Wave equation-tailored Gaussian process regression with applications to related inverse problems

Iain Henderson, Pascal Noble, Olivier Roustant

▶ To cite this version:

Iain Henderson, Pascal Noble, Olivier Roustant. Wave equation-tailored Gaussian process regression with applications to related inverse problems. 2023. hal-03941939v1

HAL Id: hal-03941939

https://hal.science/hal-03941939v1

Preprint submitted on 16 Jan 2023 (v1), last revised 7 Nov 2023 (v2)

Wave equation-tailored Gaussian process regression with applications to related inverse problems

I. Henderson *[1], P. Noble[1], and O. Roustant[1]

[1]Institut de Mathématiques de Toulouse; UMR 5219, Université de Toulouse; CNRS , INSA, F-31077 Toulouse, France

Abstract

In this article, we consider the general task of performing Gaussian process regression (GPR) on pointwise observations of solutions of the 3 dimensional free space wave equation. In a recent article, we obtained promising covariance expressions for this equation: we now explore the potential applications of these formulas. We first study the particular cases of stationarity, radial symmetry and point source, for which significant simplifications arise. We next show that the true-angle multilateration method for point source localization, as used in GPS systems, is naturally recovered by our GPR formulas in the limit of the small source radius. Additionally, we show that this GPR framework provides a new answer to the ill-posed inverse problem of reconstructing initial conditions for the wave equation from a limited number of sensors, and simultaneously enables the inference of physical parameters from these data. We finish by illustrating this "physics informed" GPR on a number of practical examples.

1 Introduction

Machine learning techniques have proved time and again that they can provide efficient solutions to difficult problems in the presence of field data. A key element to this success is the incorporation of "expert knowledge" in the corresponding statistical models. In many practical applications, this knowledge takes the form of mathematical models which are sometimes already well understood. This is e.g. common when dealing with problems arising from physics, in which case the mathematical models often take the form of Partial Differential Equations (PDEs), such as the wave equation at hand in this article. Because of the broadness of the applications PDEs offer, large efforts have been devoted to studying and solving them, both theoretically [16] and numerically [23]. These equations impose very specific (yet often simple) structures on the observed data which can be very difficult to capture or mimic with general machine learning models.

In this article, we will focus on the linear 3 dimensional homogeneous free space wave equation. This equation is the prototype for describing simple 3D phenomena which propagate at finite speed; although particularly simple in the landscape of PDEs, it is

---

*henderso@insa-toulouse.fr

in fact central for many applications emerging from different fields such as acoustics or electromagnetics. The homogeneity assumption is also commonly encountered in physics, when modelling conservation laws. Given that the main structures of the solutions of this PDE are well known, one may thus attempt to incorporate them in the machine learning models that work with such solutions.

The class of models we will deal with is that of Gaussian Process Regression (GPR), which is a Bayesian framework for function regression and interpolation [45]. It is a "kernel method", meaning that it is built upon a positive definite function, the kernel in question. In the language of Bayesian inference, GPR puts a *prior* probability distribution on a suitable function space in which the unknown function u is assumed to lie. This prior is then conditioned on available field data involving u thanks to Bayes' law, which in turn provides a *posterior* probability distribution from which statistical estimators related to u can be computed. The posterior expectation in particular plays the role of an approximant of u while the posterior covariance provides posterior error bounds. In the case of GPR, these prior and posterior distributions are in many ways generalizations to infinite dimensions of the multivariate normal distribution, and are fully specified by a mean and covariance functions. These priors are naturally obtained by modelling u as a *sample path* of a *Gaussian process* and we will thus say that we put a Gaussian process (GP) prior over u. Imposing strict *linear* constraints on a GP prior as well as on the posterior expectation it provides is straightforward in principle; we will apply this observation to the case where the linear constraint is the homogeneous wave equation itself, as in [25].

In this article, we will first be concerned with building GP priors which incorporate *beforehand* the knowledge that the sought function is in fact a solution to the same wave equation. The main consequence will be that all the possible estimators of u provided by GPR will also be solutions to the wave equation. Those priors are particular cases of general covariance formulas first described in [25]; these general covariance formulas initially take the form of convolutions. They were derived by putting generic Gaussian process priors over the initial conditions of the wave equation and propagating them through the solution map of the said equation, leading to "wave equation-tailored" covariance functions. In this article, we explore the particular cases where the initial condition priors are either stationary, radially symmetric or point sources, as then notable simplifications can be obtained. For the point source in particular, we show that the task of recovering the position of the point source using multilateration (as in e.g. GPS systems, see [18]) is naturally recovered by maximizing the likelihood attached to the GPR models we previously obtained for the wave equation, in the limit of the small source radius (Figure 1). We will also discuss applications of our model in physical parameter estimation and initial condition reconstruction. Recovering the initial position in particular is the purpose of photoacoustic tomography (PAT), an exercise for which we will provide a simple proof of concept application, in the presence of radial symmetry.

Related literature The idea of solving and "learning" linear ODEs and PDEs thanks to GPR probably goes back to [22] and has been re-explored ever since. A large part of the subsequent works inspired by [22] deal with PDEs of the form $L(u) = f$ where f is a partially known *interior* source term: that is, f and u have the same input space. We will not be interested in this case as we will impose the strict condition that $f \equiv 0$, as is e.g. the case in PAT. In our case, the initial conditions will instead play the role of the source terms. For dealing with interior source terms, see [46, 57, 3, 47, 33, 43, 44] and [2, 38] for subsequent applications to inhomogeneous wave equations. See also [8] for an alternative method applicable to non linear PDEs. Compared to these approaches, ensuring (deterministically) the homogeneity constraint $f = 0$ in the wave equation will allow us to drastically reduce the dimensionality of the problem of approximating u given scattered measurements of u.

Ensuring homogeneous PDE constraints on centered GPs is done by appropriately constraining its covariance kernel ([25], Proposition 3.4). Such PDE constrained kernels have been explicitly built for a number of classical PDEs, namely: divergence-free vector fields [36, 49], curl-free vector fields [20, 49, 54, 28], the Laplace equation [48, 35, 1], Maxwell's equations [30], the heat equation in 1D [1] and 2D [21], Helmholtz' 2D equations [1], and linear solid mechanics [27]. See also [53] where generic PDE-constrained kernels are built under stationarity assumptions. This article is the continuation of a previous work [25], where we described a covariance kernel tailored to the wave equation at hand in this article. In parallel with homoegeneous PDEs, [31, 24, 50] enforce homogeneous *boundary conditions* on the covariance kernel.

The approach presented in this article falls in the field of Bayesian methods for solving PDE related inverse problems, the literature of which is extensive; see [51, 10, 9, 11] and the many references therein. However, the method we adopt here differs from the standard Bayesian inversion methods aforementioned in that we incorporate the PDE constraint *beforehand*, i.e. directly in the prior; the PDE does not only appear in the likelihood. See [37] for a point of view similar with that of the present article, which uses PDE-tailored GP priors for building optimal finite dimensional approximations of solution spaces of PDEs.

The inverse problems we will study deal with approximating the initial conditions of (3.1) as well as the related physical parameters (wave speed, source location and source size), given scattered measurements of the solution u. A general methodology for estimating the parameters of a linear PDE using GPR is described in [43], using the forward differential operator. Here we will rather use its inverse, i.e. the Green's function. The task of approximating the initial position in particular is the purpose of photoacoustic tomography (PAT), which is a technique commonly used e.g. in biomedical imaging [4]. See e.g. [29, 5] for details on the standard mathematical techniques and models used in PAT. Note that the solution is often assumed available on a surface enclosing the source [56], in order to use Radon transforms or similar inversion formulas. Our method instead allows the sensors to be arbitrarily scattered. As the corresponding PAT problem becomes ill-posed, we do not aim for a full reconstruction of the initial conditions. Instead, we show that our method amounts to computing a projection of the solution over a well-chosen finite dimensional space. Of course, the geometry of the sensor locations plays a crucial role in the accuracy of our model, but the reconstruction formula we introduce remains nonetheless independent of any underlying geometry assumptions. It is worth noting that [40] already showed that a GPR methodology based on Radon transforms could be set up for solving x-ray tomography problems in the presence of limited (scattered) data.

Organization of the paper For self-containment, section 2 is dedicated to reminders on GPR. Section 3 is dedicated to the study of GP priors tailored to the wave equation. In section 4, we showcase some numerical applications of the previous section on wave equation data. We conclude in section 5.

Notations Let $\mathcal{D}$ be a set, $m : \mathcal{D} \to \mathbb{R}$ and $k : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$. Given $x \in \mathcal{D}$, k_x denotes the function $y \mapsto k(x, y)$. If $X = (x_1, ..., x_n)^T$ is a column vector in $\mathcal{D}^n$, we denote $m(X)$ the column vector such that $m(X)_i = m(x_i)$, $k(X, X)$ the square matrix such that $k(X, X)_{ij} = k(x_i, x_j)$ and given $x \in \mathcal{D}$, $k(X, x)$ the column vector such that $k(X, x)_i = k(x_i, x)$. The variables $(r, \theta, \phi), r \geqslant 0, \theta \in [0, \pi], \phi \in [0, 2\pi]$, denote spherical coordinates and S denotes the unit sphere of $\mathbb{R}^3$. We write $d\Omega = \sin\theta d\theta d\phi$ its surface differential element; $\gamma = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)^T \in S$ denotes the unit length vector parametrized by (θ, ϕ).

2 Background on Gaussian process regression

2.1 Gaussian Processes and positive definite functions [45]

Let $\mathcal{D}$ be a set. A Gaussian process $(U(x))_{x \in \mathcal{D}}$ is a collection of normally distributed random variables defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that for any $(x_1, ..., x_n) \in \mathcal{D}^n$, the law of $(U(x_1), ..., U(x_n))^T$ is a multivariate normal distribution. The law of a GP is characterized by its mean and covariance functions ([26], Section 8), defined by $m(x) := \mathbb{E}[U(x)]$ and $k(x, x') = \text{Cov}(U(x), U(x')) = \mathbb{E}[U(x)U(x')] - m(x)m(x')$, and we write $(U(x))_{x \in \mathcal{D}} \sim GP(m, k)$. Given $\omega \in \Omega$, the associated sample path is the deterministic function $U_\omega : x \mapsto U(x)(\omega)$. The mean function can be chosen arbitrarily, but the covariance function has to be positive definite, which means that for all $(x_1, ..., x_n) \in \mathcal{D}^n$, the matrix $(k(x_i, x_j))_{1 \leqslant i,j \leqslant n}$ is non negative definite. The mathematical properties of the GP are encoded in the function k. Furthermore, there is a bijection between positive definite functions and covariance functions of centered GPs ([26], Theorem 8.2). We will thus focus on the design of positive definite kernels. A covariance kernels is *stationary* if $k(x, x')$ only depends on the increment $x - x'$: $k(x, x') = k_S(x - x')$. Common examples are the squared exponential and Matérn kernels [45]; see e.g. equation (4.1). Informally, if the covariance function of a GP is stationary, then its sample paths "look similar at all locations" ([45], p.4).

2.2 Gaussian Process Regression [45]

2.2.1 Kriging equations. GPs can be used for function interpolation. Let u be a function defined on $\mathcal{D}$ for which we know a dataset of values $B = \{u(x_1), ..., u(x_n)\}$. Conditioning the law of a GP $(U(x))_{x \in \mathcal{D}} \sim GP(m, k)$ on the data B yields a second GP defined by $V(x) := (U(x)|U(x_i) = u(x_i), i = 1, ..., n)$. Its mean and covariance functions $\tilde{m}$ and $\tilde{k}$ are given by the so-called *Kriging* equations (2.1) and (2.2). Note $X = (x_1, ..., x_n)^T$ and assume that $K(X, X)$ is invertible, then [45]

$$\begin{cases} \tilde{m}(x) &= m(x) + k(X, x)^T k(X, X)^{-1}(u(X) - m(X)), & (2.1) \\ \tilde{k}(x, x') &= k(x, x') - k(X, x)^T k(X, X)^{-1} k(X, x'). & (2.2) \end{cases}$$

The function $\tilde{m}$ is an estimator of u and for all x in $\mathcal{D}$, $\tilde{m}(x)$ can be used for predicting the value $u(x)$. By construction, for all observation points x_i, we have $\tilde{m}(x_i) = u(x_i)$ and $\tilde{k}(x_i, x_i) = 0$. If observing noisy data $U_i = U(x_i) + \varepsilon_i$ with $(\varepsilon_1, ..., \varepsilon_n)^T \sim \mathcal{N}(0, \sigma^2 I_n)$

independent from U, one replaces $K(X,X)$ with $K(X,X)+\sigma^2 I$ in the Kriging equations and leaves the other terms $k(X,x)$ unchanged. This amounts to applying Tikhonov regularization on $k(X,X)$, which is also relevant for approximating equations (2.1) and (2.2) when $k(X,X)$ is ill-conditioned.

2.2.2 Tuning covariance kernels [45]. Covariance functions are usually chosen among a parametrized family of kernels $\{k_\theta, \theta \in \Theta \subset \mathbb{R}^q\}$. θ contains the *hyperparameters* of k_θ. One then attempts to find the value θ which fits best the observations $u_{obs} = (u_1, ..., u_n)^T$, the set of observations of u at locations $X = (x_1, ..., x_n)$. This is performed by maximizing the *marginal likelihood*, which is the probability density of the random vector $(U(x_1), ..., U(x_n))^T$ at point u_{obs}, given θ. Denote $p(u_{obs}|\theta)$ the associated marginal likelihood at θ, one searches for $\hat{\theta}$ such that $\hat{\theta} = \text{argmax}_{\theta\in\Theta} p(u_{obs}|\theta)$. Explicitly, assuming that $m \equiv 0$, then $(U(x_1), ..., U(x_n))^T \sim \mathcal{N}(0, k_\theta(X,X))$ and

$$p(u_{obs}|\theta) = \frac{1}{(2\pi)^{n/2} \det k_\theta(X,X)^{1/2}} e^{-\frac{1}{2} u_{obs}^T k_\theta(X,X)^{-1} u_{obs}}. \tag{2.3}$$

Equivalently, for noisy observations with identical noise standard deviation σ, set

$$\begin{aligned} \mathcal{L}(\theta, \sigma^2) :&= -2 \log p(u_{obs}|\theta) - n \log 2\pi \\ &= u_{\text{obs}}^T (k_\theta(X,X) + \sigma^2 I_n)^{-1} u_{\text{obs}} + \log\det(k_\theta(X,X) + \sigma^2 I_n). \end{aligned} \tag{2.4}$$

We call $\mathcal{L}(\theta,\sigma^2)$ the negative log marginal likelihood, and one may rather attempt to find $\hat{\theta}$ such that $\hat{\theta} = \arg\min_{\theta\in\Theta} \mathcal{L}(\theta,\sigma^2)$. Note that σ can also be interpreted as a hyperparameter and estimated through log marginal likelihood minimization.

2.2.3 The RKHS point of view. The Kriging equations (2.1) and (2.2) can alternatively be viewed as orthogonal projections of u in a suitable Hilbert space. Given a positive definite kernel k defined on a set $\mathcal{D}$, one may build a Reproducing Kernel Hilbert Space (RKHS) of functions defined on $\mathcal{D}$, which we denote by $\mathcal{H}_k$. The inner product of $\mathcal{H}_k$ verifies the reproducing property [55]: $\langle k(x,\cdot), k(x',\cdot)\rangle_{\mathcal{H}_k} = k(x,x')$. One may then formulate the following regularized interpolation problem [19, 55]

$$\inf_{v\in\mathcal{H}_k} ||v||_{\mathcal{H}_k} \quad \text{s.t.} \quad v(x_i) = u(x_i) \quad \forall i \in \{1, ..., n\}. \tag{2.5}$$

Then $\tilde{m}$ in equation (2.1) is the unique solution of (2.5). One can also show [55] that equation (2.1) amounts to $\tilde{m} = m + p_F(u-m)$, where p_F stands for the orthogonal projection operator on $F := \text{Span}(k(x_1,\cdot), ..., k(x_n,\cdot))$ w.r.t. the inner product of $\mathcal{H}_k$. If in particular $m \equiv 0$, then $\tilde{m} = p_F(u)$. Likewise, equation (2.2) amounts to $\tilde{k}(x,\cdot) = P_{F^\perp}(k(x,\cdot))$. Viewing the Kriging mean as an orthogonal projection over a finite dimensional deterministic space is reminiscent of Fourier series or Galerkin reconstruction approaches.

3 Gaussian processes for the 3D wave equation

3.1 General solution to the wave equation

Denote the 3D Laplace operator $\Delta = \partial_{xx}^2 + \partial_{yy}^2 + \partial_{zz}^2$ and the d'Alembert operator with the box symbol, $\Box = c^{-2}\partial_{tt}^2 - \Delta$ with constant wave speed $c > 0$. Consider then the following initial value problem in the free space $\mathbb{R}^3$

$$\begin{cases} \Box w & = 0 & \forall (x,t) \in \mathbb{R}^3 \times \mathbb{R}_+^*, \\ w(x,0) & = u_0(x), \quad \partial_t w(x,0) = v_0(x) & \forall x \in \mathbb{R}^3. \end{cases} \tag{3.1}$$

The solution of this problem is unique in the distributional sense ([14], p. 164). It can be extended to all $t \in \mathbb{R}$ and is represented as follow ([14], p. 295)

$$w(x,t) = (F_t * v_0)(x) + (\dot{F}_t * u_0)(x) \qquad \forall (x,t) \in \mathbb{R}^3 \times \mathbb{R}. \tag{3.2}$$

$(F_t)_{t\in\mathbb{R}}$ is the Green's function of the wave equation ([13], p. 202), and is a family of singular measures. $\dot{F}_t$ is its "time derivative" ([14], equation (18.16) p. 297), understood as a continuous linear form over $C^1(\mathbb{R}^3)$. Explicitly, F_t and $\dot{F}_t$ are given by

$$F_t = \frac{\sigma_{c|t|}}{4\pi c^2 t}, \qquad \text{and} \qquad \langle \dot{F}_t, \varphi \rangle_{C^1(\mathbb{R}^3)', C^1(\mathbb{R}^3)} = \partial_t \int_{\mathbb{R}^3} \varphi(x) dF_t(x). \tag{3.3}$$

where σ_R is the surface measure of the sphere of center 0 and radius R. If $u_0 \in \mathcal{C}^1(\mathbb{R}^3)$ and $v_0 \in \mathcal{C}^0(\mathbb{R}^3)$, then w as defined in (3.2) is a pointwise defined function given by the Kirschoff formula ([16], p. 72), which writes in spherical coordinates:

$$w(x,t) = \int_S t v_0(x - c|t|\gamma) + u_0(x - c|t|\gamma) - c|t|\gamma \cdot \nabla u_0(x - c|t|\gamma) \frac{d\Omega}{4\pi}. \tag{3.4}$$

3.2 Gaussian Process priors for the wave equation

3.2.1 General covariance structure. Suppose that the initial conditions u_0 and v_0 are realizations of two independent centered Gaussian processes, $U^0 \sim GP(0, k_{\mathrm{u}})$ and $V^0 \sim GP(0, k_{\mathrm{v}})$. That is, $u_0 = U^0_\omega$ and $v_0 = V^0_\omega$ for some $\omega \in \Omega$. This assumption is relevant e.g. when u_0 and v_0 are unknown, in which case U^0 and V^0 are interpreted as GP priors over u_0 and v_0. We will assume that the sample paths of V^0 are continuous and that of U^0 are continuously differentiable, in order to use the formula (3.4) (see [25], Section 4.2 for more details and discussions on these assumptions). By solving (3.1), one obtains a time-space random field $W(x,t)$ defined by

$$W(x,t) : \Omega \ni \omega \longmapsto (F_t * V^0_\omega)(x) + (\dot{F}_t * U^0_\omega)(x). \tag{3.5}$$

The next recent result is taken from [25], which shows that the random field W is a GP as well; more importantly, it describes its covariance function.

Proposition 1. *Denote $z = (x,t)$ and $z' = (x',t')$ the space-time variables. Let k_{u} (resp. k_{v}) be a positive definite function such that the sample paths of the associated GP are continuously differentiable (resp. continuous). In particular, $k_{\mathrm{v}} \in C^0(\mathbb{R}^3 \times \mathbb{R}^3)$ and $k_{\mathrm{u}}(x, .), k_{\mathrm{u}}(., x') \in C^1(\mathbb{R}^3)$ for all $x, x' \in \mathbb{R}^3$. Define then the two functions*

$$k_{\mathrm{v}}^{\mathrm{wave}}(z,z') = [(F_t \otimes F_{t'}) * k_{\mathrm{v}}](x,x'), \tag{3.6}$$

$$k_{\mathrm{u}}^{\mathrm{wave}}(z,z') = [(\dot{F}_t \otimes \dot{F}_{t'}) * k_{\mathrm{u}}](x,x'). \tag{3.7}$$

Then $(W(z))_{z\in\mathbb{R}^3\times\mathbb{R}}$ is a centered GP whose covariance kernel is given by

$$k_{\mathrm{w}}(z,z') = k_{\mathrm{v}}^{\mathrm{wave}}(z,z') + k_{\mathrm{u}}^{\mathrm{wave}}(z,z'). \tag{3.8}$$

Expressions of equations (3.6) and (3.7) in terms of integrals of k_{u}, its first derivatives and k_{v} over the unit sphere can be found in [25], p. 21. They are derived from the Kirschoff formula (3.4). It should be mentioned that the converse of Proposition 1 holds: under mild assumptions, any second order random random field $(W(z))_{z\in\mathbb{R}^3\times\mathbb{R}}$ whose covariance function is of the form (3.8) has its sample paths solution to the wave equation, in the sense of distributions, almost surely ([25], Proposition 3.4).

Observe now that for all $z = (x,t) \in \mathbb{R}^3 \times \mathbb{R}$, we have $\Box k_{\mathrm{w}}(z,\cdot) = 0$: using equation (2.1), one then deduces that all the Kriging mean obtained using the kernel k_{w} always verifies $\Box\tilde{m} = 0$. For this reason, we call WIGPR ("Wave equation informed GPR") the act of performing GPR with a covariance kernel of the form (3.8). Note that the inheritance of the distributional PDE constraint over sample paths of the conditioned GP is proved in [25], Proposition 3.7. We now explore particular cases of the covariance expressions (3.6) and (3.7) under different assumptions over k_{v} and k_{u}.

3.2.2 Stationary initial conditions. Many standard covariance kernels used for GPR are stationary [45]. For this reason, we study equation (3.6) when k_{v} is stationary. For conciseness, we restrict ourselves to the case where $u_0 = 0$, i.e. $k_{\mathrm{u}} = 0$.

Proposition 2. *Assume that* k_{v} *is continuous and stationary:* $k_{\mathrm{v}}(x,x') = k_S(x-x')$. *(i) Then* $k_{\mathrm{v}}^{\mathrm{wave}}$ *is stationary in space and*

$$[(F_t \otimes F_{t'}) * k_{\mathrm{v}}](x,x') = (F_t * F_{t'} * k_S)(x-x'). \tag{3.9}$$

(ii) Moreover, the measure $F_t * F_{t'}$ *is absolutely continuous over* $\mathbb{R}^3$. *Denoting* $|h|$ *the Euclidean norm of* $h \in \mathbb{R}^3$ *and identifying* $F_t * F_{t'}$ *with its density, we have*

$$(F_t * F_{t'})(h) = \frac{sgn(t)sgn(t')}{8\pi c^2|h|}\mathbb{1}_{\left[c\left||t|-|t'|\right|,\,c(|t|+|t'|)\right]}(|h|). \tag{3.10}$$

Proof of Proposition 2. (i) : Assume for simplicity that $c = 1$. Using the definition of the convolution against the measure $F_t \otimes F_{t'}$ (see e.g. [52], Exercise 26.1 p. 282),

$$\begin{aligned}
[(F_t \otimes F_{t'}) * k_{\mathrm{v}}](x,x') &= \int_{\mathbb{R}^3\times\mathbb{R}^3} k(x-s_1, x'-s_2)dF_t(s_1)dF_{t'}(s_2) \\
&= \int_{\mathbb{R}^3\times\mathbb{R}^3} k_S(x-x'-s_1+s_2)dF_t(s_1)dF_{t'}(s_2).
\end{aligned}$$

But S is invariant under the change of variable $S \ni \gamma \longmapsto -\gamma$ and thus for any continuous function f, $\int_{\mathbb{R}^3} f(s_2)dF_{t'}(s_2) = \int_{\mathbb{R}^3} f(-s_2)dF_{t'}(s_2)$. This yields

$$[(F_t \otimes F_{t'}) * k_{\mathrm{v}}](x,x') = \int_{\mathbb{R}^3\times\mathbb{R}^3} k_S(x-x'-s_1-s_2)dF_t(s_1)dF_{t'}(s_2).$$

Applying the definition of the convolution of measures (see e.g. [7], p. 101) to $F_t * F_{t'}$,

$$\begin{aligned}
(F_t * F_{t'} * k_S)(h) &= \int_{\mathbb{R}^3} k_S(h-s)d(F_t * F_{t'})(s) \\
&= \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} k_S(h-s_1-s_2)dF_t(s_1)dF_{t'}(s_2).
\end{aligned}$$

Setting $h = x - x'$ finishes the proof of Point (i).
(ii) : Without loss of generality we assume that $c = 1$. The computation is carried out in the Fourier domain. Recall that F_t and $\dot{F}_t$ are tempered distributions whose Fourier transforms are given by ([14], equation (18.12) p. 294)

$$\mathcal{F}(F_t)(\xi) = \frac{\sin(ct|\xi|)}{c|\xi|} \quad \text{and} \quad \mathcal{F}(\dot{F}_t)(\xi) = \cos(ct|\xi|). \tag{3.11}$$

We then obtain that ([14], Theorem 14.33)

$$\mathcal{F}(F_t * F_{t'})(\xi) = \mathcal{F}(F_t)(\xi)\mathcal{F}(F_{t'})(\xi) = \frac{\sin(t|\xi|)\sin(t'|\xi|)}{|\xi|^2} = \frac{\cos(a|\xi|) - \cos(b|\xi|)}{2|\xi|^2}. \tag{3.12}$$

with $a = t - t', b = t + t$. We then compute the inverse Fourier transform of the quantity above. Let $h \in \mathbb{R}^3$. In spherical coordinates, noting the unit vectors $\gamma_h = h/|h|$ and $\gamma = \xi/|\xi| = \xi/r$, we define f_a by

$$\begin{aligned} f_a(h) &= \int_{\mathbb{R}^3} e^{i\langle h,\xi\rangle} \frac{\cos(a|\xi|)}{|\xi|^2} d\xi = \int_0^{+\infty} \int_0^{2\pi} \int_0^{\pi} e^{ir\langle h,\gamma\rangle} \frac{\cos(ar)}{r^2} r^2 \sin\theta d\theta d\phi dr \qquad (3.13) \\ &= \int_0^{+\infty} \cos(ar) \int_S e^{ir|h|\langle \gamma,\gamma_h\rangle} d\Omega dr. \end{aligned}$$

Above, we used the spherical coordinate change $\xi = r\gamma, d\xi = r^2 \sin\theta d\theta d\phi dr$. We now make use of radial symmetry in the interior integral, as follow. Note e_3 the third vector of the canonical basis of $\mathbb{R}^3$ and M an orthogonal matrix such that $M\gamma_h = e_3$. We perform the change of variable $\gamma' = M\gamma$, using that $MS := \{M\gamma, \gamma \in S\} = S$ and that the corresponding Jacobian is equal to 1:

$$\begin{aligned} \int_S e^{ir|h|\langle \gamma,\gamma_h\rangle} d\Omega &= \int_{MS} e^{ir|h|\langle M^T\gamma',\gamma_h\rangle} d\Omega' = \int_S e^{ir|h|\langle \gamma', M\gamma_h\rangle} d\Omega' \qquad (3.14) \\ &= \int_S e^{ir|h|\langle \gamma, e_3\rangle} d\Omega = 2\pi \int_0^{\pi} e^{ir|h|\cos(\theta)} \sin(\theta) d\theta \\ &= 2\pi \left[-\frac{e^{ir|h|\cos(\theta)}}{ir|h|} \right]_0^{\pi} = 2\pi \frac{e^{ir|h|} - e^{-ir|h|}}{ir|h|} = 4\pi \frac{\sin(r|h|)}{r|h|}, \qquad (3.15) \end{aligned}$$

and thus

$$\begin{aligned} f_a(h) &= 4\pi \int_0^{\infty} \frac{\cos(ar)\sin(|h|r)}{r|h|} dr = 4\pi \int_0^{\infty} \frac{\sin((|h| + a)r) + \sin((|h| - a)r)}{2r|h|} dr \\ &= \frac{2\pi}{|h|} \int_0^{\infty} \frac{\sin(\alpha r)}{r} + \frac{\sin(\beta r)}{r} dr, \qquad (3.16) \end{aligned}$$

with $\alpha = |h| + a, \beta = |h| - a$. Finally, we have the Dirichlet integral

$$\int_0^{\infty} \frac{\sin(\alpha r)}{r} dr = \operatorname{sgn}(\alpha)\frac{\pi}{2}. \tag{3.17}$$

We define the function f_b exactly as f_a, and compute it by replacing a by b in every step above. Putting (3.12), (3.16) and (3.17) together, the inverse Fourier transform of

$\mathcal{F}(F_t * F_{t'})$ is an absolutely continuous measure whose density f is given by

$$f(h) = \frac{1}{(2\pi)^3}\frac{1}{2}\big(f_a(h) - f_b(h)\big) = \frac{1}{16\pi|h|}\Big(\mathrm{sgn}(|h| + t - t') + \mathrm{sgn}(|h| - t + t') \\ - \mathrm{sgn}(|h| + t + t') - \mathrm{sgn}(|h| - t - t')\Big) \tag{3.18}$$

$$=: \frac{1}{16\pi|h|}K(|h|, t, t'). \tag{3.19}$$

$K(|h|, t, t')$ is defined in equation (3.19). Note that $K(|h|, -t, t') = -K(|h|, t, t')$ and likewise with t', thus $K(|h|, t, t') = \mathrm{sgn}(t)\mathrm{sgn}(t')K(|h|, T, T')$ with $T = |t|, T' = |t'|$. Using the symmetries in t and t' in equation (3.18) and the fact that $\mathrm{sgn}(s) = 1$ if $s > 0$, we obtain

$$\begin{aligned} K(|h|, T, T') = {} & \mathrm{sgn}(|h| + |T - T'|) + \mathrm{sgn}(|h| - |T - T'|) \\ & - \mathrm{sgn}(|h| + T + T') - \mathrm{sgn}(|h| - T - T') \\ = {} & 1 + \mathrm{sgn}(|h| - |T - T'|) - 1 - \mathrm{sgn}(|h| - T - T') \\ = {} & \mathrm{sgn}(|h| - |T - T'|) - \mathrm{sgn}(|h| - T - T'). \end{aligned} \tag{3.20}$$

From equation (3.20), one checks that $K(|h|, T, T') = 0$ if $|h| < |T - T'|$ or $|h| > T + T'$ and $K(|h|, T, T') = 2$ if $|T - T'| < |h| < T + T'$. Thus, $K(|h|, T, T') = 2 \times \mathbb{1}_{[|T-T'|, T+T']}(|h|)$. Identifying the measure $F_t * F_{t'}$ with its density, we obtain

$$(F_t * F_{t'})(h) = \frac{\mathrm{sgn}(t)\mathrm{sgn}(t')}{8\pi|h|}\mathbb{1}_{\left[\big||t|-|t'|\big|, |t|+|t'|\right]}(|h|). \tag{3.21}$$

which concludes the proof. $\square$

Formally, one can obtain similar formulas for $k_{\mathrm{u}}^{\mathrm{wave}}$ by differentiating the formulas above with respect to t and t', as $\dot{F}_t = \partial_t F_t$ ($\dot{F}_t * \dot{F}_{t'}$ will only be a generalized function though). We underline that the proof of Point (*ii*) Proposition 2 makes use of the specificities of the dimension 3. First in equation (3.13), where the scalars r^2 cancel each other out; second in (3.15) where an exact antiderivative of the integrated function can be computed. None of these two simplifications hold in higher dimension or in dimension 2, and formulas as simple as equation (3.10) are not expected to hold.

Remark 3.1. Expression (3.10) with $h = x - x'$ is the covariance kernel of the solution process U with initial condition the "formal" white noise process V^0 with the stationary Dirac delta covariance kernel $k_{\mathrm{v}}(x, x') = \delta_0(x - x')$:

$$[(F_t \otimes F_{t'}) * k_{\mathrm{v}}](x, x') = (F_t * F_{t'} * \delta_0)(x - x') = (F_t * F_{t'})(x - x'). \tag{3.22}$$

Somewhat surprisingly, although formula (3.10) yields a summable function over $\mathbb{R}^3$ when t and t' are fixed, it can not be used for practical computations as the diagonal terms of the related covariance matrices are all singularities: $(F_t * F_t)(0) = +\infty$... Yet, formula (3.10) may be used together with explicit kernels k_S to yield usable expressions. For instance, if $k_{\mathrm{v}}(x, x') = k_S(x - x') = C\exp(-|x - x'|^2/2L^2)$, we state without proof that

$$(F_t * F_{t'} * k_S)(h) =$$

$$\mathrm{sgn}(tt')\frac{\sqrt{2\pi}}{2}\frac{CL^3}{c^2}\left(\frac{\Phi\big(\frac{R_1+|h|}{L}\big)-\Phi\big(\frac{R_1-|h|}{L}\big)}{2|h|}-\frac{\Phi\big(\frac{R_2+|h|}{L}\big)-\Phi\big(\frac{R_2-|h|}{L}\big)}{2|h|}\right). \tag{3.23}$$

where $h = x - x'$, $\Phi(s) = (2\pi)^{-1/2}\int_{-\infty}^{s}\exp(-t^2/2)dt$, $R_1 = c\big||t| - |t'|\big|$, $R_2 = c(|t| + |t'|)$. Such a kernel always takes finite values: when h goes to 0, the above formula reduces to well defined derivatives.

Although these formulas are interesting in their own right, the study of propagation phenomena is usually done thanks to compactly supported initial conditions, which can never be modelled with a stationary GP. We partially deal with compactly supported initial conditions in Section 3.2.3, within the context of radial symmetry.

3.2.3 Radially symmetric initial conditions. Assume that the sample paths of the GP prior V^0 enjoy radial symmetry around some $x_0 \in \mathbb{R}^3$. This can be expressed in terms of differential operators in (r, θ, ϕ), the spherical coordinate system around x_0:

$$\mathbb{P}(\{\omega \in \Omega : \partial_\theta V^0_\omega = 0\}) = 1, \quad \text{and} \quad \mathbb{P}(\{\omega \in \Omega : \partial_\phi V^0_\omega = 0\}) = 1. \tag{3.24}$$

Then by Proposition 3.4 of [25], k_{v} verifies (in the sense of distributions)

$$\forall x \in \mathcal{D}, \quad \partial_\theta(k_{\mathrm{v}}(x,\cdot)) = 0 \quad \text{and} \quad \partial_\phi(k_{\mathrm{v}}(x,\cdot)) = 0.$$

Thus, there exists a kernel k^0_{v} on $\mathbb{R}_+$ such that $k_{\mathrm{v}}(x,x') = k^0_{\mathrm{v}}(r^2, r'^2)$, with $r = |x|$, $r' = |x'|$ (directly using the squares r^2 and r'^2 will simplify the computations). Similarly, assume that the sample paths of U^0 exhibit radial symmetry and write $k_{\mathrm{u}}(x,x') = k^0_{\mathrm{u}}(r^2, r'^2)$. We then have the following convolution-free formulas:

Proposition 3. *Set* $K_{\mathrm{v}}(r,r') = \int_0^r\int_0^{r'} k^0_{\mathrm{v}}(s,s')dsds'$. *Then for all* $z = (x,t) \in \mathbb{R}^3 \times \mathbb{R}$ *and* $z' = (x',t') \in \mathbb{R}^3 \times \mathbb{R}$,

$$k^{\mathrm{wave}}_{\mathrm{v}}(z,z') = \frac{sgn(tt')}{16c^2rr'}\sum_{\varepsilon,\varepsilon'\in\{-1,1\}}\varepsilon\varepsilon' K_{\mathrm{v}}\big((r+\varepsilon c|t|)^2, (r'+\varepsilon' c|t'|)^2\big), \tag{3.25}$$

$$k^{\mathrm{wave}}_{\mathrm{u}}(z,z') =$$

$$\frac{1}{4rr'}\sum_{\varepsilon,\varepsilon'\in\{-1,1\}}(r+\varepsilon c|t|)(r'+\varepsilon' c|t'|) \times k^0_{\mathrm{u}}\big((r+\varepsilon c|t|)^2, (r'+\varepsilon' c|t'|)^2\big). \tag{3.26}$$

The expressions (3.25) and (3.26) are interesting in that they are much faster to compute than (3.7) and (3.6), which require to compute convolutions.

Proof. Without loss of generality, we assume that $c = 1$ and $x_0 = 0$. We first derive expression (3.25). Let f be a function defined on $\mathbb{R}_+$ and g the function defined on $\mathbb{R}^3$ by $g(x) = f(|x|^2)$. Let F be an antiderivative of f and let $x \in \mathbb{R}^3$. As in (3.14), let M be an orthogonal matrix such that $M(x/|x|) = e_3$ and use the change of variable $\gamma' = M\gamma$. As $MS = S$, we have

$$(F_t * g)(x) = \frac{1}{4\pi t}\int_S g(x - t\gamma)t^2 d\Omega = \frac{t}{4\pi}\int_S f(|x - t\gamma|^2)d\Omega$$

$$
\begin{aligned}
&= \frac{t}{4\pi}\int_S f(|x|^2+t^2-2|t|\langle x,\gamma\rangle)d\Omega \\
&= \frac{t}{4\pi}\int_{MS} f\Big(|x|^2+t^2-2|t||x|\Big\langle \frac{x}{|x|}, M^T\gamma'\Big\rangle\Big)d\Omega' \\
&= \frac{t}{4\pi}\int_S f(|x|^2+t^2-2|t||x|\langle e_3,\gamma\rangle)d\Omega \\
&= \frac{t}{4\pi}\int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi} f(|x|^2+t^2-2|t||x|\cos(\theta))\sin(\theta)d\theta d\phi \\
&= \frac{t}{2}\int_{\theta=0}^{\pi} f(|x|^2+t^2-2|t||x|\cos(\theta))\sin(\theta)d\theta \\
&= \frac{t}{4|x||t|}\Big[F(|x|^2+t^2-2|t||x|\cos(\theta))\Big]_{\theta=0}^{\theta=\pi} \\
&= \frac{\operatorname{sgn}(t)}{4|x|}\Big(F((|x|+|t|)^2)-F((|x|-|t|)^2)\Big) \\
&= \frac{\operatorname{sgn}(t)}{4|x|}\sum_{\varepsilon\in\{-1,1\}}\varepsilon F((|x|+\varepsilon|t|)^2). \qquad (3.27)
\end{aligned}
$$

Introduce now the functions

$$
k_0^1(r,r') := \int_0^{r'} k_0(r,s)ds \qquad \text{and} \qquad K_{\mathrm{v}}(r,r') := \int_0^{r}\int_0^{r'} k(s,s')ds'ds. \qquad (3.28)
$$

We apply twice result (3.27) on k_{v}: first by setting $g(x') = k_0(|x-t\gamma|^2, |x'|^2)$ where $x-t\gamma$ is fixed, which integrates to $F(s) = k_0^1(|x-t\gamma|^2, s)$. Second, by setting $g(x) = k_0^1(|x|^2, (|x'|+\varepsilon|t'|)^2)$ where $|x'|+\varepsilon'|t'|$ is fixed, which integrates to $F(s) = K_{\mathrm{v}}(s, (|x'|+\varepsilon'|t'|)^2)$. In detail, we obtain

$$
\begin{aligned}
[(F_t\otimes F_{t'})*k_{\mathrm{v}}](x,x') &= \frac{1}{4\pi t}\frac{1}{4\pi t'}\int_S\int_S k_0(|x-t\gamma|^2, |x'-t'\gamma'|^2)t'^2 d\Omega' t^2 d\Omega \\
&= \frac{1}{4\pi t}\frac{\operatorname{sgn}(t')}{4|x'|}\int_S \sum_{\varepsilon'\in\{-1,1\}} \varepsilon' k_0^1(|x-t\gamma|^2, (|x'|+\varepsilon'|t'|)^2)t^2 d\Omega \\
&= \frac{\operatorname{sgn}(tt')}{16rr'}\sum_{\varepsilon,\varepsilon'\in\{-1,1\}} \varepsilon\varepsilon' K_{\mathrm{v}}\big((r+\varepsilon|t|)^2, (r'+\varepsilon'|t'|)^2\big). \qquad (3.29)
\end{aligned}
$$

We can then use this result to compute

$$
[(\dot{F}_t\otimes\dot{F}_{t'})*k_{\mathrm{u}}](x,x') = \partial_t\partial_{t'}[(F_t\otimes F_{t'})*k_{\mathrm{u}}](x,x'). \qquad (3.30)
$$

First, we compute it for $t\neq 0$ and $t'\neq 0$ by differentiating (3.29) w.r.t. t and t', using that for $t\neq 0$, $d|t|/dt = \operatorname{sgn}(t)$ and $d\operatorname{sgn}(t)/dt = 0$. This yields

$$
\begin{aligned}
&[(\dot{F}_t\otimes\dot{F}_{t'})*k_{\mathrm{u}}](x,x') \\
&\quad = \frac{1}{4rr'}\sum_{\varepsilon,\varepsilon'\in\{-1,1\}} (r+\varepsilon|t|)(r'+\varepsilon'|t'|)k_{\mathrm{u}}^0\big((r+\varepsilon|t|)^2, (r'+\varepsilon'|t'|)^2\big). \qquad (3.31)
\end{aligned}
$$

For the case where either $t = 0$ or $t' = 0$, note first from equation (3.11) that $\mathcal{F}(\dot{F}_0)(\xi) = 1$ and thus $\dot{F}_0 = \delta_0$, the Dirac mass at 0, which is the neutral element for the convolution. Therefore, when we have both $t = 0$ and $t' = 0$:

$$[(\dot{F}_0 \otimes \dot{F}_0) * k_{\mathrm{u}}](x, x') = [(\delta_0 \otimes \delta_0) * k_{\mathrm{u}}](x, x') = [\delta_{(0,0)} * k_{\mathrm{u}}](x, x') = k_{\mathrm{u}}(x, x').$$

which is also the result provided by (3.31) evaluated at $t = t' = 0$. When $t' = 0$ and $t \neq 0$, we still have $d|t|/dt = \mathrm{sgn}(t)$ and $d\mathrm{sgn}(t)/dt = 0$, yielding

$$\begin{aligned}
[(\dot{F}_t \otimes \dot{F}_0) * k_{\mathrm{u}}](x, x') &= [(\dot{F}_t \otimes \delta_0) * k_{\mathrm{u}}](x, x') = \partial_t[(F_t \otimes \delta_0) * k_{\mathrm{u}}](x, x') \\
&= \partial_t \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} k_{\mathrm{u}}^0(|x - y|^2, |x' - y'|^2) F_t(dy)\delta_0(dy') \\
&= \partial_t \frac{1}{4\pi t}\int_S k_{\mathrm{u}}^0(|x - t\gamma|^2, |x'|^2) t^2 d\Omega \\
&= \partial_t \frac{\mathrm{sgn}(t)}{4r} \sum_{\varepsilon \in \{-1,1\}} \varepsilon k_0^1((r + \varepsilon|t|)^2, |x'|^2) \\
&= \frac{1}{2r} \sum_{\varepsilon \in \{-1,1\}} (r + \varepsilon|t|) k_{\mathrm{u}}^0((r + \varepsilon|t|)^2, |x'|^2).
\end{aligned}$$

which is also the result provided by (3.31) evaluated at $t' = 0$. The same arguments apply to show that expression (3.31) is valid when $t = 0$ and $t' \neq 0$. Therefore the expression (3.31) is valid whatever the value of $t, t' \in \mathbb{R}$. □

3.2.4 Compactly supported initial conditions. Suppose that v_0 is compactly supported on a ball $B(x_0, R)$. The Strong Huygens Principle for the 3 dimensional wave equation ([16], p. 80) states that $F_t * v_0$ is supported on the ring $B(x_0, R + c|t|) \backslash B(x_0, (R - c|t|)_+)$, where $x_+ := \max(0, x)$. From a GP modelling perspective, assuming that $\mathrm{Supp}(V^0) \subset B(x_0, R)$ amounts to imposing that $V^0(x) = 0$ *a.s.* if $x \notin B(x_0, R)$. This is equivalent to $\mathrm{Var}(V^0(x)) = k_{\mathrm{v}}(x, x) = 0$ since V^0 is assumed centered. The same reasoning in terms of support can be applied to u_0 and U^0. In the next proposition, we explore the consequences of such compactness assumptions on the radial formulas (3.25) and (3.26). The new formulas are readily deduced from Proposition 3, but we state them on their own as they are the ones used in Section 4.

Proposition 4. *Let $R_{\mathrm{v}} > 0$ and $R_{\mathrm{u}} > 0$. Let $\alpha \in (0, 1)$ and $\varphi_\alpha : \mathbb{R}_+ \to [0, 1]$ be a C^∞ decreasing function such that $\varphi_\alpha(s) = 1$ if $s < \alpha$ and $\varphi_\alpha(s) = 0$ if $s \geqslant 1$. Set the truncated kernels*

$$k_{\mathrm{v}}^{R_{\mathrm{v}}}(x, x') = k_{\mathrm{v}}^{0,R_{\mathrm{v}}}(r^2, r'^2) = k_{\mathrm{v}}^0(r^2, r'^2)\mathbb{1}_{[0,R_{\mathrm{v}}]}(r)\mathbb{1}_{[0,R_{\mathrm{v}}]}(r'), \tag{3.32}$$

$$k_{\mathrm{u}}^{R_{\mathrm{u}}}(x, x') = k_{\mathrm{u}}^{0,R_{\mathrm{u}}}(r^2, r'^2) = k_{\mathrm{u}}^0(r^2, r'^2)\varphi(r/R_{\mathrm{u}})\varphi(r'/R_{\mathrm{u}}). \tag{3.33}$$

Assume now that $V^0 \sim GP(0, k_{\mathrm{v}}^{R_{\mathrm{v}}})$ and $U^0 \sim GP(0, k_{\mathrm{u}}^{R_{\mathrm{u}}})$. Then, defining the function $K_{\mathrm{v}}(r, r') = \int_0^r \int_0^{r'} k_{\mathrm{v}}^0(s, s') ds ds'$, the two following formulas hold

$$k_{\mathrm{v}}^{\mathrm{wave}}(z, z') = \frac{sgn(tt')}{16c^2 rr'} \times$$

$$\sum_{\varepsilon,\varepsilon' \in \{-1,1\}} \varepsilon\varepsilon' K_{\mathrm{v}}\Big(\min\big((r+\varepsilon c|t|)^2, R_{\mathrm{v}}^2\big), \min\big((r'+\varepsilon' c|t'|)^2, R_{\mathrm{v}}^2\big)\Big), \tag{3.34}$$

$$k_{\mathrm{u}}^{\mathrm{wave}}(z,z') = \frac{1}{4rr'} \times \sum_{\varepsilon,\varepsilon' \in \{-1,1\}} (r+\varepsilon c|t|)(r'+\varepsilon' c|t'|) k_{\mathrm{u}}^{0,R_{\mathrm{u}}}\big((r+\varepsilon c|t|)^2, (r'+\varepsilon' c|t'|)^2\big). \tag{3.35}$$

Proof. When using the kernel $k_{\mathrm{v}}^{0,R_{\mathrm{v}}}$, we can directly use equation (3.25) by substituting K_{v} with $K_{\mathrm{v}}^{R_{\mathrm{v}}}(r,r') := \int_0^r \int_0^{r'} k_{\mathrm{v}}^{0,R_{\mathrm{v}}}(s,s')dsds'$ and observing that for all $r, r' \geqslant 0$,

$$K_{\mathrm{v}}^{R_{\mathrm{v}}}(r^2, r'^2) := \int_0^{r^2} \int_0^{r'^2} k_{\mathrm{v}}^{0,R_{\mathrm{v}}}(s,s')dsds' = K_{\mathrm{v}}\Big(\min\big(r^2, R_{\mathrm{v}}^2\big), \min\big(r'^2, R_{\mathrm{v}}^2\big)\Big)$$

which directly proves (3.34). Additionally, (3.35) is only a substitution of k_{u}^0 with $k_{\mathrm{u}}^{0,R_{\mathrm{u}}}$ in (3.26): all the mathematical steps are justified as $\varphi \in C^\infty(\mathbb{R}_+)$. ☐

Notice that the truncated kernels $k_{\mathrm{v}}^{R_{\mathrm{v}}}$ and $k_{\mathrm{u}}^{R_{\mathrm{u}}}$ are the covariance kernels of the truncated processes $V_{\mathrm{trunc}}^0(x) = \mathbb{1}_{[0,R_{\mathrm{v}}]}(|x-x_0|)V^0(x)$ and $U_{\mathrm{trunc}}^0(x) = \varphi\big(|x-x_0|/R_{\mathrm{u}}\big)U^0(x)$ respectively. For $k_{\mathrm{u}}^{R_{\mathrm{u}}}$, the truncation procedure has to be smooth enough as computing $(\dot{F}_t * \dot{F}_{t'}) * k_{\mathrm{u}}^{R_{\mathrm{u}}}$ requires to differentiate $k_{\mathrm{u}}^{R_{\mathrm{u}}}$. In contrast, we used a blunt truncation for $k_{\mathrm{v}}^{R_{\mathrm{v}}}$. Strictly speaking, the sample paths of $V_{\mathrm{trunc}}^0(x)$ are not continuous and Proposition 1 cannot be used on this GP. However, as discussed in [25], Section 4.2.1, it is easily checked that for $V_{\mathrm{trunc}}^0(x)$, all the computations leading to equation (3.6) still hold, and thus equation (3.34) also holds.

We also observe that such compactly supported kernels can never be stationary as their sample paths are compactly supported. Using equation (3.34), one can indeed check that $k_{\mathrm{v}}^{\mathrm{wave}}(z,z) = \mathrm{Var}(V(z)) = 0$ as soon as $(r - c|t|)^2 > R_{\mathrm{v}}^2$, ie $V(z) = 0$ *a.s.* and likewise for $k_{\mathrm{u}}^{\mathrm{wave}}$: this is the expression of the strong Huygens principle on the kernels $k_{\mathrm{v}}^{\mathrm{wave}}$ and $k_{\mathrm{u}}^{\mathrm{wave}}$. Such compactly supported kernels may lead to sparse covariance matrices which may then be used for computational speedups (a topic we leave aside in this article).

3.2.5 Estimation of physical parameters. The wave kernel (3.8), using for k_{u} and k_{v} radially symmetric kernels supported in $B(x_0^{\mathrm{u}}, R_{\mathrm{u}})$ and $B(x_0^{\mathrm{v}}, R_{\mathrm{v}})$ respectively, has for hyperparameters $\theta = (c, x_0^{\mathrm{u}}, R_{\mathrm{u}}, \theta_{k_{\mathrm{u}}^0}, x_0^{\mathrm{v}}, R_{\mathrm{v}}, \theta_{k_{\mathrm{v}}^0})$ Among those, $(c, x_0^{\mathrm{u}}, R_{\mathrm{u}}, x_0^{\mathrm{v}}, R_{\mathrm{v}})$ all correspond to physical parameters. Their estimation via likelihood maximisation is numerically investigated in Section 4. Note that finding the correct R_{u} and R_{v} is not a well posed problem: if $\mathrm{Supp}(U^0) \subset B(x_0^{\mathrm{u}}, R_{\mathrm{u}})$ then $\mathrm{Supp}(U^0) \subset B(x_0^{\mathrm{u}}, \alpha R_{\mathrm{u}})$ for any $\alpha \geqslant 1$ and αR_{u} is also a suitable candidate for R_{u}. This is discussed in Section 4.

Remark 3.2 (GPR, radial symmetry and the 1D wave equation). It is known that the radially symmetric 3D wave equation is equivalent to the 1D wave equation, by introducing $\tilde{w}(r,t) = rw(x,t)$, $r = |x|$. However, the joint problem of approximating a radially symmetric solution w of Problem (3.1) with GPR *and* searching for the correct source location parameters $(x_0^{\mathrm{u}}, R_{\mathrm{u}}, x_0^{\mathrm{v}}, R_{\mathrm{v}})$ cannot be reduced to the one dimensional case, as the source centers x_0^{u} and x_0^{v} both lie in $\mathbb{R}^3$.

3.3 The Point Source Limit

The case of the point source deserves a study on its own as it plays a central role for linear PDEs, both in theory [13] and in applications. For the wave equation, modelling the source term as a point source (i.e. a Dirac mass) is relevant in a number of real life cases: a localized detonation in acoustics, an electric point source in electromagnetics, a mass point in mechanics and so forth. In this section, we will not make use of the Kriging equations (2.1) and (2.2) as reconstructing an initial condition that is a point source is actually of little interest. Also, reconstructing the wave equation's Green's function thanks to a pointwise approximation such as GPR is expected to yield poor results because this Green's function in particular is not even defined pointwise: it is a family of singular measures, see equation (3.3). However, estimating the physical parameters attached to it, essentially the position parameter x_0, is a relevant question and an attainable goal. This is the topic of this section, where we study the behaviour of the log marginal likelihood that comes with WIGPR when the initial condition reduces to a point source. We will restrict ourselves to the case $u_0 = 0$ in equation (3.1) and thus focus on the kernel $k_{\mathrm{v}}^{\mathrm{wave}}(z, z')$. We begin by clarifying the setting in which we will work.

3.3.1 Setting, assumptions and objectives.

(i) Note $x_1, ..., x_q$ the q sensor locations and assume that we have N time measurements in $[0, T]$ corresponding to times $0 = t_1 < ... < t_N = T$ for each sensor; we have overall $n = Nq$ pointwise observations of a function w that is a solution of the problem (3.1). The space-time observation locations (x_i, t_j) are stored in a vector $Z = (Z_1|\cdots|Z_q)^T$ where $Z_i := ((x_i, t_1), ..., (x_i, t_N))$ corresponds to the i^{th} sensor. The observations are then stored in the column vector $w_{\mathrm{obs}} = (w(Z_1)|...|w(Z_q))^T$.

(ii) We assume that the initial condition v_0 corresponding to w is almost a point source: in particular it is supported on a small ball $B(x_0^*, R^*)$ where $R^* \ll 1$.

(iii) We are interested in finding x_0^*, the correct source location. To do so, we study the log marginal likelihood associated to the observations w_{obs}, using a covariance kernel associated to initial conditions truncated around a ball $B(x_0, R)$ to be estimated. Set first $k_{x_0}^{\mathrm{R}}(x, x') := (4\pi R^3/3)^{-2} k_{\mathrm{v}}(x, x') \mathbb{1}_{B(x_0,R)}(x) \mathbb{1}_{B(x_0,R)}(x')$ where k_{v} is a given a covariance function. The pre-factor $(4\pi R^3/3)^{-2}$ is an anticipation of the upcoming Proposition 5. We will then use the wave kernel

$$k_{\mathrm{x}_0}^{\mathrm{wave,R}}((x, t), (x', t')) = [(F_t \otimes F_{t'}) * k_{x_0}^{\mathrm{R}}](x, x'). \tag{3.36}$$

We then view (x_0, R) as hyperparameters of $k_{\mathrm{x}_0}^{\mathrm{wave,R}}$, and we denote (x_0^*, R^*) the real source position and size.

(iv) We assume that except for x_0, all the other hyperparameters θ of $k_{\mathrm{x}_0}^{\mathrm{wave,R}}$ are fixed. In particular, we assume that $c = c^*$ and $R = R^*$.

In that framework, the log-marginal likelihood $p(w_{\mathrm{obs}}|\theta)$ only depends on x_0. We thus write $K_{x_0} := k_{x_0}^{\mathrm{R}}(X, X)$ and $\mathcal{L}(\theta, \lambda) = \mathcal{L}(x_0, \lambda)$, λ being a Tikhonov regularization parameter (see equation (3.37) below). The log-marginal likelihood then writes

$$\mathcal{L}(\theta, \lambda) = \mathcal{L}(x_0, \lambda) = w_{\mathrm{obs}}^T (K_{x_0} + \lambda I_n)^{-1} w_{\mathrm{obs}} + \log\det(K_{x_0} + \lambda I_n). \tag{3.37}$$

3.3.2 Level sets of $\mathcal{L}(x_0, \lambda)$ and GPS localization. In Figure 1, we provide a 3 dimensional image which displays the numerical values of the map $x_0 \mapsto \mathcal{L}(x_0, \lambda)$ that are below a suitable threshold, on a test case. This figure constitutes visual evidence that in the limit $R \to 0$, recovering a point source location from minimizing the log marginal likelihood provided by the kernel (3.36) reduces to the classic true-angle multilateration method used for example in GPS systems (see e.g. [18]). In this localization method, the user who is located on a sphere (Earth) sends signals to satellites gravitating around the Earth. From the corresponding time measurements, the distance between the satellite and the user is deduced, which in turn defines a sphere (one for each satellite) on which the user is located. The location of the user lies at the intersection of those spheres, and the Earth. At least three satellites are needed for this intersection to be reduced to a point.

On Figure 1, three facts in particular are noteworthy; our task will be to explain them mathematically. First, as a function of x_0, $\mathcal{L}(x_0, \lambda)$ reaches local minima over the whole surface of spheres centered on each sensor. Second, at the intersection of two of those spheres, the local minima are smaller. Third, the spheres all intersect at a single point x_0^*, which is the global minima of $\mathcal{L}(x_0, \lambda)$ and the real source location.

On our way to explaining these three facts, we begin with a convergence statement describing the point source limit, from a covariance point of view.

Proposition 5. *Let k be a continuous positive definite function defined on $\mathbb{R}^3 \times \mathbb{R}^3$ and let $x_0 \in \mathbb{R}^3$. For $R > 0$, define $k_{x_0}^{\mathrm{R}}$ its truncation around x_0 by*

$$k_{x_0}^{\mathrm{R}}(x, x') = k(x, x')\mathbb{1}_{B(x_0,R)}(x)\mathbb{1}_{B(x_0,R)}(x')/(4\pi R^3/3)^2.$$

Let $t, t' \in \mathbb{R}$. Then $(F_t \otimes F_{t'}) * k_{x_0}^{\mathrm{R}}$ defines an absolutely continuous Radon measure over $\mathbb{R}^3 \times \mathbb{R}^3$. Furthermore we have the following weak-$\star$ convergence in the space of Radon measures (i.e. the dual of $C_c(\mathbb{R}^3 \times \mathbb{R}^3)$, the latter space being the space of continuous functions over $\mathbb{R}^3 \times \mathbb{R}^3$ with compact support):

$$[(F_t \otimes F_{t'}) * k_{x_0}^{\mathrm{R}}] \xrightarrow[R \to 0]{C_c(\mathbb{R}^3 \times \mathbb{R}^3)'} k(x_0, x_0) \times (\tau_{x_0} F_t) \otimes (\tau_{x_0} F_{t'}), \tag{3.38}$$

where $\tau_x \mu$, the translation of μ by x, is defined by $\int f(y)\tau_x\mu(dy) := \int f(x+y)\mu(dy)$.

Proof. The proof is carried out by direct computations. First, equation (3.4) yields

$$[(F_t \otimes F_{t'}) * k_{x_0}^{\mathrm{R}}](x, x') = tt' \int_{S \times S} k_{x_0}^{\mathrm{R}}(x - c|t|\gamma, x' - c|t'|\gamma') \frac{d\Omega d\Omega'}{(4\pi)^2}. \tag{3.39}$$

The integrated function in equation (3.39) is piecewise continuous over $\mathbb{R}^3 \times \mathbb{R}^3$ and the integral in (3.39) is well defined, whatever the values of x and x'. Let f be a continuous compactly supported function on $\mathbb{R}^3 \times \mathbb{R}^3$. We define

$$I_R := \langle (F_t \otimes F_{t'}) * k_{x_0}^{\mathrm{R}}, f\rangle/(4\pi R^3/3)^2,$$

and wish to show that $I_R \to k(x_0, x_0)\langle \tau_{x_0} F_t \otimes \tau_{x_0} F_{t'}, f\rangle$ when $R \to 0$. Using equation

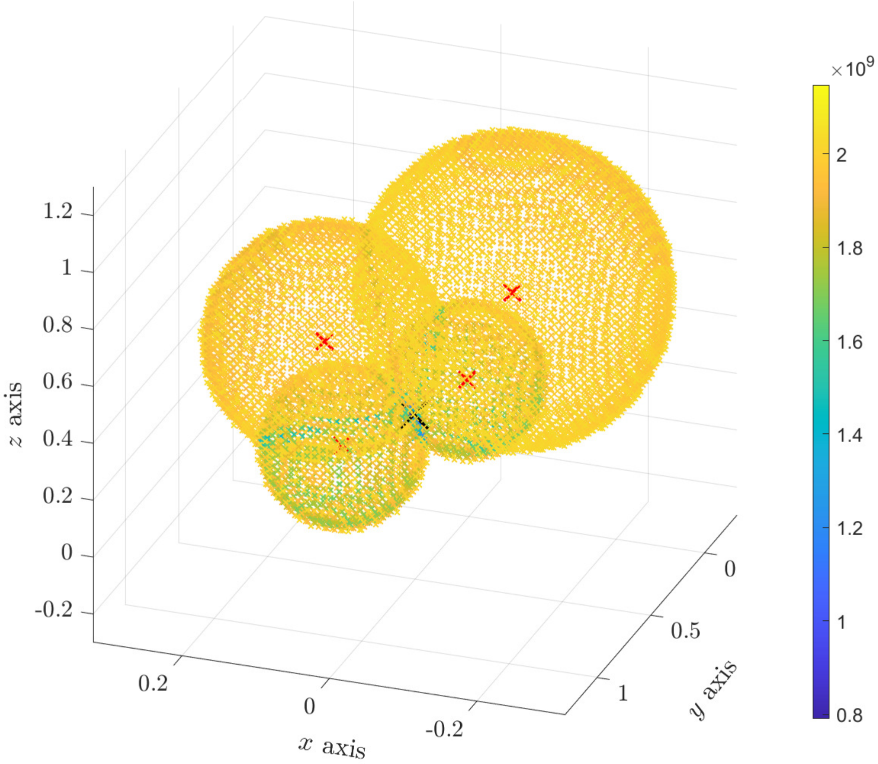


Figure 1: Negative log marginal likelihood as a function of $x_0 \in \mathbb{R}^3$. Are only represented values of the negative log marginal likelihood that are below 2.035×10^9. There only remains thin spherical shells. Red crosses: sensor locations. Black cross: source position. The source is located at the intersection of spheres centered at the sensor locations.

(52) from [25] and Fubini's theorem, we have

$$\begin{aligned}
I_R &= \frac{1}{(\frac{4}{3}\pi R^3)^2} \int_{\mathbb{R}^3 \times \mathbb{R}^3} f(x,x')[(F_t \otimes F_{t'}) * k^{\mathrm{R}}_{x_0}](x,x')dxdx' \\
&= \frac{1}{(\frac{4}{3}\pi R^3)^2} \int_{\mathbb{R}^3 \times \mathbb{R}^3} f(x,x')tt' \int_{S \times S} k^{\mathrm{R}}_{x_0}(x - c|t|\gamma, x' - c|t'|\gamma') \frac{d\Omega d\Omega'}{(4\pi)^2} dxdx' \\
&= \frac{1}{(\frac{4}{3}\pi R^3)^2}\, tt' \int_{S \times S} \int_{\mathbb{R}^3 \times \mathbb{R}^3} \Bigg(f(x,x') k_{x_0}(x - c|t|\gamma, x' - c|t'|\gamma') \\
&\qquad\qquad \times \mathbb{1}_{[0,R]}(|x - c|t|\gamma - x_0|)\mathbb{1}_{[0,R]}(|x' - c|t'|\gamma' - x_0|)\Bigg) dxdx' \frac{d\Omega d\Omega'}{(4\pi)^2}.
\end{aligned}$$

The first indicator function restricts the integration domain of x to $B(x_0 + c|t|\gamma, R)$, and symmetrically for the second indicator function and x'. For x in $B(x_0 + c|t|\gamma, R)$, in spherical coordinates around $x_0 + c|t|\gamma$, write $x = x_0 + c|t|\gamma + R\rho\gamma_x$ with $\rho \in [0,1]$, $\gamma_x \in S$ and associated surface differential element $d\Omega_x$. We do symmetrically for $x' \in B(x_0 + c|t'|\gamma', R)$, which yields

$$I_R = tt' \int_{S\times S} \int_{S\times S} \int_0^1 \int_0^1 \Bigg(f(x_0 + c|t|\gamma + R\rho\gamma_x, x_0 + c|t'|\gamma' + R\rho'\gamma_{x'}) \\ \times k(x_0 + R\rho\gamma_x, x_0 + R\rho'\gamma_{x'}) \Bigg) \times 9\rho^2 d\rho \rho'^2 d\rho' \frac{d\Omega_x d\Omega_{x'}}{(4\pi)^2} \frac{d\Omega d\Omega'}{(4\pi)^2}.$$

The integration domain above is a compact subset of $\mathbb{R}^{10}$. Since f is continuous and k is assumed continuous in the vicinity of (x_0, x_0), Lebesgue's dominated convergence theorem can be applied when $R \to 0$, which yields

$$\begin{aligned} I_R &\xrightarrow[R\to 0]{} tt' k(x_0, x_0) \int_{S\times S} f(x_0 + c|t|\gamma, x_0 + c|t'|\gamma') \frac{d\Omega d\Omega'}{(4\pi)^2} \times \left(3\int_0^1 \rho^2 d\rho\right)^2 \\ &= k(x_0, x_0) \langle \tau_{x_0} F_t \otimes \tau_{x_0} F_{t'}, f \rangle. \end{aligned}$$

which concludes the proof. □

As before, the kernel $k_{x_0}^{\mathrm{R}}$ of Proposition 5 is the covariance kernel of the truncated process $V^0_{\mathrm{trunc}}(x) = \mathbb{1}_{B(x_0,R)}(x) V^0(x)/(4\pi R^3/3)$. The limit object we obtain in equation (3.38) is not a function but a singular measure, and thus it cannot be a covariance function. This means that we do not obtain a Gaussian process in the point source limit. More precisely, the Gaussian process associated to the covariance function $k_{\mathrm{x_0}}^{\mathrm{wave,R}}$ degenerates into a Gaussian measure [7] over the locally convex space $C_c(\mathbb{R}^3 \times \mathbb{R}^3)$ when R goes to zero, though we leave aside this observation for now. On a formal level though, Proposition 5 provides an entry point for studying the log marginal likelihood (3.37) associated with the kernel (3.36) when R is small. Indeed, Proposition 5 states that for small values of R, the kernel (3.36) behaves like a rank one kernel, i.e. a kernel of the form $k(z, z') = f(z)f(z')$ for some particular function f. This observation will prove to be enough for explaining the patterns observed in Figure 1.

Properly dealing with the limit $R \to 0$ implies that we use a mathematical framework compatible with general Radon measures, as indicated by Proposition 5. This also implies an additional layer of technicality. Instead, we introduce regularized (mollified) versions of both the limit object in Proposition 5 and $\mathcal{L}(x_0, \lambda)$. and study these regularized terms. This is the content of Propositions 6 and 7, which are statements on the regularized log marginal likelihood $\mathcal{L}_{\mathrm{reg}}(x_0, \lambda)$ introduced in equation (3.40). Note however that proving a rigorous mathematical statement linking the behaviours of $\mathcal{L}(x_0, \lambda)$ and $\mathcal{L}_{\mathrm{reg}}(x_0, \lambda)$ is an open question.

3.3.3 Point source mollification. We start with regularizing F_t thanks to a mollifier $\varphi(x)$ on $\mathbb{R}^3$ which we choose to be radially symmetric as in [17], section 4.2.1. Define $\varphi_R(y) = \varphi(y/R)/R^3$, then a $\mathcal{C}_c^\infty$ regularization of F_t is obtained by setting $f_t^{\mathrm{R}}(x) := (F_t * \varphi_R)(x)$ for all x in $\mathbb{R}^3$. As F_t, f_t^{R} exhibits radial symmetry. We will next use the following regularizations:

- Note $k^{\mathrm{reg}}_{x_0}((x,t),(x',t')) := f^{\mathrm{R}}_t(x-x_0)f^{\mathrm{R}}_{t'}(x'-x_0)$, which plays the role of a regularized version of the limit measure in Proposition 5. The same proposition states that in some sense, when R approaches 0, $k^{\mathrm{wave,R}}_{\mathrm{x}_0}$ is close to $k^{\mathrm{reg}}_{x_0}$. Denote also $F_{x_0} := (F^1_{x_0}|\cdots|F^q_{x_0})^T$, with $F^i_{x_0} := (f^{\mathrm{R}}_{t_1}(x_i-x_0),...,f^{\mathrm{R}}_{t_N}(x_i-x_0))$. The covariance matrix corresponding to the hyperparameter x_0 is then given by $K^{\mathrm{reg}}_{x_0} = k^{\mathrm{reg}}_{x_0}(Z,Z) = F_{x_0}F^T_{x_0}$. In particular it is rank one.

- We also assume that $w(x_i,t_j)$ can be approximated by $\tilde{w}(x_i,t_j) = f^{\mathrm{R}}_{t_j}(x_i-x^*_0)$ as in the point source limit, $v_0 = \delta_{x^*_0}$ and in that case we would have $w(x_i,t_j) = (F_{t_j} * v_0)(x_i) = F_{t_j}(x_i-x^*_0)$ (forgetting for a second that F_t is not defined pointwise). We thus introduce the column vector of "approximated observations" $W = \big(\tilde{w}(x_i,t_j)\big)_{i,j}$ and we assume that W is ordered as $W = (W_1|\cdots|W_q)^T$ where W_i corresponds to the i^{th} sensor: $W_i = (\tilde{w}(x_i,t_1),...,\tilde{w}(x_i,t_N)) \in \mathbb{R}^N$.

We may then introduce the "regularized" log marginal likelihood built by replacing k with $k^{\mathrm{reg}}_{x_0}$ and w_{obs} by W:

$$\mathcal{L}_{\mathrm{reg}}(x_0,\lambda) := W^T(K^{\mathrm{reg}}_{x_0}+\lambda I_n)^{-1}W + \log\det(K^{\mathrm{reg}}_{x_0}+\lambda I_n), \tag{3.40}$$

with $K^{\mathrm{reg}}_{x_0} = k^{\mathrm{reg}}_{x_0}(Z,Z) = F_{x_0}F^T_{x_0}$. We will then study $\mathcal{L}_{\mathrm{reg}}(x_0,\lambda)$ in the place of $\mathcal{L}(x_0,\lambda)$; as stated before, we expect that $\mathcal{L}(x_0,\lambda)$ behaves similarly to $\mathcal{L}_{\mathrm{reg}}(x_0,\lambda)$, although proofs of such statements are lacking for the moment.

We begin with a proposition which describes the asymptotic behaviour of $\mathcal{L}_{\mathrm{reg}}(x_0,\lambda)$ in the limit of $\lambda \to 0$. This limit corresponds to noiseless observations, and the limit object in Proposition 6 provides an explanation of the patterns of Figure 1.

Proposition 6 (Asymptotic behaviour of $\mathcal{L}_{\mathrm{reg}}(x_0,\lambda)$ when $\lambda \to 0$)**.** *Let $\varepsilon > 0$ and $E_\varepsilon := \{x_0 \in \mathbb{R}^3 : ||F_{x_0}||^2_{\mathbb{R}^n} > \varepsilon\}$. Define the correlation coefficient between F_{x_0} and W by $r(x_0) = Corr(F_{x_0},W) = \langle F_{x_0},W\rangle_{\mathbb{R}^n}/(||W||_{\mathbb{R}^n}||F_{x_0}||_{\mathbb{R}^n})$. We set $r(x_0) = 0$ if $F_{x_0} = 0$. Then we have the following pointwise convergence:*

$$\forall x_0 \in \mathbb{R}^3, \quad \big|\lambda\mathcal{L}_{\mathrm{reg}}(x_0,\lambda) - ||W||^2_{\mathbb{R}^n}\big(1-r(x_0)^2\big)\big| = O_{\lambda\to 0}(\lambda\log\lambda),$$

and the uniform convergence on E_ε

$$\sup_{x_0\in E_\varepsilon}\big|\lambda\mathcal{L}_{\mathrm{reg}}(x_0,\lambda) - ||W||^2_{\mathbb{R}^n}\big(1-r(x_0)^2\big)\big| = O_{\lambda\to 0}(\lambda\log\lambda).$$

The set E_ε is the set of values of x_0 for which the vectors F_{x_0} are uniformly large enough for the Euclidean norm. This is interpreted by saying that the elements x_0 of E_ε are potential source positions for which the chosen sensor locations should capture a signal with sufficient L^2 energy (at least ε across all sensors) over the window $[0,T]$, in the case where the source is indeed located at x_0. Loosely speaking, such locations x_0 are "visible" source positions. From a covariance perspective, we have that $\rho(K^{reg}_{x_0}) = ||F_{x_0}||^2_{\mathbb{R}^n}$, where ρ denotes the spectral radius.

Proof of Proposition 6. Suppose first that $||F_{x_0}||^2_{\mathbb{R}^n} = 0$. Then by definition, $r(x_0) = 0$ and $\mathcal{L}_{\mathrm{reg}}(x_0,\lambda) = ||W||^2_{\mathbb{R}^n}/\lambda + n\log\lambda$ which indeed shows that

$$\big|\lambda\mathcal{L}_{\mathrm{reg}}(x_0,\lambda) - ||W||^2_{\mathbb{R}^n}\big| = O_{\lambda\to 0}(\lambda\log\lambda). \tag{3.41}$$

Now, let $\varepsilon > 0$ and assume that $||F_{x_0}||^2_{\mathbb{R}^n} \geqslant \varepsilon$. We first deal with the first term in equation (3.40). Using the Sherman–Morrison formula ([39], Section 2.7.1), we may invert $(K^{\text{reg}}_{x_0} + \lambda I_n)$ explicitly:

$$(K^{\text{reg}}_{x_0} + \lambda I_n)^{-1} = \frac{1}{\lambda} I_n - \frac{1}{\lambda^2} \frac{F_{x_0} F^T_{x_0}}{1 + \frac{1}{\lambda} F^T_{x_0} F_{x_0}} = \frac{1}{\lambda}\Big(I_n - \frac{F_{x_0} F^T_{x_0}}{\lambda + ||F_{x_0}||^2_{\mathbb{R}^n}} \Big).$$

The determinant term in equation (3.40) is also easily derived. Indeed, $F_{x_0} F^T_{x_0}$ has only one non zero eigenvalue equal to $||F_{x_0}||^2_{\mathbb{R}^n}$, since $(F_{x_0} F^T_{x_0}) F_{x_0} = F_{x_0} (F^T_{x_0} F_{x_0}) = ||F_{x_0}||^2_{\mathbb{R}^n} F_{x_0}$:

$$\log \det(K^{\text{reg}}_{x_0} + \lambda I_n) = (n-1) \log \lambda + \log(\lambda + ||F_{x_0}||^2_{\mathbb{R}^n}). \tag{3.42}$$

(The same argument shows that $\rho(K^{\text{reg}}_{x_0}) = ||F_{x_0}||^2_{\mathbb{R}^n}$.) Thus,

$$\begin{aligned}
\mathcal{L}_{\text{reg}}(x_0, \lambda) &= W^T (K^{\text{reg}}_{x_0} + \lambda I_n)^{-1} W + \log \det(K^{\text{reg}}_{x_0} + \lambda I_n) \\
&= \frac{1}{\lambda}\left(||W||^2_{\mathbb{R}^n} - \frac{\langle F_{x_0}, W \rangle^2_{\mathbb{R}^n}}{\lambda + ||F_{x_0}||^2_{\mathbb{R}^n}} \right) + (n-1)\log\lambda + \log(\lambda + ||F_{x_0}||^2_{\mathbb{R}^n}) \\
&= \frac{||W||^2_{\mathbb{R}^n}}{\lambda}\left(1 - \frac{\langle F_{x_0}, W \rangle^2_{\mathbb{R}^n}}{||W||^2_{\mathbb{R}^n}(\lambda + ||F_{x_0}||^2_{\mathbb{R}^n})} \right) + (n-1)\log\lambda + \log(\lambda + ||F_{x_0}||^2_{\mathbb{R}^n}).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\lambda \mathcal{L}_{\text{reg}}(x_0, \lambda) - ||W||^2_{\mathbb{R}^n}(1 - r(x_0)^2) \\
&\qquad = ||W||^2_{\mathbb{R}^n}\left(\frac{\langle F_{x_0}, W \rangle^2_{\mathbb{R}^n}}{||W||^2_{\mathbb{R}^n} ||F_{x_0}||^2_{\mathbb{R}^n}} - \frac{\langle F_{x_0}, W \rangle^2_{\mathbb{R}^n}}{||W||^2_{\mathbb{R}^n}(\lambda + ||F_{x_0}||^2_{\mathbb{R}^n})} \right) \\
&\qquad\quad + (n-1)\lambda \log\lambda + \lambda \log(\lambda + ||F_{x_0}||^2_{\mathbb{R}^n}).
\end{aligned} \tag{3.43}$$

Moreover, for the term in equation (3.43) which is multiplied by $||W||^2_{\mathbb{R}^n}$,

$$\begin{aligned}
\frac{\langle F_{x_0}, W \rangle^2_{\mathbb{R}^n}}{||W||^2_{\mathbb{R}^n} ||F_{x_0}||^2_{\mathbb{R}^n}} - \frac{\langle F_{x_0}, W \rangle^2_{\mathbb{R}^n}}{||W||^2_{\mathbb{R}^n}(\lambda + ||F_{x_0}||^2_{\mathbb{R}^n})} &= \frac{\langle F_{x_0}, W \rangle^2_{\mathbb{R}^n}}{||W||^2_{\mathbb{R}^n}}\left(\frac{1}{||F_{x_0}||^2_{\mathbb{R}^n}} - \frac{1}{\lambda + ||F_{x_0}||^2_{\mathbb{R}^n}} \right) \\
&= \frac{\langle F_{x_0}, W \rangle^2_{\mathbb{R}^n}}{||W||^2_{\mathbb{R}^n}} \frac{\lambda}{||F_{x_0}||^2_{\mathbb{R}^n}(\lambda + ||F_{x_0}||^2_{\mathbb{R}^n})} \\
&\leqslant r(x_0)^2 \frac{\lambda}{\lambda + ||F_{x_0}||^2_{\mathbb{R}^n}} \leqslant \frac{\lambda}{||F_{x_0}||^2_{\mathbb{R}^n}} \leqslant \frac{\lambda}{\varepsilon},
\end{aligned} \tag{3.44}$$

and obviously, since $\lambda \geqslant 0$,

$$\frac{\langle F_{x_0}, W \rangle^2_{\mathbb{R}^n}}{||W||^2_{\mathbb{R}^n} ||F_{x_0}||^2_{\mathbb{R}^n}} - \frac{\langle F_{x_0}, W \rangle^2_{\mathbb{R}^n}}{||W||^2_{\mathbb{R}^n}(\lambda + ||F_{x_0}||^2_{\mathbb{R}^n})} \geqslant 0. \tag{3.45}$$

Also, one sees that $F_{x_0} = 0$ as soon as $\sup_i |x_0 - x_i| > cT + R$, ie x_0 is too far from the receivers for them to capture non zero signal during the time interval $[0, T]$. Thus the function $x_0 \longmapsto ||F_{x_0}||^2_{\mathbb{R}^n}$ is zero outside of a compact set. It is obviously continuous on $\mathbb{R}^3$ and is thus bounded on $\mathbb{R}^3$ by some constant $M > 0$. Using this together with equations (3.44) and (3.45) inside equation (3.43), and assuming that $\lambda \leqslant 1$ yields

$$\big|\lambda \mathcal{L}_{\text{reg}}(x_0, \lambda) - ||W||^2_{\mathbb{R}^n}(1 - r(x_0)^2)\big| \leqslant \frac{\lambda}{\varepsilon} ||W||^2_{\mathbb{R}^n} + (n-1)|\lambda \log \lambda| + \lambda \log(M+1),$$

which shows the uniform convergence statement as well as the pointwise one (together with (3.41)). $\square$

Remark 3.3. In the proof of Proposition 6, the determinant term in (3.40) has no influence in the limit object and only pollutes the rate of convergence. Discarding it leads to a $O_{\lambda \to 0}(\lambda)$ rate of convergence.

It also makes sense to inspect the case $N \to \infty$, which is the content of the next proposition; the obtained limit object is similar to that of Proposition 6. The limit $N \to \infty$ corresponds to having the sampling frequency of the sensors go to infinity. In this case, the discrete objects in Proposition 6 behave as Riemann sums if the time steps t_k are equally spaced and we obtain integrals in the limit $N \to \infty$. Notation wise, we highlight the dependence in N in $\mathcal{L}_{\mathrm{reg}}(x_0, \lambda)$ by noting it instead $\mathcal{L}^N_{\mathrm{reg}}(x_0, \lambda)$.

Proposition 7 (Asymptotic behaviour of $\mathcal{L}^N_{\mathrm{reg}}(x_0, \lambda)$ when $N \to \infty$)**.** *Define the following vector valued functions in* $L^2([0,T], \mathbb{R}^q)$*:*

$$\forall t \in [0,T], \quad I_{\mathrm{w}}(t) := \big(\tilde{w}(x_1, t), ..., \tilde{w}(x_q, t)\big)^T,$$
$$\forall t \in [0,T], \quad I_{x_0}(t) := \big(f^{\mathrm{R}}_t(x_1 - x_0), ..., f^{\mathrm{R}}_t(x_q - x_0)\big)^T.$$

Denote $|| \cdot ||_{L^2}$ *and* $\langle , \rangle_{L^2}$ *the norm and the dot product of the usual Euclidean structure of* $L^2([0,T], \mathbb{R}^q)$*. Assume that the observations are such that* $||I_{\mathrm{w}}||_{L^2} > 0$*. Introduce then the correlation function, defined whenever* $||I_{x_0}||_{L^2} > 0$*:*

$$r_\infty(x_0) := \frac{\langle I_{\mathrm{w}}, I_{x_0} \rangle_{L^2}}{||I_{\mathrm{w}}||_{L^2} ||I_{x_0}||_{L^2}}. \tag{3.46}$$

Assume that for all $k \in \{1, ..., N\}$, $t_k = T(k-1)/(N-1)$*, i.e. the* t_k *are equally spaced in* $[0,T]$*. Then for all* x_0 *such that* $||I_{x_0}||_{L^2} \neq 0$*, we have the following pointwise convergence at* x_0

$$\frac{\lambda}{N} \mathcal{L}^N_{\mathrm{reg}}(x_0, \lambda) \xrightarrow[N \to \infty]{} ||I_{\mathrm{w}}||^2_{L^2} \big(1 - r_\infty(x_0)^2\big) + q\lambda \log \lambda. \tag{3.47}$$

Proof of Proposition 7. In all concerned mathematical objects, we highlight the N dependency with an exponent, i.e. W^N, $F^N_{x_0}$, etc. We use the exact same tools as in the previous proof, namely that we the following equality holds:

$$\mathcal{L}^N_{\mathrm{reg}}(x_0, \lambda) = \frac{||W^N||^2_{\mathbb{R}^n}}{\lambda} \left(1 - \frac{\langle F^N_{x_0}, W^N \rangle^2_{\mathbb{R}^n}}{||W^N||^2_{\mathbb{R}^n} \big(\lambda + ||F^N_{x_0}||^2_{\mathbb{R}^n}\big)}\right)$$
$$+ (n-1) \log \lambda + \log(\lambda + ||F^N_{x_0}||^2_{\mathbb{R}^n}).$$

But we also have $||W^N||^2_{\mathbb{R}^n} = \sum_{i=1}^q \sum_{k=1}^N \tilde{w}(x_i, t_k)^2$, $||F^N_{x_0}||^2_{\mathbb{R}^n} = \sum_{i=1}^q \sum_{k=1}^N f^{\mathrm{R}}_{t_k}(x_i - x_0)^2$ and $\langle F^N_{x_0}, W^N \rangle_{\mathbb{R}^n} = \sum_{i=1}^q \sum_{k=1}^N f^{\mathrm{R}}_{t_k}(x_i - x_0) \tilde{w}(x_i, t_k)$. Since the time steps are equally spaced, we can study the limit $N \to \infty$ of the above objects thanks to Riemann sums.

When $N \to \infty$,

$$\frac{1}{N}||W^N||^2_{\mathbb{R}^n} \longrightarrow \sum_{i=1}^{q} \int_0^T \tilde{w}(x_i,t)^2 dt = ||I_{\mathrm{w}}||^2_{L^2}, \tag{3.48}$$

$$\frac{1}{N}||F^N_{x_0}||^2_{\mathbb{R}^n} \longrightarrow \sum_{i=1}^{q} \int_0^T f_t(x_i - x_0)^2 dt = ||I_{x_0}||^2_{L^2}, \tag{3.49}$$

$$\frac{1}{N}\langle W^N, F^N_{x_0}\rangle_{\mathbb{R}^n} \longrightarrow \sum_{i=1}^{q} \int_0^T \tilde{w}(x_i,t) f_t(x_i - x_0) dt = \langle I_{\mathrm{w}}, I_{x_0}\rangle_{L^2}. \tag{3.50}$$

Assume that x_0 is such that $||I_{x_0}||_{L^2} \neq 0$, then because of equation (3.49), the quantity $||F^N_{x_0}||_{\mathbb{R}^n}$ is bounded from below by a constant $C > 0$ for sufficiently large N (say $C = ||I_{x_0}||_{L^2}/2$). From the three equations above, we then have the following convergence:

$$\frac{\langle F^N_{x_0}, W^N\rangle^2_{\mathbb{R}^n}}{||W^N||^2_{\mathbb{R}^n}(\lambda + ||F^N_{x_0}||^2_{\mathbb{R}^n})} = \frac{(\frac{1}{N}\langle F^N_{x_0}, W^N\rangle_{\mathbb{R}^n})^2}{\frac{1}{N}||W^N||^2_{\mathbb{R}^n}(\frac{\lambda}{N} + \frac{1}{N}||F^N_{x_0}||^2_{\mathbb{R}^n})} \xrightarrow[N\to\infty]{} r_\infty(x_0). \tag{3.51}$$

Likewise, since $n = qN$, when $N \to \infty$ we have that

$$\begin{aligned}&\frac{(n-1)\log\lambda}{N} + \frac{1}{N}\log(\lambda + ||F_{x_0}||^2_{\mathbb{R}^n})\\ &\qquad = \frac{(Nq-1)\log\lambda}{N} + \frac{\log N}{N} + \frac{1}{N}\log\left(\frac{\lambda}{N} + \frac{1}{N}||F_{x_0}||^2_{\mathbb{R}^n}\right) \xrightarrow[N\to\infty]{} q\log\lambda.\end{aligned}$$

which, together with equation (3.51), shows the announced result. □

3.3.4 Discussion: location of the point source. Propositions 6 and 7 enable us to explain the patterns observed in Figure 1 where the correct source position is located at the intersection of spheres centered on receivers.

For that purpose, we analyze the limit term in Proposition 6 (or the one in Proposition 7). We denote $L(x_0)$ the said limit object from Proposition 6:

$$L(x_0) = ||W||^2_{\mathbb{R}^n}\left(1 - r(x_0)^2\right) = ||W||^2_{\mathbb{R}^n}\left(1 - \frac{\left(\sum_{i=1}^q \langle F^i_{x_0}, W_i\rangle_{\mathbb{R}^n}\right)^2}{||W||^2_{\mathbb{R}^n}||F_{x_0}||^2_{\mathbb{R}^n}}\right).$$

Note T_i the time of arrival of the point source wave at sensor i: $|x_i - x_0^*| = c^*T_i$. Define also $S_i := S(x_i, cT_i)$, the sphere centered on x_i, and A_i the thin spherical shell of thickness $2R$ that surrounds S_i, given by $A_i := \overline{B(x_0, cT_i + R)\backslash B(x_0, cT_i - R)}$. Then:

(i) $L(x_0)$ reaches a local minima over the whole sphere S_i. When x_0 is located inside A_i, the subvectors W_i and $F^i_{x_0}$ of W and F_{x_0} respectively become almost colinear because f_t^{R} is radially symmetric. They become exactly colinear when $x_0 \in S_i$. This maximizes the term $\langle F^i_{x_0}, W_i\rangle$ in virtue of the Cauchy-Schwarz inequality. When x_0 lies on one and only one of those spherical shells A_i, the other terms $\langle F^j_{x_0}, W_j\rangle$ are all zero.

(ii) The local minima of $L(x_0)$ located at the intersection of 2 or more spheres S_i are smaller. More generally, when I is a subset of $\{1, ..., q\}$ and when $x_0 \in \bigcap_{i\in I} A_i \backslash \bigcap_{j\notin I} A_j$, the term $\sum_{i\in I}\langle F^i_{x_0}, W_i\rangle$ is (almost) maximized while $\sum_{j\notin I}\langle F^j_{x_0}, W_j\rangle = 0$, which explains why the intersection of spheres are darker coloured than the other parts of the spheres in Figure 1.

(iii) The spheres S_i intersect at a single point, which is exactly x_0^* as well as the global minima of $L(x_0)$. The quantity $r(x_0)$ reaches a global maximum when all subvectors W_i and $F_{x_0}^i$ are colinear, which is the case only when $x_0 \in \bigcap_i S_i$. When there are at least 4 sensors, the intersection of all the spheres $\bigcap_i S_i$ is reduced to at most one point. Recall that we have assumed that $c = c^*$: this implies that $x_0^* \in \bigcap_i S_i$, and thus the minimum of $L(x_0)$ is located at $x_0 = x_0^*$.

Note that if the speed c in $k_{x_0}^{\mathrm{R}}$ does not correspond to the real speed c^*, the intersection $\bigcap_i S_i$ will be empty. Additionally, from an optimization point of view, numerically solving $\inf_{x_0} \mathcal{L}(x_0, \lambda)$ is obviously highly non convex and none of our numerical experiments lead to the correct solution.

3.4 Initial condition reconstruction and error bounds

3.4.1 Initial condition reconstruction procedure. Consider a set of space locations $(x_i)_{1 \leqslant i \leqslant q}$ and moments $(t_j)_{1 \leqslant j \leqslant N}$ (imagine q sensors each collecting measurements at time t_j for all j). Consider now the following inverse problem

$$\begin{gathered}\text{Build and approximation of } u_0 \text{ and } v_0 \text{ from a finite set of measurements} \\ \{w(x_i, t_j)\}_{i,j} \text{ where } (w, u_0, v_0) \text{ are subject to (3.1).}\end{gathered} \tag{3.52}$$

We now show that WIGPR provides an answer to the problem (3.52). This is not surprising, because the covariance models described in the previous section were derived by putting GP priors over u_0 and v_0. As already observed in Section 3.2.1, performing GPR on any data with kernel (3.8) automatically produces a prediction $\tilde{m}$ that verifies $\Box \tilde{m} = 0$ in the sense of distributions. Therefore, this function $\tilde{m}$ is the solution ofthe Cauchy problem (3.1) for some initial conditions $\tilde{u}_0$ and $\tilde{v}_0$:

$$\tilde{m}(x,t) = (F_t * \tilde{v}_0)(x) + (\dot{F}_t * \tilde{u}_0)(x). \tag{3.53}$$

These initial conditions are simply given by $\tilde{u}_0(x) = \tilde{m}(x, 0)$ and $\tilde{v}_0(x) = \partial_t \tilde{m}(x, 0)$. If the data $\{w(x_i, t_j)\}_{i,j}$ on which GPR is performed is comprised of observations of a function w that is another solution of problem (3.1), the initial conditions $(\tilde{u}_0, \tilde{v}_0)$ can be understood as approximations of the initial conditions (u_0, v_0) corresponding to w. More precisely, following Section 2.2.3, we have $\tilde{m} = p_F(w)$ and thus

$$\tilde{u}_0(x) = \tilde{m}(x, 0) = p_F(w)(x, 0) \qquad \forall x \in \mathbb{R}^3, \tag{3.54}$$

$$\tilde{v}_0(x) = \partial_t \tilde{m}(x, 0) = \partial_t p_F(w)(x, 0) = p_F(\partial_t w)(x, 0) \qquad \forall x \in \mathbb{R}^3, \tag{3.55}$$

where F denotes the finite dimensional space $\mathrm{Span}(k_{\mathrm{w}}(z_1, \cdot), ..., k_{\mathrm{w}}(z_n, \cdot))$ and p_F is the orthogonal projector on F w.r.t. the Hilbert space structure of $H_{k_{\mathrm{w}}}$. Here, z_m is of the form $z_m = (x_i, t_k) \in \mathbb{R}^4$. This use of WIGPR provides a flexible framework for tackling the problem (3.52), as the sensors are not constrained in number or location by any integration formula such as Radon transforms. Taking a look at equations (3.54) and (3.55), we can qualitatively discuss the matter of optimal sensor locations for WIGPR. Indeed, we expect that $\tilde{m}$ will provide a better approximation of w when the functions $k_{\mathrm{w}}(z_i, \cdot)_{i=1,\ldots,n}$ are as orthogonal as possible in $\mathcal{H}_{k_{\mathrm{w}}}$, since $\tilde{m}$ is an orthogonal projection on F w.r.t. the $\mathcal{H}_{k_{\mathrm{w}}}$ inner product. The optimal situation is when given two different sensors x_i and x_j, the following should hold for most times t_k, t_l:

$$\langle k_{\mathrm{w}}((x_i,t_k),\cdot), k_{\mathrm{w}}((x_j,t_l),\cdot)\rangle_{\mathcal{H}_{k_{\mathrm{w}}}} = k_{\mathrm{w}}((x_i,t_k),(x_j,t_l)) \ll 1. \tag{3.56}$$

A close inspection of the explicit covariance expressions (equations (52) and (53) from [25]) shows that the property (3.56) can be obtained for most times t_k and t_l when the sensors are far apart from each other, as soon as the kernels k_{u} and k_{v} are such that $k(x,x') \longrightarrow 0$ when $|x-x'| \longrightarrow +\infty$ (which is common, see e.g. the kernel (4.1)). Computing optimal sensor locations and obtaining quantitative guaranties of the accuracy of the reconstruction provided by WIGPR is a hard question left for future research.

3.4.2 Time-dependent error bounds in terms of the initial condition reconstructions. Now that we have showed that WIGPR provides approximations of the initial conditions of (3.1), we underline the fact that these initial condition reconstructions induce a control of the spatial error between the target function u and the Kriging mean $\tilde{m}$, at all times. Indeed, we have the following L^p control in terms of the initial condition reconstruction error. Given $p \in [1,+\infty]$, denote $W^{1,p}(\mathbb{R}^3)$ the Sobolev space of functions $f \in L^p(\mathbb{R}^3)$ whose weak derivatives $\partial_{x_i} f, 1 \leqslant i \leqslant d$, exist and lie in $L^p(\mathbb{R}^3)$.

Proposition 8. *For any $p \in [1,+\infty]$ and any pair $v_0 \in L^p(\mathbb{R}^3)$, $u_0 \in W^{1,p}(\mathbb{R}^3)$ we have the following L^p estimates for all $t \in \mathbb{R}$:*

$$||F_t * v_0||_p \leqslant |t|\ ||v_0||_p, \tag{3.57}$$

$$||\dot{F}_t * u_0||_p \leqslant ||u_0||_p + C_p c|t|\ ||\nabla u_0||_p, \tag{3.58}$$

where $C_p = \left(\int_S |\gamma|_q^p \ d\Omega/4\pi\right)^{1/p} \leqslant 3^{1/q} \leqslant 3, 1/p+1/q = 1$ $(C_\infty = 1, C_1 \leqslant 1)$. Assume that the correct speed c is known and plugged in k_{w}, equations (3.57) *and* (3.58) *then lead to the following L^p error estimate between the target w and its approximant $\tilde{m}$:*

$$||w(\cdot,t)-\tilde{m}(\cdot,t)||_p \leqslant |t|\ ||v_0-\tilde{v}_0||_p + ||u_0-\tilde{u}_0||_p + C_p c|t|\ ||\nabla(u_0-\tilde{u}_0)||_p, \tag{3.59}$$

where $\tilde{u_0}$ and $\tilde{v}_0$ are defined in (3.54) *and* (3.55)*, and $\tilde{m}$ is given in equation* (3.53)*.*

Equations (3.57) and (3.58) are simple stability estimates for the 3D wave equation, although we have not found them in that form in the literature (notably the explicit control constants $|t|$ and $C_p c|t|$). They fall in the category of Strichartz estimates with L^p control for the space variable and L^∞ control for the time variable. We thus provide a proof of the Proposition 8.

Proof of Proposition 8. We have $(F_t * v_0)(x) = t\int_S v_0(x-c|t|\gamma)d\Omega/4\pi$, where $d\Omega/4\pi$ is the normalized Lebesgue measure on the unit sphere S. Assume first that $p \in [1,+\infty[$. Jensen's inequality on the function $t \longmapsto |t|^p$ yields

$$\begin{aligned}||F_t * v_0||_p^p = t^p \int_{\mathbb{R}^3} |(F_t * v_0)(x)|^p dx &= |t|^p \int_{\mathbb{R}^3} \left| \int_S v_0(x-c|t|\gamma)\frac{d\Omega}{4\pi}\right|^p dx \\ &\leqslant |t|^p \int_{\mathbb{R}^3}\int_S |v_0(x-c|t|\gamma)|^p \frac{d\Omega}{4\pi}dx = |t|^p \int_S\int_{\mathbb{R}^3} |v_0(x-c|t|\gamma)|^p dx \frac{d\Omega}{4\pi} \\ &\leqslant \int_S ||v_0||_p^p \frac{d\Omega}{4\pi} = |t|^p ||v_0||_p^p,\end{aligned} \tag{3.60}$$

which yields equation (3.57). Next,

$$
\begin{aligned}
(\dot{F}_t * u_0)(x) = \partial_t (F_t * u_0)(x) &= \partial_t \left(t \int_S u_0(x - c|t|\gamma) \frac{d\Omega}{4\pi} \right) \\
&= \int_S u_0(x - c|t|\gamma) \frac{d\Omega}{4\pi} + t \int_S -c\gamma \cdot \nabla u_0(x - c|t|\gamma) \frac{d\Omega}{4\pi} \quad =: I_1(x) + I_2(x).
\end{aligned}
$$

The functions I_1 and I_2 are defined in the equation above. We have $||\dot{F}_t * u_0||_p = ||I_1 + I_2||_p \leqslant ||I_1||_p + ||I_2||_p$. As in (3.60), $||I_1||_p \leqslant ||u_0||_p$. From Jensen's inequality,

$$
||I_2||_p^p = |ct|^p \int_{\mathbb{R}^3} \left| \int_S \gamma \cdot \nabla u_0(x - c|t|\gamma) \frac{d\Omega}{4\pi} \right|^p dx \leqslant |ct|^p \int_{\mathbb{R}^3} \int_S |\gamma \cdot \nabla u_0(x - c|t|\gamma)|^p \frac{d\Omega}{4\pi} dx.
$$

Next, we use Hölder's inequality in $\mathbb{R}^3$: $|\gamma \cdot \nabla u_0| \leqslant |\nabla u_0|_p \times |\gamma|_q$ with $1/p + 1/q = 1$, where $|v|_p = (|v_1|^p + |v_2|^p + |v_3|^p)^{1/p}$ and likewise for $|v|_q$. Thus,

$$
\begin{aligned}
||I_2||_p^p &\leqslant c^p |t|^p \int_{\mathbb{R}^3} \int_S |\nabla u_0(x - c|t|\gamma)|_p^p \times |\gamma|_q^p \frac{d\Omega}{4\pi} dx \\
&\leqslant c^p |t|^p \int_S |\gamma|_q^p \int_{\mathbb{R}^3} |\nabla u_0(x - c|t|\gamma)|_p^p dx \frac{d\Omega}{4\pi} = c^p |t|^p \left(\int_S |\gamma|_q^p \frac{d\Omega}{4\pi} \right) ||\nabla u_0||_p^p.
\end{aligned}
$$

which yields equation (3.57). Finally, the case $p = +\infty$ is trivial. Equation (3.59) is then the result of equations (3.57) and (3.58) applied to the function

$$
w(x,t) - \tilde{m}(x,t) = [F_t * (v_0 - \tilde{v}_0)](x) + [\dot{F}_t * (u_0 - \tilde{u}_0)](x).
$$

This finishes the proof. $\square$

Equation (3.59) shows that L^p approximations of the initial conditions provide an L^p control between the solution w and the approximation $\tilde{m}$, for any time t. This is one reason why in our numerical applications (Section 4), we focus on initial condition reconstruction.

When c is unknown and estimated by $\hat{c}$ through maximizing the log marginal likelihood, we have instead (highlighting the dependence in c by writing $F_t^c = \sigma_{c|t|}/4\pi c^2 t$)

$$
\begin{aligned}
||w(\cdot,t) - \tilde{m}(\cdot,t)||_p &= ||F_t^c * u_0 - F_t^{\hat{c}} * \tilde{u}_0 + \dot{F}_t^c * v_0 - \dot{F}_t^{\hat{c}} * \tilde{v}_0||_p \\
&= ||F_t^c * (u_0 - \tilde{u}_0) + (F_t^c - F_t^{\hat{c}}) * \tilde{u}_0 + \dot{F}_t^c * (v_0 - \tilde{v}_0) + (\dot{F}_t^c - \dot{F}_t^{\hat{c}}) * \tilde{v}_0||_p,
\end{aligned}
$$

and thus

$$
\begin{aligned}
||w(\cdot,t) - \tilde{m}(\cdot,t)||_p \leqslant& |t| \; ||v_0 - \tilde{v}_0||_p + ||u_0 - \tilde{u}_0||_p + C_p c|t| \; ||\nabla(u_0 - \tilde{u}_0)||_p \\
&+ ||(F_t^c - F_t^{\hat{c}}) * \tilde{u}_0||_p + ||(\dot{F}_t^c - \dot{F}_t^{\hat{c}}) * \tilde{v}_0||_p. \qquad (3.61)
\end{aligned}
$$

The terms containing $F_t^c - F_t^{\hat{c}}$ and $\dot{F}_t^c - \dot{F}_t^{\hat{c}}$ may be further controlled in terms of $|c - \hat{c}|$ with additional assumptions such as Lipschitz continuity of u_0 and v_0. Likewise, the quantity $||w(\cdot,t) - \tilde{m}(\cdot,t)||_p$ may be further controlled if additional assumptions are made on u_0 and/or v_0. We leave such results to the interested reader.

4 Numerical experiments

In this section, we showcase WIGPR on functions w that are solutions of Problem (3.1), using the kernels (3.34) and (3.35) separately as well as together, as in equation (3.8). The goal is twofold: reconstructing the target function w, which more or less amounts to reconstructing its initial conditions (Proposition 8), and estimating the physical parameters attached. Note that with badly estimated physical parameters, the reconstruction step is more or less bound to fail, especially with inaccurate wave speed c and/or source centers x_0^{u} and x_0^{v}.

Running an extensive numerical study of the capabilities and limitations of WIGPR is a large task in itself. For the time being we will settle for simple test cases; in particular we only consider compactly supported and radially symmetric initial conditions, for which we can use the formulas (3.34) and (3.35) which can be evaluated numerically with a low computational cost. We will denote with a star the corresponding true source position x_0^* and celerity c^*. whereas their starless counterpart will denote the hyperparameters of the WIGPR kernels. The estimated hyperparameters will be denoted with a hat, e.g. $\hat{c}$.

Two test cases for WIGPR are considered here. A first test case for $k_{\mathrm{u}}^{\mathrm{wave}}$ described in Subsection 4.1, for which $u_0 \neq 0$ and $v_0 = 0$. This would correspond to PAT, though real life PAT test cases would be very unlikely to enjoy radial symmetry properties. A second test case for $k_{\mathrm{u}}^{\mathrm{wave}} + k_{\mathrm{v}}^{\mathrm{wave}}$ described in Subsection 4.2, for which $u_0 \neq 0$ and $v_0 \neq 0$. For each test case, the full procedure described below is performed.

Numerical simulation and database generation Given initial conditions u_0 and v_0, we numerically simulate the solution w over a given time period. We use a basic two step explicit finite difference time domain (FDTD) numerical scheme for the wave equation as described in [6], equation A.24, over the cube $[0,1]^3$. We also use first order Engquist-Majda transparent boundary conditions [15], in order to mimic a full space simulation. We use a sample rate $SR = 200\ Hz$ (time step $\Delta t = 1/200$ s), a space step $\Delta x = 43\ mm$, and a wave speed $c^* = 0.5\ m/s$. The simulation duration is $T = 1.5\ s$.

30 sensors are scattered in the cube $[0.2, 0.8]^3$ using a Latin hypercube repartition and a minimax space filling algorithm. Signal outputs correspond to time series for each sensor, with a sample rate of 50 Hz, so 75 data points altogether spanned over the time interval $[0, T]$ for each sensor. This leads to $30 \times 75 = 2250$ observations. Each signal is then polluted by a centered Gaussian white noise with standard deviation $\sigma_{\mathrm{noise}} = 0.45$ (resp. 0.09) for the test case #1 (resp. test case #2). These values correspond to around 10% of the maximal amplitude of the signals, see Figures 2a and 4c.

Perform WIGPR on simulated data We perform WIGPR on portions of the dataset obtained above, using the `kergp` package [12] from R [42]. For that we use kernels (3.34) and/or (3.35) which are "fast" to evaluate, with K_{v} and k_{u}^0 both 1D 5/2−Matérn kernels. This Matérn kernel is stationary and writes, in term of the increment $h = x - x'$,

$$k_{5/2}(h) = \sigma^2\big(1 + |h|/\rho + |h|^2/3\rho^2\big)\exp\big(-|h|/\rho\big). \tag{4.1}$$

It has two hyperparameters on its own, ρ and σ^2. ρ is the length scale of the kernel (4.1) and should correspond to the typical variation length scale of the function approximated with GPR; σ^2 is the variance of the kernel. We tackle two different questions related to WIGPR which are respectively the estimation of physical parameters and the sensitivity to sensor locations.

(P_1) We first study how well the physical parameters (c^*, x_0^*, R^*) can be estimated with WIGPR. For this, we first select N_s time series corresponding to the first N_s sensors with $N_s \in \{3, 5, 10, 15, 20, 25, 30\}$. The corresponding Kriging database contains $75 \times N_s$ data points. For this database, we perform negative log marginal likelihood minimization to estimate the corresponding hyperparameters, which are

$$\theta = \begin{cases} (x_0^{\mathrm{u}}, R_{\mathrm{u}}, \theta_{k_{\mathrm{u}}^0}, c, \lambda) \in \mathbb{R}^8 & \text{if } v_0 = 0 \text{ and } u_0 \neq 0, \\ (x_0^{\mathrm{u}}, R_{\mathrm{u}}, \theta_{k_{\mathrm{u}}^0}, x_0^{\mathrm{v}}, R_v, \theta_{k_{\mathrm{v}}^0}, c, \lambda) \in \mathbb{R}^{14} & \text{if } v_0 \neq 0 \text{ and } u_0 \neq 0. \end{cases}$$

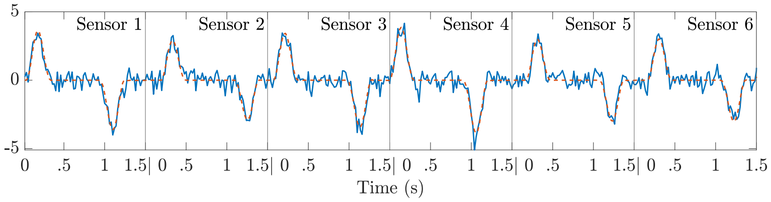
λ corresponds to σ^2 in Section 2.2.2, and is viewed as an additional hyperparameter in the log marginal likelihood. We use a COBYLA optimization algorithm to optimize $\mathcal{L}(\theta, \lambda)$ and a multistart procedure with $n_{\mathrm{mult}} = 100$ different starting points. That is, 100 different values of θ_0 are scattered over an hypercube $H \subset \mathbb{R}^8$ or $H \subset \mathbb{R}^{14}$, and the COBYLA log marginal likelihood optimization procedure is run using each value of θ_0 as a starting point. The resulting hyperparameter value providing the minimal negative log marginal likelihood is selected. The multistart procedure mitigates the risk of getting stuck in local maxima. COBYLA is a gradient-free optimization method used in `kergp` and is available in the `nloptr` package from R. We then reconstruct the initial conditions using WIGPR, which we evaluate in terms of the indicators in equation (4.2).

(P_2) Next, we study the sensibility of the reconstruction step with respect to the sensor locations. Consider 40 different Latin hypercube layouts of the 30 sensors, each obtained with a minimax space filling algorithm. For each layout, we provide the correct set of hyperparameter values to the model; these values are described in each test case. We then reconstruct the initial conditions using GPR and N_s sensors, with $N_s \in \{3, 5, 10, 15, 20, 25, 30\}$. L^p relative errors (see equation (4.2)) are computed between the reconstructed initial condition and the real initial condition. For each number of sensors N_s, statistics over the 40 different datasets for these L^p errors are summarized in boxplots (see e.g. Figure 3a). Each box plot shows the median, the first and the third quartiles of a dataset corresponding to results obtained on the 40 different receiver dispositions. The dots inside a circle correspond to the median of each boxplot. The black crosses are the mean of each box plot, which are linked together with the dashed line. The circles are outliers.

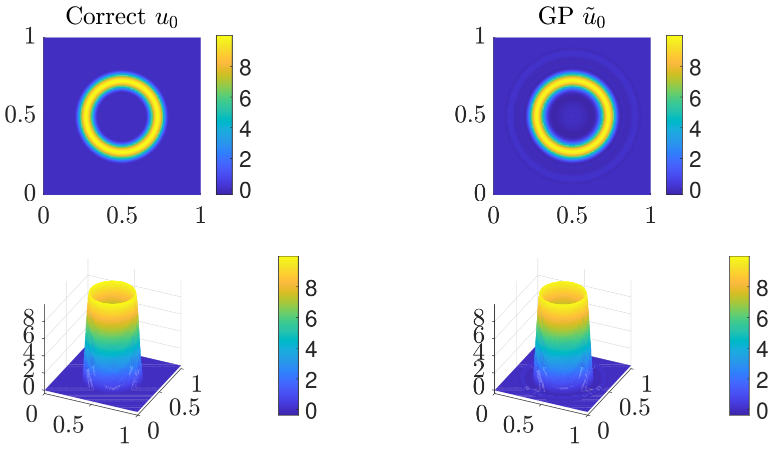
In both cases, the approximated initial position $\tilde{u}_0$ is recovered by evaluating the WIGPR Kriging mean at $t = 0$ over a 3D grid and the initial speed $\tilde{v}_0$ is recovered by evaluating the Kriging mean at $t = 0$ and $t = \Delta t = 10^{-7}$ over the same 3D grid: $\tilde{v}_0 \simeq (\tilde{m}(\cdot, \Delta t) - \tilde{m}(\cdot, 0))/\Delta t$. Figures are displayed using MATLAB [34].

Numerical indicators For (P_1), we indicate in Tables 1 and 2 the distances between the true physical parameters and the estimated ones, depending on the number of sensors used. Additionally, for every $p \in \{1, 2, \infty\}$, we indicate relative L^p reconstruction errors $e_{p,\mathrm{rel}}$ defined below depending on the number of sensors used:

$$e_{p,\mathrm{rel}}^{\mathrm{u}} = ||u_0 - \tilde{u}_0||_p / ||u_0||_p \quad \text{and} \quad e_{p,\mathrm{rel}}^{\mathrm{v}} = ||v_0 - \tilde{v}_0||_p / ||v_0||_p. \tag{4.2}$$



(a) Test case #1, excerpt of captured signals



(b) Test case #1: True u_0 (left column) vs WIGPR u_0 (right column). 15 sensors were used. The images correspond to the 3D functions evaluated at $z = 0.5$.

Figure 2: Visualization of signal and WIGPR results for the test case #1

A relative error of over 100% means that $||u_0 - \tilde{u}_0||_p \geqslant ||u_0||_p$, in which case the trivial estimator $\hat{u}_0 = 0$ performs better than the estimator $\tilde{u}_0$, in the L^p sense. Note that we deal with three dimensional functions, for which approximation errors are typically larger than for their one dimensional counterpart. Thus, relatively large errors may still correspond to pertinent approximations. For (P_2) are plotted boxplots of the relative L^p errors over the 40 different sensor layouts, depending on the number of sensors used. Integrals for the L^p error plots are approximated using Riemann sums over a sufficiently large 3D grid with space step $dx = 0.01$.

The datasets, the code for generating the datasets and the code for performing WIGPR are available at the following address:

`https://github.com/iain-pl-henderson/wave_gpr`

4.1 Test case for $k_{\mathrm{u}}^{\mathrm{wave}}$

In this test case, v_0 is assumed null and thus we set $k_{\mathrm{v}} = 0$, which yields $k_{\mathrm{v}}^{\mathrm{wave}} = 0$. We thus use $k_{\mathrm{u}}^{\mathrm{wave}}$ defined in (3.35) for GPR. We use the 1D Matérn kernel (4.1) for k_{u}^0 in equation (3.35). The initial condition u_0 is a radial ring cosine described as follows. We set $x_0^* = (0.5, 0.5, 0.5)^T$, $R_1 = 0.15, R_2 = 0.3$ and $A = 5$, the corresponding initial conditions (IC) are given by $v_0(x) = 0$ and

$$u_0(x) = A\mathbb{1}_{[R_1,R_2]}(|x - x_0^*|)\left(1 + \cos\left(\frac{2\pi(|x - x_0^*| - \frac{R_1+R_2}{2})}{R_2 - R_1}\right)\right).$$

See Figure 2b, left column, for a graphical representation of u_0. See Figure 2a for an excerpt of the corresponding Kriging database. For problem (P_1), the optimization domain is chosen to be the following hypercube of $\mathbb{R}^8$

$$\begin{aligned} \theta &= (x_0, R, \rho, \sigma^2, c, \lambda) \\ &\in [0,1]^3 \times [0.03, 0.5] \times [0.02, 2] \times [0.1, 5] \times [0.2, 0.8] \times [10^{-8}, 1]. \end{aligned} \tag{4.3}$$

For problem (P_2), the hyperparameter θ_0 provided to the model is

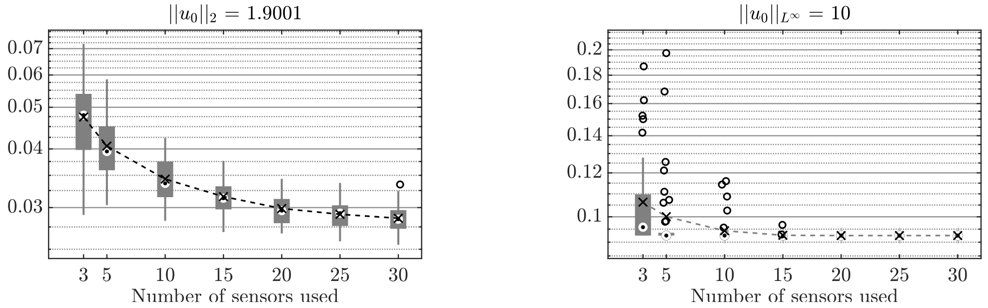
$$\theta_0 = (x_0, R, (\rho, \sigma^2), c, \lambda) = ((0.65, 0.3, 0.5), 0.3, (0.2, 3), 0.5, \sigma^2_{\text{noise}}), \tag{4.4}$$

with $\sigma^2_{\text{noise}} = 0.45^2 = 0.2025$. The value of 0.2 provided for ρ is a visual estimation of the length scale of u_0 based on Figure 2b.

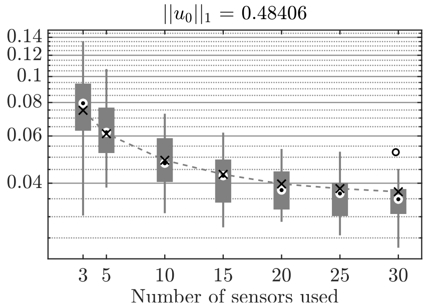
4.1.1 Discussion on the numerical results. For problem (P_1), Table 1 shows that the physical parameters x_0 and c are well estimated. The source size parameter R is overestimated, as could be expected from Section 3.2.5. The relative errors show that the overall function reconstruction is overall satisfying, with relative errors below 15% for $N_s = 20, 25$. The noise level is often overestimated. For problem (P_2) (figures 3a, 3b and 3c), the relative errors stagnate below 10%. The IQR (interquartile range, i.e. the difference between the 3^{rd} and the 1^{st} quartiles) remains below 2%. This means that for this test case, the reconstruction step is not very sensitive to the sensors layout when they are scattered as a Latin hypercube.

N_{sensors}	3	5	10	15	20	25	30	Target
$\lvert\hat{x_0} - x_0^*\rvert$	0.204	0.003	0.004	0.008	0.003	0.004	0.015	0
$\hat{R}_{\text{u}}$	0.386	0.432	0.462	0.431	0.414	0.471	0.452	0.25
$\lvert\hat{c} - c^*\rvert$	0.084	0.004	0.005	0.005	0.006	0.001	0.004	0
$\hat{\sigma}^2_{\text{noise}}$	0.917	0.879	0.93	0.99	0.361	0.988	0.377	0.2025
$\hat{\rho}$	0.02	0.02	0.025	0.02	0.035	0.024	0.032	~ 0.05
$\hat{\sigma}^2$	2.367	3.513	4.903	3.168	4.446	4.619	4.79	?
$e^{\text{u}}_{1,\text{rel}}$	1.275	0.157	0.128	0.168	0.11	0.103	0.248	0
$e^{\text{u}}_{2,\text{rel}}$	1.056	0.095	0.082	0.124	0.088	0.064	0.213	0
$e^{\text{u}}_{\infty,\text{rel}}$	1.037	0.132	0.128	0.198	0.136	0.101	0.321	0

Table 1: Hyperparameter estimation and relative errors, test case #1



(a) L^2 rel. error for u_0 (test case #1) (b) L^∞ rel. error for u_0 (test case #1)



(c) L^1 rel. error for u_0 (test case #1)

Figure 3: Box plots for the sensibility analysis, test case #1

4.2 Test case for $k_{\mathrm{u}}^{\mathrm{wave}} + k_{\mathrm{v}}^{\mathrm{wave}}$

For this test case, the initial position is a raised cosine, while the initial speed is a ring cosine. We set $x_0^{\mathrm{u}*} = (0.65, 0.3, 0.5)^T$, $R_u = 0.25$, $A_{\mathrm{u}} = 2.5$, $x_0^{\mathrm{v}*} = (0.3, 0.6, 0.7)^T$, $R_1^{\mathrm{v}} = 0.05$, $R_2^{\mathrm{v}} = 0.15$ and $A_{\mathrm{v}} = 30$. The corresponding IC are given by

$$\begin{cases} u_0(x) &= A_{\mathrm{u}} \mathbb{1}_{[0,R_u]}(|x - x_0^{\mathrm{u}*}|) \left(1 + \cos\left(\frac{\pi|x - x_0^{\mathrm{u}*}|}{R_u}\right)\right), \\ v_0(x) &= A_{\mathrm{v}} \mathbb{1}_{[R_1^{\mathrm{v}},R_2^{\mathrm{v}}]}(|x - x_0^{\mathrm{v}*}|) \left(1 + \cos\left(\frac{2\pi\left(|x - x_0^{\mathrm{v}*}| - \frac{R_1^{\mathrm{v}}+R_2^{\mathrm{v}}}{2}\right)}{R_2^{\mathrm{v}} - R_1^{\mathrm{v}}}\right)\right). \end{cases}$$

See Figures 4a and 4b, left columns, for graphical representations of u_0 and v_0. See Figure 4c for a visualization of the database. For problem (P_1), the optimization domain is chosen to be the following hypercube

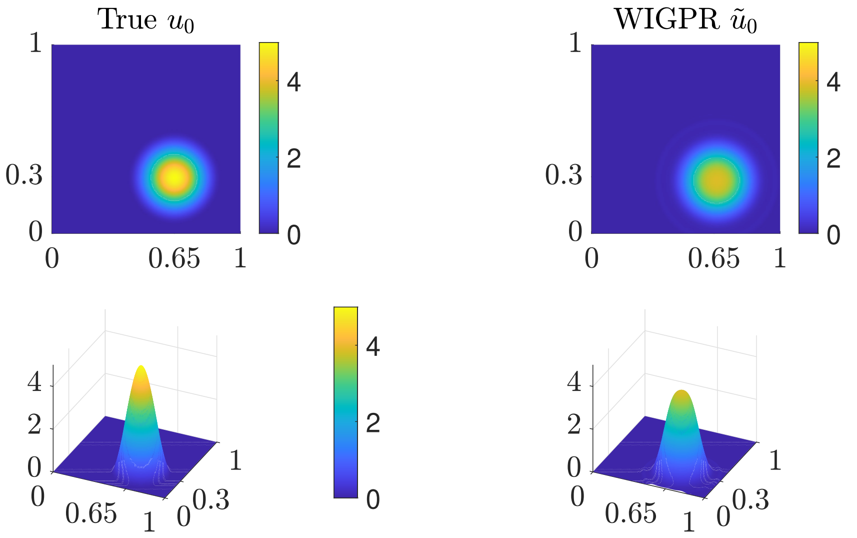
$$\begin{aligned} \theta =& (x_0^{\mathrm{u}}, R_u, (\rho_{\mathrm{u}}, \sigma_{\mathrm{u}}^2), x_0^{\mathrm{v}}, R_{\mathrm{v}}, (\rho_{\mathrm{v}}, \sigma_{\mathrm{v}}^2), c, \lambda) \\ &\in [0,1]^3 \times [0.05, 0.4] \times [0.02, 2] \times [0.1, 5] \\ &\times [0,1]^3 \times [0.05, 0.4] \times [0.02, 2] \times [0.1, 5] \times [0.2, 0.8] \times [10^{-8}, 2 \times 10^{-2}]. \end{aligned} \tag{4.5}$$

For problem (P_2), the hyperparameter value θ_0 provided to the model is

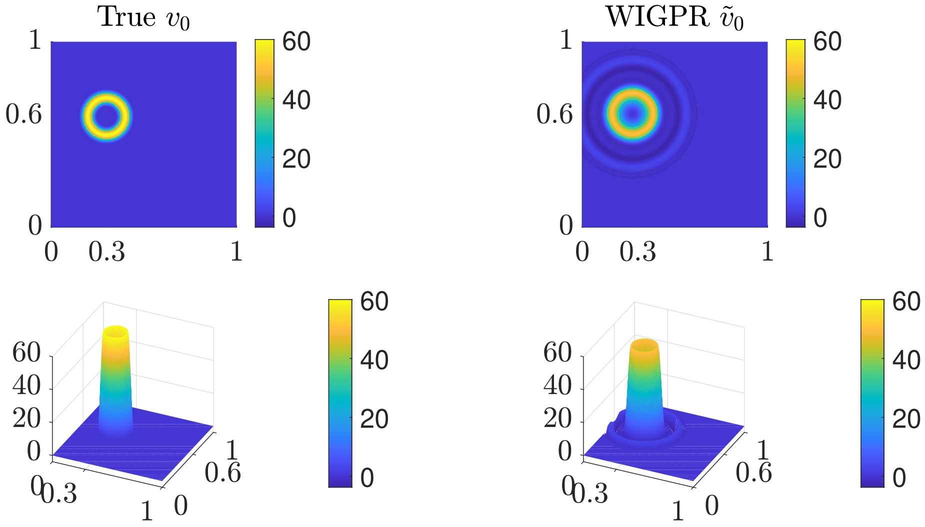
$$\theta_0 = ((0.65, 0.3, 0.5), 0.3, (0.06, 3), (0.3, 0.6, 0.7), 0.15, (0.025, 3.5), 0.5, \sigma_{\mathrm{noise}}^2), \tag{4.6}$$

with $\sigma_{\mathrm{noise}}^2 = 0.0081$. The provided values for $(\rho_{\mathrm{u}}, \sigma_{\mathrm{u}}^2)$ and $(\rho_{\mathrm{v}}, \sigma_{\mathrm{v}}^2)$ are the estimated values from (P_1).

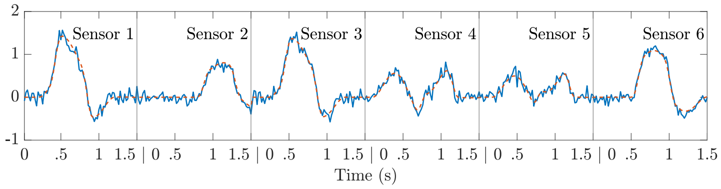
4.2.1 Discussion of the numerical results. Table 2 shows that the physical parameters x_0^{u}, x_0^{v} and c are well estimated. The source radii R_{u} and R_{v} are overestimated, as expected from Section 3.2.5. The noise level is generally overestimated. The reconstruction of the initial position u_0 yielded satisfactory results with L^2 and L^∞ relative errors below 25%, and an L^1 relative error below 35% ($N_s = 10, 15, 20, 25, 30$). The higher L^1 relative error means that the reconstructed function $\tilde{u}_0$ is supported on a larger set than the true function u_0, as the L^1 norm favours sparsity. For the initial speed v_0, the numerical indicators are not as good, reaching minimal values for $N_s = 25$. The corresponding errors for the L^1, L^2 and L^∞ errors are 64%, 28% and 64% respectively. Note though that Figure 4b (corresponding to $N_s = 20$) shows that WIGPR still managed to capture the ring structure of v_0; the corresponding L^1 error for $N_s = 20$ is 150% (Table 2), confirming that the misestimated support radius R_{v} is heavily penalized by the L^1 norm. The reconstruction of v_0 for $N_s = 30$ failed (Table 2). For problem (P_2), the numerical indicators are better. For u_0, Figures 5a, 5c and 5e show that relative error medians stagnate below 5% for $N_s \geqslant 15$. The corresponding IQR are around 2%. For v_0 (Figures 5b, 5d and 5f), the L^1, L^2 and L^∞ relative error medians stagnate at $30\%, 25\%$ and 40% respectively. The corresponding IQR stagnate at $10\%, 5\%$ and 10% respectively.



(a) True u_0 vs WIGPR u_0. The images correspond to the 3D solutions evaluated at $z = 0.5$.

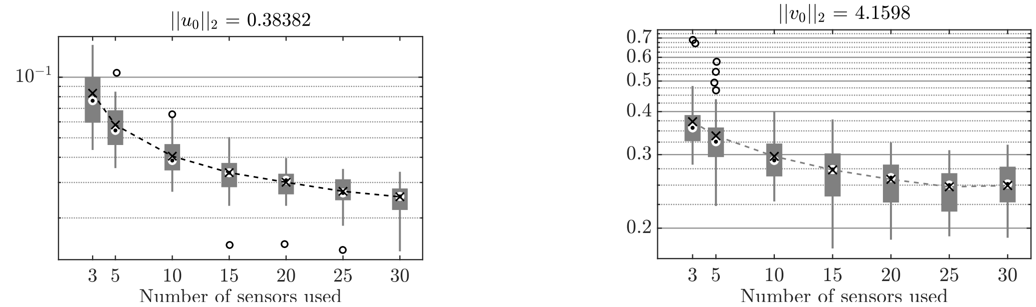


(b) True v_0 vs WIGPR v_0. The images correspond to the 3D solutions evaluated at $z = 0.7$.



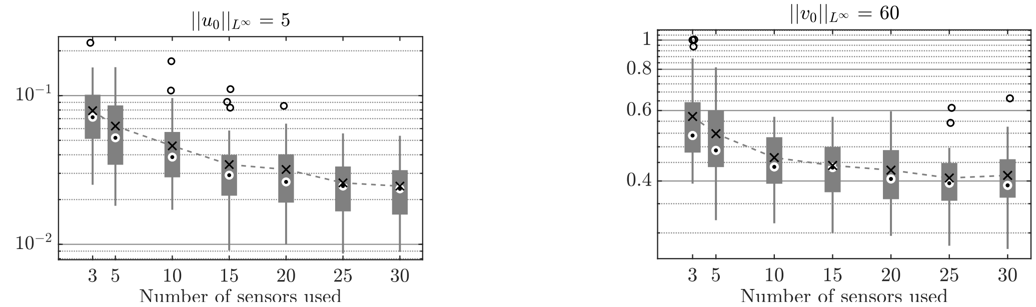
(c) Test case #1, excerpt of captured signals

Figure 4: Test case #2: top and lateral view of the reconstructions of u_0 (Figure 4a) and v_0 (Figure 4b) provided by WIGPR, in comparison with u_0 and v_0. Left columns: true IC. Right columns: WIGPR IC reconstructions. 20 sensors were used.



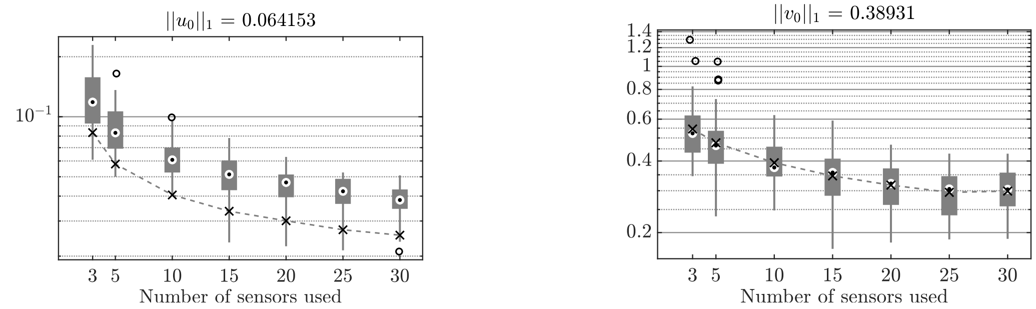
(a) L^2 rel. error for u_0 (test case #2)

(b) L^2 rel. error for v_0 (test case #2)



(c) L^∞ rel. error for u_0 (test case #2)

(d) L^∞ rel. error for v_0 (test case #2)



(e) L^1 rel. error for u_0 (test case #2)

(f) L^1 rel. error for v_0 (test case #2)

Figure 5: Box plots for the sensibility analysis, test case#2

N_{sensors}	3	5	10	15	20	25	30	Target
$\lvert\hat{x}_0^{\mathrm{u}} - x_0^{\mathrm{u}}{}^*\rvert$	0.163	0.144	0.013	0.024	0.023	0.033	0.015	0
$\hat{R}_{\mathrm{u}}$	0.4	0.274	0.384	0.309	0.352	0.286	0.313	0.25
$\lvert\hat{x}_0^{\mathrm{v}} - x_0^{\mathrm{v}}{}^*\rvert$	0.163	0.18	0.035	0.028	0.037	0.006	0.05	0
$\hat{R}_{\mathrm{v}}$	0.252	0.166	0.313	0.356	0.348	0.266	0.339	0.15
$\lvert\hat{c} - c^*\rvert$	0.165	0.156	0.028	0.036	0.042	0.011	0.04	0
$\hat{\sigma}^2_{\mathrm{noise}}$	0.0178	0.0184	0.0188	0.0161	0.0187	0.0145	0.0116	0.0081
$\hat{\rho}_{\mathrm{u}}$	0.034	0.069	0.102	0.027	0.031	0.061	0.034	~ 0.05
$\hat{\sigma}^2_{\mathrm{u}}$	4.649	4.472	4.575	2.493	0.678	3.272	2.541	?
$\hat{\rho}_{\mathrm{v}}$	0.057	0.027	0.044	0.053	0.085	0.022	0.012	~ 0.02
$\hat{\sigma}^2_{\mathrm{v}}$	3.91	2.538	3.05	1.545	4.886	3.575	4.346	?
$e^{\mathrm{u}}_{1,\mathrm{rel}}$	2.414	1.676	0.243	0.311	0.358	0.315	0.317	0
$e^{\mathrm{u}}_{2,\mathrm{rel}}$	1.276	1.053	0.174	0.223	0.228	0.261	0.205	0
$e^{\mathrm{u}}_{\infty,\mathrm{rel}}$	0.732	0.608	0.136	0.174	0.231	0.212	0.228	0
$e^{\mathrm{v}}_{1,\mathrm{rel}}$	2.865	2.796	1.315	1.42	1.51	0.645	9.784	0
$e^{\mathrm{v}}_{2,\mathrm{rel}}$	1.492	1.812	0.694	0.616	0.736	0.284	35.75	0
$e^{\mathrm{v}}_{\infty,\mathrm{rel}}$	1.083	1.608	0.817	0.763	0.845	0.635	2416.682	0

Table 2: Hyperparameter estimation and relative errors, test case #2

5 Conclusion and perspectives

In Section 3, we described several Gaussian process priors tailored to the wave equation, which may then be used in the context of Gaussian process regression (WIGPR). These priors are particular cases of general covariance formulas which were first derived in a previous work. The priors we have studied in this article correspond to the cases where either stationarity or radial symmetry assumptions over the initial conditions hold. In that framework, the physical parameter of the PDE system (e.g. source location or wave celerity) can be interpreted as hyperparameters of the WIGPR prior, as in [43]. We then showed that in the limit of the small source radius, the multilateration method for point source localization was naturally recovered by the hyperparameter estimation step of WIGPR. We furthermore showed that WIGPR naturally provides a reconstruction of the initial conditions of the wave equation, as should be expected when putting priors over them.

The radial symmetry WIGPR formulas from Section 3 were then showcased in Section 4, where two practical questions were tackled. First, WIGPR can correctly estimate certain physical parameters attached to the corresponding wave equation, namely the wave speed and source position. When these parameters are well estimated, WIGPR is capable of providing non trivial reconstructions of the initial condition, which we studied in terms of L^1, L^2 and L^∞ relative errors. The second question consisted in studying the sensibility of the reconstruction step of WIGPR w.r.t. the sensor locations. we observed that when the sensors are spread according to a Latin hypercube, the reconstruction step is not very sensitive to the layout of the sensors, assuming that the correct set of hyperparameters is provided to the model.

Future possible investigations concern the practical use of the more general formula (3.8) without any radial symmetry assumptions, e.g. for PAT applications. To compute the convolutions efficiently, one may then resort to multidimensional fast Fourier transforms. Moreover, in this first study, we have only used simple methods for GP numerical evaluation. More advanced GP techniques such as inducing points [41] should now be used to handle large size datasets such as the ones we have used in Section 4.

The case of the two dimensional wave equation is also of practical interest, e.g. in oceanography [32], and presents many different properties than its 3D counterpart ([16], p. 80). It would thus deserve a theoretical and practical study in its own right when coupled with GPR.

Acknowledgements. Research of all the authors was supported by SHOM (Service Hydrographique et Océanographique de la Marine) project “Machine Learning Methods in Oceanography” no-20CP07. The authors thank Rémy Baraille in particular for his personal involvement in the project.

References

[1] C. G. Albert and K. Rath. Gaussian process regression for data fulfilling linear differential equations with localized sources. *Entropy*, 22(2), 2020.

[2] P. A. Alvarado, M. A. Alvarez, G. Daza-Santacoloma, A. Orozco, and G. Castellanos-Dominguez. A latent force model for describing electric propagation in deep brain stimulation: A simulation study. In *36th Annu. Conf. Proc. IEEE Eng. Med. Biol. Soc.*, pages 2617–2620, 2014.

[3] M. Álvarez, D. Luengo, and N. Lawrence. Linear latent force models using Gaussian processes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35:2693–2705, 2013.

[4] H. Ammari, editor. *Mathematical modeling in biomedical imaging II. Optical, ultrasound, and opto-acoustic tomographies*, volume 2035 of *Lecture Notes in Mathematics*. Springer, Berlin, Heidelberg, 2012.

[5] M. A. Anastasio, J. Zhang, D. Modgil, and P. J. La Rivière. Application of inverse source concepts to photoacoustic tomography. *Inverse Problems*, 23(6):S21–S35, 2007.

[6] S. Bilbao. *Wave and Scattering Methods for Numerical Simulations*. John Wiley & Sons, Ltd, 2004.

[7] V. I. Bogachev. *Gaussian measures*. Number 62 in Mathematical Surveys and Monographs. American Mathematical Soc., 1998.

[8] Y. Chen, B. Hosseini, H. Owhadi, and A. M. Stuart. Solving and learning nonlinear PDEs with Gaussian processes. *J. Comput. Phys.*, 447:Paper No. 110668, 29, 2021.

[9] S. L. Cotter, M. Dashti, and A. M. Stuart. Approximation of Bayesian inverse problems for PDEs. *SIAM J. Numer. Anal.*, 48(1):322–345, 2010.

[10] M. Dashti, K. J. Law, A. M. Stuart, and J. Voss. Map estimators and their consistency in bayesian nonparametric inverse problems. *Inverse Problems*, 29(9):095017, 2013.

[11] M. Dashti and A. M. Stuart. The Bayesian approach to inverse problems. In R. Ghanem, D. Higdon, and H. Owhadi, editors, *Handbook of Uncertainty Quantification*, pages 311–428, Cham, 2017. Springer International Publishing.

[12] Y. Deville, D. Ginsbourger, and O. R. C. N. Durrande. *kergp: Gaussian Process Laboratory*, 2021. R package version 0.5.5.

[13] D. G. Duffy. *Green’s functions with applications*. Chapman and Hall/CRC, second edition edition, 2015.

[14] J. J. Duistermaat and J. A. C. Kolk. *Distributions*, pages 33–44. Birkhäuser Boston, Boston, 2010.

[15] B. Engquist and A. Majda. Absorbing boundary conditions for the numerical simulation of waves. *Math. Comp.*, 31(139):629–651, July 1977.

[16] L. Evans. *Partial Differential Equations.* Graduate studies in mathematics. American Mathematical Society, 1998.

[17] L. C. Evans and R. F. Garzepy. *Measure theory and fine properties of functions, Revised Edition (1st ed.).* Chapman and Hall/CRC, 2015.

[18] B. T. Fang. Trilateration and extension to global positioning system navigation. *J. Guid. Control Dyn.*, 9(6):715–717, 1986.

[19] G. Fasshauer. Meshfree approximation methods with MATLAB. In *Interdisciplinary Mathematical Sciences*, 2007.

[20] E. J. Fuselier Jr. *Refined error estimates for matrix-valued radial basis functions.* PhD thesis, Texas A&M University, 2007.

[21] D. Ginsbourger, O. Roustant, and N. Durrande. On degeneracy and invariances of random fields paths with applications in Gaussian process modelling. *J. Statist. Plann. Inference*, page 170 :117 – 128, 2016.

[22] T. Graepel. Solving Noisy Linear Operator Equations by Gaussian Processes: Application to Ordinary and Partial Differential Equations. In *Proc. 20th Int. Conf. Mach. Learn.*, pages 234–241. AAAI Press, 2003.

[23] C. Grossmann, H.-G. Roos, and M. Stynes. *Numerical treatment of partial differential equations.* Springer, 2007.

[24] M. Gulian, A. Frankel, and L. Swiler. Gaussian process regression constrained by boundary value problems. *Comput. Methods Appl. Mech. Engrg.*, 388:114117, 2022.

[25] I. Henderson, P. Noble, and O. Roustant. Characterization of the second order random fields subject to linear distributional PDE constraints. preprint, Sept. 2022.

[26] S. Janson. *Gaussian Hilbert Spaces.* Cambridge Tracts in Mathematics. Cambridge University Press, 1997.

[27] C. Jidling, J. Hendriks, N. Wahlstrom, A. Gregg, T. Schon, C. Wensrich, and A. Wills. Probabilistic modelling and reconstruction of strain. *Nucl. Instrum. Methods Phys. Res. B: Beam Interact. Mater. At.*, 436:141–155, 2018.

[28] C. Jidling, N. Wahlström, A. Wills, and T. B. Schön. Linearly constrained Gaussian processes. In *Adv. Neural Inf. Process. Syst.*, volume 30. Curran Associates, Inc., 2017.

[29] P. Kuchment and L. Kunyansky. Mathematics of photoacoustic and thermoacoustic tomography. In *Handbook of mathematical methods in imaging. Vol. 1, 2, 3*, pages 1117–1167. Springer, New York, 2015.

[30] M. Lange-Hegermann. Algorithmic linearly constrained Gaussian processes. In *Adv. Neural Inf. Process. Syst.*, volume 31. Curran Associates, Inc., 2018.

[31] M. Lange-Hegermann. Linearly constrained Gaussian processes with boundary conditions. In *Proc. of The 24th Int. Conf. Artif. Intell. Stat.*, volume 130 of *Proc. of Mach. Learn. Res.*, pages 1090–1098. PMLR, 13–15 Apr 2021.

[32] D. Lannes and P. Bonneton. Derivation of asymptotic two-dimensional time-dependent equations for surface water wave propagation. *Phys. Fluids*, 21(1):016601, 2009.

[33] A. F. López-Lopera, N. Durrande, and M. Álvarez. Physically-inspired Gaussian process models for post-transcriptional regulation in drosophila. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, 18:656–666, 2021.

[34] The Mathworks, Inc., Natick, Massachusetts. *MATLAB version 9.8.0.1721703 (R2020a) Update 7*, 2020.

[35] F. M. Mendes and E. A. da Costa Júnior. Bayesian inference in the numerical solution of Laplace's equation. *AIP Conf. Proc.*, 1443(1):72–79, 2012.

[36] F. J. Narcowich and J. Ward. Generalized Hermite interpolation via matrix-valued conditionally positive definite functions. *Math. Comp.*, 63:661–687, 1994.

[37] H. Owhadi. Bayesian numerical homogenization. *Multiscale Model. Simul.*, 13(3):812–828, 2015.

[38] M. Á. P. A. Alvarado and A. Orozco. A three spatial dimension wave latent force model for describing excitation sources and electric potentials produced by deep brain stimulation. *arXiv*, 2016.

[39] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes 3rd edition: The art of scientific computing.* Cambridge university press, 2007.

[40] Z. Purisha, C. Jidling, N. Wahlström, T. B. Schön, and S. Särkkä. Probabilistic approach to limited-data computed tomography reconstruction. *Inverse Problems*, 35(10):105004, sep 2019.

[41] J. Quiñonero Candela and C. E. Rasmussen. A unifying view of sparse approximate Gaussian process regression. *J. Mach. Learn. Res.*, 6:1939–1959, 2005.

[42] R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2020.

[43] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Machine learning of linear differential equations using Gaussian processes. *J. Comput. Phys.*, 348:683–693, 2017.

[44] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Numerical Gaussian processes for time-dependent and nonlinear partial differential equations. *SIAM J. Sci. Comput.*, 40(1):A172–A198, 2018.

[45] C. E. Rasmussen and C. Williams. *Gaussian Processes for Machine Learning.* the MIT Press, 2006.

[46] S. Särkkä. Linear operators and stochastic partial differential equations in Gaussian process regression. In *Artificial Neural Networks and Machine Learning – ICANN 2011*, pages 151–158, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[47] S. Särkkä, M. Álvarez, and N. Lawrence. Gaussian process latent force models for learning and stochastic control of physical systems. *IEEE Trans. on Automat. Control*, 64:2953–2960, 2019.

[48] R. Schaback. Solving the Laplace equation by meshless collocation using harmonic kernels. *Adv. Comput. Math.*, 31:457–470, 2009.

[49] M. Scheuerer and M. Schlather. Covariance models for divergence-free and curl-free random vector fields. *Stoch. Models*, 28:433 – 451, 2012.

[50] A. Solin and M. Kok. Know your boundaries: Constraining gaussian processes by variational harmonic features. In *Proc.* 22^{nd} *Int. Conf. Artif. Intell. Stat.*, volume 89 of *Proc. of Mach. Learn. Res.*, pages 2193–2202. PMLR, 16–18 Apr 2019.

[51] A. M. Stuart. Inverse problems: A Bayesian perspective. *Acta Numer.*, 19:451–559, 2010.

[52] F. Treves. *Topological Vector Spaces, Distributions and Kernels.* Dover books on mathematics. Dover Publications, 2006.

[53] R. C. Vergara, D. Allard, and N. Desassis. A general framework for SPDE-based stationary random fields. *Bernoulli*, 28(1):1–32, 2022.

[54] N. Wahlstrom, M. Kok, T. B. Schön, and F. Gustafsson. Modeling magnetic fields using Gaussian processes. *Proc. - ICASSP IEEE Int. Conf. Acoust. Speech Signal Process.*, pages 3522–3526, 2013.

[55] H. Wendland. *Scattered data approximation*, volume 17. Cambridge university press, 2004.

[56] M. Xu and L. V. Wang. Universal back-projection algorithm for photoacoustic computed tomography. *Phys. Rev. E*, 71(1):016706, 2005.

[57] M. Álvarez, D. Luengo, and N. D. Lawrence. Latent force models. In *Proc.* 12^{th} *Int. Conf. Artif. Intell. Stat.*, volume 5 of *Proc. of Mach. Learn. Res.*, pages 9–16, Hilton Clearwater Beach Resort, Clearwater Beach, Florida USA, 16–18 Apr 2009. PMLR.