


## Article

# Gaussian Process Regression for Data Fulfilling Linear Differential Equations with Localized Sources <sup>†</sup>

Christopher G. Albert <sup>1,\*</sup>  and Katharina Rath <sup>1,2</sup>

<sup>1</sup> Max-Planck-Institut für Plasmaphysik, Boltzmannstr. 2, 85748 Garching, Germany; katharina.rath@ipp.mpg.de

<sup>2</sup> Department of Statistics, Ludwig-Maximilians-Universität München, Ludwigstr. 33, 80539 Munich, Germany

\* Correspondence: albert@alumni.tugraz.at

<sup>†</sup> This paper is an extended version of our paper published in MaxEnt 2019.

Received: 15 December 2019; Accepted: 24 January 2020; Published: 27 January 2020



**Abstract:** Specialized Gaussian process regression is presented for data that are known to fulfill a given linear differential equation with vanishing or localized sources. The method allows estimation of system parameters as well as strength and location of point sources. It is applicable to a wide range of data from measurement and simulation. The underlying principle is the well-known invariance of the Gaussian probability distribution under linear operators, in particular differentiation. In contrast to approaches with a generic covariance function/kernel, we restrict the Gaussian process to generate only solutions of the homogeneous part of the differential equation. This requires specialized kernels with a direct correspondence of certain kernel hyperparameters to parameters in the underlying equation and leads to more reliable regression results with less training data. Inhomogeneous contributions from linear superposition of point sources are treated via a linear model over fundamental solutions. Maximum likelihood estimates for hyperparameters and source positions are obtained by nonlinear optimization. For differential equations representing laws of physics the present approach generates only physically possible solutions, and estimated hyperparameters represent physical properties. After a general derivation, modeling of source-free data and parameter estimation is demonstrated for Laplace's equation and the heat/diffusion equation. Finally, the Helmholtz equation with point sources is treated, representing scalar wave data such as acoustic pressure in the frequency domain.

**Keywords:** Gaussian process regression; physics-informed methods; kernel methods; field reconstruction; source localization; partial differential equations; meshless methods

## 1. Introduction

The larger context of the present work is the goal to construct reduced complexity models as emulators or surrogates that retain mathematical and physical properties of the underlying system. In recent terminology, such approaches are examples of “physics informed machine learning”. Similar to usual numerical models, the aim here is to represent infinite systems by exploiting finite information in some optimal sense. In the spirit of structure preserving numerics, one tries to move errors to the “right place” to retain laws such as conservation of mass, energy, or momentum. Here, we treat data known to fulfill a given linear differential equation. This article is an extended version of a conference paper [1] presented at the MaxEnt workshop 2019. The revised text adds hyperparameter optimization, results for the heat equation and detailed comparisons to existing methods.

This article deals with Gaussian process (GP) regression on data with additional information known in the form of linear, generally partial differential equations (PDEs). An illustrative example is

the reconstruction of an acoustic sound pressure field and source parameters from discrete microphone measurements. GPs, a special class of random fields, are used in a probabilistic rather than a stochastic sense: estimate a fixed but unknown field from possibly noisy local measurements. Uncertainties in this reconstruction are modeled by a normal distribution.

Using GPs to fit data from PDEs has been a topic of research for some time, especially in the field of geostatistics [2]. A general analysis for deterministic source densities including a number of important properties is given by [3]. In these earlier works GPs are usually referred to as “Kriging” and covariance functions/kernels as “covariograms”. A number of more recent works from various fields [4–8] use the linear operator of a PDE to relate the kernels of source and response field. One of the two is usually modeled by a generic squared exponential kernel. Although the authors of [4,6,7] use such a kernel for the response field and a kernel modified by a differential operator for the source field, [5] models the source field by a generic kernel and applies the inverse (integral) operator to obtain a kernel for the measured response. In contrast to the present approach such methods are suited best for source fields that are non-vanishing across the whole domain. In terms of deterministic numerical methods, one could say that these approaches with volumetric charge densities correspond to meshless variants of the finite element method (FEM).

The approach in the present work instead relies on Gaussian processes that generate *exact* solutions of the homogeneous part of the differential equation [9–11]. This is efficient for problems with mostly source-free domains and requires specialized kernels where possible singularities (virtual sources) are moved outside the domain of interest. In particular, boundary conditions on a finite domain can be either supplied or reconstructed in this fashion. Localized internal point sources are then superimposed as a linear model, using again fundamental solutions in the free field. One can thus interpret this approach as a probabilistic variant of a procedure related to the boundary element method (BEM), known as the *method of fundamental solutions* (MFS) or regularized BEM [12–14]. As in the BEM, the MFS also builds on fundamental solutions, but allows to place sources outside the boundary rather than localizing them on a layer. Thus, the MFS avoids singularities in boundary integrals of the BEM, while retaining a similar ratio of numerical effort and accuracy for smooth solutions. To the best of the author’s knowledge, the probabilistic variant of the MFS via GPs has first been introduced by [9] to solve the boundary value problem of the Laplace equation and dubbed *Bayesian boundary elements estimation method* ((BE)<sup>2</sup>M). This work also provides a detailed treatment of kernels for the 2D Laplace equation. A more extensive and general treatment of the Bayesian context as well as kernels and their connection to fundamental solutions is available in [10] under the term *probabilistic meshless methods* (PMM).

Although the authors of [9] treat boundary data of a the homogeneous Laplace equation and the authors of [10] provides a detailed mathematical foundation, the present work aims to extend the recent work on added point sources in [11], unify the derivation of specialized kernels, and demonstrate usefulness in applications. First, a general derivation is given on how to model PDE data by superposing a GP and a linear model for localized sources. Then, the construction of kernels for the homogeneous part of partial differential equations via according fundamental solutions is described in general. Finally, concrete application examples are given for Laplace/Poisson, heat/diffusion and Helmholtz equation for which the derivation of several kernels is presented. Performance is compared to regression with a generic squared exponential kernel, including hyperparameter optimization in all cases. For the Helmholtz equation we estimate strength and positions of sources by nonlinear optimization.

## 2. GP Regression for Data from Linear PDEs

Gaussian process (GP) regression [15] is a tool to represent and update incomplete information on scalar fields  $u(\mathbf{x})$ , i.e., a real number  $u$  depending on a (multidimensional) independent variable  $\mathbf{x}$

(the more general case of complex valued fields and vector fields is left open for future investigations in this context). A GP with mean  $m(\mathbf{x})$  and covariance function or kernel  $k(\mathbf{x}, \mathbf{x}')$  is denoted as

$$u(\mathbf{x}) \sim \mathcal{G}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')). \quad (1)$$

The choice of an appropriate kernel  $k(\mathbf{x}, \mathbf{x}')$  restricts realizations of (1) to respect regularity properties of  $u(\mathbf{x})$  such as continuity or characteristic length scales. Often regularity of  $u$  does not appear by chance, but rather reflects an underlying law. We will exploit such laws in the construction and application of GPs describing  $u$  for the case described by linear (partial) differential equations:

$$\hat{L}u(\mathbf{x}) = q(\mathbf{x}). \quad (2)$$

where  $\hat{L}$  is a linear differential operator and  $q(\mathbf{x})$  is a source term. In the laws of physics, dimensions of  $\mathbf{x}$  usually consist of space and/or time. Physical scalar fields  $u$  include, e.g., electrostatic potential  $\Phi$ , temperature  $T$ , or pressure  $p$ . Corresponding laws include Gauss' law of electrostatics for  $\Phi$  with weighted Laplacian  $\hat{L} = \varepsilon\Delta$ , thermodynamics for  $T$  with heat/diffusion operator  $\hat{L} = \frac{\partial}{\partial t} - D\Delta$  and frequency-domain acoustics for  $p$  with Helmholtz operator  $\hat{L} = \Delta + k_0^2$ . These operators contain free parameters, namely, permeability  $\varepsilon$ , wavenumber  $k_0$ , and diffusivity  $D$ , respectively. While  $\varepsilon$  may be absorbed inside  $q$  in a uniform material model of electrostatics, estimation of parameters such as  $D$  or  $k_0$  is useful for material characterization.

Consider first the source-free (homogeneous) case

$$\hat{L}u_h(\mathbf{x}) = 0. \quad (3)$$

An unknown field  $u_h(\mathbf{x})$  that fulfills (3) shall be modeled by the Gaussian process

$$u_h(\mathbf{x}) \sim \mathcal{G}(0, k(\mathbf{x}, \mathbf{x}')). \quad (4)$$

Application of a linear operator  $\hat{L}$  yields a modified Gaussian process

$$\hat{L}u_h(\mathbf{x}) \sim \mathcal{G}(0, \hat{L}k(\mathbf{x}, \mathbf{x}')\hat{L}'), \quad (5)$$

where  $\hat{L}'$  acts from the right side with respect to  $\mathbf{x}'$ . In order to fulfill (3) we require (5) to vanish identically, i.e., yield a deterministic zero. Consequently, the kernel  $k(\mathbf{x}, \mathbf{x}')$  needs to satisfy

$$\hat{L}k(\mathbf{x}, \mathbf{x}')\hat{L}' = 0. \quad (6)$$

A discussion on derivation of such kernels is found in Section 2.

For the general case (2), with unknown source density  $q(\mathbf{x})$ , we introduce a linear model

$$q(\mathbf{x}) = \sum_i \varphi_i(\mathbf{x})q_i = \boldsymbol{\varphi}^T(\mathbf{x})\mathbf{q}, \quad (7)$$

with basis functions  $\varphi_i(\mathbf{x})$  and a normally distributed prior

$$\mathbf{q} \sim \mathcal{N}(\mathbf{q}_0, \Sigma_q), \quad (8)$$

with mean  $\mathbf{q}_0$  and prior covariance  $\Sigma_q$  for coefficients  $q_i$  representing source strengths.

For a particular solution  $u_p(\mathbf{x})$  fulfilling the inhomogeneous Equation (2) with source model (8), a linear model induced by the operator  $\hat{L}$  follows as

$$u_p(\mathbf{x}) = \mathbf{h}(\mathbf{x})^T \mathbf{q}, \text{ with } \hat{L}h_i(\mathbf{x}) = \varphi_i(\mathbf{x}). \quad (9)$$

Here, coefficients  $q_i$  remain the same as in (8) and new basis functions  $h_i(\mathbf{x})$  fulfil the differential equation with source density  $\varphi_i(\mathbf{x})$ . In case of point monopole sources  $\varphi_i(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{x}_i^q)$  placed at positions  $\mathbf{x}_i^q$ , each  $h_i(\mathbf{x})$  represents a fundamental solution evaluated for the respective source, so

$$h_i(\mathbf{x}) = G(\mathbf{x}, \mathbf{x}_i^q), \quad (10)$$

where  $G(\mathbf{x}, \mathbf{x}_i^q)$  is a Green's function for operator  $\hat{L}$ . In the remaining work with localized sources we take this approach. As  $G(\mathbf{x}, \mathbf{x}_i^q)$  is usually only available for simple geometries and boundary conditions the discussed linear model alone is limited in its application. We can however represent much more general fields by a superposition of a locally source-free background  $u_h(\mathbf{x})$  and point source contributions  $u_p(\mathbf{x})$ . Boundary conditions induced by external sources are then covered by  $u_h(\mathbf{x})$ , and internal sources entering  $u_p(\mathbf{x})$  are treated via simple free-field Green's functions. Following the technique of [16] discussed in [15] (Chapter 2.7), the superposition  $u(\mathbf{x}) = u_h(\mathbf{x}) + u_p(\mathbf{x})$  of the GP  $u_h(\mathbf{x})$  and the linear model  $u_p(\mathbf{x})$  is distributed according to the Gaussian process

$$u(\mathbf{x}) \sim \mathcal{G}(\mathbf{h}(\mathbf{x})^T \mathbf{q}_0, k(\mathbf{x}, \mathbf{x}') + \mathbf{h}(\mathbf{x})^T \Sigma_q \mathbf{h}(\mathbf{x}')). \quad (11)$$

We will now verify that (11) indeed models the original differential Equation (2) correctly, thereby generalizing the analysis for a deterministic source density in [3]. With  $\hat{L}k(\mathbf{x}, \mathbf{x}')\hat{L}' = 0$ , we obtain

$$\hat{L}u(\mathbf{x}) \sim \mathcal{G}(\hat{L}\mathbf{h}(\mathbf{x})^T \mathbf{q}_0, \hat{L}\mathbf{h}(\mathbf{x})^T \Sigma_q \mathbf{h}(\mathbf{x}')\hat{L}') = \mathcal{G}(\boldsymbol{\varphi}(\mathbf{x})^T \mathbf{q}_0, \boldsymbol{\varphi}(\mathbf{x})^T \Sigma_q \boldsymbol{\varphi}(\mathbf{x})). \quad (12)$$

This is indeed the GP representing the linear source model (8) that we assumed and yields a consistent representation of  $u(\mathbf{x})$  and  $q(\mathbf{x})$  inside (2).

Using the limit of a vague prior with  $\mathbf{q}_0 = 0$  and  $|\Sigma_q^{-1}| \rightarrow 0$ , i.e., minimum information / infinite prior covariance [15,16], posteriors for mean  $\bar{u}$  and covariance matrix  $\text{cov}(u, u)$  based on given training data  $\mathbf{y} = u(X) + \sigma_n$  with measurement noise variance  $\sigma_n^2$  are

$$\bar{u}(X_\star) = K_\star^T K_y^{-1}(\mathbf{y} - H^T \bar{\mathbf{q}}) + H_\star^T \bar{\mathbf{q}} = K_\star^T K_y^{-1} \mathbf{y} + R^T \bar{\mathbf{q}}, \quad (13)$$

$$\text{cov}(u(X_\star), u(X_\star)) = K_{\star\star} - K_\star^T K_y^{-1} K_\star + R^T (H K_y^{-1} H^T)^{-1} R. \quad (14)$$

where  $X = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$  contains the training points and  $X_\star = (\mathbf{x}_{\star 1}, \mathbf{x}_{\star 2}, \dots, \mathbf{x}_{\star N_\star})$  the evaluation or test points. Functions of  $X$  and  $X_\star$  are to be understood as vectors or matrices resulting from evaluation at different positions, i.e.,  $\bar{u}(X_\star) \equiv (\bar{u}(\mathbf{x}_{\star 1}), \bar{u}(\mathbf{x}_{\star 2}), \dots, \bar{u}(\mathbf{x}_{\star N_\star}))$  is a tuple of predicted expectation values. The matrix  $K \equiv k(X, X)$  is the covariance of the training data with entries  $K_{ij} \equiv k(\mathbf{x}_i, \mathbf{x}_j)$ . Entries of the predicted covariance matrix for  $u$  evaluated at the test points  $\mathbf{x}_{\star i}$  are  $\text{cov}(u(X_\star), u(X_\star))_{ij} \equiv \text{cov}(u(\mathbf{x}_{\star i}), u(\mathbf{x}_{\star j}))$ . Furthermore,  $K_y \equiv K + \sigma_n^2 I$ ,  $K_\star \equiv k(X, X_\star)$ ,  $K_{\star\star} \equiv k(X_\star, X_\star)$ ,  $R \equiv H_\star - H K_y^{-1} K_\star$ , and entries of  $H$  are  $H_{ij} \equiv h_i(\mathbf{x}_j)$ ,  $H_{\star ij} \equiv h_i(\mathbf{x}_{\star j})$ . Posterior mean and covariance of source strengths are given from the linear model [16] in the limit of a vague prior,

$$\bar{\mathbf{q}} = (H K_y^{-1} H^T)^{-1} H K_y^{-1} \mathbf{y}, \quad (15)$$

$$\text{cov}(\mathbf{q}, \mathbf{q}) = (H K_y^{-1} H^T)^{-1}. \quad (16)$$

In the absence of sources, the matrix  $R$  vanishes, and (13) and (14) reduce to posteriors of a GP with zero prior mean and are directly used to model homogeneous solutions  $u_h(\mathbf{x})$  of (3).

#### Construction of Kernels for Homogeneous PDEs

For the representation of solutions  $u_h(\mathbf{x})$  of homogeneous differential Equations (3), the weight-space view ([15] Chapter 2.1) of Gaussian process regression is useful. There the kernel  $k$

is represented via a tuple  $\boldsymbol{\phi}(\mathbf{x}) = (\phi_1(\mathbf{x}), \phi_2(\mathbf{x}), \dots)$  of basis functions  $\phi_i(\mathbf{x})$  that underlie a linear regression model

$$u(\mathbf{x}) = \boldsymbol{\phi}(\mathbf{x})^T \mathbf{w} = \sum_i \phi_i(\mathbf{x}) w_i. \quad (17)$$

Bayesian inference starting from a Gaussian prior with covariance matrix  $\Sigma_p$  for weights  $\mathbf{w}$  yields a Mercer kernel

$$k(\mathbf{x}, \mathbf{x}') \equiv \boldsymbol{\phi}^T(\mathbf{x}) \Sigma_p \boldsymbol{\phi}(\mathbf{x}') = \sum_{i,j} \phi_i(\mathbf{x}) \Sigma_p^{ij} \phi_j(\mathbf{x}'). \quad (18)$$

The existence of such a representation is guaranteed by Mercer's theorem in the context of reproducing kernel Hilbert spaces (RKHS) [14]. More generally one can also define kernels on an uncountably infinite number of basis functions in analogy to (17) via

$$f(\mathbf{x}) = (\hat{\phi}w)(\mathbf{x}) = \langle \phi(\mathbf{x}, \boldsymbol{\zeta}), w(\boldsymbol{\zeta}) \rangle = \int \phi(\mathbf{x}, \boldsymbol{\zeta}) w(\boldsymbol{\zeta}) d\boldsymbol{\zeta}, \quad (19)$$

where  $\hat{\phi}$  is a linear operator acting on elements  $w(\boldsymbol{\zeta})$  of an infinite-dimensional weight space parametrized by an auxiliary index variable  $\boldsymbol{\zeta}$ , that may be multidimensional. We represent  $\hat{\phi}$  via an inner product  $\langle \phi(\mathbf{x}, \boldsymbol{\zeta}), w(\boldsymbol{\zeta}) \rangle$  in the respective function space given by an integral over  $\boldsymbol{\zeta}$ . The infinite-dimensional analog to the prior covariance matrix is a prior covariance operator  $\hat{\Sigma}_p$  that defines the kernel as a bilinear form

$$k(\mathbf{x}, \mathbf{x}') \equiv \langle \phi(\mathbf{x}, \boldsymbol{\zeta}), \hat{\Sigma}_p \phi(\mathbf{x}', \boldsymbol{\zeta}') \rangle \equiv \int \phi(\mathbf{x}, \boldsymbol{\zeta}) \Sigma_p(\boldsymbol{\zeta}, \boldsymbol{\zeta}') \phi(\mathbf{x}', \boldsymbol{\zeta}') d\boldsymbol{\zeta} d\boldsymbol{\zeta}'. \quad (20)$$

Kernels of the form (20) are known as convolution kernels. Such a kernel is at least positive semidefinite, and positive definiteness follows in the case of linearly independent basis functions  $\phi(\mathbf{x}, \boldsymbol{\zeta})$  [14].

For treatment of PDEs, the possible choices of index variables in Equation (18) or Equation (20) include separation constants of analytical solutions, or the frequency variable of an integral transform. In accordance with [10], using basis functions that satisfy the underlying PDE, a probabilistic meshless method (PMM) is constructed. In particular, if  $\boldsymbol{\zeta}$  parameterizes positions of sources, and  $\phi(\mathbf{x}, \boldsymbol{\zeta}) = G(\mathbf{x}, \boldsymbol{\zeta})$  in (20) is chosen to be a fundamental solution/Green's function  $G(\mathbf{x}, \boldsymbol{\zeta})$  of the PDE, one may call the resulting scheme a *probabilistic method of fundamental solutions (pMFS)*. In [10], sources are placed across the whole computational domain, and the resulting kernel is called *natural*. Here, we will instead place sources in the exterior to fulfill the homogeneous interior problem, as in the classical MFS [12–14]. Technically, this is also achieved by setting  $\Sigma_p(\boldsymbol{\zeta}, \boldsymbol{\zeta}') = 0$  for either  $\boldsymbol{\zeta}$  or  $\boldsymbol{\zeta}'$  lies in the interior. For discrete sources localized at  $\boldsymbol{\zeta} = \boldsymbol{\zeta}_i$  one obtains again discrete basis functions  $\phi_i(\mathbf{x}) = G(\mathbf{x}, \boldsymbol{\zeta}_i)$  for (18).

### 3. Application Cases

Here, the general results described in the previous sections are applied to specific equations. First, a specialized kernel fulfilling the given linear differential equation is constructed according to (18), and second, numerical experiments on physical examples are performed comparing the specialized kernel to a squared exponential kernel. Regression is performed based on values measured at a set of sampling points  $\mathbf{x}_i$  and may also include optimization of hyperparameters  $\boldsymbol{\theta}$  appearing as auxiliary variables inside the kernel  $k(\mathbf{x}, \mathbf{x}'; \boldsymbol{\theta})$ . The optimization step is, as usually, performed such that the marginal likelihood of the GP is maximized (maximum likelihood or ML values). In the Bayesian sense, this corresponds to a maximum a-posteriori (MAP) estimate for a flat prior. Accordingly,  $\boldsymbol{\theta}_{ML}$  is fixed rather than providing a joint probability distribution function including  $\boldsymbol{\theta}$  as random variables. We note that depending on the setting this choice may lead to underestimation of uncertainties in the reconstruction of  $u(\mathbf{x})$ , in particular for sparse, low-quality measurements.

### 3.1. Laplace's Equation in Two Dimensions

First, we explore construction of kernels fulfilling (5) for a homogeneous problem in a finite and infinite dimensional index space, depending on the mode of separation. Consider Laplace's equation:

$$\Delta u(\mathbf{x}) = 0. \quad (21)$$

In contrast to the Helmholtz equation, Laplace's equation has no scale, i.e., permits all length scales in the solution. In the 2D case using polar coordinates the Laplacian becomes

$$\frac{1}{r} \frac{\partial}{\partial r} \left( r \frac{\partial u(r, \theta)}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2 u(r, \theta)}{\partial \theta^2} = 0. \quad (22)$$

A well-known family of solutions for this problem based on the separation of variables is

$$u(r, \theta) = r^{\pm m} e^{\pm im\theta}, \quad (23)$$

with separation constant  $m$ , leading to real-valued combinations

$$r^m \cos(m\theta), r^m \sin(m\theta), r^{-m} \cos(m\theta), r^{-m} \sin(m\theta). \quad (24)$$

As our aim is to work in bounded regions, we discard the solutions with negative exponent that diverge at  $r = 0$ . Choosing a diagonal prior that weights sine and cosine terms equivalently [9] and introducing a length scale  $\ell$  as a free parameter we obtain a kernel according to (18) with

$$\begin{aligned} k(\mathbf{x}, \mathbf{x}'; \ell, \sigma_m) &= \sum_{m=0}^{\infty} \left( \frac{rr'}{\ell^2} \right)^m \sigma_m^2 (\cos(m\theta) \cos(m\theta') + \sin(m\theta) \sin(m\theta')) \\ &= \sum_{m=0}^{\infty} \left( \frac{rr'}{\ell^2} \right)^m \sigma_m^2 \cos(m(\theta - \theta')). \end{aligned} \quad (25)$$

A flat prior  $\sigma_m^2 = \sigma_u^2$  for all polar harmonics and a characteristic length scale  $\ell$  as another hyperparameter yields

$$k(\mathbf{x}, \mathbf{x}'; \ell, \sigma_u) = \sigma_u^2 \frac{1 - \frac{rr'}{\ell^2} \cos(\theta - \theta')}{1 - 2\frac{rr'}{\ell^2} \cos(\theta - \theta') + \frac{(rr')^2}{\ell^4}} = \sigma_u^2 \frac{1 - \frac{\mathbf{x} \cdot \mathbf{x}'}{\ell^2}}{1 - 2\frac{\mathbf{x} \cdot \mathbf{x}'}{\ell^2} + \frac{|\mathbf{x}|^2 |\mathbf{x}'|^2}{\ell^4}}. \quad (26)$$

This kernel is not stationary, but isotropic around a fixed coordinate origin. Introducing a mirror point  $\bar{\mathbf{x}}'$  with polar angle  $\bar{\theta}' = \theta'$  and radius  $\bar{r}' = \ell^2/r'$  we notice that (26) can be written as

$$k(\mathbf{x}, \mathbf{x}'; \ell, \sigma_u) = \sigma_u^2 \frac{|\bar{\mathbf{x}}'|^2 - \mathbf{x} \cdot \bar{\mathbf{x}}'}{(\mathbf{x} - \bar{\mathbf{x}}')^2}, \quad (27)$$

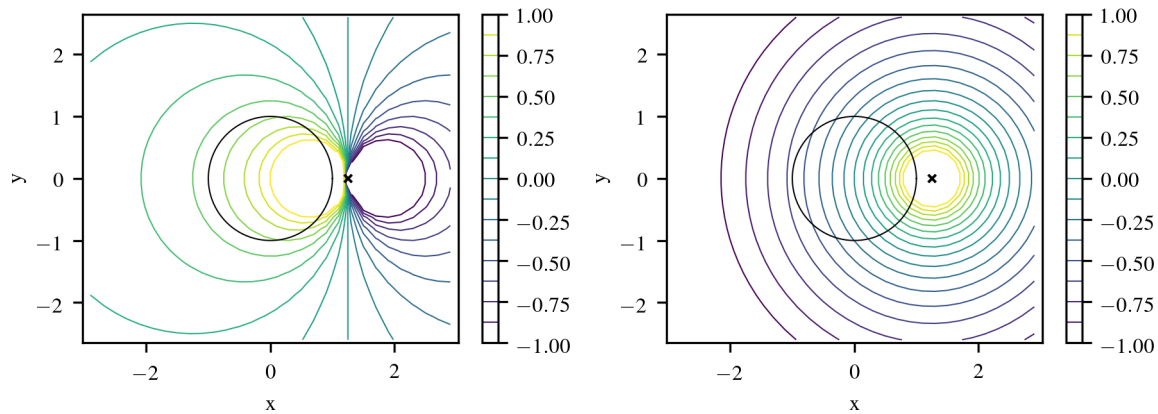
making a dipole singularity apparent at  $\mathbf{x} = \bar{\mathbf{x}}'$ . In addition,  $k$  is normalized to 1 at  $\mathbf{x} = 0$ . Choosing  $\ell > R_0$  larger than the radius  $R_0$  of a circle centered in the origin and enclosing the computational domain, we have  $\bar{r}' > \ell^2/\ell = \ell > R_0$ . Thus, all mirror points and the according singularities are moved outside the domain. This behavior is illustrated in Figure 1 where computing the covariance kernel with respect to point  $\mathbf{x}' = (0.8, 0)$  leads to distances  $> 1$  everywhere inside the unit circle.



Choosing a slowly decaying  $\sigma_m^2 = \sigma_u^2 / m$ , excluding  $m = 1$  and adding a constant term yields a logarithmic kernel instead [9] with

$$\begin{aligned} k(\mathbf{x}, \mathbf{x}'; \ell, \sigma_u) &= \sigma_u^2 \left( 1 - \frac{1}{2} \ln \left( 1 - 2 \frac{\mathbf{x} \cdot \mathbf{x}'}{\ell^2} + \frac{|\mathbf{x}|^2 |\mathbf{x}'|^2}{\ell^4} \right) \right) \\ &= \sigma_u^2 \left( 1 - \ln \left( \frac{|\mathbf{x} - \bar{\mathbf{x}}'|}{|\bar{\mathbf{x}}'|} \right) \right). \end{aligned} \quad (28)$$

Instead of a dipole singularity that expression features a monopole singularity at  $\mathbf{x} - \bar{\mathbf{x}}'$  that is again avoided by placing it outside the domain for any pair of  $\mathbf{x}$  and  $\mathbf{x}'$  (Figure 1).



**Figure 1.** Kernels  $k(\mathbf{x}, \mathbf{x}')$  evaluated at  $\mathbf{x} = (x, y)$  and  $\mathbf{x}' = (0.8, 0)$ . **Left:** dipole response of (27), **right:** monopole response of (28). Singularities are moved outside the domain of interest.

Using instead Cartesian coordinates  $x, y$  to separate the Laplacian provides harmonic functions like

$$u(x, y) = e^{\pm \kappa x} e^{\pm i \kappa y}. \quad (29)$$

Here, all solutions yield finite values at  $x = 0$ , so we do not have to exclude any of them a priori. Introducing, again, a diagonal covariance operator in (20) and taking the real part yields

$$k(\mathbf{x}, \mathbf{x}'; \sigma^2(\kappa)) = \int \varphi(\mathbf{x}, \kappa) \sigma^2(\kappa) \varphi(\mathbf{x}', \kappa) d\kappa = \text{Re} \int_{-\infty}^{\infty} \sigma^2(\kappa) e^{\kappa(x \pm x')} e^{i\kappa(y \pm y')} d\kappa. \quad (30)$$

Setting  $\sigma^2(\kappa) \equiv e^{-2\kappa^2}$  and choosing a characteristic length scale  $\ell$  together with a possible rotation angle  $\theta_0$  of the coordinate frame yields the kernel

$$k(\mathbf{x}, \mathbf{x}'; \ell, \theta_0, \sigma_u) = \frac{\sigma_u^2}{2} \text{Re} \exp \left( \frac{((x + x') \pm i(y - y'))^2 e^{i2\theta_0}}{\ell^2} \right). \quad (31)$$

Other sign combinations do not yield a positive definite kernel – similar to the polar kernel (27) before we couldn't obtain an fully stationary expression that depends only on differences between coordinates of  $\mathbf{x}$  and  $\mathbf{x}'$ .

For demonstration purposes we consider an analytical solution to a boundary value problem of Laplace's equation on a square domain  $\Omega$  with corners at  $(x, y) = (\pm 1, \pm 1)$ . The reference solution is

$$u_{\text{ref}}(x, y) = \frac{1}{2} \left( e^y \cos(x) + e^{2x} \cos(2y) \right) \quad (32)$$

and depicted in the upper left of Figure 2 together with the extension outside the boundaries. This figure also shows results from a GP fitted based on data with artificial noise of  $\sigma_n = 0.1$  measured at 8 points using kernel (27) with optimized maximum-likelihood (ML) values for hyperparameters

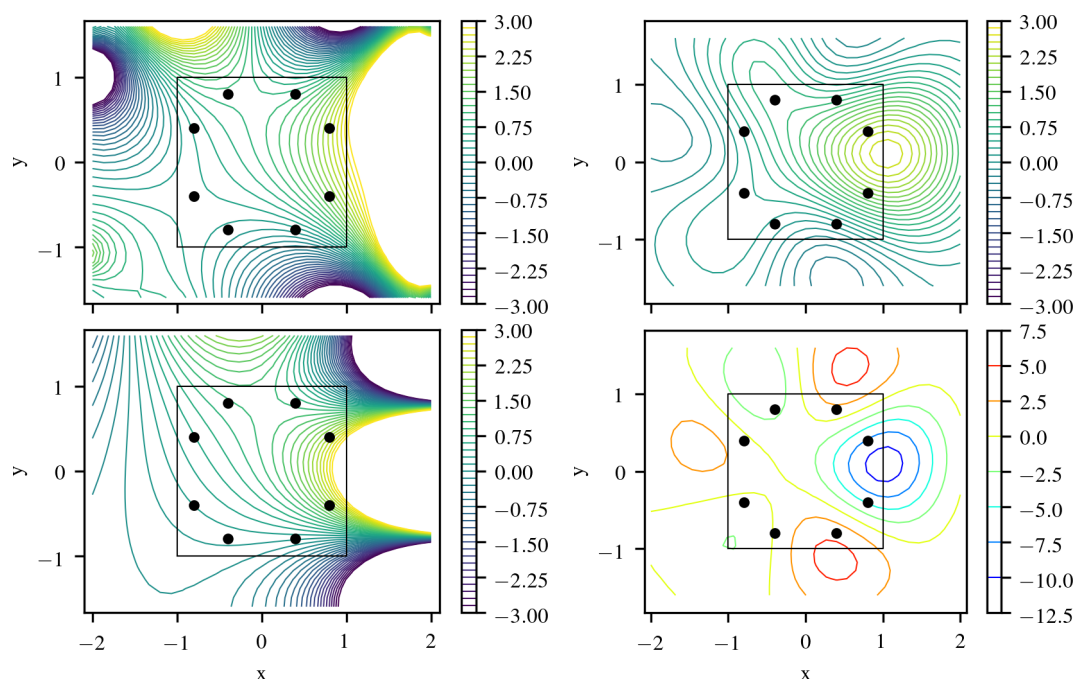
$\ell$  and  $\sigma_u$  but fixed  $\sigma_n$ . Inside  $\Omega$  the solution is represented with errors below 5%. This is also reflected in the error predicted by the posterior variance of the GP that remains small in the region enclosed by measurement points. The analogy in classical analysis is the theorem that the solution of a homogeneous elliptic equation is fully determined by boundary values.

In comparison, a reconstruction using a generic squared exponential kernel

$$k(\mathbf{x}, \mathbf{x}'; \ell, \sigma_u) = \sigma_u^2 \exp\left(\frac{-(\mathbf{x} - \mathbf{x}')^2}{2\ell^2}\right) \quad (33)$$

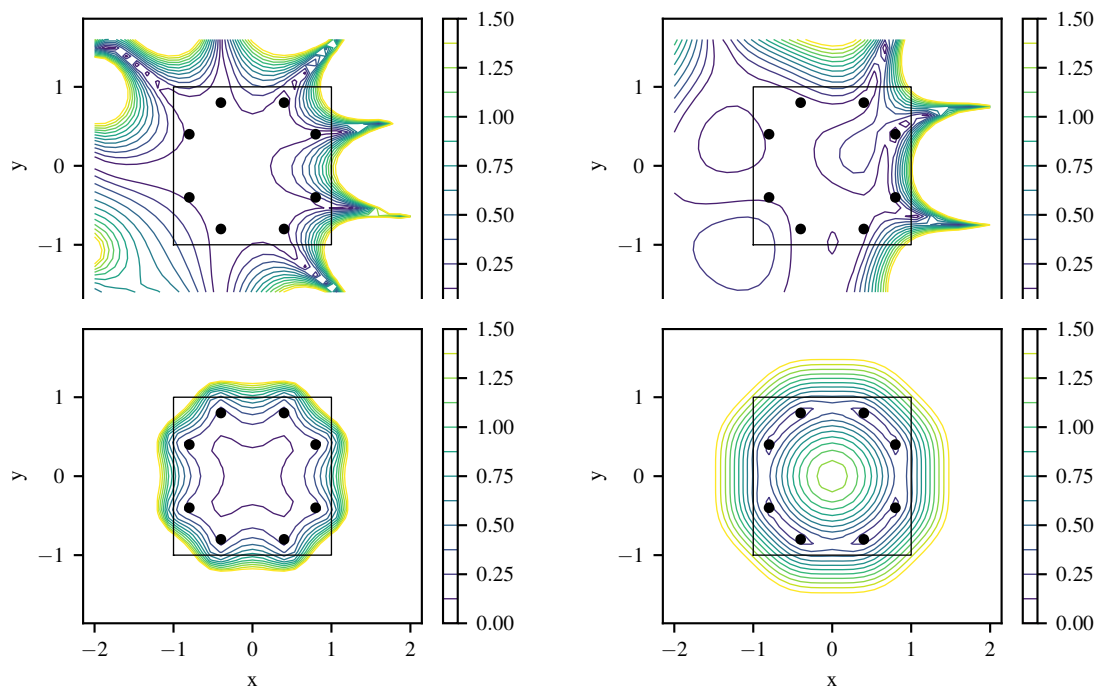
yields a much worse approximation quality in Figures 2 and 3. This is in contrast to earlier investigations [1] where a fixed length scale hyperparameter  $\ell = 2$  was used. Although the specialized GP with kernel (27) could identify this length scale during hyperparameter optimization, using a generic kernel (33) leads to an underestimation of  $\ell$  and requires twice the number of training points to achieve a similar fit quality and profits from scattered training points, as it has no information about the nature of the boundary value problem (Figures 4 and 5).

In addition, the posterior covariance of that reconstruction is not able to capture the vanishing error inside the enclosed domain due to given boundary data. More severely, in contrast to the specialized GP, the posterior mean  $\bar{u}$  does not satisfy Laplace's equation  $\Delta \bar{u} = 0$  exactly. This leads to a violation of the classical result that (differences of) solutions of Laplace's equation may not have extrema inside  $\Omega$ , showing up in the difference to the reconstruction in Figures 3 and 4. This kind of error is quantified by computation of the reconstructed charge density  $\bar{q} = \Delta \bar{u}$ . This is fine if data from Poisson's equation  $\Delta u = q$  with distributed charges should be fitted instead. However, to keep  $\Delta u = 0$  exact in  $\Omega$ , one requires more specialized kernels such as (27).

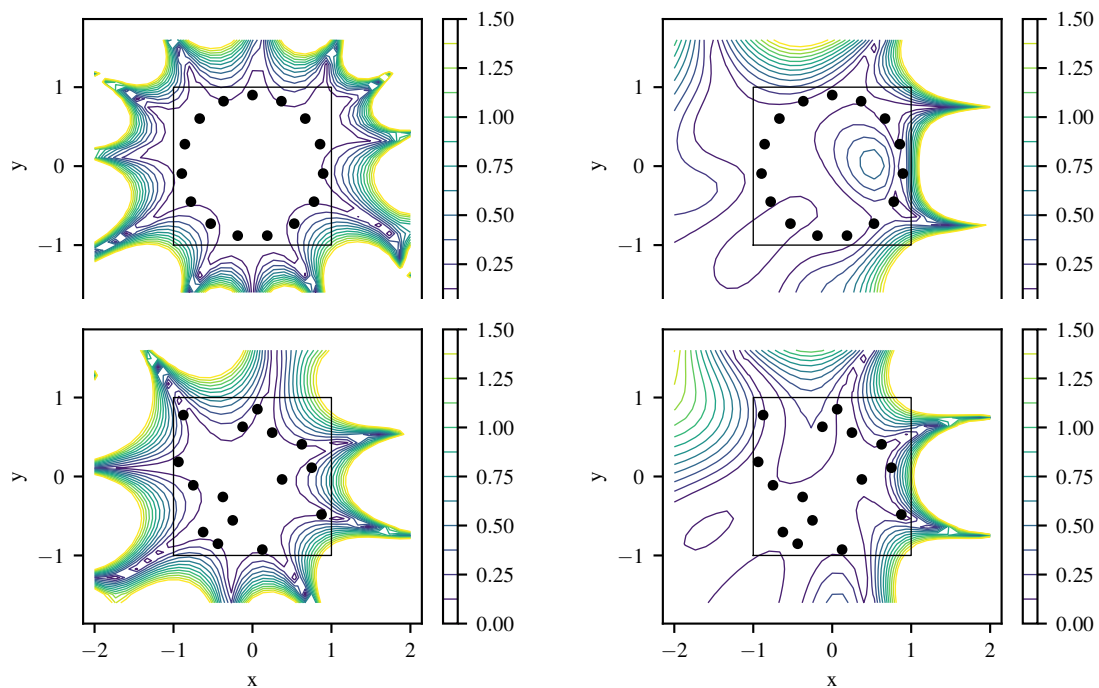


**Figure 2.** GP reconstruction of Laplace's equation with specialized locally source-free Mercer kernel (27) (top left) and generic squared exponential kernel (top right). Sources lie outside the black square region and 8 measurement positions are marked by black dots. Reference analytical solution (bottom left). Source density  $\bar{q} = \Delta \bar{u}$  of prediction via a generic squared exponential kernel (bottom right).

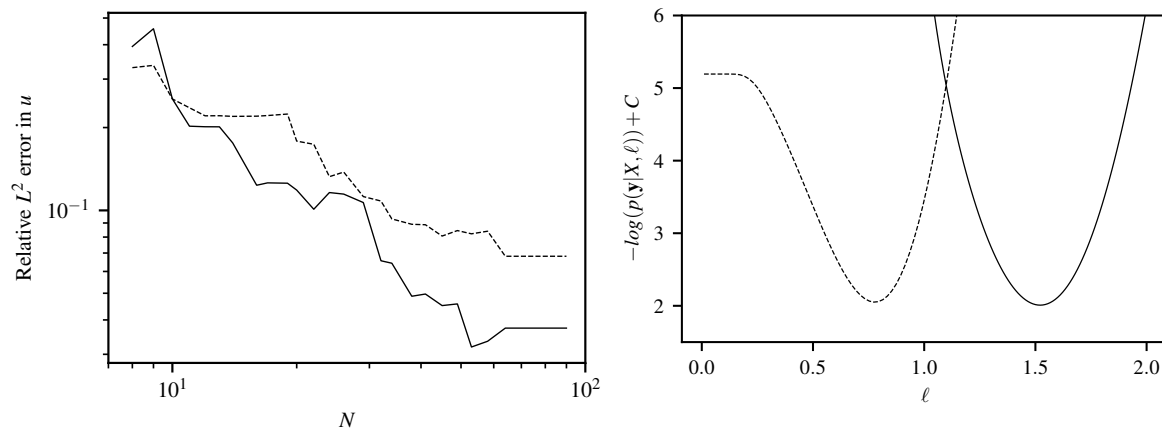




**Figure 3.** Absolute error (**top left**) and predicted 95% confidence interval (**bottom left**) with specialized locally source-free Mercer kernel (27) in comparison to absolute error (**top right**) and predicted 95% confidence interval (**bottom right**) with generic squared exponential kernel for 8 training points.



**Figure 4.** Absolute error as in Figure 3 for 15 training points on a circle (**top**) and for quasi-random points (**bottom**). As the generic squared exponential kernel does not fulfill the given differential equation, even for a larger number of training points, the source density doesn't vanish in the domain.



**Figure 5.** (Left) Comparison of relative  $L^2$  error in  $u$  for specialized kernel (solid line) and squared exponential kernel (dashed line) for Laplace's equation for  $N$  quasi-random training points. (Right) Negative log likelihood from 8 training data of Figure 2 with optimum at  $\ell = 1.52$  for specialized kernel (solid line) and at  $\ell = 0.78$  for the squared exponential kernel (dashed line).

### 3.2. Heat Equation: Physical Parameter Estimation

Let us now consider the 1D homogeneous heat/diffusion equation over position  $x$  and time  $t$ ,

$$\frac{\partial u(x, t)}{\partial t} - D\Delta u(x, t) = 0 \quad (34)$$

for  $(x, t) \in \mathbb{R} \times \mathbb{R}^+$ . Here, the diffusivity  $D$  is a physical parameter determining how fast solutions spread in space. Integrating the fundamental solution

$$G(x, t, \xi, \tau) = \frac{1}{\sqrt{4\pi D(t - \tau)}} \exp\left(-\frac{(x - \xi)^2}{4D(t - \tau)}\right) \quad (35)$$

from  $\xi = -\infty$  to  $\infty$  at  $\tau = 0$ , i.e., placing sources everywhere in space at a single initial time, and adding a scale hyperparameter  $\sigma_u$  leads to the convolution kernel

$$k_n(x, t, x', t'; D, \sigma_u) = \frac{\sigma_u^2}{\sqrt{4\pi D(t + t')}} \exp\left(-\frac{(x - x')^2}{4D(t + t')}\right). \quad (36)$$

In terms of  $x$ , this is a stationary squared exponential kernel and the natural kernel over the domain  $x \in \mathbb{R}$ . The kernel broadens with increasing  $t$  and  $t'$ . Nonstationarity in time can also be considered natural to the heat equation, as its solutions show a preferred time direction on each side of the singularity  $t = 0$ . The only difference of (36) to the fundamental solution (35) is the positive sign between  $t$  and  $t'$ . As both  $t$  and  $t'$  are positive,  $k$  is guaranteed to take finite values and, in contrast to (35), does not become singular at  $(x, t) = (x', t')$ .

As for the Laplace equation it is also convenient to define a non-stationary kernel by cutting out a domain that is known to be free of sources. In case heat sources are known to exist only left of the origin we evaluate the integral over the fundamental solution over  $(-\infty, 0)$  to

$$k(x, t, x', t'; D, \sigma_u) = k_n(x, t, x', t'; D, \sigma_u) \left[1 + \frac{g(x, t, x', t'; D)}{2}\right], \quad (37)$$

where

$$g(x, t, x', t'; D) \equiv \operatorname{erf}\left(-\frac{x/t + x'/t'}{2\sqrt{D}\sqrt{1/t + 1/t'}}\right) \quad (38)$$

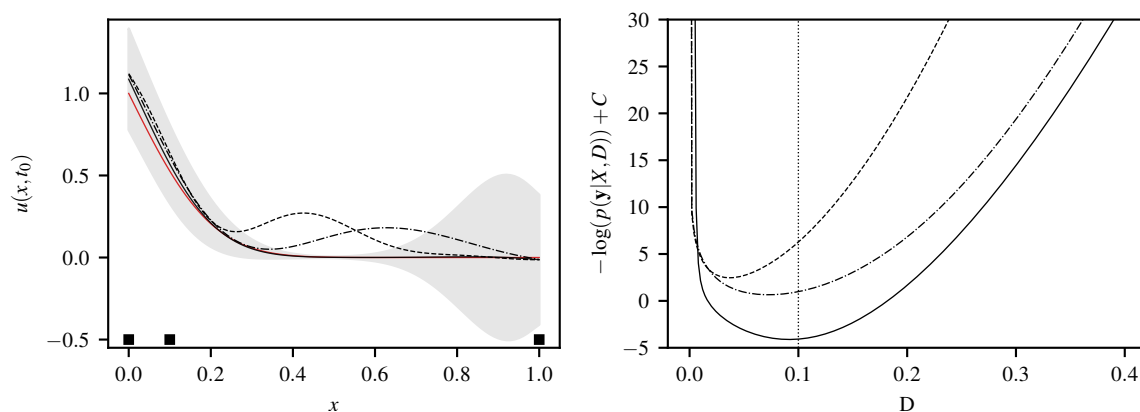
is defined via the error function erf. Choosing instead a source-free region domain interval  $(a, b)$  we integrate over  $\mathbb{R} \setminus (a, b)$  and obtain

$$k(x, t, x', t'; D, \sigma_u) = k_n(x, t, x', t'; D, \sigma_u) \left[ 1 - \frac{g(x - b, t, x' - b, t'; D) - g(x - a, t, x' - a, t'; D)}{2} \right]. \quad (39)$$

Incorporating the prior knowledge that there are no domain sources is expected to improve the reconstruction.

As a physical example, we consider a rod with temperatures held fixed at two ends and a given initial temperature distribution. We model this as an initial-boundary value problem for (34) on the interval  $x \in (0, 1)$  with Dirichlet boundary data  $u(0) = 1$  and  $u(1) = 0$ . As initial conditions, we set  $u(x, 0) = 0$  everywhere except at the left end where  $u(0, 0) = 1$ . The actual diffusivity is chosen as  $D = 0.1$ , and we let  $u(x, t)$  evolve from  $t_0 = 0$  until  $t_1 = 1$ . With increasing  $t$  the initial conditions are smoothed out as  $u$  approaches the stationary solution  $u(x, t \rightarrow \infty) = 1 - x$ . Measurements of  $u$  are performed at three positions  $x = 0, 0.1, 1$  at four times  $t = 10^{-5}, 0.25, 0.5, 0.75$ , yielding 12 training points in total. In Figure 6 the resulting reconstruction of  $u(x, t = 0.125)$  is plotted for each of the three kernels defined above. Kernel (39) allowing initial sources on both sides of the interval yields the best reconstruction. Furthermore, it is the only one that reproduces meaningful uncertainty bands based on the 95% confidence interval  $\bar{u} \pm 1.96\sigma$ , whereas the ones for (36) and (36) span the whole plot domain. Estimation of diffusivity  $D$  is also most reliable with kernel (39). The according negative log likelihood can be seen on the right plot in Figure 6. Although all three kernels produce well posed optimization problems, only (39) has the minimum at the correct position  $D = 0.1$ .

The reason for the requirement of kernel (39) is clear from the statement of the problem: keeping  $u$  fixed on both sides of the interval can only be achieved by restricting the heat flux in a predefined way that requires sources on both sides at  $t = 0$ . However, the domain itself should not contain any heat sources at any time. If we had placed an open boundary condition on the right side, kernel (37) would have been the more natural choice instead.



**Figure 6.** (Left) GP reconstruction of  $u(x, t = 0.125)$  for 1D heat equation Dirichlet problem based on measurement points (■) at  $x = 0, 0.1, 1$ , reference in red. Kernels (36), (37) and (39) marked by dashed, dash-dotted and solid lines, respectively. 95% confidence interval bands shown only for (39), producing the best fit. (Right) negative log likelihood over diffusivity  $D$ .

### 3.3. Helmholtz Equation: Source and Wavenumber Reconstruction

Finally, to demonstrate the full method, we consider the Helmholtz equation with sources:

$$\Delta u(\mathbf{x}) + k_0^2 u(\mathbf{x}) = q(\mathbf{x}). \quad (40)$$

In 1D, solutions for the homogeneous equation with  $\mathbf{x} = x$  are given by linear combinations of  $\cos(k_0 x)$ ,  $\sin(k_0 x)$ . Choosing a diagonal prior in (18) leads to a stationary kernel

$$k(x, x'; k_0, \sigma_u) = \cos(k_0 x) \sigma_u \cos(k_0 x') + \sin(k_0 x) \sigma_u \sin(k_0 x') = \sigma_u \cos(k_0(x - x')), \quad (41)$$

as presented in [11]. For the two-dimensional case in polar coordinates, a family of solutions based on the separation of variables is

$$\cos(m\theta), \sin(m\theta), J_m(k_0 r), Y_m(k_0 r), \quad (42)$$

where  $J_m$  and  $Y_m$  are Bessel functions of first and second kind, respectively. Similar to the simpler 1D case, by applying Neumann's addition theorem, we obtain a specialized kernel

$$k(\mathbf{x}, \mathbf{x}'; k_0, \sigma_u) = \sigma_u^2 J_0(k_0 |\mathbf{x} - \mathbf{x}'|). \quad (43)$$

In the 3D case, one would proceed in a similar fashion with spherical Bessel functions, which yields the kernel that was already postulated in [11]. In contrast to the case of Laplace's equation in the previous section, these source-free Helmholtz kernels do not possess singularities at any finite distance from the origin, i.e., no virtual exterior sources in the Mercer kernel (20). As a consequence they provide smoothing regularization on the order of the wavelength  $\lambda_0 = 2\pi/k_0$  to reconstructed fields and boundary conditions that may or may not be desired. Internal sources at positions  $\mathbf{x}_k^q$  are linearly modeled according to (10) with basis

$$h_i(\mathbf{x}) = G(\mathbf{x}, \mathbf{x}_i^q) = H_0^{(2)}(k_0 |\mathbf{x} - \mathbf{x}_i^q|), \quad (44)$$

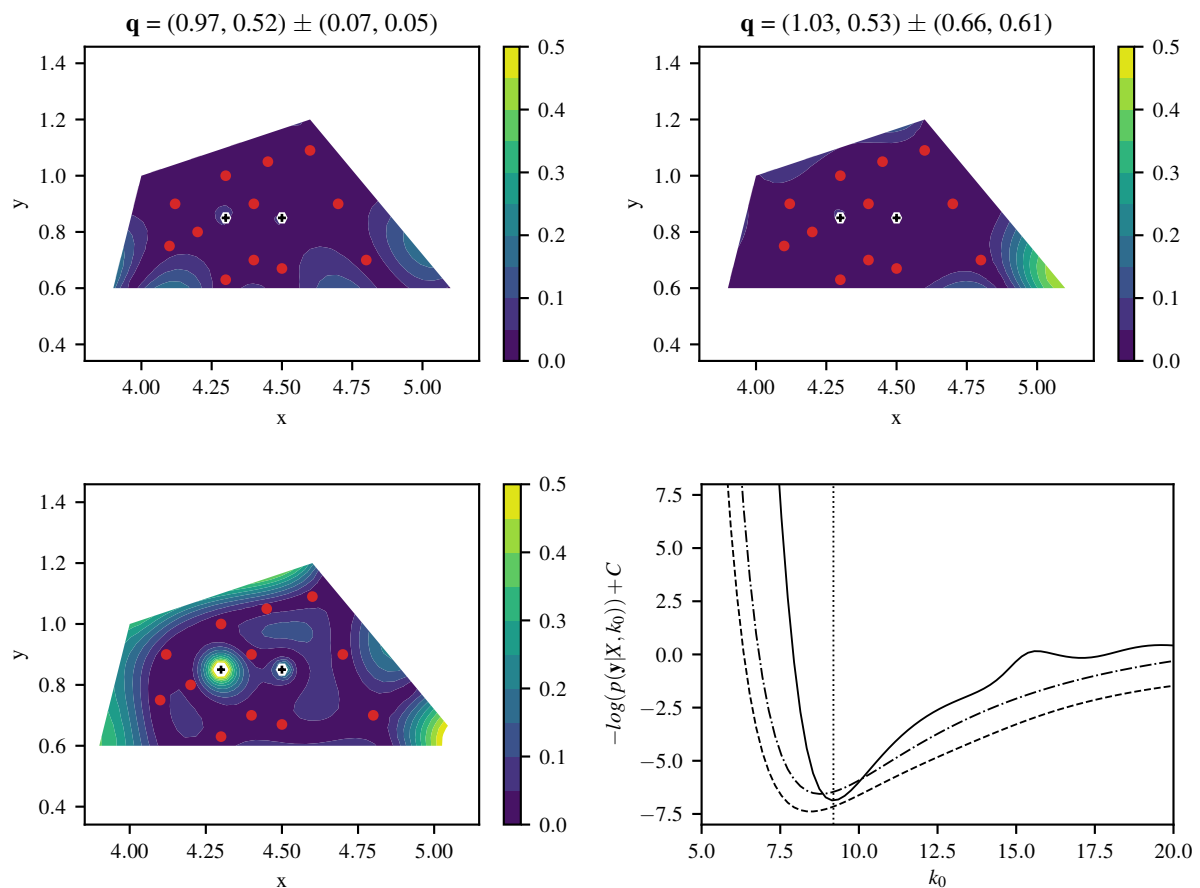
where  $H_0^{(2)}$  is the Hankel function of the second kind. The method of source strength reconstruction is improved compared to [11], as it constitutes a linear problem according to (15). Nonlinear optimization is instead applied to  $\sigma_u$  and wavenumber  $k_0$  as free hyperparameters to be estimated during the GP regression. The set-up is the same as in [11]: a 2D cavity with various boundary conditions and two sound sources of strengths 0.5 and 1.0, respectively. Results for sound pressure fulfilling (40) are normalized to have a maximum  $p/p_0 = 1$ . We compare three variants of GP regression for these data:

- (1) Superposition of specialized kernel GP for homogeneous part  $u_h$  and linear source model for  $u_p$ .
- (2) Superposition of generic squared-exponential kernel GP for  $u_h$  and linear source model for  $u_p$ .
- (3) Generic squared-exponential kernel GP model for the full field  $u$ .

Naturally, after the presented analysis, only (1) can be the "correct" way of regression for this kind of data from a PDE with point sources. Variant (2) is a "hybrid" that should be able to identify point sources, while polluting the source-free part with volumetric contributions. Considering that (2) helps to separate the effect from this pollution from the effect of adding a linear source model. Variant (3) is expected to show worse performance compared to (1) and (2), as neither source-free part nor singularities of  $u$  at point source positions can be modeled correctly.

Figure 7 shows the local absolute field reconstruction error based on 12 training data points with artificial noise of  $\sigma_n = 0.01$ . Hyperparameters  $k_0$  and  $\sigma_u$  are set to optimized ML values, and  $\sigma_n$  is fixed to its actual value. The upper left plot shows results for variant (1) with the specialized kernel (43). Variant (3) with a generic squared exponential kernel (33) of length scale  $\ell = \pi/(\sqrt{2}k_0)$  to model  $u$  yields a much higher field reconstruction error as depicted in the lower left of Figure 7. The field reconstruction using the generic kernel is improved when a linear model for the inhomogeneous term

is included (variant (2)), as shown in the upper right of Figure 7. However, the original differential Equation (40) is only fulfilled exactly when using a specialized kernel with  $\hat{L}k(\mathbf{x}, \mathbf{x}')\hat{L}' = 0$ . As expected, variant (1) produces the best reconstruction at a given number of training points (Figure 8). There the first 12 points are chosen as marked in Figure 7, and more points are generated from a quasi-random Halton sequence. The obtained negative log-likelihood (Figure 7, lower right) depending on  $k_0$  and  $\sigma_u$  at its ML value demonstrates the well-posedness of estimating  $k_0$  having the physical meaning of a wavenumber. Variants (2) and (3) lead to a slightly less peaked estimate for a spatial length scale hyperparameter without a direct physical interpretation.

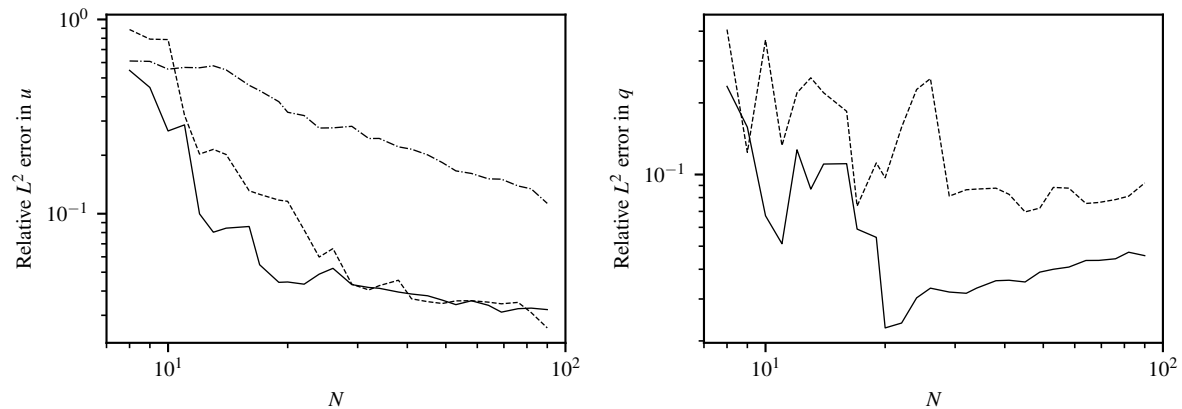


**Figure 7.** Reconstruction error for the Helmholtz equation from 12 training points for specialized kernel (**top left**), squared exponential kernel with linear source model (**top right**) and squared exponential kernel (**bottom right**); reconstructed source strengths  $\mathbf{q}$  with 95% confidence interval via posterior (15) and (16). Negative log likelihood (**bottom right**) with optimum  $k_0^{\text{ML}} = 9.19$  for specialized kernel (solid line), sq.exp. kernel with linear source model (dashed), and sq.exp. kernel alone (dash-dotted).

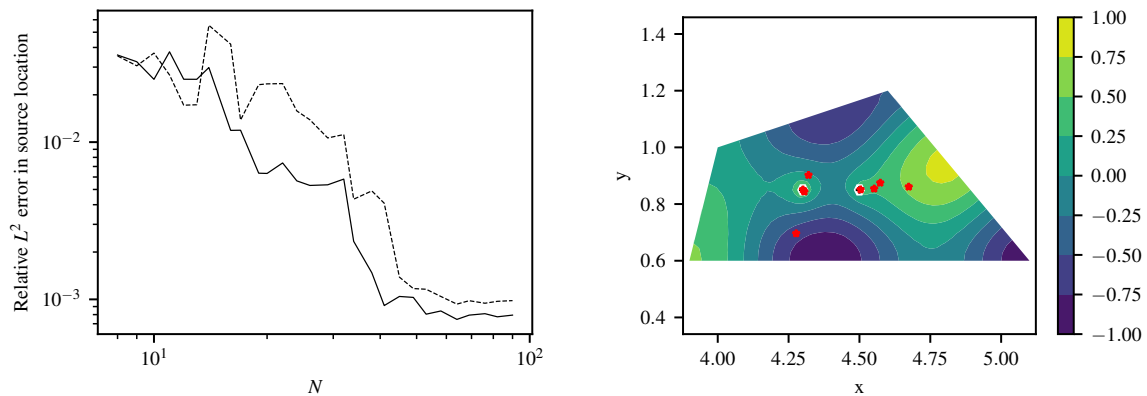
For estimation of source positions, nonlinear optimization is applied to source positions as free hyperparameters within the given boundaries, employing an evolutionary algorithm CMA-ES [17]. The results of source strength and position estimation using (15) and (16) in the configuration with 12 training points is given in Table 1. Both estimates match the exact values reasonably well. At an increasing number of training data the reconstruction becomes more accurate, stagnating at an error between 0.1% and 1% and showing the advantage of the specialized kernel more clearly (Figures 8 and 9). The relative  $L^2$  error in source positions for specialized and generic squared exponential kernel with linear source model is depicted in the left plot of Figure 9. Again, results from the specialized kernel are usually more accurate and stable compared to using a squared-exponential kernel for the source-free part of the field at a given number of training points.

**Table 1.** Comparison and results for estimation of source strength  $\mathbf{q}$  and source position  $\mathbf{x}_i^q$  for 12 training data points for specialized and squared exponential kernel with linear source model.

Exact Values	Specialized Kernel	sq. exp. Kernel
$\mathbf{q} = (1.0, 0.5)$	$\mathbf{q} = (0.97, 0.52)$	$\mathbf{q} = (1.03, 0.53)$
$\mathbf{x}_1^q = (4.3, 0.85)$	$\mathbf{x}_1^q = (4.31, 0.85)$	$\mathbf{x}_1^q = (4.30, 0.82)$
$\mathbf{x}_2^q = (4.5, 0.85)$	$\mathbf{x}_2^q = (4.65, 0.90)$	$\mathbf{x}_2^q = (4.61, 0.84)$



**Figure 8.** Comparison of relative  $L^2$  error in  $u$  (left) and  $q$  (right) for specialized kernel (solid line), squared exponential kernel (dash-dotted) and squared exponential kernel with linear source model (dashed) for Helmholtz equation with  $N$  quasi-random training points. As the squared exponential kernel alone (without linear source model) cannot reproduce point sources, no result is shown for the point source strength estimation in the right plot for this case.



**Figure 9.** (Left) Comparison of relative  $L^2$  error in source position for specialized kernel (solid line) and squared exponential kernel with linear source model (dashed) for Helmholtz equation with  $N$  quasi-random training points. (Right) reconstructed field using specialized kernel (43) and showing convergence of estimated source location for  $N = (12, 15, 20, 30)$  quasi-random training points.

#### 4. Summary and Outlook

A framework for application of Gaussian process regression to data from underlying linear partial differential equations with localized sources has been presented. The method is based on superposition of a Gaussian process that generates exact solutions of the homogeneous equation, complemented by a linear model for sources. For the homogeneous part, specialized kernels are constructed from fundamental solutions via Mercer's theorem. For source contributions, fundamental solutions are used as basis functions in the linear model. Examples for suitable kernels have been given for Laplace's equation, heat equation and Helmholtz equation. Regression has been shown to yield better results compared to using a squared exponential kernel at the same number of training



points in the considered application cases. Advantages of the specialized kernel approach are the possibility to represent exact absence of sources as well as physical interpretability of hyperparameters. This comes at the cost of requiring non-standard, possibly nonstationary kernels. The presented method has been demonstrated to be able to accurately estimate system parameters such as diffusivity and wavenumber, as well as position and strength of point sources using only around 10 training data points in two-dimensional domains.

In a next step, reconstruction of vector fields via GPs could be formulated, taking laws such as Maxwell's equations or Hamilton's equations of motion into account. A starting point could be squared exponential kernels for divergence- and curl-free vector fields [18]. Such kernels have been used in [19] to perform statistical reconstruction, and [20] apply them to GPs for source identification in the Laplace/Poisson equation. To model Hamiltonian dynamics in phase-space, vector-valued GPs could possibly be extended to represent not only volume-preserving (divergence-free) maps but retain full symplectic properties, conserving all integrals of motion such as energy or momentum.

**Author Contributions:** Conceptualization, C.G.A. and K.R.; investigation, C.G.A. and K.R.; methodology, C.G.A. and K.R.; software, C.G.A. and K.R.; supervision, C.G.A.; visualization, C.G.A. and K.R.; writing—original draft, C.G.A. and K.R.; writing—review and editing, C.G.A. and K.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study is a contribution to the Reduced Complexity Models grant number ZT-I-0010 funded by the Helmholtz Association of German Research Centers and received funding from the Munich School of Data Science (MUDS).

**Acknowledgments:** The authors would like to thank Dirk Nille, Roland Preuss, and Udo von Toussaint for insightful discussions.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Albert, C.G. Gaussian Processes for Data Fulfilling Linear Differential Equations. In Proceedings of the 39th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Munich, Germany, 30 June–5 July 2019. [\[CrossRef\]](#)
2. Dong, A. Kriging Variables that Satisfy the Partial Differential Equation  $\Delta Z = Y$ . *Geostatistics* **1989**, *4*, 237–248. [\[CrossRef\]](#)
3. van den Boogaart, K.G. Kriging for Processes Solving Partial Differential Equations. In Proceedings of the IAMG2001, Cancun, Mexico, 10–12 September 2001; pp. 1–21.
4. Graepel, T. Solving Noisy Linear Operator Equations by Gaussian Processes: Application to Ordinary and Partial Differential Equations. In Proceedings of the Twentieth International Conference on Machine Learning, Washington, DC, USA, 21–24 August 2003; pp. 234–241.
5. Särkkä, S. Linear Operators and Stochastic Partial Differential Equations in Gaussian Process Regression. In Proceedings of the 21st International Conference on Artificial Neural Networks, Espoo, Finland, 14–17 June 2011; pp. 151–158. [\[CrossRef\]](#)
6. Raissi, M.; Perdikaris, P.; Karniadakis, G.E. Inferring Solutions of Differential Equations Using Noisy Multi-Fidelity Data. *J. Comput. Phys.* **2017**, *335*, 736–746. [\[CrossRef\]](#)
7. Raissi, M.; Perdikaris, P.; Karniadakis, G.E. Machine Learning of Linear Differential Equations Using Gaussian Processes. *J. Comput. Phys.* **2017**, *348*, 683–693. [\[CrossRef\]](#)
8. Yang, X.; Tartakovsky, G.; Tartakovsky, A. Physics-Informed Kriging: A Physics-Informed Gaussian Process Regression Method for Data-Model Convergence. *arXiv* **2018**, arXiv:1809.03461.
9. Mendes, F.M.; da Costa Junior, E.A. Bayesian Inference in the Numerical Solution of Laplace's Equation. *AIP Conf. Proc.* **2012**, *1443*, 72–79. [\[CrossRef\]](#)
10. Cockayne, J.; Oates, C.; Sullivan, T.; Girolami, M. Probabilistic Numerical Methods for Partial Differential Equations and Bayesian Inverse Problems. *arXiv* **2016**, arXiv:1605.07811.
11. Albert, C. Physics-Informed Transfer Path Analysis With Parameter Estimation Using Gaussian Processes. In Proceedings of the 23rd International Congress on Acoustics, Aachen, Germany, 9–13 September 2019; pp. 459–466. [\[CrossRef\]](#)

12. Lackner, K. Computation of Ideal MHD Equilibria. *Comput. Phys. Commun.* **1976**, *12*, 33–44. [[CrossRef](#)]
13. Golberg, M.A. The Method of Fundamental Solutions for Poisson's Equation. *Eng. Anal. Bound. Elem.* **1995**, *16*, 205–213. [[CrossRef](#)]
14. Schaback, R.; Wendland, H. Kernel Techniques: From Machine Learning to Meshless Methods. *Acta Numer.* **2006**, *15*, 543–639. [[CrossRef](#)]
15. Rasmussen, C.E.; Williams, C.K.I. *Gaussian Processes for Machine Learning*; MIT Press: Cambridge, MA, USA, 2006.
16. O'Hagan, A. Curve Fitting and Optimal Design for Prediction. *J. R. Stat. Soc. Ser. B* **1978**, *40*, 1–24. [[CrossRef](#)]
17. Hansen, N.; Akimoto, Y.; Baudis, P. CMA-ES/Pycma on Github. Available online: <https://doi.org/10.5281/zenodo.2559634> (accessed on 13 December 2019).
18. Narcowich, F.J.; Ward, J.D. Generalized Hermite Interpolation via Matrix-Valued Conditionally Positive Definite Functions. *Math. Comput.* **1994**, *63*, 661. [[CrossRef](#)]
19. Macêdo, I.; Castro, R. Learning Divergence-Free and Curl-Free Vector Fields with Matrix-Valued Kernels. Available online: [http://preprint.impa.br/FullText/Macedo\\_Thu\\_Oct\\_21\\_16\\_38\\_10\\_BRDT\\_2010/macedo-MVRBFs.pdf](http://preprint.impa.br/FullText/Macedo_Thu_Oct_21_16_38_10_BRDT_2010/macedo-MVRBFs.pdf) (accessed on 13 December 2019).
20. Cobb, A.D.; Everett, R.; Markham, A.; Roberts, S.J. Identifying Sources and Sinks in the Presence of Multiple Agents with Gaussian Process Vector Calculus. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '18), London, UK, 19–23 August 2018; pp. 1254–1262.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).