BRNO UNIVERSITY OF TECHNOLOGY

Faculty of Electrical Engineering
and Communication

MASTER'S THESIS

Brno, 2021

Bc. Martin Ptáček

BRNO UNIVERSITY OF TECHNOLOGY

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

FAKULTA ELEKTROTECHNIKY
A KOMUNIKAČNÍCH TECHNOLOGIÍ

DEPARTMENT OF RADIO ELECTRONICS

ÚSTAV RADIOELEKTRONIKY

SPATIAL FUNCTION ESTIMATION WITH UNCERTAIN SENSOR LOCATIONS

SPATIAL FUNCTION ESTIMATION WITH UNCERTAIN SENSOR LOCATIONS

MASTER'S THESIS

DIPLOMOVÁ PRÁCE

AUTHOR
AUTOR PRÁCE

Bc. Martin Ptáček

SUPERVISOR
VEDOUCÍ PRÁCE

prof. Dr. Ing. Franz Hlawatsch
VIENNA UNIVERSITY OF TECHNOLOGY

doc. RNDr. Jitka Poměnková, Ph.D.
BRNO UNIVERITY OF TECHNOLOGY

BRNO 2021

Master's Thesis

Master's study program **Telecommunications**

Department of Radio Electronics

Student: Bc. Martin Ptáček **ID:** 182742

Year of study: 2 **Academic year:** 2020/21

TITLE OF THESIS:

Spatial Function Estimation with Uncertain Sensor Locations

INSTRUCTION:

The goal of the diploma thesis is to investigate the possibilities of using Gaussian Process Regression (GPR) for estimation of a spatial function while operating on uncertain sensor positions. This investigation shall include:

1. General overview of the GPR and its derivation from the Bayesian viewpoint
2. Thorough examination of the known methods accounting for sensor positions uncertainty as suggested in the literature
3. Comparing the examined methods from the theoretical viewpoint
4. Verifying the functionality of the presented methods by implementing them and presenting the simulation results.

RECOMMENDED LITERATURE:

[1] C. E. Rasmussen and Ch. K. I. Williams. Gaussian processes for machine learning. Adaptive computation and machine learning. MIT Press, Cambridge, Mass, 2006.

[2] M. Jadaliha, Y. Xu, J. Choi, N. S. Johnson, and W. Li. Gaussian Process Regression for Sensor Networks Under Localization Uncertainty. IEEE Transactions on Signal Processing, 61(2):223–237, January 2013.

[3] L. S. Muppirisetty, T. Svensson, and H. Wymeersch. Spatial Wireless Channel Prediction under Location Uncertainty. IEEE Transactions on Wireless Communications, 15(2):1031–1044, February 2016.

Date of project specification: 8.2.2021 **Deadline for submission:** 5.8.2021

Supervisor: doc. RNDr. Jitka Poměnková, Ph.D.
Consultant: prof. Dr. Ing. Franz Hlawatsch

doc. Ing. Tomáš Frýza, Ph.D.
Chair of study program board

ABSTRACT

In this thesis, we investigate the task of spatial function estimation from the viewpoint of Gaussian Process Regression (GPR) while accounting for uncertain training positions (uncertain sensor positions, uncertain inputs). We first present the theory behind GPR with known training positions. The theory is then applied to derive the expressions for the GPR predictive distribution at a test position under training position uncertainty. Because these expressions are intractable, they are evaluated approximately using the Monte Carlo sampling method. This method is demonstrated to improve the prediction performance over the standard usage of GPR not accounting for uncertainty and also compared to a simplified approach present in the literature.

We furthermore investigate the possibilities of performing GPR under training position uncertainty while using closed form expressions for prediction reported in the literature. It turns out that significant approximations are needed to obtain these closed form expressions, which makes the resulting posterior distribution inherently approximate. In fact, the resulting GPR method uses the standard form of GPR for prediction along with a modified expression of the covariance function. A simulation shows that the prediction results of this method are similar to those of standard GPR not accounting for uncertainty. On the other hand, the posterior variance indicating the prediction uncertainty was increased, which is the desired effect of incorporating uncertainty of training positions.

ABSTRAKT

Tato práce se zabývá úlohou odhadování prostorové funkce z hlediska regrese pomocí Gaussovských procesů (GPR) za současné nejistoty tréninkových pozic (pozic senzorů). Nejdříve je zde popsána teorie v pozadí GPR metody pracující se známými tréninkovými pozicemi. Tato teorie je poté aplikována při odvození výrazů prediktivní distribuce GPR v testovací pozici při uvážení nejistoty tréninkových pozic. Kvůli absenci analytického řešení těchto výrazů byly výrazy aproximovány pomocí metody Monte Carlo. U odvozené metody bylo demonstrováno zlepšení kvality odhadu prostorové funkce oproti standardnímu použití GPR metody a také oproti zjednodušenému řešení uvedenému v literatuře.
Dále se práce zabývá možností použití metody GPR s nejistými tréninkovými pozicemi v kombinaci s výrazy s dostupným analytickým řešením. Ukazuje se, že k dosažení těchto výrazů je třeba zavést značné předpoklady, což má od počátku za následek nepřesnost prediktivní distribuce. Také se ukazuje, že výsledná metoda používá standardní výrazy GPR v kombinaci s upravenou kovarianční funkcí. Simulace dokazují, že tato metoda produkuje velmi podobné odhady jako základní GPR metoda uvažující známé tréninkové pozice. Na druhou stranu prediktivní variance (nejistota odhadu) je u této metody zvýšena, což je žádaný efekt uvážení nejistoty tréninkových pozic.

KEYWORDS

Gaussian Process Regression (GPR), Uncertain sensor (training) positions, Uncertain inputs, Machine learning, Monte Carlo sampling

PTÁČEK, Martin. *Spatial Function Estimation with Uncertain Sensor Locations*. Brno, 2021. Available from: https://www.vutbr.cz/studenti/zav-prace/detail/137441. Master's Thesis. Brno University of Technology, Faculty of Electrical Engineering and Communication, Department of Radio Electronics. Advised by Franz Hlawatsch, Ao.Univ.Prof. Dipl.-Ing. Dr.techn.

Author's Declaration

Author:	Bc. Martin Ptáček
Author's ID:	182742
Paper type:	Master's Thesis
Academic year:	2020/21
Topic:	Spatial Function Estimation with Uncertain Sensor Locations

I declare that I have written this paper independently, under the guidance of the advisor and using exclusively the technical references and other sources of information cited in the paper and listed in the comprehensive bibliography at the end of the paper.

As the author, I furthermore declare that, with respect to the creation of this paper, I have not infringed any copyright or violated anyone's personal and/or ownership rights. In this context, I am fully aware of the consequences of breaking Regulation 11 of the Copyright Act No. 121/2000 Coll. of the Czech Republic, as amended, and of any breach of rights related to intellectual property or introduced within amendments to relevant Acts such as the Intellectual Property Act or the Criminal Code, Act No. 40/2009 Coll. of the Czech Republic, Section 2, Head VI, Part 4.

Brno ....

author's signature*

*The author signs only in the printed version.

ACKNOWLEDGEMENT

I would like to thank my supervisor, Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Franz Hlawatsch, for suggesting the topic of my diploma thesis and sharing his expertise. I am especially grateful for his constant helpfulness and support during the preparation of my diploma thesis.

I am also grateful to my co-supervisor, doc. RNDr. Jitka Poměnková, Ph.D., for much appreciated organizational help, comments and support.

Contents

List of Figures

List of Tables

Introduction

Spatial function estimation or spatial regression is a task receiving a lot of attention in multiple fields including astronomy, geostatistics, mining, meteorology, and recently machine learning and telecommunications. In the field of telecommunications, spatial function estimation finds its application in estimating channel quality metrics, for example building received signal strength indicator (RSSI) maps, which can be further used for resource allocation and network performance optimization. [1] [2]

Spatial function estimation

A spatial function can be imagined as a distribution of some scalar quantity, such as temperature, illuminance, underground gold density, or wind power in an area. For these examples, we consider the area (space) to be two-dimensional. An example of a spatial function is shown in Figure 1.

Within the task of spatial function estimation, we want to predict the spatial function value at a given position called the *test position* based on noisy observations of the spatial function at several known positions termed the *training positions*. An example of a set of training positions together with the corresponding spatial function observations is shown in Figure 1. The extension to multiple test positions in order to obtain a kind of *map* of the estimated function is straightforward by performing the prediction separately for each test position.

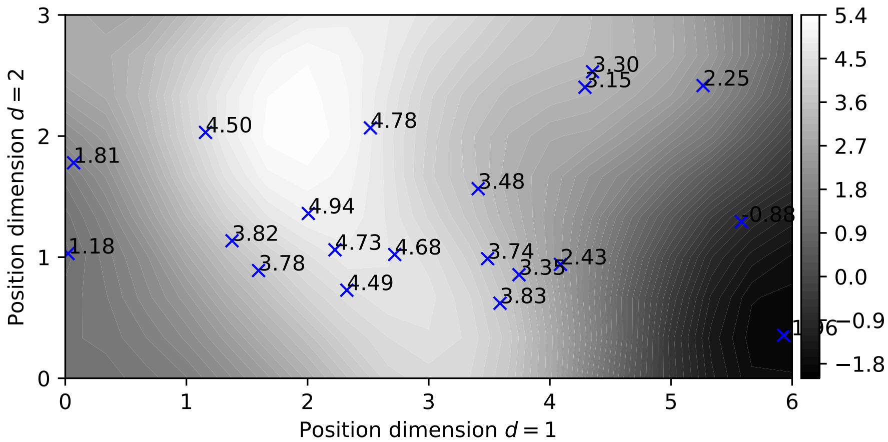


Fig. 1: An example of a spatial function (represented by gray levels) and training positions (indicated by blue crosses) along with the observed function values.

Gaussian Process Regression

Gaussian process regression (GPR) [1], [3] is a method for spatial regression that is used in many fields including geostatistics, meteorology, astronomy and many more. Recently it has gained popularity because of its success in applications as a general machine learning method. The advantages of GPR are its flexibility and the availability of a closed form expression for the predictions. Another significant advantage is the ability of GPR to indicate the level of belief (uncertainty) for the provided predictions. A significant disadvantage of GPR is its computational complexity, which grows cubically with the number of training positions. This issue can be dealt with using multiple methods [2]. Another disadvantage of GPR is its sensitivity to errors (uncertainty) in the training positions (inputs). This is the problem addressed in this thesis. An introduction into GPR will be provided in Chapter 1. [2]

Training position uncertainty

The presence of uncertainty in the training positions has a strong effect on the quality of GPR predictions. One aspect is that the predicted function value is likely to contain larger errors. A second aspect is that the prediction uncertainty provided by the standard GPR framework is too confident, i.e. smaller than the true uncertainty. This is important especially when GPR is used for decision making in critical processes [3].

Problem formulation

We will investigate a GPR scenario where the training positions are uncertain while the test position is known. This type of problem was chosen because it comprises most of the difficulties of the inference task while allowing for easy visualization and evaluation of the prediction quality.

In contrast to the channel prediction and machine learning literature, we assume complete prior knowledge of the statistical properties of the spatial function including the hyperparameters of the Gaussian process. Thus, estimation (learning) of these hyperparameters is beyond the scope of this thesis.

In addition, we also assume that the statistical distribution of the uncertain training positions is known. In some sections, we will furthermore assume a specific form of the distribution of the training positions.

Thesis outline

In Chapter 1, we describe the theory behind GPR and interpret GPR from the Bayesian perspective. In Chapter 2, we incorporate training positions uncertainty and derive integral expressions of the posterior distribution and of the posterior mean and variance. A Monte Carlo sampling method is then used to evaluate these expressions. The resulting GPR performance is then compared via simulations to the performance of standard GPR. We also describe a simplified approach for incorporating training position uncertainty and compare it to the original approach. In Chapter 3, we investigate the possibility of performing GPR under training position uncertainty using closed form expressions proposed in the literature. We formulate the underlying assumptions and compare the performance of the resulting GPR method to that of the methods from Chapter 2.

1 Gaussian Process Regression

Gaussian process regression (GPR) is a relatively old framework dating back to the times of Wiener and Kolmogorov in 1940's and even further back to the work of Danish astronomer Thiele in 1880. GPR finds its applications in geostatistics, where it is termed *kriging* after a mining engineer Krige, whose findings were further developed by Matheron in 1960's [4]. GPR is also applied in meteorology and many other fields including epidemiology [5]. More recently it has been considered in the general regression and machine learning context, where it gained popularity since the publication of Rasmussen in 1996 [6] comparing GPR to other machine learning methods. The idea of GPR in the context of spatial function estimation is to exploit the similarity of a spatial function in positions that are close to each other, i.e. spatially close positions are expected to attain a similar value of the spatial function. This similarity is statistically modeled using a covariance function, which plays a key role in GPR.

The main sources for this chapter are [1, ch. 2] and [7, sec. 2.3] while some derivations were carried out from scratch.

1.1 The Gaussian Process

Definition 1. A *Gaussian Process* (GP) is a collection of random variables, any finite number of which have a joint Gaussian distribution.

The GP will be denoted as f to express that it is a spatial random function of argument $\boldsymbol{x}$, which is a D-dimensional vector such that $\boldsymbol{x} \in \mathbb{R}^D$. Each one-dimensional random variable contained in the GP will then be denoted as $f(\boldsymbol{x})$, where $\boldsymbol{x}$ denotes the position of the random variable within the complete random process.

To completely define a GP we use a *mean function* $m(\boldsymbol{x})$ and a *covariance function* $k(\boldsymbol{x}, \boldsymbol{x}')$. The mean function is defined as

$$m(\boldsymbol{x}) \triangleq \mathrm{E}\{f(\boldsymbol{x})\} = \int_{\mathbb{R}} f' \, p_{f(\boldsymbol{x})}(f') \, df' \ , \tag{1.1}$$

where $p_{f(\boldsymbol{x})}(f')$ denotes the *Probability density function* (pdf) of the random variable $f(\boldsymbol{x})$ evaluated at f'. The covariance function is defined as

$$\begin{aligned} k(\boldsymbol{x}, \boldsymbol{x}') &\triangleq \operatorname{cov}\{f(\boldsymbol{x}), f(\boldsymbol{x}')\} = \mathrm{E}\{(f(\boldsymbol{x}) - m(\boldsymbol{x}))(f(\boldsymbol{x}') - m(\boldsymbol{x}'))\} \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} (f' - m(\boldsymbol{x}))(f'' - m(\boldsymbol{x}')) \, p_{f(\boldsymbol{x}) f(\boldsymbol{x}')}(f', f'') \, df' \, df'' \ , \end{aligned} \tag{1.2}$$

where $p_{f(\boldsymbol{x}), f(\boldsymbol{x}')}(f', f'')$ denotes the joint pdf of the random variables $f(\boldsymbol{x})$ and $f(\boldsymbol{x}')$ evaluated at f' and f''. Finally, the complete GP is expressed as

$$p(f(\boldsymbol{x})) = \mathcal{GP}(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}')) \ , \tag{1.3}$$

where the position $\boldsymbol{x}$ on the left side of (1.3) describes no specific position but generally the complete space of positions $\mathbb{R}^D$.

1.2 Observation Model

We define a set of I known positions $\boldsymbol{x}^{(i)}$ with $i = 1, 2, \ldots, I$ called *training positions*, which we stacked into a column vector of all training positions

$$\boldsymbol{x}_\mathrm{t} \triangleq \mathrm{col}(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots, \boldsymbol{x}^{(I)}) \in \mathbb{R}^{DI} \ . \tag{1.4}$$

Correspondingly, we have a set of GP random variables at the training positions $\boldsymbol{x}^{(i)}$, which are arranged into the random vector

$$\boldsymbol{f} \triangleq (f(\boldsymbol{x}^{(1)}) \quad f(\boldsymbol{x}^{(2)}) \quad \ldots \quad f(\boldsymbol{x}^{(I)}))^\mathrm{T} \in \mathbb{R}^I \ . \tag{1.5}$$

According to the Definition 1, $\boldsymbol{f}$ has a multivariate Gaussian distribution

$$p(\boldsymbol{f}) = \mathcal{N}(\boldsymbol{f};\ \boldsymbol{m},\ \boldsymbol{K}) \ , \tag{1.6}$$

where $\boldsymbol{m}$ denotes the vector of the GP mean values at the training positions $\boldsymbol{x}^{(i)}$, i.e.,

$$\boldsymbol{m} = \mathrm{E}\{\boldsymbol{f}\} = (m(\boldsymbol{x}^{(1)}) \quad m(\boldsymbol{x}^{(2)}) \quad \cdots \quad m(\boldsymbol{x}^{(I)}))^\mathrm{T} \ , \in \mathbb{R}^I \tag{1.7}$$

where $m(\boldsymbol{x}^{(i)})$ is defined in (1.1). $\boldsymbol{K}$ denotes the GP covariance matrix at the training positions $\boldsymbol{x}^{(i)}$, i.e.,

$$\boldsymbol{K} \triangleq \mathrm{cov}\{\boldsymbol{f}\} = \begin{pmatrix} k(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(1)}) & k(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}) & \cdots & k(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(I)}) \\ k(\boldsymbol{x}^{(2)}, \boldsymbol{x}^{(1)}) & k(\boldsymbol{x}^{(2)}, \boldsymbol{x}^{(2)}) & \cdots & k(\boldsymbol{x}^{(2)}, \boldsymbol{x}^{(I)}) \\ \vdots & \vdots & \ddots & \vdots \\ k(\boldsymbol{x}^{(I)}, \boldsymbol{x}^{(1)}) & k(\boldsymbol{x}^{(I)}, \boldsymbol{x}^{(2)}) & \cdots & k(\boldsymbol{x}^{(I)}, \boldsymbol{x}^{(I)}) \end{pmatrix} \in \mathbb{R}^{I \times I} \ , \tag{1.8}$$

where $k(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})$ is the GP covariance function defined in (1.2).

We assume we do not have access to the realization of the random vector $\boldsymbol{f}$, but we have access to a vector of observations $\boldsymbol{y}$ defined as

$$\boldsymbol{y} \triangleq (y^{(1)} \quad y^{(2)} \quad \ldots \quad y^{(I)})^\mathrm{T} \in \mathbb{R}^I \tag{1.9}$$

that is a noisy version of $\boldsymbol{f}$, i.e.,

$$\boldsymbol{y} = \boldsymbol{f} + \boldsymbol{\epsilon} \ , \tag{1.10}$$

where $\boldsymbol{\epsilon} \triangleq (\epsilon^{(1)} \quad \epsilon^{(2)} \quad \ldots \quad \epsilon^{(I)})^\mathrm{T} \in \mathbb{R}^I$ is the vector of observation errors (measurement noise). $\boldsymbol{\epsilon}$ is assumed to be zero-mean isotropic Gaussian according to

$$p(\boldsymbol{\epsilon}) = \mathcal{N}(\boldsymbol{\epsilon};\ \mathbf{0},\ \sigma_\epsilon^2 \mathbf{I}_I) \ , \tag{1.11}$$

where $\mathbf{I}_I$ denotes the identity matrix of dimensions $I \times I$. As the observation errors $\epsilon^{(i)}$ are jointly Gaussian and uncorrelated, they are also independent. It is furthermore assumed that $\boldsymbol{\epsilon}$ is independent of $\boldsymbol{f}$. From (1.10) and our statistical assumptions, it then follows that $\boldsymbol{y}$ is a Gaussian random vector distributed according to

$$p(\boldsymbol{y}) = \mathcal{N}(\boldsymbol{y};\, \boldsymbol{m},\, \boldsymbol{Q}) \;, \tag{1.12}$$

with a covariance matrix $\boldsymbol{Q}$ of the random vector $\boldsymbol{y}$ given as

$$\boldsymbol{Q} = \boldsymbol{Q} = \text{cov}\{\boldsymbol{f}\} + \text{cov}\{\boldsymbol{\epsilon}\} = \boldsymbol{K} + \sigma_\epsilon^2 \mathbf{I}_I \;, \tag{1.13}$$

where we used the independence of random vectors $\boldsymbol{f}$ and $\boldsymbol{\epsilon}$.

1.3 Regression Problem

In the regression problem, we are interested in the distribution of the GP at a single known *test position* $\boldsymbol{x}^{(*)} \in \mathbb{R}^D$, which is the distribution of the random variable $f(\boldsymbol{x}^{(*)})$. This random variable will be further denoted as $f_* \triangleq f(\boldsymbol{x}^{(*)})$. The test position $\boldsymbol{x}^{(*)}$ is in general not contained in the vector of training positions $\boldsymbol{x}$ but the regression method works the same even in case the test position is equal to one of the training positions $\boldsymbol{x}^{(i)}$. Without performing any observations $y^{(i)}$ of the GP at the training positions $\boldsymbol{x}^{(i)}$, the distribution of f_* corresponds to our *prior* knowledge about the GP, i.e. is the GP distribution. According to (1.3), the distribution of f_* is

$$p(f_*) = \mathcal{N}(f_*;\, m_*,\, k_*) \;, \tag{1.14}$$

where m_* is the mean value of the GP at the test position $\boldsymbol{x}^{(*)}$,

$$m_* \triangleq \mathrm{E}\{f_*\} = m(\boldsymbol{x}_*) \;, \tag{1.15}$$

and k_* is the variance of the GP at $\boldsymbol{x}^{(*)}$,

$$k_* \triangleq \text{var}\{f_*\} = k(\boldsymbol{x}^{(*)}, \boldsymbol{x}^{(*)}) \;. \tag{1.16}$$

To incorporate the GP observations $y^{(i)}$ into the regression problem within the Bayesian framework, we consider the conditional distribution of the test position GP random variable f_* given the random vector of observations $\boldsymbol{y}$, i.e., the **posterior** pdf $p(f_*|\boldsymbol{y})$. Using Bayes' theorem, we obtain

$$p(f_*|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|f_*)\, p(f_*)}{p(\boldsymbol{y})} = \frac{p(f_*, \boldsymbol{y})}{p(\boldsymbol{y})} \;, \tag{1.17}$$

which will also be called the *predictive distribution.* This posterior pdf is the key to estimating the desired parameter, i.e., the GP random variable at the test position, f_*. Next, we consider some of the individual terms in (1.17).

The conditional pdf of random observations $\boldsymbol{y}$ given the GP random value at test position f_*, $p(\boldsymbol{y}|f_*)$, will be termed *likelihood*. The likelihood will be further studied in Section 1.5. The unconditional pdf $p(\boldsymbol{y})$ of the vector of observations $\boldsymbol{y}$ is often called the *evidence*. According to (1.12), $p(\boldsymbol{y})$ is Gaussian.

In contrast to the typical usage of Bayesian inference, here we have a direct access to the *joint distribution* of f_* and $\boldsymbol{y}$. According to the definition of the GP, this distribution is Gaussian and given by

$$p(f_*, \boldsymbol{y}) = p\left(\begin{pmatrix} f_* \\ \mathbf{y} \end{pmatrix}\right) = \mathcal{N}\left(\begin{pmatrix} f_* \\ \boldsymbol{y} \end{pmatrix}; \begin{pmatrix} m_* \\ \boldsymbol{m} \end{pmatrix}, \begin{pmatrix} k_* & \boldsymbol{c}^{\mathrm{T}} \\ \boldsymbol{c} & \boldsymbol{Q} \end{pmatrix}\right) , \tag{1.18}$$

where $\boldsymbol{c} \triangleq \operatorname{cov}\{\boldsymbol{y},\ f_*\}$ (a column vector of dimension I) is the cross-covariance of the vector of observations $\boldsymbol{y} = \boldsymbol{f} + \boldsymbol{\epsilon}$ and the GP random variable f_* at the test position $\boldsymbol{x}^{(*)}$. We obtain

$$\boldsymbol{c} = \operatorname{cov}\{\boldsymbol{f} + \boldsymbol{\epsilon},\ f_*\} = \operatorname{cov}\{\boldsymbol{f},\ f_*\} + \operatorname{cov}\{\boldsymbol{\epsilon},\ f_*\} \in \mathbb{R}^I . \tag{1.19}$$

Since $\boldsymbol{\epsilon}$ and f_* are independent, $\operatorname{cov}\{\boldsymbol{\epsilon},\ f_*\} = \mathbf{0}$ and thus we obtain

$$\boldsymbol{c} = \operatorname{cov}\{\boldsymbol{f},\ f_*\} = (k(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(*)}) \quad k(\boldsymbol{x}^{(2)}, \boldsymbol{x}^{(*)}) \quad \dots \quad k(\boldsymbol{x}^{(I)}, \boldsymbol{x}^{(*)}))^{\mathrm{T}} . \tag{1.20}$$

1.4 Posterior Distribution

Since f_* and $\boldsymbol{y}$ are jointly Gaussian according to (1.18), the mean of the posterior distribution (1.17) is obtained using formula in Appendix A.1 as

$$\mu_{f_*|y} = \mathrm{E}\{f_*|\boldsymbol{y}\} = m_* + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{Q}^{-1}(\boldsymbol{y} - \boldsymbol{m}) . \tag{1.21}$$

Similarly, the posterior variance is using Appendix A.1 given as

$$\operatorname{var}\{f_*|\boldsymbol{y}\} = \sigma^2_{f_*|y} = \frac{1}{\Lambda_{f_*}} = k_* - \boldsymbol{c}^{\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{c} . \tag{1.22}$$

These results are derived in Appendix B using *completing the square* method and formulated firstly using precision submatrices and then using covariance submatrices of the joint distribution in (1.18).

1.5 Likelihood Function

Next we shall express the likelihood function $p(\boldsymbol{y}|f_*)$. We must note that the likelihood function is considered as a function of GP test position value f_* while having the observations vector $\boldsymbol{y}$ given. Therefore the likelihood function is not a probability distribution as it in general does not integrate to one over the parameter f_* as in

$$\int_{\mathbb{R}} p_{\boldsymbol{y}|f_*}(\boldsymbol{y}|f'_*)\, df'_* \neq 1 \ , \tag{1.23}$$

where $p_{\boldsymbol{y}|f_*}(\boldsymbol{y}|f'_*)$ denotes the probability density of the vector $\boldsymbol{y}$ given the GP realization f'_* at the test position $\boldsymbol{x}^{(*)}$.

Using the product rule we express the likelihood function similarly to (1.17) as

$$p(\boldsymbol{y}|f_*) = \frac{p(f_*, \boldsymbol{y})}{p(f_*)} \ . \tag{1.24}$$

Further considering the terms of the joint distribution submatrices and the prior (1.14) we can express the likelihood function similarly as in (B.3) obtaining

$$\begin{aligned}
p(\boldsymbol{y}|f_*) &= \frac{(2\pi)^{-\frac{I+1}{2}} |\boldsymbol{\Lambda}_\mathrm{N}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\begin{pmatrix} f_* \\ \boldsymbol{y} \end{pmatrix} - \begin{pmatrix} m_* \\ \boldsymbol{m} \end{pmatrix}\right)^\mathrm{T} \boldsymbol{\Lambda}_\mathrm{N} \left(\begin{pmatrix} f_* \\ \boldsymbol{y} \end{pmatrix} - \begin{pmatrix} m_* \\ \boldsymbol{m} \end{pmatrix}\right)\right)}{(2\pi k_*)^{-\frac{1}{2}} \exp(-\frac{1}{2}(f_* - m_*) k_*^{-1} (f_* - m_*))} \\
p(\boldsymbol{y}|f_*) &= (2\pi)^{-\frac{I}{2}} |\boldsymbol{\Lambda}_\mathrm{N}|^{\frac{1}{2}} k_*^{\frac{1}{2}} \\
&\quad \cdot \exp\left(-\frac{1}{2}\left(\begin{pmatrix} f_* \\ \boldsymbol{y} \end{pmatrix} - \begin{pmatrix} m_* \\ \boldsymbol{m} \end{pmatrix}\right)^\mathrm{T} \boldsymbol{\Lambda}_\mathrm{N} \left(\begin{pmatrix} f_* \\ \boldsymbol{y} \end{pmatrix} - \begin{pmatrix} m_* \\ \boldsymbol{m} \end{pmatrix}\right)\right) \\
&\quad \cdot \exp\left(\frac{1}{2}(f_* - m_*) k_*^{-1} (f_* - m_*)\right) \ .
\end{aligned} \tag{1.25}$$

Observing the two rightmost exponents in (1.25) we see that they both contain a quadratic function of f_*. Therefore the likelihood function takes the form of a Gaussian distribution up to a normalizing constant. Further we use *completing the square* method to obtain the mean and variance of the normalized version of the likelihood function

$$p_\mathrm{norm}(\boldsymbol{y}|f_*) = \mathcal{N}(\boldsymbol{y};\ \mu_{y|f_*},\ \sigma^2_{y|f_*}) \propto p(\boldsymbol{y}|f_*) \tag{1.26}$$

that is a valid probability distribution

$$\begin{aligned}
p_\mathrm{norm}(\boldsymbol{y}|f_*) &= (2\pi\sigma^2_{y|f_*})^{-\frac{1}{2}} \exp\left(-\frac{(f_* - \mu_{y|f_*})^2}{2\sigma^2_{y|f_*}}\right) \\
&= (2\pi\sigma^2_{y|f_*})^{-\frac{1}{2}} \exp\left(-\frac{(f_*^2 - 2 f_* \mu_{y|f_*} + \mu^2_{y|f_*})}{2\sigma^2_{y|f_*}}\right) \ .
\end{aligned} \tag{1.27}$$

It shall be noted that the normalized likelihood function parameters $\mu_{y|f_*}$ and $\sigma^2_{y|f_*}$ are scalars and therefore indicate that we are considering the likelihood as a function of f_*. If we were considering the likelihood as a function of $\boldsymbol{y}$, the likelihood would be a multivariate Gaussian distribution with vector mean $\boldsymbol{\mu}_{y|f_*}$ and covariance matrix as parameters, which ensures distinct notation of these parameters.

Normalized likelihood function mean and variance with precision submatrices

Using the precision submatrices in (B.2) we can develop the rightmost exponent of (1.25) as

$$\begin{aligned}
&-\frac{1}{2}(f_*-m_*)\Lambda_{f_*}(f_*-m_*)-\frac{1}{2}(f_*-m_*)\boldsymbol{\Lambda}_{f_*,y}^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{m})\\
&-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{m})^{\mathrm{T}}\boldsymbol{\Lambda}_{f_*,y}(f_*-m_*)-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{m})^{\mathrm{T}}\boldsymbol{\Lambda}_{y}(\boldsymbol{y}-\boldsymbol{m})\\
&+\frac{1}{2}(f_*-m_*)k_*^{-1}(f_*-m_*)\\
&=-\frac{1}{2}(f_*-m_*)\Lambda_{f_*}(f_*-m_*)-(f_*-m_*)\boldsymbol{\Lambda}_{f_*,y}^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{m})\\
&-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{m})^{\mathrm{T}}\boldsymbol{\Lambda}_{y}(\boldsymbol{y}-\boldsymbol{m})+\frac{1}{2}(f_*-m_*)k_*^{-1}(f_*-m_*)\\
&=-\frac{1}{2}f_*^2(\Lambda_{f_*}-k_*^{-1})+f_*(\Lambda_{f_*}m_*-\boldsymbol{\Lambda}_{f_*,y}^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{m})-k_*^{-1}m_*)+\text{const.}\ ,
\end{aligned} \tag{1.28}$$

where we in the last step disregarded the terms independent of f_*. Taking the term quadratic in f_* from (1.28) and placing it equal to the term quadratic in f_* in rightmost exponent in (1.27) we get

$$-\frac{1}{2}f_*^2(\Lambda_{f_*}-k_*^{-1})=-\frac{1}{2\sigma_{y|f_*}^2}f_*^2\ , \tag{1.29}$$

which can be rearranged to express the variance of the normalized likelihood $p_{\text{norm}}(\boldsymbol{y}|f_*)$ as

$$\sigma_{y|f_*}^2=\frac{1}{\Lambda_{f_*}-k_*^{-1}}\ . \tag{1.30}$$

Next, taking the linear term in f_* from (1.28) and placing it equal to the term linear in f_* in the rightmost exponent in (1.27) we get

$$f_*(\Lambda_{f_*}m_*-\boldsymbol{\Lambda}_{f_*,y}^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{m})-k_*^{-1}m_*)=f_*\frac{\mu_{y|f_*}}{\sigma_{y|f_*}^2}\ , \tag{1.31}$$

which can be rearranged to express the mean value of the normalized likelihood $p_{\text{norm}}(\boldsymbol{y}|f_*)$ as

$$\mu_{y|f_*}=\sigma_{y|f_*}^2(\Lambda_{f_*}m_*-\boldsymbol{\Lambda}_{f_*,y}^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{m})-k_*^{-1}m_*) \tag{1.32}$$

$$=\frac{1}{\Lambda_{f_*}-k_*^{-1}}(\Lambda_{f_*}m_*-\boldsymbol{\Lambda}_{f_*,y}^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{m})-k_*^{-1}m_*)$$

$$=m_*-\frac{\boldsymbol{\Lambda}_{f_*,y}^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{m})}{\Lambda_{f_*}-k_*^{-1}}\ , \tag{1.33}$$

where we used the expression for normalized likelihood variance in (1.30).

Normalized likelihood function mean and variance with covariance submatrices

The normalized likelihood function variance in (1.30) can also be expressed only using joint covariance submatrices in (B.1) by inserting (B.15) obtaining

$$\sigma^2_{y|f_*} = \frac{1}{(k_* - \boldsymbol{c}^{\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{c})^{-1} - k_*^{-1}} \ . \tag{1.34}$$

The same can be done for the normalized likelihood function mean in (1.33) to express it using covariance terms by inserting (B.15) and (B.17) obtaining

$$\begin{aligned} \mu_{y|f_*} &= m_* - \frac{\boldsymbol{\Lambda}^{\mathrm{T}}_{f_*,y}(\boldsymbol{y}-\boldsymbol{m})}{\Lambda_{f_*} - k_*^{-1}} \\ &= m_* + \frac{\boldsymbol{c}^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{m})}{k_*\boldsymbol{Q} - \boldsymbol{c}^{\mathrm{T}}\boldsymbol{c}}((k_* - \boldsymbol{c}^{\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{c})^{-1} - k_*^{-1})^{-1} \ . \end{aligned} \tag{1.35}$$

1.6 Bayesian View of GPR

In the derivation of the GPR posterior distribution $p(f_*|\boldsymbol{y})$ in 1.4 we directly used the known joint distribution $p(f_*, \boldsymbol{y})$ as our basis. The joint distribution is in more common cases of Bayesian inference not directly known. In these cases we resort to decomposition of joint distribution using the product rule into product of likelihood and prior given by

$$p(f_*, \boldsymbol{y}) = p(\boldsymbol{y}|f_*)\, p(f_*) \ , \tag{1.36}$$

where we will consider the likelihood $p(\boldsymbol{y}|f_*)$ and prior $p(f_*)$ to be known, which is true even in this case given the derivation in 1.5 and the prior expression in (1.14).

Considering the Bayes' rule expression in (1.17) we can now express the posterior distribution

$$p(f_*|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|f_*)\, p(f_*)}{p(\boldsymbol{y})} \propto p(\boldsymbol{y}|f_*)\, p(f_*) \ , \tag{1.37}$$

where in the last step we disregarded the normalizing constant not dependent of f_*. Inserting now the normal distribution expressions from (1.26) and (1.14) into (1.37) results in

$$p(f_*|\boldsymbol{y}) \propto \mathcal{N}(f_*;\ \mu_{y|f_*},\ \sigma^2_{y|f_*})\mathcal{N}(f_*;\ m_*,\ k_*) \ , \tag{1.38}$$

which can be further expanded while disregarding the normalizing constants as

$$\begin{aligned} p(f_*|\boldsymbol{y}) &\propto \exp\left(-\frac{(f_* - \mu_{y|f_*})^2}{2\sigma^2_{y|f_*}}\right) \exp\left(-\frac{(f_* - m_*)^2}{2k_*}\right) \\ &\propto \exp\left(-\frac{f_*^2}{2\sigma^2_{y|f_*}} + \frac{f_*\mu_{y|f_*}}{\sigma^2_{y|f_*}} - \frac{\mu^2_{y|f_*}}{2\sigma^2_{y|f_*}} - \frac{f_*^2}{2k_*} + \frac{f_* m_*}{k_*} - \frac{m_*^2}{2k_*}\right) \ . \end{aligned} \tag{1.39}$$

Because of the presence of the quadratic term f_*^2 in the exponent we know the posterior distribution is Gaussian the same way as in (B.5) in the form

$$p(f_*|\boldsymbol{y}) = \mathcal{N}(f_*;\, \mu_{f_*|y},\, \sigma^2_{f_*|y}) = (2\pi\sigma^2_{f_*|y})^{-\frac{1}{2}}\exp\left(-\frac{f_*^2 - 2f_*\mu_{f_*|y} + \mu^2_{f_*|y}}{2\sigma^2_{f_*|y}}\right) , \quad (1.40)$$

which suggests that we can employ the same method as in previous sections, *completing the square*, to express the posterior mean $\mu_{f_*|y}$ and variance $\sigma^2_{f_*|y}$.

Comparing the terms containing f_*^2 in exponents of (1.39) and (1.40) we obtain

$$-\frac{f_*^2}{2\sigma^2_{y|f_*}} - \frac{f_*^2}{2k_*} = -\frac{f_*^2}{2\sigma^2_{f_*|y}} . \quad (1.41)$$

From here we can express the posterior variance as

$$\sigma^2_{y|f_*} = \frac{1}{\sigma^{-2}_{y|f_*} + k_*^{-1}} , \quad (1.42)$$

which after inserting the result for normalized likelihood function variance (1.30) becomes

$$\sigma^2_{y|f_*} = \frac{1}{\Lambda_{f_*} - k_*^{-1} + k_*^{-1}} = \frac{1}{\Lambda_{f_*}} . \quad (1.43)$$

Now expressing the GP precision Λ_{f_*} using covariance submatrices as in (B.15) we obtain

$$\sigma^2_{y|f_*} = k_* - \boldsymbol{c}^{\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{c} , \quad (1.44)$$

which is identical result as in (1.22).

Further comparing the terms linear in f_* in exponents of (1.39) and (1.40) we obtain

$$\frac{f_*\mu_{y|f_*}}{\sigma^2_{y|f_*}} + \frac{f_* m_*}{k_*} = \frac{f_*\mu_{f_*|y}}{\sigma^2_{f_*|y}} , \quad (1.45)$$

from where we can express the posterior mean as

$$\mu_{f_*|y} = \left(\frac{\mu_{y|f_*}}{\sigma^2_{y|f_*}} + \frac{m_*}{k_*}\right)\sigma^2_{f_*|y} , \quad (1.46)$$

which can be expanded using precision terms for posterior precision in (B.7) and normalized likelihood mean in (1.32) as

$$\begin{aligned}\mu_{f_*|y} &= \left(\sigma^{-2}_{y|f_*}\sigma^2_{y|f_*}(\Lambda_{f_*}m_* - \boldsymbol{\Lambda}^{\mathrm{T}}_{f_*,y}(\boldsymbol{y}-\boldsymbol{m}) - k_*^{-1}m_*) + \frac{m_*}{k_*}\right)\frac{1}{\Lambda_{f_*}} \\ &= \frac{\Lambda_{f_*}m_* - \boldsymbol{\Lambda}^{\mathrm{T}}_{f_*,y}(\boldsymbol{y}-\boldsymbol{m})}{\Lambda_{f_*}} \\ &= m_* - \Lambda_{f_*}^{-1}\boldsymbol{\Lambda}^{\mathrm{T}}_{f_*,y}(\boldsymbol{y}-\boldsymbol{m}) .\end{aligned} \quad (1.47)$$

This is the same expression as derived in (B.11) directly from the joint distribution $p(f_*, \boldsymbol{y})$, i.e., showing the equivalence of the Bayesian approach.

2 Sampling Methods for GPR Under Position Uncertainty

Until now we considered the training positions $\boldsymbol{x}^{(i)}$ and test position $\boldsymbol{x}^{(*)}$ as being perfectly known known vectors. This is an unrealistic assumption in practical applications and may lead to significant errors in spatial function estimation. The position information obtained using a localization technique is subject to measurement errors, which lead to position estimation errors. In this section we shall consider methods for coping with position uncertainties to enhance the performance of GPR in this more general setup.

2.1 Incorporating Uncertain Training Positions

This approach for incorporating location uncertainty into GPR is based on [2]. To address the uncertainty in the training positions, we will consider the vector of the stacked training positions $\boldsymbol{x}_\mathrm{t}$ defined in (1.4) as being random from now on. Until now we considered implicit training positions $\boldsymbol{x}_\mathrm{t}$ of the random variables $\boldsymbol{f}$ within the GP. From now on, it needs to be explicitly specified which positions in the GP we consider. This will be represented by conditioning the GP random variables on the vector of training positions $\boldsymbol{x}_\mathrm{t}$. Thus, the random vector $\boldsymbol{f}$ from (1.5) is written as

$$\boldsymbol{f}|\boldsymbol{x}_\mathrm{t} = (f(\boldsymbol{x}^{(1)}) \quad f(\boldsymbol{x}^{(2)}) \quad \ldots \quad f(\boldsymbol{x}^{(I)}))^\mathrm{T} \in \mathbb{R}^I \ , \tag{2.1}$$

where $\boldsymbol{x}^{(i)}$ for $i = 1, \ldots, I$ are the individual random training positions contained in $\boldsymbol{x}_\mathrm{t}$. The vector of random training positions $\boldsymbol{x}_\mathrm{t}$ will be considered to have a known prior distribution $p(\boldsymbol{x}_\mathrm{t})$. The localization technique outputs an estimate of the training positions. We will consider the estimate to be the observation $\tilde{\boldsymbol{x}}_\mathrm{t}$ of the true training positions in the form

$$\tilde{\boldsymbol{x}}_\mathrm{t} = \mathrm{col}(\tilde{\boldsymbol{x}}^{(1)}, \tilde{\boldsymbol{x}}^{(2)}, \ldots, \tilde{\boldsymbol{x}}^{(I)}) \in \mathbb{R}^{DI} \ . \tag{2.2}$$

This observation involves a random position observation noise vector

$$\boldsymbol{w} = \mathrm{col}(\boldsymbol{w}^{(1)}, \boldsymbol{w}^{(2)}, \ldots, \boldsymbol{w}^{(I)}) \in \mathbb{R}^{DI} \tag{2.3}$$

according to

$$\tilde{\boldsymbol{x}}_\mathrm{t} = \boldsymbol{x}_\mathrm{t} + \boldsymbol{w} \ , \tag{2.4}$$

which can be decomposed into the individual training position observation vectors

$$\tilde{\boldsymbol{x}}^{(i)} = \boldsymbol{x}^{(i)} + \boldsymbol{w}^{(i)} \ , \qquad i = 1, \ldots, I \ . \tag{2.5}$$

The pdf $p(\boldsymbol{w})$ of the observation error vector will be considered to be known; it will be denoted by $p_{\boldsymbol{w}}(\cdot)$ when evaluated at a specific position.

The observation vector of training positions $\tilde{\boldsymbol{x}}_{\mathrm{t}}$ is related to the vector of the true training positions $\boldsymbol{x}_{\mathrm{t}}$ according to a known conditional pdf $p(\tilde{\boldsymbol{x}}_{\mathrm{t}}|\boldsymbol{x}_{\mathrm{t}})$, which can be expressed using the known pdf $p_{\boldsymbol{w}}(\cdot)$ of the observation noise vector $\boldsymbol{w}$ as

$$p(\tilde{\boldsymbol{x}}_{\mathrm{t}}|\boldsymbol{x}_{\mathrm{t}}) = p_{\boldsymbol{w}}(\tilde{\boldsymbol{x}}_{\mathrm{t}} - \boldsymbol{x}_{\mathrm{t}}) \ . \tag{2.6}$$

The posterior pdf $p(\boldsymbol{x}_{\mathrm{t}}|\tilde{\boldsymbol{x}}_{\mathrm{t}})$, i.e., the pdf of the vector of the true training positions $\boldsymbol{x}_{\mathrm{t}}$ given the observation vector $\tilde{\boldsymbol{x}}_{\mathrm{t}}$, will be important later. It can be obtained from the prior pdf $p(\boldsymbol{x}_{\mathrm{t}})$ and the likelihood $p(\tilde{\boldsymbol{x}}_{\mathrm{t}}|\boldsymbol{x}_{\mathrm{t}})$ using Bayes' theorem as

$$p(\boldsymbol{x}_{\mathrm{t}}|\tilde{\boldsymbol{x}}_{\mathrm{t}}) = \frac{p(\tilde{\boldsymbol{x}}_{\mathrm{t}}|\boldsymbol{x}_{\mathrm{t}})p(\boldsymbol{x}_{\mathrm{t}})}{p(\tilde{\boldsymbol{x}}_{\mathrm{t}})} = \frac{p(\tilde{\boldsymbol{x}}_{\mathrm{t}}|\boldsymbol{x}_{\mathrm{t}})p(\boldsymbol{x}_{\mathrm{t}})}{\int_{\mathbb{R}^{DI}} p(\tilde{\boldsymbol{x}}_{\mathrm{t}}|\boldsymbol{x}'_{\mathrm{t}})p(\boldsymbol{x}'_{\mathrm{t}})\, d\boldsymbol{x}'_{\mathrm{t}}} \ , \tag{2.7}$$

where $\boldsymbol{x}'_{\mathrm{t}}$ denotes the integration variable corresponding to $\boldsymbol{x}_{\mathrm{t}}$.

In contrast to the setup in Section 1.2, the available observations for GPR now consist only of the vector of spatial function observations $\boldsymbol{y}$ at the true training positions $\boldsymbol{x}_{\mathrm{t}}$ and the vector of observation of training positions $\tilde{\boldsymbol{x}}_{\mathrm{t}}$.

We will consider the posterior predictive pdf in (1.17) and further condition it on the observation vector of training positions $\tilde{\boldsymbol{x}}_{\mathrm{t}}$. This added condition represents the fact that in the random training positions setup the GP random variables are explicitly linked to the specific training positions. The posterior pdf of the GP at the test position $\boldsymbol{x}^{(*)}$ therefore takes the form $p(f_*|\boldsymbol{y}, \tilde{\boldsymbol{x}}_{\mathrm{t})}$, and using the product rule it can be further developed as

$$p(f_*|\boldsymbol{y}, \tilde{\boldsymbol{x}}_{\mathrm{t}}) = \frac{p(f_*, \boldsymbol{y}|\tilde{\boldsymbol{x}}_{\mathrm{t}})}{p(\boldsymbol{y}|\tilde{\boldsymbol{x}}_{\mathrm{t}})} \ . \tag{2.8}$$

Using the sum rule, we now add to the probability densities in the numerator and denominator in the right-hand side of (2.8) the true vector of training positions $\boldsymbol{x}_{\mathrm{t}}$, i.e.,

$$\begin{aligned} p(f_*|\boldsymbol{y}, \tilde{\boldsymbol{x}}_{\mathrm{t}}) &= \frac{\int_{\mathbb{R}^{DI}} p(f_*, \boldsymbol{y}, \boldsymbol{x}_{\mathrm{t}}|\tilde{\boldsymbol{x}}_{\mathrm{t}})\, d\boldsymbol{x}_{\mathrm{t}}}{\int_{\mathbb{R}^{DI}} p(\boldsymbol{y}, \boldsymbol{x}_{\mathrm{t}}|\tilde{\boldsymbol{x}}_{\mathrm{t}})\, d\boldsymbol{x}_{\mathrm{t}}} \\ &= \frac{\int_{\mathbb{R}^{DI}} p(f_*|\boldsymbol{y}, \boldsymbol{x}_{\mathrm{t}}, \tilde{\boldsymbol{x}}_{\mathrm{t}})p(\boldsymbol{y}, \boldsymbol{x}_{\mathrm{t}}|\tilde{\boldsymbol{x}}_{\mathrm{t}})\, d\boldsymbol{x}_{\mathrm{t}}}{\int_{\mathbb{R}^{DI}} p(\boldsymbol{y}|\boldsymbol{x}_{\mathrm{t}}, \tilde{\boldsymbol{x}}_{\mathrm{t}})p(\boldsymbol{x}_{\mathrm{t}}|\tilde{\boldsymbol{x}}_{\mathrm{t}})\, d\boldsymbol{x}_{\mathrm{t}}} \\ &= \frac{\int_{\mathbb{R}^{DI}} p(f_*|\boldsymbol{y}, \boldsymbol{x}_{\mathrm{t}}, \tilde{\boldsymbol{x}}_{\mathrm{t}})p(\boldsymbol{y}|\boldsymbol{x}_{\mathrm{t}}, \tilde{\boldsymbol{x}}_{\mathrm{t}})p(\boldsymbol{x}_{\mathrm{t}}|\tilde{\boldsymbol{x}}_{\mathrm{t}})\, d\boldsymbol{x}_{\mathrm{t}}}{\int_{\mathbb{R}^{DI}} p(\boldsymbol{y}|\boldsymbol{x}_{\mathrm{t}}, \tilde{\boldsymbol{x}}_{\mathrm{t}})p(\boldsymbol{x}_{\mathrm{t}}|\tilde{\boldsymbol{x}}_{\mathrm{t}})\, d\boldsymbol{x}_{\mathrm{t}}} \ , \end{aligned} \tag{2.9}$$

where we in the second and third step used the product rule. Considering the conditional independence of f_* and $\boldsymbol{y}$ of $\tilde{\boldsymbol{x}}_{\mathrm{t}}$ given $\boldsymbol{x}_{\mathrm{t}}$, we can simplify (2.9) into

$$p(f_*|\boldsymbol{y}, \tilde{\boldsymbol{x}}_{\mathrm{t}}) = \frac{\int_{\mathbb{R}^{DI}} p(f_*|\boldsymbol{y}, \boldsymbol{x}_{\mathrm{t}},)p(\boldsymbol{y}|\boldsymbol{x}_{\mathrm{t}})p(\boldsymbol{x}_{\mathrm{t}}|\tilde{\boldsymbol{x}}_{\mathrm{t}})\, d\boldsymbol{x}_{\mathrm{t}}}{\int_{\mathbb{R}^{DI}} p(\boldsymbol{y}|\boldsymbol{x}_{\mathrm{t}})p(\boldsymbol{x}_{\mathrm{t}}|\tilde{\boldsymbol{x}}_{\mathrm{t}})\, d\boldsymbol{x}_{\mathrm{t}}} \ . \tag{2.10}$$

Examining the terms in (2.10), we recognize that

- $p(f_*|\boldsymbol{y}, \boldsymbol{x}_\mathrm{t})$ is the GPR posterior pdf as in (B.5) while considering the training positions given by $\boldsymbol{x}_\mathrm{t}$,;
- $p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t})$ is similar to (1.12), i.e., the pdf of the GP observations at the training positions $\boldsymbol{x}_\mathrm{t}$;
- $p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})$ is the posterior pdf of the training positions given the observation of training positions, cf. (2.7).

In contrast to the posterior pdf $p(f_*|\boldsymbol{y})$ of GPR with known training positions in (B.5), the posterior pdf $p(f_*|\boldsymbol{y}, \tilde{\boldsymbol{x}}_\mathrm{t})$ is not Gaussian, but we will approximate in by a Gaussian pdf, i.e.,

$$p(f_*|\boldsymbol{y}, \tilde{\boldsymbol{x}}_\mathrm{t}) \sim \mathcal{N}(f_*;\ \mathrm{E}\{f_*|\boldsymbol{y}, \tilde{\boldsymbol{x}}_\mathrm{t}\},\ \mathrm{var}\{f_*|\boldsymbol{y}, \tilde{\boldsymbol{x}}_\mathrm{t}\})\ . \tag{2.11}$$

Posterior mean

The mean value of the posterior pdf $p(f_*|\boldsymbol{y}, \tilde{\boldsymbol{x}}_\mathrm{t})$ is

$$\mathrm{E}\{f_*|\boldsymbol{y}, \tilde{\boldsymbol{x}}_\mathrm{t}\} = \int_{\mathbb{R}} f_* p(f_*|\boldsymbol{y}, \tilde{\boldsymbol{x}}_\mathrm{t})\, df_*\ . \tag{2.12}$$

By inserting (2.10) into (2.12) and then rearranging the integrals we obtain

$$\begin{aligned} \mathrm{E}\{f_*|\boldsymbol{y}, \tilde{\boldsymbol{x}}_\mathrm{t}\} &= \frac{\int_{\mathbb{R}} f_* (\int_{\mathbb{R}^{DI}} p(f_*|\boldsymbol{y}, \boldsymbol{x}_\mathrm{t}) p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t}) p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})\, d\boldsymbol{x}_\mathrm{t})\, df_*}{\int_{\mathbb{R}^{DI}} p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t}) p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})\, d\boldsymbol{x}_\mathrm{t}} \\ &= \frac{\int_{\mathbb{R}^{DI}} (\int_{\mathbb{R}} f_* p(f_*|\boldsymbol{y}, \boldsymbol{x}_\mathrm{t})\, df_*) p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t}) p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})\, d\boldsymbol{x}_\mathrm{t}}{\int_{\mathbb{R}^{DI}} p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t}) p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})\, d\boldsymbol{x}_\mathrm{t}}\ . \end{aligned} \tag{2.13}$$

The inner integral in the numerator is the posterior mean of f_* conditioned on the vector of training positions $\boldsymbol{x}_\mathrm{t}$ is given by

$$\mu_*(\boldsymbol{x}_\mathrm{t}) \triangleq \mathrm{E}\{f_*|\boldsymbol{y}, \boldsymbol{x}_\mathrm{t}\} = \int_{\mathbb{R}} f_* p(f_*|\boldsymbol{y}, \boldsymbol{x}_\mathrm{t})\, df_*\ , \tag{2.14}$$

which is a standard GPR posterior mean value according to (1.21). After plugging that into (2.13) we get

$$\mathrm{E}\{f_*|\boldsymbol{y}, \tilde{\boldsymbol{x}}_\mathrm{t}\} = \frac{\int_{\mathbb{R}^{DI}} \mu_*(\boldsymbol{x}_\mathrm{t}) p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t}) p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})\, d\boldsymbol{x}_\mathrm{t}}{\int_{\mathbb{R}^{DI}} p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t}) p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})\, d\boldsymbol{x}_\mathrm{t}}\ . \tag{2.15}$$

Posterior variance

To express the variance of the posterior pdf, we first consider

$$\mathrm{E}\{f_*^2|\boldsymbol{y}, \tilde{\boldsymbol{x}}_\mathrm{t}\} = \int_{\mathbb{R}} f_*^2 p(f_*|\boldsymbol{y}, \tilde{\boldsymbol{x}}_\mathrm{t})\, df_*\ , \tag{2.16}$$

which after inserting (2.10) and then rearranging the integrals becomes

$$\begin{aligned}
\mathrm{E}\{f_*^2|\boldsymbol{y},\tilde{\boldsymbol{x}}_\mathrm{t}\} &= \frac{\int_{\mathbb{R}} f_*^2 (\int_{\mathbb{R}^{DI}} p(f_*|\boldsymbol{y},\boldsymbol{x}_\mathrm{t}) p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t}) p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})\, d\boldsymbol{x}_\mathrm{t})\, df_*}{\int_{\mathbb{R}^{DI}} p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t}) p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})\, d\boldsymbol{x}_\mathrm{t}} \\
&= \frac{\int_{\mathbb{R}^{DI}} (\int_{\mathbb{R}} f_*^2 p(f_*|\boldsymbol{y},\boldsymbol{x}_\mathrm{t})\, df_*) p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t}) p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})\, d\boldsymbol{x}_\mathrm{t}}{\int_{\mathbb{R}^{DI}} p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t}) p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})\, d\boldsymbol{x}_\mathrm{t}} \;. \qquad (2.17)
\end{aligned}$$

The inner integral in the numerator is the posterior mean of f_*^2 conditioned on the vector of training positions $\boldsymbol{x}_\mathrm{t}$ as

$$\mathrm{E}\{f_*^2|\boldsymbol{y},\boldsymbol{x}_\mathrm{t}\} = \int_{\mathbb{R}} f_*^2 p(f_*|\boldsymbol{y},\boldsymbol{x}_\mathrm{t})\, df_* = \mu_*^2(\boldsymbol{x}_\mathrm{t}) + \sigma_*^2(\boldsymbol{x}_\mathrm{t}) \;, \qquad (2.18)$$

where $\mu_*(\boldsymbol{x}_\mathrm{t})$ was defined in (2.14) and

$$\sigma_*^2(\boldsymbol{x}_\mathrm{t}) \triangleq \mathrm{var}\{f_*|\boldsymbol{y},\boldsymbol{x}_\mathrm{t}\} \;. \qquad (2.19)$$

Note that, similarly to $\mu_*(\boldsymbol{x}_\mathrm{t})$, $\sigma_*^2(\boldsymbol{x}_\mathrm{t})$ is a standard GPR posterior variance according to (1.22). After plugging (2.18) into (2.17), we get

$$\begin{aligned}
\mathrm{E}\{f_*^2|\boldsymbol{y},\tilde{\boldsymbol{x}}_\mathrm{t}\} &= \frac{\int_{\mathbb{R}^{DI}} \mathrm{E}\{f_*^2|\boldsymbol{y},\boldsymbol{x}'_\mathrm{t}\} p(\boldsymbol{y}|\boldsymbol{x}'_\mathrm{t}) p(\boldsymbol{x}'_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})\, d\boldsymbol{x}'_\mathrm{t}}{\int_{\mathbb{R}^{DI}} p(\boldsymbol{y}|\boldsymbol{x}'_\mathrm{t}) p(\boldsymbol{x}'_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})\, d\boldsymbol{x}'_\mathrm{t}} \\
&= \frac{\int_{\mathbb{R}^{DI}} (\mu_*^2(\boldsymbol{x}'_\mathrm{t}) + \sigma_*^2(\boldsymbol{x}'_\mathrm{t})) p(\boldsymbol{y}|\boldsymbol{x}'_\mathrm{t}) p(\boldsymbol{x}'_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})\, d\boldsymbol{x}'_\mathrm{t}}{\int_{\mathbb{R}^{DI}} p(\boldsymbol{y}|\boldsymbol{x}'_\mathrm{t}) p(\boldsymbol{x}'_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})\, d\boldsymbol{x}'_\mathrm{t}} \;. \qquad (2.20)
\end{aligned}$$

Now expressing the variance of the posterior distribution as

$$\mathrm{var}\{f_*|\boldsymbol{y},\tilde{\boldsymbol{x}}_\mathrm{t}\} = \mathrm{E}\{(f_* - \mathrm{E}\{f_*|\boldsymbol{y},\tilde{\boldsymbol{x}}_\mathrm{t}\})^2|\boldsymbol{y},\tilde{\boldsymbol{x}}_\mathrm{t}\} = \mathrm{E}\{f_*^2|\boldsymbol{y},\tilde{\boldsymbol{x}}_\mathrm{t}\} - (\mathrm{E}\{f_*|\boldsymbol{y},\tilde{\boldsymbol{x}}_\mathrm{t}\})^2 \quad (2.21)$$

and by further inserting (2.20), we obtain

$$\mathrm{var}\{f_*|\boldsymbol{y},\tilde{\boldsymbol{x}}_\mathrm{t}\} = \frac{\int_{\mathbb{R}^{DI}} (\mu_*^2(\boldsymbol{x}'_\mathrm{t}) + \sigma_*^2(\boldsymbol{x}'_\mathrm{t})) p(\boldsymbol{y}|\boldsymbol{x}'_\mathrm{t}) p(\boldsymbol{x}'_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})\, d\boldsymbol{x}'_\mathrm{t}}{\int_{\mathbb{R}^{DI}} p(\boldsymbol{y}|\boldsymbol{x}'_\mathrm{t}) p(\boldsymbol{x}'_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})\, d\boldsymbol{x}'_\mathrm{t}} - (\mathrm{E}\{f_*|\boldsymbol{y},\tilde{\boldsymbol{x}}_\mathrm{t}\})^2 \;, \qquad (2.22)$$

where $\mathrm{E}\{f_*|\boldsymbol{y},\tilde{\boldsymbol{x}}_\mathrm{t}\}$ is already known from (2.15).

2.2 Monte Carlo Evaluation

The integrals in the expressions for the posterior mean (2.15) and the posterior variance (2.22) cannot be computed in closed form in general. In our, simulations we will therefore need an approximation method to compute them. One such method is the *Monte Carlo (MC) evaluation* as explained in [8, sec. 2.1]. To explain the basic idea, we will first consider the approximation of the expected value of some scalar function $\phi(q) \in \mathbb{R}$ of a random variable $q \in \mathbb{R}$ with respect to a pdf $p(q)$,

$$\mathrm{E}^{p(q)}\{\phi(q)\} = \int_{\mathbb{R}} \phi(q) p(q)\, dq \;, \qquad (2.23)$$

using a point-mass function

$$p_{\mathrm{MC}}(q) \triangleq \frac{1}{s}\sum_{i=1}^{s}\delta(q-q_i)\ , \tag{2.24}$$

to approximate $p(q)$. Here $q_i, i = 1, \ldots, s$ is a set of iid random samples drawn according to the pdf $p(q)$ and $\delta(q)$ is the *Dirac delta function.* Using $p_{\mathrm{MC}}(q)$ as an approximation of $p(q)$, we can express the approximate expectation of $\phi(q)$ by inserting (2.24) into (2.23) as

$$\begin{aligned}
\mathrm{E}^{p_{\mathrm{MC}}(q)}\{\phi(q)\} &\triangleq \int_{\mathbb{R}} \phi(q)p_{\mathrm{MC}}(q)\,dq \\
&= \int_{\mathbb{R}} \phi(q)\frac{1}{s}\sum_{i=1}^{s}\delta(q-q_i)\,dq \\
&= \frac{1}{s}\sum_{i=1}^{s}\int_{\mathbb{R}} \phi(q)\delta(q-q_i)\,dq \\
&= \frac{1}{s}\sum_{i=1}^{s}\phi(q_i)\ .
\end{aligned} \tag{2.25}$$

This approximate expectation can be shown under mild conditions [8, Sec.2.1] to converge to the true expectation as N goes to infinity, i.e.,

$$\lim_{s\to\infty} \mathrm{E}^{p_{\mathrm{MC}}(q)}\{\phi(q)\} = \lim_{s\to\infty}\frac{1}{s}\sum_{i=1}^{s}\phi(q_i) = \mathrm{E}^{p(q)}\{\phi(q)\} = \int_{\mathbb{R}}\phi(q)p(q)\,dq\ . \tag{2.26}$$

$\mathrm{E}^{p_{\mathrm{MC}}(q)}\{\phi(q)\}$ can also be shown to be an unbiased estimator of $\phi(q)$ even for finite s. If we further assume the variance of the random variable $\phi(q)$, i.e.,

$$\mathrm{var}^{p(q)}\{\phi(q)\} = \mathrm{E}^{p(q)}\{\phi^2(q)\} - \mathrm{E}^{p(q)}\{\phi(q)\}^2 < \infty \tag{2.27}$$

then the variance of the MC approximation can be shown to decrease with increasing s as

$$\mathrm{var}^{p(q)}\{\mathrm{E}^{p_{\mathrm{MC}}(q)}\{\phi(q)\}\} = \frac{\mathrm{var}^{p(q)}\{\phi(q)\}}{s}\ . \tag{2.28}$$

Monte Carlo approximation of the posterior mean

Let us return to our original problem of approximating the posterior mean $\mathrm{E}\{f_*|\boldsymbol{y},\tilde{\boldsymbol{x}}_{\mathrm{t}}\}$ in (2.15). This mean can be reconsidered as the ratio of two expectations with respect to the pdf $p(\boldsymbol{x}_{\mathrm{t}}|\tilde{\boldsymbol{x}}_{\mathrm{t}})$,

$$\mathrm{E}\{f_*|\boldsymbol{y},\tilde{\boldsymbol{x}}_{\mathrm{t}}\} = \frac{\int_{\mathbb{R}^{DI}}\mu_*(\boldsymbol{x}_{\mathrm{t}})p(\boldsymbol{y}|\boldsymbol{x}_{\mathrm{t}})p(\boldsymbol{x}_{\mathrm{t}}|\tilde{\boldsymbol{x}}_{\mathrm{t}})\,d\boldsymbol{x}_{\mathrm{t}}}{\int_{\mathbb{R}^{DI}}p(\boldsymbol{y}|\boldsymbol{x}_{\mathrm{t}})p(\boldsymbol{x}_{\mathrm{t}}|\tilde{\boldsymbol{x}}_{\mathrm{t}})\,d\boldsymbol{x}_{\mathrm{t}}} = \frac{\mathrm{E}^{p(\boldsymbol{x}_{\mathrm{t}}|\tilde{\boldsymbol{x}}_{\mathrm{t}})}\{\mu_*(\boldsymbol{x}_{\mathrm{t}})p(\boldsymbol{y}|\boldsymbol{x}_{\mathrm{t}})\}}{\mathrm{E}^{p(\boldsymbol{x}_{\mathrm{t}}|\tilde{\boldsymbol{x}}_{\mathrm{t}})}\{p(\boldsymbol{y}|\boldsymbol{x}_{\mathrm{t}})\}}\ . \tag{2.29}$$

This corresponds to our MC evaluation example (2.23): $p(q)$ corresponds to $p(\boldsymbol{x}_{\mathrm{t}}|\tilde{\boldsymbol{x}}_{\mathrm{t}})$ and $\phi(q)$ corresponds to $\mu_*(\boldsymbol{x}_{\mathrm{t}})p(\boldsymbol{y}|\boldsymbol{x}_{\mathrm{t}})$ in the numerator of (2.29) and to $p(\boldsymbol{y}|\boldsymbol{x}_{\mathrm{t}})$ in the denominator of (2.29).

Accordingly, we will approximate $p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})$ with a point-mass function

$$p_\mathrm{MC}(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t}) = \frac{1}{s}\sum_{i=1}^{s}\delta(\boldsymbol{x}_\mathrm{t} - \boldsymbol{x}_{\mathrm{t},i}) \tag{2.30}$$

using a set of samples $\boldsymbol{x}_{\mathrm{t},i}, i = 1,\ldots,s$ drawn according to $p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})$. Then, we can approximate the posterior mean $\mathrm{E}\{f_*|\boldsymbol{y},\tilde{\boldsymbol{x}}_\mathrm{t}\}$ in (2.29) using the approximation relation (2.25) as

$$\begin{aligned}\mathrm{E}^\mathrm{MC}\{f_*|\boldsymbol{y},\tilde{\boldsymbol{x}}_\mathrm{t}\} &\triangleq \frac{\mathrm{E}^{p_\mathrm{MC}(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})}\{\mu_*(\boldsymbol{x}_\mathrm{t})p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t})\}}{\mathrm{E}^{p_\mathrm{MC}(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})}\{p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t})\}} \\ &= \frac{\frac{1}{s}\sum_{i=1}^{s}\mu_*(\boldsymbol{x}_{\mathrm{t},i})p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t} = \boldsymbol{x}_{\mathrm{t},i})}{\frac{1}{s}\sum_{i=1}^{s}p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t} = \boldsymbol{x}_{\mathrm{t},i})} \\ &= \frac{\sum_{i=1}^{s}\mu_*(\boldsymbol{x}_{\mathrm{t},i})p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t} = \boldsymbol{x}_{\mathrm{t},i})}{\sum_{i=1}^{s}p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t} = \boldsymbol{x}_{\mathrm{t},i})}\,. \end{aligned}\tag{2.31}$$

All the terms in (2.31) can be calculated using the standard GPR expressions with known training positions, i.e., $\mu_*(\boldsymbol{x}_{\mathrm{t},i})$ can be calculated according to (1.21) and $p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t} = \boldsymbol{x}_{\mathrm{t},i})$ according to (1.12).

Monte Carlo approximation of the posterior variance

Similarly to the posterior mean, we will also approximate the posterior variance $\mathrm{var}\{f_*|\boldsymbol{y},\tilde{\boldsymbol{x}}_\mathrm{t}\}$ in (2.22). We express the left term on the right-hand side of (2.22) as the ratio of two expectations with respect to the probability density $p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})$, i.e.,

$$\begin{aligned}\mathrm{var}\{f_*|\boldsymbol{y},\tilde{\boldsymbol{x}}_\mathrm{t}\} &= \frac{\int_{\mathbb{R}^{DI}}(\mu_*^2(\boldsymbol{x}_\mathrm{t}) + \sigma_*^2(\boldsymbol{x}_\mathrm{t}))p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t})p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})\,d\boldsymbol{x}_\mathrm{t}}{\int_{\mathbb{R}^{DI}}p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t})p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})\,d\boldsymbol{x}_\mathrm{t}} - (\mathrm{E}\{f_*|\boldsymbol{y},\tilde{\boldsymbol{x}}_\mathrm{t}\})^2 \\ &= \frac{\mathrm{E}^{p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})}\{(\mu_*^2(\boldsymbol{x}_\mathrm{t}) + \sigma_*^2(\boldsymbol{x}_\mathrm{t}))p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t})\}}{\mathrm{E}^{p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})}\{p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t})\}} - (\mathrm{E}\{f_*|\boldsymbol{y},\tilde{\boldsymbol{x}}_\mathrm{t}\})^2\,. \end{aligned}\tag{2.32}$$

This again corresponds to our MC evaluation example (2.23): $p(q)$ again corresponds to $p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})$, and $\phi(q)$ corresponds to $(\mu_*^2(\boldsymbol{x}_\mathrm{t}) + \sigma_*^2(\boldsymbol{x}_\mathrm{t}))p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t})$ in the numerator of (2.32) and to $p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})$ in the denominator of (2.32).

accordingly we again approximate $p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})$ with the point-mass function $p_\mathrm{MC}(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})$ in (2.30). Then, using the approximation relation (2.25), we obtain the following approximation of the posterior variance in (2.32) as

$$\begin{aligned}\mathrm{var}^\mathrm{MC}\{f_*|\boldsymbol{y},\tilde{\boldsymbol{x}}_\mathrm{t}\} &\triangleq \frac{\mathrm{E}^{p_\mathrm{MC}(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})}\{(\mu_*^2(\boldsymbol{x}_\mathrm{t}) + \sigma_*^2(\boldsymbol{x}_\mathrm{t}))p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t})\}}{\mathrm{E}^{p_\mathrm{MC}(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})}\{p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t})\}} - (\mathrm{E}^\mathrm{MC}\{f_*|\boldsymbol{y},\tilde{\boldsymbol{x}}_\mathrm{t}\})^2 \\ &= \frac{\frac{1}{s}\sum_{i=1}^{s}(\mu_*^2(\boldsymbol{x}_{\mathrm{t},i}) + \sigma_*^2(\boldsymbol{x}_{\mathrm{t},i}))p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t} = \boldsymbol{x}_{\mathrm{t},i})}{\frac{1}{s}\sum_{i=1}^{s}p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t} = \boldsymbol{x}_{\mathrm{t},i})} - (\mathrm{E}^\mathrm{MC}\{f_*|\boldsymbol{y},\tilde{\boldsymbol{x}}_\mathrm{t}\})^2 \\ &= \frac{\sum_{i=1}^{s}(\mu_*^2(\boldsymbol{x}_{\mathrm{t},i}) + \sigma_*^2(\boldsymbol{x}_{\mathrm{t},i}))p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t} = \boldsymbol{x}_{\mathrm{t},i})}{\sum_{i=1}^{s}p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t} = \boldsymbol{x}_{\mathrm{t},i})} - (\mathrm{E}^\mathrm{MC}\{f_*|\boldsymbol{y},\tilde{\boldsymbol{x}}_\mathrm{t}\})^2\,, \end{aligned}\tag{2.33}$$

where we replaced $\mathrm{E}\{f_*|\boldsymbol{y}, \tilde{\boldsymbol{x}}_\mathrm{t}\}$ by its MC approximation $\mathrm{E}^{\mathrm{MC}}\{f_*|\boldsymbol{y}, \tilde{\boldsymbol{x}}_\mathrm{t}\}$ as given by (2.31). Here again, all the terms in (2.33) can be calculated using the standard GPR expressions with known training positions. We note that in [2], identical results were derived using the *importance sampling* method. The derivation from [2] is reviewed in Appendix C.

2.3 Sampling from the Posterior Position Distribution

In order to use the MC approximation $\mathrm{E}^{\mathrm{MC}}\{f_*|\boldsymbol{y}, \tilde{\boldsymbol{x}}_\mathrm{t}\}$ in (2.31) and $\mathrm{var}^{\mathrm{MC}}\{f_*|\boldsymbol{y}, \tilde{\boldsymbol{x}}_\mathrm{t}\}$ in (2.33) we have to sample from the posterior distribution of training positions $p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})$. According to (2.7), we have

$$p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t}) \propto p(\tilde{\boldsymbol{x}}_\mathrm{t}|\boldsymbol{x}_\mathrm{t})p(\boldsymbol{x}_\mathrm{t}) \ , \tag{2.34}$$

where the denominator in (2.7) has been disregarded because it does not depend on $\boldsymbol{x}_\mathrm{t}$. The likelihood function $p(\tilde{\boldsymbol{x}}_\mathrm{t}|\boldsymbol{x}_\mathrm{t})$ is given by (2.6), i.e.,

$$p(\tilde{\boldsymbol{x}}_\mathrm{t}|\boldsymbol{x}_\mathrm{t}) = p_{\boldsymbol{w}}(\tilde{\boldsymbol{x}}_\mathrm{t} - \boldsymbol{x}_\mathrm{t}) \ . \tag{2.35}$$

We assume the observation noise pdf $p_{\boldsymbol{w}}(\cdot)$ is zero-mean isotropic Gaussian with a known variance σ_v^2,

$$p(\boldsymbol{w}) = \mathcal{N}(\boldsymbol{w};\ \mathbf{0},\ \sigma_\mathrm{v}^2\mathbf{I}_{DI}) \ , \tag{2.36}$$

which implies that the individual noise components for the same training position as well as the noises for different training positions are independent. It follows that the distribution of the noise for one (dth) spatial coordinate of one training position is

$$p(w_d^{(i)}) = \mathcal{N}(w_d^{(i)};\ 0,\ \sigma_\mathrm{v}^2) \,, \qquad d = 1, 2, \ldots, D \,, \qquad i = 1, 2, \ldots, I \ . \tag{2.37}$$

With (2.5) the likelihood for a single spatial coordinate of a single training position is obtained as

$$p(\tilde{x}_d^{(i)}|x_d^{(i)}) = \mathcal{N}(\tilde{x}_d^{(i)};\ x_d^{(i)},\ \sigma_\mathrm{v}^2) \ . \tag{2.38}$$

According to (2.34), we also need to specify the prior distribution of the training positions, $p(\boldsymbol{x}_\mathrm{t})$. There are a few reasonable ways to express the prior, namely considering improper prior, Gaussian prior or uniform prior. Further we shall consider the individual true training positions are iid resulting in We assume that the individual training positions are iid, i.e.,

$$p(\boldsymbol{x}_\mathrm{t}) = \prod_{i=1}^{I} p(\boldsymbol{x}^{(i)}) \ . \tag{2.39}$$

This means that we can generate the samples of the individual training positions $\boldsymbol{x}^{(i)}$ separately. Moreover, we assume that the individual spatial coordinates of each training position are independent but not necessarily identically distributed, i.e., for $D = 2$,

$$p(\boldsymbol{x}^{(i)}) = p(x_1^{(i)})p(x_2^{(i)}) \ . \tag{2.40}$$

By inserting (2.40) into (2.39) we obtain

$$p(\boldsymbol{x}_{\mathrm{t}}) = \prod_{i=1}^{I} p(x_1^{(i)})p(x_2^{(i)}) \ . \tag{2.41}$$

This means that we can draw samples for each training position dimension separately. The posterior pdf of each spatial dimension of each training position is given by (cf. (2.34))

$$p(x_d^{(i)}|\tilde{x}_d^{(i)}) \propto p(\tilde{x}_d^{(i)}|x_d^{(i)})p(x_d^{(i)}) \,, \qquad d = 1, 2, \ldots, D \ , \tag{2.42}$$

with $p(\tilde{x}_d^{(i)}|x_d^{(i)})$ given by (2.38). Next, we shall consider three different choices of the prior distribution $p(x_d^{(i)})$ used to obtain the posterior distribution in (2.42) in order to be able to draw samples from it.

Improper prior

Our first choice is the improper prior, which expresses our lack of prior knowledge about $x_d^{(i)}$. It is given by

$$p_{\mathrm{I}}(x_d^{(i)}) = 1 \ , \tag{2.43}$$

which is not a valid probability distribution as it does not integrate to one. The joint prior distribution of all the training positions is then obtained from (2.41) as

$$p_{\mathrm{I}}(\boldsymbol{x}_{\mathrm{t}}) = \prod_{i=1}^{I} p(x_1^{(i)})p(x_2^{(i)}) = 1 \ . \tag{2.44}$$

Inserting this into (2.42) and using (2.38) yields for the posterior distribution

$$\begin{aligned} p_{\mathrm{I}}(x_d^{(i)}|\tilde{x}_d^{(i)}) &\propto p(\tilde{x}_d^{(i)}|x_d^{(i)})p_{\mathrm{I}}(x_d^{(i)}) \\ &\propto p(\tilde{x}_d^{(i)}|x_d^{(i)}) \\ &\propto (2\pi\sigma_{\mathrm{v}}^2)^{-\frac{1}{2}}\exp\left(-\frac{(\tilde{x}_d^{(i)} - x_d^{(i)})^2}{2\sigma_{\mathrm{v}}^2}\right) \ , \end{aligned} \tag{2.45}$$

which is recognized to be a Gaussian probability distribution for the random variable $x_d^{(i)}$. Therefore we can replace the proportionality with equality and get

$$p_{\mathrm{I}}(x_d^{(i)}|\tilde{x}_d^{(i)}) = (2\pi\sigma_{\mathrm{v}}^2)^{-\frac{1}{2}}\exp\left(-\frac{(x_d^{(i)} - \tilde{x}_d^{(i)})^2}{2\sigma_{\mathrm{v}}^2}\right) = \mathcal{N}(x_d^{(i)};\, \tilde{x}_d^{(i)},\, \sigma_{\mathrm{v}}^2) \ . \tag{2.46}$$

This is a valid distribution, from which we can draw samples using methods implemented in many programming and simulation environments. We also note that using the improper prior amounts to a classical (non-Bayesian) estimation framework because the prior distribution is totally noninformative.

Gaussian prior

Another option for the prior distribution of $x_d^{(i)}$, as suggested in [2], is the Gaussian distribution

$$p_{\mathrm{G}}(x_d^{(i)}) = \mathcal{N}(x_d^{(i)};\ \mu_d,\ \sigma_{\mathrm{x}}^2)\ . \tag{2.47}$$

Here, the mean μ_d may depend on the dimension index d whereas the variance σ_{x}^2 is fixed Inserting (2.47) and (2.38) into (2.42), we get for the posterior distribution

$$\begin{aligned}
p_{\mathrm{G}}(x_d^{(i)}|\tilde{x}_d^{(i)}) &\propto p(\tilde{x}_d^{(i)}|x_d^{(i)})p_{\mathrm{G}}(x_d^{(i)}) \\
&\propto (2\pi\sigma_{\mathrm{v}}^2)^{-\frac{1}{2}}\exp\left(-\frac{(\tilde{x}_d^{(i)}-x_d^{(i)})^2}{2\sigma_{\mathrm{v}}^2}\right)(2\pi\sigma_{\mathrm{x}}^2)^{-\frac{1}{2}}\exp\left(-\frac{(x_d^{(i)}-\mu_d)^2}{2\sigma_{\mathrm{x}}^2}\right) \\
&\propto \exp\left(-\frac{(\tilde{x}_d^{(i)}-x_d^{(i)})^2}{2\sigma_{\mathrm{v}}^2}-\frac{(x_d^{(i)}-\mu_d)^2}{2\sigma_{\mathrm{x}}^2}\right) \\
&\propto \exp\left(-\left(\frac{1}{2\sigma_{\mathrm{v}}^2}+\frac{1}{2\sigma_{\mathrm{x}}^2}\right)(x_d^{(i)})^2+\left(\frac{\tilde{x}_d^{(i)}}{\sigma_{\mathrm{v}}^2}+\frac{\mu_d}{\sigma_{\mathrm{x}}^2}\right)x_d^{(i)}-\left(\frac{(\tilde{x}_d^{(i)})^2}{2\sigma_{\mathrm{v}}^2}+\frac{\mu_d^2}{2\sigma_{\mathrm{x}}^2}\right)\right)\ .
\end{aligned} \tag{2.48}$$

Noticing that the exponent is a mixed quadratic-linear-constant function of $x_d^{(i)}$, we recognize this distribution to be Gaussian. To obtain the mean $\mu_{d,\mathrm{post}}$ and variance $\sigma_{\mathrm{x,post}}^2$ of this Gaussian distribution, we employ the *competing the square* method. We formally set

$$\begin{aligned}
p_{\mathrm{G}}(x_d^{(i)}|\tilde{x}_d^{(i)}) &= \mathcal{N}(x_d^{(i)};\ \mu_{d,\mathrm{post}},\ \sigma_{\mathrm{x,post}}^2) \\
&\propto \exp\left(-\frac{(x_d^{(i)}-\mu_{d,\mathrm{post}})^2}{2\sigma_{\mathrm{x,post}}^2}\right) \\
&\propto \exp\left(-\frac{1}{2\sigma_{\mathrm{x,post}}^2}(x_d^{(i)})^2+\frac{\mu_{d,\mathrm{post}}}{\sigma_{\mathrm{x,post}}^2}x_d^{(i)}-\frac{\mu_{d,\mathrm{post}}^2}{2\sigma_{\mathrm{x,post}}^2}\right)\ .
\end{aligned} \tag{2.49}$$

Comparing the quadratic terms in equations (2.48) and (2.49), we obtain

$$\frac{1}{2\sigma_{\mathrm{x,post}}^2} = \frac{1}{2\sigma_{\mathrm{v}}^2}+\frac{1}{2\sigma_{\mathrm{x}}^2} \tag{2.50}$$

and in turn

$$\sigma_{\mathrm{x,post}}^2 = \frac{1}{\dfrac{1}{\sigma_{\mathrm{v}}^2}+\dfrac{1}{\sigma_{\mathrm{x}}^2}}\ . \tag{2.51}$$

Similarly, comparing the linear terms in equations (2.48) and (2.49), we obtain

$$\frac{\mu_{d,\text{post}}}{\sigma^2_{\text{x,post}}} = \frac{\tilde{x}_d^{(i)}}{\sigma^2_{\text{v}}} + \frac{\mu_d}{\sigma^2_{\text{x}}} \tag{2.52}$$

and thus

$$\mu_{d,\text{post}} = \sigma^2_{\text{x,post}} \left(\frac{\tilde{x}_d^{(i)}}{\sigma^2_{\text{v}}} + \frac{\mu_d}{\sigma^2_{\text{x}}} \right) . \tag{2.53}$$

Based on these expressions of the Gaussian distribution parameters $\mu_{d,\text{post}}$ and $\sigma^2_{\text{x,post}}$, we can sample from the posterior distribution

$$p_{\text{G}}(x_d^{(i)} | \tilde{x}_d^{(i)}) = \mathcal{N}(x_d^{(i)};\ \mu_{d,\text{post}},\ \sigma^2_{\text{x,post}}) \tag{2.54}$$

for each spatial dimension d.

Uniform prior

In our simulations, we will assume that the support of $\boldsymbol{x}_d^{(i)}$ is a subspace $\mathcal{X}_d \subset \mathbb{R}$. Therefore, the pdf of $x_d^{(i)}$ is zero outside $\mathcal{X}_d$. Moreover, having no further prior knowledge, we assume the distribution of $x_d^{(i)}$ to be uniform within $\mathcal{X}_d$, i.e.,

$$p_{\text{U}}(x_d^{(i)}) = \mathcal{U}(\mathcal{X}_d) = \begin{cases} \frac{1}{|\mathcal{X}_d|} & \text{for } x_d^{(i)} \in \mathcal{X}_d \\ 0 & \text{for } x_d^{(i)} \notin \mathcal{X}_d \ , \end{cases} \tag{2.55}$$

where $|\mathcal{X}_d| = \int_{\mathcal{X}_d} dx_d^{(i)}$. By plugging this position prior into (2.42) and using (2.38), we obtain the posterior pdf

$$\begin{aligned} p_{\text{U}}(x_d^{(i)} | \tilde{x}_d^{(i)}) &\propto p(\tilde{x}_d^{(i)} | x_d^{(i)}) p_{\text{U}}(x_d^{(i)}) \\ &\propto \begin{cases} (2\pi\sigma^2_{\text{v}})^{-\frac{1}{2}} \exp\left(-\frac{(\tilde{x}_d^{(i)} - x_d^{(i)})^2}{2\sigma^2_{\text{v}}} \right) \frac{1}{|\mathcal{X}_d|} & \text{for } x_d^{(i)} \in \mathcal{X}_d \\ 0 & \text{for } x_d^{(i)} \notin \mathcal{X}_d \end{cases} \\ &\propto \begin{cases} \exp\left(-\frac{(x_d^{(i)} - \tilde{x}_d^{(i)})^2}{2\sigma^2_{\text{v}}} \right) & \text{for } x_d^{(i)} \in \mathcal{X}_d \\ 0 & \text{for } x_d^{(i)} \notin \mathcal{X}_d \ . \end{cases} \end{aligned} \tag{2.56}$$

Samples can be drawn from this distribution by first drawing samples from the Gaussian distribution and then rejecting those samples that are outside $\mathcal{X}_d$. This is illustrated for a one dimensional training position sample generation in Figure 2.1.

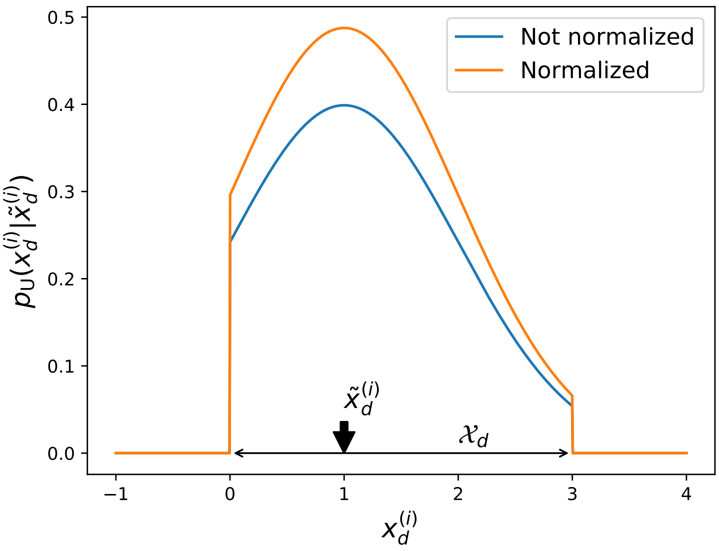


Fig. 2.1: Posterior position pdf $p_U(x_d^{(i)}|\tilde{x}_d^{(i)})$, for a uniform prior, with support $|\mathcal{X}_d| = [0, 3]$ for one spatial coordinate.

2.4 Simulation

We simulated the method of GPR operating under training positions uncertainty, as derived in previous sections of this chapter. For the simulation we used a similar setup as in [2]. The Gaussian process was considered to be zero-mean, i.e., $m(\boldsymbol{x}) = 0$, with squared exponential covariance function

$$k(\boldsymbol{x}, \boldsymbol{x}') = \sigma^2 \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{x}'\|^2}{2\sigma_x^2}\right) . \tag{2.57}$$

Here, $\sigma^2 = 2$ is the variance of each GP random variable and $\sigma_x = \sqrt{2}$ is the spatial covariance length scale. This covariance function is positive definite [1, p. 86] and is known to result in a positive definite matrix $\boldsymbol{K}$ according to (1.8). The complete GP can be expressed similarly to (1.3) as

$$f(\boldsymbol{x}) \sim \mathcal{GP}(0, k(\boldsymbol{x}, \boldsymbol{x}')) . \tag{2.58}$$

We performed the GPR on a subset $\mathcal{X}_1 \times \mathcal{X}_2 \subset \mathbb{R}^2$, where $\mathcal{X}_1 = [0, 6]$ and $\mathcal{X}_2 = [0, 3]$. We randomly generated $I = 20$ random training positions $\boldsymbol{x}^{(i)}$, $i = 1, \ldots, I$ using a uniform distribution according to (2.55), i.e., $p(x_1^{(i)}) = \mathcal{U}(\mathcal{X}_1)$ and $p(x_2^{(i)}) = \mathcal{U}(\mathcal{X}_2)$. We can sample from these distributions using common packages in most programming environments, e.g. package *NumPy* in PYTHON as used for our simulations. We sample from this uniform distribution to obtain $\boldsymbol{x}_{t,s}$, which will be further considered to be the realization of the vector of random training positions $\boldsymbol{x}_t$ and also referred to as the vector of the *true training positions* $\boldsymbol{x}_{t,s}$. The vector

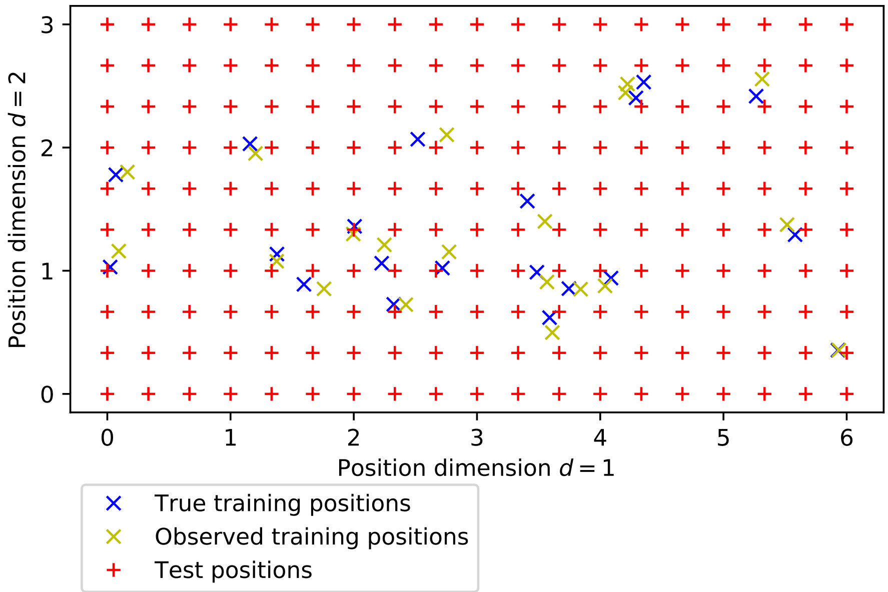


Fig. 2.2: Example of the region of interest along with the true training positions (blue), the observed training positions (yellow) and the test positions (red).

of the true training positions is created by stacking the individual training position samples in single dimension according to

$$\boldsymbol{x}_{\mathrm{t}} = \mathrm{col}(x_1^{(1)}, x_2^{(1)}, x_1^{(2)}, \ldots, x_2^{(I)}) \in (\mathcal{X}_1 \times \mathcal{X}_2)^I \subset \mathbb{R}^{2I} \ . \tag{2.59}$$

The random training positions observation $\tilde{\boldsymbol{x}}_{\mathrm{t}}$ is then generated by adding Gaussian noise to the true training positions $\boldsymbol{x}_{\mathrm{t,s}}$ according to (2.38) as

$$p(\tilde{\boldsymbol{x}}_{\mathrm{t}}|\boldsymbol{x}_{\mathrm{t}} = \boldsymbol{x}_{\mathrm{t,s}}) = \mathcal{N}(\tilde{\boldsymbol{x}}_{\mathrm{t}};\ \boldsymbol{x}_{\mathrm{t,s}},\ \sigma_{\mathrm{v}}^2 \mathbf{I}_{DI}) \ , \tag{2.60}$$

where the observation noise standard deviation is chosen as $\sigma_{\mathrm{v}} = \sqrt{0.1}$. From this conditional distribution we draw a sample $\tilde{\boldsymbol{x}}_{\mathrm{t,s}}$ representing the known vector of training positions observation.

To observe the GP realization and GPR results we will consider a stacked vector of individual known test positions $\boldsymbol{x}_{\mathrm{test}}$ being in a rectangular grid covering the region of interest $\mathcal{X}$. The density of the test positions grid is 3 samples per spatial unit giving us

$$T = (6 \cdot 3 + 1)(3 \cdot 3 + 1) = 190 \tag{2.61}$$

test positions $\boldsymbol{x}_{\mathrm{test}}^{(j)}$ with $j = 1, \ldots, T$ stacked to a vector $\boldsymbol{x}_{\mathrm{test}} \in \mathbb{R}^{DT}$ similarly as for $\boldsymbol{x}_{\mathrm{t}}$ in (1.4). A sample of the vector of the true training positions $\boldsymbol{x}_{\mathrm{t,s}}$, observed training positions $\tilde{\boldsymbol{x}}_{\mathrm{t,s}}$ and test positions $\boldsymbol{x}_{\mathrm{test}}$ are visualized in figure 2.2.

Generating the GP realization

For simulation purposes we will be considering the vector of GP random variables similarly to (1.6) while now omitting the conditioning on the vector of random training positions $\boldsymbol{x}_{\mathrm{t}}$. This vector $\boldsymbol{f}$ is then distributed according to

$$p(\boldsymbol{f}) = \mathcal{N}(\boldsymbol{f};\, \boldsymbol{m},\, \boldsymbol{K}) \ , \tag{2.62}$$

with the difference to (1.6) that to the GP random variables at true training positions $\boldsymbol{x}_{\mathrm{t,s}}$ we include the test positions $\boldsymbol{x}_{\mathrm{test}}$. All these positions are stacked to a known vector

$$\boldsymbol{x}_{\mathrm{t,ext}} = \mathrm{col}(\boldsymbol{x}_{\mathrm{t,s}}, \boldsymbol{x}_{\mathrm{test}}) \in \mathbb{R}^{D(I+T)} \ . \tag{2.63}$$

This gives us the vector of GP random variables at $\boldsymbol{x}_{\mathrm{t,ext}}$ positions extended from (1.5) as

$$\boldsymbol{f}_{\mathrm{ext}} = \mathrm{col}(\boldsymbol{f}, \boldsymbol{f}_{\mathrm{test}}) \in \mathbb{R}^{I+T} \tag{2.64}$$

with GP random variables at training positions $\boldsymbol{x}_{\mathrm{t,s}}$

$$\boldsymbol{f} = (f(\boldsymbol{x}_{\mathrm{s}}^{(1)}) \quad f(\boldsymbol{x}_{\mathrm{s}}^{(2)}) \quad \dots \quad f(\boldsymbol{x}_{\mathrm{s}}^{(I)}))^{\mathrm{T}} \in \mathbb{R}^{I} \tag{2.65}$$

and GP random variables at test positions $\boldsymbol{x}_{\mathrm{test}}$

$$\boldsymbol{f}_{\mathrm{test}} = (f(\boldsymbol{x}_{\mathrm{test}}^{(1)}) \quad f(\boldsymbol{x}_{\mathrm{test}}^{(2)}) \quad \dots \quad f(\boldsymbol{x}_{\mathrm{test}}^{(T)}))^{\mathrm{T}} \in \mathbb{R}^{T} \ . \tag{2.66}$$

The extended vector of GP random variables $\boldsymbol{f}_{\mathrm{ext}}$ is then distributed according to

$$p(\boldsymbol{f}_{\mathrm{ext}}) = \mathcal{N}(\boldsymbol{f}_{\mathrm{ext}};\, \boldsymbol{m}_{\mathrm{ext}},\, \boldsymbol{K}_{\mathrm{ext}}) \tag{2.67}$$

with vector of GP mean values $\boldsymbol{m}_{\mathrm{ext}} \in \mathbb{R}^{I+T}$ similarly extended from (1.7) and covariance matrix $\boldsymbol{K}_{\mathrm{ext}} \in \mathbb{R}^{(I+T)\times(I+T)}$ extended from (1.8).

A realization of the GP random vector $\boldsymbol{f}_{\mathrm{ext}}$ at positions $\boldsymbol{x}_{\mathrm{t,ext}}$ is generated as suggested in [1, sec. A.2] using Cholesky decomposition $\boldsymbol{L}$ of the covariance matrix $\boldsymbol{K}_{\mathrm{ext}}$ as

$$\boldsymbol{L}\boldsymbol{L}^{\mathrm{T}} = \boldsymbol{K}_{\mathrm{ext}} + \epsilon \mathbf{I}_{D(I+T)} \ , \tag{2.68}$$

where the regularization term $\epsilon \mathbf{I}_{D(I+T)}$ with chosen small parameter $\epsilon = 10^{-3}$ ensures positive definiteness of the matrix on the right-hand side of (2.68) in contrast to the positive semidefinite matrix $\boldsymbol{K}_{\mathrm{ext}}$. This is required because Cholesky decomposition needs to operate on positive definite matrix. We will consider a random vector $\boldsymbol{u} \in \mathbb{R}^{(I+T)}$ distributed according to isotropic Gaussian as in

$$p(\boldsymbol{u}) = \mathcal{N}(\boldsymbol{u};\, \mathbf{0},\, \mathbf{I}_{(I+T)}) \ , \tag{2.69}$$

which we can obtain samples of using common packages in most programming environments, e.g. package *NumPy* in PYTHON as used for our simulations. The random vector $\boldsymbol{f}_{\text{ext}}$ is than related to the random vector $\boldsymbol{u}$ according to

$$\boldsymbol{f}_{\text{ext}} = \boldsymbol{m}_{\text{ext}} + \boldsymbol{L}\boldsymbol{u} \ . \tag{2.70}$$

The relation for $\boldsymbol{f}_{\text{ext}}$ can be shown to produce the desired distribution of $\boldsymbol{f}_{\text{ext}}$ as in (2.67) by firstly expressing its mean

$$\mathrm{E}^{p(\boldsymbol{u})}\{\boldsymbol{m}_{\text{ext}} + \boldsymbol{L}\boldsymbol{u}\} = \mathrm{E}^{p(\boldsymbol{u})}\{\boldsymbol{m}_{\text{ext}}\} + \boldsymbol{L}\mathrm{E}^{p(\boldsymbol{u})}\{\boldsymbol{u}\} = \boldsymbol{m}_{\text{ext}} \ , \tag{2.71}$$

where we used the independence of $\boldsymbol{m}_{\text{ext}}$ on $\boldsymbol{u}$ and that $\boldsymbol{u}$ is zero-mean. Further expressing the covariance

$$\begin{aligned}
\operatorname{cov}\{\boldsymbol{m}_{\text{ext}} + \boldsymbol{L}\boldsymbol{u}\} &= \mathrm{E}^{p(\boldsymbol{u})}\{(\boldsymbol{m}_{\text{ext}} + \boldsymbol{L}\boldsymbol{u} - \boldsymbol{m}_{\text{ext}})(\boldsymbol{m}_{\text{ext}} + \boldsymbol{L}\boldsymbol{u} - \boldsymbol{m}_{\text{ext}})^{\mathrm{T}}\} \\
&= \mathrm{E}^{p(\boldsymbol{u})}\{(\boldsymbol{L}\boldsymbol{u})(\boldsymbol{L}\boldsymbol{u})^{\mathrm{T}}\} \\
&= \mathrm{E}^{p(\boldsymbol{u})}\{\boldsymbol{L}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{L}^{\mathrm{T}}\} \\
&= \boldsymbol{L}\mathrm{E}^{p(\boldsymbol{u})}\{\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\}\boldsymbol{L}^{\mathrm{T}} \\
&= \boldsymbol{L}\mathbf{I}_{D(I+T)}\boldsymbol{L}^{\mathrm{T}} \\
&= \boldsymbol{L}\boldsymbol{L}^{\mathrm{T}} \\
&\sim \boldsymbol{K}_{\text{ext}} \ ,
\end{aligned} \tag{2.72}$$

where we used the identity covariance matrix of vector $\boldsymbol{u}$. By obtaining a sample of random vector $\boldsymbol{u}$ we calculate the realization $\boldsymbol{f}_{\text{ext,s}}$ of the random vector $\boldsymbol{f}_{\text{ext}}$ using (2.70). The realization $\boldsymbol{f}_{\text{test,s}}$ of GP random variables at test positions $\boldsymbol{f}_{\text{test}}$ as a subvector of $\boldsymbol{f}_{\text{ext,s}}$ is than displayed in part a) of figure 2.3 including indication of the true training positions $\boldsymbol{x}_{\text{t,s}}$.

Having the true realization $\boldsymbol{f}_{\text{ext,s}}$ of GP vector $\boldsymbol{f}_{\text{ext}}$ we can obtain the probability density of the random observation vector $\boldsymbol{y}$ of GP random variables at the true training positions $\boldsymbol{x}_{\text{t,s}}$ by taking the generated values of GP at training positions $\boldsymbol{f}_{\text{s}}$ and adding observation noise as in (1.10). From the observation noise distribution (1.11) and from (1.10) we can express the conditional distribution of the GP observation vector $\boldsymbol{y}$ given a GP realization vector at training positions $\boldsymbol{f}_{\text{s}}$ as

$$p(\boldsymbol{y}|\boldsymbol{f} = \boldsymbol{f}_{\text{s}}) = \mathcal{N}(\boldsymbol{y};\ \boldsymbol{f}_{\text{s}},\ \sigma_\epsilon^2\mathbf{I}_I) \ , \tag{2.73}$$

where the additive zero-mean Gaussian noise is chosen to have standard deviation $\sigma_\epsilon = 0.01$. By sampling from this distribution we obtain the realization GP observation $\boldsymbol{y}_{\text{s}}$ at true training positions $\boldsymbol{x}_{\text{t,s}}$.

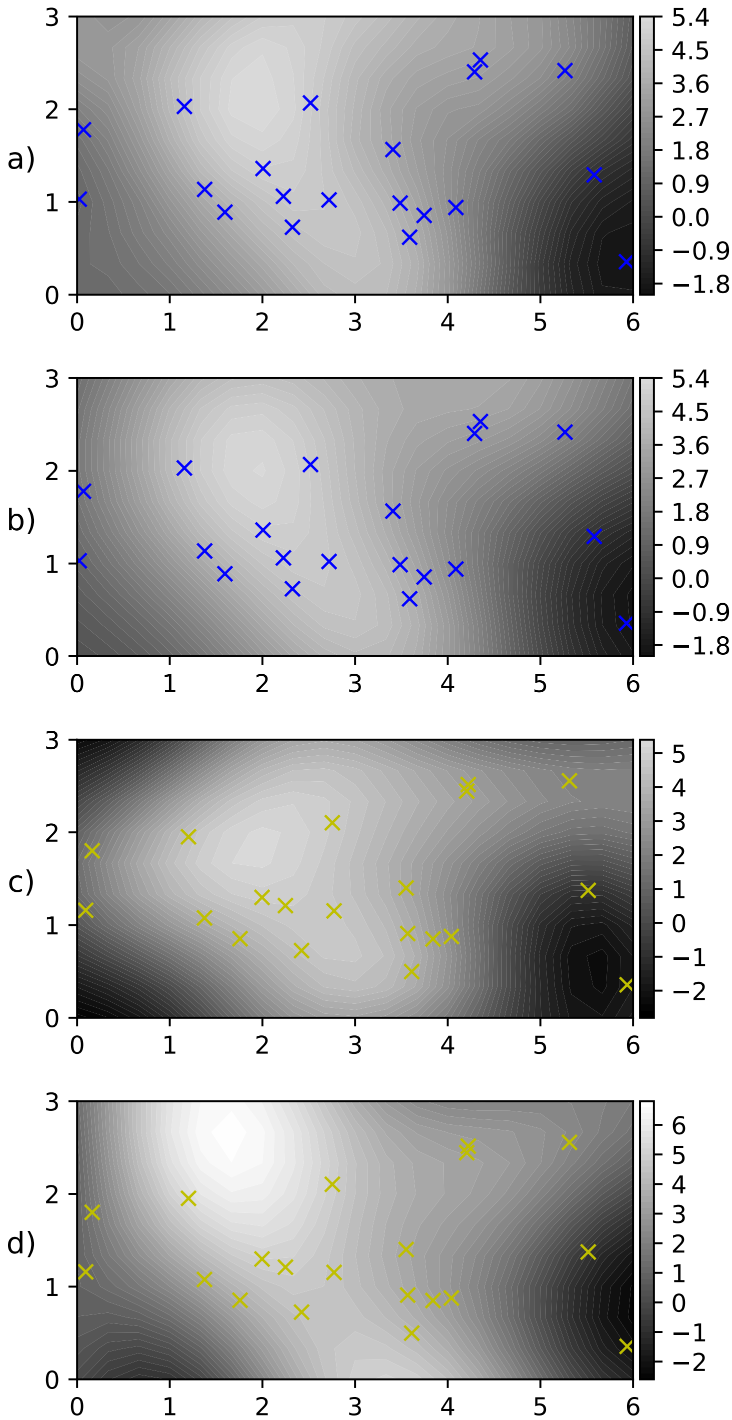


Fig. 2.3: GP estimation:

a) Realization of a GP and true training positions (indicated by blue crosses),

b) GP estimates using the true training positions,

c) GP estimates using the observed training positions (indicated by yellow crosses) directly,

d) GP estimates using the observed training positions and the MC approximation with uniform position prior.

The same scale is used in (a)-(d).

GPR using the true training positions

As the next step we will perform GPR using the true training positions $\boldsymbol{x}_{\mathrm{t,s}}$ as derived in section 1.4 according to the GPR posterior mean given by (1.21)

$$\mu_{f_*|y} = m_* + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{Q}^{-1}(\boldsymbol{y} - \boldsymbol{m}) \ . \tag{2.74}$$

It shall be noted that for the regression we can now use only true training positions $\boldsymbol{x}_{\mathrm{t,s}}$, GP observation at that positions $\boldsymbol{y}_{\mathrm{s}}$ and information about the GP itself as in (2.58).

We start by calculating the covariance matrix of GP observation at training positions $\boldsymbol{x}_{\mathrm{t,s}}$ as in (1.13)

$$\boldsymbol{Q} = \mathrm{cov}\{\boldsymbol{y}\} = \mathrm{cov}\{\boldsymbol{f}\} + \mathrm{cov}\{\boldsymbol{\epsilon}\} = \boldsymbol{K} + \sigma_\epsilon^2\mathbf{I}_I \tag{2.75}$$

while employing the chosen covariance function in (2.57) to calculate the individual elements of matrix $\boldsymbol{K}$. This covariance matrix $\boldsymbol{Q}$ stays the same for all the individual test positions within $\boldsymbol{x}_{\mathrm{test}}$. Further we shall calculate the cross-covariance vector as in (1.20) for a single test position $\boldsymbol{x}_{\mathrm{test}}^{(j)}$ given by

$$\begin{aligned}\boldsymbol{c}^{(j)} &= \mathrm{cov}\{\boldsymbol{y},\ f(\boldsymbol{x}_{\mathrm{test}}^{(j)})\} \\ &= (k(\boldsymbol{x}^{(1)}, \boldsymbol{x}_{\mathrm{test}}^{(j)}) \quad k(\boldsymbol{x}^{(2)}, \boldsymbol{x}_{\mathrm{test}}^{(j)}) \quad \ldots \quad k(\boldsymbol{x}^{(I)}, \boldsymbol{x}_{\mathrm{test}}^{(j)}))^{\mathrm{T}} \in \mathbb{R}^I \ .\end{aligned} \tag{2.76}$$

We are now ready to insert the results of (2.75) and (2.76) into the GPR posterior mean expression in (2.74) while considering a zero-mean GP obtaining

$$\mu_{f(\boldsymbol{x}_{\mathrm{test}}^{(j)})|y} = (\boldsymbol{c}^{(j)})^{\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{y}_{\mathrm{s}} \ . \tag{2.77}$$

Evaluating this posterior mean expression for every test position $\boldsymbol{x}_{\mathrm{test}}^{(j)}$ in $\boldsymbol{x}_{\mathrm{test}}$ we obtain a vector of GPR mean predictions $\boldsymbol{\mu}_{f(\boldsymbol{x}_{\mathrm{test}})|y} \in \mathbb{R}^T$, which is visualized in part b) of figure 2.3.

Further we shall evaluate the posterior variance according to (1.22)

$$\sigma^2_{f_*|y} = k_* - \boldsymbol{c}^{\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{c} \ , \tag{2.78}$$

from where we after inserting the results of (2.75), (2.76) and noticing that $k_* = \sigma^2$ obtain

$$\sigma^2_{f(\boldsymbol{x}_{\mathrm{test}}^{(j)})|y} = \sigma^2 - (\boldsymbol{c}^{(j)})^{\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{c}^{(j)} \ . \tag{2.79}$$

Evaluating this posterior variance expression for every test position $\boldsymbol{x}_{\mathrm{test}}^{(j)}$ in $\boldsymbol{x}_{\mathrm{test}}$ we obtain the GPR variance prediction visualized in part a) of figure 2.4.

GPR using the observed training positions directly

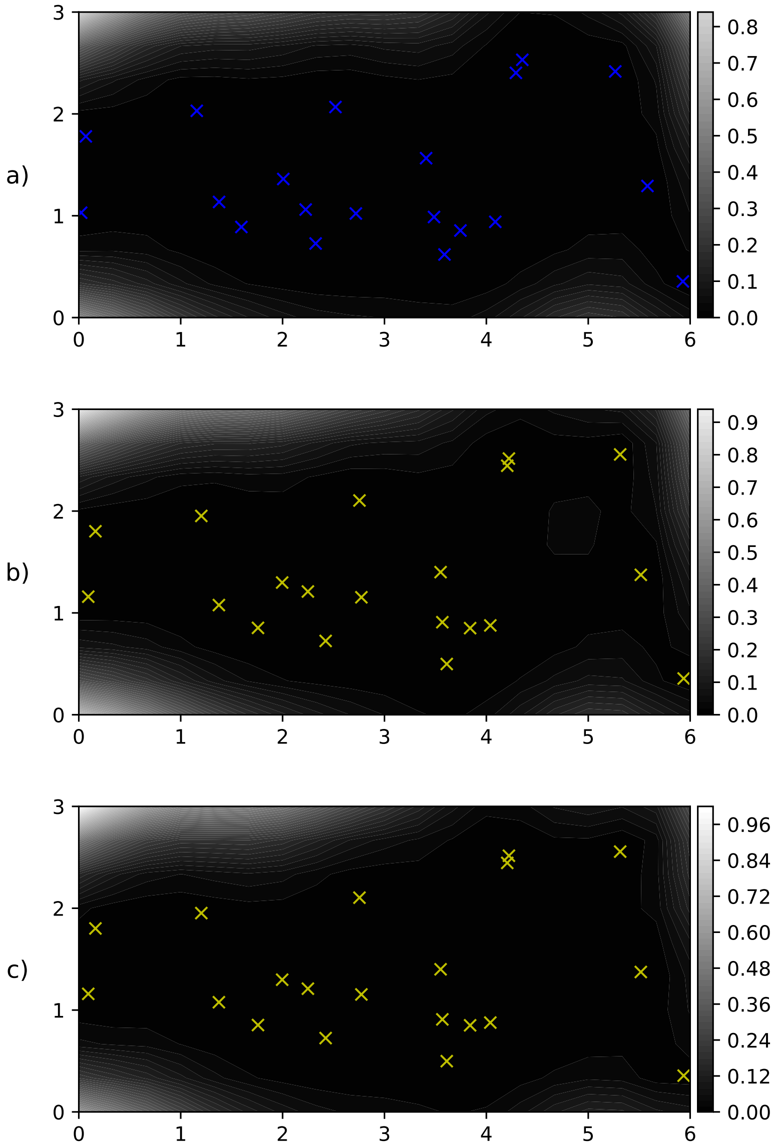


Fig. 2.4: GP variance estimation:

a) GP variance estimation using true training positions,

b) GP variance estimation using the observed training positions directly,

c) GP variance estimation using the observed training positions and the MC approximation with uniform position prior.

The same scale is used in (a)-(c).

Next, we are going to use the noisy training positions observation $\tilde{\boldsymbol{x}}_{\mathrm{t,s}}$ instead of the true training positions $\boldsymbol{x}_{\mathrm{t,s}}$ within the standard GPR framework i.e. considering the noisy positions observation to be the true training positions. The posterior mean can be obtained by using training position observation $\tilde{\boldsymbol{x}}_{\mathrm{t,s}}$ for calculating the GP observation covariance matrix $\boldsymbol{Q}$ in (2.75) and cross-covariance vector $\mathrm{cov}\{f(\boldsymbol{x}_{\mathrm{test}}^{(j)}),\, \boldsymbol{y}\}$ in (2.76). These results are then used for calculating the posterior mean $\mu_{f(\boldsymbol{x}_{\mathrm{test}}^{(j)})|y}$ according to (2.77) for each individual test position $\boldsymbol{x}_{\mathrm{test}}^{(j)}$. This posterior mean is depicted in part c) of figure 2.3. Same procedure applies to calculating the posterior variance $\sigma^2_{f(\boldsymbol{x}_{\mathrm{test}}^{(j)})|y}$ according to (2.79) with a result displayed in part b) of figure 2.4. We observe even in this single realization that the posterior mean prediction has larger posterior mean error than in part b), where we used true training positions.

GPR using the observed training positions while accounting for their uncertainty

Further we are going to use the observation of training positions $\tilde{\boldsymbol{x}}_{\mathrm{t,s}}$ while accounting for the position uncertainty. This shall be done by employing MC sampling of training positions distribution as in (2.31)

$$\mathrm{E}^{\mathrm{MC}}\{f_*|\boldsymbol{y}, \tilde{\boldsymbol{x}}_{\mathrm{t}}\} = \frac{\sum_{i=1}^{s} \mu_*(\boldsymbol{x}_{\mathrm{t},i}) p(\boldsymbol{y}|\boldsymbol{x}_{\mathrm{t}} = \boldsymbol{x}_{\mathrm{t},i})}{\sum_{i=1}^{s} p(\boldsymbol{y}|\boldsymbol{x}_{\mathrm{t}} = \boldsymbol{x}_{\mathrm{t},i})} \,. \tag{2.80}$$

We will start by generating $s = 100$ training positions vector samples $\tilde{\boldsymbol{x}}_{\mathrm{tMC},i} \in \mathbb{R}^{DI}$ with $i = 1, \ldots, s$ from the position posterior distribution described in section 2.3. For doing this we choose the uniform training positions prior as the most appropriate for our simulation setup.

For each individual training positions sample $\tilde{\boldsymbol{x}}_{\mathrm{tMC},i}$ we calculate the covariance matrix of GP observation $\boldsymbol{Q}_i$ as in (2.75) and a cross-covariance vector $\mathrm{cov}\{f(\boldsymbol{x}_{\mathrm{test}}^{(j)}), \boldsymbol{y}_{\mathrm{MC},i}\}$ with regards to a single test position $\boldsymbol{x}_{\mathrm{test}}^{(j)}$ and a vector of GP random values $\boldsymbol{y}_{\mathrm{MC},i}$ at sample training positions $\tilde{\boldsymbol{x}}_{\mathrm{tMC},i}$ as in (2.76). The posterior mean $\mu_{f(\boldsymbol{x}_{\mathrm{test}}^{(j)})|y}(\tilde{\boldsymbol{x}}_{\mathrm{tMC},i})$ at test position $\boldsymbol{x}_{\mathrm{test}}^{(j)}$ for given training positions sample $\tilde{\boldsymbol{x}}_{\mathrm{tMC},i}$ is then calculated using (2.77) and the respective posterior variance $\sigma^2_{f(\boldsymbol{x}_{\mathrm{test}}^{(j)})|y}(\tilde{\boldsymbol{x}}_{\mathrm{tMC},i})$ using (2.79). Further we shall evaluate the marginal likelihood $p(\boldsymbol{y}|\boldsymbol{x}_{\mathrm{t}})$, which is in contrast to (1.12) conditioned on $\boldsymbol{x}_{\mathrm{t}}$ to indicate that the true training positions considered are not implicit. By inserting our intermediate results we can express it as

$$p(\boldsymbol{y}|\tilde{\boldsymbol{x}}_{\mathrm{t}} = \tilde{\boldsymbol{x}}_{\mathrm{tMC},i}) = \mathcal{N}(\boldsymbol{y};\, \boldsymbol{0},\, \boldsymbol{Q}_i) \,. \tag{2.81}$$

All s training positions samples are then combined to evaluate the MC approximation of the posterior mean by inserting into (2.80) obtaining

$$\mathrm{E}^{\mathrm{MC}}\{f(\boldsymbol{x}_{\mathrm{test}}^{(j)})|\boldsymbol{y}=\boldsymbol{y}_{\mathrm{s}},\tilde{\boldsymbol{x}}_{\mathrm{t}}=\tilde{\boldsymbol{x}}_{\mathrm{t,s}}\}=\frac{\sum_{i=1}^{s}\mu_{f(\boldsymbol{x}_{\mathrm{test}}^{(j)})|y}(\tilde{\boldsymbol{x}}_{\mathrm{tMC},i})p(\boldsymbol{y}=\boldsymbol{y}_{\mathrm{s}}|\tilde{\boldsymbol{x}}_{\mathrm{t}}=\tilde{\boldsymbol{x}}_{\mathrm{tMC},i})}{\sum_{i=1}^{s}p(\boldsymbol{y}=\boldsymbol{y}_{\mathrm{s}}|\tilde{\boldsymbol{x}}_{\mathrm{t}}=\tilde{\boldsymbol{x}}_{\mathrm{tMC},i})}\,. \tag{2.82}$$

This posterior mean is then evaluated for all test positions contained in $\boldsymbol{x}_{\mathrm{test}}$. We can see a visualization of these predictions in part d) of figure 2.3. It can be seen even from this single realization that the prediction performs better in contrast to the direct usage of training positions observation in part c).

Similarly as in (2.82) we obtain the MC approximation of the posterior variance according to (2.33) as

$$\begin{aligned}\mathrm{var}^{\mathrm{MC}}\{f(\boldsymbol{x}_{\mathrm{test}}^{(j)})|\boldsymbol{y}=\boldsymbol{y}_{\mathrm{s}},\tilde{\boldsymbol{x}}_{\mathrm{t}}=\tilde{\boldsymbol{x}}_{\mathrm{t,s}}\}=&\\ \frac{\sum_{i=1}^{s}(\mu_{f(\boldsymbol{x}_{\mathrm{test}}^{(j)})|y}(\tilde{\boldsymbol{x}}_{\mathrm{tMC},i})+\sigma^2_{f(\boldsymbol{x}_{\mathrm{test}}^{(j)})|y}(\tilde{\boldsymbol{x}}_{\mathrm{tMC},i}))p(\boldsymbol{y}=\boldsymbol{y}_{\mathrm{s}}|\tilde{\boldsymbol{x}}_{\mathrm{t}}=\tilde{\boldsymbol{x}}_{\mathrm{tMC},i})}{\sum_{i=1}^{s}p(\boldsymbol{y}=\boldsymbol{y}_{\mathrm{s}}|\tilde{\boldsymbol{x}}_{\mathrm{t}}=\tilde{\boldsymbol{x}}_{\mathrm{tMC},i})}&\\ -(\mathrm{E}^{\mathrm{MC}}\{f(\boldsymbol{x}_{\mathrm{test}}^{(j)})|\boldsymbol{y}=\boldsymbol{y}_{\mathrm{s}},\tilde{\boldsymbol{x}}_{\mathrm{t}}=\tilde{\boldsymbol{x}}_{\mathrm{t,s}}\})^2\,,&\end{aligned} \tag{2.83}$$

which after evaluating for all test positions contained in $\boldsymbol{x}_{\mathrm{test}}$ gives results displayed in part c) of figure 2.4.

Performance evaluation

To evaluate how each of the regression methods performed we will use the *root-mean-square* error (RMSE) of predictions at the test positions in $\boldsymbol{x}_{\mathrm{test}}$. In general, a prediction for a single test position $\boldsymbol{x}_{\mathrm{test}}^{(j)}$ will be denoted as a posterior mean $\boldsymbol{\mu}^{(j)}$. For this posterior mean we will be inserting the prediction obtained by each of considered GPR methods. Still in general, the GPR for a single simulation run can be expressed as

$$\mathrm{E}_{\mathrm{RMS}}(\boldsymbol{\mu},\boldsymbol{f}_{\mathrm{test,s}})=\sqrt{\frac{\sum_{j=1}^{T}(\boldsymbol{\mu}^{(j)}-\boldsymbol{f}_{\mathrm{test,s}}^{(j)})^2}{T}}\,, \tag{2.84}$$

where $\boldsymbol{\mu}^{(j)}$ denotes the jth element of the vector of mean predictions $\boldsymbol{\mu}$ at test positions $\boldsymbol{x}_{\mathrm{test}}$ and $\boldsymbol{f}_{\mathrm{test,s}}^{(j)}$ is the jth element of the vector of GP realizations, both corresponding to the test position $\boldsymbol{x}_{\mathrm{test}}^{(j)}$. In this form the $\mathrm{E}_{\mathrm{RMS}}$ is an estimator of the RMSE within a single simulation with constant training positions and T test positions.

Because of the numerous sources of randomness in our simulation it is necessary to carry out multiple simulation runs and after that evaluate the RMSE for each of the discussed GPR methods. We performed $R = 100$ GPR simulation runs, each time generating new training positions and GP realization. The resulting RMSE metric comprising all the simulation runs is calculated according to

$$\mathrm{E}_{\mathrm{RMS}} = \sqrt{\frac{\sum_{r=1}^{R} ({\mathrm{E}_{\mathrm{RMS}}}^{(r)})^2}{R}}\,, \tag{2.85}$$

where ${\mathrm{E}_{\mathrm{RMS}}}^{(r)}$ denotes the RMSE result of rth simulation run.

Firstly we evaluate the RMSE of the GPR operating with the true training positions $\boldsymbol{x}_{\mathrm{t,s}}$. Similarly we evaluate the RMSE for GPR using training position observations $\tilde{\boldsymbol{x}}_{\mathrm{t},s}$ directly and for GPR accounting for training positions uncertainty. The resulting RMSE for each of performed GPR scenarios is summed up in table 2.1. We can see that our results show a significantly decreased RMSE when employing GPR accounting for position uncertainty in comparison to the GPR using noisy positions observation directly. Nevertheless, our results differ from the ones provided in [2, tab. 2], which could be caused by some differences in simulation setup, namely the ambiguity in number of MC samples used, number of simulations performed and different training positions prior distribution.

GPR method	Evaluated RMSE	RMSE from [2, tab. 2]
GPR using the true training positions	0.3837	0.1281
GPR using the observed training positions directly	1.516	0.9126
GPR using the observed training positions while accounting for their uncertainty	0.8683	0.3403

Tab. 2.1: RMSE of the considered GPR methods.

A second metric to evaluate the quality of predictions as suggested in [3] and [1] is the *mean negative log-predictive density* (MNLPD), which employs not only the posterior mean as for the RMSE but also the posterior variance. The general idea is to consider the Gaussian approximation of the posterior pdf at the test position and to evaluate its probability density at the point of spatial function true realization value. The closer the posterior mean to the true realization value, the higher is the density. And, the posterior variance determines how quickly the posterior density decreases when the true realization goes away from the posterior mean. Therefore this metric penalizes for prediction errors more strongly in test positions, where predictive variance is lower. To obtain the negative log predictive density for a single test position $\boldsymbol{x}_{\mathrm{test}}^{(j)}$ for which we have already calculated the posterior mean $\boldsymbol{\mu}^{(j)}$ and the posterior variance $\boldsymbol{\sigma}^{(j),2}$ we start by expressing the posterior pdf evaluated at the true spatial function value $\boldsymbol{f}^{(j)}_{\text{test,s}}$ as

$$p(f(\boldsymbol{x}^{(j)}_{\text{test}}) = \boldsymbol{f}^{(j)}_{\text{test,s}}|\boldsymbol{y}) = \frac{1}{\sqrt{2\pi\boldsymbol{\sigma}^{(j),2}}}\exp\left(-\frac{(\boldsymbol{f}^{(j)}_{\text{test,s}} - \boldsymbol{\mu}^{(j)})^2}{2\boldsymbol{\sigma}^{(j),2}}\right) , \tag{2.86}$$

which can be in the following steps

$$\begin{aligned}\sqrt{2\pi\boldsymbol{\sigma}^{(j),2}}p(f(\boldsymbol{x}^{(j)}_{\text{test}}) = \boldsymbol{f}^{(j)}_{\text{test,s}}|\boldsymbol{y}) &= \exp\left(-\frac{(\boldsymbol{f}^{(j)}_{\text{test,s}} - \boldsymbol{\mu}^{(j)})^2}{2\boldsymbol{\sigma}^{(j),2}}\right)\\ \frac{1}{2}\log(2\pi\boldsymbol{\sigma}^{(j),2}) + \log(p(f(\boldsymbol{x}^{(j)}_{\text{test}}) = \boldsymbol{f}^{(j)}_{\text{test,s}}|\boldsymbol{y})) &= -\frac{(\boldsymbol{f}^{(j)}_{\text{test,s}} - \boldsymbol{\mu}^{(j)})^2}{2\boldsymbol{\sigma}^{(j),2}}\end{aligned} \tag{2.87}$$

transformed into the negative log-predictive density

$$-\log(p(f(\boldsymbol{x}^{(j)}_{\text{test}}) = \boldsymbol{f}^{(j)}_{\text{test,s}}|\boldsymbol{y})) = \frac{1}{2}\left(\log(2\pi) + \log(\boldsymbol{\sigma}^{(j),2}) + \frac{(\boldsymbol{f}^{(j)}_{\text{test,s}} - \boldsymbol{\mu}^{(j)})^2}{\boldsymbol{\sigma}^{(j),2}}\right) . \tag{2.88}$$

To obtain an estimator of the mean negative log-predictive density we average over all the T test positions within one simulation to obtain

$$\begin{aligned}&\mathrm{E}_{\mathrm{MNLPD}}(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{f}_{\text{test,s}}) =\\ &\qquad \frac{1}{2T}\sum_{j=1}^{T}\left(\log(2\pi) + \log(\boldsymbol{\sigma}^{(j),2}) + \frac{(\boldsymbol{f}^{(j)}_{\text{test,s}} - \boldsymbol{\mu}^{(j)})^2}{\boldsymbol{\sigma}^{(j),2}}\right) .\end{aligned} \tag{2.89}$$

Similarly as for the RMSE we carried out 100 GPR simulations to obtain a MNLPD comparison of discussed GPR methods, now with averaging of the results of each of the simulations. The resulting MNLPD values can be found in table 2.2. The lower the MNLPD value the better are the predictive mean and predictive variance matched. We can see that GPR method accounting for training positions uncertainty performed much better than the GPR method using the noisy training positions directly.

2.5 Disregarding the Dependence between $\tilde{\boldsymbol{x}}_{\text{t}}$ and $\boldsymbol{y}$

In our derivation of the posterior pdf in (2.10) under training positions uncertainty we accounted for the statistical dependence between training positions $\tilde{\boldsymbol{x}}_{\text{t}}$ and the spatial function observations $\boldsymbol{y}$. By omitting this dependence (as shown in [14], [16]) we can reach a simpler form of the predictive pdf, but at the cost of not utilizing all

GPR method	Evaluated MNLPD
GPR using the true training positions	0.4632
GPR using the observed training positions directly	35.54
GPR using the observed training positions while accounting for their uncertainty	16.99

Tab. 2.2: MNLPD of the considered GPR methods.

the information provided by our statistical model. We start our derivation similarly as in (2.8) but firstly using the sum rule as

$$\begin{aligned} p(f_*|\boldsymbol{y}, \tilde{\boldsymbol{x}}_\mathrm{t}) &= \int_{\mathbb{R}^{DI}} p(f_*, \boldsymbol{x}_\mathrm{t}|\boldsymbol{y}, \tilde{\boldsymbol{x}}_\mathrm{t})\, d\boldsymbol{x}_\mathrm{t} \\ &= \int_{\mathbb{R}^{DI}} p(f_*|\boldsymbol{x}_\mathrm{t}, \boldsymbol{y}, \tilde{\boldsymbol{x}}_\mathrm{t}) p(\boldsymbol{x}_\mathrm{t}|\boldsymbol{y}, \tilde{\boldsymbol{x}}_\mathrm{t})\, d\boldsymbol{x}_\mathrm{t} \ . \end{aligned} \tag{2.90}$$

Here we can again use the conditional independence of f_* of $\tilde{\boldsymbol{x}}_\mathrm{t}$ given $\boldsymbol{x}_\mathrm{t}$ just as in (2.10) to rewrite the term $p(f_*|\boldsymbol{x}_\mathrm{t}, \boldsymbol{y}, \tilde{\boldsymbol{x}}_\mathrm{t})$ equivalently as $p(f_*|\boldsymbol{x}_\mathrm{t}, \boldsymbol{y})$. By expressing the term $p(\boldsymbol{x}_\mathrm{t}|\boldsymbol{y}, \tilde{\boldsymbol{x}}_\mathrm{t})$ as

$$\begin{aligned} p(\boldsymbol{x}_\mathrm{t}|\boldsymbol{y}, \tilde{\boldsymbol{x}}_\mathrm{t}) &= \frac{p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t}, \tilde{\boldsymbol{x}}_\mathrm{t}) p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})}{p(\boldsymbol{y}|\tilde{\boldsymbol{x}}_\mathrm{t})} \\ &= \frac{p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t}, \tilde{\boldsymbol{x}}_\mathrm{t}) p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})}{\int_{\mathbb{R}^{DI}} p(\boldsymbol{y}, \boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})\, d\boldsymbol{x}_\mathrm{t}} \\ &= \frac{p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t}, \tilde{\boldsymbol{x}}_\mathrm{t}) p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})}{\int_{\mathbb{R}^{DI}} p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t}, \tilde{\boldsymbol{x}}_\mathrm{t}) p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})\, d\boldsymbol{x}_\mathrm{t}} \\ &= \frac{p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t}) p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})}{\int_{\mathbb{R}^{DI}} p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t}) p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})\, d\boldsymbol{x}_\mathrm{t}} \ , \end{aligned} \tag{2.91}$$

where we in the first step used the Bayes' rule, in the second step used the sum rule, in the third step used the product rule and in the last step used the conditional independence of $\boldsymbol{y}$ of $\tilde{\boldsymbol{x}}_\mathrm{t}$ given $\boldsymbol{x}_\mathrm{t}$. From (2.91) we can see that $p(\boldsymbol{x}_\mathrm{t}|\boldsymbol{y}, \tilde{\boldsymbol{x}}_\mathrm{t}) \neq p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})$. However, if we within this section take the assumption that $\boldsymbol{x}_\mathrm{t}$ is independent of $\boldsymbol{y}$ given $\tilde{\boldsymbol{x}}_\mathrm{t}$, i.e. $p(\boldsymbol{x}_\mathrm{t}|\boldsymbol{y}, \tilde{\boldsymbol{x}}_\mathrm{t}) = p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})$, we can express (2.90) as

$$p(f_*|\boldsymbol{y}, \tilde{\boldsymbol{x}}_\mathrm{t}) = \int_{\mathbb{R}^{DI}} p(f_*|\boldsymbol{x}_\mathrm{t}, \boldsymbol{y}, \tilde{\boldsymbol{x}}_\mathrm{t}) p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})\, d\boldsymbol{x}_\mathrm{t} \ . \tag{2.92}$$

This equation can be found in [14, eq.14], where it is in a more general form considering also an uncertain test position. Here the authors later perform a simulation of a scenario with uncertain training positions and posterior mean prediction at a known test position. In this setup the provided equation for posterior pdf [14, eq.14] collapses into the derived form (2.92).

MC approximation of the posterior mean

The equation (2.92) is used to express the posterior mean using a similar derivation as for (2.15) resulting in

$$\begin{aligned}
\mathrm{E}\{f_*|\boldsymbol{y},\tilde{\boldsymbol{x}}_\mathrm{t}\} &= \int_{\mathbb{R}} f_* p(f_*|\boldsymbol{y},\tilde{\boldsymbol{x}}_\mathrm{t})\, df_* \\
&= \int_{\mathbb{R}} f_* \int_{\mathbb{R}^{DI}} p(f_*|\boldsymbol{x}_\mathrm{t},\boldsymbol{y},\tilde{\boldsymbol{x}}_\mathrm{t}) p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})\, d\boldsymbol{x}_\mathrm{t}\, df_* \\
&= \int_{\mathbb{R}^{DI}} \int_{\mathbb{R}} f_* p(f_*|\boldsymbol{x}_\mathrm{t},\boldsymbol{y},\tilde{\boldsymbol{x}}_\mathrm{t})\, df_* p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})\, d\boldsymbol{x}_\mathrm{t} \\
&= \int_{\mathbb{R}^{DI}} \mu_*(\boldsymbol{x}_\mathrm{t}) p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})\, d\boldsymbol{x}_\mathrm{t} \\
&= \mathrm{E}^{p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})}\{\mu_*(\boldsymbol{x}_\mathrm{t})\}\ .
\end{aligned} \tag{2.93}$$

This term is again approximated using a MC method with samples of vector of training positions $\boldsymbol{x}_{\mathrm{t},i}$ drawn according to the pdf $p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})$. These samples provide us with a pmf approximation $p_\mathrm{MC}(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})$ of $p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})$ just as in (2.30). Using this pmf we can express the approximate posterior mean as

$$\mathrm{E}^\mathrm{MC}\{f_*|\boldsymbol{y},\tilde{\boldsymbol{x}}_\mathrm{t}\} = \mathrm{E}^{p_\mathrm{MC}(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})}\{\mu_*(\boldsymbol{x}_\mathrm{t})\} = \frac{1}{s}\sum_{i=1}^{s}\mu_*(\boldsymbol{x}_{\mathrm{t},i})\ . \tag{2.94}$$

The individual random training positions in [14] were considered to be independent and to have isotropic Gaussian distribution. This is equivalent to our discussion in training positions sampling section 2.3 considering improper prior. However, as there is no necessary assumption about the Gaussianity of the training positions distribution, we used the uniform prior for generating our samples of training position. This was motivated by the fact that this prior is most suited to our simulation scenario and that the results of this section hold also for the improper prior case.

In figure 2.5 you can see the results of a single GPR simulation. Here we compare the realization of a GP to the posterior mean obtained with (2.94) (disregarding statistical dependence between $\tilde{\boldsymbol{x}}_\mathrm{t}$ and $\boldsymbol{y}$) and to the posterior mean obtained with (2.31), which accounts for the statistical dependence between $\tilde{\boldsymbol{x}}_\mathrm{t}$ and $\boldsymbol{y}$. We can observe that omitting the statistical dependence assumption leads to a significantly different posterior mean prediction. That means the statistical independence assumption has a significant effect on the prediction quality.

MC approximation of the posterior variance

Considering the posterior pdf according to (2.92) we can find the expression for the posterior variance by following the derivation of (2.22). This results in the exact

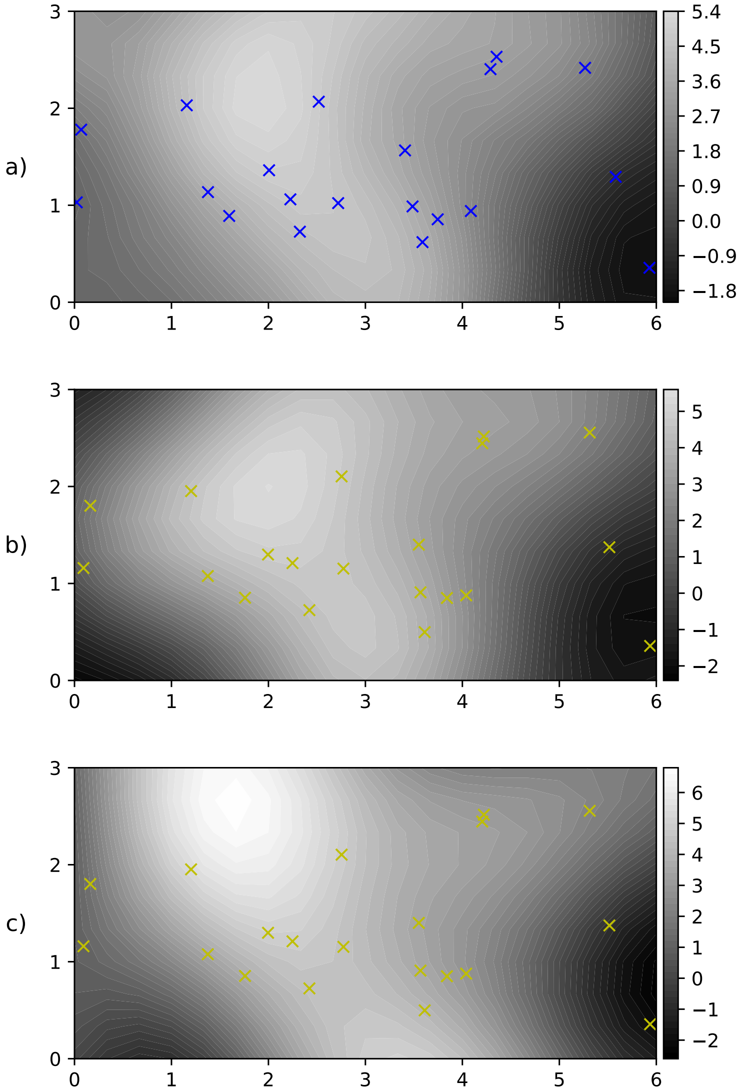


Fig. 2.5: GP estimation:
a) Realization of a GP and true training positions (indicated by blue crosses),
b) GP estimates obtained from (2.94) (disregarding the statistical dependence between $\tilde{\boldsymbol{x}}_{\mathrm{t}}$ and $\boldsymbol{y}$),
c) GP estimates obtained from (2.31) accounting for the statistical dependence between $\tilde{\boldsymbol{x}}_{\mathrm{t}}$ and $\boldsymbol{y}$).
The same scale is used in (a)-(c).

posterior variance term

$$\mathrm{var}\{f_*|\boldsymbol{y},\tilde{\boldsymbol{x}}_\mathrm{t}\} = \int_{\mathbb{R}^{DI}} (\mu_*^2(\boldsymbol{x}_\mathrm{t}) + \sigma_*^2(\boldsymbol{x}_\mathrm{t}))p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})\, d\boldsymbol{x}_\mathrm{t} - (\mathrm{E}\{f_*|\boldsymbol{y},\tilde{\boldsymbol{x}}_\mathrm{t}\})^2 \ , \qquad (2.95)$$

where $\mathrm{E}\{f_*|\boldsymbol{y},\tilde{\boldsymbol{x}}_\mathrm{t}\}$ is already known from (2.93). By approximating the training positions posterior pdf $p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})$ by the pmf $p_\mathrm{MC}(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})$ as in (2.30) we can approximate the posterior variance in (2.95) as

$$\mathrm{var}^\mathrm{MC}\{f_*|\boldsymbol{y},\tilde{\boldsymbol{x}}_\mathrm{t}\} = \frac{1}{s}\sum_{i=1}^{s}(\mu_*^2(\boldsymbol{x}_{\mathrm{t},i}) + \sigma_*^2(\boldsymbol{x}_{\mathrm{t},i})) - (\mathrm{E}^\mathrm{MC}\{f_*|\boldsymbol{y},\tilde{\boldsymbol{x}}_\mathrm{t}\})^2 \qquad (2.96)$$

with samples of the vector of training positions drawn according to $p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})$.

In figure 2.6 you can see a visualization of the variance prediction for the same simulation as in figure 2.5. It is a comparison between the posterior variance obtained with (2.96) (disregarding statistical dependence between $\tilde{\boldsymbol{x}}_\mathrm{t}$ and $\boldsymbol{y}$) and the posterior variance obtained with (2.22), which accounts for the statistical dependence between $\tilde{\boldsymbol{x}}_\mathrm{t}$ and $\boldsymbol{y}$. We can see that omitting the statistical dependence assumption leads to a significantly higher posterior variance expressing a higher level of prediction uncertainty. This leads to a conclusion that the amount of information lost by omitting the statistical dependence is significant.

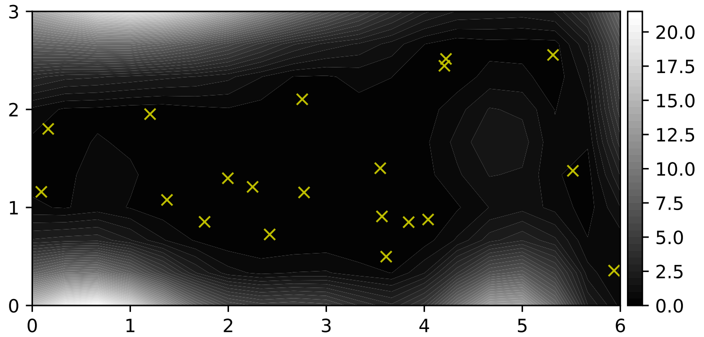
Performance evaluation

Similarly as in section 2.4 we compared the currently studied GPR methods using the RMSE an MNLPD metrics. Again, 100 simulations were carried out to obtain results displayed in table 2.3. Considering the RMSE metric we can notice that disregarding of the statistical dependence between $\tilde{\boldsymbol{x}}_\mathrm{t}$ and $\boldsymbol{y}$ leads to an increase in the posterior mean error. The situation regarding MNLPD metric is different as it favors the simplified GPR method over the original one. This can be explained by the high predictive variance provided by the simplified GPR method that allows for higher RMSE while reducing the MNLPD metric.

GPR method	Evaluated RMSE	Evaluated MNLPD
GPR according to (2.94)	1.026	0.9206
GPR according to (2.31)	0.8683	16.99

Tab. 2.3: Comparison in terms of the RMSE and MNLPD metrics of the simplified GPR method according to (2.94) (disregarding the statistical dependence between $\tilde{\boldsymbol{x}}_\mathrm{t}$ and $\boldsymbol{y}$) and the originally derived GPR method according to (2.31) (accounting for this statistical dependence between $\tilde{\boldsymbol{x}}_\mathrm{t}$ and $\boldsymbol{y}$).

a)



b)

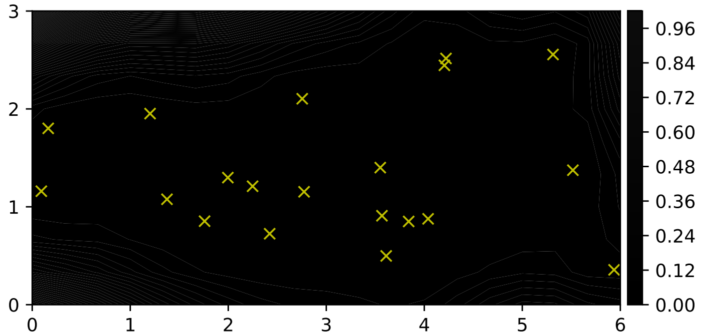


Fig. 2.6: GPR variance estimation:
a) using the approximation according to (2.96) (disregarding the statistical dependence between $\tilde{\boldsymbol{x}}_{\mathrm{t}}$ and $\boldsymbol{y}$),
b) using the approximation according to (2.33) accounting for the statistical dependence between $\tilde{\boldsymbol{x}}_{\mathrm{t}}$ and $\boldsymbol{y}$).
The same scale is used in (a) and (b).

3 GPR Closed-form Prediction under Position Uncertainty

One of the important advantages of the GPR is the availability of the closed-form expression for the posterior pdf mean and variance. In Chapter 2 we described a method to incorporate training positions uncertainty. But, at the same time, we lost the closed-form expressions and needed to approximate the posterior pdf using the Monte Carlo sampling. In this chapter we will investigate the possibilities of performing GPR operating on uncertain training positions while using closed-form expressions describing the posterior pdf. This chapter was mostly inspired by [3, Ch.3], [15], [14] and [16], while some necessary assumptions were formulated independently.

3.1 Incorporating Uncertain Training Positions

The setup considered will be almost identical to the one in Chapter 2 but here we will take the perspective of the individual random training positions $\boldsymbol{x}^{(i)}$ rather than the whole batch $\boldsymbol{x}_{\mathrm{t}}$. The true training position $\boldsymbol{x}^{(i)}$ is unknown but a localization technique is considered to provide us with an observation of it corrupted by an additive noise as in (2.4), i.e.

$$\tilde{\boldsymbol{x}}^{(i)} = \boldsymbol{x}^{(i)} + \boldsymbol{w}^{(i)} \ . \tag{3.1}$$

The noise $\boldsymbol{w}^{(i)}$ will be considered iid for individual training positions with a known distribution $p(\boldsymbol{w}^{(i)})$. Combined with the prior distribution of the training position $p(\boldsymbol{x}^{(i)})$ we can express the posterior pdf $p(\boldsymbol{x}^{(i)}|\tilde{\boldsymbol{x}}^{(i)})$ as shown for the complete batch of training positions in 2.3. For now we will consider the posterior pdf $p(\boldsymbol{x}^{(i)}|\tilde{\boldsymbol{x}}^{(i)})$ known but still general.

The random variable corresponding to a GP at the random training position $\boldsymbol{x}^{(i)}$ observed as $\tilde{\boldsymbol{x}}^{(i)}$ is then expressed as

$$f(\boldsymbol{x}^{(i)})|\tilde{\boldsymbol{x}}^{(i)} = f(\tilde{\boldsymbol{x}}^{(i)} + \boldsymbol{w}^{(i)}) \ . \tag{3.2}$$

Approximation with a GP

As stated in [9, page 48], [3, page 51], (3.2) considered as a function of $\tilde{\boldsymbol{x}}^{(i)}$ is not a GP anymore. This can be shown considering two random training positions $\boldsymbol{x}^{(i)}$ and $\boldsymbol{x}^{(j)}$. Considering the joint distribution of the GP $f(\boldsymbol{x})$ at these two positions conditioned on them, i.e. we suppress the positions randomness, we obtain a

Gaussian distribution as discussed in Chapter 1 in the form

$$p(f(\boldsymbol{x}^{(i)}), f(\boldsymbol{x}^{(j)})|\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}) = \mathcal{N}\left(\begin{pmatrix} f_i \\ f_j \end{pmatrix}; \begin{pmatrix} m(\boldsymbol{x}^{(i)}) \\ m(\boldsymbol{x}^{(j)}) \end{pmatrix}, \begin{pmatrix} k(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(i)}) & k(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}) \\ k(\boldsymbol{x}^{(j)}, \boldsymbol{x}^{(i)}) & k(\boldsymbol{x}^{(j)}, \boldsymbol{x}^{(j)}) \end{pmatrix}\right) . \tag{3.3}$$

Here $m(\boldsymbol{x})$ and $k(\boldsymbol{x}, \boldsymbol{x}')$ are the mean function and the covariance function of the GP $f(\boldsymbol{x})$. To obtain the joint distribution as in (3.3) but conditioned on the training positions observations $\tilde{\boldsymbol{x}}^{(i)}$ and $\tilde{\boldsymbol{x}}^{(j)}$ we proceed as

$$\begin{aligned} &p(f(\boldsymbol{x}^{(i)}), f(\boldsymbol{x}^{(j)})|\tilde{\boldsymbol{x}}^{(i)}, \tilde{\boldsymbol{x}}^{(j)}) \\ &= \int_{\mathbb{R}^D}\int_{\mathbb{R}^D} p(f(\boldsymbol{x}^{(i)}), f(\boldsymbol{x}^{(j)}), \boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}|\tilde{\boldsymbol{x}}^{(i)}, \tilde{\boldsymbol{x}}^{(j)})\, d\boldsymbol{x}^{(i)}\, d\boldsymbol{x}^{(j)} \\ &= \int_{\mathbb{R}^D}\int_{\mathbb{R}^D} p(f(\boldsymbol{x}^{(i)}), f(\boldsymbol{x}^{(j)})|\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}, \tilde{\boldsymbol{x}}^{(i)}, \tilde{\boldsymbol{x}}^{(j)}) p(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}|\tilde{\boldsymbol{x}}^{(i)}, \tilde{\boldsymbol{x}}^{(j)})\, d\boldsymbol{x}^{(i)}\, d\boldsymbol{x}^{(j)} \\ &= \int_{\mathbb{R}^D}\int_{\mathbb{R}^D} p(f(\boldsymbol{x}^{(i)}), f(\boldsymbol{x}^{(j)})|\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}) p(\boldsymbol{x}^{(i)}|\tilde{\boldsymbol{x}}^{(i)}) p(\boldsymbol{x}^{(j)}|\tilde{\boldsymbol{x}}^{(j)})\, d\boldsymbol{x}^{(i)}\, d\boldsymbol{x}^{(j)} , \end{aligned} \tag{3.4}$$

where we in the first step used the sum rule, in the second the product rule and in the third notice the conditional independence of $f(\boldsymbol{x}^{(i)}), f(\boldsymbol{x}^{(j)})$ of $\tilde{\boldsymbol{x}}^{(i)}, \tilde{\boldsymbol{x}}^{(j)}$ given $\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}$. In this step we also used the independence of the individual training positions $\boldsymbol{x}^{(i)}$ and $\boldsymbol{x}^{(j)}$.

The resulting expression in (3.4) can be seen as not a multivariate Gaussian distribution in general. Nevertheless we will approximate it to be a Gaussian distribution. This allows us to consider a new GP $g(\tilde{\boldsymbol{x}}^{(i)}) \approx f(\boldsymbol{x}^{(i)})|\tilde{\boldsymbol{x}}^{(i)}$ operating on the random position observations. With this GP we are able to use the standard GPR formulas for predictions from Chapter 1. For this we will use, as in (1.3), a new version of the mean and the covariance function defining the GP as

$$p(g(\tilde{\boldsymbol{x}}^{(i)})) = \mathcal{GP}(m'(\tilde{\boldsymbol{x}}^{(i)}), k'(\tilde{\boldsymbol{x}}^{(i)}, \tilde{\boldsymbol{x}}^{(j)})) . \tag{3.5}$$

To obtain them we start by expressing the pdf of the random variable $f(\boldsymbol{x}^{(i)})|\tilde{\boldsymbol{x}}^{(i)}$ as

$$\begin{aligned} p(f(\boldsymbol{x}^{(i)})|\tilde{\boldsymbol{x}}^{(i)}) &= \int_{\mathbb{R}^D} p(f(\boldsymbol{x}^{(i)}), \boldsymbol{x}^{(i)}|\tilde{\boldsymbol{x}}^{(i)})\, d\boldsymbol{x}^{(i)} \\ &= \int_{\mathbb{R}^D} p(f(\boldsymbol{x}^{(i)})|\boldsymbol{x}^{(i)}, \tilde{\boldsymbol{x}}^{(i)}) p(\boldsymbol{x}^{(i)}|\tilde{\boldsymbol{x}}^{(i)})\, d\boldsymbol{x}^{(i)} \\ &= \int_{\mathbb{R}^D} p(f(\boldsymbol{x}^{(i)})|\boldsymbol{x}^{(i)}) p(\boldsymbol{x}^{(i)}|\tilde{\boldsymbol{x}}^{(i)})\, d\boldsymbol{x}^{(i)} \\ &= \mathrm{E}^{p(\boldsymbol{x}^{(i)}|\tilde{\boldsymbol{x}}^{(i)})}\{f(\boldsymbol{x}^{(i)})|\boldsymbol{x}^{(i)}\} , \end{aligned} \tag{3.6}$$

where we in the first step used the sum rule, in the second the product rule and in the third the conditional independence of $f(\boldsymbol{x}^{(i)})$ of $\tilde{\boldsymbol{x}}^{(i)}$ given $\boldsymbol{x}^{(i)}$.

GP mean function

The mean of the approximate GP $g(\tilde{\boldsymbol{x}}^{(i)})$ will be expressed as the mean of the random variable $f(\boldsymbol{x}^{(i)})|\tilde{\boldsymbol{x}}^{(i)}$, which can be obtained as

$$\begin{aligned}
m'(\tilde{\boldsymbol{x}}^{(i)}) &= \mathrm{E}\{f(\boldsymbol{x}^{(i)})|\tilde{\boldsymbol{x}}^{(i)}\} \\
&= \int_{\mathbb{R}} f p_{f(\boldsymbol{x}^{(i)})|\tilde{\boldsymbol{x}}^{(i)}}(f)\, df \\
&= \int_{\mathbb{R}} f \int_{\mathbb{R}^D} p_{f(\boldsymbol{x}^{(i)})|\boldsymbol{x}^{(i)}}(f) p(\boldsymbol{x}^{(i)}|\tilde{\boldsymbol{x}}^{(i)})\, d\boldsymbol{x}^{(i)}\, df \\
&= \int_{\mathbb{R}^D} \int_{\mathbb{R}} f p_{f(\boldsymbol{x}^{(i)})|\boldsymbol{x}^{(i)}}(f)\, df\, p(\boldsymbol{x}^{(i)}|\tilde{\boldsymbol{x}}^{(i)})\, d\boldsymbol{x}^{(i)} \\
&= \mathrm{E}^{p(\boldsymbol{x}^{(i)}|\tilde{\boldsymbol{x}}^{(i)})}\{\mathrm{E}\{f(\boldsymbol{x}^{(i)})|\boldsymbol{x}^{(i)}\}\} \\
&= \mathrm{E}^{p(\boldsymbol{x}^{(i)}|\tilde{\boldsymbol{x}}^{(i)})}\{m(\boldsymbol{x}^{(i)})|\boldsymbol{x}^{(i)}\}\ ,
\end{aligned} \tag{3.7}$$

where we in the first step inserted (3.6), in the second rearranged the integrals, in the third expressed using expectations and in the forth step used the GP mean function at a known position as in (1.1). Considering a zero-mean GP $f(\boldsymbol{x})$ as in Chapter 2, i.e. $m(\boldsymbol{x}) = 0$, we can now see that the approximate GP $g(\tilde{\boldsymbol{x}}^{(i)})$ is also zero-mean with mean function $m'(\tilde{\boldsymbol{x}}^{(i)}) = 0$.

GP covariance function

The covariance function of the approximate GP $g(\tilde{\boldsymbol{x}}^{(i)})$ will also be derived from the random variables $f(\boldsymbol{x}^{(i)})|\tilde{\boldsymbol{x}}^{(i)}$ and $f(\boldsymbol{x}^{(j)})|\tilde{\boldsymbol{x}}^{(j)}$ as

$$\begin{aligned}
k'(\tilde{\boldsymbol{x}}^{(i)}, \tilde{\boldsymbol{x}}^{(j)}) &= \mathrm{cov}\{f(\boldsymbol{x}^{(i)}),\ f(\boldsymbol{x}^{(j)})|\tilde{\boldsymbol{x}}^{(i)}, \tilde{\boldsymbol{x}}^{(j)}\} \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} f_i f_j p_{f(\boldsymbol{x}^{(i)}), f(\boldsymbol{x}^{(j)})|\tilde{\boldsymbol{x}}^{(i)}, \tilde{\boldsymbol{x}}^{(j)}}(f_i, f_j)\, df_i\, df_j \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} f_i f_j p_{f(\boldsymbol{x}^{(i)}))|\tilde{\boldsymbol{x}}^{(i)}}(f_i) p_{f(\boldsymbol{x}^{(j)}))|\tilde{\boldsymbol{x}}^{(j)}}(f_j)\, df_i\, df_j \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} f_i f_j \int_{\mathbb{R}^D} p_{f(\boldsymbol{x}^{(i)})|\boldsymbol{x}^{(i)}}(f_i) p(\boldsymbol{x}^{(i)}|\tilde{\boldsymbol{x}}^{(i)})\, d\boldsymbol{x}^{(i)} \\
&\qquad \cdot \int_{\mathbb{R}^D} p_{f(\boldsymbol{x}^{(j)})|\boldsymbol{x}^{(j)}}(f_j) p(\boldsymbol{x}^{(j)}|\tilde{\boldsymbol{x}}^{(j)})\, d\boldsymbol{x}^{(j)}\, df_i\, df_j \\
&= \int_{\mathbb{R}^D} \int_{\mathbb{R}^D} \int_{\mathbb{R}} \int_{\mathbb{R}} f_i f_j p_{f(\boldsymbol{x}^{(i)})|\boldsymbol{x}^{(i)}}(f_i) p_{f(\boldsymbol{x}^{(j)})|\boldsymbol{x}^{(j)}}(f_j)\, df_i\, df_j \\
&\qquad \cdot p(\boldsymbol{x}^{(i)}|\tilde{\boldsymbol{x}}^{(i)}) p(\boldsymbol{x}^{(j)}|\tilde{\boldsymbol{x}}^{(j)})\, d\boldsymbol{x}^{(i)}\, d\boldsymbol{x}^{(j)} \\
&= \int_{\mathbb{R}^D} \int_{\mathbb{R}^D} (k(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})|\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}) \\
&\qquad \cdot p(\boldsymbol{x}^{(i)}|\tilde{\boldsymbol{x}}^{(i)}) p(\boldsymbol{x}^{(j)}|\tilde{\boldsymbol{x}}^{(j)})\, d\boldsymbol{x}^{(i)}\, d\boldsymbol{x}^{(j)} \\
&= \mathrm{E}^{p(\boldsymbol{x}^{(i)}|\tilde{\boldsymbol{x}}^{(i)}) p(\boldsymbol{x}^{(j)}|\tilde{\boldsymbol{x}}^{(j)})}\{k(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})|\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}\}\ ,
\end{aligned} \tag{3.8}$$

where we sequentially expressed the covariance in integral form, assumed statistical independence between $p(f(\boldsymbol{x}^{(i)}))|\tilde{\boldsymbol{x}}^{(i)})$ and $p(f(\boldsymbol{x}^{(j)}))|\tilde{\boldsymbol{x}}^{(j)})$, inserted the pdf term

(3.6), rearranged the integrals, inserted the term for GP covariance function at known positions (1.2) and in the last step reformulated the expression using expectation. It shall be noted that $k(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})|\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}$ is a known covariance function of the original GP $f(\boldsymbol{x})$ operating on the known positions.

3.2 GPR Posterior Distribution

As we previously approximated the process $g(\tilde{\boldsymbol{x}}^{(i)})$ to be a GP, the derivations for posterior mean and variance from Chapter 1 hold and the predictive pdf is Gaussian with mean according to (1.21) expressed as

$$\mathrm{E}\{f(\boldsymbol{x}^{(*)})|\boldsymbol{y}, \tilde{\boldsymbol{x}}_{\mathrm{t}}\} = m(\boldsymbol{x}^{(*)}) + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{Q}^{-1}(\boldsymbol{y} - \boldsymbol{m}) \ , \tag{3.9}$$

which becomes after considering a zero-mean GP $f(\boldsymbol{x})$

$$\mathrm{E}\{f(\boldsymbol{x}^{(*)})|\boldsymbol{y}, \tilde{\boldsymbol{x}}_{\mathrm{t}}\} = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{y} \ . \tag{3.10}$$

The posterior variance is then similarly as in (1.22) expressed as

$$\mathrm{var}\{f(\boldsymbol{x}^{(*)})|\boldsymbol{y}, \tilde{\boldsymbol{x}}_{\mathrm{t}}\} = k(\boldsymbol{x}^{(*)}, \boldsymbol{x}^{(*)}) - \boldsymbol{c}^{\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{c} \ . \tag{3.11}$$

Here we must note that the cross covariance vector $\boldsymbol{c}$ and the covariance matrix $\boldsymbol{Q}$ are different to the ones used in Chapter 1 and 2 as they are now composed of elements calculated using the new covariance function mentioned previously in (3.8). The covariance matrix $\boldsymbol{Q}$ is, similarly as in (1.13), given by

$$\boldsymbol{Q} = \begin{pmatrix} k'(\tilde{\boldsymbol{x}}^{(1)}, \tilde{\boldsymbol{x}}^{(1)}) & k'(\tilde{\boldsymbol{x}}^{(1)}, \tilde{\boldsymbol{x}}^{(2)}) & \cdots & k'(\tilde{\boldsymbol{x}}^{(1)}, \tilde{\boldsymbol{x}}^{(I)}) \\ k'(\tilde{\boldsymbol{x}}^{(2)}, \tilde{\boldsymbol{x}}^{(1)}) & k'(\tilde{\boldsymbol{x}}^{(2)}, \tilde{\boldsymbol{x}}^{(2)}) & \cdots & k'(\tilde{\boldsymbol{x}}^{(2)}, \tilde{\boldsymbol{x}}^{(I)}) \\ \vdots & \vdots & \ddots & \vdots \\ k'(\tilde{\boldsymbol{x}}^{(I)}, \tilde{\boldsymbol{x}}^{(1)}) & k'(\tilde{\boldsymbol{x}}^{(I)}, \tilde{\boldsymbol{x}}^{(2)}) & \cdots & k'(\tilde{\boldsymbol{x}}^{(I)}, \tilde{\boldsymbol{x}}^{(I)}) \end{pmatrix} + \sigma_\epsilon^2 \mathbf{I}_I \ . \tag{3.12}$$

The cross covariance vector $\boldsymbol{c}$ is formed similarly as in (1.20) with the difference that it considers random training positions and known test position, which results in

$$\boldsymbol{c} = \left(k^*(\tilde{\boldsymbol{x}}^{(1)}, \boldsymbol{x}^{(*)}) \quad k^*(\tilde{\boldsymbol{x}}^{(2)}, \boldsymbol{x}^{(*)}) \quad \ldots \quad k^*(\tilde{\boldsymbol{x}}^{(I)}, \boldsymbol{x}^{(*)})\right)^{\mathrm{T}} \tag{3.13}$$

with individual elements formed using a new covariance function $k^*(\tilde{\boldsymbol{x}}^{(i)}, \boldsymbol{x}^{(*)})$ operating on random training positions $\tilde{\boldsymbol{x}}^{(i)}$ and known test position $\boldsymbol{x}^{(*)}$.

Using a similar argument as for $k'(\tilde{\boldsymbol{x}}^{(i)}, \tilde{\boldsymbol{x}}^{(j)})$ in (3.8), it can be shown that the individual covariances $k^*(\tilde{\boldsymbol{x}}^{(i)}, \boldsymbol{x}^{(*)})$ in $\boldsymbol{c}$ can be expressed as

$$\begin{aligned} k^*(\tilde{\boldsymbol{x}}^{(i)}, \boldsymbol{x}^{(*)}) &= \mathrm{cov}\{\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(*)}|\tilde{\boldsymbol{x}}^{(i)}\} \\ &= \int_{\mathbb{R}^D} (k(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(*)})|\boldsymbol{x}^{(i)}) p(\boldsymbol{x}^{(i)}|\tilde{\boldsymbol{x}}^{(i)}) \, d\boldsymbol{x}^{(i)} \\ &= \mathrm{E}^{p(\boldsymbol{x}^{(i)}|\tilde{\boldsymbol{x}}^{(i)})}\{k(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(*)})|\boldsymbol{x}^{(i)}\} \ . \end{aligned} \tag{3.14}$$

3.3 GP Covariance Functions

To be able to perform the GPR in closed-form we still need to find a closed-form expressions for the individual elements of the covariance matrix $\boldsymbol{Q}$ in (3.12) and the cross covariance vector $\boldsymbol{c}$ in (3.13), i.e. closed-form expressions for $k'(\tilde{\boldsymbol{x}}^{(i)}, \tilde{\boldsymbol{x}}^{(j)})$ and $k^*(\tilde{\boldsymbol{x}}^{(i)}, \boldsymbol{x}^{(*)})$. Considering a general form of the covariance function $k(\boldsymbol{x}, \boldsymbol{x}')$ and the training position posterior pdf $p(\boldsymbol{x}^{(i)}|\tilde{\boldsymbol{x}}^{(i)})$, a closed-form expression for k' and k^* can be obtained by approximating it with the Taylor series as suggested in [10].

To avoid approximation of the covariance function, we can choose a specific combination of $k(\boldsymbol{x}, \boldsymbol{x}')$ and $p(\boldsymbol{x}^{(i)}|\tilde{\boldsymbol{x}}^{(i)})$ leading to a closed-form of k' and k^* inherently. One of such combinations is a linear covariance function and Gaussian position posterior pdf as discussed in [3, page 43]. Another combination is the squared exponential covariance function and Gaussian position posterior pdf. This case will be further studied as it corresponds to the setup considered previously in Chapter 2. The covariance function is then in the same form as considered in (2.57), i.e. is expressed as

$$k(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})|\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)} = \sigma^2 \exp\left(-\frac{\|\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(j)}\|^2}{2\sigma_{\mathrm{x}}^2}\right) . \tag{3.15}$$

The assumed squared exponential covariance function in (3.15) can be expressed as a scaled Gaussian pdf according to

$$\begin{aligned} k(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})|\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)} &= \sigma^2 (2\pi)^{\frac{D}{2}} |\sigma_{\mathrm{x}}^2 \mathbf{I}_D|^{\frac{1}{2}} (2\pi)^{-\frac{D}{2}} |\sigma_{\mathrm{x}}^2 \mathbf{I}_D|^{-\frac{1}{2}} \\ &\qquad \cdot \exp\left(-\frac{1}{2}(\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(j)})^{\mathrm{T}} (\sigma_{\mathrm{x}}^2 \mathbf{I}_D)^{-1} (\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(j)})\right) \\ &= c\, \mathcal{N}(\boldsymbol{x}^{(i)};\, \boldsymbol{x}^{(j)},\, \sigma_{\mathrm{x}}^2 \mathbf{I}_D) , \end{aligned} \tag{3.16}$$

where $c = \sigma^2 (2\pi)^{\frac{D}{2}} |\sigma_{\mathrm{x}}^2 \mathbf{I}_D|^{\frac{1}{2}}$ is the scaling constant and D is the size of a position vector $\boldsymbol{x}$.

The Gaussian position posterior pdf corresponds to the setup in section 2.3 considering the improper prior. The posterior pdf for a single spatial dimension in (2.46) can be extended using the previous assumptions of independence in individual spatial dimensions into single training position posterior pdf as

$$p(\boldsymbol{x}^{(i)}|\tilde{\boldsymbol{x}}^{(i)}) = \mathcal{N}(\boldsymbol{x}^{(i)};\, \tilde{\boldsymbol{x}}^{(i)},\, \sigma_{\mathrm{v}}^2 \mathbf{I}_D) . \tag{3.17}$$

We could possibly also choose the Gaussian prior from section 2.3, which also results in the Gaussian posterior pdf, but the choice of improper prior simplifies the derivation as then the observation $\tilde{\boldsymbol{x}}^{(i)}$ represents the mean of the posterior pdf.

GP covariance function for single random position

For simplicity we will start by expressing the covariance function $k^*(\tilde{\boldsymbol{x}}^{(i)}, \boldsymbol{x}^{(*)})$ of a random training position and a known test position. By inserting (3.16) and (3.17) into (3.14) we obtain

$$\begin{aligned} k^*(\tilde{\boldsymbol{x}}^{(i)}, \boldsymbol{x}^{(*)}) &= \int_{\mathbb{R}^D} (k(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(*)}) | \boldsymbol{x}^{(i)}) p(\boldsymbol{x}^{(i)} | \tilde{\boldsymbol{x}}^{(i)}) \, d\boldsymbol{x}^{(i)} \\ &= \int_{\mathbb{R}^D} c \, \mathcal{N}(\boldsymbol{x}^{(i)};\ \boldsymbol{x}^{(*)},\ \sigma_{\mathrm{x}}^2 \mathbf{I}_D) \mathcal{N}(\boldsymbol{x}^{(i)};\ \tilde{\boldsymbol{x}}^{(i)},\ \sigma_{\mathrm{v}}^2 \mathbf{I}_D) \, d\boldsymbol{x}^{(i)} \ . \end{aligned} \tag{3.18}$$

To simplify the product of two Gaussian pdfs we use the formula from appendix A.2. This enables us to express it as a single scaled Gaussian pdf given as

$$\begin{aligned} k^*(\tilde{\boldsymbol{x}}^{(i)}, \boldsymbol{x}^{(*)}) &= \int_{\mathbb{R}^D} c \, z \, \mathcal{N}(\boldsymbol{x}^{(i)};\ \boldsymbol{z},\ \boldsymbol{Z}) \, d\boldsymbol{x}^{(i)} \\ &= c \, z \int_{\mathbb{R}^D} \mathcal{N}(\boldsymbol{x}^{(i)};\ \boldsymbol{z},\ \boldsymbol{Z}) \, d\boldsymbol{x}^{(i)} \\ &= c \, z \ , \end{aligned} \tag{3.19}$$

where z is the scaling constant evaluated according to (A.8). $\boldsymbol{z}$ and $\boldsymbol{Z}$ are the mean and the covariance matrix of the resulting normalized pdf, which integrates to 1. By inserting the term for constant c from (3.16) and the term for z we obtain

$$\begin{aligned} k^*(\tilde{\boldsymbol{x}}^{(i)}, \boldsymbol{x}^{(*)}) &= \sigma^2 (2\pi)^{\frac{D}{2}} |\sigma_{\mathrm{x}}^2 \mathbf{I}_D + \sigma_{\mathrm{v}}^2 \mathbf{I}_D|^{\frac{1}{2}} (2\pi)^{-\frac{D}{2}} |\sigma_{\mathrm{x}}^2 \mathbf{I}_D|^{-\frac{1}{2}} \\ &\quad \cdot \exp\left(-\frac{1}{2} (\boldsymbol{x}^{(*)} - \tilde{\boldsymbol{x}}^{(i)})^{\mathrm{T}} (\sigma_{\mathrm{x}}^2 \mathbf{I}_D + \sigma_{\mathrm{v}}^2 \mathbf{I}_D)^{-1} (\boldsymbol{x}^{(*)} - \tilde{\boldsymbol{x}}^{(i)}) \right) \\ &= \sigma^2 \left(\frac{\sigma_{\mathrm{x}}^2}{\sigma_{\mathrm{x}}^2 + \sigma_{\mathrm{v}}^2} \right)^{-\frac{D}{2}} \exp\left(-\frac{\|\boldsymbol{x}^{(*)} - \tilde{\boldsymbol{x}}^{(i)}\|^2}{2(\sigma_{\mathrm{x}}^2 + \sigma_{\mathrm{v}}^2)} \right) \ . \end{aligned} \tag{3.20}$$

GP covariance function for 2 random positions

Next we shall express the closed-form expression for the covariance function $k'(\tilde{\boldsymbol{x}}^{(i)}, \tilde{\boldsymbol{x}}^{(j)})$ of two random training positions. For that we will apply the same procedure as described in equations (3.18) to (3.20) twice. The term in (3.8) can be in this way developed as

$$\begin{aligned} k'(\tilde{\boldsymbol{x}}^{(i)}, \tilde{\boldsymbol{x}}^{(j)}) &= \int_{\mathbb{R}^D} \int_{\mathbb{R}^D} (k(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}) | \boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}) \\ &\qquad \cdot p(\boldsymbol{x}^{(i)} | \tilde{\boldsymbol{x}}^{(i)}) p(\boldsymbol{x}^{(j)} | \tilde{\boldsymbol{x}}^{(j)}) \, d\boldsymbol{x}^{(i)} \, d\boldsymbol{x}^{(j)} \\ &= \int_{\mathbb{R}^D} \int_{\mathbb{R}^D} c \, \mathcal{N}(\boldsymbol{x}^{(i)};\ \boldsymbol{x}^{(j)},\ \sigma_{\mathrm{x}}^2 \mathbf{I}_D) \\ &\qquad \cdot \mathcal{N}(\boldsymbol{x}^{(i)};\ \tilde{\boldsymbol{x}}^{(i)},\ \sigma_{\mathrm{v}}^2 \mathbf{I}_D) \mathcal{N}(\boldsymbol{x}^{(j)};\ \tilde{\boldsymbol{x}}^{(j)},\ \sigma_{\mathrm{v}}^2 \mathbf{I}_D) \, d\boldsymbol{x}^{(i)} \, d\boldsymbol{x}^{(j)} \\ &\dots \\ &= \sigma^2 \left(\frac{\sigma_{\mathrm{x}}^2}{\sigma_{\mathrm{x}}^2 + 2\sigma_{\mathrm{v}}^2} \right)^{-\frac{D}{2}} \exp\left(-\frac{\|\tilde{\boldsymbol{x}}^{(i)} - \tilde{\boldsymbol{x}}^{(j)}\|^2}{2(\sigma_{\mathrm{x}}^2 + 2\sigma_{\mathrm{v}}^2)} \right) \ . \end{aligned} \tag{3.21}$$

With the terms in (3.20) and (3.21) we can construct the cross covariance vector $\boldsymbol{c}$ and covariance matrix $\boldsymbol{Q}$ in closed-form, which enables us to perform GPR and obtain posterior mean at known test position according to (3.10) and posterior variance according to (3.10). We note that the currently described approach is similar to the direct usage of the GPR noisy observation of training positions in Chapter 2 on page 46. The difference to the previous approach is that the currently derived covariance functions k' and k^* have a widened length scale σ_{x}^2 and the vertical scale σ^2 is decreased.

3.4 Simulation Results

We performed a simulation with an identical setup as in section 2.4 to evaluate the performance of the current prediction method based on the uncertain covariance functions. We compared it to the standard GPR method operating directly on the observed training positions. We also compared it to the GPR method accounting for training positions uncertainty using MC as derived in section 2.2 with a difference in the prior distribution of training positions. Here we consider the prior distribution to be improper as in Chapter 2 page 38 because it results in a Gaussian posterior distribution with mean located at the observation of training positions. This posterior is the same as the one considered in section 3.3 and therefore it enables to perform a more relevant comparison.

In figure 3.1 you can see a comparison of the discussed methods in a single simulation performing the posterior mean evaluation. We can see that the currently derived GPR method with uncertain covariance functions (d) produces an almost identical result as the GPR operating directly on the observation of training positions (b). At the same time, this prediction shows a higher error in comparison to the true GP realization (a) than the GPR MC method (c).

The posterior variance of the discussed methods in a single simulation scenario can be found in figure 3.2. There we can observe that the current GPR method based on uncertain covariance functions (c) produces a slightly increased level of the posterior variance, i.e. a higher uncertainty in the produced predictions, which is the desired effect of incorporating training positions uncertainty.

By performing 100 simulation runs we obtained results used to evaluate the performance of discussed methods in terms of the RMSE and MNLPD metrics. The results can be found in table 3.1. As expected, The GPR method based on uncertain covariance functions shows only a little decrease in the RMSE in comparison to the GPR method using the observation of training positions directly. Still, the GPR method using MC outperforms both the other methods in terms of RMSE. The situation is different with the MNLPD metric, where the GPR method based on

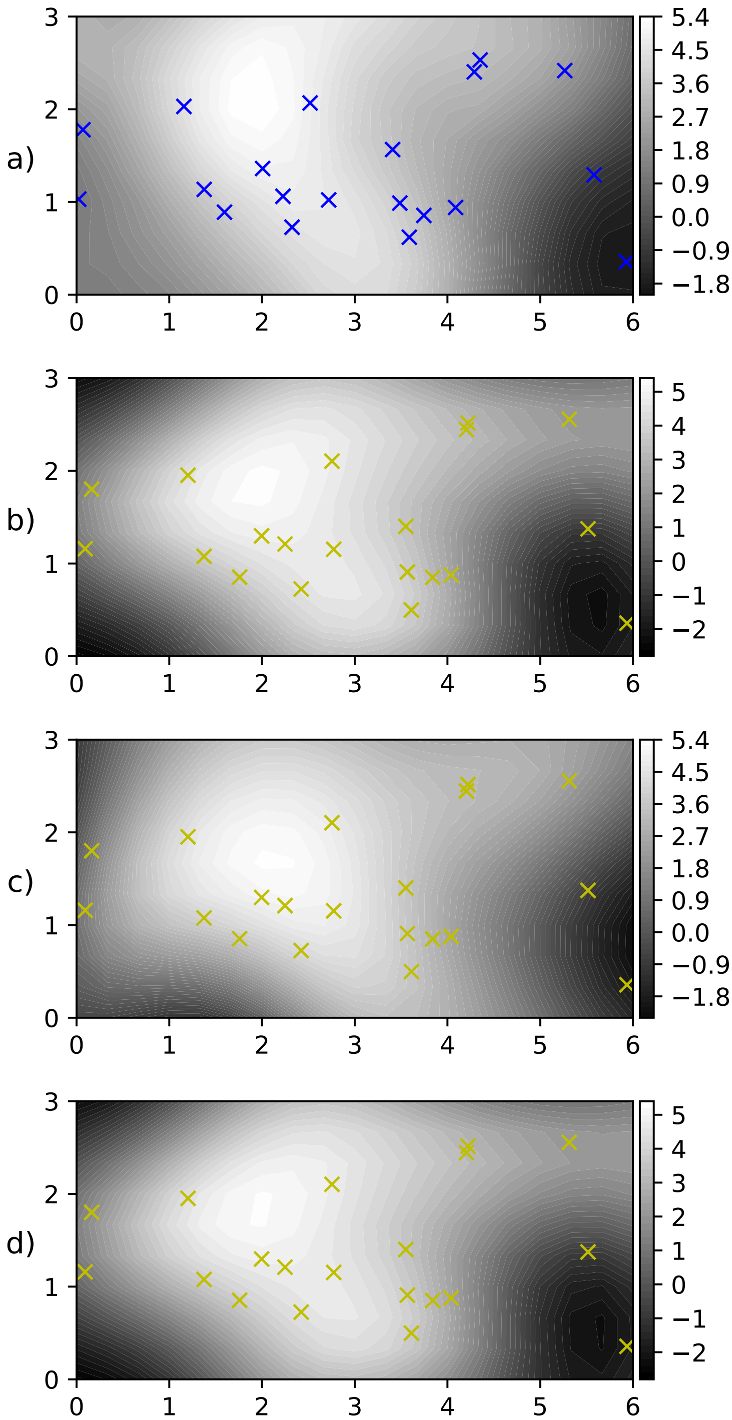


Fig. 3.1: GP estimation:
a) Realization of a GP and true training positions (indicated by blue crosses),
b) GP estimates obtained using the observed training positions (indicated by yellow crosses) directly,
c) GP estimates obtained using the observed training positions and the MC approximation with improper position prior,
d) GP estimates obtained using uncertain covariance functions.
The same scale is used in (a)-(d).

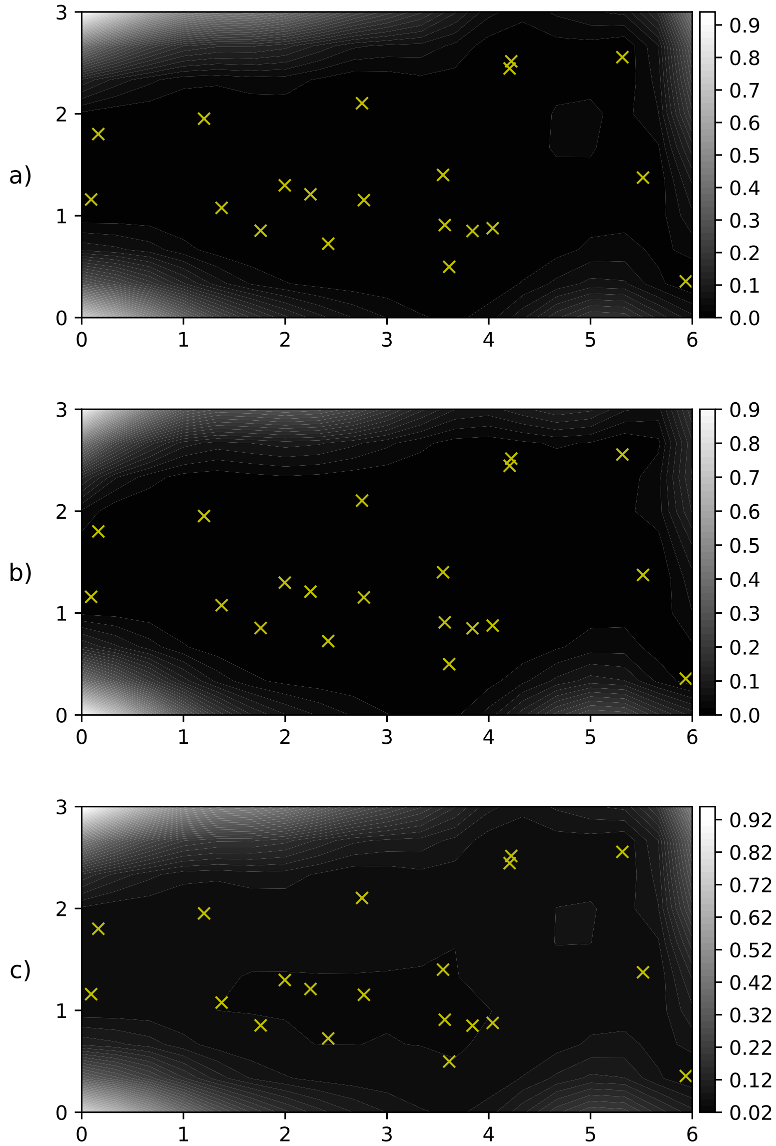


Fig. 3.2: GPR variance estimation:
a) GP variance estimation using the observed training positions (indicated by yellow crosses) directly,
b) GP variance estimation using the observed training positions and the MC approximation with improper position prior,
c) GP variance estimation using uncertain covariance functions.
The same scale is used in (a)-(c)

uncertain covariance functions shows a significant improvement over both the other methods, which accounts to the increased level of the posterior variance and therefore the prediction uncertainty.

GPR method	Evaluated RMSE	Evaluated MNLPD
GPR using the observed training positions directly	1.516	35.54
GPR using the observed training positions & the MC while accounting for uncertainty	0.8398	16.78
GPR using the observed training positions & uncertain covariance functions	1.505	5.786

Tab. 3.1: Comparison of the direct application of the noisy training positions to GPR, the Monte Carlo method derived in chapter 2 with improper positions prior and the method using uncertain covariance functions derived in this chapter in terms of the RMSE and MNLPD metrics.

3.5 Avoiding Approximations in a Different Problem Formulation

As it was shown in section 3.1 there are significant approximations needed to be able to perform GPR in closed-form under uncertain training positions. However, there exists a different prediction problem setup, where these assumptions are not needed and the expressions for the posterior mean and variance can be expressed exactly in closed-form. This setup was considered in [3, ch.3] as a first step to incorporate position uncertainty in general. In this problem setup we will consider the training positions $\boldsymbol{x}_{\mathrm{t}}$ to be known, while the test position $\boldsymbol{x}^{(*)}$ is considered random with a known pdf $p(\boldsymbol{x}^{(*)}|\tilde{\boldsymbol{x}}^{(*)})$ conditioned on a test position observation $\tilde{\boldsymbol{x}}^{(*)}$.

The predictive distribution conditioned on the test position observation $\tilde{\boldsymbol{x}}^{(*)}$ can be developed as

$$\begin{aligned} p(f_*|\boldsymbol{y}, \tilde{\boldsymbol{x}}^{(*)}) &= \int_{\mathbb{R}^D} p(f_*, \boldsymbol{x}^{(*)}|\boldsymbol{y}, \tilde{\boldsymbol{x}}^{(*)})\, d\tilde{\boldsymbol{x}}^{(*)} \\ &= \int_{\mathbb{R}^D} p(f_*|\boldsymbol{x}^{(*)}, \boldsymbol{y}, \tilde{\boldsymbol{x}}^{(*)}) p(\boldsymbol{x}^{(*)}|\boldsymbol{y}, \tilde{\boldsymbol{x}}^{(*)})\, d\tilde{\boldsymbol{x}}^{(*)} \\ &= \int_{\mathbb{R}^D} p(f_*|\boldsymbol{x}^{(*)}, \boldsymbol{y}) p(\boldsymbol{x}^{(*)}|\tilde{\boldsymbol{x}}^{(*)})\, d\tilde{\boldsymbol{x}}^{(*)} \\ &= \mathrm{E}^{p(\boldsymbol{x}^{(*)}|\tilde{\boldsymbol{x}}^{(*)})}\{p(f_*|\boldsymbol{x}^{(*)}, \boldsymbol{y})\} \end{aligned} \tag{3.22}$$

where we in the first step used the sum rule and in the second the product rule. In the third step we considered the conditional independence of f_* of $\tilde{\boldsymbol{x}}^{(*)}$ given $\boldsymbol{x}^{(*)}$ and also the conditional independence of $\boldsymbol{x}^{(*)}$ of $\boldsymbol{y}$ given $\tilde{\boldsymbol{x}}^{(*)}$. Since $p(f_*|\boldsymbol{x}^{(*)}, \boldsymbol{y})$ is not a linear function of $\boldsymbol{x}^{(*)}$, $p(f_*|\boldsymbol{y}, \tilde{\boldsymbol{x}}^{(*)})$ is not Gaussian. Just as in Chapter 2 we will approximate this posterior pdf as a Gaussian by expressing its mean and variance. The posterior mean can be developed as

$$\begin{aligned}
\mathrm{E}^{p(f_*|\boldsymbol{y},\tilde{\boldsymbol{x}}^{(*)})}\{f_*\} &= \int_{\mathbb{R}} f_* p(f_*|\boldsymbol{y}, \tilde{\boldsymbol{x}}^{(*)})\, df_* \\
&= \int_{\mathbb{R}} f_* \int_{\mathbb{R}^D} p(f_*|\boldsymbol{x}^{(*)}, \boldsymbol{y}) p(\boldsymbol{x}^{(*)}|\tilde{\boldsymbol{x}}^{(*)})\, d\tilde{\boldsymbol{x}}^{(*)}\, df_* \\
&= \int_{\mathbb{R}^D} \int_{\mathbb{R}} f_* p(f_*|\boldsymbol{x}^{(*)}, \boldsymbol{y})\, df_* p(\boldsymbol{x}^{(*)}|\tilde{\boldsymbol{x}}^{(*)})\, d\tilde{\boldsymbol{x}}^{(*)} \\
&= \int_{\mathbb{R}^D} \mu(\boldsymbol{x}^{(*)}, \boldsymbol{y}) p(\boldsymbol{x}^{(*)}|\tilde{\boldsymbol{x}}^{(*)})\, d\tilde{\boldsymbol{x}}^{(*)} \\
&= \mathrm{E}^{p(\boldsymbol{x}^{(*)}|\tilde{\boldsymbol{x}}^{(*)})}\{\mu(\boldsymbol{x}^{(*)}, \boldsymbol{y})\} \ ,
\end{aligned} \tag{3.23}$$

where we firstly inserted (3.22), then rearranged the integrals and in the last step recognized the term for the GPR posterior mean $\mu(\boldsymbol{x}^{(*)}, \boldsymbol{y})$ operating on known training and test position. Replacing this term by the GPR predictive mean in (1.21) we can further develop (3.23) as

$$\begin{aligned}
\mathrm{E}^{p(f_*|\boldsymbol{y},\tilde{\boldsymbol{x}}^{(*)})}\{f_*\} &= \mathrm{E}^{p(\boldsymbol{x}^{(*)}|\tilde{\boldsymbol{x}}^{(*)})}\{m(\boldsymbol{x}^{(*)}) + \boldsymbol{c}^{\mathrm{T}}(\boldsymbol{x}^{(*)})\boldsymbol{Q}^{-1}(\boldsymbol{y} - \boldsymbol{m})\} \\
&= \mathrm{E}^{p(\boldsymbol{x}^{(*)}|\tilde{\boldsymbol{x}}^{(*)})}\{\boldsymbol{c}^{\mathrm{T}}(\boldsymbol{x}^{(*)})\boldsymbol{Q}^{-1}\boldsymbol{y}\} \\
&= \mathrm{E}^{p(\boldsymbol{x}^{(*)}|\tilde{\boldsymbol{x}}^{(*)})}\{\boldsymbol{c}(\boldsymbol{x}^{(*)})\}^{\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{y} \ ,
\end{aligned} \tag{3.24}$$

where we considered a zero-mean GP and moved the terms independent of $\boldsymbol{x}^{(*)}$ out of the expectation. We notice that the covariance matrix $\boldsymbol{Q}$ can be calculated directly using the standard covariance function as training positions are known. What is left is to evaluate the expectation of the cross covariance vector $\boldsymbol{c}(\boldsymbol{x}^{(*)})$, which can be expressed as the expectation of the individual vector elements as

$$\mathrm{E}^{p(\boldsymbol{x}^{(*)}|\tilde{\boldsymbol{x}}^{(*)})}\{\boldsymbol{c}(\boldsymbol{x}^{(*)})\} = \begin{pmatrix} \mathrm{E}^{p(\boldsymbol{x}^{(*)}|\tilde{\boldsymbol{x}}^{(*)})}\{k(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(*)})\} \\ \mathrm{E}^{p(\boldsymbol{x}^{(*)}|\tilde{\boldsymbol{x}}^{(*)})}\{k(\boldsymbol{x}^{(2)}, \boldsymbol{x}^{(*)})\} \\ \vdots \\ \mathrm{E}^{p(\boldsymbol{x}^{(*)}|\tilde{\boldsymbol{x}}^{(*)})}\{k(\boldsymbol{x}^{(I)}, \boldsymbol{x}^{(*)})\} \end{pmatrix} . \tag{3.25}$$

If we further consider the case of the square exponential covariance function k as in (3.15) and the Gaussian posterior pdf $p(\boldsymbol{x}^{(*)}|\tilde{\boldsymbol{x}}^{(*)})$ of the test position similarly as in (3.17), the individual elements of this vector can be evaluated using the same procedure as in section 3.3, just with the difference that we take the expectation with respect to the test position and not the training position. Following the derivation of (3.20) with this slight modification we obtain the elements of the vector $\boldsymbol{c}(\boldsymbol{x}^{(*)})$ in a closed-form expression as

$$\mathrm{E}^{p(\boldsymbol{x}^{(*)}|\tilde{\boldsymbol{x}}^{(*)})}\{k(\boldsymbol{x}^{(i)},\boldsymbol{x}^{(*)})\} = \sigma^2 \left(\frac{\sigma_{\mathrm{x}}^2}{\sigma_{\mathrm{x}}^2+\sigma_{\mathrm{v}}^2}\right)^{-\frac{D}{2}} \exp\left(-\frac{\|\tilde{\boldsymbol{x}}^{(*)}-\boldsymbol{x}^{(i)}\|^2}{2(\sigma_{\mathrm{x}}^2+\sigma_{\mathrm{v}}^2)}\right) . \qquad (3.26)$$

A closed-form expression for the posterior variance can be derived in a similar manner. This derivation shows that in the scenario considering uncertain test position and known training positions and a suitable choice of the covariance function and the test position distribution, the closed-form expressions for the posterior mean and variance are exact. This confirms the results in [3, Ch.3] describing this scenario. However we must keep in mind that for uncertain training positions the estimation task is more complicated as described earlier in this chapter.

Conclusion

Summary

In Chapter 1 of this thesis, we investigated the theory behind Gaussian Process Regression (GPR) and applied it to the task of spatial function estimation with uncertain training positions. This was motivated by the expectation that properly accounting for training position uncertainty will result in a better performance than standard GPR.

In Chapter 2, based on the method suggested in [2], we derived exact integral expressions for the posterior distribution as well as the posterior mean and variance. Because these expressions can not be evaluated in closed form, we resorted to an approximate evaluation by means of Monte Carlo (MC) sampling method. We also considered different choices of the prior distribution of the training positions and described how to draw samples from the posterior distribution of the training positions. Finally, we described a simplified model disregarding certain statistical dependencies, which led to simplified expressions of the posterior mean and variance as presented in [14].

In Chapter 3, we a closed form expression for GPR under training position uncertainty as reported in [2], [3] and [9]. Whereas this approach does not require an approximation as in Chapter 2, we demonstrated that it relies on strong approximating assumptions. This makes the closed form approach inherently approximate even though no MC approximation is needed. On the other hand, the computational complexity is lower because only a single matrix inversion is required, instead of one matrix inversion for each MC sample. Finally, we discussed a relevant method in [3], where a different scenario with uncertain test position is used. We showed that in this scenario the approximating assumptions are not necessary and the terms for the posterior distribution in closed form are exact.

Performance evaluation through simulations

To support our understanding, confirm the underlying theory, and evaluate the performance of the discussed methods, we created a simulation script in PYTHON . The script allowed us to obtain and visualize numerical results of a single simulation trial and also to perform multiple simulation trials and average the results over all trials. These results were used to evaluate and compare the performance of the discussed GPR methods.

The MC GPR method discussed in Chapter 2 was observed to outperform the method using the observed training positions directly (naively) in terms of two different performance metrics (RMSE and MNLPD). The simplified GPR method disregarding certain statistical dependencies produced an increased RMSE, which indicates a poorer performance. However, the posterior variance in a single simulation trial was observed to be significantly increased in regions far from the training positions. This led to a decrease in the MNLPD metric, which indicates a better performance.

The GPR method in Chapter 3 exhibited only a slightly smaller RMSE than the method using the observed training position directly (naively). At the same time, its RMSE was significantly lower than that of the MC GPR method from Chapter 2. On the other hand, the MNLPD from Chapter 3 was observed to be significantly smaller than the MNLPD of both the method using the observed training positions directly and the MC GPR method.

Contributions

To our knowledge, a detailed derivation of the MC GPR method such as the one provided in Chapter 2 is not available in the literature. We also discovered a common misconception present in the literature regarding omitting certain statistical dependencies of the model, which is discussed in Section 2.5. We conjecture that this misconception was caused by a misinterpretation of the derivation in [3, ch. 3], which was carried out for different problem scenario. Furthermore, we formulated the approximations that are required to obtain the closed-form GPR method from Chapter 3. Finally, we believe that a comparison of the performance of the closed-form method in Chapter 3 with the MC GPR from Chapter 2 considering all the relevant statistical dependencies, as presented in Section 2.1, cannot be found in the literature.

Further topics

We investigated possible localization algorithms that could be employed to obtain training position observations, to be used as input to the considered GPR methods. One such algorithm is the *nonparametric belief propagation algorithm* considered for distributed localization of sensor nodes in [11]. In particular, combining the MC GPR method of Section 2 with this localization method would be interesting because of the nonparametric representation of the distribution of training positions using samples, which could be directly used by the MC GPR method.

Besides the MC approach considered in this thesis, another approach to dealing with uncertain training positions within GPR is provided by *stochastic variational inference* [12].

Bibliography

[1] RASMUSSEN, Carl Edward. Gaussian Processes for Machine Learning. the MIT Press, 2006.

[2] JADALIHA, Mahdi, Yunfei XU, Jongeun CHOI, Nicholas S. JOHNSON a Weiming LI. Gaussian Process Regression for Sensor Networks Under Localization Uncertainty. IEEE Transactions on Signal Processing. 2013, 61(2), 223-237. ISSN 1053-587X. Dostupné z: doi:10.1109/TSP.2012.2223695

[3] GIRARD, Agathe. Approximate methods for propagation of uncertainty with gaussian process models. 2004. Doctoral dissertation. University of Glasgow.

[4] Advances in Applied Probability. 5. 1973. ISSN 0001-8678. Dostupné také z: `https://www.cambridge.org/core/product/identifier/S0001867800039379`

[5] The Statistician. 43. 1994. ISSN 00390526. Dostupné také z: `https://www.jstor.org/stable/10.2307/2348933?origin=crossref`

[6] RASMUSSEN, Carl Edward. Evaluation of gaussian processes and other methods for non-linear regression. 1996. Ph.D. thesis. University of Toronto. Vedoucí práce Geoffrey Hinton.

[7] BISHOP, Christopher M. Pattern recognition and machine learning. [New York]: Springer, c2006. Information science and statistics. ISBN 978-0-387-31073-2.

[8] ANDRIEU, Christophe, Nando DE FREITAS, Arnaud DOUCET a Michael I. JORDAN. An Introduction to MCMC for Machine Learning. Machine Learning. 50(1/2), 5-43. ISSN 08856125. Dostupné z: doi:10.1023/A:1020281327116

[9] SEEGER, Matthias. Bayesian Gaussian Process Models: PAC-Bayesian Generalisation Error Bounds and Sparse Approximations. 2003.

[10] GIRARD, Agathe a Roderick MURRAY-SMITH. Learning a Gaussian Process Model with Uncertain Inputs. 2003.

[11] IHLER, A.T., J.W. FISHER, R.L. MOSES a A.S. WILLSKY. Nonparametric belief propagation for self-localization of sensor networks. IEEE Journal on Selected Areas in Communications. 2005, 23(4), 809-819. ISSN 0733-8716. Dostupné z: doi:10.1109/JSAC.2005.843548

[12] TITSIAS, Michalis K. Variational learning of inducing variables in sparse Gaussian processes. In Artificial Intelligence and Statistics 12. 2009, , 567-574.

[13] PETERSEN, Kaare Brandt a Michael Syskind PEDERSEN. The Matrix Cookbook. 2012.

[14] MUPPIRISETTY, L. Srikar, Tommy SVENSSON a Henk WYMEERSCH. Spatial Wireless Channel Prediction under Location Uncertainty. IEEE Transactions on Wireless Communications. 2016, 15(2), 1031-1044. ISSN 1536-1276. Dostupné z: doi:10.1109/TWC.2015.2481879

[15] CERVONE, Daniel a Natesh S. PILLAI. Gaussian Process Regression with Location Errors. ArXiv:1506.08256. 2015, 28.

[16] FROHLE, Markus, L. Srikar MUPPIRISETTY a Henk WYMEERSCH. Channel gain prediction for multi-agent networks in the presence of location uncertainty. 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2016, 2016, , 3911-3915. ISBN 978-1-4799-9988-0. Dostupné z: doi:10.1109/ICASSP.2016.7472410

Notation

$\mathbb{R}$	Real numbers
$\mathbb{R}^i$	Space spanned by a column vector of real numbers of dimension i
$\mathbb{R}^{i\times j}$	Space spanned by a matrix of real numbers with i rows and j columns
x	Scalar
$\boldsymbol{x}$	Column vector
$\boldsymbol{X}$	Matrix
$\mathbf{I}_i$	Identity matrix of dimensions $i \times i$
$\lvert x\rvert$	Absolute value of a variable x
$\lVert\boldsymbol{x}\rVert$	Norm of a vector $\boldsymbol{x}$
$\lvert\boldsymbol{X}\rvert$	Determinant of a matrix
$\boldsymbol{X}^{-1}$	Inverse of a matrix
$\cdot^{\mathrm{T}}$	Transpose of a vector or matrix
$\tilde{\cdot}$	Observation of a random quantity
$\mathrm{col}(x, y)$	Column vector composed of elements x at top and y at bottom
$p(x)$	Probability density function of a random variable x
$p(x, y)$	Joint probability density function of a random variables x and y
$p(x\|y)$	Conditional probability density function of a random variable x given random variable y
$p(x; y)$	Probability density function of a random variable x parametrized by a deterministic parameter y
$p_y(x)$	Probability density function of a random variable y evaluated at the position x
$\mathrm{E}^{p(x)}\{x\}$	Expectation of a variable x with respect to a probability density $p(x)$
$\mathrm{var}^{p(x)}\{x\}$	Variance of a random variable x with respect to a probability density $p(x)$
$\mathrm{cov}\{x,\ y\}$	Covariance of random variables x and y with respect to the corresponding joint probability density $p(x, y)$
$\mathrm{cov}\{\boldsymbol{x}\}$	Covariance matrix of a random vector $\boldsymbol{x}$ with respect to the corresponding probability density $p(\boldsymbol{x})$
Λ_x	Precision of a random variable x
$\boldsymbol{\Lambda}_{x,y}$	Cross precision vector of a random variable x with random vector $\boldsymbol{y}$
$\boldsymbol{\Lambda}_y$	Precision matrix of a random vector $\boldsymbol{y}$
μ_x	Mean of a random variable x equivalent to $\mathrm{E}^{p(x)}\{x\}$
$\boldsymbol{\mu}_x$	Mean of a random vector $\boldsymbol{x}$ equivalent to $\mathrm{E}^{p(\boldsymbol{x})}\{\boldsymbol{x}\}$
σ_x	Standard deviation of a random variable x
σ_x^2	Variance of a random variable x equivalent to $\mathrm{var}^{p(x)}\{x\}$

$\mathcal{N}(x;\, \mu,\, \sigma^2)$	Univariate Gaussian pdf with mean μ and variance σ^2 evaluated at x
$\mathcal{N}(\boldsymbol{x};\, \boldsymbol{\mu},\, \boldsymbol{C})$	Multivariate Gaussian pdf with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{C}$ evaluated at $\boldsymbol{x}$
$\mathcal{U}(\mathcal{X})$	Uniform distribution over subspace $\mathcal{X}$
$\propto$	Proportionality up to a multiplicative constant
$\approx$	Approximately equals
$\sim$	Distributed according to; formal similarity of expressions
const.	Constant factor shaped according to context
$\delta(\cdot)$	Dirac delta function
x_s	Realization, sample of a random variable x distributed according to a probability density described in the context

Symbols and abbreviations

GP Gaussian Process

GPR Gaussian Process Regression

MC Monte Carlo

RMSE Root-mean-square error

MNLPD Mean negative log-predictive density

pdf Probability density function

iid Independent and identically distributed

List of appendices

A Mathematical formulas

Here is located a list of common mathematical formulas used within the thesis. Some of the formulas were included in the relevant literature but a comprehensive overview can be found in [13].

A.1 Conditional Gaussian pdf

Consider a two random vectors $\boldsymbol{x}_\mathrm{A}$ and $\boldsymbol{x}_\mathrm{A}$. The vectors are considered to be jointly Gaussian according to

$$p(\boldsymbol{x}_\mathrm{A}, \boldsymbol{x}_\mathrm{B}) = \mathcal{N}\left(\begin{pmatrix}\boldsymbol{x}_\mathrm{A}\\ \boldsymbol{x}_\mathrm{B}\end{pmatrix};\ \begin{pmatrix}\boldsymbol{\mu}_\mathrm{A}\\ \boldsymbol{\mu}_\mathrm{B}\end{pmatrix},\ \begin{pmatrix}\boldsymbol{C}_\mathrm{A} & \boldsymbol{C}_\mathrm{C}\\ \boldsymbol{C}_\mathrm{C}^\mathrm{T} & \boldsymbol{C}_\mathrm{B}\end{pmatrix}\right)\ , \tag{A.1}$$

where $\boldsymbol{\mu}_\mathrm{A}$ and $\boldsymbol{C}_\mathrm{A}$ are the mean and the covariance matrix of vector $\boldsymbol{x}_\mathrm{A}$. $\boldsymbol{C}_\mathrm{C}$ is the cross covariance matrix of vectors $\boldsymbol{x}_\mathrm{A}$ and $\boldsymbol{x}_\mathrm{B}$. The posterior distribution of $\boldsymbol{x}_\mathrm{A}$ given $\boldsymbol{x}_\mathrm{B}$ can be shown to be given as

$$p(\boldsymbol{x}_\mathrm{A}|\boldsymbol{x}_\mathrm{B}) = \mathcal{N}(\boldsymbol{x}_\mathrm{A};\ \boldsymbol{\mu}_\mathrm{A|B},\ \boldsymbol{C}_\mathrm{A|B})\ . \tag{A.2}$$

Here, the posterior mean is given as

$$\boldsymbol{\mu}_\mathrm{A|B} = \boldsymbol{\mu}_\mathrm{A} + \boldsymbol{C}_\mathrm{C}\boldsymbol{C}_\mathrm{B}^{-1}(\boldsymbol{\mu}_\mathrm{B} - \boldsymbol{\mu}_\mathrm{B})\ . \tag{A.3}$$

The posterior variance has the form

$$\boldsymbol{C}_\mathrm{A|B} = \boldsymbol{C}_\mathrm{A} - \boldsymbol{C}_\mathrm{C}\boldsymbol{C}_\mathrm{B}^{-1}\boldsymbol{C}_\mathrm{C}^\mathrm{T}\ . \tag{A.4}$$

A.2 Product of Gaussian pdfs

Consider two Gaussian pdf considered as a function of vector $\boldsymbol{x}$, i.e.

$$p_\mathrm{A}(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x};\ \boldsymbol{\mu}_\mathrm{A},\ \boldsymbol{C}_\mathrm{A}) \tag{A.5}$$

and

$$p_\mathrm{B}(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x};\ \boldsymbol{\mu}_\mathrm{B},\ \boldsymbol{C}_\mathrm{B})\ . \tag{A.6}$$

The product of the pdfs is then given by

$$\begin{aligned} p_\mathrm{A}(\boldsymbol{x})\, p_\mathrm{B}(\boldsymbol{x}) &= \mathcal{N}(\boldsymbol{x};\ \boldsymbol{\mu}_\mathrm{A},\ \boldsymbol{C}_\mathrm{A})\, \mathcal{N}(\boldsymbol{x};\ \boldsymbol{\mu}_\mathrm{B},\ \boldsymbol{C}_\mathrm{B}) \\ &= c_\mathrm{C}\, \mathcal{N}(\boldsymbol{x};\ \boldsymbol{\mu}_\mathrm{C},\ \boldsymbol{C}_\mathrm{C})\ . \end{aligned} \tag{A.7}$$

Here, the scaling constant is evaluated as

$$\begin{aligned} c_{\mathrm{C}} &= \mathcal{N}(\boldsymbol{\mu}_{\mathrm{A}};\ \boldsymbol{\mu}_{\mathrm{B}},\ \boldsymbol{C}_{\mathrm{A}} + \boldsymbol{C}_{\mathrm{B}}) \\ &= (2\pi)^{-\frac{D}{2}} |\boldsymbol{C}_{\mathrm{A}} + \boldsymbol{C}_{\mathrm{B}}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} (\boldsymbol{\mu}_{\mathrm{A}} - \boldsymbol{\mu}_{\mathrm{B}})^{\mathrm{T}} (\boldsymbol{C}_{\mathrm{A}} + \boldsymbol{C}_{\mathrm{B}})^{-1} (\boldsymbol{\mu}_{\mathrm{A}} - \boldsymbol{\mu}_{\mathrm{B}}) \right) \ , \end{aligned} \tag{A.8}$$

where D is the dimension of the vector $\boldsymbol{x}$. The mean of the normalized product distribution is

$$\boldsymbol{\mu}_{\mathrm{C}} = (\boldsymbol{C}_{\mathrm{A}}^{-1} + \boldsymbol{C}_{\mathrm{B}}^{-1})^{-1} (\boldsymbol{C}_{\mathrm{A}}^{-1} \boldsymbol{\mu}_{\mathrm{A}} + \boldsymbol{C}_{\mathrm{B}}^{-1} \boldsymbol{\mu}_{\mathrm{B}}) \ . \tag{A.9}$$

The covariance matrix of the normalized product distribution

$$\boldsymbol{C}_{\mathrm{C}} = (\boldsymbol{C}_{\mathrm{A}}^{-1} + \boldsymbol{C}_{\mathrm{B}}^{-1})^{-1} \ . \tag{A.10}$$

B Derivation of the Gaussian posterior distribution

To derive the posterior pdf $p(f_*|\boldsymbol{y})$ we employ the *completing the square* method. In this method, we express the exponent of the posterior pdf as a quadratic form. Let us denote the covariance matrix of the joint pdf $p(f_*, \boldsymbol{y})$ in (1.18) as $\boldsymbol{N}$, i.e.,

$$\boldsymbol{N} \triangleq \begin{pmatrix} k_* & \boldsymbol{c}^{\mathrm{T}} \\ \boldsymbol{c} & \boldsymbol{Q} \end{pmatrix} . \tag{B.1}$$

We will also consider the precision matrix of the joint distribution as

$$\boldsymbol{N}^{-1} = \boldsymbol{\Lambda}_{\mathrm{N}} = \begin{pmatrix} \Lambda_{f_*} & \boldsymbol{\Lambda}_{f_*,y}^{\mathrm{T}} \\ \boldsymbol{\Lambda}_{f_*,y} & \boldsymbol{\Lambda}_y \end{pmatrix} , \tag{B.2}$$

where the precision Λ_{f_*}, cross-precision vector $\boldsymbol{\Lambda}_{f_*,y}$ and precision matrix $\boldsymbol{\Lambda}_y$ correspond to the variance k_*, cross-covariance vector $\boldsymbol{c}$ and covariance matrix $\mathrm{cov}\{\mathbf{y}\}$ respectively. We can now develop the posterior distribution (1.17) as

$$\begin{aligned} p(f_*|\boldsymbol{y}) &= \frac{(2\pi)^{-\frac{I+1}{2}}|\boldsymbol{\Lambda}_{\mathrm{N}}|^{\frac{1}{2}}\exp\left(-\frac{1}{2}\left(\begin{pmatrix} f_* \\ \boldsymbol{y} \end{pmatrix} - \begin{pmatrix} m_* \\ \boldsymbol{m} \end{pmatrix}\right)^{\mathrm{T}} \boldsymbol{\Lambda}_{\mathrm{N}} \left(\begin{pmatrix} f_* \\ \boldsymbol{y} \end{pmatrix} - \begin{pmatrix} m_* \\ \boldsymbol{m} \end{pmatrix}\right)\right)}{(2\pi)^{-\frac{I}{2}}|\boldsymbol{Q}|^{-\frac{1}{2}}\exp(-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{m})^{\mathrm{T}}\boldsymbol{Q}^{-1}(\boldsymbol{y}-\boldsymbol{m}))} \\ p(f_*|\boldsymbol{y}) &= (2\pi)^{-\frac{1}{2}}|\boldsymbol{\Lambda}_{\mathrm{N}}|^{\frac{1}{2}}|\boldsymbol{Q}|^{\frac{1}{2}} \\ &\quad \cdot \exp\left(-\frac{1}{2}\left(\begin{pmatrix} f_* \\ \boldsymbol{y} \end{pmatrix} - \begin{pmatrix} m_* \\ \boldsymbol{m} \end{pmatrix}\right)^{\mathrm{T}} \boldsymbol{\Lambda}_{\mathrm{N}} \left(\begin{pmatrix} f_* \\ \boldsymbol{y} \end{pmatrix} - \begin{pmatrix} m_* \\ \boldsymbol{m} \end{pmatrix}\right) + \frac{1}{2}(\boldsymbol{y}-\boldsymbol{m})^{\mathrm{T}}\boldsymbol{Q}^{-1}(\boldsymbol{y}-\boldsymbol{m})\right) . \end{aligned} \tag{B.3}$$

For further reference, using (B.2) we develop the term in the rightmost exponent of (B.3) containing f_* as

$$\begin{aligned} &-\frac{1}{2}\left(\begin{pmatrix} f_* \\ \boldsymbol{y} \end{pmatrix} - \begin{pmatrix} m_* \\ \boldsymbol{m} \end{pmatrix}\right)^{\mathrm{T}} \boldsymbol{\Lambda}_{\mathrm{N}} \left(\begin{pmatrix} f_* \\ \boldsymbol{y} \end{pmatrix} - \begin{pmatrix} m_* \\ \boldsymbol{m} \end{pmatrix}\right) \\ &= -\frac{1}{2}(f_* - m_*)\Lambda_{f_*}(f_* - m_*) - \frac{1}{2}(f_* - m_*)\boldsymbol{\Lambda}_{f_*,y}^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{m}) \\ &\quad - \frac{1}{2}(\boldsymbol{y}-\boldsymbol{m})^{\mathrm{T}}\boldsymbol{\Lambda}_{f_*,y}(f_* - m_*) - \frac{1}{2}(\boldsymbol{y}-\boldsymbol{m})^{\mathrm{T}}\boldsymbol{\Lambda}_y(\boldsymbol{y}-\boldsymbol{m}) . \end{aligned} \tag{B.4}$$

Posterior mean and variance using precision submatrices

Here we are considering the posterior distribution in (1.17) as a function of f_*. This conditional distribution is under the stated assumption of joint Gaussianity again Gaussian, which is indicated by the presence of quadratic term f_*^2 in the exponent. Therefore, we can express it in terms of parameters, i.e., posterior mean $\mu_{f_*|y}$ and posterior variance $\sigma^2_{f_*|y}$ as

$$\begin{aligned} p(f_*|\boldsymbol{y}) = \mathcal{N}(f_*;\, \mu_{f_*|y},\, \sigma^2_{f_*|y}) &= (2\pi\sigma^2_{f_*|y})^{-\frac{1}{2}}\exp\left(-\frac{(f_* - \mu_{f_*|y})^2}{2\sigma^2_{f_*|y}}\right) \\ &= (2\pi\sigma^2_{f_*|y})^{-\frac{1}{2}}\exp\left(-\frac{f_*^2 - 2f_*\mu_{f_*|y} + \mu^2_{f_*|y}}{2\sigma^2_{f_*|y}}\right) \ . \end{aligned} \tag{B.5}$$

Considering that equations (B.3) and (B.5) express the same quantity, the factors of f_* in the rightmost exponent of each equation must be also equal. Now we can put equal the coefficients of quadratic terms f_*^2 in each equation

$$-\frac{1}{2}\Lambda_{f_*} f_*^2 = -\frac{1}{2\sigma^2_{f_*|y}} f_*^2 \ . \tag{B.6}$$

This equation can be rearranged to express directly the posterior variance

$$\sigma^2_{f_*|y} = \frac{1}{\Lambda_{f_*}} \ . \tag{B.7}$$

Further we shall consider the terms linear in f_* contained in rightmost exponent of (B.5) and of (B.4) and place them equal as follows

$$\begin{aligned} \frac{f_*\mu_{f_*|y}}{\sigma^2_{f_*|y}} &= -\frac{1}{2} f_*\Lambda_{f_*}(-m_*) - \frac{1}{2}(-m_*)\Lambda_{f_*} f_* \\ &\quad - \frac{1}{2} f_* \boldsymbol{\Lambda}^{\mathrm{T}}_{f_*,y}(\boldsymbol{y} - \boldsymbol{m}) - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{m})^{\mathrm{T}}\boldsymbol{c} f_* \ . \end{aligned} \tag{B.8}$$

Noticing the equality of the two rightmost terms in (B.8) we can further simplify into

$$\begin{aligned} \frac{f_*\mu_{f_*|y}}{\sigma^2_{f_*|y}} &= f_*\Lambda_{f_*} m_* - f_*\boldsymbol{\Lambda}^{\mathrm{T}}_{f_*,y}(\boldsymbol{y} - \boldsymbol{m}) \\ &= f_*(\Lambda_{f_*} m_* - \boldsymbol{\Lambda}^{\mathrm{T}}_{f_*,y}(\boldsymbol{y} - \boldsymbol{m})) \ . \end{aligned} \tag{B.9}$$

Now we can express the posterior mean value as

$$\mu_{f_*|y} = \sigma^2_{f_*|y}(\Lambda_{f_*} m_* - \boldsymbol{\Lambda}^{\mathrm{T}}_{f_*,y}(\boldsymbol{y} - \boldsymbol{m})) \ . \tag{B.10}$$

Considering that posterior variance equals inverse prior precision $\sigma^2_{f_*|y} = \Lambda^{-1}_{f_*}$ as shown in (B.7) we can unify the expression to use precision terms as

$$\mu_{f_*|y} = m_* - \Lambda^{-1}_{f_*}\boldsymbol{\Lambda}^{\mathrm{T}}_{f_*,y}(\boldsymbol{y} - \boldsymbol{m}) \ . \tag{B.11}$$

Posterior mean and variance using covariance submatrices

Until now we expressed the posterior distribution parameters using the joint distribution precision submatrices as in (B.2). To express the parameters using covariance submatrices as in (B.1) we will use the block matrix inversion formula according to [7] in a form

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{M} & -\boldsymbol{M}\boldsymbol{B}\boldsymbol{D}^{-1} \\ -\boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{M} & \boldsymbol{D}^{-1} + \boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{M}\boldsymbol{B}\boldsymbol{D}^{-1} \end{pmatrix} , \tag{B.12}$$

where $\boldsymbol{M}^{-1}$ is known as the *Schur complement* and is defined as

$$\boldsymbol{M}^{-1} = \boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C} . \tag{B.13}$$

Forming (B.1) and (B.2) into single equation according to $\boldsymbol{N}^{-1} = \boldsymbol{\Lambda}_{\mathrm{N}}$ gives us

$$\begin{pmatrix} k_* & \boldsymbol{c}^{\mathrm{T}} \\ \boldsymbol{c} & \boldsymbol{Q} \end{pmatrix}^{-1} = \begin{pmatrix} \Lambda_{f_*} & \boldsymbol{\Lambda}_{f_*,y}^{\mathrm{T}} \\ \boldsymbol{\Lambda}_{f_*,y} & \boldsymbol{\Lambda}_y \end{pmatrix} . \tag{B.14}$$

Comparing (B.14) to (B.12) we can work out the precision submatrices of interest. Firstly, considering the correspondence of precision Λ_{f_*} to submatrix $\boldsymbol{M}$ we get

$$\Lambda_{f_*} = (k_* - \boldsymbol{c}^{\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{c})^{-1} . \tag{B.15}$$

Secondly, considering the correspondence of the cross-precision vector $\boldsymbol{\Lambda}_{f_*,y}^{\mathrm{T}}$ to submatrix $-\boldsymbol{M}\boldsymbol{B}\boldsymbol{D}^{-1}$ we obtain

$$\boldsymbol{\Lambda}_{f_*,y}^{\mathrm{T}} = -(k_* - \boldsymbol{c}^{\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{c})^{-1}\boldsymbol{c}^{\mathrm{T}}\boldsymbol{Q}^{-1} \tag{B.16}$$

$$= -\frac{\boldsymbol{c}^{\mathrm{T}}}{k_*\boldsymbol{Q} - \boldsymbol{c}^{\mathrm{T}}\boldsymbol{c}} . \tag{B.17}$$

Plugging (B.15) into (B.7) we get the variance of the posterior distribution as in (B.5) expressed with the joint distribution covariance submatrices as

$$\sigma_{f_*|y}^2 = \frac{1}{\Lambda_{f_*}} = k_* - \boldsymbol{c}^{\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{c} . \tag{B.18}$$

Further, plugging (B.15) and (B.16) into (B.11) we get the mean of the posterior distribution as in (B.5) expressed with the joint distribution covariance submatrices as

$$\begin{aligned} \mu_{f_*|y} &= m_* - \Lambda_{f_*}^{-1}\boldsymbol{\Lambda}_{f_*,y}^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{m}) \\ &= m_* - (k_* - \boldsymbol{c}^{\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{c})(-(k_* - \boldsymbol{c}^{\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{c})^{-1}\boldsymbol{c}^{\mathrm{T}}\boldsymbol{Q}^{-1})(\boldsymbol{y} - \boldsymbol{m}) \\ &= m_* + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{Q}^{-1}(\boldsymbol{y} - \boldsymbol{m}) . \end{aligned} \tag{B.19}$$

C Alternative view of the MC evaluation using importance sampling

One alternative view of the application of Monte Carlo evaluation to approximate the predictive mean (2.15) is shown with *importance sampling* as suggested in [2]. Here, we keep the general problem setup of approximating expected value of $\phi(q)$ with respect to a general probability density $p(q)$ as in (2.23), i.e.,

$$\mathrm{E}^{p(q)}\{\phi(q)\} = \int \phi(q)p(q)\,dq \ . \tag{C.1}$$

The difference considered now is that we may not be able to sample directly from $p(q)$. Instead, we propose to sample from a different distribution, i.e., a *proposal distribution* $\rho(q)$ which is related to $p(q)$ according to

$$p(q) = w(q)\rho(q) \ , \tag{C.2}$$

where $w(q)$ is the *importance weight.* In order for the importance weight to be finite we must choose the proposal distribution $\rho(q)$ such that its support includes the support of $p(q)$. We can then express the approximate of distribution $p(q)$ using the point-mass function

$$p_{\mathrm{MCIS}}(q) = \frac{\sum_{i=1}^{s} w(q^{(i)})\delta(q - q^{(i)})}{\sum_{i=1}^{s} w(q^{(i)})} \ , \tag{C.3}$$

where $q^{(i)}, i = 1, \ldots, s$ is a set of random samples drawn from proposal distribution $\rho(q)$. Using $p_{\mathrm{MCIS}}(q)$ as an approximation of $p(q)$, we can express the approximate expectation of $\phi(q)$ as

$$\begin{aligned}
\mathrm{E}^{p_{\mathrm{MCIS}}(q)}\{\phi(q)\} &= \int_{\mathbb{R}} \phi(q)p_{\mathrm{MCIS}}(q)\,dq \\
&= \int_{\mathbb{R}} \phi(q)\frac{\sum_{i=1}^{s} w(q^{(i)})\delta(q - q^{(i)})}{\sum_{i=1}^{s} w(q^{(i)})}\,dq \\
&= \frac{\sum_{i=1}^{s} \int_{\mathbb{R}} \phi(q)w(q^{(i)})\delta(q - q^{(i)})\,dq}{\sum_{i=1}^{s} w(q^{(i)})} \\
&= \frac{\sum_{i=1}^{s} \phi(q^{(i)})w(q^{(i)})}{\sum_{i=1}^{s} w(q^{(i)})} \ .
\end{aligned} \tag{C.4}$$

Monte Carlo approximation of the posterior mean

We will now apply this approximation to the posterior mean (2.29) expressed as a ratio of expectations with respect to the distribution $p(\boldsymbol{y}|\boldsymbol{x}_{\mathrm{t}})p(\boldsymbol{x}_{\mathrm{t}}|\tilde{\boldsymbol{x}}_{\mathrm{t}}) = p(\boldsymbol{y}, \boldsymbol{x}_{\mathrm{t}}|\tilde{\boldsymbol{x}}_{\mathrm{t}})$, i.e.,

$$
\begin{aligned}
\mathrm{E}\{f_*|\boldsymbol{y},\tilde{\boldsymbol{x}}_\mathrm{t}\} &= \frac{\int_{\mathbb{R}^{DI}} \mu_*(\boldsymbol{x}_\mathrm{t}) p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t}) p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})\, d\boldsymbol{x}_\mathrm{t}}{\int_{\mathbb{R}^{DI}} p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t}) p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})\, d\boldsymbol{x}_\mathrm{t}} \\
&= \frac{\mathrm{E}^{p(\boldsymbol{y},\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})}\{\mu_*(\boldsymbol{x}_\mathrm{t})\}}{\mathrm{E}^{p(\boldsymbol{y},\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})}\{1\}} \\
&= \mathrm{E}^{p(\boldsymbol{y},\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})}\{\mu_*(\boldsymbol{x}_\mathrm{t})\}\ .
\end{aligned} \tag{C.5}
$$

We note that $\phi(q)$ from (C.1) corresponds to the term $\mu_*(\boldsymbol{x}_\mathrm{t})$ in (C.5) and the term $p(q)$ corresponds to $p(\boldsymbol{y},\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})$.

Noticing that we cannot sample directly from $p(q)$, we consider the factorization of this distribution according to (C.2). The proposal distribution $\rho(q)$ then corresponds to $p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})$ and the importance weight $w(q)$ corresponds to $p(\boldsymbol{y}|\boldsymbol{x}_\mathrm{t})$. Expressing now the posterior mean from (C.5) using the approximate point mass function (C.4) results in

$$
\mathrm{E}^{\mathrm{MCIS}}\{f_*|\boldsymbol{y},\tilde{\boldsymbol{x}}_\mathrm{t}\} = \mathrm{E}^{p_{\mathrm{MCIS}}(q)}\{\mu_*(\boldsymbol{x}_\mathrm{t})\} = \frac{\sum_{i=1}^{s} \mu_*(\boldsymbol{x}_{\mathrm{t},i}) p(\boldsymbol{y}|\boldsymbol{x}_{\mathrm{t},i})}{\sum_{i=1}^{s} p(\boldsymbol{y}|\boldsymbol{x}_{\mathrm{t},i})}\ . \tag{C.6}
$$

The set of samples $\boldsymbol{x}_{\mathrm{t},i}, i = 1,\ldots,s$ is drawn from $p(\boldsymbol{x}_\mathrm{t}|\tilde{\boldsymbol{x}}_\mathrm{t})$. (C.5) is the same result as in (2.31). Similarly we would proceed also for the posterior variance derivation using importance sampling, finally obtaining the expression (2.33).

D Content of the electronic attachment

The electronic attachment of this thesis is an archive with the following structure:

```
+---Generating_graphs
|       uniformPosPrior.py
|       uniform_pos_prior.pdf
|
\---GPR_loc_uncertainty
    |   GPR_Jadaliha_mean_prediction.pdf
    |   GPR_Jadaliha_positions.pdf
    |   GPR_Jadaliha_var_prediction.pdf
    |   GPR_mean_prediction_simplified.pdf
    |   GPR_mean_uncertain.pdf
    |   GPR_realization.pdf
    |   GPR_var_prediction_simplified.pdf
    |   GPR_var_uncertain.pdf
    |   GPR__loc_uncertainty.py
    |   resultsCurrent.tex
    |   util.py
```

The file `uniformPosPrior.py` produces the plot in Figure 2.1.
The file `GPR__loc_uncertainty.py` is producing all the following plots in this thesis. These are located as `.pdf` files in the folder `GPR_loc_uncertainty`. For that it utilizes the tools in `util.py`. The simulation results are then printed into the file `resultsCurrent.tex`, which is formatted such that it can then be an input to LaTeX.