Characterization of the second order random fields subject to linear distributional PDE constraints

Iain Henderson [*1], Pascal Noble[1], and Olivier Roustant[1]

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

Abstract

Let L be a linear differential operator acting on functions defined over an open set $\mathcal{D} \subset \mathbb{R}^d$. In this article, we characterize the measurable second order random fields $U = (U(x))_{x \in \mathcal{D}}$ whose sample paths all verify the partial differential equation (PDE) $L(u) = 0$, solely in terms of their first two moments. When compared to previous similar results, the novelty lies in that the equality $L(u) = 0$ is understood in the *sense of distributions*, which is a powerful functional analysis framework mostly designed to study linear PDEs. This framework enables to reduce to the minimum the required differentiability assumptions over the first two moments of $(U(x))_{x \in \mathcal{D}}$ as well as over its sample paths in order to make sense of the PDE $L(U_\omega) = 0$. In view of Gaussian process regression (GPR) applications, we show that when $(U(x))_{x \in \mathcal{D}}$ is a Gaussian process (GP), the sample paths of $(U(x))_{x \in \mathcal{D}}$ conditioned on pointwise observations still verify the constraint $L(u) = 0$ in the distributional sense. We finish by deriving a simple but instructive example, a GP model for the 3D linear wave equation, for which our theorem is applicable and where the previous results from the literature do not apply in general.

1   Introduction

When dealing with an unknown function of interest $u : \mathcal{D} \to \mathbb{R}$ where say $\mathcal{D} \subset \mathbb{R}^d$, it is common (as e.g. in Bayesian inverse problems) to assume that it is a sample path of a random field $U = (U(x))_{x \in \mathcal{D}}$. Incorporating prior knowledge over u, such as smoothness, is then achieved by constraining the law of U accordingly. Sometimes, this prior knowledge comes from physical considerations. If u describes a positive quantity such as mass or energy, then the random variables $U(x)$ should all be positive almost surely (a.s.). In many cases, this physical constraint can be more precisely translated as a partial differential equation (PDE). Such equations are a pivotal tool for modelling, understanding and predicting real-life phenomena such as those arising from fluid mechanics, electromagnetics or biology to name a few. The most simple (yet central) PDEs are those that are linear. In this article, we will only consider homogeneous linear PDEs, which take the form

$$L(u) := \sum_{|\alpha| \leqslant n} a_\alpha(x) \partial^\alpha u = 0. \tag{1}$$

---

*henderso@insa-toulouse.fr

Key words Generalized functions, Linear constraints, Linear Partial Differential Equations, Second order random fields

Above, u is the unknown function of interest, defined over an open set $\mathcal{D} \subset \mathbb{R}^d$, and L is a linear partial differential operator. In (1), for a multi-index $\alpha = (\alpha_1, ..., \alpha_d)^T \in \mathbb{N}^d$, we used the notations $|\alpha| = \alpha_1 + ... + \alpha_d$ and $\partial^\alpha = (\partial_{x_1})^{\alpha_1}...(\partial_{x_d})^{\alpha_d}$. Homogeneous PDEs, i.e. PDEs with a null term on the right-hand side of (1), are often encountered to describe conservation laws, such as conservation of mass, energy or momentum in closed systems [52].

In order to incorporate the knowledge that $L(u) = 0$ in the prior U, a natural question is whether one can characterize, in terms of their law, the random fields U whose sample paths are all solutions to the PDE (1). Let U be a centered second order random field with covariance function k: under the assumption that U is a Gaussian process (GP) whose sample paths are n times differentiable, [25] proved for some classes of differential operators L of order n that ([25], Sections 3.3 and 4.1)

$$\mathbb{P}(L(U) = 0) = 1 \iff \forall x \in \mathcal{D}, L(k(x, \cdot)) = 0. \tag{2}$$

This property provides a simple characterization of the GPs that incorporate the PDE constraint (1) sample path-wise. Such GPs would fall in the category of "physics-informed" GPs in the machine learning community. In the proof of this property, the fact that the sample paths are n times differentiable, i.e. that the PDE (1) can be understood *pointwise*, is central. These functions are then *strong* solutions of the PDE (1) (see Definition 3.1).

In the standard PDE approach though, equation (1) is reinterpreted by weakening the definition of the derivatives of u, thereby weakening the required regularity assumptions over u. It can indeed happen in practice that the sought solutions of the PDE $L(u) = 0$ are not n times differentiable or even continuous (see e.g. [16], Section 2.1), and they are only solutions of some weakened formulation of equation (1). This is typically the case for hyperbolic PDEs such as the wave equation presented in Section 4. We introduce here the distributional formulation of the PDE (1), where the regularity assumptions over u are relaxed to the maximum. As such, this formulation enables working with potentially singular solutions of equation (1), solutions which are not allowed to appear in more restrictive functional frameworks (see also the upcoming Remark 3.4). Another advantage of the distributional formulation is that it provides a unifying framework for dealing with linear PDEs, independently of their nature. In contrast, traditional weak or variational formulations vary greatly depending on the nature of the PDE. As an illustration, in [16], one can compare the different function spaces for weak solutions associated to elliptic PDEs (Section 6.1.2), parabolic PDEs (Section 7.1.1(b)) and hyperbolic PDEs (Section 7.2.1(b)). The distributional formulation will be our main object of interest in this article, and can be seen as a weakened form of weak formulations of PDEs. Consider equation (1), and "test it locally": that is, multiply it by a compactly supported, infinitely differentiable test function φ (i.e. $\varphi \in C_c^\infty(\mathcal{D})$) and integrate over $\mathcal{D}$:

$$\forall \varphi \in C_c^\infty(\mathcal{D}), \quad \sum_{|\alpha| \leqslant n} \int_{\mathcal{D}} \varphi(x) a_\alpha(x) \partial^\alpha u(x) dx = 0. \tag{3}$$

For each integral term above, perform $|\alpha|$ successive integrations by parts to transfer the derivatives from u to φ. Since φ is identically null on a neighbourhood of the boundary of $\mathcal{D}$, the boundary terms of each integration by parts vanish and we obtain that

$$\forall \varphi \in C_c^\infty(\mathcal{D}), \quad \int_{\mathcal{D}} u(x) \sum_{|\alpha| \leqslant n} (-1)^{|\alpha|} \partial^\alpha (a_\alpha \varphi)(x) dx = 0. \tag{4}$$

To make sense of (4), one only requires u to be locally integrable, i.e. $\int_K |u(x)| dx < +\infty$ for all compact set $K \subset \mathcal{D}$. We then say that a locally integrable function u is a solution to $L(u) = 0$ in the sense of distributions, or distributional sense, if u verifies (4). In this case, u is a solution

to equation (1) in the sense of "all smooth local averages" (i.e. for all $\varphi \in C_c^\infty(\mathcal{D})$), though not pointwise in general: taking $\varphi(x) = \delta_0(x - x_0)$ is not allowed without additional assumptions over u.

The distributional formulation of the PDE $L(u) = 0$ is "compliant with physics" too, as pointed out by W. Rudin ([46], p. 150): most of the sensors we use in practice are only capable of computing local averages of the physical quantity they are measuring. Suppose one wishes to check experimentally that a temperature field obeys the heat equation, by using a set of thermometers: then one will actually only deal with the distributional formulation of the heat equation.

The natural question that follows from this new definition is whether one can characterize, in terms of their law, the random fields whose sample paths are solutions to the PDE $L(u) = 0$ *in the distributional sense*. The answer is yes, and is the main content of this article. Under the assumptions that U is a measurable centered second order random field and that its standard deviation function $\sigma : x \longmapsto \sqrt{k(x, x)}$ is locally integrable, we show in Proposition 3.5 that

$$\mathbb{P}(L(U) = 0 \text{ in the distrib. sense}) = 1 \iff \forall x \in \mathcal{D}, L(k(x, \cdot)) = 0 \text{ in the distrib. sense.} \quad (5)$$

Related literature

It is known, at least since the fifties, that some covariance functions are naturally linked to certain stochastic partial derivative equations (SPDEs), i.e. PDEs where the source term is random. For example, it was already observed in 1954 by [56] that the covariance function of a stationary GP U verifying the two dimensional SPDE $(\alpha^2 - \partial_{xx}^2 - \partial_{yy}^2)^{3/4} U = W_S$, where W_S is a spatial white noise process, is exponential, i.e. of the form $\text{Cov}(U(x + h), U(x)) = C \exp(-\alpha|h|)$. Already for this SPDE, the differentiation has to be understood in a weakened sense as white noise processes are not random fields in the usual sense. In [45], SPDEs describing the random motion of micro-particles are introduced to link certain covariance functions, Matérn in particular, with an underlying physical model. We also refer to [36] for a large overview of the possible applications and recent developments pertaining to random fields defined by SPDEs. A general framework for the study of SPDEs was recently reintroduced in [8], which was then used to classify the stationary *generalized* random fields that are solutions of a wide class of linear SPDEs. In particular, [8] provides a description of all the second order stationary generalized random fields that are solutions to certain homogeneous PDEs, and the 3D wave equation in particular (which we also study in Section 4), in terms of their covariance operator. Loosely speaking, generalized random fields are function-indexed random fields where the covariance function is replaced by a covariance operator. From a functional analysis point of view, this is actually very close to the tools we use here, although in this article we constrain ourselves to work with (standard) random fields with well-defined sample paths, as these are the objects that arise the most in the random function models met in practice. The two other key differences between this work and [8] are that (i) we do not focus on *stationary* random field models for u and (ii) we focus on the *homogeneous* case for PDE $L(u) = 0$.

The literature concerning random fields that are PDE-constrained at the level of the sample paths is rather sparse. In [25], general theorems are exposed for many different classes of linear operators acting on suitable spaces of functions. These theorems take the form of equation (2), and can in turn be applied to certain differential operators (see [25], Sections 3.3 and 4.1). [50] builds covariance functions that ensure that the sample path of a given two or three dimensional random vector field are either divergence or curl free. This result is notable because "any" three dimensional vector field can be decomposed as a sum of divergence and curl free vector fields through the Helmholtz-Hodge decomposition theorem. Moreover, divergence or curl free vector fields are commonly encountered in fluid mechanics. [20] extends the results of [50] to random

3

fields on the sphere of $\mathbb{R}^3$, which has been rediscovered later in [18]. In [15], stationary GPs are represented in terms of a random wavevector. [15] then characterizes the stationary GPs whose sample paths verify a homogeneous linear PDE, in terms of the spectral measure of the GP and in terms of its random wavevector. [15] additionally requires that the sample paths be infinitely differentiable, that the PDE's coefficients be constant and that only even orders of differentiation appear in the PDE. This is then applied to a few wave models. A simple algorithm for building linearly constrained GPs is proposed in [30], based on formal GPR derivations upon (1); however, partly because the assumed regularity of u is not fully addressed, the claim that the sample paths of the underlying GP are indeed linearly constrained is left unproved. This is clarified in [33], where the requirement that $u \in C^\infty(\mathbb{R}^d)$ is made explicit and the enforcement of the PDE on the sample paths is proved for GPs whose sample paths are smooth. The algorithm from [30] is then supplemented in [33], where parametrizations of the solution spaces of (1) thanks to Gröbner bases are proposed. In [34], the same author completes the approach from [33] by incorporating boundary conditions on hypersurfaces in the Gröbner basis parametrization. With the idea to apply GPR to rigid body dynamics, [23] enforces Gauss' principle of least constraint on the sample paths of a GP.

One can understand our main result (Proposition 3.5) as a characterization of the "physics-informed" random fields that incorporate the distributional PDE constraint $L(u) = 0$ at the level of the sample paths. It turns out that the design of similar "physics-informed priors" has received a lot a attention from the machine learning community since the early 2000' ([26]), in the context of Gaussian process regression (GPR); see Section 3.3.1 for a description of this technique. GPR is a Bayesian framework for function regression and interpolation which is well suited for handling linear constraints, partly because GPs are "stable under linear combinations", see Section 2.1.3. The recurring idea is to assume that the function u in equation (1) is a sample path of a (centered) GP U and to draw the consequences of equation (1) on the covariance function of U. The covariance function of U is then expected to incorporate the constraint $L(u) = 0$ in some sense. The majority of these works (except those mentioned above) do not aim at analysing whether the obtained covariance function indeed yields sample path PDE constraints over U: this is justified by the fact that they are only concerned with imposing the constraints on the function provided by GPR to approximate u. This approximation of u, which we denote by $\tilde{m}$, is called the Kriging mean in the GPR context; see equation (33) for a definition.

While they do not primarily focus on investigating sample path PDE constraints (contrarily to this article), the works coming from the GPR community are still very connected to this article. Indeed, they are concerned with designing explicit covariance functions that verify constraints of the form $L(k(\cdot, x)) = 0$ for all $x \in \mathcal{D}$ (the PDE is understood in the strong sense in these works). Indeed, this constraint ensures that all the possible regression functions $\tilde{m}$ provided by the corresponding GPR model verify the constraint $L(\tilde{m}) = 0$ (as seen in equation (33)). Note that "$L(k(\cdot, x)) = 0 \ \forall x \in \mathcal{D}$" is the right-hand side of equation (2): actually knowing covariance functions that verify this constraint is a necessary complement to the condition we prove in this article (Proposition 3.5 is otherwise useless in practice). Explicit PDE constrained covariance functions were designed for a number of classical PDEs, namely: divergence-free vector fields [38, 50], curl-free vector fields [21, 50], the Laplace equation [47, 37, 3], Maxwell's equations [55, 30, 33] (although [55, 30] only exploit curl/divergence free constraints), the 1D heat equation [3], Helmholtz' 2D equations [3], and linear solid mechanics [29]. [34] and [27] enforce homogeneous boundary conditions on the covariance function.

We finish with a brief overview of the alternative "physics-informed" GPR models. Contrarily to the equation (1) considered here, one may put a random source term f in the PDE and study instead the SPDE $L(u) = f$: see [42], [4] and [40] for entry points on the related literature. A recent article [9] extended the use of GPR to nonlinear PDEs by imposing the nonlinear interpolation constraints on the collocation points, setting the way forward for many possible

applications of GPR to nonlinear realistic PDE models, as found e.g. in fluid mechanics. In [39], the variational formulation (see [16], Section 6.1.2 for a definition) of certain linear PDEs has been incorporated into a GPR framework. This approach requires the use of Gaussian generalized random fields (see [6], Section 2.2.1.1), or "functional Gaussian processes" following [39]. The variational formulation of a PDE differs from its distributional formulation in the choice of the space of test functions.

Contribution and organisation of the paper. Consider the PDE in equation (1), where the coefficients of the differential operator L have possibly limited smoothness. Consider also a centered second order measurable stochastic process $U = (U(x))_{x \in \mathcal{D}}$ with covariance function $k(x, x')$ (see Sections 2.1.2 and 2.1.3). Under the assumption that its standard deviation function $\sigma : x \longmapsto k(x, x)^{1/2}$ is locally integrable, we show in Proposition 3.5 that the announced equation (5) holds. The result is then compared to a previous result from [25], which ensures pointwise linear differential degeneracy of the sample paths of U under stronger assumptions. We then provide a simple corollary which states that linear distributional differential constraints are preserved when a GP U is conditioned on pointwise observations, in view of GPR applications.

As an application example, we derive a general Gaussian process model for the homogeneous 3D free space wave equation, for which the solutions are not smooth in general. This equation is central for describing finite speed propagation phenomena as found e.g. in acoustics. Plugging this model in a GPR framework yields potential applications in different inverse problems related to this PDE, such as thermoacoustic tomography (i.e. initial condition reconstruction, [31], Section 19.3.1.1), source localization or propagation speed estimation, following e.g. the GPR methodology from [42] or [25], Section 4.2.

This model is derived by putting GP priors over the initial conditions of the wave equation and in Proposition 4.1, we obtain "explicit" formulas for the covariance function of the solution process, in the form of convolutions. From Propositions 3.5 and 4.1, we obtain that the sample paths of the corresponding (nonstationary) GP all verify the wave equation in the distributional sense. When the covariance functions of the initial conditions are not smooth enough, the result from [25] cannot be applied. Explicitly, for this PDE, choosing the commonly used 3/2-Matérn covariance functions for the initial position is enough to land outside the scope of the result from [25] (Section 4.2.1).

We emphasize that the covariance expressions exposed in Proposition 4.1 are original and interesting in themselves, as they can be used for efficient GPR for the wave equation. Specifically, the key difference with the wave equation covariance functions presented in [8] is that here, no stationarity assumptions are made on the solution stochastic process U. In particular the spectral measure provided by Bochner's theorem [43], which is the key tool used in [8], is not available anymore. We thus resort to more standard integration techniques to prove Proposition 4.1.

The paper is organized as follow. For self-containment, Section 2 is dedicated to reminders on random fields and generalized functions. This Section and all the proofs are detailed enough so that this article is accessible to the analyst, the probability theorist and the statistician. In Section 3, we state and prove our new necessary and sufficient condition on random fields that are subject to linear distributional differential constraints. Section 4 is dedicated to the study of a GP model for the wave equation. We conclude in Section 5.

2   Background

2.1   Random fields

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. For convenience, we will assume that it is complete, i.e. that $\mathcal{A}$ contains the subsets of sets $A \in \mathcal{A}$ such that $\mathbb{P}(A) = 0$.

2.1.1 Basic definitions. Let $\mathcal{D} \subset \mathbb{R}^d$ be an open set. In this article, a random field $U = (U(x))_{x \in \mathcal{D}}$ is a collection of real random variables defined on Ω. We define its sample path at point $\omega \in \Omega$ to be the deterministic function $x \longmapsto U(x)(\omega)$, and we denote it by U_ω. Given an operator acting on the sample paths of U, an event of the form $\{L(U) \in A\}$ will always be understood sample path wise: that is, by definition, $\{L(U) \in A\} := \{\omega \in \Omega : L(U_\omega) \in A\}$. Such sets are not automatically measurable; still, they are measurable as soon as they contain an event of probability 1 (as the ones in Propositions 3.2 and 3.5), since $(\Omega, \mathcal{A}, \mathbb{P})$ is a complete probability space.

2.1.2 Measurable random fields. In view of our main theorem, a necessary notion is that of the measurability of the random field U. U is said to be measurable ([13], p. 60 or [35], p. 34) if it is measurable seen as a bivariate map $U : (\Omega \times \mathcal{D}, \mathcal{A} \otimes \mathcal{B}(\mathcal{D})) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})), (\omega, x) \mapsto U(x)(\omega)$. Here, $\mathcal{B}(S)$ denotes the Borel σ-algebra of S and $\mathcal{A} \otimes \mathcal{B}(\mathcal{D})$ denotes the product σ-algebra of $\mathcal{A}$ and $\mathcal{B}(\mathcal{D})$.

To work with measurable random fields, one will often consider random fields U which are continuous *in probability*, i.e. for all $x \in \mathcal{D}$ and $\varepsilon > 0, \mathbb{P}(|U(x) - U(x + h)| > \varepsilon) \to 0$ when $h \to 0$. Indeed, continuity in probability implies the existence of a measurable modification of U, i.e. a measurable random field $\tilde{U}$ such that $\mathbb{P}(\tilde{U}(x) = U(x)) = 1$ for all $x \in \mathcal{D}$ ([13], Theorem 2.6, p. 61). One then implicitly works with $\tilde{U}$. In this article, we will directly assume that we deal with measurable random fields instead of assuming any continuity regularity on the sample paths of the said stochastic process. This is because pointwise continuity is not really relevant when working with PDEs in a weak sense; actually, one of the main points of working with weakened formulations is to avoid strong (i.e. pointwise) formulations. Note however that ensuring measurability outside of the above mentioned theorem, though possible, rapidly becomes tedious (see e.g. [12], Theorem 2.3). A famous theorem from Kolmogorov ([10], Theorem 3.3 p. 73 and Theorem 3.4 p. 74) provides sufficient conditions for almost sure continuity of the sample paths, which in turn implies continuity in probability of the random field. This condition is phrased in terms of a sufficient Hölder control of the expectation of the increments of the process. Refinements in the case of Gaussian processes exist: see e.g. [2], Theorem 1.4.1, p. 20. On a final note, the measurability assumption is discussed in [53] (Theorem 3.3), where it is shown to be a necessary condition for the existence of Karhunen-Loève expansions of second order random fields.

2.1.3 Second order random fields, Gaussian processes.. Note $L^2(\mathbb{P})$ the Hilbert space of real valued random variables X such that $\mathbb{E}[X^2] < +\infty$. A stochastic process $(U(x))_{x \in \mathcal{D}}$ is said to be second order if for all $x \in \mathcal{D}$, $U(x) \in L^2(\mathbb{P})$. One can then define its mean and covariance functions by $m(x) = \mathbb{E}[U(x)]$ and $k(x, x') = \mathbb{E}[(U(x) - m(x))(U(x') - m(x'))]$ respectively. One can then also define its standard deviation function

$$\sigma : x \mapsto \sqrt{k(x, x)}. \tag{6}$$

A Gaussian process $(U(x))_{x \in \mathcal{D}}$ over $\mathcal{D}$ is a random field over $\mathcal{D}$ such that for any $(x_1, ..., x_n) \in \mathcal{D}^n$ and any $(a_1, ..., a_n) \in \mathbb{R}^n, \sum_i a_i U(x_i)$ is a Gaussian random variable; that is, 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 ([28], Section 8). We write $(U(x))_{x \in \mathcal{D}} \sim GP(m, k)$. Given a GP $(U(x))_{x \in \mathcal{D}}$, we will sometimes use the space $\mathcal{L}(U) = \overline{\mathrm{Span}(U(x), x \in \mathcal{D})}$, i.e. the Hilbert subspace of $L^2(\mathbb{P})$ induced by U. Since $L^2(\mathbb{P})$-limits of Gaussian random variables drawn from the same GP remain Gaussian ([28], Section 1.3), $\mathcal{L}(U)$ only encompasses Gaussian random variables.

Whereas m can be any function, the covariance function k has to be symmetric and positive definite: for all $(x_1, ..., x_n)$ in $\mathcal{D}^n$, the matrix $(k(x_i, x_j))_{1 \leqslant i, j \leqslant n}$ is symmetric and nonnegative definite.

Symmetric positive definite functions verify the Cauchy-Schwarz inequality [43] :

$$\forall x, x' \in \mathcal{D}, \quad |k(x, x')| \leqslant \sqrt{k(x, x)}\sqrt{k(x', x')}. \tag{7}$$

Note that there is a one-to-one correspondence between positive definite functions and the laws of centered GPs ([13], Theorem 3.1). We provide below two examples of radial Matérn covariance functions ([43], pp. 84-85), which will be useful in Section 4. Set $r = ||x - x'||$, the Euclidean distance between x and x', then the following two functions are valid covariance functions, given any $l > 0$:

$$k_{1/2}(x, x') = \exp(-r/l), \qquad k_{3/2}(x, x') = (1 + r/l) \exp(-r/l). \tag{8}$$

These covariance functions are widely used in machine learning, especially $k_{3/2}$. Almost surely, the sample paths of a GP with a Matérn covariance function k_ν with $\nu = m + 1/2$, $m \in \mathbb{N}$, are of differentiability class C^m and not C^{m+1}. They are thus commonly used to model functions with finite smoothness.

2.2 Tools from functional analysis

We refer to [46] and [54] for further details on generalized functions and Radon measures. In this whole subsection, $\mathcal{D}$ is an open set of $\mathbb{R}^d$.

2.2.1 Class C^m functions, test functions, locally integrable functions. Given $m \in \mathbb{N}$, $C^m(\mathcal{D})$ denotes the space of real-valued functions defined over $\mathcal{D}$ of class C^m, and $C_c^m(\mathcal{D})$ denotes the subspace of $C^m(\mathcal{D})$ of functions φ whose support $\mathrm{Supp}(\varphi)$ is compact. Recall that $\mathrm{Supp}(\varphi)$ is the closure of the set $\{x : \varphi(x) \neq 0\}$. The space $C_c^\infty(\mathcal{D})$, which we will rather denote $\mathscr{D}(\mathcal{D})$, is the space of compactly supported infinitely differentiable functions supported on $\mathcal{D}$, also known as test functions. $L_{loc}^1(\mathcal{D})$ denotes the space of measurable scalar functions f defined on $\mathcal{D}$ that are locally integrable, i.e. such that $\int_K |f| < +\infty$ for all compact sets $K \subset \mathcal{D}$. Two locally integrable functions are equal in $L_{loc}^1(\mathcal{D})$ when they are equal almost everywhere (a.e.) in the sense of the Lebesgue measure over $\mathbb{R}^d$. $L_{loc}^1(\mathcal{D})$ is a very large space which contains the space of piecewise continuous functions, but also all the local Lebesgue spaces $L_{loc}^p(\mathcal{D}), p \geqslant 1$ and thus all the Sobolev spaces of nonnegative exponent. It is in fact the largest space of functions that can be alternatively viewed as continuous linear forms over $\mathscr{D}(\mathcal{D})$ (see Section 2.2.4 below).

2.2.2 Generalized functions. We endow $\mathscr{D}(\mathcal{D})$ with its usual LF-space topology, defined for example in [54], Chapter 13. LF stands for "strict inductive limit of Fréchet spaces". As it will appear in several places later on, we briefly describe the LF topology following [54], although this is not necessary for understanding the article. Assume that a vector space E can be written as $E = \bigcup_n E_n$ where (E_n) is an increasing sequence of Fréchet spaces (i.e. metrizable complete locally convex topological vector spaces), such that the natural injection $E_n \to E_{n+1}$ is a linear homeomorphism over its range. The LF topology over E is defined as follow: a convex set $V \subset E$

is a neighborhood of 0 if and only if $V \cap E_n$ is a neighborhood of 0 for all n. It is remarkable that LF topologies are *not* metrizable except if for some n_0, $E_n = E_{n_0}$ for all $n \geqslant n_0$ ([54], Remark 13.1). In return, this allows for some other very nice topological properties to hold, e.g., LF spaces are complete ([54], Theorem 13.1).

For $\mathscr{D}(\mathcal{D})$, the LF topology is the one corresponding to the decomposition $\mathscr{D}(\mathcal{D}) = \bigcup_i \mathscr{D}_{K_i}(\mathcal{D})$, where $\mathscr{D}_{K_i}(\mathcal{D}) := \{\varphi \in C^\infty(\mathcal{D}) : \mathrm{Supp}(\varphi) \subset K_i\}$, and $(K_i)_{i \in \mathbb{N}}$ is an increasing sequence of compact subsets of $\mathcal{D}$ such that $\bigcup_i K_i = \mathcal{D}$ ([54], pp. 131-133). This LF topology does not depend on the choice of $(K_i)_{i \in \mathbb{N}}$. An example of metric inducing the Fréchet topology of $\mathscr{D}_{K_i}(\mathcal{D})$ is the following:

$$d_i(\varphi, \psi) := \sup_{N \in \mathbb{N}} 2^{-N} \frac{p_{N,i}(\varphi - \psi)}{1 + p_{N,i}(\varphi - \psi)}, \quad p_{N,i}(\varphi) := \max_{|\alpha| \leqslant N} \sup_{x \in K_i} |\partial^\alpha \varphi(x)|. \tag{9}$$

It is given in [46], Section 1.46 p. 34 and Remark 1.38(c) p. 29. Explicitly, a sequence $(\varphi_n) \subset \mathscr{D}(\mathcal{D})$ converges to $\varphi \in \mathscr{D}(\mathcal{D})$ if there exists a compact set $K \subset \mathcal{D}$ such that $\mathrm{Supp}(\varphi_n) \subset K$ for all $n \in \mathbb{N}$ and for all $\alpha \in \mathbb{N}^d$, $||\partial^\alpha \varphi_n - \partial^\alpha \varphi||_\infty \to 0$ ([46], Theorem 6.5(f) and the remark following p. 154).

We call generalized function any continuous linear form on $\mathscr{D}(\mathcal{D})$, i.e. any element of $\mathscr{D}(\mathcal{D})'$, the topological dual of $\mathscr{D}(\mathcal{D})$. We will rather denote it by $\mathscr{D}'(\mathcal{D})$ as in [54], Notation 21.1. The topology of $\mathscr{D}(\mathcal{D})$ is such that $T \in \mathscr{D}'(\mathcal{D})$ if and only if for all compact set $K \subset \mathcal{D}$, there exists $C_K > 0$ and a nonnegative integer n_K such that

$$\forall \varphi \in \mathscr{D}(\mathcal{D}) \text{ such that } \mathrm{Supp}(\varphi) \subset K, \ \ |T(\varphi)| \leqslant C_K \sum_{|\alpha| \leqslant n_K} ||\partial^\alpha \varphi||_\infty. \tag{10}$$

We recall that we use the following notations: for a multi-index $\alpha = (\alpha_1, ..., \alpha_d) \in \mathbb{N}^d$, we denote $|\alpha| = \alpha_1 + ... + \alpha_d$ and $\partial^\alpha := (\partial_{x_1})^{\alpha_1}...(\partial_{x_d})^{\alpha_d}$ where $\partial_{x_i}^{\alpha_i}$ is the α_i^{th} derivative with reference to the ith coordinate x_i. Generalized functions are also called "distributions", a terminology we will only use when there is no risk of confusion with probability distributions. The duality bracket will be denoted $\langle, \rangle$: for all $\varphi \in \mathscr{D}(\mathcal{D})$ and $T \in \mathscr{D}'(\mathcal{D})$, we have $\langle T, \varphi \rangle := T(\varphi)$.

2.2.3 Generalized functions and differentiation. Any generalized function T can be infinitely differentiated ([46], Section 6.12, p. 158 or [54], pp. 248-250) according to the following definition

$$\partial^\alpha T : \varphi \longmapsto \langle T, (-1)^{|\alpha|} \partial^\alpha \varphi \rangle. \tag{11}$$

The derivative $\partial^\alpha T$ is then also a continuous linear form over $\mathscr{D}(\mathcal{D})$, i.e. $\partial^\alpha T \in \mathscr{D}'(\mathcal{D})$.

2.2.4 Regular generalized functions. Any function $f \in L^1_{loc}(\mathcal{D})$ can be injectively identified to a generalized function T_f ([54], p. 224 or [46], Section 6.11, p. 157) defined as follow

$$\forall \varphi \in \mathscr{D}(\mathcal{D}), \quad \langle T_f, \varphi \rangle := \int_{\mathcal{D}} f(x) \varphi(x) dx. \tag{12}$$

The map $L^1_{loc}(\mathcal{D}) \ni f \longmapsto T_f$ is linear and injective; any generalized function T that is of the form T_f for some $f \in^1_{loc}(\mathcal{D})$ is said to be regular. Throughout this article, we will use the abusive notation $\langle T_f, \varphi \rangle = \langle f, \varphi \rangle$, as if $\langle, \rangle$ were the L^2 inner product. Observe that equations (11) and (12) combined provide a flexible definition of the derivatives of any function $f \in L^1_{loc}(\mathcal{D})$ up to any order. One also sees that *weak* derivatives, as encountered in Sobolev spaces ([7], Section 9.1) and weak formulations of PDEs, are particular cases of distributional derivatives: given $\alpha \in \mathbb{N}^d$ and two locally integrable functions f and f_α, f admits f_α for its αth weak derivative if and only if $\partial^\alpha T_f = T_{f_\alpha}$. One then conveniently writes $\partial^\alpha f = f_\alpha$ ($\partial^\alpha f$ is unique in $L^1_{loc}(\mathcal{D})$ from the injectivity of the mapping (12)).

8

2.2.5 Radon measures. This subsection and the ones that follow are only necessary for dealing with the wave equation in Section 4. In this article, we call positive Radon measure any positive measure over $\mathcal{D}$ that is Borel regular ([17], Definition 1.9) and that has finite mass over any compact subset of $\mathcal{D}$. A Radon measure is a linear combination of positive Radon measures. In [32], Chapter 9, it is proved that the space of Radon measures over $\mathcal{D}$ is isomorphic to the space of continuous linear forms over $C_c(\mathcal{D})$, the space of compactly supported continuous functions on $\mathcal{D}$ endowed with its usual LF-space topology described e.g. in [54], pp. 131-133. The corresponding isomorphism is given by

$$\mu \longmapsto T_\mu : \begin{cases} C_c(\mathcal{D}) & \longrightarrow \mathbb{R} \\ f & \longmapsto \int_\mathcal{D} f(x)\mu(dx). \end{cases} \tag{13}$$

We have the following facts. (i) Any signed measure that admits a density f with reference to the Lebesgue measure such that $f \in L^1_{loc}(\mathcal{D})$ is a Radon measure ([54], p. 217). (ii) Any Radon measure can be injectively identified to a generalized function T_μ by replacing $C_c(\mathcal{D})$ by $\mathscr{D}(\mathcal{D})$ in equation (13). In particular, Radon measures can be differentiated up to any order through equation (11). (iii) Any Radon measure μ, can be uniquely written as $\mu = \mu^+ - \mu^-$ where μ^+ and μ^- are positive Radon measures ([32], Chapter 9). We then define its total variation by $|\mu| := \mu^+ + \mu^-$.

2.2.6 Finite order generalized functions. A generalized function T is said to be of finite order if there exists a nonnegative integer m such that one can take $n_K = m$, independently of K, in the definition of the continuity of T, i.e. equation (10). The order of T is then the smallest of those integers m. The space of generalized functions of order m is isomorphic to $C_c^m(\mathcal{D})'$, the space of continuous linear forms over $C_c^m(\mathcal{D})$, when $C_c^m(\mathcal{D})$ is endowed with its usual LF-space topology ([54], pp. 131-133). The key property for us is that such generalized functions can be represented thanks to Radon measures. If L is of order m, there exists a family of Radon measures $\{\mu_p\}_{|p|\leqslant m}$ over $\mathcal{D}$ such that

$$T = \sum_{|p|\leqslant m} \partial^p \mu_p, \tag{14}$$

where the equality in equation (14) holds in $\mathscr{D}'(\mathcal{D})$ and $C_c^m(\mathcal{D})'$ ([54], p 259). Among the finite order generalized functions are those that are *compactly supported*, i.e. those for which the measures μ_p such that $T = \sum_{|p|\leqslant m} \partial^p \mu_p$ all have compact support.

2.2.7 Convolution with generalized functions. As above, we consider $C_c^m(\mathbb{R}^d)$ endowed with its LF-space topology. Let $f \in C_c^m(\mathbb{R}^d)$ and $T \in C_c^m(\mathbb{R}^d)'$. Note $\tau_x f$ the function $y \longmapsto f(y-x)$ and $\check{f}$ the function $y \longmapsto f(-y)$. Then ([54], p. 287, Section 27) one may define the convolution between T and f by

$$T * f : x \longmapsto \langle T, \tau_{-x}\check{f}\rangle, \tag{15}$$

and $T * f$ is a function in the classical sense, i.e. defined pointwise. When T is a regular generalized function, equation (15) reduces to the usual convolution of functions through the identification defined in equation (12). Similarly if T is in fact a Radon measure μ:

$$(T * f)(x) = \int_{\mathbb{R}^d} f(x-y)\mu(dy). \tag{16}$$

More general definitions of generalized function convolution are available ([54], Chapter 27) but this one is sufficient for our use.

2.2.8 Tensor product of generalized functions. For two generalized functions $T_1 \in \mathscr{D}'(\mathcal{D}_1)$ and $T_2 \in \mathscr{D}'(\mathcal{D}_2)$, $T_1 \otimes T_2 \in \mathscr{D}'(\mathcal{D}_1 \times \mathcal{D}_2)$ denotes their tensor product([54], pp. 416-417), which is uniquely determined by the following tensor property:

$$\forall \varphi_1 \in \mathscr{D}(\mathcal{D}_1), \forall \varphi_2 \in \mathscr{D}(\mathcal{D}_2), \ \langle T_1 \otimes T_2, \varphi_1 \otimes \varphi_2 \rangle = \langle T_1, \varphi_1 \rangle \times \langle T_2, \varphi_2 \rangle. \tag{17}$$

$T_1 \otimes T_2$ reduces to the tensor product of functions (respectively, measures) when T_1 and T_2 are functions (respectively, measures) through the identification of equation (12) (respectively, equation (13)).

3 Random fields under linear differential constraints

The results in this section state that under suitable assumptions over the first two moments of a given second order random field $U = (U(x))_{x \in \mathcal{D}}$, sample path degeneracy properties with reference to differential constraints can be read on the first two moments of U, namely the mean function and the functions $k_x : y \longmapsto k(x, y)$, where k is the covariance function of U. This is remarkable because the space induced by the sample paths of U is a priori much larger than the space spanned by the functions $k_x, x \in \mathcal{D}$. Moreover, the functions k_x are "accessible", i.e. checking that these functions indeed verify the linear constraint can usually be done with direct computations.

We begin by recalling a result from [25] in the case of pointwise defined derivatives. We next state and prove a result similar to that of [25], where we interpret the derivatives in the distributional sense.

3.1 The case of classical derivatives

We start by properly defining the notion of strong solutions of a PDE.

Definition 3.1 (Strong/classical solutions). Let L be a differential operator defined as in equation (1), with continuous coefficients. We say that a function u is a classical or strong solution to the PDE $L(u) = 0$ if u is n times differentiable and u verifies the PDE pointwise:

$$\forall x \in \mathcal{D}, \quad L(u)(x) = \sum_{|\alpha| \leqslant n} a_\alpha(x) \partial^\alpha u(x) = 0. \tag{18}$$

Note that the space of n times differentiable functions does not have the nice topological properties of $C^n(\mathcal{D})$ and in most cases met in practice, one rather requires that strong solutions lie in $C^n(\mathcal{D})$. It is however in the sense of the definition 3.1 that the theorem from [25] is best understood. This theorem, which we remind in Proposition 3.2, is the one proved and used in [25] to build a Gaussian process whose sample paths are all strong solutions to the Laplace equation on a 2D circular domain.

We first introduce some notations. Let $(U(x))_{x \in \mathcal{D}}$ be a centered Gaussian process with covariance function k. Denote $\mathcal{F}(\mathcal{D}, \mathbb{R})$ the space of real-valued pointwise-defined functions on $\mathcal{D}$ (often alternatively denoted $\mathbb{R}^{\mathcal{D}}$). We will only use $\mathcal{F}(\mathcal{D}, \mathbb{R})$ as a set, therefore we do not consider any topology over it. We refer to [51], Section 9, for details on $\mathcal{F}(\mathcal{D}, \mathbb{R})$ seen as a topological vector space. Denote $\mathcal{H}_k$ the reproducing kernel Hilbert space (RKHS) associated to k (see [6], Definition 1 p. 7 and Theorem 3, p. 19). $\mathcal{H}_k$ is a Hilbert space of pointwise-defined functions (i.e. $\mathcal{H}_k \subset \mathcal{F}(\mathcal{D}, \mathbb{R})$ as sets), such that the pointwise evaluation maps $l_x : f \mapsto f(x)$ are continuous functionals. Although belonging to $\mathcal{F}(\mathcal{D}, \mathbb{R})$ does not seem very restrictive at first glance, this clashes with the usual L^p and Sobolev spaces encountered in PDE theory, which are sets of functions defined up to a set of null Lebesgue measure.

Proposition 3.2 (sample paths of GPs under linear constraints [25])**.** *Let* $\big(U(x)\big)_{x\in\mathcal{D}} \sim GP(0,k)$. *Note for all* $x \in \mathcal{D}$ *the function* $k_x : y \longmapsto k(x,y)$. *Let* E *be a real vector space of functions defined on* $\mathcal{D}$ *that contains the sample paths of* U *almost surely and* $L : E \longrightarrow \mathcal{F}(\mathcal{D},\mathbb{R})$ *be a linear operator. Assume that for all* $x \in \mathcal{D}$, $L(U)(x) \in \mathcal{L}(U)$, *where* $\mathcal{L}(U)$ *is the closure of* $Span\{U(x), x \in \mathcal{D}\}$ *in* $L^2(\mathbb{P})$. *Then there exists a unique linear operator* $\mathscr{L} : \mathcal{H}_k \longrightarrow \mathcal{F}(\mathcal{D},\mathbb{R})$ *such that*

$$\forall x,y \in \mathcal{D}, \quad \mathbb{E}[L(U)(x)U(y)] = \mathscr{L}(k_y)(x),$$

and such that for all $x \in \mathcal{D}, h \in \mathcal{H}_k$ *and sequence* $(h_n) \subset \mathcal{H}_k$ *such that* $h_n \to h$ *for the topology of* $\mathcal{H}_k$, *we have* $\mathscr{L}(h_n)(x) \longrightarrow \mathscr{L}(h)(x)$. *Finally, the following statements are equivalent:*

(i) $\mathbb{P}(L(U) = 0) = 1$.

(ii) $\forall x \in \mathcal{D}, \mathscr{L}(k_x) = 0$.

A sufficient condition ensuring that the sample paths of a GP lie in $C^n(\mathcal{D})$ is found in [2], Theorem 1.4.2. More broadly, both necessary and sufficient conditions over the first two moments of a GP for its sample path to be (Hölder) continuous are well-known: see e.g. [1], Theorems 3.3.3 and 8.3.2.

The proof of Proposition 3.2 heavily relies on the Loève isometry ([6], Theorem 35, p. 65) between the two Hilbert spaces $\mathcal{H}_k$ and $\mathcal{L}(U)$ (see Section 2.1.3 for details on $\mathcal{L}(U)$). This theorem can be applied when L is a differential operator as discussed in [25]. However, in Proposition 3.2, the differential operator L of order n has to be valued in $\mathcal{F}(\mathcal{D},\mathbb{R})$; in particular for $u \in E$, the function $L(u)$ has to be defined pointwise in order to use the Loève isometry. To summarize, in all generality the derivatives in L have to be understood in a classical sense and E has to be contained in $\mathscr{D}^n(\mathcal{D})$, the space of n times differentiable functions on $\mathcal{D}$. Requiring that $E \subset \mathscr{D}^n(\mathcal{D})$ is a very strong assumption with reference to the sample paths of U; furthermore, this is not compliant with the usual way of studying PDEs where derivatives are understood in a weaker sense. We present in Proposition 3.5 an adaptation of Proposition 3.2 where the derivatives are understood in the distributional sense. By transferring all the derivatives on the test function, we will be liberated from any differentiability assumptions over the sample paths of U, effectively replacing $\mathscr{D}^n(\mathcal{D})$ with $L^1_{loc}(\mathcal{D})$. Finally, the random field U will not be assumed Gaussian and will only be required to be measurable second order.

3.2 The case of distributional derivatives

3.2.1 Distributional solutions of PDEs. In this section, we elaborate a bit more on the notion of distributional solutions to a given PDE. Let $L = \sum_{|\alpha|\leqslant n} a_\alpha(x)\partial^\alpha$ be a linear differential operator, and assume for the moment that its coefficients are infinitely differentiable. We briefly recall the steps described in the introduction that lead to the definition of distributional solutions presented in equation (4). Start from a strong solution u of class C^n of $L(u) = 0$, multiply this PDE by a test function $\varphi \in \mathscr{D}(\mathcal{D})$, integrate over $\mathcal{D}$ and perform $|\alpha|$ integration by parts to transfer all derivatives from u to φ. Since the support of φ is a compact subset of the open set $\mathcal{D}$, the boundary terms of each integration by parts vanish, leading to

$$\forall \varphi \in \mathscr{D}(\mathcal{D}), \quad \int_{\mathcal{D}} u(x) \sum_{|\alpha|\leqslant n} (-1)^{|\alpha|}\partial^\alpha(a_\alpha\varphi)(x)dx = 0. \tag{19}$$

Following equation (19) we introduce L^*, the formal adjoint of L, acting on $\mathscr{D}(\mathcal{D})$, defined by the following formula ([54], pp. 247-249)

$$L^* : \varphi \longmapsto \sum_{|\alpha|\leqslant n} (-1)^{|\alpha|}\partial^\alpha(a_\alpha\varphi). \tag{20}$$

Note that for equation (19) to be well defined, the assumptions that $u \in L^1_{loc}(\mathcal{D})$ and $a_\alpha \in C^{|\alpha|}(\mathcal{D})$ are sufficient. More precisely, these assumptions are enough to show that the map $L(u)$ defined by duality

$$L(u) : \begin{cases} \mathscr{D}(\mathcal{D}) & \longrightarrow \mathbb{R} \\ \varphi & \longmapsto \int_\mathcal{D} L^*(\varphi)(x)u(x)dx \end{cases} \tag{21}$$

defines a continuous linear form over $\mathscr{D}(\mathcal{D})$, i.e. $L(u) \in \mathscr{D}'(\mathcal{D})$ (see equations (10) and (26) for a rigorous proof of this statement). This definition extends the definition of distributional derivatives from Section 2.2.3 to differential operators. By construction, L and L^* verify a duality identity: given $\varphi \in \mathscr{D}(\mathcal{D})$ and $u \in L^1_{loc}(\mathcal{D})$, $\langle L(u), \varphi \rangle = \langle u, L^*(\varphi) \rangle$.

As in Section 2.2.4, the assumption that $u \in L^1_{loc}(\mathcal{D})$ is in fact a *continuity* assumption over the associated linear form $L(u)$ (a more general and theoretical analysis of such observations can be found in [54], pp. 247-251). This finally leads to the following definition, following e.g. [14], p. 10:

Definition 3.3 (Distributional solutions). A function $u \in L^1_{loc}(\mathcal{D})$ is said to be a solution to the PDE $L(u) = 0$ in the sense of distributions if $L(u) = 0$ in $\mathscr{D}'(\mathcal{D})$, i.e. when $L(u)$ is seen as en element of $\mathscr{D}'(\mathcal{D})$ through equation (21) and 0 is the null linear form over $\mathscr{D}(\mathcal{D})$.

As weak derivatives are a particular case of distributional derivatives (Section 2.2.4), one expects that the distributional solutions of a PDE that admit some weak derivatives are in fact weak solutions, i.e. solutions of some weak formulation of that PDE. Rigorous statements of this general fact have to be checked on a case-by-case basis, depending on the weak formulation at hand (a more in-depth discussion falls outside of the scope of this article). As an example, this is the case for the weak formulation of elliptic PDEs in $H^1_0(\mathcal{D})$ (see e.g. [16], Section 6.2), where $H^1_0(\mathcal{D})$ is the closure of $\mathscr{D}(\mathcal{D})$ in the Sobolev space $H^1(\mathcal{D}) := \{u \in L^2(\mathcal{D}) : \nabla u$ exists as a weak derivative and $\nabla u \in L^2(\mathcal{D})^d\}$.

Remark 3.4 (Measure-valued solutions of PDEs). Although it is not the main focus of the paper, we can even allow u in Definition 3.3 to be a Radon measure by replacing $u(x)dx$ with $\mu(dx)$ in equation (21). This will be useful in Section 4.1, where we will encounter a measure-valued PDE solution which is central from a physical viewpoint, with the wave equation's Green's function (it is *not* actually a function!). Notice that weak formulations in Sobolev spaces, say $H^1(\mathcal{D})$, are not well-equipped to work with such solutions, and our distributional framework becomes needed.

3.2.2 Random fields under distributional differential constraints.
We can now state the following proposition, based on Definition 3.3.

Proposition 3.5 (sample paths of random fields under linear differential constraints, distributional derivatives). *Let $\mathcal{D} \subset \mathbb{R}^d$ be an open set and let $L = \sum a_\alpha(x)\partial^\alpha$, $|\alpha| \leqslant n$, be a linear differential operator of order n with coefficients $a_\alpha(x) \in C^{|\alpha|}(\mathcal{D})$. Let $U = \big(U(x)\big)_{x \in \mathcal{D}}$ be a measurable second order random field with mean function $m(x)$ and covariance function $k(x, x')$. For all $x \in \mathcal{D}$, note $k_x : y \longmapsto k(x, y)$. Suppose that $m \in L^1_{loc}(\mathcal{D})$ and $\sigma \in L^1_{loc}(\mathcal{D})$, where $\sigma : x \mapsto k(x, x)^{1/2}$.*
1) Then $\mathbb{P}(U \in L^1_{loc}(\mathcal{D})) = 1$ and for all $x \in \mathcal{D}, k_x \in L^1_{loc}(\mathcal{D})$.
2) Suppose that $L(m) = 0$ in the sense of distributions. Then the following statements are equivalent:

   (i) $\mathbb{P}(L(U) = 0$ in the sense of distributions$) = 1$.

   (ii) $\forall x \in \mathcal{D}, \; L(k_x) = 0$ in the sense of distributions.

12

Explicitly, by (i) we mean that there exists a set $A \in \mathcal{A}$ with $\mathbb{P}(A) = 1$ such that for all $\omega \in A$,

$$\forall \varphi \in \mathscr{D}(\mathcal{D}), \quad \langle U_\omega, L^*\varphi \rangle = \int_\mathcal{D} U_\omega(x) L^*\varphi(x) dx = 0. \tag{22}$$

The fact that the functions $x \longmapsto U_\omega(x)$ and $y \longmapsto k_x(y)$ lie in $L^1_{loc}(\mathcal{D})$ ensure the existence of the integrals in equations (22) (see Point 2 of the proof of Proposition 3.5) as well as the continuity of the associated linear forms over $\mathscr{D}(\mathcal{D})$, following the definition of equation (21). The assumption that $a_\alpha \in C^{|\alpha|}(\mathcal{D})$ is not very strong, in the sense that it is the minimal assumption to ensure that the adjoint L^* is well-defined (equation (20)), and thus that Definition 3.3 even makes sense. Likewise, requiring that $\sigma \in L^1_{loc}(\mathcal{D})$ is not very restrictive (see Section 2.2.1). However, ensuring the measurability of the random process U is more demanding in practice, because it is difficult to ensure this property outside of having continuity in probability (see Section 2.1.2).

The following lemma will be crucial for the proof of Proposition 3.5:

Lemma 3.6. $\mathscr{D}(\mathcal{D})$ *is sequentially separable, i.e. there exists a countable subset $F \subset \mathscr{D}(\mathcal{D})$ such that for all $\varphi \in \mathscr{D}(\mathcal{D})$, there exists a sequence $(\varphi_n) \subset F$ such that $\varphi_n \to \varphi$ in $\mathscr{D}(\mathcal{D})$ for its LF topology.*

Recall that a topological space E is separable if there exists a countable subset $F \subset E$ such that its closure in E is equal to E. If the topology of E is metrizable (as e.g. for Fréchet spaces), sequential separability and separability are equivalent. If this topology is *not* metrizable (as e.g. for LF spaces), then sequential separability implies separability but the converse need not hold. Below, we provide a short proof of Lemma 3.6, as we could not find it in the literature. The weaker property that $\mathscr{D}(\mathcal{D})$ is separable is already difficult to track down, see e.g. [22], Corollaire (1).2, p. 78 or [24], p. 73, (3).

Proof. We first show that the spaces $\mathscr{D}_{K_i}(\mathcal{D})$ introduced in Section 2.2.2 are separable Fréchet spaces. The Fréchet topology of $\mathscr{D}_{K_i}(\mathcal{D})$ is the one induced by the usual Fréchet topology of $C^\infty(\mathcal{D})$ when $\mathscr{D}_{K_i}(\mathcal{D})$ is seen as a subspace of $C^\infty(\mathcal{D})$ ([54], pp. 131-132). As a Fréchet space, $C^\infty(\mathcal{D})$ is metrizable ([54], p. 85). But $C^\infty(\mathcal{D})$ is also a Montel space ([54], Proposition 34.4, p. 357): as a metrizable Montel space, it is automatically separable ([48], p. 195 or [11]). Thus $\mathscr{D}_{K_i}(\mathcal{D})$ is also separable as a subset of the separable metrizable space $C^\infty(\mathcal{D})$ ([7], Proposition 3.25, p. 73).

Denote now F_i a countable dense subset of $\mathscr{D}_{K_i}(\mathcal{D})$ and consider $F := \bigcup_{i \in \mathbb{N}} F_i$. Let $\varphi \in \mathscr{D}(\mathcal{D})$ and $i \in \mathbb{N}$ such that $\text{Supp}(\varphi) \subset K_i$, where $(K_i)_{i \in \mathbb{N}}$ is the sequence of compact sets from Section 2.2.2. Then $\varphi \in \mathscr{D}_{K_i}(\mathcal{D})$ and there exists a sequence $(\varphi_n) \subset F_i \subset F$ such that $\varphi_n \to \varphi$ in the sense of the Fréchet topology of $\mathscr{D}_{K_i}(\mathcal{D})$, i.e. the metric d_i in equation (9). From equation (9), $||\partial^\alpha \varphi_n - \partial^\alpha \varphi||_\infty \to 0$ for all $\alpha \in \mathbb{N}^d$. Since $\text{Supp}(\varphi_n) \subset K_i$ for all $n \in \mathbb{N}$, we have that $\varphi_n \to \varphi$ in $\mathscr{D}(\mathcal{D})$ (see Section 2.2.2). $\qquad \square$

We are now able to prove Proposition 3.5.

Proof. Suppose first that U is centered, i.e. $m \equiv 0$.
1) We begin by showing that the sample paths of U almost surely lie in $L^1_{loc}(\mathcal{D})$. Note first that thanks to the Cauchy-Schwarz inequality, $\mathbb{E}[|U(x)|] \leqslant \sigma(x)$. Now, let $(K_i)_{i \in \mathbb{N}}$ be an increasing sequence of compact subsets of $\mathcal{D}$ such that $\bigcup_{i \in \mathbb{N}} K_i = \mathcal{D}$. Using Tonelli's theorem, we have that for any $n \in \mathbb{N}$,

$$\mathbb{E}\left[\int_{K_i} |U(x)| dx \right] = \int_{K_i} \mathbb{E}[|U(x)|] dx \leqslant \int_{K_i} \sigma(x) dx < +\infty, \tag{23}$$

13

since $\sigma \in L^1_{loc}(\mathcal{D})$. Note that in order for the integrals above to be well defined, imposing that U is a measurable random field cannot be circumvented. Equation (23) yields a set $B_n \subset \Omega$ of probability 1 over which the random variable $\omega \longmapsto \int_{K_i} |U_\omega(x)|dx$ is finite (from Fubini's theorem again, the map $\omega \longmapsto \int_{K_i} |U_\omega(x)|dx$ is measurable). Consider now the set $B = \bigcap_{n\in\mathbb{N}} B_n$ which remains of probability 1. For all compact subset $K \subset \mathcal{D}$, there exists an integer n_K such that $K \subset K_{n_K}$ and thus for all $\omega \in B$,

$$\int_K |U_\omega(x)|dx \leqslant \int_{K_{n_K}} |U_\omega(x)|dx < +\infty, \tag{24}$$

which shows that the sample paths of U lie in $L^1_{loc}(\mathcal{D})$ almost surely. Similarly, we check that for all $x \in \mathcal{D}, k_x$ lies in $L^1_{loc}(\mathcal{D})$: for any compact set K, since $\sigma \in L^1_{loc}(\mathcal{D})$ and because of equation (7),

$$\int_K |k_x(y)|dy = \int_K |k(x,y)|dy \leqslant \sigma(x)\int_K \sigma(y)dy < \infty.$$

2) Let us check in advance that whatever $f \in L^1_{loc}(\mathcal{D})$, the map $T(f) : \varphi \longmapsto \langle f, L^*\varphi \rangle$ is a continuous linear form over $\mathscr{D}(\mathcal{D})$. Since $a_\alpha \in C^{|\alpha|}(\mathcal{D})$, we can apply Leibniz' rule on $L^*\varphi = \sum_{|\alpha|\leqslant n} (-1)^{|\alpha|}\partial^\alpha(a_\alpha\varphi)$. This yields a family $\{f_\alpha\}_{|\alpha|\leqslant n}$ of continuous functions over $\mathcal{D}$ such that

$$\forall \varphi \in \mathscr{D}(\mathcal{D}), \quad \forall x \in \mathcal{D}, \quad L^*\varphi(x) = \sum_{|\alpha|\leqslant n} f_\alpha(x)\partial^\alpha\varphi(x). \tag{25}$$

For all $f \in L^1_{loc}(\mathcal{D})$, for all compact set $K \subset \mathcal{D}$ and for all $\varphi \in \mathscr{D}(\mathcal{D})$ such that $\text{Supp}(\varphi) \subset K$, we have $\text{Supp}(L^*\varphi) \subset K$ and equation (25) yields

$$\begin{aligned}|\langle f, L^*\varphi \rangle| &\leqslant \int_{\mathcal{D}} |f(x)||L^*\varphi(x)|dx \\ &\leqslant \left(\int_K |f(x)|dx \times \max_{|\alpha|\leqslant n}\sup_{x\in K} |f_\alpha(x)|\right) \times \sum_{|\alpha|\leqslant n} ||\partial^\alpha\varphi||_\infty < +\infty. \end{aligned} \tag{26}$$

This proves that $T(f) : \varphi \longmapsto \langle f, L^*\varphi \rangle$ is a continuous linear form over $\mathscr{D}(\mathcal{D})$ (see equation (10)).

$(i) \implies (ii)$: Suppose (i). Let $\varphi \in \mathscr{D}(\mathcal{D})$. There exists a set $A \subset \Omega$ such that $\mathbb{P}(A) = 1$ and such that

$$\forall \omega \in A, \quad \langle U_\omega, L^*\varphi \rangle = \int_{\mathcal{D}} U_\omega(x)L^*\phi(x)dx = 0.$$

Multiplying equation above with the random variable $U(x')$, taking the expectation and formally permuting (for now) the integral and the expectation, we obtain

$$\begin{aligned}0 = \mathbb{E}\left[U(x')\int_{\mathcal{D}} U(x)L^*\varphi(x)dx\right] &= \int_{\mathcal{D}} L^*\varphi(x)\mathbb{E}[U(x)U(x')]dx \\ &= \int_{\mathcal{D}} L^*\varphi(x)k(x,x')dx = \langle k_{x'}, L^*\varphi \rangle.\end{aligned}$$

The integral-expectation permutation is justified by writing down the expectation as an integral and using Fubini's theorem, checking that the below quantity is finite. We use Tonelli's theorem

14

and the Cauchy-Schwarz inequality:

$$\mathbb{E}\left[\int_{\mathcal{D}} |U(x')U(x)L^*\varphi(x)|dx\right] = \int_{\mathcal{D}} |L^*\varphi(x)|\mathbb{E}[|U(x)U(x')|]dx$$

$$\leqslant \int_{\mathcal{D}} |L^*\varphi(x)|\mathbb{E}[U(x)^2]^{1/2}\mathbb{E}[U(x')^2]^{1/2}dx$$

$$\leqslant \sigma(x')\int_{\mathcal{D}} |L^*\varphi(x)|\sigma(x)dx < +\infty.$$

Indeed, since $\sigma \in L^1_{loc}(\mathcal{D})$, setting $f = \sigma$ in equation (26) shows that the last integral is indeed finite. Thus, $\forall x \in \mathcal{D}, \forall \varphi \in \mathscr{D}(\mathcal{D}), \langle k_x, L^*\varphi \rangle = 0$ which proves that $(i) \implies (ii)$.

$(ii) \implies (i)$: Suppose (ii). Let $\varphi \in \mathscr{D}(\mathcal{D})$, we have $\langle k_{x'}, L^*\varphi \rangle = 0$. Multiplying this with $L^*\varphi(x')$ and integrating with reference to x' yields

$$0 = \int_{\mathcal{D}} L^*\varphi(x')\int_{\mathcal{D}} L^*\varphi(x)k(x, x')dxdx' = \int_{\mathcal{D}}\int_{\mathcal{D}} L^*\varphi(x)L^*\varphi(x')\mathbb{E}[U(x)U(x')]dxdx'.$$

Permuting formally the expectation and the integrals (justified in equation (27)) yields

$$0 = \int_{\mathcal{D}}\int_{\mathcal{D}} L^*\varphi(x)L^*\varphi(x')\mathbb{E}[U(x)U(x')]dxdx'$$

$$= \mathbb{E}\left[\left(\int_{\mathcal{D}} L^*\varphi(x)U(x)dx\right)^2\right] = \mathbb{E}[\langle U, L^*\varphi \rangle^2],$$

and thus $\langle U, L^*\varphi \rangle = 0$ a.s. : there exists $A_\varphi \in \mathcal{A}$ with $\mathbb{P}(A_\varphi) = 1$ such that $\forall \omega \in A_\varphi, \langle U_\omega, L^*\varphi \rangle = 0$. We justify the expectation-integral permutation with the computation below

$$\int_{\mathcal{D}}\int_{\mathcal{D}} |L^*\varphi(x)L^*\varphi(x')|\mathbb{E}[|U(x)U(x')|]dxdx'$$

$$\leqslant \int_{\mathcal{D}}\int_{\mathcal{D}} |L^*\varphi(x)L^*\varphi(x')|\sigma(x)\sigma(x')dxdx'$$

$$\leqslant \left(\int_{\mathcal{D}} |L^*\varphi(x)|\sigma(x)dx\right)^2 < +\infty. \tag{27}$$

As previously, setting $f = \sigma$ in equation (26) shows that the integral above is indeed finite.

This does not finish the proof as we need to find a set A with $\mathbb{P}(A) = 1$, independently from φ, such that $\forall \omega \in A, \langle U_\omega, L^*\varphi \rangle = 0$. For this we shall use Lemma 3.6. Let

$$A := B \cap \left(\bigcap_{\varphi \in F} A_\varphi\right), \tag{28}$$

where the set F is introduced in Lemma 3.6. Then $\mathbb{P}(A) = 1$ since $\mathbb{P}(B) = 1, \mathbb{P}(A_\varphi) = 1$ and F is countable. Let $\omega \in A$. Since $U_\omega \in L^1_{loc}(\mathcal{D})$, equation (26) shows that the map $T_\omega : \varphi \longmapsto \langle U_\omega, L^*\varphi \rangle$ is a continuous linear form on $\mathscr{D}(\mathcal{D})$. In particular, Theorem 6.6(c) p. 155 from [46] states that T_ω is in fact *sequentially* continuous. Let $\varphi \in \mathscr{D}(\mathcal{D})$ and $(\varphi_n) \subset F$ be such that $\varphi_n \to \varphi$ in $\mathscr{D}(\mathcal{D})$, from Lemma 3.6. From the sequential continuity of T_ω, $T_\omega(\varphi) = \lim_{n\to\infty} T_\omega(\varphi_n) = 0$ since $\forall n \in \mathbb{N}, T_\omega(\varphi_n) = 0$. That is, we have proved that

$$\forall \omega \in A, \quad \forall \varphi \in \mathscr{D}(\mathcal{D}), \quad \langle U_\omega, L^*\varphi \rangle = T_\omega(\varphi) = 0.$$

Since $\mathbb{P}(A) = 1$, this shows that $(ii) \implies (i)$.

When U is not centered, consider the centered random field V defined by $V(x) = U(x) - m(x)$ for which the above proof can be applied. Since L is linear and m is assumed to verify $L(m) = 0$ in the sense of distributions, the probabilistic sets $A_U = \{L(U) = 0$ in the sense of distributions$\}$ and $A_V = \{L(V) = 0$ in the sense of distributions$\}$ coincide and thus, $A \subset A_U$. Finally, U and V have the same covariance function $k(x, x')$. Thus,

$$\mathbb{P}(L(U) = 0 \text{ in the distrib. sense}) = 1 \iff \mathbb{P}(L(V) = 0 \text{ in the distrib. sense}) = 1$$
$$\iff \forall x \in \mathcal{D}, L(k_x) = 0 \text{ in the distrib. sense},$$

which finishes the proof in the general case. $\qquad\square$

Remark 3.7. Distributional solutions are the weakest types of solutions for PDEs. In general, additional regularity conditions have to be imposed to obtain physically realistic solutions, such as Sobolev regularity or entropy conditions as for nonlinear hyperbolic PDEs [52]. However, every step in the above proof remains valid when replacing $\varphi \in \mathscr{D}(\mathcal{D})$ with $\varphi \in C_c^n(\mathcal{D})$. Although we have not explicited the usual topology of $C_c^n(\mathcal{D})$ in this article, we state that this is enough to show that the equalities stated in Proposition 3.5 also hold in $C_c^n(\mathcal{D})'$, the space of finite order generalized functions of order n, rather than just in $\mathscr{D}'(\mathcal{D})$. $C_c^n(\mathcal{D})'$ is a smaller space than $\mathscr{D}'(\mathcal{D})$, though less used in functional analysis than $\mathscr{D}'(\mathcal{D})$.

We partially recover Proposition 3.2 when the sample paths of U are n times differentiable with locally integrable n^{th} derivative and $k \in C^{n,n}(\mathcal{D} \times \mathcal{D})$. Indeed, in that case one can show that if $L = \sum_{|\alpha| \leqslant n} a_\alpha(x) \partial^\alpha$, then we simply have $\mathscr{L} = L$ in Proposition 3.2. Additionally, $L(U_\omega)$ and $L(k_x)$ both lie in $\mathcal{F}(\mathcal{D}, \mathbb{R}) \cap L_{loc}^1(\mathcal{D})$. In that framework, Proposition 3.2 states that

$$\forall x \in \mathcal{D}, \ L(k_x) = 0 \iff \mathbb{P}(L(U) = 0) = 1, \tag{29}$$

where the function equalities of the form $L(f) = 0$ in equation (29) are valid everywhere on $\mathcal{D}$. In contrast, for any function g that lies in $L_{loc}^1(\mathcal{D})$, we have

$$g = 0 \text{ in the sense of distributions} \iff g = 0 \ a.e. \tag{30}$$

Equation (30) is just another way of saying that the linear map $f \longmapsto T_f$ given in (12) is injective. Following equation (30), Proposition 3.5 states a slightly weaker result than (29), namely that

$$\forall x \in \mathcal{D}, \ \ L(k_x) = 0 \ a.e. \iff \mathbb{P}(L(U) = 0 \ a.e.) = 1. \tag{31}$$

If we actually have that the sample paths of U lie in $C^n(\mathcal{D})$, nullity almost everywhere implies nullity everywhere and we recover equation (29) from equation (31).

Instead of having the sample paths of U lie in $C^n(\mathcal{D})$ though, one may rather encounter the case where U is *mean-square* differentiable up to a certain order m. Under some continuity assumptions over the covariance function of U and up to suitable modifications, [49] showed that the sample paths of the mean-square differentiated process are actually weak derivatives of the sample paths of U. As observed after Definition 3.3, we thus expect that the sample paths of the mean-square differentiable random fields verifying Point 2, (ii) of Proposition 3.5 are solutions of some weak formulation of the PDE, rather than just distributional solutions.

Example (A first order PDE). Consider a continuous, nondifferentiable one dimensional covariance function $k_0 : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, for example $k_0(x, x') = \exp(-|x - x'|)$. It is then readily checked that the function $k : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ defined by $k((x, y), (x', y')) = k_0(x - y, x' - y')$ is positive definite and verifies Point 2, (ii) of Proposition 3.5 for the PDE

$$\partial_x u + \partial_y u = 0 \quad \text{in } \mathbb{R}^2. \tag{32}$$

Consider now a centered second order random field $(U(x, y))_{(x,y) \in \mathbb{R}^2}$ with covariance function k, passing to a measurable version of U if necessary (it exists from Section 2.1.2, as the continuity of k yields the continuity in probability of U). Then almost surely, its sample paths verify the PDE (32) in the sense of distributions, even though they are not expected to be differentiable. An example of random field whose covariance function is k as defined above, is the GP $(U_0(x - y))_{(x,y) \in \mathbb{R}^2}$ where $(U_0(x))_{x \in \mathbb{R}} \sim GP(0, k_0)$. These formulas can be obtained by viewing the PDE (32) as a transport equation under the condition that $U(x, 0) = U_0(x)$, following the same approach as in the upcoming Section 4.2.

3.3 A heredity property for Gaussian process regression

3.3.1 Gaussian process regression in a nutshell. 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 database B yields a second GP $\tilde{U}$ given by $\tilde{U}(x) := (U(x)|U(x_i) = u(x_i), i = 1, ..., n)$. The law of $\tilde{U}$ is known: $(\tilde{U}(x))_{x \in \mathcal{D}} \sim GP(\tilde{m}, \tilde{k})$. $\tilde{m}$ and $\tilde{k}$ are given by the so-called *Kriging* equations (33) and (34). Let $X = (x_1, ..., x_n)^T$, 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 $k(X, x)_i = k(x_i, x)$. Suppose that $K(X, X)$ is invertible, then [43]

$$
\begin{cases}
\tilde{m}(x) & = \quad m(x) + k(X, x)^T k(X, X)^{-1}(u(X) - m(X)), & (33) \\
\tilde{k}(x, x') & = \quad k(x, x') - k(X, x)^T k(X, X)^{-1} k(X, x'). & (34)
\end{cases}
$$

The Kriging standard deviation function is then given by

$$\tilde{\sigma}(x) = \tilde{k}(x, x)^{1/2}. \tag{35}$$

The so-called Kriging mean $\tilde{m}$ plays the role of an approximation of u; in particular, it interpolates u at the observation points: $\tilde{m}(x_i) = u(x_i)$ for all $i = 1, ..., n$. Moreover, the Kriging covariance $\tilde{k}$ can be used to further control the distance between u and $\tilde{m}$.

3.3.2 Conditioned Gaussian processes under linear differential constraints. We can now state the following corollary, which draws the consequences of Proposition 3.5 when applied to GPR.

Proposition 3.8 (Heredity of Proposition 3.5 to conditioned GPs)**.** *Let $\mathcal{D}$ and L be as defined in Proposition 3.5. Let $(U(x))_{x \in \mathcal{D}} \sim GP(m, k)$ be a Gaussian process that verifies the assumptions of Proposition 3.5. Suppose also that*

$$L(m) = 0 \quad and \quad \forall x \in \mathcal{D}, \ L(k_x) = 0 \quad both \ in \ the \ sense \ of \ distributions. \tag{36}$$

(i) Then whatever the integer p, the vector $u = (u_1, ..., u_p)^T \in \mathbb{R}^p$ and the vector $X = (x_1, ..., x_p)^T \in \mathcal{D}^p$ such that $k(X, X)$ is invertible, the Kriging mean $\tilde{m}(x)$ and the Kriging standard deviation function $\tilde{\sigma}$ both lie in $L^1_{loc}(\mathcal{D})$, and we have

$$L(\tilde{m}) = 0 \quad and \quad \forall x \in \mathcal{D}, \ L(\tilde{k}_x) = 0 \quad both \ in \ the \ sense \ of \ distributions.$$

where $\tilde{m}$ and $\tilde{k}$ are defined in equations (33) and (34).
(ii) As such, the sample paths of the conditioned Gaussian process $(\tilde{U}(x))_{x \in \mathcal{D}}$ defined by $\tilde{U}(x) = (U(x)|U(x_i) = u_i \ \forall i = 1, ..., p)$ are almost surely solutions of the equation $L(f) = 0$ in the sense of distributions:

$$\mathbb{P}(L(\tilde{U}) = 0 \ in \ the \ sense \ of \ distributions) = 1.$$

Proof. Note first that for all $x \in \mathcal{D}, \tilde{k}(x, x) \leqslant k(x, x)$ ([19], p. 117). Thus the function $\tilde{\sigma} : x \longmapsto \tilde{k}(x, x)^{1/2}$ also lies in $L^1_{loc}(\mathcal{D})$. Point (i) is then a direct consequence of the definition of $\tilde{m}$ and $\tilde{k}$ in equations (33) and (34), and the linearity of L. Proposition 3.5 can then be applied conjointly with (i), which yields point (ii) since the mean and covariance functions of the GP $\tilde{U}$ are $\tilde{m}$ and $\tilde{k}$ (see equations (33) and (34)). $\qquad\square$

Proposition 3.8 shows that when U is a GP, the results of Proposition 3.5 are inherited on the conditioned posterior process $\tilde{U}$. One weak consequence of Proposition 3.8 is that if GPR is performed with a covariance function k that verifies point (ii) of Proposition 3.5, then all the possible Kriging means provided by GPR remain solutions of the PDE $L(\tilde{m}) = 0$.

4   Gaussian processes and the 3 dimensional wave equation

The formalism we used in the previous section is necessary to tackle hyperbolic PDEs as in some cases, their solutions only verify the PDE in a weaker sense, e.g. the distributional sense ([16], Sections 2.1.1 and 7.2). Hyperbolic PDEs are typically encountered when describing finite speed propagation phenomena and their prototype is the wave equation (see equation (37)); this equation is central in a number of fields such as acoustics, electromagnetics and quantum mechanics. In this section, we derive a GP model for the solutions of the homogeneous 3D wave equation, with explicit covariance formulas in the form of convolutions.

We show on one example that the model we obtain below is capable of dealing with an initial speed v_0 that is piecewise continuous and an initial position u_0 that has piecewise continuous derivatives, when the initial discontinuity surfaces are "nice enough". This is an advantage with reference to the previous models, where the sample paths actually had to be sufficiently differentiable to obtain sample path degeneracy with reference to the PDE.

4.1   General solution to the 3 dimensional wave equation

Denote the 3D Laplace operator $\Delta = \partial^2_{xx} + \partial^2_{yy} + \partial^2_{zz}$ and the d'Alembert operator $\Box = 1/c^2\partial^2_{tt} - \Delta$ with constant wave speed $c > 0$. We focus on the general initial value problem in the free space $\mathbb{R}^3$

$$\begin{cases} \Box w & = 0 & \forall(x, t) \in \mathbb{R}^3 \times \mathbb{R}^*_+, \\ w(x, 0) & = u_0(x) & \forall x \in \mathbb{R}^3, \\ (\partial_t w)(x, 0) & = v_0(x) & \forall x \in \mathbb{R}^3. \end{cases} \qquad (37)$$

Throughout this article, we will refer to u_0 as the initial position and v_0 as the initial speed. The solution of this problem is unique in the distributional sense ([14], p. 164). It can be extended to all $t \in \mathbb{R}$ ([14], p. 295) and is represented as follow ([14], p. 295 again)

$$w(x, t) = (F_t * v_0)(x) + (\dot{F}_t * u_0)(x) \qquad \forall(x, t) \in \mathbb{R}^3 \times \mathbb{R}, \qquad (38)$$

where F_t and $\dot{F}_t$ are known generalized functions. That is, the function $w(x, t)$ above is a solution of the system (37), in which t is now allowed to lie in $\mathbb{R}$ rather than $\mathbb{R}^*_+$. The existence of such an extension is possible because of the time reversibility of the wave equation (in the language of semigroup theory, its semigroup can be embedded in a group, [41], Theorem 4.5 p. 222). In dimension 3, F_t and $\dot{F}_t$ are compactly supported generalized functions of order 0 and 1 respectively. They are given by

$$F_t = \frac{\sigma_{c|t|}}{4\pi c^2 t} \quad \text{and} \quad \dot{F}_t = \partial_t F_t \quad \forall t \in \mathbb{R}, \qquad (39)$$

where σ_R is the surface measure of the sphere of center 0 and radius R; $\dot{F}_t = \partial_t F_t$ means that for all $f \in C_c^1(\mathbb{R}^3), \langle \dot{F}_t, f \rangle = \partial_t \langle F_t, f \rangle$. We make these expressions more explicit in equation (40), using spherical coordinates. It is worth noting that $(F_t)_{t \in \mathbb{R}}$ corresponds to the *Green's function* of the wave equation, in the sense that it verifies the system (37) with $u_0 = 0$ and $v_0 = \delta_0$ where δ_0 is the Dirac mass ([14], pp. 294-295). As discussed in Remark 3.4, $(F_t)_{t \in \mathbb{R}}$ is a family of singular measures and this PDE system has to be understood in the distributional sense. Note also that equations (39) show that F_t and $\dot{F}_t$ are supported on the sphere of radius $c|t|$: "the support of F_t propagates at finite speed c". This property is known as the Huygens principle for the three dimensional wave equation, see [16], p. 80.

Suppose that $u_0 \in C^1(\mathbb{R}^3)$ and $v_0 \in C^0(\mathbb{R}^3)$, then w as defined in equation (38) is a pointwise defined function (Section 2.2.7) and in that case an explicit formula for such convolutions is reminded in equation (15) (yet one may actually make sense out of (38) when u_0 and v_0 are only required to be any generalized functions, see [54], Chapter 27).

Equation (38) can be written using means over spheres. Denote $(r, \theta, \phi), r \geqslant 0, \theta \in [0, \pi], \phi \in [0, 2\pi]$ the spherical coordinates, $S(0, 1)$ the unit sphere of $\mathbb{R}^3$ and $\gamma = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)^T$ the corresponding parametrization of $S(0, 1)$ ($||\gamma||_2 = 1$). We write $d\Omega = \sin \theta d\theta d\phi$ the surface differential element of $S(0, 1)$. The formulas (38) and (39) then lead to the Kirschoff formula ([16], p. 72):

$$w(x, t) = \int_{S(0,1)} t v_0(x - c|t|\gamma) + u_0(x - c|t|\gamma) - c|t|\gamma \cdot \nabla u_0(x - c|t|\gamma) \frac{d\Omega}{4\pi} \qquad (40)$$

4.2   Gaussian process modelling of the solution

Suppose now that u_0 and v_0 are unknown, and only pointwise values of w are observed. We thus model u_0 and v_0 as random functions and put Gaussian process priors over u_0 and v_0. More precisely, we make the following assumptions.

(A_1) Suppose that the initial conditions u_0 and v_0 of Problem (37) are sample paths drawn from two independent Gaussian processes $U^0 \sim GP(0, k_{\mathrm{u}})$ and $V^0 \sim GP(0, k_{\mathrm{v}})$: $\exists \omega \in \Omega, \forall x \in \mathbb{R}^3, u_0(x) = U_\omega^0(x)$ and $v_0(x) = V_\omega^0(x)$.

(A_2) Suppose that all sample paths of U^0 lie in $C^1(\mathbb{R}^3)$ and that those of V^0 lie in $C^0(\mathbb{R}^3)$, almost surely. A sufficient condition for this is given in [2], Theorem 1.4.2. This theorem states that under mild technical assumptions, the paths of $(U(x))_{x \in \mathcal{D}} \sim GP(0, k)$ lie in C^l a.s. as soon as $k \in C^{2l}(\mathcal{D} \times \mathcal{D})$, *which we assume from now on.* This is e.g. fulfilled by the Matérn covariance functions from equation (8), with $l = 0$ for $k_{1/2}$ and $l = 1$ for $k_{3/2}$.

We now analyse the consequence of these two assumptions. First, they imply that by solving (37), one obtains a time-space stochastic process $W(x, t)$ defined by

$$W(x, t) : \Omega \ni \omega \longmapsto (F_t * V_\omega^0)(x) + (\dot{F}_t * U_\omega^0)(x). \qquad (41)$$

Here again, V_ω^0 denotes the sample path of V^0 at $\omega \in \Omega$ and likewise for U_ω^0. In particular, thanks to assumption (A_2), equation (41) defines a random variable for all (x, t). Note the space-time variable $z = (x, t)$ and note the random variables

$$V(z) : \omega \longmapsto (F_t * V_\omega^0)(x) \quad \text{and} \quad U(z) : \omega \longmapsto (\dot{F}_t * U_\omega^0)(x), \qquad (42)$$

that is, $W(z) = U(z) + V(z)$. We show in the next proposition that the random fields U, V and W are GPs as well. In particular we describe their covariance functions.

Proposition 4.1. *Define the two functions*

$$k_{\mathrm{v}}^{\mathrm{wave}}(z, z') = [(F_t \otimes F_{t'}) * k_{\mathrm{v}}](x, x'), \tag{43}$$

$$k_{\mathrm{u}}^{\mathrm{wave}}(z, z') = [(\dot{F}_t \otimes \dot{F}_{t'}) * k_{\mathrm{u}}](x, x'). \tag{44}$$

(i) Then $U = (U(z))_{z \in \mathbb{R}^3 \times \mathbb{R}}$ and $V = (V(z))_{z \in \mathbb{R}^3 \times \mathbb{R}}$ as defined in (42) are two independent centered GPs with covariance functions $k_{\mathrm{u}}^{\mathrm{wave}}$ and $k_{\mathrm{v}}^{\mathrm{wave}}$ respectively. Consequently, $(W(z))_{z \in \mathbb{R}^3 \times \mathbb{R}}$ is a centered GP whose covariance function is given by

$$k_W(z, z') = k_{\mathrm{v}}^{\mathrm{wave}}(z, z') + k_{\mathrm{u}}^{\mathrm{wave}}(z, z'). \tag{45}$$

(ii) Conversely, any measurable centered second order random field with covariance function k_W has its sample paths solution of the wave equation (37), in the sense of distributions, almost surely.

The formulas (43) and (44) can easily be derived formally, by running computations as if F_t and $\dot{F}_t$ were regular generalized functions (Section 2.2.4). This is somewhat justified because any generalized function can be approximated with a sequence of smooth compactly supported functions, by a "cutting and regularizing" argument ([54], Theorem 28.2, Chapter 28). However, checking that this procedure passes to the limit everywhere is tedious. Here, we rather make use of representations of F_t and $\dot{F}_t$ thanks to Radon measures (Sections 2.2.5 and 2.2.6) and use Fubini's theorem. We refer to Sections 2.2.6 and 2.2.8 for the definition of $\dot{F}_t \otimes \dot{F}_{t'}$, and Section 2.2.7 for the definition of $(\dot{F}_t \otimes \dot{F}_{t'}) * k_{\mathrm{u}}$.

Proof. (i) : first we prove that U and V are GPs. Since U^0 and V^0 are GPs, $\mathcal{L}(U^0)$ and $\mathcal{L}(V^0)$ are only comprised of Gaussian random variables (see Section 2.1.3). We then rely on the Kirschoff formula (40), writing the integrals as limits of Riemann sums. We start with V, that is, we focus on the first term in Kirschoff's formula (40). To show that V is a Gaussian process, we only need to show that for any z, $V(z) \in \mathcal{L}(V^0)$ as this will ensure the Gaussian process property. Since the sample paths of V^0 are continuous almost surely, there exists a sequence of numbers $(a_k^n) \subset \mathbb{R}$ and points $(y_k^n) \subset S(0, 1)$ such that for almost any $\omega \in \Omega$,

$$V(z)(\omega) = (F_t * V_{\omega}^0)(x) = t \int_{S(0,1)} V^0(x - c|t|\gamma)(\omega) \frac{d\Omega}{4\pi}$$

$$= \frac{t}{4\pi} \int_0^{2\pi} \int_0^{\pi} V^0(x - c|t|\gamma(\theta, \phi))(\omega) \sin(\theta) d\theta d\phi = \lim_{n \to \infty} \sum_{k=1}^n a_k^n V^0(x - c|t|y_k^n)(\omega).$$

This shows that $V(z)$ is the a.s. limit of the sequence of centered Gaussian random variables $(Y_n) \subset \mathcal{L}(V^0)$, where $Y_n = \sum_{k=1}^n a_k^n V^0(x - c|t|y_k^n)$; Y_n is Gaussian because V^0 is a GP. Almost sure convergence implies convergence in law. From [35], Proposition 1.1, $V(z)$ is normally distributed and the convergence also takes place in $L^2(\mathbb{P})$. Therefore, $V(z) \in \mathcal{L}(V^0)$ and V is a Gaussian process. From the same proposition, $V(z)$ is centered because the variables Y_n are centered. Note that since F_t is supported on the compact set $S(0, c|t|)$, we only required the sample paths of V^0 to be continuous rather than continuous and compactly supported.

We apply the same reasoning to U, by applying the above steps to the second part of Kirschoff's formula (40). One's ability to write out the integrals as a limit of Riemann sums is ensured when the sample paths of U^0 lie in $C^1(\mathbb{R}^3)$.

Finally, since U^0 and V^0 are independent, $\mathcal{L}(U^0)$ and $\mathcal{L}(V^0)$ are orthogonal in $L^2(\mathbb{P})$. Since $\mathcal{L}(U) \subset \mathcal{L}(U^0)$ and likewise for V, U and V are independent Gaussian processes as for Gaussian random variables, independence is equivalent to null covariance. Finally, the sum of independent

Gaussian random variables is a Gaussian random variable. Therefore $\mathcal{L}(W) \subset \mathcal{L}(U) + \mathcal{L}(V)$ is only comprised of Gaussian random variables and W is a Gaussian process. Now, we prove that

$$\mathbb{E}[U(z)U(z')] = [(\dot{F}_t \otimes \dot{F}_{t'}) * k_{\mathrm{u}}](x, x'). \tag{46}$$

The main argument is Fubini's theorem for Radon measures. For this we use the fact that $\dot{F}_t$ is a distribution of order 1 and can be identified to a sum of derivatives of measures (see equation (14)): for all $t \in \mathbb{R}$, there exists $\{\mu_\alpha^t\}_{\alpha \in \mathbb{N}^3, |\alpha| \leqslant 1}$ a family of Radon measures such that

$$\dot{F}_t = \sum_{|\alpha| \leqslant 1} \partial^\alpha \mu_\alpha^t \quad \text{in the sense of distributions.} \tag{47}$$

Moreover, $\dot{F}_t$ is compactly supported, therefore all the measures μ_α^t are also compactly supported. First, we write $U_\omega(z)$ in integral form:

$$U_\omega(z) = \left(\dot{F}_t * U_\omega^0\right)(x) = \langle \dot{F}_t, \tau_{-x}\check{U}_\omega^0 \rangle = \left\langle \sum_{|\alpha| \leqslant 1} \partial^\alpha \mu_\alpha^t, \tau_{-x}\check{U}_\omega^0 \right\rangle \tag{48}$$

$$= \sum_{|\alpha| \leqslant 1} \langle \mu_\alpha^t, (-1)^{|\alpha|} \partial^\alpha \tau_{-x}\check{U}_\omega^0 \rangle = \sum_{|\alpha| \leqslant 1} \int_{\mathbb{R}^3} (-1)^{|\alpha|} \partial^\alpha U_\omega^0(x - y)\mu_\alpha^t(dy). \tag{49}$$

Before applying Fubini's theorem, we need to check an integrability condition. Let $\alpha \in \mathbb{N}^3$ be such that $|\alpha| \leqslant 1$. Recall that $|\mu_t^\alpha|$ is defined in Section 2.2.5; denote also $\sigma_{\partial^\alpha U^0}(x) = \sqrt{\mathrm{Var}(\partial^\alpha U_0(x))}$. Since the sample paths of U^0 lie in $C^1(\mathcal{D})$ a.s., those of $\partial^\alpha U^0$ lie in $C^0(\mathcal{D})$ and thus the function $x \longmapsto \mathrm{Var}(\partial^\alpha U_0(x))$ also lies in $C^0(\mathcal{D})$ ([5], chapter 1, Section 4.3). Therefore the function $x \longmapsto \sigma_{\partial^\alpha U^0}(x)$ also lies in $C^0(\mathcal{D})$. We now check that the integral I below is finite. We use Tonelli's theorem and the Cauchy-Schwarz inequality:

$$I := \int_\Omega \sum_{|\alpha| \leqslant 1} \int_{\mathbb{R}^3} \left|\partial^\alpha U_\omega^0(x - y)\right| |\mu_\alpha^t|(dy) \sum_{|\alpha'| \leqslant 1} \int_{\mathbb{R}^3} \left|\partial^{\alpha'} U_\omega^0(x' - y')\right| |\mu_{\alpha'}^{t'}|(dy') \mathbb{P}(d\omega)$$

$$= \sum_{|\alpha|, |\alpha'| \leqslant 1} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \int_\Omega \left|\partial^\alpha U_\omega^0(x - y)\partial^{\alpha'} U_\omega^0(x' - y')\right| \mathbb{P}(d\omega) |\mu_\alpha^t|(dy) |\mu_{\alpha'}^{t'}|(dy')$$

$$= \sum_{|\alpha|, |\alpha'| \leqslant 1} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \mathbb{E}\big[|\partial^\alpha U^0(x - y)\partial^{\alpha'} U^0(x' - y')|\big] |\mu_\alpha^t|(dy) |\mu_{\alpha'}^{t'}|(dy')$$

$$\leqslant \sum_{|\alpha|, |\alpha'| \leqslant 1} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \left(\mathbb{E}\big[\partial^\alpha U^0(x - y)^2\big]\mathbb{E}\big[\partial^{\alpha'} U^0(x' - y')^2\big]\right)^{1/2} |\mu_\alpha^t|(dy) |\mu_{\alpha'}^{t'}|(dy')$$

$$\leqslant \left(\sum_{|\alpha| \leqslant 1} \int_{\mathbb{R}^3} \left(\mathbb{E}\big[\partial^\alpha U^0(x - y)^2\big]\right)^{1/2} |\mu_\alpha^t|(dy)\right) \times \left(\sum_{|\alpha| \leqslant 1} \int_{\mathbb{R}^3} \left(\mathbb{E}\big[\partial^\alpha U^0(x - y)^2\big]\right)^{1/2} |\mu_\alpha^{t'}|(dy)\right)$$

$$\leqslant \left(\sum_{|\alpha| \leqslant 1} (|\mu_\alpha^t| * \sigma_{\partial^\alpha U^0})(x)\right) \times \left(\sum_{|\alpha| \leqslant 1} (|\mu_\alpha^{t'}| * \sigma_{\partial^\alpha U^0})(x')\right) < +\infty.$$

For all multi-index α, the scalar $(|\mu_\alpha^t| * \sigma_{\partial^\alpha U^0})(x)$ is finite because $x \longmapsto \sigma_{\partial^\alpha U^0}(x)$ is continuous and $|\mu_\alpha^t|$ is compactly supported. Note also that from Assumption (A_2), the GP U^0 is *mean square differentiable* up to order 1, which implies ([44], Section III.1.4) that we have, for all multi-indexes α, α' such that $|\alpha|, |\alpha'| \leqslant 1$, x and x':

$$\mathbb{E}\big[\partial^\alpha U^0(x)\partial^{\alpha'} U^0(x')\big] = \partial_1^\alpha \partial_2^{\alpha'} k_{\mathrm{u}}(x, x'). \tag{50}$$

where ∂_1 (respectively ∂_2) denotes derivatives with reference to the first (respectively second) argument of k_{u}. We may thus permute integrals and differential operators in $\mathbb{E}\big[U(z)U(z')\big]$:

$$
\begin{aligned}
\mathbb{E}\big[U(z)U(z')\big] &= \mathbb{E}\Bigg[\sum_{|\alpha|\leqslant 1} \int_{\mathbb{R}^3} (-1)^{|\alpha|}\partial^\alpha U^0(x-y)\mu_\alpha^t(dy)) \sum_{|\alpha'|\leqslant 1} \int_{\mathbb{R}^3} (-1)^{|\alpha'|}\partial^{\alpha'} U^0(x-y)\mu_{\alpha'}^{t'}(dy') \Bigg] \\
&= \sum_{|\alpha|,|\alpha'|\leqslant 1} \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} (-1)^{|\alpha|}(-1)^{|\alpha'|}\partial_1^\alpha \partial_2^{\alpha'} \mathbb{E}\big[U^0(x-y)U^0(x'-y')\big]\mu_\alpha^t(dy)\mu_{\alpha'}^{t'}(dy') \\
&= \sum_{|\alpha|,|\alpha'|\leqslant 1} \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} (-1)^{|\alpha|}(-1)^{|\alpha'|}\partial_1^\alpha \partial_2^{\alpha'} k_{\mathrm{u}}(x-y,x'-y')\mu_\alpha^t(dy)\mu_{\alpha'}^{t'}(dy') \\
&= \Bigg[\Big(\sum_{|\alpha|\leqslant 1} \partial^\alpha \mu_\alpha^t \otimes \sum_{|\alpha'|\leqslant 1} \partial^{\alpha'}\mu_{\alpha'}^{t'}\Big) * k_{\mathrm{u}}\Bigg](x,x') = [(\dot{F}_t \otimes \dot{F}_{t'}) * k_{\mathrm{u}}](x,x'),
\end{aligned}
$$

which proves (46).

One proves that $\mathbb{E}\big[V(z)V(z')\big] = [(F_t\otimes F_{t'})*k_{\mathrm{v}}](x,x')$ the exact same way, which is actually simpler as F_t is directly a measure. To conclude,

$$
\begin{aligned}
k_W(z,z') &= \mathrm{Cov}(W(z),W(z')) \\
&= \mathbb{E}[(W(z)W(z')] = \mathbb{E}\big[\big(U(z)+V(z)\big)\big(U(z')+V(z')\big)\big] \\
&= \mathbb{E}\big[U(z)U(z')\big] + \mathbb{E}\big[U(z)V(z')\big] + \mathbb{E}\big[V(z)U(z')\big] + \mathbb{E}\big[V(z)V(z')\big] \\
&= [(\dot{F}_t\otimes \dot{F}_{t'})*k_{\mathrm{u}}](x,x') + [(F_t\otimes F_{t'})*k_{\mathrm{v}}](x,x'). \qquad (51)
\end{aligned}
$$

The cross terms are null because $U(z)$ and $V(z')$ are independent as well as $U(z')$ and $V(z)$.

(ii) : with expression (45), one checks that for any fixed z', the function $z \longmapsto k_W(z,z')$ is of the form (38) and thus verifies $\Box k_{x'} = 0$ in the sense of distributions. (ii) is then a direct consequence of Proposition 3.5. $\qquad\Box$

Remark 4.2. If U and V are not independent, then the two terms $[(\dot{F}_t \otimes F_{t'}) * k_{uv}](x,x')$ and $[(F_t \otimes \dot{F}_{t'}) * k_{vu}](x,x')$ must be added to equation (45), where $k_{uv}(x,x')$ denotes the cross co-variance between U and V : $k_{uv}(x,x') = \mathrm{Cov}(U(x),V(x'))$ and $k_{vu}(x,x') = \mathrm{Cov}(V(x),U(x')) = k_{uv}(x',x)$.

More explicitly, we have the following Kirschoff-like integral formulas for $k_{\mathrm{v}}^{\mathrm{wave}}$ and $k_{\mathrm{u}}^{\mathrm{wave}}$:

$$
[(F_t \otimes F_{t'}) * k_{\mathrm{v}}](x,x') = tt' \int_{S(0,1)\times S(0,1)} k_v(x-c|t|\gamma, x'-c|t'|\gamma')\frac{d\Omega d\Omega'}{(4\pi)^2}, \qquad (52)
$$

$$
\begin{aligned}
[(\dot{F}_t \otimes \dot{F}_{t'}) * k_{\mathrm{u}}](x,x') = \int_{S(0,1)\times S(0,1)} \Big(&k_u(x-c|t|\gamma, x'-c|t'|\gamma') \\
&- c|t|\nabla_1 k_u(x-c|t|\gamma, x'-c|t'|\gamma')\cdot\gamma \\
&- c|t'|\nabla_2 k_u(x-c|t|\gamma, x'-c|t'|\gamma')\cdot\gamma' \\
&+ c^2 tt'\gamma^T \nabla_1\nabla_2 k_u(x-c|t|\gamma, x'-c|t'|\gamma')\gamma'\Big)\frac{d\Omega d\Omega'}{(4\pi)^2}. \qquad (53)
\end{aligned}
$$

Above, $\nabla_1 k_u(x,x')$ is the gradient vector of k_u with reference to x, $\nabla_2 k_u(x,x')$ is the gradient vector of k_u with reference to x' and $\nabla_1\nabla_2 k_u(x,x')$ is the matrix whose entry (i,j) is given by

$$
\nabla_1\nabla_2 k_u(x,x')_{ij} = \partial_{x_i^1}\partial_{x_j^2} k_u(x,x'). \qquad (54)
$$

($\partial_{x_i^1}$ is the derivative with reference to the ith coordinate of x, $\partial_{x_j^2}$ is the derivative with reference to the jth coordinate of x').

4.2.1 Extending the covariance functions $k_{\mathrm{u}}^{\mathrm{wave}}$ and $k_{\mathrm{v}}^{\mathrm{wave}}$ to initial conditions u_0 and v_0 with piecewise regularity. The formulas (52) and (53) are valid in a more general context than that of assumptions (A_1) and (A_2). We provide below examples where these formulas yield valid covariance functions (in particular, functions defined for *all* values of (x, t) and (x', t')) corresponding to initial conditions with some forms of piecewise discontinuities. Assume, for example, that the initial speed v_0 is compactly supported on a ball $B(x_0, R)$ centered on some point x_0 with radius R. This is a natural model when v_0 is assumed to be a localized source. For the process V^0, this translates as $V^0(x) = 0$ *a.s.* if x is outside the ball $B(x_0, R)$. One can thus truncate the covariance function of V^0 accordingly, e.g. choosing the following function for k_{v} (see Section 2.1.3 for $k_{1/2}$)

$$k_{\mathrm{v}}(x, x') = k_{1/2}(x, x') \mathbb{1}_{[0,R]}(||x - x_0||) \mathbb{1}_{[0,R]}(||x' - x_0||). \tag{55}$$

Above, $||x||$ denotes the Euclidean norm of x. Such a covariance function indeed verifies $k_{\mathrm{v}}(x, x) = \mathrm{Var}(V^0)(x)) = 0$ if $||x - x_0|| > R$ and the GP corresponding to k_{v} is $V^0(x) = V_{1/2}(x) \mathbb{1}_{[0,R]}(||x - x_0||)$, where $V_{1/2}$ is a continuous modification of a GP with covariance function $k_{1/2}$. Note that the sample paths of V^0 are piecewise continuous and that V as defined in (42) is well-defined and measurable. The integrals in (52) still make sense and point (ii) from Proposition is still valid: the sample paths of the process V whose covariance function is $k_{\mathrm{v}}^{\mathrm{wave}}$ (or any other measurable centered second order random field with this covariance function) remains a solution of the wave equation in the distributional sense. One can perform the same kind of discussions on $k_{\mathrm{u}}^{\mathrm{wave}}$: for example, equation (53) shows that when $k_{\mathrm{u}} \in C^{1,1}(\mathbb{R}^3 \times \mathbb{R}^3) \backslash C^{2,2}(\mathbb{R}^3 \times \mathbb{R}^3)$, $k_{\mathrm{u}}^{\mathrm{wave}}$ is only expected to lie in $C^{1,1}(\mathbb{R}^3 \times \mathbb{R}^3)$; the sample paths of the GP with covariance function $k_{\mathrm{u}}^{\mathrm{wave}}$ will be at most of class C^1 and thus cannot be strong solutions of equation (37). This is the case when k_{u} is the $k_{3/2}$ Matérn covariance function from equation (8).

More generally, one can incorporate a finite number of discontinuities on k_{v} and on the derivatives of k_{u} so that they remain piecewise continuous: the integrals above will remain well defined and the sample paths of the corresponding GPs will remain distributional solutions to the wave equation, even though they will not be sufficiently differentiable to be strong solutions.

5 Conclusion and perspectives

In Section 3, we have presented a new result that provides a simple characterization of the measurable second order random fields $(U(x))_{x \in \mathcal{D}}$ whose sample paths verify homogeneous linear differential constraints within the framework of generalized functions. This characterization is valid for any linear differential operator L, provided that its coefficients fulfil minimal smoothness requirements, and no stationarity assumptions over $(U(x))_{x \in \mathcal{D}}$ are required. Motivated by physical applications, we described in Section 4 a Gaussian process model of the wave equation which is central to describe propagation phenomena. This PDE served as an application case for Proposition 3.5, and the GP model was derived by putting a GP prior on the wave equation's initial conditions. In Proposition 4.1, we presented covariance formulas that are tailored to the wave equation and take the form of convolutions; these expressions are interesting in themselves and call for physics-informed GPR applications for this equation. In particular, we showed that these formulas can model piecewise continuously differentiable solutions for the wave equation. Moreover, this setting provides a natural way to incorporate any type of information, both numerical or experimental. In a forthcoming paper, we will show how to use GPR conjointly with the covariance functions from Proposition 4.1 in a numerical setting, in order to construct approximate solutions of the wave equation based on scattered observations. This in turn provides a natural method for solving different inverse problems, by using the likelihood of GPR as well as the reconstructed solution. An other application concerns the design of transparent boundary

conditions (TBC): this provides artificial boundary conditions on a computational domain so that the computed solution is exactly an approximation of a solution on the whole space. Those conditions are usually nonlocal and restricted to simple geometries. A GPR strategy is meshless therefore suitable to design TBC on any type of computational domains.

Proposition 3.5 constitutes a first step towards understanding PDE constrained random fields in an weakened sense; different functional analysis frameworks can now be considered, obvious extensions being the weak or variational formulations of equation (1). These formulations are obtained by transferring only a part of the derivatives of the PDE to the test function and are for instance the canonical way of studying elliptic PDEs ([16], Section 6.1.2). The natural spaces arising from these formulations are Sobolev spaces rather than $\mathscr{D}'(\mathcal{D})$. An attached question, as studied in [49], is that of the Sobolev regularity of a given second order random field; a current research topic is whether or not one may relax the continuity assumptions required in [49]. Finally, the matter of using random fields for modelling and approximating solutions of nonlinear PDEs is a natural direction for future research.

References

[1] R. J. Adler. *The Geometry of Random Fields.* Classics in Applied Mathematics. Philadelphia, PA: SIAM, 2010.

[2] R. J. Adler and J. E. Taylor. *Random Fields and Geometry.* Springer Monographs in Mathematics. New York, NY: Springer, 2007.

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

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

[5] J.-M. Azaïs and M. Wschebor. *Level sets and extrema of random processes and fields.* Hoboken, NJ: Wiley & Sons, Inc., 2009.

[6] A. Berlinet and C. Thomas-Agnan. *Reproducing Kernel Hilbert Spaces in Probability and Statistics.* New York, NY: Springer, 2004.

[7] H. Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations.* Universitext. New York, NY: Springer, 2010.

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

[9] Y. Chen, B. Hosseini, H. Owhadi, and A. M. Stuart. Solving and learning nonlinear PDEs with Gaussian processes. *J. Comput. Phys.*, 447:110668, 2021.

[10] G. Da Prato and J. Zabczyk. *Stochastic equations in infinite dimensions*, volume 44 of *Encyclopedia of Mathematics and its Applications.* Cambridge: Cambridge University Press, 1992.

[11] J. Dieudonné. Sur les espaces de Montel métrisables. *C. R. Acad. Sci., Paris*, 238:194–195, 1954.

[12] J. L. Doob. Stochastic processes depending on a continuous parameter. *Trans. Amer. Math. Soc.*, 42(1):107–140, 1937.

[13] J. L. Doob. *Stochastic processes*. Wiley Classics Library. New York, NY: John Wiley & Sons, Inc., 1990.

[14] J. J. Duistermaat and J. A. C. Kolk. *Distributions*. Boston, MA: Birkhäuser Boston, Inc., 2010.

[15] A. Estrade and J. Fournier. Anisotropic gaussian wave models. *ALEA Lat. Am. J. Probab. Math. Stat.*, 17, 2020.

[16] L. C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. Providence, RI: American Mathematical Society, second edition, 2010.

[17] L. C. Evans and R. F. Gariepy. *Measure theory and fine properties of functions*. Textbooks in Mathematics. Boca Raton, FL: CRC Press, revised edition, 2015.

[18] M. Fan, D. Paul, T. C. Lee, and T. Matsuo. Modeling tangential vector fields on a sphere. *J. Amer. Statist. Assoc.*, 113(524):1625–1636, 2018.

[19] G. E. Fasshauer. *Meshfree approximation methods with MATLAB*, volume 6 of *Interdisciplinary Mathematical Sciences*. Hackensack, NJ: World Scientific Publishing Co. Pte. Ltd., 2007.

[20] J. Fiedler. *Distances, Gegenbauer expansions, curls, and dimples: On dependence measures for random fields*. PhD thesis, Faculty of Mathematics and Computer Science, 2016.

[21] E. J. Fuselier, Jr. *Refined error estimates for matrix-valued radial basis functions*. Ann Arbor, MI: ProQuest LLC, 2006. Thesis (Ph.D.)–Texas A&M University.

[22] J. Gapaillard and J. Michaux. Sur les processus linéaires définis sur un espace nucléaire. *Ann. Fac. Sci. Toulouse Math. (5)*, 8(1):75–92, 1986/87.

[23] A. Geist and S. Trimpe. Learning constrained dynamics with gauss' principle adhering gaussian processes. In *Proceedings of the 2nd Conference on Learning for Dynamics and Control*, volume 120 of *Proc. Mach. Learn. Res.*, pages 225–234. PMLR, 10–11 Jun 2020.

[24] I. M. Gel'fand and N. Y. Vilenkin. *Generalized functions. Vol. 4: Applications of harmonic analysis*. New York-London: Academic Press, 1964.

[25] 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*, 170:117–128, 2016.

[26] T. Graepel. Solving noisy linear operator equations by gaussian processes: Application to ordinary and partial differential equations. In *Machine Learning, Proceedings of the Twentieth International Conference (ICML 2003), August 21-24, 2003, Washington, DC, USA*, pages 234–241. AAAI Press, 2003.

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

[28] S. Janson. *Gaussian Hilbert spaces*, volume 129 of *Cambridge Tracts in Mathematics*. Cambridge: Cambridge University Press, 1997.

[29] 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.

[30] 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.

[31] 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. New York: Springer, 2015.

[32] S. Lang. *Real and functional analysis*, volume 142 of *Graduate Texts in Mathematics*. New York, NY: Springer-Verlag, third edition, 1993.

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

[34] M. Lange-Hegermann. Linearly constrained Gaussian processes with boundary conditions. In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proc. Mach. Learn. Res.*, pages 1090–1098. PMLR, 13–15 Apr 2021.

[35] J.-F. Le Gall. *Mouvement brownien, martingales et calcul stochastique*, volume 71 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Heidelberg: Springer, 2013.

[36] F. Lindgren, D. Bolin, and H. Rue. The spde approach for gaussian and non-gaussian fields: 10 years and still running. *Spat. Stat.*, page 100599, 2022.

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

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

[39] N. Nguyen and J. Peraire. Gaussian functional regression for linear partial differential equations. *Comput. Methods Appl. Mech. Engrg.*, 287:69–89, 2015.

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

[41] A. Pazy. *Semigroups of linear operators and applications to partial differential equations*, volume 44 of *Applied Mathematical Sciences*. New York: Springer-Verlag, 1983.

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

[43] C. E. Rasmussen and C. K. I. Williams. *Gaussian processes for machine learning*. Adaptive Computation and Machine Learning. Cambridge, MA: MIT Press, 2006.

[44] K. Ritter. *Average-case analysis of numerical problems*, volume 1733 of *Lecture Notes in Mathematics*. Berlin: Springer-Verlag, 2000.

[45] L. Roques, D. Allard, and S. Soubeyrand. Spatial statistics and stochastic partial differential equations: A mechanistic viewpoint. *Spat. Stat.*, page 100591, 2022.

[46] W. Rudin. *Functional analysis*. International Series in Pure and Applied Mathematics. New York, NY: McGraw-Hill, Inc., second edition, 1991.

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

[48] H. H. Schaefer and M. P. Wolff. *Topological vector spaces*, volume 3 of *Graduate Texts in Mathematics*. New York, NY: Springer-Verlag, second edition, 1999.

[49] M. Scheuerer. Regularity of the sample paths of a general second order random field. *Stochastic Process. Appl.*, 120(10):1879–1897, 2010.

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

[51] L. Schwartz. Sous-espaces Hilbertiens d'espaces vectoriels topologiques et noyaux associés (noyaux reproduisants). *J. Anal. Math.*, 13(1):115–256, 1964.

[52] D. Serre. *Systems of conservation laws. Vol 1.* Cambridge: Cambridge University Press, 1999.

[53] I. Steinwart. Convergence types and rates in generic Karhunen-Loeve expansions with applications to sample path properties. *Potential Anal.*, 51(3):361–395, 2019.

[54] F. Trèves. *Topological vector spaces, distributions and kernels*. New York-London: Academic Press, 1967.

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

[56] P. Whittle. On stationary processes in the plane. *Biometrika*, pages 434–449, 1954.