Magnetic interpolation using Gaussian processes

by

Wolf Nederpel

June 23, 2020

Magnetic interpolation using Gaussian processes

Dutch Title:

Magnetische interpolatie met behulp van Gaussische processen

by

Wolf Nederpel

A thesis on behalf of the Delft Institute for Applied Mathematics
to obtain the degree of Bachelor of Science in Applied Mathematics
at the Delft University of Technology.

Student number:	4697146	
Project duration:	April 20, 2020 – June 29, 2020	
Thesis committee:	Prof. dr. ir. A.W. Heemink,	TU Delft, Contact person
	Dr. J. Söhl,	TU delft, Staff member
	Dr. ir. E.S.A.M. Lepelaars,	TU Delft, Supervisor
	Ir. A.R.P.J. Vijn,	TU Delft, Supervisor

TU Delft

Abstract

In this thesis a model for interpolation of magnetic fields is constructed using Gaussian processes. This model takes the curl- and divergence-free properties of magnetic fields into account. The Gaussian process regression, or kriging, is tested on both simulated and real data. It is also attempted to reconstruct the magnetization of objects based on measurement data.

Acknowledgements

I would like to thank Aad Vijn and Eugene Lepelaar for their support and flexibility throughout this project. They gave me freedom to explore while keeping me focused and were always available for questions. Even though physical meetings were impossible our meetings were always helpful and pleasant. They also corrected many drafts.
I would also like to thank Veronica and my housemates Jop and Eefje for proof reading this document.

Contents

CHAPTER 1

Introduction

The Earth's magnetic field and localisation have been connected ever since the compass was first invented. The Earth can be thought of as a large magnet, as the molten iron in its core generates a magnetic field around the earth. This magnetic field can then be used by a compass to find the magnetic north pole. This magnetic field can also influence other objects.

Just like how one can turn an iron object like a screwdriver into a temporary magnet by rubbing a magnet against it, the Earth's magnetic field magnetizes iron (or more generally, ferromagnetic) objects which move through it. This effect can be used for localisation purposes. Steel objects and buildings, as ferromagnetic objects, are magnetized and therefore have a magnetic field around them. A field, known as the distortion field, which is the difference between the actual magnetic field at a location and the Earth's magnetic field sits around steel objects and in buildings. This field can for example be used to localise objects or to find ones position in a building.

As an important object of study for localisation, the Dutch Organisation for Applied Scientific Research (TNO) is working together with Delft University of Technology (TUD) to study these distortion fields. This collaboration aims to develop methods to better understand and model how objects become magnetized and what the effect of this is. As a part of this large project, the question arose whether it was possible to obtain more information from and interpolate measurement data using Gaussian process regression, or kriging. This thesis serves then as an initial research into the capabilities of kriging and its applications concerning the field of magnetostatics.

To this end this research starts by describing the fields which we are trying to model and their physical properties. Then, Gaussian processes are introduced and the role of the covariance function is discussed. In Chapters 4 and 5 two kernels are derived which encapsulate the physical properties of the magnetic field and a method of finding the hyperparameters in these kernels is presented. After this several experiments and their results are shown. Both experiments with simulated and real data were executed. As a final test it was attempted to recreate the magnetization of the test object.

CHAPTER 2

Magnetostatics

2.1. Introduction to Magnetostatics

This section is mainly based on Jackson 2007. An important object of study in the field of magnetostatics is the magnetic point dipole. It is defined to be the limit of a magnetized object as its size approaches zero, while the magnetic moment $\mathbf{m}$ of the object remains constant. The magnetic moment of a dipole is defined to be the vector relating the torque τ $[N\ m]$ to the magnetic flux density $\mathbf{B}$ $[T]$ by

$$\tau = \mathbf{m} \times \mathbf{B}. \tag{2.1}$$

With these magnetic dipoles one can define the magnetization field of an object $\mathbf{M}$ $[A\ m^2]$ to be the volume average of all the microscopic magnetic moments in a material. This is the field that a ferromagnetic object obtains when moving through the Earth's magnetic field, resulting in a magnetic field of the object.

During the nineteenth century, an important breakthrough was accomplished when James Clerk Maxwell managed to capture the behaviour of magnetism, electricity and electromagnetic radiation in Maxwell 1865. The well known Maxwell equations are a number of coupled partial differential equations, which describe magnetic and electric fields and are given by

$$\nabla \cdot \mathbf{D} = \rho \tag{2.2}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{2.3}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{2.4}$$

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}. \tag{2.5}$$

Here $\mathbf{E}$ $[V\ m^{-1}]$ is the electric field, $\mathbf{D}$ $[C\ m^{-2}]$ is the electric displacement field, $\mathbf{B}$ $[T]$ is the magnetic flux density, $\mathbf{H}$ $[A\ m^{-1}]$ is the magnetic field and $\mathbf{J}$ $[A\ m^{-2}]$ is the current density. Our main focus lies on (2.3) and (2.5), which, in a time independent and current free situation, reduce down to

$$\nabla \cdot \mathbf{B} = 0 \tag{2.6}$$

$$\nabla \times \mathbf{H} = \mathbf{0}. \tag{2.7}$$

These **H**- and **B**-fields are the result of some magnetized object. The most important equations that will be exploited during this research are (2.6) and (2.7) together with the coupling

$$\mathbf{H} = \frac{1}{\mu_0}\mathbf{B} - \mathbf{M}, \tag{2.8}$$

where $\mathbf{H}$ and $\mathbf{B}$ are the respectively curl- and divergence-free vector fields which are equal up to multiplication by a constant μ_0 outside of magnetized material such as in air or water. This μ_0 is called the magnetic vacuum permeability and is given by CODATA 2018 to be $1.25663706212 \cdot 10^{-6}$. As these fields are similar in that sense, they are generally both referred to as the magnetic field.

2.2. Derivation of H-field of a dipole

In order to generate data for the simulated experiments that are discussed later in this paper, the **B**- and **H**-field of a magnetic point dipole, or dipole, are used. The derivation below is focused on finding the **H**- and **M**-field, but as $\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M})$, this also leads to a description of the **B**-field.

To find the **H**-field produced by a (magnetic ponit) dipole, we start by finding the **M**-field of a three-dimensional rod, which we then shrink to a point while keeping its magnetic moment constant. This is the definition of a magnetic point dipole. To find the **H**-field we first use the earlier stated equation

$$\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M}), \tag{2.9}$$

which is described in more detail in Chapter 2. We also know that $\mathbf{M} = 0$ outside of magnetic material (e.g. in air or water, where one can measure a magnetic field). Inside magnetic material **H** and **M** are related by

$$\nabla \cdot \mathbf{H} = -\nabla \cdot \mathbf{M}, \tag{2.10}$$

because $\nabla \cdot \mathbf{B} = 0$. Curl-free vector fields have the property that a line integral along a path P is only dependent on the endpoints, A and B, of P such that

$$\int_P \mathbf{H}(\mathbf{x}) d\mathbf{x} = \varphi(A) - \varphi(B). \tag{2.11}$$

Where $\varphi : \mathbb{R}^3 \to \mathbb{R}$ and so

$$\mathbf{H} = -\nabla\varphi. \tag{2.12}$$

Substituting this in (2.10) gives

$$\Delta\varphi = \nabla \cdot \mathbf{M}. \tag{2.13}$$

The solution to this equation is given by the convolution between the greens function for the Laplace operation Δ and $\nabla \cdot \mathbf{M}$

$$\begin{aligned} \varphi(\mathbf{r}) &= \iiint_{\mathbb{R}^3} \Phi(\mathbf{r} - \mathbf{r}')\nabla \cdot \mathbf{M}(\mathbf{r}')d\mathbf{r}' \\ &= \frac{1}{4\pi} \iiint_{\mathbb{R}^3} \frac{1}{\|\mathbf{r} - \mathbf{r}'\|} \nabla \cdot \mathbf{M}(\mathbf{r}')d\mathbf{r}', \end{aligned} \tag{2.14}$$

where **r** is the location of the measurement.

We will use the **M**-field to calculate φ from which **H** is determined, but first we must determine **M** of a dipole.

To this end, we start by finding the **M**-field of a uniformly magnetized rod in the x-direction with magnitude M_0 (implying that the magnetization is zero outside of the rod). To find the **M**-field of a dipole from this, the rod is shrunk down to a point while its magnetic dipole moment is kept constant. After this process magnetization is other directions is added. Using cylindrical coordinates (x, ρ, θ) the magnetization of the rod, Ω with length $2l$ and radius α, is given by

$$\mathbf{M} = \frac{M_0}{\mathrm{Vol}(\Omega)} \mathbf{u}_x \cdot \left[(U(x + l) - U(x - l)) \cdot U(\alpha - \rho) \right], \tag{2.15}$$

where U is a Heaviside function so that

$$U(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1, & \text{if } x \geq 0 \end{cases}, \tag{2.16}$$

and $\mathbf{u}_x$ is the unit vector in the x-direction, as the magnetization is zero outside of the rod. The magnetic dipole moment $\mathbf{m}$ is found by integrating $\mathbf{M}$ over the entire volume which gives

$$\mathbf{m} = \frac{M_0}{\text{Vol}(\Omega)} \mathbf{u}_x \cdot \text{Vol}(\Omega) = M_0 \mathbf{u}_x, \tag{2.17}$$

as the magnetization is constant over the rod. Note how the magnetic dipole moment is independent of the volume, so as we shrink the rod, the magnetization of the rod must increase as the magnetization is inversely proportional to the volume. If we first let the radius α go to zero we set (now in Cartesian coordinates)

$$\mathbf{M} = c\mathbf{u}_x \cdot [(U(x+l) - U(x-l))\delta(y)\delta(z)], \tag{2.18}$$

where c is yet to be determined and δ denotes the Dirac-delta function. Integrating over the entire volume and using the properties of the delta function we now find

$$\mathbf{m} = c\mathbf{u}_x \cdot 2l. \tag{2.19}$$

So the choice $c = \frac{M_0}{2l}$ gives

$$\mathbf{m} = M_0 \mathbf{u}_x \tag{2.20}$$

as required.

Now to find the potential, we first find that by filling in the value of c in $\mathbf{M}$ and using the rules for Heaviside functions that

$$\nabla \cdot \mathbf{M} = \frac{M_0}{2l} \mathbf{u}_x \cdot [(U(x+l) - U(x-l))\delta(y)\delta(z)] \tag{2.21}$$

$$= \frac{M_0}{2l} \cdot [(\delta(x+l) - \delta(x-l))\delta(y)\delta(z)], \tag{2.22}$$

as $\mathbf{M}$ only has non-zero entries in the x-component. Using equation (2.14), this gives a potential given by

$$\varphi(\mathbf{r}) = \frac{1}{4\pi} \iiint_{\mathbb{R}^3} \frac{1}{\|\mathbf{r} - \mathbf{r}'\|} \frac{M_0}{2l} \cdot \left[(\delta(x'+l) - \delta(x'-l))\delta(y')\delta(z')\right] d\mathbf{r}' \tag{2.23}$$

$$= \frac{M_0}{8l\pi} \iiint_{\mathbb{R}^3} \frac{(\delta(x'+l) - \delta(x'-l))\delta(y')\delta(z')}{\sqrt{(x-x')^2 + (y-y')^2 + (z-z')^2}} d\mathbf{r}' \tag{2.24}$$

$$= \frac{M_0}{8l\pi} \left[\frac{1}{\sqrt{(x+l)^2 + y^2 + z^2}} - \frac{1}{\sqrt{(x-l)^2 + y^2 + z^2}} \right]. \tag{2.25}$$

Now, to find the limit as $l \to 0$ we will use the definition of the central derivative,

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x-h)}{2h}. \tag{2.26}$$

We first rewrite the potential as

$$\varphi(\mathbf{r}) = \frac{M_0}{4\pi} \frac{\frac{1}{\sqrt{(x+l)^2+y^2+z^2}} - \frac{1}{\sqrt{(x-l)^2+y^2+z^2}}}{2l}, \tag{2.27}$$

now, using the definition with $f(x) = \frac{1}{\sqrt{x^2+y^2+z^2}}$. we find

$$\varphi(\mathbf{r}) = \frac{M_0}{4\pi} f'(x) \tag{2.28}$$

$$= \frac{M_0}{4\pi} \left(\frac{-1}{2} (x^2+y^2+z^2)^{\frac{-3}{2}} 2x \right) \tag{2.29}$$

$$= \frac{M_0}{4\pi} \cdot \frac{-x}{\|\mathbf{r}\|^3}. \tag{2.30}$$

We have now found the potential of a dipole with magnetic moment $\mathbf{m} = M_x \mathbf{u}_x$. To find the potential for a dipole with a more general moment $\mathbf{m} = M_x \mathbf{u}_x + M_y \mathbf{u}_y + M_z \mathbf{u}_z$, note how we used $f(\mathbf{r}) = \frac{1}{\|\mathbf{r}\|}$ and $\varphi(\mathbf{r}) = \frac{1}{4\pi} M_0 f_x$ to find our previous result. The more general result can be obtained as

$$\varphi(\mathbf{r}) = \frac{1}{4\pi} \left(M_x f_x + M_y f_y + M_z f_z \right) \tag{2.31}$$

$$= \frac{1}{4\pi} \left(\frac{-x M_x}{\|\mathbf{r}\|^3} + \frac{-y M_y}{\|\mathbf{r}\|^3} + \frac{-z M_z}{\|\mathbf{r}\|^3} \right) \tag{2.32}$$

$$= \frac{-\mathbf{m} \cdot \mathbf{r}}{4\pi \|\mathbf{r}\|^3}. \tag{2.33}$$

From this we can find the H-field corresponding to a dipole with a general magnetic moment to be

$$\mathbf{H}(\mathbf{r}) = \nabla \varphi(\mathbf{r}) \tag{2.34}$$

$$= -\nabla \frac{\mathbf{m} \cdot \mathbf{r}}{4\pi \|\mathbf{r}\|^3} \tag{2.35}$$

$$= \frac{1}{4\pi} \left[\frac{3(\mathbf{m} \cdot \mathbf{r})\mathbf{r}}{\|\mathbf{r}\|^5} - \frac{\mathbf{m}}{\|\mathbf{r}\|^3} \right]. \tag{2.36}$$

Which is motivated by

$$-\frac{\partial}{\partial x} \frac{\mathbf{m} \cdot \mathbf{r}}{4\pi \|\mathbf{r}\|^3} = \frac{-1}{4\pi} \frac{\partial}{\partial x} \frac{M_x \cdot x + M_y \cdot y + M_z \cdot z}{(x^2+y^2+z^2)^{\frac{3}{2}}} \tag{2.37}$$

$$= \frac{-1}{4\pi} \left(\frac{-3}{2} (x^2+y^2+z^2)^{-\frac{5}{2}} 2x \left(M_x \cdot x + M_y \cdot y + M_z \cdot z \right) \right.$$

$$\left. + (x^2+y^2+z^2)^{\frac{-3}{2}} \cdot M_x \right) \tag{2.38}$$

$$= \frac{1}{4\pi} \left(\frac{3(\mathbf{m} \cdot \mathbf{r})x}{\|r\|^5} - \frac{M_x}{\|r\|^3} \right), \tag{2.39}$$

as similar results can be obtained for a derivation with respect to y or z. The field equation (2.36) is used in Chapter 6 to execute simulated experiments.

CHAPTER 3

Gaussian Processes

The well known Gaussian (normal) distribution has no shortage of relevance. Many observations one makes can be accurately modelled as being sampled from a Gaussian distribution. A stochastic variable X is distributed according to a Gaussian distribution when its probability density function (pdf) is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\left(\frac{x-\mu}{\sigma}\right)^2\right). \tag{3.1}$$

It turns out that, for example, the height of males of a certain age are distributed according to a Gaussian distribution with some mean μ and variance σ. The fact that many such things can be modeled as being the result of a Gaussian distribution is due to the central limit theorem. When not only the height of males of a certain age, but also the length of their feet is modelled, we might realise that there is a certain dependency between the two. People who are quite tall might generally also have large feet. In other words, there is a covariance between height and foot length.

The resulting distribution of height ($\mathbf{X}_1$) and foot length ($\mathbf{X}_2$) is a multivariate Gaussian distribution, who's pdf is given by

$$f_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{-\frac{k}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right), \tag{3.2}$$

where k is the number of dimensions of our stochastic variable $\mathbf{X}$, $\boldsymbol{\mu}$ is the vector containing the means of each component of $\mathbf{X}$, $\Sigma_{i,j}$ is the covariance between $\mathbf{X}_i$ and $\mathbf{X}_j$ ($i, j \in 0, ..., k$), and $|\Sigma|$ denotes the determinant of Σ. As a Gaussian distribution is completely determined by a mean $\boldsymbol{\mu}$ and a covariance matrix Σ, we write $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$. To understand the properties of the covariance matrix, we need the following definition:

DEFINITION 3.1. *A symmetric real $n \times n$ matrix A is called positive semi-definite when*

$$\mathbf{u}^\top A\mathbf{u} \geq 0 \qquad \forall \mathbf{u} \in \mathbb{R}^n \backslash 0. \tag{3.3}$$

A is called positive definite when

$$\mathbf{u}^\top A\mathbf{u} > 0 \qquad \forall \mathbf{u} \in \mathbb{R}^n \backslash 0. \tag{3.4}$$

The covariance matrix is a symmetric matrix, as the covariance between height and foot length should be the same as the covariance between foot length and height, and a positive semi-definite matrix (Theorem 3.1). For (3.2) to exist, Σ must be positive definite, as the determinant of symmetric positive semi-definite matrix can be 0 while that of a symmetric positive definite matrix is guaranteed to be positive as required since $f_{\mathbf{X}}$ is function to $\mathbb{R}$.

THEOREM 3.1. *Every covariance matrix Σ of a zero-mean random vector, given by $\Sigma_{i,j} = E(X_i X_j)$, or equivalently $\Sigma = E(\mathbf{X}\mathbf{X}^\top)$, is positive semi-definite. If none of the elements of $\mathbf{X}$ is a linear combination of the others, Σ is positive definite.*

PROOF. Let $\mathbf{u}$ be a real vector and $\mathbf{X}$ be a stochastic vector both of length n. Then

$$\begin{aligned}
\mathbf{u}^\top \Sigma \mathbf{u} &= \mathbf{u}^\top E(\mathbf{X}\mathbf{X}^\top)\mathbf{u} && (3.5)\\
&= E(\mathbf{u}^\top \mathbf{X}\mathbf{X}^\top \mathbf{u}) && (3.6)\\
&= E(\mathbf{u}^\top \mathbf{X}(\mathbf{u}^\top \mathbf{X})^\top) && (3.7)\\
&= E(\mathbf{u}^\top \mathbf{X}\mathbf{u}^\top \mathbf{X})) && (3.8)\\
&= E((\mathbf{u}^\top \mathbf{X})^2) \geq 0 && (3.9)\\
& && (3.10)
\end{aligned}$$

As $\mathbf{u}^\top \mathbf{X}$ is of size 1×1. Since

$$E((\mathbf{u}^\top \mathbf{X})^2) = 0 \iff \mathbf{u}^\top \mathbf{X} = 0 \tag{3.11}$$

the statement holds. □

An important notion that is used in this research is that of a conditional distribution. Say we know the multivariate Gaussian distribution of height and foot length, then what is the height distribution of people with feet of, say, 26 cm? The pdf of this distribution is, by definition, given by

$$f_{\mathbf{X}_1|\mathbf{X}_2}(\mathbf{x}_1|\mathbf{x}_2 = 26) = \frac{f_{\mathbf{X}}([\mathbf{x}_1\ 26]^\top)}{f_{\mathbf{X}_2}(26)}. \tag{3.12}$$

In this formulation 26 can be replaced by any reasonable foot length. The notation $\mathbf{X}_1|\mathbf{X}_2$ references the fact that $\mathbf{X}_2$ is known while $\mathbf{X}_1$ is not. The mean and covariance of the conditional distribution for are presented and derived in Lemma 3.1. This conditional distribution, which gives the distribution of a stochastic variable when other stochastic variables are known, will allow us to use Gaussian distributions for regression.

LEMMA 3.1. *Suppose that a random vector $\mathbf{X} = (\mathbf{X}_1, ..., \mathbf{X}_k)^\top$ is distributed as a zero-mean multivariate Gaussian*

$$\mathbf{X} \sim N(\mathbf{0}, \Sigma). \tag{3.13}$$

If we partition $\mathbf{X} = (\mathbf{X}^{(1)}\ \mathbf{X}^{(2)})$ and

$$\Sigma = \begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{bmatrix} \tag{3.14}$$

accordingly, then the conditional distribution of $\mathbf{X}^{(1)}$ given $\mathbf{X}^{(2)}$ is given by

$$\mathbf{X}^{(1)}|\mathbf{X}^{(2)} \sim N(\Sigma_{2,1}\Sigma_{1,1}^{-1}\mathbf{x}^{(1)},\ \Sigma_{2,2} - \Sigma_{2,1}\Sigma_{1,1}^{-1}\Sigma_{1,2}). \tag{3.15}$$

PROOF. Writing

$$\Sigma^{-1} = V = \begin{bmatrix} V_{1,1} & V_{1,2} \\ V_{2,1} & V_{2,2} \end{bmatrix} \tag{3.16}$$

gives

$$\begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{bmatrix} \cdot \begin{bmatrix} V_{1,1} & V_{1,2} \\ V_{2,1} & V_{2,2} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \tag{3.17}$$

and specifically the relations

$$\Sigma_{1,1}V_{1,1} + \Sigma_{1,2}V_{2,1} = I \tag{3.18}$$
$$\Sigma_{1,1}V_{1,2} + \Sigma_{1,2}V_{2,2} = 0 \tag{3.19}$$
$$\Sigma_{2,1}V_{1,1} + \Sigma_{2,2}V_{2,1} = 0 \tag{3.20}$$
$$\Sigma_{2,1}V_{1,2} + \Sigma_{2,2}V_{2,2} = I. \tag{3.21}$$

The exponent of the multivariate normal (3.2) can now be reformulated as

$$\mathbf{x}^\top \Sigma^{-1}\mathbf{x} = \mathbf{x}^{(1)\top}V_{1,1}\mathbf{x}^{(1)} + \mathbf{x}^{(1)\top}V_{1,2}\mathbf{x}^{(2)} + \mathbf{x}^{(2)\top}V_{2,1}\mathbf{x}^{(1)} + \mathbf{x}^{(2)\top}V_{2,2}\mathbf{x}^{(2)}. \tag{3.22}$$

In order to find the conditional distribution, we must complete the square in $\mathbf{x}^{(2)}$ in this expression. Writing

$$\mathbf{x}^\top \Sigma^{-1}\mathbf{x} = (\mathbf{x}^{(2)} - \mathbf{m})^\top M(\mathbf{x}^{(2)} - \mathbf{m}) + \mathbf{c} \tag{3.23}$$
$$= \mathbf{x}^{(2)\top}M\mathbf{x}^{(2)} - \mathbf{x}^{(2)\top}M\mathbf{m} - \mathbf{m}^\top M\mathbf{x}^{(2)} + \mathbf{m}^\top M\mathbf{m} + \mathbf{c}, \tag{3.24}$$

and comparing with (3.22) we find

$$\mathbf{x}^{(2)\top}M\mathbf{x}^{(2)} = \mathbf{x}^{(2)\top}V_{2,2}\mathbf{x}^{(2)} \implies M = V_{2,2} \tag{3.25}$$
$$-\mathbf{x}^{(2)\top}V_{2,2}\mathbf{m} = \mathbf{x}^{(2)\top}V_{2,1}\mathbf{x}^{(1)} \implies \mathbf{m} = -V_{2,2}^{-1}V_{2,1}\mathbf{x}^{(1)} \tag{3.26}$$
$$(V_{2,2}^{-1}V_{2,1}\mathbf{x}^{(1)})^\top V_{2,2}V_{2,2}^{-1}V_{2,1}\mathbf{x}^{(1)} + \mathbf{c} = \mathbf{x}^{(1)\top}V_{2,1}^\top V_{2,2}^{-1}V_{2,1}\mathbf{x}^{(1)} + \mathbf{c} \tag{3.27}$$
$$= \mathbf{x}^{(1)\top}V_{1,1}\mathbf{x}^{(1)} \implies \mathbf{c} = \mathbf{x}^{(1)\top}(V_{1,1} - V_{2,1}^\top V_{2,2}^{-1}V_{2,1})\mathbf{x}^{(1)}, \tag{3.28}$$

after which (3.23) gives

$$\mathbf{x}^\top \Sigma^{-1}\mathbf{x} = (\mathbf{x}^{(2)} + V_{2,2}^{-1}V_{2,1}\mathbf{x}^{(1)})^\top V_{2,2}(\mathbf{x}^{(2)} + V_{2,2}^{-1}V_{2,1}\mathbf{x}^{(1)}) + \mathbf{x}^{(1)\top}(V_{1,1} - V_{2,1}^\top V_{2,2}^{-1}V_{2,1})\mathbf{x}^{(1)}. \tag{3.29}$$

This exponent is the sum of two parts, one which, when compared with the exponent in (3.2), can be considered a function of $\mathbf{x}^{(2)}$ once $\mathbf{x}^{(1)}$ is given, and one which is a function of $\mathbf{x}^{(1)}$ alone. Also notice how both functions could be a valid exponent of a multivariate Gaussian distribution. As by (3.12) we have that

$$f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}^{(1)}|\mathbf{X}^{(2)}}(\mathbf{x}^{(1)}|\mathbf{x}^{(2)})f_{\mathbf{X}^{(2)}}(\mathbf{x}^{(2)}) \tag{3.30}$$

where

$$f_{\mathbf{X}^{(2)}}(\mathbf{x}^{(2)}) \propto \exp\left(\mathbf{x}^{(1)\top}(V_{1,1} - V_{2,1}^\top V_{2,2}^{-1}V_{2,1})\mathbf{x}^{(1)}\right) \tag{3.31}$$

and

$$f_{\mathbf{X}^{(1)}|\mathbf{X}^{(2)}}(\mathbf{x}^{(1)}|\mathbf{x}^{(2)}) \propto \exp\left((\mathbf{x}^{(2)} + V_{2,2}^{-1}V_{2,1}\mathbf{x}^{(1)})^\top V_{2,2}(\mathbf{x}^{(2)} + V_{2,2}^{-1}V_{2,1}\mathbf{x}^{(1)})\right) \tag{3.32}$$

giving that

$$\mathbf{x}^{(1)}|\mathbf{x}^{(2)} \sim N(-V_{2,2}^{-1}V_{2,1}\mathbf{x}^{(1)}, V_{2,2}^{-1}). \tag{3.33}$$

This can be translated back in to term of Σ using (3.18 - 3.21). First, from (3.19) $V_{1,2} = -\Sigma_{1,1}\Sigma_{1,2}V_{2,2}$ after which we can substitute $V_{1,2}$ in (3.21) to find

$$-\Sigma_{1,1}\Sigma_{1,2}V_{2,2}V_{2,1} + V_{2,2} + \Sigma_{2,2}V_{2,2} = I \tag{3.34}$$

which implies

$$V_{2,2}^{-1} = \Sigma_{2,2} - \Sigma_{2,1}\Sigma_{1,1}^{-1}\Sigma_{1,2}, \tag{3.35}$$

giving an expression for the conditional covariance matrix. Now for the mean we find, by the transpose of (3.19) and the symmetry of the covariance matrix, that

$$V_{2,1}\Sigma_{1,1} + V_{2,2}\Sigma_{2,1} = 0. \tag{3.36}$$

Left multiplication with $V_{2,2}^{-1}$ and right multiplication with $Sigma_{1,1}^{-1}$ now yields

$$V_{2,2}^{-1}V_{2,1} + \Sigma_{2,1}\Sigma_{1,1}^{-1} = 0 \tag{3.37}$$

$$V_{2,2}^{-1}V_{2,1} = -\Sigma_{2,1}\Sigma_{1,1}^{-1}. \tag{3.38}$$

Substituting (3.35) and (3.38) into (3.33) gives the final result of

$$\mathbf{x}^{(1)}|\mathbf{x}^{(2)} \sim N(\Sigma_{2,1}\Sigma_{1,1}^{-1}\mathbf{x}^{(1)}, \Sigma_{2,2} - \Sigma_{2,1}\Sigma_{1,1}^{-1}\Sigma_{1,2}). \tag{3.39}$$

□

3.1. Gaussian Process regressions

This section is mainly based on Rasmussen and Williams 2005. The intuitive idea behind kriging, or Gaussian process regression, is that a function f can be thought of as a very long (infinite) vector. In implementations functions are often approximated by vectors, where there is a vector of inputs and a corresponding vector of outputs. As there are more elements added to this vector, the vector will approach the true function. To link this idea to a Gaussian process, which is the idea of a multivariate Gaussian distribution but for an infinite dimensional stochastic vector, the assumption is made that the function values are distributed according to a Gaussian process.

DEFINITION 3.2. *A Gaussian process, or GP, is a collection of random variables, any finite number of which have a joint Gaussian distribution.*

A Gaussian process is determined by its mean $\mu(\mathbf{x})$ and covariance function $k(\mathbf{x}, \mathbf{x}')$, just like a Gaussian distribution. So when $f(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ is distributed according to a Gaussian process we write

$$f(\mathbf{x}) \sim GP(\mu(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')). \tag{3.40}$$

This means that the function value at some point $\mathbf{x}$ is given by some distribution with a mean of $\mu(\mathbf{x})$ and a covariance with other inputs $\mathbf{x}'$ given by $k(\mathbf{x}, \mathbf{x}')$. This covariance function has analogous properties to the covariance matrix in (3.2).

The mean μ is often taken to be zero for simplicity, this practice is followed in this work. The covariance function determines the 'type' of function a sample of such a distribution is. To see the importance of the covariance function, note how it can for example encode smoothness by translating the deterministic idea of continuity, $|f(\mathbf{x}) - f(\mathbf{x}')|$ is small when $|\mathbf{x} - \mathbf{x}'|$ is small, to a stochastic idea, $f(\mathbf{x})$ and $f(\mathbf{x}')$ are highly correlated when $|\mathbf{x} - \mathbf{x}'|$ is small. The covariance function can also encode the assumption that a function is periodic by making points a certain distance apart highly correlated again. By determining how and when function values are correlated, the covariance function can encode many such 'types' of functions. In Chapter 4 of Rasmussen and Williams 2005 several covariance functions and how to make new ones of old are discussed. This is how we can use the covariance function to encode a priori knowledge about a function. Once that is established, the conditional distribution introduced in (3.12) can be used to find the distribution of functions of the correct 'type' once the function value at certain points is known. Covariance functions are discussed

in more detail in Section 3.2.

So this covariance function can be used to encode a priori knowledge about the type of function at hand. In order to use Gaussian processes for regression purposes, a posterior distribution of functions as a result of some (noise free) measurement data $\{(\mathbf{x}_i,\ f(\mathbf{x}_i) \mid i = 1, \dots, n\}$ is required. where f is the function for which currently unknown output values at points $\{\mathbf{x}_{i*} | i = 1, ..., m\}$ (known as input or test points) are to be predicted and where n and m are respectively the number of data and input points.

We introduce the following notation

$$[K(X,X)]_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j) \qquad \forall\ i,j \in 0, ..., n, \tag{3.41}$$

$$[K(X_*,X)]_{i,j} = k(\mathbf{x}_{i*}, \mathbf{x}_j) \qquad \forall\ i \in 0, ..., m\ \&\ \forall\ j \in 0, ..., n, \tag{3.42}$$

$$[K(X,X_*)]_{i,j} = k(\mathbf{x}_i, \mathbf{x}_{j*}) \qquad \forall\ i \in 0, ..., n\ \&\ \forall\ j \in 0, ..., m, \tag{3.43}$$

$$[K(X_*,X_*)]_{i,j} = k(\mathbf{x}_{i*}, \mathbf{x}_{j*}) \qquad \forall\ i \in 0, ..., m\ \&\ \forall\ j \in 0, ..., m, \tag{3.44}$$

$$\mathbf{f} = [f(\mathbf{x}_1), ..., f(\mathbf{x}_n)]^\top, \tag{3.45}$$

$$\mathbf{f}_* = [f(\mathbf{x}_{1*}), ..., f(\mathbf{x}_{m*})]^\top. \tag{3.46}$$

With this notation, X is the matrix who's rows are given by the measurement locations and X_* is the matrix who's rows are given by the input points. Now, by the assumption that f is a realization of a Gaussian process we find

$$\mathbf{f} \ \sim N(\mathbf{0}, K(X,X)), \tag{3.47}$$

$$\mathbf{f}_* \ \sim N(\mathbf{0}, K(X_*,X_*)). \tag{3.48}$$

Combined with the matrices $K(X,X_*)$ and $K(X_*,X)$ which give the covariance between $\mathbf{f}$ and $\mathbf{f}_*$, yields

$$\begin{bmatrix}\mathbf{f}\\ \mathbf{f}_*\end{bmatrix} \ \sim\ N\left(\mathbf{0}, \begin{bmatrix}K(X,X) & K(X,X_*)\\ K(X_*,X) & K(X_*,X_*)\end{bmatrix}\right). \tag{3.49}$$

From this distribution, the conditional distribution of $\mathbf{f}_*$ given the test inputs X_*, training inputs X and the known function values $\mathbf{f}$ as derived in Lemma 3.1 can be found to be

$$\mathbf{f}_* | X_*, X, \mathbf{f} \ \sim\ N\left(K(X_*,X)K(X,X)^{-1}\mathbf{f},\ K(X_*,X_*) - K(X_*,X)K(X,X)^{-1}K(X,X_*)\right). \tag{3.50}$$

From this posterior distribution, one can use the mean as the predicted value of $f(\mathbf{x}_*)$, and the variance as a measure of how certain the model is about the predicted value.

When we consider a model where the measurement data comes from some noisy observation $y_i = f(\mathbf{x})_i + \epsilon_i$, where $\epsilon_i \ \sim\ N(0, \sigma^2_{noise})$. We find that

$$\mathrm{Cov}(y_i, y_j) = \mathrm{Cov}(f(\mathbf{x})_i + \epsilon_i, f(\mathbf{x})_j + \epsilon_j) = \mathrm{Cov}(f(\mathbf{x})_i, f(\mathbf{x})_j) + \mathrm{Cov}(\epsilon_i, \epsilon_j) \tag{3.51}$$

$$= k(\mathbf{x_i}, \mathbf{x_j}) + \mathrm{Var}(\epsilon_i)\delta_{ij} = k(\mathbf{x_i}, \mathbf{x_j}) + \sigma^2_{noise}\delta_{ij}. \tag{3.52}$$

Writing $K = K(X,X)$, $K_* = K(X,X_*) = K(X_*,X)^\top$ and $\mathbf{y} = [y_1, ..., y_n]^\top$ we now find

$$\mathbf{f}_* | X_*, X, \mathbf{y} \ \sim\ N(K_*^\top (K + \sigma^2_{noise} I)^{-1}\mathbf{y},\ K(X_*,X_*) - K_*^\top (K + \sigma^2_{noise} I)^{-1} K_*). \tag{3.53}$$

Now if one wishes to know the function value at some point $\mathbf{x}_*$ a reasonable predictor under the assumptions is the most likely value to come out of this distribution. For a Gaussian

distribution this is the mean. The mean is thus the value used in the experiments presented in this paper as the predicted value which gives

$$\mathbf{f}_* = K_*^\top (K + \sigma_{noise} I)^{-1} \mathbf{y}. \tag{3.54}$$

3.2. Covariance Functions and reproducing kernel Hilbert spaces

Not all functions can be used as covariance functions. First of all, the matrix K as described above (known as the Gram matrix or covariance matrix) must be symmetric as covariance is symmetric in its inputs. Secondly, the covariance matrix must be positive semi-definite. Functions which give rise to such matrices are positive semi-definite kernels corresponding to reproducing kernel Hilbert spaces (RKHS) of which the function these kernels produce are a member. First we will state the definition of a RKHS and the important Moore-Aronszajn theorem.

DEFINITION 3.3. *(Reproducing kernel Hilbert space). Let $\mathcal{H}$ be a Hilbert space of real function f defined on an index set (a set which labels elements of another set) $\mathcal{X}$. Then $\mathcal{H}$ is called a reproducing kernel Hilbert space endowed with an inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ (and norm $\|f\|_{\mathcal{H}} = \sqrt{\langle f, f \rangle_{\mathcal{H}}}$) if there exists a function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ with the following properties:*

(1) for every $\mathbf{x}$, $k(\mathbf{x}, \mathbf{x}')$ as a function of $\mathbf{x}'$ belongs to $\mathcal{H}$, and
(2) k has the reproducing property $\langle f(\cdot), k(\cdot, \mathbf{x}) \rangle_{\mathcal{H}} = f(\mathbf{x})$.

THEOREM 3.2. *(Moore-Aronszajn theorem, Aronszajn [1950]). Let $\mathcal{X}$ be an index set. Then for every positive definite function $k(\cdot, \cdot)$ on $\mathcal{X} \times \mathcal{X}$ there exists a unique RKHS, and vice versa.*

An example of a Hilbert space and a kernel is L_2 with $k(\mathbf{x}, \mathbf{x}') = \delta(\mathbf{x} - \mathbf{x}')$. The kernel is what we call the representer of evaluation as $\langle f(\cdot), k(\cdot, \mathbf{x}) \rangle_{L_2} = \int f(\mathbf{x}') k(\mathbf{x}', \mathbf{x}) d\mathbf{x}' = f(\mathbf{x})$. L_2 is not a RKHS as property 1. of the definition is not satisfied.

A commonly used kernel in Gaussian process regression, which encodes the assumption that the prior functions are smooth, is the squared exponential (SE) kernel, given by

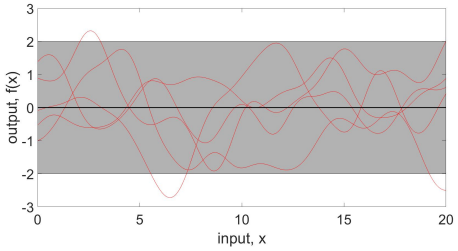
$$k_{SE}(\mathbf{x}, \mathbf{x}') = \sigma_{mag}^2 \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2l^2} \right). \tag{3.55}$$

Intuitively, if for evaluation of a L_2 function in a point $\mathbf{x}$ only information about that very point is needed, it makes sense that for evaluation of a smooth function also information is needed about points close to $\mathbf{x}$ to guarantee smoothness. The kernel is indefinitely differentiable, which means that a GP with this kernel has mean square derivatives of all orders (Section 3.3), making the resulting functions very smooth.

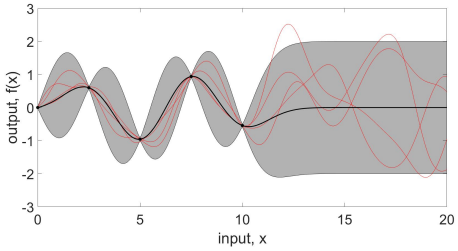
Another example of a covariance function is the periodic covariance function (with a period of 2π, for one dimensional input) given by

$$k_{periodic}(x, x') = \sigma_{mag}^2 \exp\left(-\frac{\sin\left(\frac{x - x'}{2}\right)}{l^2} \right). \tag{3.56}$$

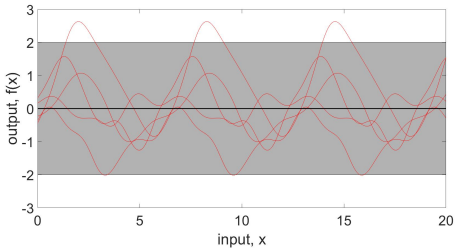
The magnitude-scale σ_{mag} and length-scale l are called hyperparameters and influence the signal variance and the number of fluctuations respectively. The hyperparameters will be named according to their role. Although several kernels have hyperparameters with the same role, it should be understood that their value might differ between kernels even though they



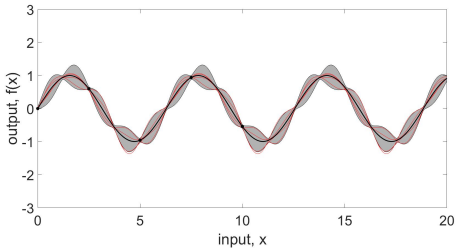
(A) a priori distribution, SE kernel.



(B) a postriori distibution, SE kernel.



(C) a postriori distibution, periodic kernel.



(D) a postriori distibution, periodic kernel.

FIGURE 3.1. Example regression, data is shown by the black dots, functions drawn from the distribution are shown in red, the mean of which is shown in black, and the shaded region shows twice the standard deviation. Hyper-parameters were set as $\sigma_{mag} = 1$ and $l = 1.2$. The data is sampled from a standard sine function to which some slight noise was added ($\sigma_{noise} = 0.05$).

are named the same.

An illustration of Gaussian process regression, which shows the importance of choosing the right covariance function corresponding to the a priori knowledge of the function one is modeling, is shown in Figure 3.1. As the data originates from an ordinary sine function with a period of 2π, the periodic kernel yield a much better approximation then the SE kernel. Note how the periodic kernel is capable of using a priori information about the function to make predictions far away from data points.

From (3.53) it can roughly be seen that the resulting prediction,

$$f(\mathbf{x}_*) = K_*^\top (K + \sigma^2 I)^{-1} \mathbf{y}, \tag{3.57}$$

can be written as a linear combination of kernel evaluations at $(\mathbf{x}_*, \mathbf{x}_i)$, or,

$$f = \sum_{i=1}^{n} k(\cdot, \mathbf{x}_i)\alpha_i. \tag{3.58}$$

This formulation is made more clear when the same estimator is derived from a regularization point of view using the representer theorem. This is discussed more in depth and connected to vector-valued functions in Alvarez et al. n.d. and Macêdo and Castro 2008. From this formulation it can be seen that the resulting prediction is in fact a member of the RKHS of which the kernel is a part. This confirms that Gaussian process regression takes place in a RKHS corresponding to the kernel which is chosen beforehand based on a priori knowledge of the function.

In this thesis, the focus is on finding values of a vector field. As discussed in Chapter 2, this vector field can be seen as either curl- or divergence-free. To this end, two kernels are required which encode the a priori knowledge that a vector field is curl- or divergence-free. This will yield much more accurate results than assuming the vector field components to simply be smooth functions (Solin et al. 2015), as with the curl- and divergence-free kernels it can be assured that the resulting vector field adhere to equations (2.7) and (2.6). In the same work, an overview of the applications of these kernels is given. The kernels are derived in Chapter 4.

3.3. linear operations on Gaussian processes

An important notion we will exploit in Chapter 4 is that of the derivative of a function which is distributed according to a Gaussian process. Just as a Gaussian distribution is closed under linear transformation, a Gaussian process is closed under linear operations. As by Agrell 2019, if we have a linear operator $L_{\mathbf{x}}[f]$ (e.g. differentiation, integration) acting on $\mathbf{x}$ and f is distributed according to a Gaussian process with mean m and kernel k, then

$$L_{\mathbf{x}}[f] \sim GP(L_{\mathbf{x}}[m],\ L_{\mathbf{x}'}[L_{\mathbf{x}}[k]]), \tag{3.59}$$

if $L_{\mathbf{x}}[f]$ exists. If we take this linear operator to be partial differentiation in some direction, this gives

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}_i} \sim GP\left(\frac{m(\mathbf{x})}{\partial \mathbf{x}_i},\ \frac{\partial^2}{\partial \mathbf{x}_i \partial \mathbf{x}_i'} k(\mathbf{x}, \mathbf{x}')\right). \tag{3.60}$$

Hence, as the SE kernel (3.55) is infinitely differentiable, it will yield functions which are infinitely differentiable. The fact that the derivative of a function distributed as GP is still distributed as a GP is integral to the derivations in Chapter 4.

CHAPTER 4

Derivation of curl- and divergence-free kernels

As discussed in Chapter 2, a magnetic field can be seen either as a curl- or divergence-free vector field. In order to make accurate predictions with kriging we need to incorporate this knowledge into a kernel. Here, two kernels are derived which take these properties into account.

To model vector valued functions $\mathbf{f}(\mathbf{x}) = [\mathbf{f}_1(\mathbf{x}), \dots, \mathbf{f}_d(\mathbf{x})]^\top$, one can model each component separately as a one dimensional GP

$$\mathbf{f}_i(\mathbf{x}) \sim GP(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \qquad \forall i \in 1, \dots, d. \tag{4.1}$$

However, to incorporate dependencies between the components, we consider a matrix-valued kernel

$$\mathbf{f}(\mathbf{x}) \sim GP(\boldsymbol{m}(\mathbf{x}), K(\mathbf{x}, \mathbf{x}')), \tag{4.2}$$

where the components of $\mathbf{m}$ and $K(\mathbf{x}, \mathbf{x}')$ are given by

$$\boldsymbol{m}_i(\mathbf{x}) = \mathbb{E}\left[\mathbf{f}_i(\mathbf{x})\right], \tag{4.3}$$

$$[K(\mathbf{x}, \mathbf{x}')]_{i,j} = \mathbb{E}\left[\left(\mathbf{f}_i(\mathbf{x}) - \boldsymbol{m}_i(\mathbf{x})\right)\left(\mathbf{f}_j(\mathbf{x}') - \boldsymbol{m}_j(\mathbf{x}')\right)\right] \qquad \forall i, j \in 1, \dots, d. \tag{4.4}$$

More in depth and theoretical research into matrix valued kernels and proof that the derived kernels below truly yield curl- and divergence-free kernels is found in Narcowich and Ward 1994 and Jr. 2005. Proving that the presented kernels are curl- and divergence-free comes down to checking whether the columns of the kernel are as such. This is due to (3.58) which extends to matrix valued kernels as

$$\mathbf{f}(\mathbf{x}) = \sum_{i=1}^{n} K(\mathbf{x}, \mathbf{x}')\boldsymbol{\alpha}_i. \tag{4.5}$$

An overview of matrix-valued kernels can be found in Alvarez et al. n.d.

4.1. Curl-Free Kernel Derivation

In this derivation we follow Solin et al. 2015. Our goal here is to derive a kernel which encodes the prior assumption of a vector field being curl-free.
As the **H**-field is a curl-free vector field, we know by (2.12) that it can be written as

$$\mathbf{H} = -\nabla\varphi, \tag{4.6}$$

$$(\mathbf{H}_{x_1}, \mathbf{H}_{x_2}, \mathbf{H}_{x_3}) = -\left(\frac{\partial\varphi}{\partial x_1}, \frac{\partial\varphi}{\partial x_2}, \frac{\partial\varphi}{\partial x_3}\right). \tag{4.7}$$

Under the assumption that φ is a (zero-mean) smooth, function we can model it as a Gaussian process with a squared exponential kernel (3.55)

$$\varphi(\mathbf{x}) \sim GP(0, k_{SE}(\mathbf{x}, \mathbf{x}')). \tag{4.8}$$

As by Section 3.3, $\mathbf{H}(\mathbf{x})$ is now also distributed as a GP

$$\mathbf{H}(\mathbf{x}) \sim GP(0, K_{\mathbf{H}}(\mathbf{x}, \mathbf{x}')). \tag{4.9}$$

Where $K_{\mathbf{H}}$ is a matrix valued kernel and $[K_{\mathbf{H}}(\mathbf{x},\mathbf{x}')]_{i,j}$ is given by

$$[K_{\mathbf{H}}(\mathbf{x},\mathbf{x}')]_{i,j} = \mathrm{Cov}(\mathbf{H}_i(\mathbf{x}),\mathbf{H}_j(\mathbf{x}')) \tag{4.10}$$

$$= \mathrm{Cov}(\frac{\partial\varphi(\mathbf{x})}{\partial x_i},\frac{\partial\varphi(\mathbf{x}')}{\partial x'_j}) \tag{4.11}$$

$$= \frac{\partial^2}{\partial x_i \partial x'_j}\mathrm{Cov}(\varphi(\mathbf{x}),\varphi(\mathbf{x'})) \tag{4.12}$$

$$= \frac{\partial^2}{\partial x_i \partial x'_j}k(\mathbf{x},\mathbf{x}') \tag{4.13}$$

$$= \frac{\partial^2}{\partial x_i \partial x'_j}\left[\sigma^2_{mag}\exp\left(-\frac{(x_1-x'_1)^2+(x_2-x'_2)^2+(x_3-x'_3)^2}{2l^2}\right)\right] \tag{4.14}$$

$$= \sigma^2_{mag}\frac{\partial}{\partial x_i}\left[\frac{(x_j-x'_j)}{l^2}\exp\left(-\frac{(x_1-x'_1)^2+(x_2-x'_2)^2+(x_3-x'_3)^2}{2l^2}\right)\right] \tag{4.15}$$

$$= \sigma^2_{mag}\left[\frac{\delta_{i,j}}{l^2}-\frac{(x_i-x'_i)(x_j-x'_j)}{l^4}\right]\exp\left(-\frac{(x_1-x'_1)^2+(x_2-x'_2)^2+(x_3-x'_3)^2}{2l^2}\right) \tag{4.16}$$

$$K_{\mathbf{H}}(\mathbf{x},\mathbf{x}') = \frac{\sigma^2_{mag}}{l^2}\left[\mathbf{I}_{3,3}-\frac{(\mathbf{x}-\mathbf{x'})(\mathbf{x}-\mathbf{x'})^\top}{l^2}\right]\exp\left(-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2l^2}\right), \tag{4.17}$$

where, using the linearity of the covariance, (4.12) follows roughly from

$$\begin{aligned}\mathrm{Cov}(\frac{\partial\varphi(\mathbf{x})}{\partial x_i},\varphi(\mathbf{x})) &= \lim_{h\to 0}\mathrm{Cov}(\frac{\varphi(\mathbf{x})-\varphi(\mathbf{x}+e_i h)}{h},\varphi(\mathbf{x}')) \\ &= \lim_{h\to 0}\frac{\mathrm{Cov}(\varphi(\mathbf{x}),\varphi(\mathbf{x}')-\mathrm{Cov}(\varphi(\mathbf{x}+e_i h),\varphi(\mathbf{x}'))}{h} \\ &= \frac{\partial}{\partial x_i}\mathrm{Cov}(\varphi(\mathbf{x}),\varphi(\mathbf{x}')).\end{aligned} \tag{4.18}$$

Where e_i is the unit vector in the direction of x_i. Now we have a kernel which encodes the a priori assumption that the **H**-field is curl-free.

To make this kernel more flexible, we can set different length-scale parameters to allow for e.g. a vector field which changes less quickly in the z-direction than in the x-direction. To this end, we model the potential as a squared-exponential with different length-scale parameters for each direction. This gives

$$\varphi(\mathbf{x}) \sim GP(0,k(\mathbf{x},\mathbf{x}')) \quad \text{with} \quad k(\mathbf{x},\mathbf{x}') = \sigma^2_{mag}\exp\left(-\frac{(x_1-x'_1)^2}{2l_1^2}-\frac{(x_2-x'_2)^2}{2l_2^2}-\frac{(x_3-x'_3)^2}{2l_3^2}\right). \tag{4.19}$$

Following the derivation above, we now find

$$[\mathrm{Cov}(\mathbf{H}(\mathbf{x},\mathbf{H}(\mathbf{x}'))]_{i,j} = \sigma^2\left[\frac{\delta_{i,j}}{l_i^2}-\frac{(x_i-x'_i)(x_j-x'_j)}{l_i^2 l_j^2}\right]\exp\left(-\frac{(x_1-x'_1)^2}{2l_1^2}-\frac{(x_2-x'_2)^2}{2l_2^2}-\frac{(x_3-x'_3)^2}{2l_3^2}\right). \tag{4.20}$$

4.2. Divergence-Free Kernel Derivation

In this derivation, we follow Wahlström 2015 appendix A.2.
Similarly to how the curl-free kernel is derived in Chapter 4.1 using the scalar magnetic potential φ, we now need the vector magnetic potential $\mathbf{A}$ to derive the divergence-free kernel. The relevance of the vector potential is due to the following important identity

$$\nabla \cdot (\nabla \times \mathbf{A}) \equiv 0. \tag{4.21}$$

This identity implies that a divergence-free vector field can be obtained by taking to curl of another vector field. Hence we can define $\mathbf{B}$ as the curl of some vector field. This gives

$$\mathbf{B} = \nabla \times \mathbf{A} \iff \mathbf{B}_i(\mathbf{x}) = \sum_{j=1}^{3}\sum_{k=1}^{3} \epsilon_{ijk} \frac{\partial}{\partial x_j} \mathbf{A}_k(\mathbf{x}), \tag{4.22}$$

where ϵ_{ijk} is the Levi-Civita symbol defined by

$$\epsilon_{ijk} = \begin{cases} 1 & \text{if } (ijk) = (123), (231) \text{ or } (312) \\ -1 & \text{if } (ijk) = (321), (213) \text{ or } (132) \\ 0 & \text{if } i = j, j = k \text{ or } k = i. \end{cases} \tag{4.23}$$

By modelling the vector potential as a GP where each component of $\mathbf{A}$ is considered an independent smooth function which share hyperparameters and have different length-scales in different directions as

$$\mathbf{A}(\mathbf{x}) \sim GP(0, K_{SE}(\mathbf{x}, \mathbf{x}')), \tag{4.24}$$

$$K_{SE}(\mathbf{x}, \mathbf{x}') = I_{3,3} \cdot \sigma_{mag}^2 \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{x}')^\top \mathrm{diag}(\mathbf{l})^{-2}(\mathbf{x} - \mathbf{x}')\right), \tag{4.25}$$

$$= I_{3,3} \cdot k_{SE'}(\mathbf{x}, \mathbf{x}'). \tag{4.26}$$

Where $\mathrm{diag}(\mathbf{l})$ denotes the square diagonal matrix with $\mathbf{l}$ on its diagonal. We find that $\mathbf{B}(\mathbf{x})$ is also a GP (3.60)

$$\mathbf{B}(\mathbf{x}) \sim GP(0, K_{\mathbf{B}}(\mathbf{x}, \mathbf{x}')), \tag{4.27}$$

where $K_{\mathbf{B}}(\mathbf{x}, \mathbf{x}')$ is a matrix valued kernel who's components are given by

$$[K_{\mathbf{B}}(\mathbf{x}, \mathbf{x}')]_{i,j} = \mathrm{Cov}\left[B_i(\mathbf{x}), B_j(\mathbf{x}')\right] \tag{4.28}$$

$$= \mathrm{Cov}\left[\sum_{k,l}^{3} \epsilon_{ikl} \frac{\partial}{\partial x_k} \mathbf{A}_l(\mathbf{x}), \sum_{m,n}^{3} \epsilon_{jmn} \frac{\partial}{\partial x_m} \mathbf{A}_n(\mathbf{x})\right] \tag{4.29}$$

$$= \sum_{k,l,m,n}^{3} \epsilon_{ikl}\epsilon_{jmn} \frac{\partial^2}{\partial x_k \partial x_m} K_{ln}(\mathbf{x}, \mathbf{x}'). \tag{4.30}$$

This follows from (4.18) and the linearity of the covariance function. Continuing, we find

$$[K_{\mathbf{B}}(\mathbf{x}, \mathbf{x}')]_{i,j} = \sum_{k,l,m,n}^{3} \epsilon_{ikl}\epsilon_{jmn}\delta_{ln} \frac{\partial^2}{\partial x_k \partial x_m} k_{SE'}(\mathbf{x}, \mathbf{x}') \tag{4.31}$$

$$= \sum_{k,m}^{3} (\delta_{ij}\delta_{km} - \delta_{im}\delta_{kj}) \frac{\partial^2}{\partial x_k \partial x_m} k_{SE'}(\mathbf{x}, \mathbf{x}'). \tag{4.32}$$

$$= \sum_{k,m}^{3} (\delta_{ij}\delta_{km} - \delta_{im}\delta_{kj}) \left(\frac{\delta_{km}}{l_k^2} - \frac{x_k - x_k'}{l_k^2} \frac{x_m - x_m'}{l_m^2}\right) k_{SE'}(\mathbf{x}, \mathbf{x}'), \tag{4.33}$$

where (4.33) follows from (4.20) and (4.32) follows as

$$\epsilon_{ikl}\epsilon_{jmn}\delta_{ln} = \begin{cases} 1 & \text{if } (ikl) \;\&\; (jmn) = (123)\ (321) \text{ or } (231) \\ & \text{or if } (ikl) \;\&\; (jmn) = (213)\ (231) \text{ or } (321) \\ -1 & \text{if } (k = m \text{ or } i = j) \text{ and not } (i = m \text{ or } k = j) \text{ (as then } j = m \text{ or } i = k) \\ 0 & \text{else.} \end{cases} \tag{4.34}$$

Now with the following equations

$$\sum_{k,m}^{3} \delta_{ij}\delta_{km}\delta_{km} = 3\delta_{ij}, \tag{4.35}$$

$$\sum_{k,m}^{3} \delta_{im}\delta_{kj}\delta_{km} = \delta_{ij}, \tag{4.36}$$

$$\sum_{k,m}^{3} \delta_{ij}\delta_{km}\frac{(x_k - x'_k)(x_m - x'_m)}{l_k^2 l_m^2} = \sum_{k}^{3} \delta_{i,j}\frac{(x_k - x'_k)^2}{l_k^4}, \tag{4.37}$$

$$\sum_{k,m}^{3} \delta_{im}\delta_{kj}\frac{(x_k - x'_k)(x_m - x'_m)}{l_k^2 l_m^2} = \frac{(x_i - x'_i)(x_j - x'_j)}{l_i^2 l_j^2}, \tag{4.38}$$

we can simplify (4.33) to

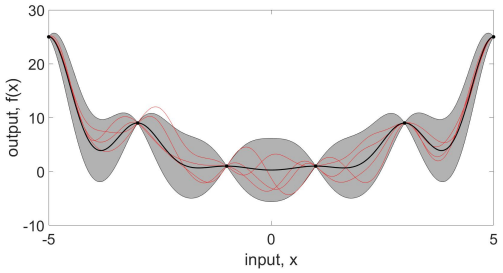
$$[K_{\mathbf{B}}(\mathbf{x}, \mathbf{x}')]_{i,j} = \left(2\frac{\delta_{ij}}{l_i^2} - \delta_{ij}\sum_{k}^{3}\left(\frac{x_k - x'_k}{l_k^2}\right)^2 + \frac{(x_i - x'_i)(x_j - x'_j)}{l_i^2 l_j^2} \right) \cdot k_{SE'}(\mathbf{x}, \mathbf{x}'). \tag{4.39}$$

This yields a kernel which encodes the a priori assumption of the **B**-field being divergence-free.

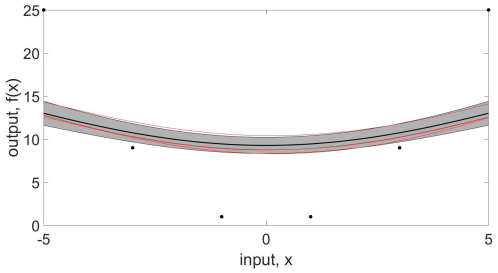
CHAPTER 5

Optimizing hyperparameters

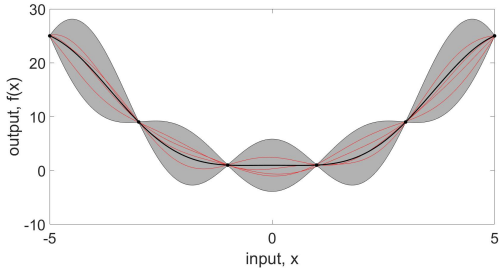
The parameters σ_{mag} and l or $\mathbf{l}$, together with σ_{noise} and σ_{mean}, as presented in (4.20), (4.39) and (6.1) are known as hyperparameters as they are part of a non-parameteric model. These hyperparameters further specify the type of function one is modelling. Finding accurate values for these parameters is, similar to finding a fitting kernel, integral to the accuracy of the prediction. The influence these parameters on the distribution of functions is shown in Figure 5.1 and 5.2. σ_{noise} is not considered a hyperparameter in this work. This is because this parameter can also be found from data and optimizing hyperparameters is computationally expensive.



(A) $(l, \sigma_{mag}, \sigma_{noise}) = (0.5, 3, 0.001)$



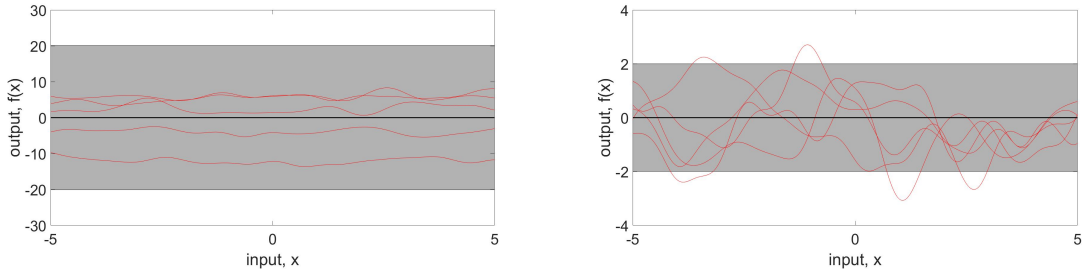
(B) $(l, \sigma_{mag}, \sigma_{noise}) = (10, 3, 1)$



(C) $(l, \sigma_{mag}, \sigma_{noise}) = (2, 30, 0.01)$

FIGURE 5.1. Regression on 6 data points generated from $f(x) = X^2$ shown by black dots using the SE kernel (3.55) with varying hyperparameters. The mean of the posterior distribution is shown in black, realisations of which are shown in red and twice the standard deviation is shown in grey.

As these hyperparameters influence the prediction capabilities of Gaussian process regression, it is important to find accurate values. In this paper this was done by optimizing the log likelihood (also called log marginal likelihood) $\log(f(\mathbf{y}|X, \boldsymbol{\theta}))$ of the data given the model with hyperparameters $\boldsymbol{\theta}$ and the known inputs X. As this will allow us to use efficient solvers. Using that by assumption, $\mathbf{y} \sim N(\mathbf{0}, K + \sigma^2_{noise} I)$, and with (3.2) the log likelihood is given by



(A) $(l, \sigma_{mag}, \sigma_{noise}, \sigma_{mean}) = (1, 1, 0.01, 10)$ (B) $(l, \sigma_{mag}, \sigma_{noise}, \sigma_{mean}) = (1, 1, 0.01, 0)$

FIGURE 5.2. priori distribution of the SE kernel (3.55) plus a constant factor σ^2_{mean}.

$$\log(f(\mathbf{y}|X, \boldsymbol{\theta})) = \log\left((2\pi)^{-\frac{n}{2}}|(K + \sigma^2_{noise}I)|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\mathbf{y}^\top(K + \sigma^2_{noise}I)^{-1}\mathbf{y}\right)\right) \tag{5.1}$$

$$= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|K + \sigma^2_{noise}I| - \frac{1}{2}\mathbf{y}^\top(K + \sigma^2_{noise}I)^{-1}\mathbf{y}. \tag{5.2}$$

This exists as $K + \sigma^2_{noise}I$ is a positive definite matrix because K is positive semi-definite and $\sigma^2_{noise}I$ is positive definite. In optimizing this likelihood, the partial derivatives with respect to the hyperparameters are useful. They are given by (Rasmussen and Williams 2005)

$$\frac{\partial}{\partial\theta} log(f(\mathbf{y}|X, \boldsymbol{\theta})) = -\frac{1}{2}\frac{\partial \log|K + \sigma^2_{noise}I|}{\partial\theta_i} - \frac{1}{2}\mathbf{y}^\top\frac{\partial(K + \sigma^2_{noise}I)^{-1}}{\partial\theta_i}\mathbf{y} \tag{5.3}$$

$$= -\frac{1}{2}\text{trace}\left((K + \sigma^2_{noise}I)^{-1}\frac{\partial(K + \sigma^2_{noise}I)}{\partial\theta_i}\right) +$$

$$\frac{1}{2}\mathbf{y}^\top(K + \sigma^2_{noise}I)^{-1}\frac{\partial(K + \sigma^2_{noise}I)}{\partial\theta_i}(K + \sigma^2_{noise}I)^{-1}\mathbf{y} \tag{5.4}$$

$$= \frac{1}{2}\text{trace}\left(\left(\boldsymbol{\alpha}\boldsymbol{\alpha}^\top - (K + \sigma^2_{noise}I)^{-1}\right)\frac{\partial(K + \sigma^2_{noise}I)}{\partial\theta_i}\right), \tag{5.5}$$

where $\boldsymbol{\alpha} = (K + \sigma^2_{noise}I)^{-1}\mathbf{y}$. (5.5) follows from the identities

$$\frac{\partial A^{-1}}{\partial\theta} = -A^{-1}\frac{\partial A}{\partial\theta}a^{-1} \tag{5.6}$$

and

$$\frac{\partial}{\partial\theta} log|A| = \text{trace}\left(A^{-1}\frac{\partial A}{\partial\theta}\right). \tag{5.7}$$

For a symmetric positive definite matrix A. Identity (5.6) holds as

$$\frac{\partial I}{\partial\theta} = \frac{\partial AA^{-1}}{\partial\theta} \tag{5.8}$$

$$= \frac{\partial A}{\partial\theta}A^{-1} + A\frac{\partial A^{-1}}{\partial\theta} \tag{5.9}$$

$$\implies \frac{\partial A^{-1}}{\partial\theta} = -A^{-1}\frac{\partial A}{\partial\theta}A^{-1}. \tag{5.10}$$

To see why identity (5.7) holds, notice that the eigenvectors $\mathbf{u}_i$ corresponding to eigenvalues λ_i of A are orthogonal

$$\mathbf{u}_i^\top \mathbf{u}_j = \delta_{ij}. \tag{5.11}$$

After differentiation we find

$$\frac{\partial \mathbf{u}_i^\top}{\partial \theta}\mathbf{u}_j + \mathbf{u}_i^\top \frac{\partial \mathbf{u}_j}{\partial \theta} = 0. \tag{5.12}$$

Now, note that the determinant of a matrix is given by the product of its eigenvalues,

$$|A| = \prod_i \lambda_i, \tag{5.13}$$

which gives

$$\frac{\partial \log(|A|)}{\partial \theta} = \sum_i \frac{\partial \log(\lambda_i)}{\partial \theta} = \sum_i \frac{1}{\lambda_i}\frac{\partial \lambda_i}{\partial \theta}. \tag{5.14}$$

To find a similar expression for the right side of the identity consider

$$A^{-1} = \sum_i \lambda_i^{-1}\mathbf{u}_i\mathbf{u}_i^\top, \tag{5.15}$$

$$\frac{\partial A}{\partial \theta} = \sum_j \frac{\partial \lambda_j}{\partial \theta}\mathbf{u}_j\mathbf{u}_j^\top + \sum_j \lambda_j \left(\frac{\partial \mathbf{u}_j^\top}{\partial \theta}\mathbf{u}_j + \mathbf{u}_j^\top \frac{\partial \mathbf{u}_j}{\partial \theta}\right), \tag{5.16}$$

$$A^{-1}\frac{\partial A}{\partial \theta} = \sum_{i,j} \lambda^{-1}\frac{\partial \lambda_j}{\partial \theta}\mathbf{u}_i\mathbf{u}_i^\top\mathbf{u}_j\mathbf{u}_j^\top + \sum_{i,j}\mathbf{u}_i\mathbf{u}_i^\top\left(\frac{\partial \mathbf{u}_j^\top}{\partial \theta}\mathbf{u}_j + \mathbf{u}_j^\top \frac{\partial \mathbf{u}_j}{\partial \theta}\right), \tag{5.17}$$

which together with

$$\text{trace}\left[\mathbf{u}_i\mathbf{u}_i^\top\mathbf{u}_j\mathbf{u}_j^\top\right] = (\mathbf{u}_i^\top\mathbf{u}_j)(\mathbf{u}_i^\top\mathbf{u}_j) = \delta_{ij} \tag{5.18}$$

$$\text{trace}\left[\mathbf{u}_i\mathbf{u}_i\top\left(\frac{\partial \mathbf{u}_j}{\partial \theta}\mathbf{u}_j^\top\mathbf{u}_j\frac{\partial \mathbf{u}_j^\top}{\partial \theta}\right)\right] = \left(\mathbf{u}_i^\top\frac{\partial \mathbf{u}_j}{\partial \theta}\right)(\mathbf{u}_j^\top\mathbf{u}_i) + (\mathbf{u}_i^\top\mathbf{u}_j)\left(\frac{\partial \mathbf{u}_i^\top}{\partial \theta}\mathbf{u}_j\right) \tag{5.19}$$

$$= \delta_{ij}\left(\mathbf{u}_i^\top\frac{\partial \mathbf{u}_j}{\partial \theta} + \frac{\partial \mathbf{u}_i^\top}{\partial \theta}\mathbf{u}_j\right) \tag{5.20}$$

$$= 0, \tag{5.21}$$

give that

$$\text{trace}A^{-1}\frac{\partial A}{\partial \theta} = \sum_i \frac{1}{\lambda_i}\frac{\partial \lambda_i}{\partial \theta} \tag{5.22}$$

for the right hand size of (5.7) aswell.

Continuing with the log likelihood (5.2), the partial derivatives of K are elementwise derivatives. As we do not consider the standard deviation of the error σ_{noise} a hyperparameter and writing $a_n = \frac{1}{l_n^2}$, the partial derivatives for the curl-free kernel with separate length-scales

can be calculated from

$$\left[\frac{\partial k(\mathbf{x},\mathbf{x}')}{\partial \sigma_{mag}}\right]_{i,j} = 2\sigma_{mag}\left[\frac{\delta_{i,j}}{l_i^2} - \frac{(x_i - x_i')(x_j - x_j')}{l_i^2 l_j^2}\right]\exp\left(-\frac{(x_1 - x_1')^2}{2l_1^2} - \frac{(x_2 - x_2')^2}{2l_2^2} - \frac{(x_3 - x_3')^2}{2l_3^2}\right) \tag{5.23}$$

$$\begin{aligned}\left[\frac{\partial k(\mathbf{x},\mathbf{x}')}{\partial l_n}\right]_{i,j} = \left[\frac{\partial k(\mathbf{x},\mathbf{x}')}{\partial a_n}\frac{\partial a_n}{\partial l_n}\right]_{i,j} = -\frac{2}{l_n^3}\sigma_{mag}^2 \exp\left(-\frac{(x_1 - x_1')^2}{2l_1^2} - \frac{(x_2 - x_2')^2}{2l_2^2} - \frac{(x_3 - x_3')^2}{2l_3^2}\right) \\ \left(\left[\frac{\delta_{i,j}}{l_i^2} - \frac{(x_i - x_i')(x_j - x_j')}{l_i^2 l_j^2}\right]\frac{-(x_n - x_n')^2}{2} + \delta_{i,j,n}\left(1 - \frac{(x_i - x_i')(x_j - x_j')}{2l_i^2 f}\right)\right. \\ \left. - \delta_{i,n}(1 - \delta_{j,n})\left(\frac{(x_i - x_i')(x_j - x_j')}{l_i^2}\right) + \delta_{j,n}(1 - \delta_{i,n})\left(\frac{(x_i - x_i')(x_j - x_j')}{l_j^2}\right)\right)\end{aligned} \tag{5.24}$$

$$\left[\frac{\partial k(\mathbf{x},\mathbf{x}')}{\partial \sigma_{mean}}\right]_{i,j} = 2\sigma_{mean} I. \tag{5.25}$$

This means that the log likelihood can be effectively optimized using gradient based optimizations. It is possible for the optimization method to terminate at a local minimum. All local minima correspond to a certain interpretation of data. For example, one minimum might correspond to a model which allows variation at a small length-scale with low noise, while another minimum might correspond to a model with a high amount of noise and a high length-scale. To avoid local minima the optimization process was executed multiple times with different initial values, in problems where local minima are still a problem, global optimization tools can be used such as a genetic algorithm. In the case of the divergence-free kernel, no gradient was available and as such a constraint optimization process was used in this thesis.

CHAPTER 6

Experiments

The model used in our experiments is a model in which the non-zero mean of the field is integrated out (Wahlström et al. 2013) and contributes constantly to the covariance function, the variations of the magnetic field are modeled by the curl-free kernel as derived in Section 4.1. This model can be written as

$$\begin{aligned} H(\mathbf{x}) &\sim GP(\mathbf{0}, K_{\mathbf{H}}(\mathbf{x}, \mathbf{x}') + \sigma_{mean}^2 I), \\ \mathbf{y}_i &= H(\mathbf{x}_i) + \epsilon, \end{aligned} \tag{6.1}$$

where $\epsilon \sim N(\mathbf{0}, \sigma_{noise}^2 I)$ corresponds to sensor noise and σ_{noise} has the same unit as the field measured.

We conducted several experiments with the final goal of testing whether we can make a Gaussian process model capable of accurately interpolating measurements of the magnetic signature of a steel object. the model was first tested on the field of a single simulated magnetic point dipole, after which the model was tested on a field generated by multiple different simulated dipoles. After this we moved on to real data, measured from a steel object under different external fields. The hyperparameters were first trained on a data set after which we tried to replicate that same data set. As a final experiment it was tested whether those found hyperparameters truly could represent the test object. This was done by trying to replicate other measurements taken of the object under a different external magnetic field. Unless mentioned otherwise, we will use the curl-free kernel with separate length-scales as specified in section 4.1.

6.1. Error measure definition

In order to quantify the results of the experiments, we use the root-mean-squared error (RMSE) and normalized root-mean-squared error (NRMSE). Given predictions $\{\hat{y}_i | i = 1, ..., n\}$ of corresponding true values $\{y_i | i = 1, .., n\}$ the RMSE is given by

$$\text{RMSE} = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y} - y)^2}{n}} \tag{6.2}$$

and the NRMSE is given by

$$\text{NRMSE} = \frac{\text{RMSE}}{\max(y) - \min(y)}. \tag{6.3}$$

This leads the unit of the RMSE to be the unit of the measured phenomena (in our case $[Am^{-1}]$ is case of measurements of the **H**-field and $[T]$ in case of measurements of the **B**-field) and the NRMSE to be dimensionless. The NRMSE is often interpreted to be a percentage and as such this measure can be used to compare the quality of regression across different datasets. In this work each component of each prediction is considered a separate approximation of its corresponding true component for the error assessment. In the case that

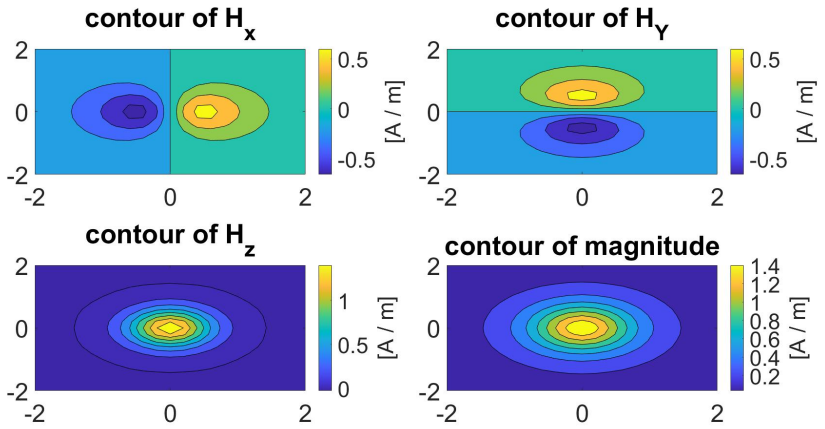


FIGURE 6.1. **H**-field resulting from a magnetic dipole in the origin with $\mathbf{m} = [0,0,1]^\top$ at 1 meter above the dipole.

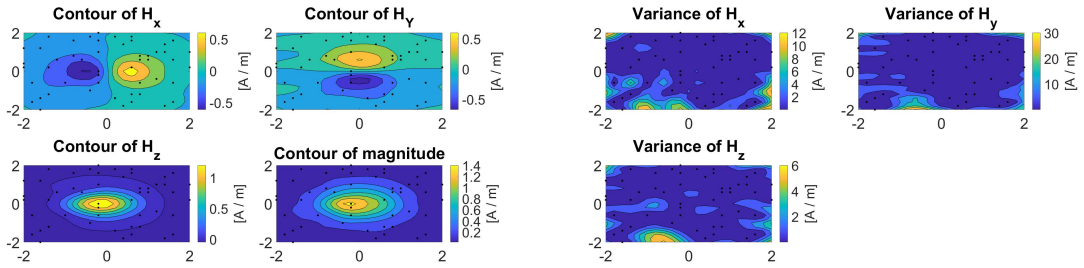
data points are sampled at random from the field, the corresponding (N)RMSE's presented are calculated three times and averaged to avoid uniquely negative or positive distributions of data points. It should be kept in mind that, as the NRMSE is normalised by the difference between the maximum and minimum value y obtains, a data set with relatively large extreme values might have an artificially low NRMSE.

6.2. Simulated experiments

The derivation of the **H**-field equation resulting from a dipole can be found in Section 2.2. The field resulting from multiple dipoles can be found by simply summing the fields of all the dipoles separately. this is a fundamental assumption in magnetostatics.

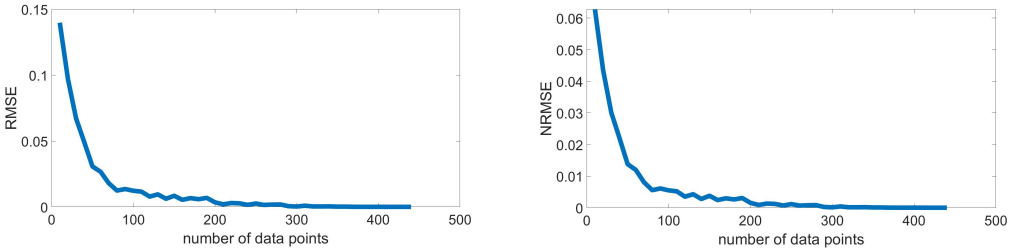
6.2.1. Single Dipole. As a first experiment we defined a dipole in the origin with magnetic moment $\mathbf{m} = [0,0,1]^\top$. Contour plots of the resulting **H**-field can be found in Figure 6.1. This field is defined on $21^2 = 441$ evenly spaced points in a plane with $(x,y,z) \in [-2,2] \times [-2,2] \times \{1\}$. First, the hyperparameters were estimated by maximizing the log likelihood as mentioned in chapter 5, using all available points, this yielded $\sigma_{mag} = 2.3394$, $l_1 = 0.4647$, $l_2 = 0.6066$, $l_3 = 1.0062$, $\sigma_{mean} = 0$ the noise hyperparameter σ_{noise} was set to 0.0001 as some noise is required for numerical stability.
From this field we picked a number of randomly distributed points which served as the known, noise free data. all the points were used as a validation set.
The RMSE and NRMSE for different numbers of data points are presented in Figure 6.3. As the data points are drawn at random from the field, the (N)RMSE was calculated 3 times and averaged. Contour plots of the regression results with 50 data points are shown in Figure 6.2. This shows that the model is capable of making accurate predictions about this simple **H** field from sparse and irregular data.

6.2.2. Multiple Dipoles. To increase the complexity of the field we executed a similar experiment but with four dipoles with four different magnetic moments. The dipoles are defined at the locations $[1,1,0]^\top$, $[1,-1,0]^\top$, $[-1,-1,0]^\top$ and $[-1,1,0]^\top$, with respective magnetic moments $[0,0,1]^\top$, $[0,1,0]^\top$, $[1,0,0]^\top$ and $[-1,-1,-1]^\top$. The **H**-field these dipoles generate is shown in Figure 6.4.



(A) Predicted **H**-field. (B) Variance of predicted **H**-field.

FIGURE 6.2. Regression results on a single dipole with 50 data points shown in black at 1 meter above the dipole.



(A) RMSE $[Am^{-1}]$ (B) NRMSE

FIGURE 6.3. Error of regression on a field generated by a single dipole at 1 meter above the dipole.

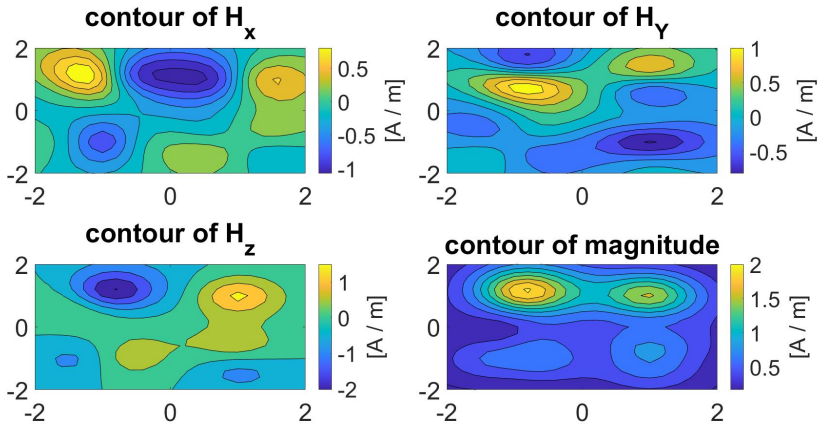
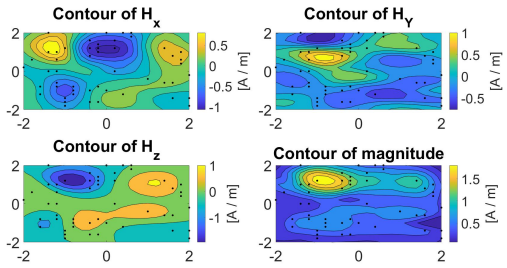
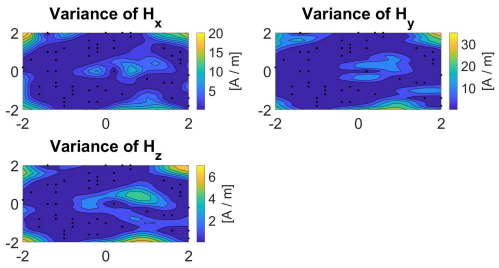


FIGURE 6.4. **H**-field generated by dipoles at $[1,1,0]^\top$, $[1,-1,0]^\top$, $[-1,-1,0]^\top$ and $[-1,1,0]^\top$ with magnetic moments $[0,0,1]^\top$, $[0,1,0]^\top$, $[1,0,0]^\top$ and $[-1,-1,-1]^\top$ at 1 meter above the dipoles.

For this field the hyperparameters were optimized using all 441 points in the plane. local minima proved to be a problem for this field, so a genetic algorithm was used to find $\sigma_{mag} = 2.6706$, $l_1 = 0.5684$, $l_2 = 0.4078$, $l_3 = 0.9911$, $\sigma_{mean} = 0$, where σ_{noise} was set to 0.0001 again.

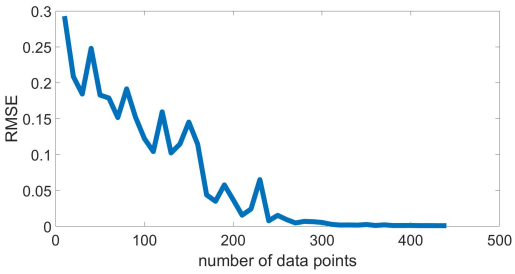


(A) Predicted **H**-field.

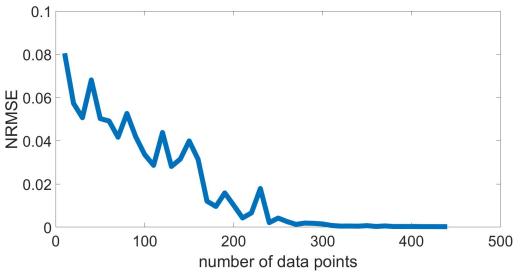


(B) Variance of predicted **H**-field.

FIGURE 6.5. Regression results on multiple dipoles with 50 data points shown in black at at 1 meter above the dipoles.



(A) RMSE $[Am^{-1}]$



(B) NRMSE

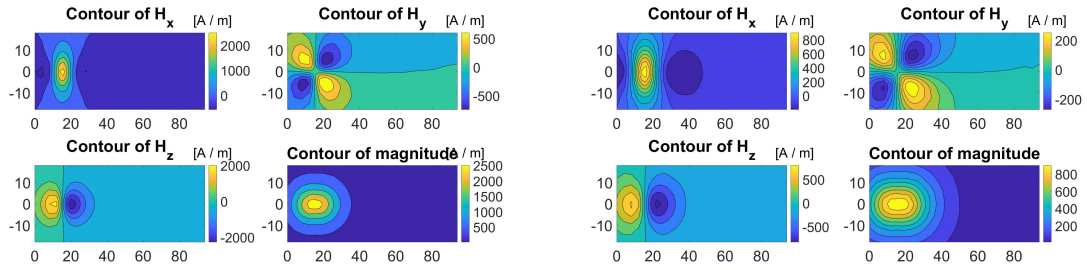
FIGURE 6.6. Error of regression on a field generated by multiple dipoles at 1 meter above the dipoles.

The (N)RMSE for different numbers of data points is shown in Figure 6.6, a contour plot of the regression using 50 data points can be found in Figure 6.5.

6.3. Real Data

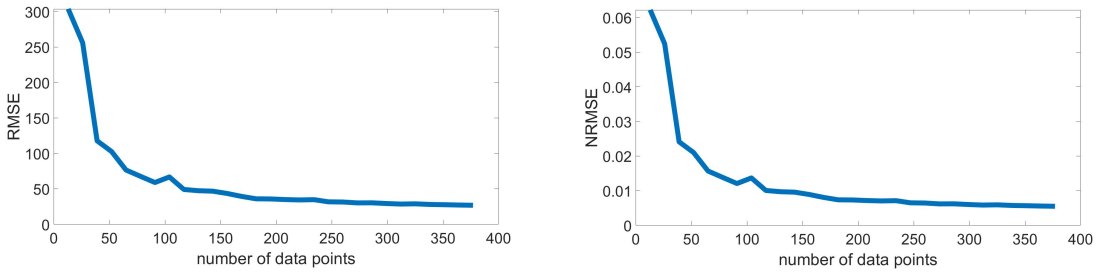
Judging the results of these experiments good, the model was tested on real data. The data is of the distortion field obtained from a steel object under different external magnetic fields. The magnetic field of the object is measured by two sensor arrays at 9 and 13 meters below the object. The data consist of 754 points, half of which lie in the plane at 9 meters below the object, and half of which lie at 13 meters below the object. The data points lie in an irregular grid spanning $[0, 92] \times [-18, 18] \times \{z\}$, where the grid gets finer around the 'interesting' part of the data. We have several data sets of the distortion field under various magnetic conditions. The data provided was of the **B**-field, but as $\frac{\mathbf{B}}{\mu_0} - \mathbf{M} = \mathbf{H}$ and $\mathbf{M} = 0$ outside of magnetized material, the data is still of curl- (and divergence-) free field. We omit the division by μ_0 to make the data more readable. With this data we set up the following questions:

- Is the model still capable of making accurate predictions on real data?
- Is is possible to find accurate hyperparameters using one dataset that translate well to other sets?
- Is is possible to measurements from one plane to make accurate predictions about another plane?
- How do the predictions of the **B**- and **H**-field compare?



(A) Training field at 9m. (B) Training field at 13m.

FIGURE 6.7. Training field.



(A) RMSE $[Am^{-1}]$ (B) NRMSE

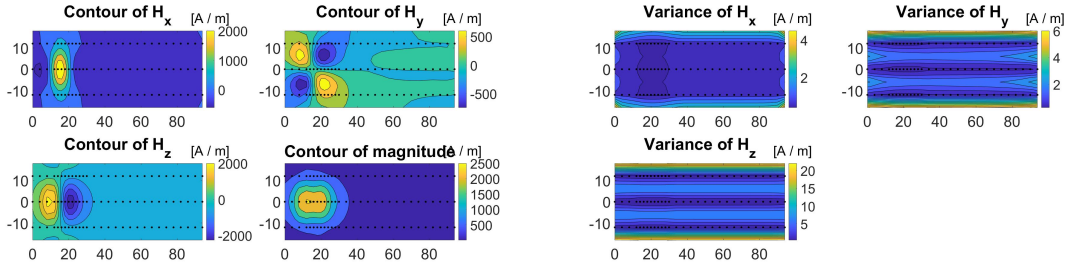
FIGURE 6.8. Error of predicted **H**-field on real data using randomly sampled data points.

- Is it possible to use Gaussian process regression to recreate the **M**-field?

6.3.1. Assessment of the Regressions on Real Data. A contour plot of the field used to train our model can be found in Figure 6.7.

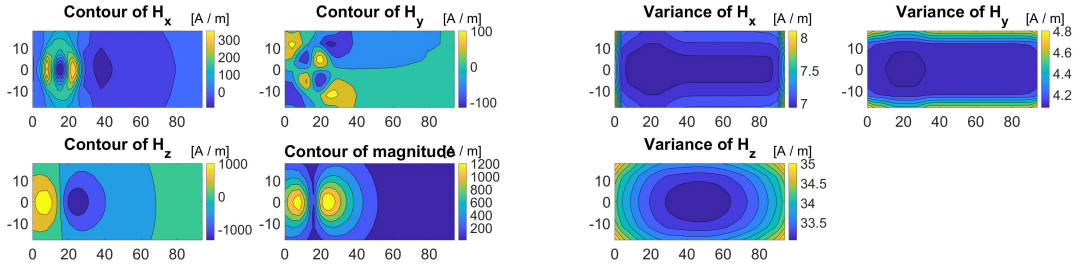
From this field hyperparameters were optimized to be $\sigma_{mean} = 23.9713$, $l_1 = 5.9916$, $l_2 = 7.8230$, $l_3 = 3.3120$, $\sigma_{mean} = 0.2450$. The noise variation was calculated using the right most part of data by considering it a measurement of a constant field and finding its variation, this yielded $\sigma_{noise} = 0.9987$. These parameters will be used for each subsequent test using the curl-free kernel. The model was first tested using randomly sampled data from the plane at a depth of 9 meters to predict the field at 9 meters. The (N)RMSE plots of this test are shown in Figure 6.8. This gives us confidence that our model can still make accurate predictions on real data. An example of the regression using 57 data points is shown in Figure 6.9 where the data is sampled from three lines through the field at 9 meters. Note how the model predicts extreme values at points far away from data points.

It was also attempted to recreate the field at 13 meters using data from the field at 9 meters. This lead to the results shown in Figure 6.10. This might not look promising, with a NRMSE of 0.0805, but the model also indicates that it is less certain of its results as shown by the relatively high variance. This does not compensate for the poor prediction, as the percentage of H_i which fell into the the interval $\hat{H}_i \pm \sqrt{\Sigma_{i,i}}$ was 6.6%. In the case of Figure 6.8 this percentage was 19.36%. Another way to find the magnetic field at a deeper plane is described in Pinheiro 1994, where knowledge of a single component in the entirety of a



(A) Predicted field. (B) confidence of prediction.

FIGURE 6.9. Prediction of **H**-field at 9m using 57 data points from 9m shown in black.



(A) Predicted field. (B) Confidence of regression.

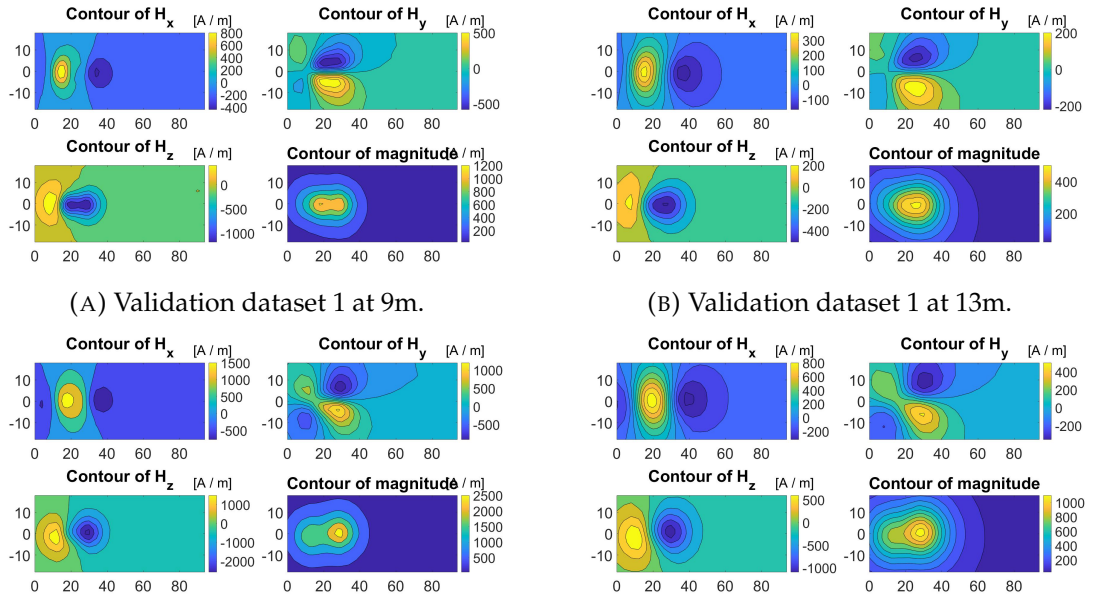
FIGURE 6.10. Prediction of field at 13m using all available data from 9m.

plane is used to find the exact magnetic field in other planes. This theory can be combined with GP's by interpolating data points from a certain plane to the entirety of that plane. We recommend setting $\sigma_{mean} = 0$ to assure the field disappearing far away from the object.

6.3.2. Testing the Hyperparameters. Other than the training dataset shown in Figure 6.7, we also had access to other data sets which were used as validation datasets. Each set is obtained from the same object, under a different external magnetic field. Two of these fields are shown in Figure 6.11, in total there are 5 additional data sets next to the original training data set. We will use these validation sets to test whether the hyperparameters can truly represent the test object.

To test whether the hyperparameters obtained from the training set carry over well to the validation sets, Figure 6.12 was created which shows the different NRMSE's of regression on the training data set and on the additional data sets. As the NRMSE is usually interpreted as a percentage, this comparison across datasets can be made. In Figure 6.13 the regression on validation set 2 at 9 meters is shown using 50 data points shown in black, alongside the true field.

As can be seen the resulting NRMSE differ by a small number of percentage points, and the curves are quite different from the linear lines one sees when poor hyperparameters are used. This result gives us confidence that the found parameters can accurately represent the



(A) Validation dataset 1 at 9m.

(B) Validation dataset 1 at 13m.

(C) Validation dataset 2 at 9m.

(D) Validation dataset 2 at 13m.

FIGURE 6.11. Validation datasets 1 and 2.

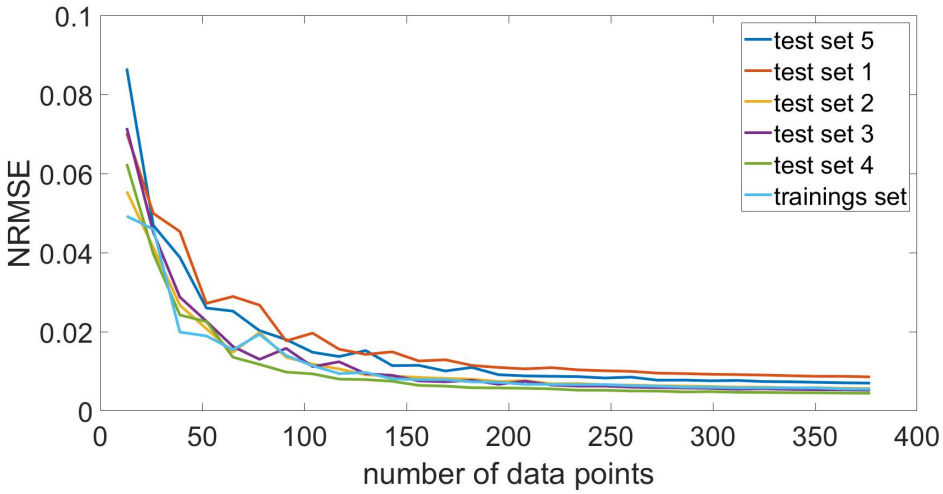
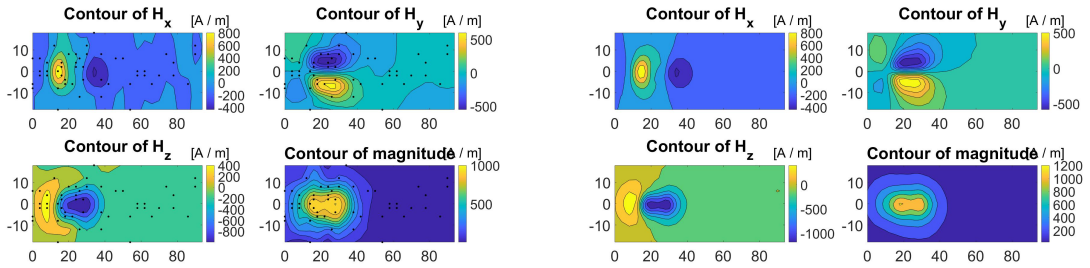


FIGURE 6.12. A comparisson of the NRMSE's of regression on different datasets.

test object.

6.4. Divergence-Free Kernel

The model used in the experiments described above uses the curl-free kernel as derived in Chapter 4.1 to model the variations of the **H**-field. Similarly, a divergence-free kernel can be used to model the variations of the **B**-field. The divergence-free kernel is derived in Chapter



(A) Predicted field. (B) True field.

FIGURE 6.13. Example regression on validation set 2 at 9m with data points shown in black, alongside the true field.

4.2 and is given by

$$[k_{\mathbf{B}}(\mathbf{x},\mathbf{x}')]_{i,j} = \sigma_{mag}^2 \left(2\frac{\delta_{ij}}{l_i^2} - \delta_{ij} \sum_k^3 \left(\frac{x_k - x_k'}{l_k^2} \right)^2 + \frac{(x_i - x_i')(x_j - x_j')}{l_i^2 l_j^2} \right) \cdot \exp\left(-\frac{(x_1 - x_1')^2}{2l_1^2} - \frac{(x_2 - x_2')^2}{2l_2^2} - \frac{(x_3 - x_3')^2}{2l_3^2} \right). \tag{6.4}$$

(6.5)

The results obtained using the divergence-free kernel were generally slightly worse than those obtained using the rotation-free kernel. The optimized hyperparameters of the divergence-free kernel found using the training set were $\sigma_{mag} = 80$, $l_1 = 4.8606$, $l_2 = 5.4658$, $l_3 = 5.0$, $\sigma_{mean} = 0$. The noise parameter σ_{noise} was again set to be 0.9987 as calculated using the training set. The NRMSE of the regression using the divergence-free model with these hyperparameters on the training set, as well as on the test sets is shown in Figure 6.14. The worse results could be explained by the parameters found being in a (worse) local optimum as no gradient based optimizer was used but a constrained optimizer. The Figure does show that the found parameters translate well to other data sets of the same object.

The divergence-free kernel was, however, slightly better at predicting the field at 13 meters deep using only data from 9 meters deep. Where the curl-free kernel obtained a NRMSE of 0.0805, the divergence-free kernel obtained 0.0764. The model also shows its uncertainty better. Where for the curl-free 6.6% of the H_i fell into the interval $\hat{H} \pm \sqrt{\Sigma_{i,i}}$, 34.8% fell into that interval using the divergence-free kernel. Visually, the prediction was also better as seen in Figure 6.15. Again, the field at a deeper plane can also be found using the theory in Pinheiro 1994.

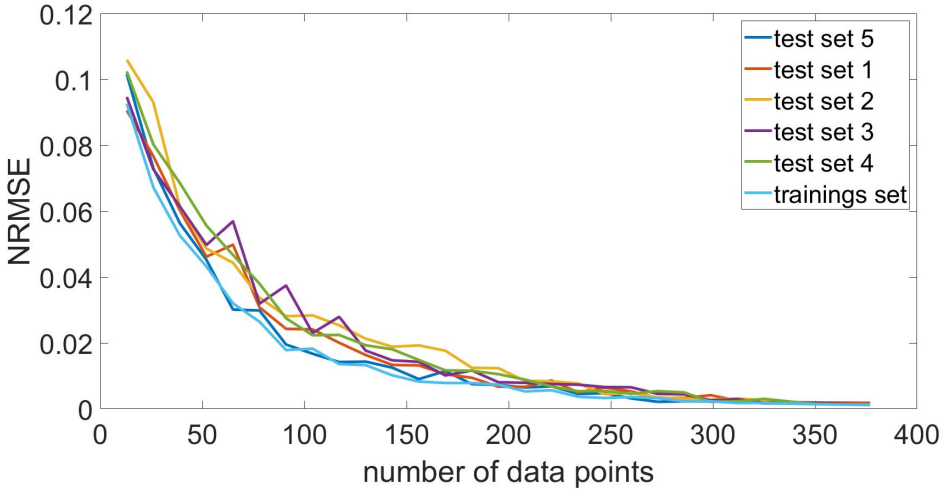
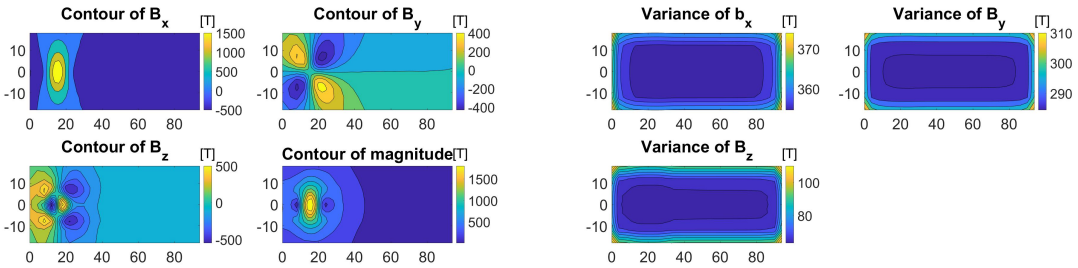


FIGURE 6.14. The NRMSE's of regression on different datasets using the divergence-free model.



(A) Predicted field. (B) Confidence of prediction.

FIGURE 6.15. Predicted **H**-field at 13m using all available data from 9m of the training set shown in black and the divergence-free model.

CHAPTER 7

Predicting the M-field

As we currently have a model of the **H**- and **B**-field and since $\mathbf{M} = \frac{\mathbf{B}}{\mu_0} - \mathbf{H}$ we can also formulate a model which is capable of predicting the **M**-field. To do this we follow the methodology used by Wahlström et al. 2013. Where the **M**- and **B**-field are modeled jointly.

7.1. Model

We consider a model where we normalize the **B**-field by μ_0 to equalize the measurements of the **H**- and **B**-field. We have measurement data $\mathbf{y}$ of the true magnetic field $\mathbf{f}$ which is corrupted by Gaussian noise

$$\begin{aligned} \mathbf{y}_{\mathbf{B},i} =& \mathbf{f}_{\mathbf{B}}(\mathbf{x}_i) + \epsilon_i, \\ \mathbf{y}_{\mathbf{H},i} =& \mathbf{f}_{\mathbf{H}}(\mathbf{x}_i) + \epsilon_i, \qquad \epsilon_i \sim N(0, \sigma_{noise} I_3), \end{aligned} \tag{7.1}$$

where $\mathbf{y}_{\mathbf{H},i} = \mathbf{y}_{\mathbf{B},i}$ and $\mathbf{f}_{\mathbf{B}}(\mathbf{x}_i) = \mathbf{f}_{\mathbf{H}}(\mathbf{x}_i)$ as the point of measurement $\mathbf{x}_i$ lies outside of magnetized material and the **B**- and **H**-field are equalized. This information is taken into account by considering a noise free measurement of the **M**-field as

$$0 = \mathbf{y}_{\mathbf{M},i} = \mathbf{f}_{\mathbf{M}}(\mathbf{x}_i) = \mathbf{f}_{\mathbf{B}}(\mathbf{x}_i) - \mathbf{f}_{\mathbf{H}}(\mathbf{x}_i). \tag{7.2}$$

This connection between the **M**-, **B**- and **H**-field makes the joint estimation possible.
To preform regression we assume that both $\mathbf{f}_{\mathbf{H}}$ and $\mathbf{f}_{\mathbf{B}}$ (and thus also $\mathbf{f}_{\mathbf{M}}$ via (7.2)) are distributed according to a GP with common, yet unknown mean $\boldsymbol{\beta} \sim N(\mathbf{0}, \sigma_{mean} I_3)$ and covariance function which preserve the curl- and divergence-free properties of the field as derived in Sections 4.1 and 4.2. Thus we write

$$\mathbf{f}_{\mathbf{B}} \sim GP(\boldsymbol{\beta}, K_{\mathbf{B}}(\mathbf{x}, \mathbf{x}')), \qquad \mathbf{f}_{\mathbf{H}} \sim GP(\boldsymbol{\beta}, K_{\mathbf{H}}(\mathbf{x}, \mathbf{x}')). \tag{7.3}$$

This model can be written as a standard vector valued GP model

$$\begin{aligned} \mathbf{y}_i &= \mathbf{f}(\mathbf{x}_i) + \epsilon_i, \qquad \epsilon_i \sim N(\mathbf{0}, \Sigma) \\ \mathbf{f}(\mathbf{x}) &\sim GP(\boldsymbol{\beta}, K(\mathbf{x}, \mathbf{x}')), \end{aligned} \tag{7.4}$$

by writing

$$\mathbf{y}_i = \begin{bmatrix} \mathbf{y}_{\mathbf{B},i} \\ \mathbf{y}_{\mathbf{M},i} \end{bmatrix} \qquad \text{and} \qquad \Sigma = \begin{bmatrix} \sigma_{noise} I_3 & 0_{3,3} \\ 0_{3,3} & 0_{3,3} \end{bmatrix}, \tag{7.5}$$

where $0_{3,3}$ the 3×3 matrix filled with zeros, and correspondingly

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} \mathbf{f}_{\mathbf{B}}(\mathbf{x}) \\ \mathbf{f}_{\mathbf{M}}(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} I_3 & 0_{3,3} \\ I_3 & -I_3 \end{bmatrix} \begin{bmatrix} \mathbf{f}_{\mathbf{B}}(\mathbf{x}) \\ \mathbf{f}_{\mathbf{H}}(\mathbf{x}) \end{bmatrix} \sim GP(\beta, K(\mathbf{x}, \mathbf{x}')). \tag{7.6}$$

This gives, writing k for $k(\mathbf{x}, \mathbf{x}')$,

$$\beta = \begin{bmatrix} I_3 & 0_{3,3} \\ I_3 & -I_3 \end{bmatrix} \cdot \begin{bmatrix} N(\mathbf{0}, \sigma_{mean} I_3) \\ N(\mathbf{0}, \sigma_{mean} I_3) \end{bmatrix} = \begin{bmatrix} N(\mathbf{0}, \sigma_{mean} I_3) \\ \mathbf{0} \end{bmatrix}, \tag{7.7}$$

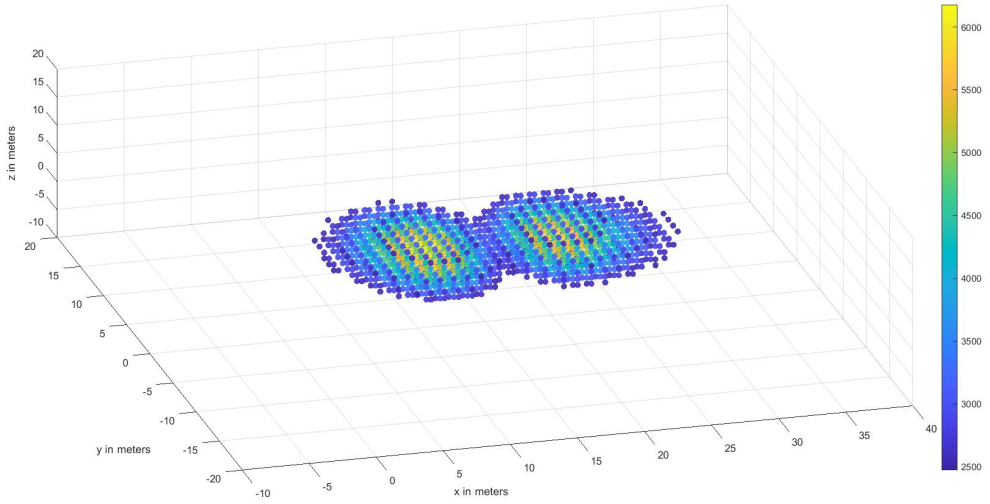


FIGURE 7.1. Predicted **M**-field of training dataset using all available data points. Shown are all the points at which the magnitude of the field is larger than 40% of the maximum magnitude, colored corresponding to their magnitude. The shown point cloud is roughly in an area of $(x, y, z) \in [5, 25] \times [-5, 5] \times [0, 7]$.

and

$$K = \begin{bmatrix} I_3 & 0_{3,3} \\ I_3 & -I_3 \end{bmatrix} \cdot \begin{bmatrix} K_{\mathbf{B}} & 0_{3,3} \\ 0_{3,3} & K_{\mathbf{H}} \end{bmatrix} \cdot \begin{bmatrix} I_3 & I_3 \\ 0_{3,3} & -I_3 \end{bmatrix} = \begin{bmatrix} K_{\mathbf{B}} & K_{\mathbf{B}} \\ K_{\mathbf{B}} & K_{\mathbf{B}} + K_{\mathbf{H}} \end{bmatrix}. \tag{7.8}$$

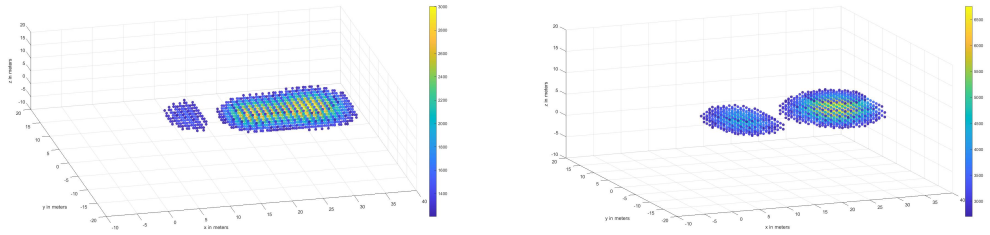
This makes used of the fact that $\mathbf{f}(\mathbf{x}) \sim GP(\boldsymbol{\mu}, K)$ implies $C\mathbf{f}(\mathbf{x}) \sim GP(C\boldsymbol{\mu}, CKC^\top)$ by (3.58). Finally, to create a GP with zero mean, β is integrated out to find

$$K = \begin{bmatrix} K_{\mathbf{B}} + \sigma_{mean} I_3 & K_{\mathbf{B}} \\ K_{\mathbf{B}} & K_{\mathbf{B}} + K_{\mathbf{H}} \end{bmatrix}. \tag{7.9}$$

The resulting model is one where $f : \mathbb{R}^6 \to \mathbb{R}^6$ with covariance function given by (7.9).

7.2. Results

To acquire results with this model, a slight amount of noise was added to the otherwise noise free 'measurements' of the **M**-field to ensure numerical stability. The hyperparameters found earlier for the curl- and divergence-free kernels were used again. A plot of the predicted **M** can be seen in Figure 7.1, where the points in space at which the magnitude of the field is larger then 40% of the maximal magnitude are shown with a color representing the magnitude of the predicted **M** field at the point. It has been checked in the region $(x, y, z) \in [-50, 50] \times [-50, 50] \times [-20, 40]$ that the shown cloud is the only cloud. The same experiment has also been conducted using the validation sets rather then the training set. The results of this, for validation sets 1 and 2, are shown in Figure 7.2. On all the datasets the results are somewhat centered around where the measured object would be, although systematically to0 low, with the magnetization centered roughly around in an area of $(x, y, z) \in [5, 25] \times [-5, 5] \times [2, 5]$. To validate the results it was attempted to recreate the **B**-field using the constructed **M**-field.



(A) Validation set 1. (B) Validation set 2.

FIGURE 7.2. Predicted **M**-field of validation sets 1 and 2 using all available data points. Shown are all the points at which the magnitude of the field is larger than 40% of the maximum magnitude, colored corresponding to their magnitude. The shown point clouds are roughly in an area of $(x, y, z) \in [5, 25] \times [-5, 5] \times [0, 7]$.

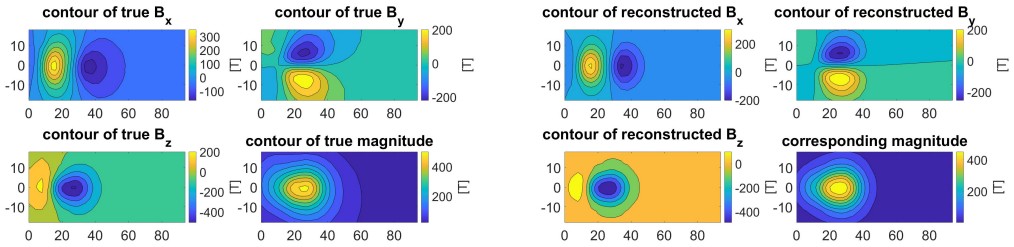
dataset	NRMSE
training set	0.0395
validation set 1	0.0411
validation set 2	0.0449
validation set 3	0.0433
validation set 4	0.0434
validation set 5	0.0330

TABLE 1. NRMSE between the reconstructed **B**-field at 13 meters deep and the measured **B**-field.

7.3. Validation of the results

As we have no information about the true magnetization of the object, we have to validate the results by recreating the **B**-field resulting from the approximated **M**-field. In Section 2.2 the **H**-field of a dipole is derived and if we multiply this field with μ_0 we get a description of the **B**-field of a dipole with a certain magnetic moment. When we define a dipole at each point in space at which we have calculated the magnetic field, with a magnetic moment equal to the magnetic field multiplied by a constant σ_{geom} dependent on the geometry of the structure in which the **M**-field lives, we get the **B**-field resulting from the predicted magnetization field. As we cannot take all the points at which the **M**-field was predicted to be non-zero into account, only points where the magnitude of the **M**-field is larger then 20% of the maximum magnitude were included. The factor σ_{mean} was found by optimizing the NRMSE between the measured **B**-field and the reconstructed **B**-field based on the predicted **M**-field both at 9 meters deep of the training set. This yielded a factor $\sigma_{geom} = 0.1336$. This parameter can now be used to find the NRMSE at 13 meters deep for all datasets. The results of this are shown in Table 1.

These results are surprisingly decent, given that the found **M**-fields do not represent the measured object that accurately. A reason for this might be that an **M**-field is not uniquely determined by the **B**-field which generates it. A common example of this is that the magnetic field generated by a uniformly magnetized sphere with a certain radius r is equal to that of the field generated by any other uniformly magnetized sphere with radius r', as long as they share the same magnetic moment (Fitzpatrick 2004) (or equivalently, equal up



(A) Measured **B**-field (B) Reconstructed **B**-field

FIGURE 7.3. Measured and reconstructed **B** field at 13 meters deep for validation set 2.

to multiplication with a constant as long as they are magnetized in the same direction). This implies that our model does not give an approximation of the the true magnetization field, but of a possible magnetization field. An example of the reconstructed **B**-field and the true **B**-field for validation set 2 is shown in Figure 7.3. To reconstruct a well-defined **M** more research is needed in which the geometry of the measured object is possibly taken into account such that the region in which $\mathbf{f_H} = \mathbf{f_B}$ can be assured is extended. A combination of dipole models and Gaussian processes can also be used.

CHAPTER 8

Conclusions and recommendations

Gaussian process regression is a powerful tool for many purposes. The goal behind this research was to investigate the usefulness of this promising field in the area of magnetostatics and magnetic interpolation. It has been shown that with the right covariance function and accurate hyperparameters, GP regression is capable of finding unique and accurate results with sparse and irregular data.

In chapter 6 several questions arose as to what was possible with Gaussian processes, they were:

- Is the model still capable of making accurate predictions on real data?
- Is is possible to find accurate hyperparameters using one dataset that translate well to other sets?
- Is is possible to measurements from one plane to make accurate predictions about another plane?
- How do the predictions of the **B**- and **H**-field compare?
- Is it possible to use Gaussian process regression to recreate the **M**-field?

As summarised in figure 6.12, the there is little difference in the regression performance between the trainig sets and the validation sets and the general performance gives a respectable NRMSE for sparse and scattered data. It should be noted that the NRMSE might give a skewed view of the true accuracy as most of our data sets contain relatively large extreme values which artificially lowers the NRMSE. As shown most clearly in Figure 6.9, the model is capable of making accurate predictions about extreme values far away from data points. The predictions of the **B**-field generally turned out slightly worse than those of the **H**-field.

The shortcomings of GP regression were also visible. As shown in Figures 6.10 and 6.15, our proposed models were not capable of making accurate prediction about a plane with only data originating from a different nearby plane. In future research GP regression could be combined with the theory in Pinheiro 1994 to improve these results.

With Gaussian process modeling we were also able to predict the magnetization which brought about the modelled distortion fields. Although the results of this approximate a valid possible **M**-field, it cannot be guaranteed that this is an approximation of the true **M**-field. Further research in this area is required to make the problem well-defined and improve the results. Possible extensions include taking the geometry of the modelled object into account and combining GP's with existing dipole models.

As both the process of finding hyperparameters and generating general regression results is computationally expensive it is recommended to investigate efficient implementations such as in Solin and Särkkä 2014.

Bibliography

Agrell, Christian (2019). "Gaussian Processes with Linear Operator Inequality Constraints." In: *Journal of Machine Learning Research* 20, pp. 1–36.

Alvarez, Mauricio A. et al. (n.d.). *Kernels for Vector-Valued Functions: a Review*. URL: arXiv:1106.6251v2.

Fitzpatrick, Richard (2004). "Lecture notes on Classical Electromagnetism." In: pp. 110–113.

Jackson, John David (2007). *Classical Electrodynamics*. Wiley India.

Jr., Edward J. Fuselier (2005). "Refined error estimates for matrix-valued radial basis functions." PhD thesis. URL: https://oaktrust.library.tamu.edu/handle/1969.1/5788.

Macêdo, Ives and Rener Castro (2008). "Learning divergence-free and curl-free vector fields with matrix-valued kernels." In: *Instituto Nacional de Matematica Pura e Aplicada*.

Maxwell, James Clerk (1865). "A dynamical theory of the electromagnetic field." In: *Philosophical Transactions of the Royal Society of London* 155.1, pp. 459–512.

Narcowich, Francis J. and Joseph D. Ward (1994). "Generalized Hermite interpolation via matrix-valued conditionally positive definite functions." In: *Mathematics of Computation* 63, pp. 661–687.

Pinheiro, José Carlos Alves (1994). "Vectorial magnetic field of a vessel from the vertical field." In: *Oceans; MTS/IEEE Conference proceedings*, pp. 519–521.

Rasmussen, Carl and Christopher Williams (2005). *Gaussian Procceses for Machine Learning*. Cambridge, Massachusetts: MIT Press.

Solin, Arno et al. (2015). "Modeling and interpolation of the ambient magnetic field by Gaussian processes." In: URL: arXiv:1509.04634v1.

Solin, Arno and Simo Särkkä (2014). "Hilbert Space Methods for Reduced-Rank Gaussian Process Regression." In: URL: https://link.springer.com/article/10.1007/s11222-019-09886-w.

Wahlstöm, Niklas (2015). "Modeling of Magnetic Fields and Extended Objects for Localization Applications." PhD thesis. URL: http://user.it.uu.se/~thosc112/team/wahlstrom2015phd.pdf.

Wahlström, Niklas et al. (2013). *Proceedings of the 38th International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.