{d} + \rho^2 \frac{L_f}{2} \left(\frac{c + \delta}{\delta} \right)^{\frac{2}{d} - 1} \frac{2}{d} \frac{-c}{\delta^2} = 0.$$

Consequently

$$\frac{L_g}{2} + \rho^2 \frac{L_f}{2} \delta^{1 - \frac{2}{d}} \frac{-c}{\delta^2} = 0.$$

So

$$L_g = L_f \frac{\rho^2 c}{\delta^{1 + \frac{2}{d}}}.$$

Finally,

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

And

$$m^* = \sqrt[d+2]{c\rho^2 \frac{L_f}{L_g}}.$$

□

9 REAL DATA PROBLEMS

We consider the following real data problems. The first three of them (**Euler**, **Airfoil** [Bernstein et al., 2011], **MachAngle**) are devoted to calculation of lift and drag coefficients of an airfoil depending on flight conditions and airfoil geometry. To evaluate these outputs we use different solvers as high and low fidelity data sources. The next two problems (**Press** [Burnaev and Zaytsev, 2015], **Disk** [Zaytsev, 2016]) investigates dependence of maximum stress and maximum displacement on geometry of these tools. As there are three fidelities for **Press** problem we use in each experiment only two of them. The last two problems ([Kandasamy et al., 2016], **SVM**, **Supernova**) consider modeling of a dependence of goodness of fit on model parameters.

Euler. Eleven input variables parametrize geometry of an airfoil.

Airfoil. The geometry of an airfoil and the flight regime (the speed and the angle of attack) are described by 52 input variables. We employ a dimension reduction procedure similar to the PCA, and model the dependence on six input factors [Bernstein et al., 2011].

MachAngle. Two input variables are the Mach number and the angle of attack for a specific airfoil. Low fidelity solver provides almost linear dependence.

Press. We model the maximum stress and the maximum displacement for a C-shaped press [Burnaev and Zaytsev, 2015]. Six input variables describe the geometry of the press, and the fidelity of output depends on a mesh size. We generate three different data samples that correspond to high, moderate and low fidelity outputs. We refer to the case when we model the high fidelity output by $u(\mathbf{x})$ and the moderate fidelity output by $f(\mathbf{x})$ as `Press12`, and the case when we model the high fidelity output by $u(\mathbf{x})$ and the low fidelity output by $f(\mathbf{x})$ as `Press13`.

Disk. We model the maximum stress and the maximum displacement of a rotating disk in an engine [Zaytsev, 2016]. Six input variables describe the geometry of the disk. We use two different solvers to obtain high and low fidelity values.

SVM. We model the dependence of the SVM classifier accuracy from the **sklearn** [Pedregosa et al., 2011] on the kernel bandwidth and the margin coefficient for the "MAGIC Gamma Telescope" dataset [Kandasamy et al., 2016]. We have two input variables. As a measure of accuracy we use the area under the ROC curve as suggested by the authors of the dataset [Bock et al., 2004]. To generate the low fidelity dataset we estimate the accuracy of the classifier constructed using 500 training points, and to generate the high fidelity dataset we estimate the accuracy of the classifier constructed using 2000 training points.

Supernova. We model the dependecy of the likelihood of the supernova redshift data on the three fundamental physical constants, similarly to [Kandasamy et al., 2016, Davis et al., 2007]. To get a variable fidelity data we vary the grid size for a one-dimensional integration: we generate the low fidelity data using the grid of size 3 and the high fidelity data using the grid of size 1000. We note that if the size of the grid is greater than 3, then the high and low fidelity functions become indistinguishable.

References

[Abramowitz and Stegun, 1964] Abramowitz, M. and Stegun, I. A. (1964). *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, volume 55. Courier Corporation.

[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.

[Bock et al., 2004] Bock, R., Chilingarian, A., Gaug, M., Hakl, F., Hengstebeck, T., Jiřina, M., Klaschka, J., Kotrč, E., Savickỳ, P., Towers, S., et al. (2004). Methods for multidimensional event classification: a case study using images from a Cherenkov gamma-ray telescope. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 516(2):511–528.

[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.

[Chang et al., 2011] Chang, S.-H., Cosman, P., and Milstein, L. (2011). Chernoff-type bounds for the Gaussian error function. *Communications, IEEE Transactions on*, 59(11):2939–2944.

[Davis et al., 2007] Davis, T., Mörtsell, E., Sollerman, J., Becker, A., Blondin, S., Challis, P., Clocchiatti, A., Filippenko, A., Foley, R., Garnavich, P., et al. (2007). Scrutinizing exotic cosmological models using essence supernova data combined with other cosmological probes. *The Astrophysical Journal*, 666(2):716.

[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.

[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.

[Pedregosa et al., 2011] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

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

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

[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.