



HAL
open science

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

Iain Henderson, Pascal Noble, Olivier Roustant

► **To cite this version:**

Iain Henderson, Pascal Noble, Olivier Roustant. Characterization of the second order random fields subject to linear distributional PDE constraints. 2022. hal-03770715v1

HAL Id: hal-03770715

<https://hal.science/hal-03770715v1>

Preprint submitted on 6 Sep 2022 (v1), last revised 16 Jan 2023 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

Iain Henderson ^{*1}, Pascal Noble¹, and Olivier Roustant¹

¹Institut de Mathématiques de Toulouse, Université Paul Sabatier, 118 Rte de Narbonne, 31400 Toulouse, France

May 2022

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 U_ω 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 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 is a Gaussian process (GP), the sample paths of U 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 a function of interest $u : \mathcal{D} \rightarrow \mathbb{R}$ is unknown, 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, etc, 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

*henderson@insa-toulouse.fr

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

those that are linear. In this article, we will only consider homogeneous linear PDEs, which take the form

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

Above, u is the unknown function of interest, defined over an open set $D \subset \mathbb{R}^d$, and L is a linear partial differential operator. In (1), for a multi-index $\alpha = (\alpha_1, \dots, \alpha_d)^T \in \mathbb{N}^d$, we used the notations $|\alpha| = \alpha_1 + \dots + \alpha_d$ and $\partial^\alpha = (\partial_{x_1})^{\alpha_1} \dots (\partial_{x_d})^{\alpha_d}$. 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 [46].

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 kernel k : under the assumption that U is a Gaussian process (GP) whose sample paths are n times differentiable, [23] proved for some classes of differential operators L of order n that ([23], Sections 3.3 and 4.1)

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

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; 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 on the distributional formulation of the PDE (1), where the regularity assumptions over u are relaxed to the maximum: this formulation will be our main object of interest in this article. Consider equation (1), and "test it locally": that is, multiply it by a compactly supported, smooth test function $\varphi \in C_c^\infty(D)$ and integrate over D :

$$\forall \varphi \in C_c^\infty(D), \sum_{|\alpha| \leq n} \int_D \varphi(x) a_\alpha(x) \partial^\alpha u(x) dx = 0 \quad (3)$$

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 D , the boundary terms of each integration by parts vanish and we obtain that

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

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 $K \subset D$. 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 $\varphi \in C_c^\infty(D)$), though not pointwise in general (taking $\varphi(x) = \delta(x - x_0)$ is not allowed).

The distributional formulation of the PDE $Lu = 0$ is "compliant with physics" too, as pointed out by W. Rudin ([41], 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 $\sigma : x \mapsto \sqrt{k(x, x)}$ is locally integrable, we show in Proposition 3.4 that

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

Related literature

[49] classifies the stationary generalized random fields that are solutions of a wide class of linear stochastic partial derivative equations (SPDEs), i.e. PDEs where the source term is a random function. In particular, [49] provides a description of all the second order stationary generalized random fields that are solutions to some 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 [49] 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 [23], 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 [23], Sections 3.3 and 4.1). In a previous paper, [45] builds covariance structures 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. [17] extends the results of [45] to random fields on the sphere of \mathbb{R}^3 , which has been rediscovered later in [15]. A simple algorithm for building linearly constrained GPs is proposed in [29], 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 [30], 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 [29] is then supplemented in [30], where parametrizations of the solution spaces of (1) thanks to Gröbner bases are proposed. In [31], the same author completes the approach from [30] by incorporating boundary conditions on hypersurfaces in the Gröbner basis parametrization. With the idea to apply GPR to rigid body dynamics, [20] enforces Gauss' principle of least constraint on the sample paths of a GP.

One can understand our main result (Proposition 3.4) 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' ([24]),

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, that is well suited for handling linear constraints (this is because GPs are "stable under linear combinations", see Section 2.1.3). The recurring idea is to suppose 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 structure 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 structure 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 kernels 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 kernels that verify this constraint is a necessary complement to the condition we prove in this article (Proposition 3.4 is otherwise useless in practice). Explicit PDE constrained kernels were designed for a number of classical PDEs, namely : divergence-free vector fields [35][45], curl-free vector fields [18][45], the Laplace equation [42][34][2], Maxwell's equations [51][29] [30] (although [51] and [29] only exploit curl/divergence free constraints), the 1D heat equation [2], Helmholtz' 2D equations [2], and linear solid mechanics [28]. [31] and [25] enforce homogeneous boundary conditions on the covariance kernel.

We finish by providing 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 $Lu = f$: see [38], [3] and [37] for entry points on the related literature. A recent article [7] 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 [36], the variational formulation (see [13], 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 [5], Section 2.2.1.1), or "functional Gaussian processes" following [36]. The variational formulation of the PDE $Lu = 0$ differs from the distributional formulation (4) in the choice of the space of test functions.

Contribution and organisation of the paper. Given an open set $\mathcal{D} \subset \mathbb{R}^d$, consider a differential operator as in equation (1) whose variable coefficients, defined on \mathcal{D} , 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 $\sigma : x \mapsto \sqrt{k(x, x)}$ is locally integrable, we show in Proposition 3.4 that

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

The result is then compared to a previous result from [23], 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 inherited to pointwise conditioned versions of U when U is a GP, 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. In our specific case, the need for its study naturally arose from acoustical oceanography considerations (see e.g. [33] for an introduction on that topic) in the context of a project funded by the Service Hydrographique et Océanographique de la Marine (SHOM).

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 kernel of the solution process, in the form of convolutions. From Propositions 3.4 and 4.1, we obtain that the sample paths of the corresponding (non stationary) GP all verify the wave equation in the distributional sense. When the covariance kernels of the initial conditions are not smooth enough, the result from [23] cannot be applied. Explicitly, for this PDE, choosing the commonly used 3/2-Matérn kernel for the initial position is enough to land outside the scope of the result from [23] (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 kernels presented in [49] is that here, no stationarity assumptions are made on the solution stochastic process U . In particular the spectral measure provided by Bochner's theorem [39], which is the key tool used in [49], 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 $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. For convenience, we will suppose that it is complete, i.e. that \mathcal{F} contains the subsets of sets $A \in \mathcal{F}$ such that $\mathbb{P}(A) = 0$.

2.1.1 Basic definitions. Let $\mathcal{D} \subset \mathbb{R}^d$ 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 (or sample path) at point $\omega \in \Omega$ to be the deterministic function $x \mapsto U(x)(\omega)$, and we denote it by U_ω . Given an operator acting on the sample paths of U , an event of the form $\{T(U) \in A\}$ will always be understood sample path wise : that is, by definition, $\{T(U) \in A\} := \{\omega \in \Omega : T(U_\omega) \in A\}$. Such sets are not automatically \mathbb{P} -measurable; still, they are measurable as soon as they contain an event of probability 1 (as the ones in Propositions 3.2 and 3.4), since $(\Omega, \mathcal{F}, \mathbb{P})$ 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 ([11], p. 60 or [32], p. 34) if it is measurable seen as a bivariate map $U : (\Omega \times \mathcal{D}, \mathcal{F} \otimes \mathcal{B}(\mathcal{D})) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $(\omega, x) \mapsto U(x)(\omega)$.

Here, $\mathcal{B}(S)$ denotes the Borel σ -algebra of S and $\mathcal{F} \otimes \mathcal{B}(\mathcal{D})$ denotes the product σ -algebra of \mathcal{F} 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 $\varepsilon > 0$, $\mathbb{P}(|U(x) - U(x+h)| > \varepsilon) \rightarrow 0$ when $h \rightarrow 0$. 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 $x \in \mathcal{D}$ ([11], Theorem 2.6, p. 61). One then implicitly works with \tilde{U} . In this article, we will directly suppose 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 weak 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. [10], Theorem 2.3). A famous theorem from Kolmogorov ([8], 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. [4], chapter 1, corollary 1.7.

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

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

A Gaussian process U over \mathcal{D} is a random field over \mathcal{D} such that for any $(x_1, \dots, x_n) \in \mathcal{D}^n$ and any $(a_1, \dots, a_n) \in \mathbb{R}^n$, $\sum_i a_i U(x_i)$ is a Gaussian random variable; that is, the law of $(U(x_1), \dots, U(x_n))^T$ is a multivariate normal distribution. The law of a GP is characterized by its mean and covariance functions ([27], Section 8). We write $(U(x))_{x \in \mathcal{D}} \sim GP(m, k)$. Given a GP U , we will sometimes use the space $\mathcal{L}(U) = \overline{\text{Span}(U(x), x \in \mathcal{D})}$, 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 ([27], 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: $k(x, x') = k(x', x)$ and

$$\sum_{i,j=1}^n a_i a_j k(x_i, x_j) \geq 0 \quad \forall (x_1, \dots, x_n) \in \mathcal{D}^n, \quad \forall (a_1, \dots, a_n) \in \mathbb{R}^n \quad (7)$$

Symmetric positive definite functions verify the Cauchy-Schwarz inequality [39] :

$$\forall x, x' \in \mathcal{D}, \quad |k(x, x')| \leq \sqrt{k(x, x)} \sqrt{k(x', x')} \quad (8)$$

Note that there is a one-to-one correspondence between positive definite kernels and the laws of centered GPs ([11], Theorem 3.1). We provide below two examples of radial Matérn covariance kernels ([39], pp. 84-85), which will be useful in Section 4. Set $r = \|x - x'\|$ the Euclidean distance between x and x' , the following two functions are valid covariance kernels :

$$k_{1/2}(x, x') = \exp\left(-\frac{r}{l}\right) \quad k_{3/2}(x, x') = \left(1 + \frac{r}{l}\right) \exp\left(-\frac{r}{l}\right) \quad (9)$$

These covariance kernels are widely used in machine learning, especially the kernel $k_{3/2}$. The sample paths of a GP with covariance kernel $k_{m/2}$ lie in $C^m(\mathcal{D})$ and not $C^{m+1}(\mathcal{D})$. They are thus commonly encountered to model functions with finite smoothness. Notice that the mean and covariance functions are functions in the classical sense : that is, they are defined everywhere and not up to a set of Lebesgue measure 0. This will be important when discussing admissible covariance kernels in Section 4.

2.2 Tools from functional analysis

We refer to [41] and [48] 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 Locally integrable functions, test functions. The space $C_c^\infty(\mathcal{D})$, which we will rather denote $\mathcal{D}(\mathcal{D})$, designates the space of compactly supported infinitely differentiable functions supported on \mathcal{D} , a.k.a. 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 elements of $L_{loc}^1(\mathcal{D})$ are identified 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 of functions which contains the space of piecewise continuous functions, but also all the local Lebesgue spaces $L_{loc}^p(\mathcal{D}), p \geq 1$. It is in fact the largest space of functions that can be alternatively viewed as continuous linear forms over $\mathcal{D}(\mathcal{D})$ (see Section 2.2.4 below).

2.2.2 Generalized functions. Endow $\mathcal{D}(\mathcal{D})$ with its usual LF-space topology, defined for example in [48], Chapter 13. We call generalized function any continuous linear form on $\mathcal{D}(\mathcal{D})$, i.e. any element of $\mathcal{D}(\mathcal{D})'$, the topological dual of $\mathcal{D}(\mathcal{D})$. We will rather denote it by $\mathcal{D}'(\mathcal{D})$ as in [48], Notation 21.1. The topology of $\mathcal{D}(\mathcal{D})$ is such that $T \in \mathcal{D}'(\mathcal{D})$ if and only if for all compact set $K \subset \mathcal{D}$, there exists $C_K > 0$ and a non negative integer n_K such that

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

Above, we used the following notations. For a multi-index $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}^d$, we use the usual notations $|\alpha| = \alpha_1 + \dots + \alpha_d$ and $\partial^\alpha := \partial_{x_1}^{\alpha_1} \dots \partial_{x_d}^{\alpha_d}$ where $\partial_{x_i}^{\alpha_i}$ is the α_i^{th} derivative w.r.t the i^{th} 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 $\langle \cdot, \cdot \rangle$: for all $\varphi \in \mathcal{D}(\mathcal{D})$ and $T \in \mathcal{D}'(\mathcal{D})$, we have $\langle T, \varphi \rangle = T(\varphi)$.

2.2.3 Generalized functions and differentiation. Any generalized function L can be indefinitely differentiated ([41], Section 6.12, p. 158 or [48], pp. 248-250) with the following definition

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

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

2.2.4 Regular generalized functions. Any function $f \in L_{loc}^1(\mathcal{D})$ can be injectively identified to a generalized function T_f ([48], p. 224 or [41], Section 6.11, p. 157) defined as follow

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

The map $L_{loc}^1(\mathcal{D}) \ni f \mapsto T_f$ is linear and injective; any generalized function L that is of the form T_f for some $f \in L_{loc}^1(\mathcal{D})$ 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. Note that equations (11) and (12) combined provide a flexible definition of the derivatives of any function $f \in L_{loc}^1(\mathcal{D})$ up to any order.

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 paper, we call positive Radon measure any positive measure over \mathcal{D} that is Borel regular ([14], 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 [47], 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 [48]. The corresponding isomorphism is given by

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

We have the following facts. (i) Any signed measure that admits a density f w.r.t. the Lebesgue measure such that $f \in L_{loc}^1(\mathcal{D})$ is a Radon measure ([48], p. 217). (ii) Any Radon measure can be injectively identified to a generalized function T_μ by replacing $C_c(\mathcal{D})$ by $\mathcal{D}(\mathcal{D})$ 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 ([47], Chapter 9). We then define its total variation by $|\mu| := \mu^+ + \mu^-$.

Remark 2.1 (Literature guide on Radon measures). What is meant behind the terminology of Radon measures varies between authors. [14] calls Radon measure what we call *positive* Radon measure in this article. [47] proves that continuous linear forms over $C_c(\mathcal{D})$ are differences of Radon measures in the sense of the Radon measures defined in [14], but [47] never uses the term of Radon measures, positive or not. Likewise, [48] calls *positive* Radon measure any *positive* linear form over $C_c(\mathcal{D})$ which, thanks to the proof from [47], reduces to Radon measures in the sense of [14].

2.2.6 Finite order generalized functions. A generalized function L is said to be of finite order if there exists a non negative integer k such that one can take $n_K = k$, independently of K , in the definition of the continuity of L , i.e. equation (10). The order of L is then the smallest of those integers k . The space of generalized functions of order k is isomorphic to $C_c^k(\mathcal{D})'$, the space of continuous linear forms over $C_c^k(\mathcal{D})$, when $C_c^k(\mathcal{D})$ is endowed with its usual LF-space topology. The key property for us is that such generalized functions can be represented thanks to Radon measures. If L is of order k , there exists a family of Radon measures $\{\mu_p\}_{|p| \leq k}$ over \mathcal{D} such that

$$T = \sum_{|p| \leq k} \partial^p \mu_p \quad (14)$$

where the equality in (14) holds in $\mathcal{D}'(\mathcal{D})$ and $C_c^k(\mathcal{D})'$ ([48], 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 $T = \sum_{|p| \leq k} \partial^p \mu_p$ all have compact support.

2.2.7 Convolution with generalized functions. As above, we consider $C_c^k(\mathbb{R}^d)$ endowed with its usual LF-space topology. Let $f \in C_c^k(\mathbb{R}^d)$ and $T \in C_c^k(\mathbb{R}^d)'$. Note $\tau_x f$ the function $y \mapsto f(y-x)$ and \check{f} the function $y \mapsto f(-y)$. Then ([48], p. 287, Section 27) one may define the convolution between L and f by

$$T * f : x \mapsto \langle T, \tau_{-x} \check{f} \rangle \quad (15)$$

and $T * f$ is a function in the classical sense, i.e. defined pointwise. When L lies in $L_{loc}^1(\mathcal{D})$, equation (15) reduces to the usual convolution of functions through the identification defined in equation (12). Similarly if L is in fact a Radon measure μ :

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

More general definitions of generalized function convolution are available ([48], Chapter 27) but this one is sufficient for our use.

2.2.8 Tensor product of generalized functions. For two generalized functions $T_1 \in \mathcal{D}'(\mathcal{D}_1)$ and $T_2 \in \mathcal{D}'(\mathcal{D}_2)$, $T_1 \otimes T_2 \in \mathcal{D}'(\mathcal{D}_1 \times \mathcal{D}_2)$ denotes their tensor product ([48], pp. 416-417), which is uniquely determined by the following tensor property :

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

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

3 Random fields under linear differential constraints

The results in this Section state that under suitable assumptions on the first two moments of a given second order random field $U = (U(x))_{x \in \mathcal{D}}$, sample path degeneracy properties w.r.t. differential constraints can be read on the these two first moments, namely the mean function and the functions $k_x : y \mapsto k(x, y)$, 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 [23] in the case of pointwise defined derivatives. We next state and prove a result similar to that of [23], 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 (1), with continuous coefficients. We say that a function u is a classical or strong solution to the PDE $Lu = 0$ if u is n times differentiable and u verifies the PDE pointwise :

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

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 [23] is best understood. This theorem, which we remind in Proposition 3.2, is the one proved and used in [23] 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 kernel k . Note $\mathcal{F}(\mathcal{D}, \mathbb{R})$ the space of real-valued pointwise-defined functions on \mathcal{D} . Denote \mathcal{H}_k the reproducing kernel Hilbert space (RKHS) associated to k (see [5], Definition 1 p. 7 and Theorem 3, p. 19). \mathcal{H}_k is a Hilbert space of pointwise-defined functions, 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 often clashes with the usual L^p and Sobolev spaces encountered in PDE theory, which are sets of functions defined up to a set of Lebesgue measure 0.

Proposition 3.2 (sample paths of GPs under linear constraints [23]). *Let $(U(x))_{x \in \mathcal{D}} \sim GP(0, k)$ be a centered GP. Note for all $x \in \mathcal{D}$ the function $k_x : y \mapsto k(x, y)$. Let E be a real vector space of functions defined on \mathcal{D} that contains the sample paths of U almost surely and $T : E \rightarrow \mathcal{F}(\mathcal{D}, \mathbb{R})$ be a linear operator. Suppose that for all $x \in \mathcal{D}$, $T(U)(x) \in \mathcal{L}(U)$. Then there exists a unique linear operator $\mathcal{T} : \mathcal{H}_k \rightarrow \mathcal{F}(\mathcal{D}, \mathbb{R})$ such that for all $x, x' \in \mathcal{D}$,*

$$\mathbb{E}[T(U)(x)U(x')] = \mathcal{T}(k_{x'})(x)$$

and $\forall x \in \mathcal{D}, \forall h_n \xrightarrow{\mathcal{H}_k} h, \mathcal{T}(h_n)(x) \rightarrow \mathcal{T}(h)(x)$. Moreover, the following statements are equivalent :

- (i) $\mathbb{P}(T(U) = 0) = 1$
- (ii) $\forall x \in \mathcal{D}, \mathcal{T}(k_x) = 0$
- (iii) $\mathcal{T}(\mathcal{H}_k) = \{0\}$

The proof of Proposition 3.2 relies heavily on the Loève isometry ([5], Theorem 35, p. 65) between the two Hilbert spaces \mathcal{H}_k and $\mathcal{L}(U)$ (see Section 2.1.3 for $\mathcal{L}(U)$). This theorem can be applied when L is a differential operator as discussed in [23]. However, in Proposition 3.2, the differential operator L of order n has to be valued in the space of pointwise defined functions $\mathcal{F}(\mathcal{D}, \mathbb{R})$; 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 $\mathcal{D}^n(\mathcal{D})$, the space of n times differentiable functions on \mathcal{D} . Requiring that $E \subset \mathcal{D}^n(\mathcal{D})$ is a very strong assumption w.r.t. 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 3.4 an adaptation of Proposition 3.2 where we make use of weaker definitions of derivatives. We also relax the assumptions made on U and its sample paths. In these propositions, $(U(x))_{x \in \mathcal{D}}$ is not supposed Gaussian and is only required to be 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 $L = \sum_{|\alpha| \leq n} a_\alpha(x) \partial^\alpha$ be a linear differential operator, and suppose for the moment that its coefficients are smooth. 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 $Lu = 0$, multiply this PDE by a test function $\varphi \in \mathcal{D}(\mathcal{D})$, 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 and we have that

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

Following equation (19) we introduce L^* , the formal adjoint of L , acting on $\mathcal{D}(\mathcal{D})$, defined by the following formula ([48], pp. 247-249)

$$T^* : \varphi \mapsto \sum_{|\alpha| \leq n} (-1)^{|\alpha|} \partial^\alpha (a_\alpha \varphi) \quad (20)$$

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

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

defines a continuous linear form over $\mathcal{D}(\mathcal{D})$, i.e. $L(u) \in \mathcal{D}'(\mathcal{D})$ (see equations (10) and (27) for a rigorous proof of this statement). By construction, L and L^* verify the following duality identity

$$\forall \varphi \in \mathcal{D}(\mathcal{D}), \quad \langle T(u), \varphi \rangle = \langle u, T^*(\varphi) \rangle \quad (22)$$

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 [48], pp. 247-251). This finally leads to the following definition, following e.g. [26] :

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 $\mathcal{D}'(\mathcal{D})$, i.e. when $L(u)$, is seen as an element of $\mathcal{D}'(\mathcal{D})$ through equation (21), is the null linear form over $\mathcal{D}(\mathcal{D})$.

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

Proposition 3.4 (sample paths of random fields under linear differential constraints, distributional derivatives). *Let $\mathcal{D} \subset \mathbb{R}^d$ be an open set and let $T = \sum a_\alpha(x) \partial^\alpha$, $|\alpha| \leq n$, be a linear differential operator of order n with coefficients $a_\alpha(x) \in C^{|\alpha|}(\mathcal{D})$. Let $U = (U(x))_{x \in \mathcal{D}}$ be a measurable second order stochastic process with mean function $m(x)$ and covariance kernel $k(x, x')$. For all $x \in \mathcal{D}$, note $k_x : y \mapsto k(x, y)$. Suppose that its mean function m lies in $L^1_{loc}(\mathcal{D})$ as well as its standard deviation function σ (see equation (6)).*

1) *Then on a set of probability 1, the sample paths of U lie in $L^1_{loc}(\mathcal{D})$ as well as the functions k_x for all $x \in \mathcal{D}$.*

2) *Suppose that $T(m) = 0$ in the sense of distributions. Then the following statements are equivalent :*

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

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

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

$$\forall \varphi \in \mathcal{D}(\mathcal{D}), \quad \langle U_\omega, T^* \varphi \rangle = \int_{\mathcal{D}} U_\omega(x) T^* \varphi(x) dx = 0 \quad (23)$$

The fact that the functions $x \mapsto U_\omega(x)$ and $y \mapsto k_x(y)$ lie in $L^1_{loc}(\mathcal{D})$ ensure the existence of the integrals in equations (23) (see point 2 of the proof of Proposition 3.4) as well as the continuity of the associated linear forms over $\mathcal{D}(\mathcal{D})$, 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 the 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). On the other hand, 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 3.4 :

Lemma 3.5. *$\mathcal{D}(\mathcal{D})$ is a separable topological space, i.e. there exists a countable subset $F \subset \mathcal{D}(\mathcal{D})$ that is dense in $\mathcal{D}(\mathcal{D})$ for its LF-topology.*

This result is difficult to track down in the literature, although it seems to be known in parts of the functional analysis community : see [19], Corollaire (1).2, p. 78 for a reference in French. It can also be found in the technical reference [21], p. 73, (3); but the proof presented in this reference (more precisely, the last theorem on page 58, Section 6.5 in [22] on which it depends) relies on the inconvenient axiom of choice. For self-containment, we thus provide below a proof of this lemma based on the notions of Fréchet, LF and Montel spaces. For more details on these spaces, we refer to [48], Chapters 10, 13 and 34.

Proof. $\mathcal{D}(\mathcal{D})$ is an LF-space as the inductive limit of the Fréchet spaces $\mathcal{D}_{K_i}(\mathcal{D}) := \{\varphi \in C^\infty(\mathcal{D}) : \text{Supp}(\varphi) \subset K_i\}, i \in \mathbb{N}$, where $K_1 \subset K_2 \subset \dots$ are compact subsets of \mathcal{D} such that $\bigcup_i K_i = \mathcal{D}$ ([48], pp. 131-133). As such, $\mathcal{D}(\mathcal{D})$ is separable iff $\mathcal{D}_{K_i}(\mathcal{D})$ is separable for all $i \in \mathbb{N}$ [50], which we now show. The Fréchet topology of $\mathcal{D}_{K_i}(\mathcal{D})$ is the one induced by the usual Fréchet topology of $C^\infty(\mathcal{D})$ when $\mathcal{D}_{K_i}(\mathcal{D})$ is seen as a subspace of $C^\infty(\mathcal{D})$ ([48], pp. 131-132). As a Fréchet space, $C^\infty(\mathcal{D})$ is metrizable ([48], p. 85). But $C^\infty(\mathcal{D})$ is also a Montel space ([48], Proposition 34.4, p. 357) : as a metrizable Montel space, it is automatically separable ([43], p. 195 or [9]). Thus $\mathcal{D}_{K_i}(\mathcal{D})$ is also separable as a subset of the separable metric space $C^\infty(\mathcal{D})$ ([6], Proposition 3.25, p. 73). \square

We are now able to prove Proposition 3.4.

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, $\mathbb{E}[|U(x)|] \leq \sigma(x)$. Now, let $(K_n)_{n \in \mathbb{N}}$ be an increasing sequence of compact subsets of \mathcal{D} such that $\bigcup_{n \in \mathbb{N}} K_n = \mathcal{D}$. Using Tonelli's theorem, we have that for any $n \in \mathbb{N}$,

$$\mathbb{E} \left[\int_{K_n} |U(x)| dx \right] = \int_{K_n} \mathbb{E}[|U(x)|] dx \leq \int_{K_n} \sigma(x) dx < +\infty \quad (24)$$

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 stochastic process cannot be circumvented. Using the property that " $\mathbb{E}[|X|] < +\infty \implies |X| < +\infty$ almost surely", this yields a set $B_n \subset \Omega$ of probability 1

such that the random variable $\omega \mapsto \int_{K_n} |U_\omega(x)| dx$ takes finite values over B_n (from Fubini's theorem again, the map $\omega \mapsto \int_{K_n} |U_\omega(x)| dx$ is indeed 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 $\omega \in B$,

$$\int_K |U_\omega(x)| dx \leq \int_{K_{n_K}} |U_\omega(x)| dx < +\infty \quad (25)$$

which shows that the sample paths of U lie in $L^1_{loc}(\mathcal{D})$ almost surely.

Now, we check that for all $x \in \mathcal{D}$, $k_x \in L^1_{loc}(\mathcal{D})$: for any compact set K , since $\sigma \in L^1_{loc}(\mathcal{D})$ and because of (8),

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

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

$$\forall \varphi \in \mathcal{D}(\mathcal{D}), \quad \forall x \in \mathcal{D}, \quad T^* \varphi(x) = \sum_{|\alpha| \leq n} f_\alpha(x) \partial^\alpha \varphi(x) \quad (26)$$

For all $f \in L^1_{loc}(\mathcal{D})$, for all compact set $K \subset \mathcal{D}$ and for all $\varphi \in \mathcal{D}(\mathcal{D})$ such that $\text{Supp}(\varphi) \subset K$, (26) yields

$$\begin{aligned} |\langle f, T^* \varphi \rangle| &\leq \int_{\mathcal{D}} |f(x)| |T^* \varphi(x)| dx \\ &\leq \left(\int_K |f(x)| dx \times \max_{|\alpha| \leq n} \sup_{x \in K} |f_\alpha(x)| \right) \sum_{|\alpha| \leq n} \|\partial^\alpha \varphi\|_\infty < +\infty \end{aligned} \quad (27)$$

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

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

$$\int_{\mathcal{D}} U_\omega(x) T^* \varphi(x) dx = 0$$

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

$$\begin{aligned} 0 &= \mathbb{E} \left[U(x') \int_{\mathcal{D}} U(x) T^* \varphi(x) dx \right] = \int_{\mathcal{D}} T^* \varphi(x) \mathbb{E}[U(x) U(x')] dx \\ &= \int_{\mathcal{D}} T^* \varphi(x) k(x, x') dx = \langle k_{x'}, T^* \varphi \rangle \end{aligned}$$

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

and the Cauchy-Schwarz inequality :

$$\begin{aligned} \mathbb{E} \left[\int_{\mathcal{D}} |U(x')U(x)T^*\varphi(x)|dx \right] &= \int_{\mathcal{D}} |T^*\varphi(x)|\mathbb{E}[|U(x)U(x')|]dx \\ &\leq \int_{\mathcal{D}} |T^*\varphi(x)|\mathbb{E}[U(x)^2]^{1/2}\mathbb{E}[U(x')^2]^{1/2}dx \\ &\leq \sigma(x') \int_{\mathcal{D}} |T^*\varphi(x)|\sigma(x)dx < +\infty \end{aligned}$$

Indeed, since $\sigma \in L^1_{loc}(\mathcal{D})$, setting $f = \sigma$ in (27) shows that the last integral is finite. Thus, $\forall x \in \mathcal{D}, \forall \varphi \in \mathcal{D}(\mathcal{D}), \langle k_x, T^*\varphi \rangle = 0$ which proves that (i) \implies (ii).

(ii) \implies (i) : Suppose (ii). Let $\varphi \in \mathcal{D}(\mathcal{D})$, we have $\langle k_{x'}, T^*\varphi \rangle = 0$. Multiplying this with $\overline{T^*\varphi(x')}$ and integrating w.r.t. x' yields

$$0 = \int_{\mathcal{D}} T^*\varphi(x') \int_{\mathcal{D}} T^*\varphi(x)k(x, x')dx dx' = \int_{\mathcal{D}} \int_{\mathcal{D}} T^*\varphi(x)T^*\varphi(x')\mathbb{E}[U(x)U(x')]dx dx'$$

Permuting formally the expectancy and the integrals (justified in equation (28)) yields

$$\begin{aligned} 0 &= \int_{\mathcal{D}} \int_{\mathcal{D}} T^*\varphi(x)T^*\varphi(x')\mathbb{E}[U(x)U(x')]dx dx' \\ &= \mathbb{E} \left[\left(\int_{\mathcal{D}} T^*\varphi(x)U(x)dx \right)^2 \right] = \mathbb{E}[\langle U, T^*\varphi \rangle^2] \end{aligned}$$

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

$$\begin{aligned} \int_{\mathcal{D}} \int_{\mathcal{D}} |T^*\varphi(x)T^*\varphi(x')|\mathbb{E}[|U(x)U(x')|]dx dx' & \tag{28} \\ &\leq \int_{\mathcal{D}} \int_{\mathcal{D}} |T^*\varphi(x)T^*\varphi(x')|\sigma(x)\sigma(x')dx dx' \\ &\leq \left(\int_{\mathcal{D}} |T^*\varphi(x)|\sigma(x)dx \right)^2 < +\infty \end{aligned}$$

As previously, setting $f = \sigma$ in (27) shows that the integral above is 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 $\forall \omega \in A, \langle U_\omega, T^*\varphi \rangle = 0$. For this we use the fact that $\mathcal{D}(\mathcal{D})$ is a separable topological space, see Lemma 3.5. Let $F \subset \mathcal{D}(\mathcal{D})$ be a countable dense subset of $\mathcal{D}(\mathcal{D})$, and let

$$A := B \cap \left(\bigcap_{\varphi \in F} A_\varphi \right) \tag{29}$$

Then $\mathbb{P}(A) = 1$ since F is countable. Let $\omega \in A$. Since $U_\omega \in L^1_{loc}(\mathcal{D})$, (27) shows that the map $L_\omega : \varphi \longmapsto \langle U_\omega, T^*\varphi \rangle$ is a continuous linear form on $\mathcal{D}(\mathcal{D})$. In particular, Theorem 2.1.4 from [26] states that L_ω is in fact *sequentially* continuous. Let $\varphi \in \mathcal{D}(\mathcal{D})$ and a sequence $(\varphi_n) \subset F$ such that $\varphi_n \rightarrow \varphi$ in the sense of the LF topology of $\mathcal{D}(\mathcal{D})$. Then $L_\omega(\varphi) = \lim_n L_\omega(\varphi_n) = 0$ since $\forall n, L_\omega(\varphi_n) = 0$. That is, we have proved that

$$\forall \omega \in A, \forall \varphi \in \mathcal{D}(\mathcal{D}), \langle U_\omega, T^*\varphi \rangle = L_\omega(\varphi) = 0$$

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

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

$$\begin{aligned} \mathbb{P}(T(U) = 0 \text{ in the sense of distributions}) &= 1 \\ \iff \mathbb{P}(T(V) = 0 \text{ in the sense of distributions}) &= 1 \iff \forall x \in \mathcal{D}, T(k_x) = 0 \end{aligned}$$

which finishes the proof in the general case. □

Remark 3.6. 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 non linear hyperbolic PDEs [46]. However, every step in the above proof remains valid when replacing $\varphi \in \mathcal{D}(\mathcal{D})$ with $\varphi \in C_c^n(\mathcal{D})$. Although we have not clarified 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 3.4 also hold in $C_c^n(\mathcal{D})'$, the space of finite order generalized functions of order n , rather than just in $\mathcal{D}'(\mathcal{D})$. $C_c^n(\mathcal{D})'$ is a smaller space than $\mathcal{D}'(\mathcal{D})$, though less used in functional analysis than $\mathcal{D}'(\mathcal{D})$.

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 $T = \sum_{|\alpha| \leq n} a_\alpha(x) \partial^\alpha$, then we simply have $\mathcal{T} = T$ in Proposition 3.2. Additionally, $T(U_\omega)$ and $T(k_x)$ both lie in $\mathcal{F}(\mathcal{D}, \mathbb{R}) \cap L_{loc}^1(\mathcal{D})$. In that framework, Proposition 3.2 states that

$$\forall x \in \mathcal{D}, T(k_x) = 0 \iff \mathbb{P}(T(U) = 0) = 1 \quad (30)$$

where the function equalities of the form $T(f) = 0$ in (30) are valid everywhere on \mathcal{D} . In contrast, for any function g that lies in $L_{loc}^1(\mathcal{D})$, we have

$$g = 0 \text{ in the sense of distributions} \iff g = 0 \text{ a.e.} \quad (31)$$

Equation (31) is just another way of saying that the linear map $f \mapsto T_f$ given in (12) is injective. Following equation (31), Proposition 3.4 states a slightly weaker result than (30), namely that

$$\forall x \in \mathcal{D}, T(k_x) = 0 \text{ a.e.} \iff \mathbb{P}(T(U) = 0 \text{ a.e.}) = 1 \quad (32)$$

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 (30) from equation (32).

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 $B = \{u(x_1), \dots, u(x_n)\}$. 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} with $\tilde{U}(x) := (U(x) | U(x_i) = u(x_i), i = 1, \dots, n)$. 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 $X = (x_1, \dots, x_n)^T$, 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 [39]

$$\begin{cases} \tilde{m}(x) &= m(x) + k(X, x)^T k(X, X)^{-1} (u(X) - m(X)) \end{cases} \quad (33)$$

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

The Kriging standard deviation function is then given by

$$\tilde{\sigma}(x) = \sqrt{\tilde{k}(x, x)} \quad (35)$$

The so-called Kriging mean \tilde{m} is then an approximation of u ; in particular, it interpolates u at the observation points : $\tilde{m}(x_i) = u(x_i)$ for all i . 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 3.4 when applied to GPR.

Proposition 3.7 (Hereditiy of Proposition 3.4 to conditioned GPs). *Let \mathcal{D} and L be as defined in Proposition 3.4. Let $(U(x))_{x \in \mathcal{D}} \sim GP(m, k)$ be a Gaussian process that verifies the hypotheses of Proposition 3.4. Suppose also that*

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

(i) *Then whatever the integer p , the vector $u = (u_1, \dots, u_p)^T \in \mathbb{R}^p$ and the vector $X = (x_1, \dots, x_p)^T \in \mathcal{D}^p$ 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*

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

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 $\tilde{U}(x) = (U(x) | U(x_i) = u_i \forall i = 1, \dots, p)$ are almost surely solutions of the equation $T(f) = 0$ in the sense of distributions :*

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

Proof. Note first that for all $x \in \mathcal{D}$, $\tilde{k}(x, x) \leq k(x, x)$ ([16], p. 117). Thus the function $\tilde{\sigma} : x \mapsto \sqrt{\tilde{k}(x, x)}$ 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 3.4 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 Section 3.3.1, equations (33) and (34)). \square

Proposition 3.7 shows that when U is a GP, the results of Proposition 3.4 are inherited on the conditioned posterior process \tilde{U} . One weak consequence of Proposition 3.7 is that if GPR is performed with a kernel k that verifies point (ii) of Proposition 3.4, then the predictions provided by GPR are all solutions of the PDE $T(\tilde{m}) = 0$.

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 ([13], Section 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 discontinuity surfaces are "nice enough". This is an advantage w.r.t. the previous models, where the sample paths actually had to be sufficiently differentiable to obtain sample path degeneracy w.r.t. the PDE.

4.1 General Solution to the 3 Dimensional Wave Equation

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

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

Throughout this article, we will refer to u_0 as the initial position and v_0 as the initial speed. The problem (37) is a Cauchy problem with initial conditions (IC) u_0 and v_0 . It admits a unique solution which can be extended to all times $t \in \mathbb{R}$, and is represented as follow ([13], Section 4.3.1.b, example 4)

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

where F_t and \dot{F}_t are known generalized functions. Notably, F_t corresponds to the *Green's function* of the wave equation [12]. In dimension 3, F_t and \dot{F}_t are compactly supported generalized functions of order 0 and 1 respectively. They are given by the formulas

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

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.

Suppose that $u_0 \in C^1(\mathbb{R}^3)$ and $v_0 \in C^0(\mathbb{R}^3)$, then w as defined in (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 [48], Chapter 27).

Equation (38) can be written down using means over spheres. Denote (r, θ, ϕ) the spherical coordinates, $S(0, 1)$ the unit sphere of \mathbb{R}^3 and $\gamma = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)^T \in S(0, 1)$ the corresponding parametrization of $S(0, 1)$ ($\|\gamma\| = 1$). We write $d\Omega = \sin \theta d\theta d\phi$ the surface

differential element of $S(0, 1)$. Combining formulas (38) and (39) leads to the Kirschhoff formula ([13], p. 72)

$$w(x, t) = \int_{S(0,1)} tv_0(x - c|t|\gamma) + u_0(x - c|t|\gamma) - c|t|\gamma \cdot \nabla u_0(x - c|t|\gamma) \frac{d\Omega}{4\pi} \quad (40)$$

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. Thus, we 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.

- (A₁) Suppose that the initial conditions u_0 and v_0 of Problem (37) are sample paths drawn from two independent Gaussian processes $U^0 \sim GP(0, k_u)$ and $V^0 \sim GP(0, k_v) : \exists \omega \in \Omega, \forall x \in \mathbb{R}^3, u_0(x) = U_\omega^0(x)$ and $v_0(x) = V_\omega^0(x)$.
- (A₂) 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 [1], 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 kernels from equation (9).

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

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

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 (A₂), (41) defines a random variable for all (x, t) . Note the space-time variable $z = (x, t)$ and note the random variables

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

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 kernels.

Proposition 4.1. *Define the two functions*

$$k_v^{\text{wave}}(z, z') = [(F_t \otimes F_{t'}) * k_v](x, x') \quad (43)$$

$$k_u^{\text{wave}}(z, z') = [(\dot{F}_t \otimes \dot{F}_{t'}) * k_u](x, x') \quad (44)$$

(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 kernels k_u^{wave} and k_v^{wave} respectively. Consequently, $(W(z))_{z \in \mathbb{R}^3 \times \mathbb{R}}$ is a centered GP whose covariance kernel is given by

$$k_W(z, z') = k_v^{\text{wave}}(z, z') + k_u^{\text{wave}}(z, z') \quad (45)$$

(ii) Conversely, any centered second order random field with covariance kernel k_W has its sample paths solution of the 3 dimensional 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 ([48], 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. Also, see section 2.2.8 for the definition of $\dot{F}_t \otimes \dot{F}_{t'}$.

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 Kirschhoff formula (40), writing the integrals as limits of Riemann sums. We start with V , that is, we focus on the first term in Kirschhoff'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_α^n and points y_k^n such that for almost any $\omega \in \Omega$,

$$\begin{aligned} V(z)(\omega) &= (F_t * V_\omega^0)(x) = t \int_{S(0,1)} V^0(x - c|t|\gamma)(\omega) \frac{d\Omega}{4\pi} \\ &= \frac{t}{4\pi} \int_0^{2\pi} \int_0^\pi V^0(x - c|t|\gamma(\theta, \phi))(\omega) \sin(\theta) d\theta d\phi = \lim_{n \rightarrow \infty} \sum_{\alpha=1}^n a_\alpha^n V^0(x - y_k^n)(\omega) \end{aligned}$$

This shows that $V(z)$ is an a.s. limit of a sequence of centered Gaussian random variables ($Y_n \in \mathcal{L}(V^0)$); a.s. convergence implies convergence in law. From [32], 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 Kirschhoff'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 (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

$$\mathbb{E}[U(z)U(z')] = [(\dot{F}_t \otimes \dot{F}_{t'}) * k_u](x, x') \quad (46)$$

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 the a sum of derivatives of measures (see equation (14)) : for all $t \in \mathbb{R}$, there exists $\{\mu_\alpha^t\}_{\alpha \in \mathbb{N}^3, |\alpha| \leq 1}$ a family of Radon measures such that

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

Moreover, \dot{F}_t is compactly supported, therefore all the measures μ_α^t are also compactly sup-

ported. First, we write $U_\omega(z)$ in integral form :

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

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

Before applying Fubini's theorem, we need to check an integrability condition. Let $\alpha \in \mathbb{N}^3$ such that $|\alpha| \leq 1$. Recall that $|\mu_\alpha^t|$ is defined in Section 2.2.5; denote also $\sigma_{\partial^\alpha U^0}(x) = \sqrt{\text{Var}(\partial^\alpha U_0(x))}$. 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 $x \mapsto \text{Var}(\partial^\alpha U_0(x))$ also lies in $C^0(\mathcal{D})$ ([4], chapter 1, Section 4.3). Therefore the function $x \mapsto \sigma_{\partial^\alpha U^0}(x)$ 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:

$$\begin{aligned} I &:= \int_{\Omega} \sum_{|\alpha| \leq 1} \int_{\mathbb{R}^3} \left| \partial^\alpha U_\omega^0(x-y) \right| |\mu_\alpha^t(dy)| \sum_{|\alpha'| \leq 1} \int_{\mathbb{R}^3} \left| \partial^{\alpha'} U_\omega^0(x'-y') \right| |\mu_{\alpha'}^{t'}(dy')| \mathbb{P}(d\omega) \\ &= \sum_{|\alpha|, |\alpha'| \leq 1} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \int_{\Omega} \left| \partial^\alpha U_\omega^0(x-y) \partial^{\alpha'} U_\omega^0(x'-y') \right| \mathbb{P}(d\omega) |\mu_\alpha^t(dy)| |\mu_{\alpha'}^{t'}(dy')| \\ &= \sum_{|\alpha|, |\alpha'| \leq 1} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \mathbb{E} \left[\left| \partial^\alpha U^0(x-y) \partial^{\alpha'} U^0(x'-y') \right| \right] |\mu_\alpha^t(dy)| |\mu_{\alpha'}^{t'}(dy')| \\ &\leq \sum_{|\alpha|, |\alpha'| \leq 1} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \left(\mathbb{E}[\partial^\alpha U^0(x-y)^2] \mathbb{E}[\partial^{\alpha'} U^0(x'-y')^2] \right)^{1/2} |\mu_\alpha^t(dy)| |\mu_{\alpha'}^{t'}(dy')| \\ &\leq \left(\sum_{|\alpha| \leq 1} \int_{\mathbb{R}^3} \left(\mathbb{E}[\partial^\alpha U^0(x-y)^2] \right)^{1/2} |\mu_\alpha^t(dy)| \right) \times \left(\sum_{|\alpha'| \leq 1} \int_{\mathbb{R}^3} \left(\mathbb{E}[\partial^{\alpha'} U^0(x'-y')^2] \right)^{1/2} |\mu_{\alpha'}^{t'}(dy')| \right) \\ &\leq \left(\sum_{|\alpha| \leq 1} (|\mu_\alpha^t| * \sigma_{\partial^\alpha U^0})(x) \right) \times \left(\sum_{|\alpha'| \leq 1} (|\mu_{\alpha'}^{t'}| * \sigma_{\partial^{\alpha'} U^0})(x') \right) < +\infty \end{aligned}$$

For all multi-index α , the scalar $(|\mu_\alpha^t| * \sigma_{\partial^\alpha U^0})(x)$ is finite because $x \mapsto \sigma_{\partial^\alpha U^0}(x)$ is continuous and $|\mu_\alpha^t|$ is compactly supported. Note also that from hypothesis (A_2) , the GP U^0 is *mean square differentiable* up to order 1, which implies ([40], Section III.1.4) that we have, for all multi-indexes α, α' such that $|\alpha|, |\alpha'| \leq 1$, x and x' :

$$\mathbb{E}[\partial^\alpha U^0(x) \partial^{\alpha'} U^0(x')] = \partial_1^\alpha \partial_2^{\alpha'} k_u(x, x') \quad (50)$$

where ∂_1 (resp. ∂_2) denotes derivatives w.r.t. the first (resp. second) argument of k_u . We may thus permute integrals and differential operators in $\mathbb{E}[U(z)U(z')]$:

$$\begin{aligned} \mathbb{E}[U(z)U(z')] &= \mathbb{E} \left[\sum_{|\alpha| \leq 1} \int_{\mathbb{R}^3} (-1)^{|\alpha|} \partial^\alpha U^0(x-y) \mu_\alpha^t(dy) \sum_{|\alpha'| \leq 1} \int_{\mathbb{R}^3} (-1)^{|\alpha'|} \partial^{\alpha'} U^0(x'-y') \mu_{\alpha'}^{t'}(dy') \right] \\ &= \sum_{|\alpha|, |\alpha'| \leq 1} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} (-1)^{|\alpha|} (-1)^{|\alpha'|} \partial_1^\alpha \partial_2^{\alpha'} \mathbb{E}[U^0(x-y)U^0(x'-y')] \mu_\alpha^t(dy) \mu_{\alpha'}^{t'}(dy') \\ &= \sum_{|\alpha|, |\alpha'| \leq 1} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} (-1)^{|\alpha|} (-1)^{|\alpha'|} \partial_1^\alpha \partial_2^{\alpha'} k_u(x-y, x'-y') \mu_\alpha^t(dy) \mu_{\alpha'}^{t'}(dy') \\ &= \left[\left(\sum_{|\alpha| \leq 1} \partial^\alpha \mu_\alpha^t \otimes \sum_{|\alpha'| \leq 1} \partial^{\alpha'} \mu_{\alpha'}^{t'} \right) * k_u \right](x, x') = [(\dot{F}_t \otimes \dot{F}_{t'}) * k_u](x, x') \end{aligned}$$

which proves (46).

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

$$\begin{aligned}
k_W(z, z') &= \text{Cov}(W(z), W(z')) \\
&= \mathbb{E}[(W(z)W(z'))] = \mathbb{E}[(U(z) + V(z))(U(z') + V(z'))] \\
&= \mathbb{E}[U(z)U(z')] + \mathbb{E}[U(z)V(z')] + \mathbb{E}[V(z)U(z')] + \mathbb{E}[V(z)V(z')] \\
&= [(\dot{F}_t \otimes \dot{F}_{t'}) * k_u](x, x') + [(F_t \otimes F_{t'}) * k_v](x, x')
\end{aligned} \tag{51}$$

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 $z \mapsto k_W(z, z')$ is of the form (38) and thus verifies $\square k_{x'} = 0$ in the sense of distributions. (ii) is then a direct consequence of Proposition 3.4. \square

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 covariance between U and V : $k_{uv}(x, x') = \text{Cov}(U(x), V(x'))$ and $k_{vu}(x, x') = \text{Cov}(V(x), U(x')) = k_{uv}(x', x)$.

More explicitly, we have the following Kirschhoff-like integral formulas for k_v^{wave} and k_u^{wave} :

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

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

Above, $\nabla_1 k_u(x, x')$ is the gradient vector of k_u w.r.t. x , $\nabla_2 k_u(x, x')$ is the gradient vector of k_u w.r.t. x' and $\nabla_1 \nabla_2 k_u(x, x')$ is the matrix whose entry (i, j) is given by

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

($\partial_{x_i^1}$ is the derivative w.r.t. the i^{th} coordinate of x , $\partial_{x_j^2}$ is the derivative w.r.t. the j^{th} coordinate of x').

4.2.1 Extending the kernels k_u^{wave} and k_v^{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 (A₁) and (A₂). We provide below examples where these formulas yield valid covariance kernels (in particular, functions defined for *all* values of (x, t) and (x', t') ; see the end of Section 2.1.3) 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 kernel of V^0 accordingly, e.g. choosing the following function for k_v (see Section 2.1.3 for $k_{1/2}$)

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

Above, $\|x\|$ denotes the Euclidean norm of x . Such a kernel indeed verifies $k_v(x, x) = \text{Var}(V^0)(x) = 0$ 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_v^{wave} (or any other measurable centered second order random field with this covariance kernel) remains a solution of the wave equation in the distributional sense. One can perform the same kind of discussions on k_u^{wave} : for example, equation (53) shows that when $k_u \in C^{1,1}(\mathbb{R}^3 \times \mathbb{R}^3) \setminus C^{2,2}(\mathbb{R}^3 \times \mathbb{R}^3)$, k_u^{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 kernel k_u^{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 commonly used $k_{3/2}$ Matérn covariance kernel.

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 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 are required. Motivated by acoustical oceanography considerations, we described in Section 4 a Gaussian process model of the wave equation. This PDE served as an application case for Proposition 3.4, 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.

Proposition 3.4 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 and/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 ([13], Section 6.1.2). The natural spaces arising from these formulations are Sobolev spaces rather than $\mathcal{D}'(\mathcal{D})$. An attached question, as first studied in [44], 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 hypotheses required in [44]. Finally, the matter of using random fields for modelling and approximating solutions of nonlinear PDEs is a natural direction for future research.

Acknowledgements. We would like to thank the Service Hydrographique et Océanographique de la Marine (SHOM) for funding this work, and Rémy Baraille in particular for his personal

involvement in the project. We thank the GMM laboratory of INSA Toulouse for providing us access to their servers on which we ran our numerical computations.

References

- [1] R. J. Adler and J. E. Taylor. *Random Fields and Geometry*. Springer-Verlag New York, NY, 2007.
- [2] C. G. Albert and K. Rath. Gaussian process regression for data fulfilling linear differential equations with localized sources. *Entropy*, 22(2), 2020.
- [3] M. Álvarez, D. Luengo, and N. Lawrence. Linear latent force models using Gaussian processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35:2693–2705, 2013.
- [4] J.-M. Azaïs and M. Wschebor. *Level sets and extrema of random processes and fields*. Wiley & Sons, 2009.
- [5] A. Berlinet and C. Thomas-Agnan. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Springer US, 2004.
- [6] H. Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Universitext. Springer New York, 2010.
- [7] Y. Chen, B. Hosseini, H. Owhadi, and A. M. Stuart. Solving and learning nonlinear PDEs with Gaussian processes. *Journal of Computational Physics*, 447:110668, 2021.
- [8] G. Da Prato and J. Zabczyk. *Stochastic equations in infinite dimensions*. Cambridge university press, 2014.
- [9] J. Dieudonné. Sur les espaces de Montel métrisables. *C. R. Acad. Sci., Paris*, 238:194–195, 1954.
- [10] J. L. Doob. Stochastic processes depending on a continuous parameter. *Transactions of the American Mathematical Society*, 42(1):107–140, 1937.
- [11] J. L. Doob. *Stochastic processes*. Wiley Classics Library. Wiley-Interscience, 1990.
- [12] D. G. Duffy. *Green’s functions with applications*. Chapman and Hall/CRC, second edition edition, 2015.
- [13] L. Evans. *Partial Differential Equations*. Graduate studies in mathematics. American Mathematical Society, 1998.
- [14] L. C. Evans and R. F. Garzepy. *Measure theory and fine properties of functions, Revised Edition (1st ed.)*. Chapman and Hall/CRC, 2015.
- [15] M. Fan, D. Paul, T. C. Lee, and T. Matsuo. Modeling tangential vector fields on a sphere. *Journal of the American Statistical Association*, 113(524):1625–1636, 2018.
- [16] G. Fasshauer. Meshfree approximation methods with MATLAB. In *Interdisciplinary Mathematical Sciences*, 2007.
- [17] J. Fiedler. *Distances, Gegenbauer expansions, curls, and dimples: On dependence measures for random fields*. PhD thesis, 2016.

- [18] E. J. Fuselier Jr. *Refined error estimates for matrix-valued radial basis functions*. PhD thesis, Texas A&M University, 2007.
- [19] J. Gapaillard and J. Michaux. Sur les processus linéaires définis sur un espace nucléaire. In *Annales de la Faculté des sciences de Toulouse: Mathématiques*, volume 8, pages 75–92.
- [20] A. Geist and S. Trimpe. Learning constrained dynamics with gauss’ principle adhering gaussian processes. In A. M. Bayen, A. Jadbabaie, G. Pappas, P. A. Parrilo, B. Recht, C. Tomlin, and M. Zeilinger, editors, *Proceedings of the 2nd Conference on Learning for Dynamics and Control*, volume 120 of *Proceedings of Machine Learning Research*, pages 225–234. PMLR, 10–11 Jun 2020.
- [21] I. M. Gel’fand and G. E. Shilov. *Generalized functions, Vol. 4: applications of harmonic analysis*. Academic Press, 1964.
- [22] I. M. Gel’fand and G. Shilov. *Generalized Functions, Volume 2: Spaces of Fundamental and Generalized Functions. Vol. 261*. Academic Press, New York, NY. MR0230128, 1968.
- [23] D. Ginsbourger, O. Roustant, and N. Durrande. On degeneracy and invariances of random fields paths with applications in Gaussian process modelling. *Journal of Statistical Planning and Inference*, page 170 :117 – 128, 2016.
- [24] T. Graepel. Solving Noisy Linear Operator Equations by Gaussian Processes: Application to Ordinary and Partial Differential Equations. In *Proceedings of the 20th International Conference on Machine Learning*, pages 234–241. AAAI Press, 2003.
- [25] M. Gulian, A. Frankel, and L. Swiler. Gaussian process regression constrained by boundary value problems. *Computer Methods in Applied Mechanics and Engineering*, 388:114117, 2022.
- [26] L. Hörmander. *The analysis of linear partial differential operators I: Distribution theory and Fourier analysis*. Springer, 2015.
- [27] S. Janson. *Gaussian Hilbert Spaces*. Cambridge Tracts in Mathematics. Cambridge University Press, 1997.
- [28] C. Jidling, J. Hendriks, N. Wahlstrom, A. Gregg, T. Schon, C. Wensrich, and A. Wills. Probabilistic modelling and reconstruction of strain. *Nuclear Instruments & Methods in Physics Research Section B-beam Interactions With Materials and Atoms*, 436:141–155, 2018.
- [29] C. Jidling, N. Wahlström, A. Wills, and T. B. Schön. Linearly constrained Gaussian processes. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.
- [30] M. Lange-Hegermann. Algorithmic linearly constrained Gaussian processes. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.
- [31] M. Lange-Hegermann. Linearly constrained Gaussian processes with boundary conditions. In A. Banerjee and K. Fukumizu, editors, *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pages 1090–1098. PMLR, 13–15 Apr 2021.

- [32] J.-F. Le Gall. *Mouvement brownien, martingales et calcul stochastique*. Springer, 2013.
- [33] H. Medwin, C. S. Clay, and T. K. Stanton. *Fundamentals of acoustical oceanography*, 1999.
- [34] 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.
- [35] F. J. Narcowich and J. Ward. Generalized Hermite interpolation via matrix-valued conditionally positive definite functions. *Mathematics of Computation*, 63:661–687, 1994.
- [36] N. Nguyen and J. Peraire. Gaussian functional regression for linear partial differential equations. *Computer Methods in Applied Mechanics and Engineering*, 287:69–89, 2015.
- [37] H. Owhadi. Bayesian numerical homogenization. *Multiscale Modeling & Simulation*, 13(3):812–828, 2015.
- [38] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Machine learning of linear differential equations using Gaussian processes. *Journal of Computational Physics*, 348:683–693, 2017.
- [39] C. E. Rasmussen and C. Williams. *Gaussian Processes for Machine Learning*. the MIT Press, 2006.
- [40] K. Ritter. *Average-Case Analysis of Numerical Problems*. Lecture Notes in Mathematics. Springer Berlin Heidelberg, 2007.
- [41] W. Rudin. *Functional Analysis*. McGraw-Hill, New York, 1991.
- [42] R. Schaback. Solving the Laplace equation by meshless collocation using harmonic kernels. *Advances in Computational Mathematics*, 31:457–470, 2009.
- [43] H. Schaefer. *Topological Vector Spaces*. Graduate Texts in Mathematics. Springer-Verlag New York, 1999.
- [44] M. Scheuerer. Regularity of the sample paths of a general second order random field. *Stochastic Processes and their Applications*, 120(10):1879–1897, 2010.
- [45] M. Scheuerer and M. Schlather. Covariance models for divergence-free and curl-free random vector fields. *Stochastic Models*, 28:433 – 451, 2012.
- [46] D. Serre. *Systems of conservation laws*. Cambridge University Press, 1999.
- [47] S.Lang. *Real and Functional Analysis*. Graduate Texts in Mathematics. Springer-Verlag New York, 1993.
- [48] F. Trèves. *Topological Vector Spaces, Distributions and Kernels*. Dover books on mathematics. Dover Publications, 2006.
- [49] R. C. Vergara, D. Allard, and N. Desassis. A general framework for SPDE-based stationary random fields. *Bernoulli*, 28(1):1–32, 2022.
- [50] G. Vidossich. Characterization of separability for LF-spaces. *Annales de l’institut Fourier*, tome 18(2):p. 87–90, 1968.
- [51] N. Wahlstrom, M. Kok, T. B. Schön, and F. Gustafsson. Modeling magnetic fields using Gaussian processes. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3522–3526, 2013.