JMLR: Workshop and Conference Proceedings 1:1–13, 2018 COPA 2018

# Interpolation error of Gaussian process regression for misspecified case

**A. Zaytsev** A.ZAYTSEV@SKOLTECH.RU
**E. Romanenkova** SHULGINA@PHYSTECH.EDU
**D. Ermilov** DMITRII.ERMILOV@SKOLTECH.RU
*Skoltech, IITP RAS, MIPT*

**Editor:** Editor's name

## Abstract

An interpolation error is an integral of the squared error of a regression model over a domain of interest. We consider the interpolation error for the case of misspecified Gaussian process regression: a used covariance function differs from a true one. We derive the interpolation error for a grid design of experiments for an arbitrary covariance function. Then we consider particular types of covariance functions from theoretical and practical points of view. For $Matern_{\frac{1}{2}}$ covariance function poor estimation of parameters only slightly affects the quality of interpolation. For the most common covariance functions including $Matern_{\frac{3}{2}}$ and squared exponential covariance functions poor choose of parameters of covariance functions leads to a bad quality of interpolation.

**Keywords:** Gaussian process regression, interpolation error estimation, model misspecification

## 1. Introduction

Gaussian process regression or kriging is widely used for construction of regression models? see Rasmussen and Williams (2006); Burnaev et al. (2016); Cressie (2015). The main assumption of these approaches is that the target function is a realization of Gaussian process model with a given spectral density (or equivalently a covariance function) and mean functions.

In machine learning it is of great importance to get a measure of the quality of a regression model. Popular choice in literature is an interpolation error, see Golubev and Krymova (2013); Le Gratiet and Garnier (2015): an expected squared error of interpolation integrated over a domain of interest for a given approach of a regression model construction.

There are a number of problem statements relevant to this general problem. Classical approaches imply that the true model is known and coincides with the one used for the construction of a regression model Stein (2012). Modern approaches more often consider a minimax problem statement Zaytsev and Burnaev (2017) or a misspecified problem statement Vaart and Zanten (2011). In the minimax problem statement, we assume that the true model belongs to a certain class of models and try to find the interpolation error in the worst case Golubev and Krymova (2013). In the misspecified problem statement, we specify how a used model differs from the true model Panov (2016).

Let us elaborate in more details the misspecified problem statement for Gaussian process regression. For real problems, one doesn't know the true Gaussian process regression model, while the usual assumption is that the spectral density (a Fourier transform of the covariance function) belongs to a given parametric family and the mean value is zero. After selection of a parametric family one estimates parameters of a spectral density using approaches similar to the maximum likelihood approach or Bayesian approach Zaytsev et al. (2014) Quality of estimation of parameters varies Zaytsev et al. (2014); Bachoc (2018). Moreover, smoothness of the target function is often unknown. So it is hard to select a parametric family of spectral densities. Thus, bad estimates of spectral density and wrong choice of a parametric family lead to a difference between the true regression model and the used regression model.

Our goal is to obtain the exact expression for interpolation error in a misspecification case. The assumptions are similar to those used in the state of the art: Gaussian process is stationary, the design of experiments is an infinite grid with a given step along each dimension. The grid designs of experiments are often used due to their low computational complexities Belyaev et al. (2015). Moreover, numerical experiments show that these assumptions don't significantly affect the results Zaytsev and Burnaev (2017). Using obtained expression as a tool we are able to consider widely used setups for Gaussian process regression taking into account possible model misspecification. We consider the squared exponential function and the Matern covariance functions with $\nu = \frac{1}{2}$ Minasny and McBratney (2005).

The article has the following sections:

- Section 2 describes the prior results in this area in more details;
- Section 3 describes results for usage of known covariance function and minimax case;
- Section 4 describes results for the case when the true covariance function differs from the used one and examines in more details the case of model misspecification for the Matern covariance function;
- Section 5 contains results of numerical experiments;

## 2. Related work

Classical approaches imply that the true model is known and coincides with the one used for the construction of a regression model. The first results in this area go back to Kolmogorov (1941) and Wiener (1949). A.Kolmogorov and N.Wiener simultaneously obtained mean squared errors at a point in an interpolation and an extrapolation problem statements with all training points lying on a grid. An article Le Gratiet and Garnier (2015) considered the integrated mean squared interpolation error for a Gaussian process with noise if the sample size tends to infinity.

Modern approaches more often consider a minimax problem statement. An article Golubev and Krymova (2013) considered the minimax interpolation error for a Sobolev class of Gaussian processes for a segment if the training sample is an infinite grid. More recent article Zaytsev and Burnaev (2017) considered a multivariate scenario while considering

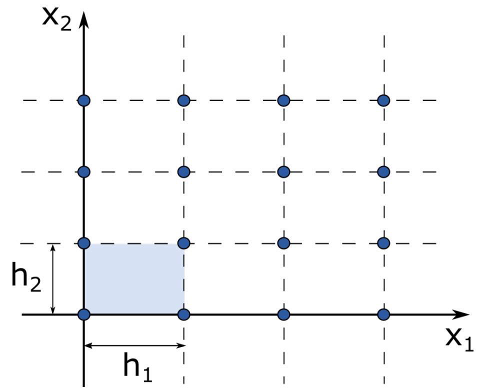



Figure 1: A design of experiments $D_H$ for $d = 2$.

Gaussian processes with an upper bound only for a sum of squares of the first partial derivatives of Gaussian process realization.

Another branch of modern results considers a misspecified problem statement. For a review of results for a squared error at a single point see book Stein (2012) More general papers van der Vaart and van Zanten (2008); Vaart and Zanten (2011) consider the case of mean squared error for an area, while their results are not directly applicable in practice-related problems due to complex assumptions. Note also that these articles as well as Suzuki (2012); Castillo et al. (2008) provide the upper bound. An article Bachoc (2013) considered the empirical comparison of the interpolation error for cross validation and maximum likelihood estimates, while theoretical properties of these approaches are investigated in more details in Bachoc (2018). There the focus is not on the interpolation error itself, but on the quality of parameter estimation.

## 3. Interpolation error and minimax interpolation error

Let us introduce interpolation for the case with no misspecification and the minimax case. All results in this section are provided in a way similar to Zaytsev and Burnaev (2017).

For $\mathbb{R}^d$ there is a stationary Gaussian process $f(\cdot)$ with the covariance function $R(\mathbf{x})$. The spectral density Stein (2012) is defined as

$$F(\boldsymbol{\omega}) = \frac{1}{2\pi} \int_{\mathbb{R}^d} e^{-\mathrm{i}\boldsymbol{\omega}\mathbf{x}} R(\mathbf{x}) d\mathbf{x}.$$

We observe the random process at the infinite grid $D_H = \{x_{\mathbf{k}} = H\mathbf{k}, \mathbf{k} \in \mathbb{Z}^d\}$. $H$ is a diagonal matrix with elements at the diagonal $\mathrm{diag}\{h_1, \ldots, h_d\}$. An example of such two-dimensional design of experiments is given at Figure 1.

We investigate interpolation error of $f(\mathbf{x})$ using the best regression model $\tilde{f}(\mathbf{x})$. In Gaussian case this model depends linearly on observations

$$\tilde{f}(\mathbf{x}) = H \sum_{\mathbf{k} \in \mathbb{Z}^d} K(\mathbf{x} - \mathbf{x}_{\mathbf{k}}) f(\mathbf{x}_k),$$

where $K(\mathbf{x})$ is a kernel function obtained as a solution of Kolmogorov-Wiener–Hopf equations.

For a set $\Omega_H = \prod_{i=1}^{d}[0, h_i]$ we are interested in evaluation of the integral of the expectation of squared differences between true value of a random process and its interpolation:

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

The following theorem holds:

**Theorem 1** *For a Gaussian random process $f(\mathbf{x})$ with a spectral density $F(\boldsymbol{\omega})$, observed at $D_H$ the interpolation error for the best regression model has the form:*

$$\sigma^2(\tilde{f}, F) = \int_{\Omega_H} F(\boldsymbol{\omega}) \left[\left(1 - \hat{K}(\boldsymbol{\omega})\right)^2 + \sum_{\mathbf{k} \neq 0} \hat{K}^2\left(\boldsymbol{\omega} + H^{-1}\mathbf{k}\right)\right] d\boldsymbol{\omega},$$

*where $\hat{K}(\boldsymbol{\omega})$ is a Fourier transform of $K(\boldsymbol{\omega})$. Moreover, $\hat{K}(\boldsymbol{\omega})$ has the form*

$$\hat{K}(\boldsymbol{\omega}) = F(\boldsymbol{\omega}) / \sum_{\mathbf{k}} F\left(\boldsymbol{\omega} + H^{-1}\mathbf{k}\right).$$

Often the true spectral density is unknown. So, we are interested in the minimax interpolation error:

$$R^H(L) = \inf_{\tilde{f}} \sup_{F \in \mathcal{F}(L)} \sigma^2(\tilde{f}, F), \qquad (1)$$

where $\mathcal{F}(L)$ defines a set of spectral densities that correspond to smooth enough Gaussian processes:

$$\sup_{F \in \mathcal{F}(L)} \mathbb{E}\|f^{(1)}(\mathbf{x})\|^2 \leq L,$$

$f^{(1)}(\mathbf{x})$ is a vector of first partial derivatives of Gaussian process with a spectral density $F(\boldsymbol{\omega})$.

The following theorem holds:

**Theorem 2** *The minimax interpolation error $R^H(L)$ from (1) has the form:*

$$R^H(L) = \frac{L}{2\pi^2} \max_{i=\overline{1,d}} h_i^2.$$

Given the theorems above we can get the interpolation error for certain covariance functions: exponential and squared exponential covariance functions.

**Corollary 3** *For Gaussian process at $\mathbb{R}$ with the exponential spectral density of the form $F_\theta(\omega) = \frac{\theta}{\theta^2 + \omega^2}$ the interpolation error (5) for the best interpolation has the form:*

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

*for $\theta h \to 0$.*

**Corollary 4** *For Gaussian process at $\mathbb{R}$ with the squared exponential spectral density of the form $F_\theta(\omega) = \frac{1}{\sqrt{\theta}} \exp\left(-\frac{\omega^2}{2\theta}\right)$ the interpolation error (5) for the best interpolation has the form:*

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

*for $\theta h^2 \to 0$.*

So, the minimax error decreases as $h^2$ for $h = \max_{i \in \{1,d\}} h_i$, while for some covariance functions it can decrease exponential with respect to $h$, or decrease linearly with $h$ for a non-smooth Gaussian process.

## 4. Interpolation error for misspecified case

In practice, we use a model of the Gaussian process. This model can be different from the true model. Let us consider a Gaussian process with the true spectral density $F(\boldsymbol{\omega})$, while for estimation we use a Gaussian process with the spectral density $F_\theta(\boldsymbol{\omega})$. The problem is to estimate the interpolation error for misspecified spectral density used for computation of the final approximation.

We again consider the infinite grid design of experiments $D_H$ and sample of values $\{f(\mathbf{x_k})\}$ at $D_H$ of a realization of a Gaussian process with the spectral density $F(\boldsymbol{\omega})$.

The best interpolation given the assumptions has the form:

$$\tilde{f}_\theta(\mathbf{x}) = H \sum_{\mathbf{k} \in \mathbb{Z}^d} K_\theta(\mathbf{x} - \mathbf{x_k}) f(\mathbf{x_k}).$$

We obtain the kernel $K(\cdot)$ by minimization of the mean squared error assuming that the true spectral density is $F(\boldsymbol{\omega})$. We obtain the kernel $K_\theta(\cdot)$ in a similar way, but using the true spectral density $F_\theta(\boldsymbol{\omega})$.

Our goal is to estimate the interpolation error

$$\sigma_H^2(\tilde{f}_\theta, F) = \frac{1}{\Omega_H} \int_{\Omega_H} \mathbb{E}\left[\tilde{f}_\theta(\mathbf{x}) - f(\mathbf{x})\right]^2 d\mathbf{x}.$$

**Theorem 5** *The interpolation error for the true spectral density $F(\boldsymbol{\omega})$, if we used the spectral density $F_\theta(\boldsymbol{\omega})$ for construction of the regression model given observations at $H^{-1}\mathbf{k}, \mathbf{k} \in \mathbb{Z}^d$ has the form:*

$$\sigma_H^2(\tilde{f}_\theta, F) = \int_{\Omega_H} F(\boldsymbol{\omega}) \left[\left(1 - \hat{K}_\theta(\boldsymbol{\omega})\right)^2 + \sum_{\mathbf{k} \neq 0} \hat{K}_\theta^2\left(\boldsymbol{\omega} + H^{-1}\mathbf{k}\right)\right] d\boldsymbol{\omega}.$$

So, given spectral densities $F(\boldsymbol{\omega})$ and $F_\theta(\boldsymbol{\omega})$ we can get the target interpolation error by analytical integration of (5) or numerical estimation.

Note, that we can rewrite the result of Theorem 5 in the form:

$$\sigma_H^2(\tilde{f}_\theta, F) = \int_{\Omega_H} F(\boldsymbol{\omega}) \left\{ \frac{\sum_{\mathbf{k} \neq 0} F_\theta(\boldsymbol{\omega} + H^{-1}\mathbf{k})}{\sum_{\mathbf{s}} F_\theta(\boldsymbol{\omega} + H^{-1}\mathbf{s})} \right\} d\boldsymbol{\omega}$$

As a set of coefficients $K(\mathbf{x} - \mathbf{x_k})$ minimizes the interpolation error, it holds that $\sigma_H^2(\tilde{f}_\theta, F) \geq \sigma_H^2(\tilde{f}, F)$. Now we are ready for analysis of difference of the interpolation error in the cases of misspecified and correctly specified models.

### 4.1. Interpolation error for misspecified Matern spectral density $\nu = \frac{1}{2}$

We consider the interpolation error for the misspecified case for Matern with $\nu = \frac{1}{2}$ spectral density. For $\mathbb{R}$ and a stationary Gaussian process with Matern $\nu = \frac{1}{2}$ covariance function $R_\theta(x)$:

$$R_\theta(x) = \sqrt{\frac{\pi}{2}} \exp\left(-\theta \left\|x\right\|\right). \tag{2}$$

An alternative name for this covariance function is the exponential covariance function GPy (since 2012). The spectral density that corresponds to this covariance function is the following:

$$F_\theta(\omega) = \frac{\theta}{\theta^2 + \omega^2}.$$

To construct an interpolation we use a misspecified spectral density $F_{\theta'}, \theta' \neq \theta$.

**Corollary 6** *We observe a realization of Gaussian process at $D_h \subset \mathbb{R}$ with the true exponential spectral density of the form $F_\theta(\omega) = \frac{\theta}{\theta^2+\omega^2}$. Then the interpolation error (5) for $\tilde{f}_{\theta'}$ constructed with the assumption that the true spectral density is $F_{\theta'}(\omega)$ has the form:*

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

So, for small $h$ the interpolation error doesn't depend on the coefficient $\theta'$. It is obvious that generally, the misspecified case lacks this nice property.

## 5. Computational experiments

We obtained theoretical results in previous sections under the assumption that the realization of a Gaussian process is known at an infinite grid. It is impossible to fulfill such assumption in practice. While we expect that results will be the same for a large enough finite sample and an infinite grid of points, we should validate theoretical for a finite sample.

For this purpose a realization of a Gaussian process with Matern covariance function from (2) is taken as objective function. It is a special case of exponential covariance function, so, results of section 2 are applicable to it.

### 5.1. Workflow of computational experiments

In this subsection, we provide technical details on computational experiments. For experiments we used Gaussian process regression realization from GPy (since 2012) library. We provide the code used to generate results in the article at A. Zaytsev, E. Romanenkova, D. Ermilov (since 2017) github page. For the sake of clarity and faster convergence of the empirical interpolation error to the true one, we consider one-dimensional grids of points, while our theoretical results are valid for the multivariate case.

There are three steps in computational experiment dedicated to obtaining of the interpolation error: create a realization of a Gaussian process; use a regression model with an alternative covariance function on the base of a training sample; estimate the interpolation error given the constructed regression model and a test sample.

We create realizations of a Gaussian process using the following steps:

1. Generate a grid of points $X$ of size $n$ from interval $[0, 1]$. The step of the grid $h$ is inversely proportional to the current sample size $n$.
2. Select the parameter of covariance function $\theta$.
3. Evaluate the sample covariance matrix $K$ for points $X$. We use Matern covariance function with parameter $\theta$ as a covariance function and white noise with variance $10^{-8}$.
4. Evaluate the Cholesky decomposition $L$ of the obtained covariance matrix $K$.
5. Generate a vector of i.i.d. random variables $\mathbf{y}_0$ from the standard normal distribution of size $n$.
6. Obtain multivariate normal distribution by multiplying the Cholesky decomposition $L$ by $\mathbf{y}_0$ at the previous step.

As the results of this procedure we get $\mathbf{y} = L\mathbf{y}_0 \sim \mathcal{N}(\mathbf{0}, K)$. Few examples of generated realizations are at Figures 2.

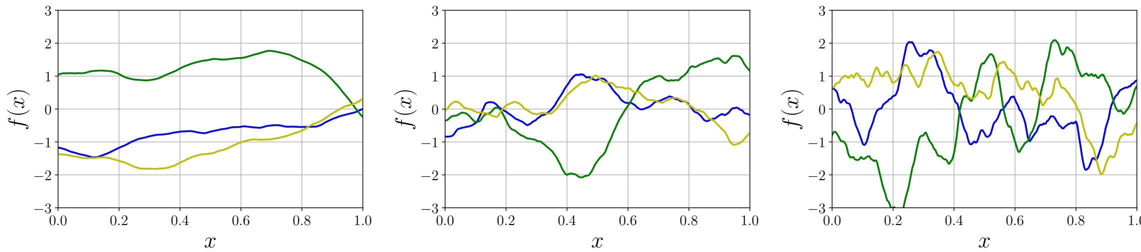
Next algorithm is used for the construction of the regression model and evaluation of the interpolation error of this model:

1. Select the parameter $\theta'$ — an assumption about the true parameter of the covariance function.
2. Specify a regression model on the grid given the parameter $\theta'$.
3. Estimate the interpolation error for the selected sample size as $\sum_{i=1}^{n_{\text{test}}} (\hat{y}_i - y_i)^2$, where $\hat{y}_i$ and $y_i$ are respectively the predicted and the true value at a test point $\mathbf{x}_i$ using a test sample with a more dense grid.

### 5.2. Simulations

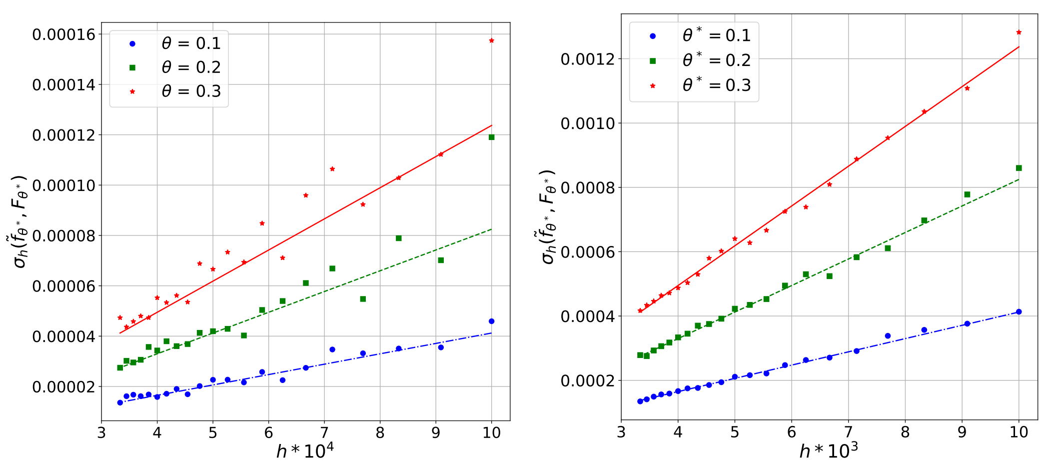
We start with the following setup: $\theta = \theta'$ for values $\theta = [0.1, 0.2, 0.3]$. After that the experiment we average the result of 10 realizations. The obtained interpolation errors fit the straight line even better. At Figure 3 we provide results for a single realization and 10 realizations: the obtained experimental errors agrees well with the theoretical results. So, the assumption of the infinite grid does not affect the interpolation error: the theoretical and the practical results are consistent.

We continue with the misspecified problem: for a fixed $\theta = 0.1$, three different $\theta' = [0.1, 1, 10]$ are used during model construction. For each of the values at each sample size, we run the "basic algorithm" 20 times and average the results. We see that results are




$(a)$ Smoothness 10 $(b)$ Smoothness 100 $(c)$ Smoothness 500

Figure 2: Realizations of Gaussian process with the squared exponential covariance function for different smoothness values




$(a)$ A single run of the algorithm $(b)$ Averaging of 10 runs

Figure 3: The interpolation errors obtained in the experiment. The solid lines indicate theoretical values: they are close to obtained in experiments even for a single run.

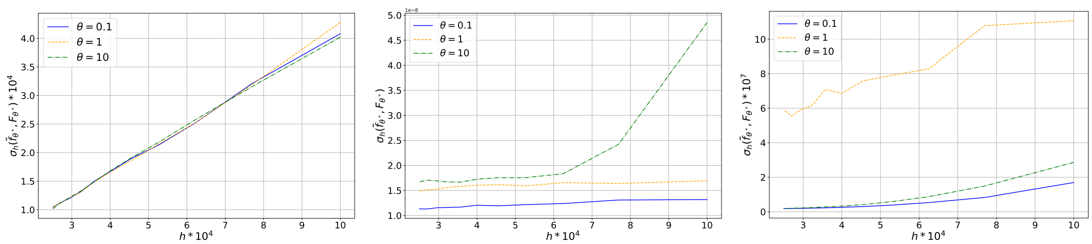
similar for different used values of $h$. While the obtained results slightly differ, Theorem 1 gives almost perfect approximation in this case.

For the exponential, $Matern_{3/2}$ and RBF kernels we examine the interpolation error for the misspecified model, that lies in the same parametric class of models. The obtained interpolation errors are at Figure 4. We also run the Wilcoxon difference test Wilcoxon (1945) to test if the results are different. For the exponential kernel we get $p-$value $> 0.92$, while for $Matern_{3/2}$ and the squared exponential kernel $p-$value $< 10^{-2}$. So, for the exponential kernel we can't reject the hypothesis that the interpolation errors are the same, no matter what value of the covariance function parameter is used, while for two other covariance functions results suggest that the hypothesis that the interpolation errors are the same seems to be wrong. Therefore, for a non-exponential kernel the interpolation error depends on the value of the parameter used for the construction of the regression model.

In the following experiment, we investigate the interpolation error for the case when the wrong parametric class of functions is used. In particular, the parameter of the model is

| h | $\theta = 0.1$ $\theta' = 0.1$ | $\theta = 0.1$ $\theta' = 1$ | $\theta = 0.1$ $\theta' = 10$ |
|---|---|---|---|
| 0.01 | 0.425 | 0.413 | 0.424 |
| 0.004 | 0.166 | 0.168 | 0.162 |
| 0.0025 | 0.102 | 0.105 | 0.104 |

Table 1: Obtained values of the interpolation error $\sigma_h^2(\hat{f}_\theta, F_\theta) \cdot 10^3$ for misspecified case averaged over 20 realizations.




($a$) The exponential covariance function ($b$) $Matern_{3/2}$ covariance function ($c$) The squared exponential covariance function

Figure 4: The interpolation errors obtained in the experiment for misspecified values of parameters for different classes of covariance functions averaged over 20 realizations. The true parameter $\theta$ is 1.

$\theta = 1$; when the model $Matern_{3/2}$ is true, models with covariance functions $Matern_{3/2}$, $Matern_{5/2}$ and exponential give the interpolation errors provided at Figure 5. We see that the error estimation derived in Corollary 3 is not applicable in the case of using different classes of the true and used function.

## 6. Conclusions

This article presents the interpolation error for misspecified regression model of Gaussian process regression. The obtained result can be used for analysis of the effect of the model misspecification on the quality of obtained regression model. For example, for Matern covariance function the interpolation error doesn't significantly depend on a used value of the parameter. This effect holds for numerical experiments for non-grid finite training samples. However, for other covariance functions there is a clear dependence on quality of estimates of parameters.

## Acknowledgments

The results presented in Section 4 were supported solely by the Russian Science Foundation grant (project 14-50-00150).

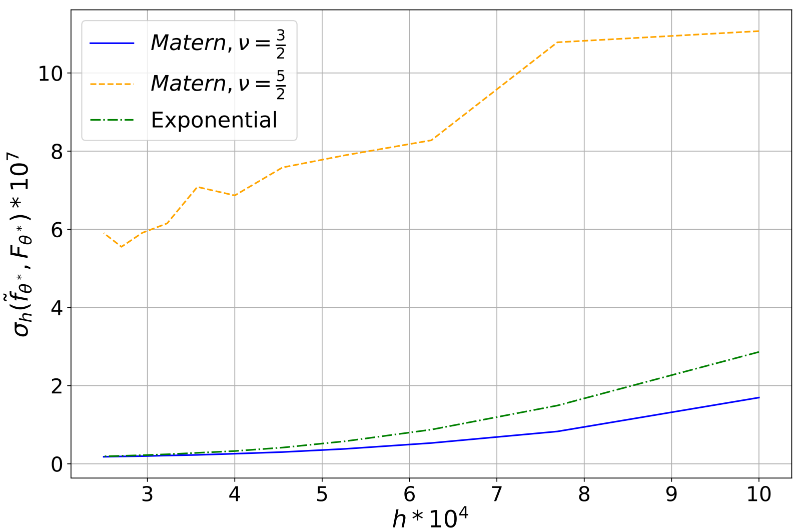



Figure 5: The interpolation errors obtained in the experiment for misspecified class of covariance functions averaged over 20 realizations.

## References


A. Zaytsev, E. Romanenkova, D. Ermilov. Interpolation errors for Gaussian process regression. https://github.com/likzet/gp_interpolation_error, since 2017.

F. Bachoc. Cross validation and maximum likelihood estimations of hyper-parameters of gaussian processes with model misspecification. *Computational Statistics & Data Analysis*, 66:55–69, 2013.

F. Bachoc. Asymptotic analysis of covariance parameter estimation for gaussian processes in the misspecified case. *Bernoulli*, 24(2):1531–1575, 2018.

M. Belyaev, E. Burnaev, and Y. Kapushev. Gaussian process regression for structured data sets. In *International Symposium on Statistical Learning and Data Sciences*, pages 106–115. Springer, 2015.

E.V. Burnaev, M.E. Panov, and A.A. Zaytsev. Regression on the basis of nonstationary Gaussian processes with Bayesian regularization. *Journal of communications technology and electronics*, 61(6):661–671, 2016.

I. Castillo et al. Lower bounds for posterior rates with Gaussian process priors. *Electronic Journal of Statistics*, 2:1281–1299, 2008.

N. Cressie. *Statistics for spatial data*. John Wiley & Sons, 2015.

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

GPy. GPy: A Gaussian process framework in python. http://github.com/SheffieldML/GPy, since 2012.

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

L. Le Gratiet and J. Garnier. Asymptotic analysis of the learning curve for gaussian process regression. *Machine Learning*, 98(3):407–433, 2015.

B. Minasny and A. McBratney. The Matérn function as a general model for soil variograms. *Geoderma*, 128(3):192–207, 2005.

M.E. Panov. Nonasymptotic approach to Bayesian semiparametric inference. In *Doklady Mathematics*, volume 93, pages 155–158. Springer, 2016.

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

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

T. Suzuki. PAC-Bayesian bound for Gaussian process regression and multiple kernel additive model. In *COLT*, pages 8–1, 2012.

A. van der Vaart and H. van Zanten. Information rates of nonparametric gaussian process methods. *Journal of Machine Learning Research*, 12(Jun):2095–2119, 2011.

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

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

F. Wilcoxon. Individual comparisons by ranking methods. *Biometrics bulletin*, 1(6):80–83, 1945.

A. Zaytsev and E. Burnaev. Minimax approach to variable fidelity data interpolation. In *Artificial Intelligence and Statistics*, pages 652–661, 2017.

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

## Appendix A. Proofs of the presented results

The proof of Theorem 5:

**Proof** The difference between this problem and the problem given in Theorem 1 is on different set of coefficient $K_\theta(\mathbf{x} - \mathbf{x_k})$ identified by the spectral density. Consequently, we are able to get the proof of Theorem 1 using the results given in Zaytsev and Burnaev (2017).

It is easy to see that

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

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

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

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

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

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

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

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

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

To get $\hat{K}_\theta(\boldsymbol{\omega})$ that minimizes the interpolation error we rewrite the equation above using $F_\theta$ instead of $F$:

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

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

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

Then

$$\sigma_H^2(\tilde{f}_\theta, F) = \int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \frac{\sum_{\mathbf{k} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}} \hat{F}_\theta(\boldsymbol{\omega} + H^{-1}\mathbf{k})}{\sum_{\mathbf{k} \in \mathbb{Z}^d} \hat{F}_\theta(\boldsymbol{\omega} + H^{-1}\mathbf{k})} d\boldsymbol{\omega}. \tag{3}$$

■

The proof of Corollary 6:

**Proof** Our goal is to evaluate

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

It holds that

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

Then

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

For three integrals presented above it holds that:

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

Moreover,

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

Finally,

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

Consequently

$$\sigma_h^2\left(\tilde{f}_{\theta'}, F_\theta\right) = \pi - \frac{\pi}{2\pi(\theta + \theta')h \coth(\pi\theta' h)} +$$
$$+ \frac{\pi}{2\pi(\theta + \theta')h \coth(\pi\theta' h)} \left(1 - \frac{\theta \exp(-2\pi h\theta') - \theta' \exp(-2\pi h\theta)}{\theta - \theta'}\right) \frac{\coth^2(\pi\theta' h) - 1}{\pi h \coth(\pi\theta' h)}.$$

We are interesting in case $h \to 0$. In this case, we can use Taylor decomposition to evaluate the final result. We get it in the following form:

$$\sigma_h^2\left(\tilde{f}_{\theta'}, F_\theta\right) = \frac{2\pi^2}{3} \theta h + O((\theta h)^2) + O((\theta' h)^2).$$

∎