

On degeneracy and invariances of random fields paths with applications in Gaussian process modelling

David Ginsbourger^{a,b,*}, Olivier Roustant^c, Nicolas Durrande^c

^a*Idiap Research Institute, Martigny, Switzerland*

^b*IMSV, Department of Mathematics and Statistics, University of Bern, Switzerland*

^c*Mines Saint-Etienne, UMR CNRS 6158, LIMOS, F-42023 Saint-Etienne, France*

Abstract

We study pathwise invariances and degeneracies of random fields with motivating applications in Gaussian process modelling. The key idea is that a number of structural properties one may wish to impose a priori on functions boil down to degeneracy properties under well-chosen linear operators. We first show in a second order set-up that almost sure degeneracy of random field paths under some class of linear operators defined in terms of signed measures can be controlled through the two first moments. A special focus is then put on the Gaussian case, where these results are revisited and extended to further linear operators thanks to state-of-the-art representations. Several degeneracy properties are tackled, including random fields with symmetric paths, centred paths, harmonic paths, or sparse paths. The proposed approach delivers a number of promising results and perspectives in Gaussian process modelling. In a first numerical experiment, it is shown that dedicated kernels can be used to infer an axis of symmetry. Our second numerical experiment deals with conditional simulations of a solution to the heat equation, and it is found that adapted kernels notably enable improved predictions of non-linear functionals of the field such as its maximum.

Keywords: Covariance kernels, Linear operators, RKHS, Structural priors.
2000 MSC: 60G60, 60G17, 62J02.

*Corresponding author

Email addresses: ginsbourger@stat.unibe.ch (David Ginsbourger),
roustant@emse.fr (Olivier Roustant), durrande@emse.fr (Nicolas Durrande)

1. Introduction

Whether for function approximation, classification, or density estimation, probabilistic models relying on random fields have been increasingly used in recent works from various research communities. Finding their applied roots in geostatistics and spatial statistics with optimal linear prediction and Kriging [1, 2], random field models for prediction have become a main stream topic in machine learning (under the *Gaussian Process Regression* terminology, see, e.g., [3]), with a spectrum ranging from metamodeling and adaptive design approaches in science and engineering [4, 5, 6]) to theoretical Bayesian statistics in function spaces (see [7, 8, 9] and references therein).

Often, a Gaussian random field model is assumed for some function f of interest, and so all prior assumptions on f are accounted for by the corresponding mean function m and covariance kernel k . The choice of m and k should thus reflect as much as possible any prior belief the modeller wishes to incorporate in the model. Such prior belief on f may of course include classical regularity properties in the first place (continuity, differentiability, Hölder regularity, etc.), but also more specific properties such as symmetries [10, 11], sparse functional ANOVA decompositions [12, 13, 14], or degeneracy under multivariate differential operators in the case of vector-valued random fields. To take a concrete example, covariance structures characterizing divergence-free and curl-free random vector fields have been recently presented and illustrated in [15]. Besides that, the idea of expressing structure with kernels has been explored in [16], where a number of practical aspects regarding positive-(semidefiniteness-preserving) operations are addressed.

Here we shall discuss how the two first moments influence mathematical properties of associated realisations (or *paths*), both in a general second order set-up and in the Gaussian case. A number of well-known random field properties driven by the covariance kernel are in the mean square sense, e.g. L^2 continuity and differentiability [17]. However, such results generally are not informative about the pathwise behaviour of underlying random fields. On the other hand, much can be said about path regularity properties of random field paths (see, e.g., classical results in [18], [19]), based in particular on the behaviour of the covariance kernel in the neighbourhood of the diagonal in the second order case. In the stationary case, it is then sufficient to look at the covariance function in the neighborhood of the origin (with similar results for the variogram in the intrinsic stationary –but not necessarily second order– case). More recently, [20] has taken a new look at path

regularity of second-order random fields, and drew conclusions about a.s. continuous differentiability in non-Gaussian settings. Also, we refer to [21] for an enlightening exposition of state-of-the-art results concerning regularity properties of random field sample paths in various frameworks.

Our focus in the present work is on *pathwise* mathematical properties of second order random fields and statistical applications thereof in the context of Gaussian process modelling. Motivated by several practical situations, we pay a particular attention to random fields $Z = (Z_{\mathbf{x}})_{\mathbf{x} \in D}$ that are supported by the null space of some linear operator T , i.e. for which

$$T(Z) = \mathbf{0} \text{ (a.s.)}. \quad (1)$$

As we first develop in general second-order settings, an impressive diversity of path properties including invariances under group actions or sparse ANOVA decompositions of multivariate paths can be encapsulated in the framework of Eq. 1. Furthermore, in the particular case of Gaussian random fields, a more general class of path properties (notably some degeneracy properties involving differential operators) can be covered through the link between operators on the paths and operators on the *reproducing kernel Hilbert space* [22] associated with the random field, and also through an additional representation of Z in terms of Gaussian measures on Banach spaces.

While Section 2 is dedicated to the exposition of the main results, proofs are presented in appendix to ease the reading. Applications in the context of random field modelling, and especially for Gaussian process modelling, are then investigated throughout Section 3. In particular, we tackle zero-integral random processes, random fields with paths invariant under group actions, random fields with additive paths, random fields with harmonic paths, and discuss further potential applications.

In Section 4, we present two original numerical experiments where the notions of degeneracy and invariance appear very useful in Gaussian process modelling under two types of structural prior information. In the first case, the objective function possesses an unknown axis of symmetry, which is inferred by maximum likelihood, relying on a family of argumentwise invariant covariance kernels. In the second case, we obtain an improved interpolation of a solution to the heat equation thanks to a bi-harmonic kernel. The proposed model enables performing harmonic conditional simulations, which has very beneficial consequences in terms of estimation of the maximum. Section 5 is dedicated to conclusions and perspectives. The main results are finally proven in Appendix A.

2. Main results

Let (D, \mathcal{D}) be a measurable space, $(\Omega, \mathcal{A}, \mathbb{P})$ be a complete probability space, and $Z = (Z_{\mathbf{x}})_{\mathbf{x} \in D}$ be a measurable real-valued stochastic process over $(\Omega, \mathcal{A}, \mathbb{P})$. Let us further assume that the paths of Z belong with probability 1 to some function space $\mathcal{F} \subset \mathcal{M}(D, \mathbb{R})$, where $\mathcal{M}(D, \mathbb{R})$ is the set of $(\mathcal{D}, \mathcal{B}(\mathbb{R}))$ -measurable functions, and consider a linear operator $T : \mathcal{F} \rightarrow \mathcal{F}$. Here both Z and $T(Z)$ are assumed second order, in the sense that their marginals possess a variance, and we aim at giving necessary and sufficient conditions in terms of the two first moments of Z for the following *degeneracy* to hold:

$$\mathbb{P}(T(Z) = \mathbf{0}) = \mathbb{P}(\forall \mathbf{x} \in D \ T(Z)_{\mathbf{x}} = 0) = 1. \quad (2)$$

We prove that in a variety of settings on T and Z , this is equivalent to having that both m and k are in the null space of T in a sense to be discussed next. In Section 2.1 we discuss equivalent conditions that do not involve any distributional assumption, and we obtain a characterization of degeneracy under a specific class of operators that prove useful for applications in Sections 3 and 4. In Section 2.2 we generalize the results to a wider class of operators T in the specific framework of Gaussian processes and Gaussian measures.

2.1. Results in the second order framework

Proposition 1. *Let $T : \mathcal{F} \rightarrow \mathcal{F}$ be a linear operator such that for all $\mathbf{x} \in D$ there exists a signed measure $\nu_{\mathbf{x}} : \mathcal{D} \rightarrow \mathbb{R}$ satisfying*

$$T(f)(\mathbf{x}) = \int f(\mathbf{u}) d\nu_{\mathbf{x}}(\mathbf{u}), \quad (3)$$

and assume that

$$\sup_{\mathbf{x} \in D} \int_D \sqrt{k(\mathbf{u}, \mathbf{u}) + m(\mathbf{u})^2} d|\nu_{\mathbf{x}}|(\mathbf{u}) < +\infty. \quad (4)$$

Then the following are equivalent:

- a)** $\forall \mathbf{x} \in D \ \mathbb{P}(T(Z)_{\mathbf{x}} = 0) = 1$ (“ $T(Z) = \mathbf{0}$ up to a modification”)
- b)** $\forall \mathbf{x} \in D \ T(m)(\mathbf{x}) = 0$ and $(T \otimes T(k))(\mathbf{x}, \mathbf{x}) = 0$.

*Assuming further that $T(Z)$ is separable, **a)** and **b)** are also equivalent to*

- c)** $\mathbb{P}(T(Z) = \mathbf{0}) = \mathbb{P}(\forall \mathbf{x} \in D \ T(Z)_{\mathbf{x}} = 0) = 1$ (“ $T(Z) = \mathbf{0}$ almost surely”).

Proof. See Appendix A. □

Remark 1. A special class of T 's fulfilling the requirements of Proposition 1 is given by the “composition of composition operators” discussed in [23]:

$$T(f) = \sum_{i=1}^q \alpha_i f \circ v_i, \quad (5)$$

where the functions v_i are assumed $(\mathcal{D}, \mathcal{D})$ -measurable ($1 \leq i \leq q$), and the α_i coefficients are scalars, that can be generalized to $(\mathcal{D}, \mathcal{B}(\mathbb{R}))$ -measurable functions. As we will see in Section 3 and 4, this simple class of operators is actually sufficient to cover a number of meaningful applications.

Remark 2. In Proposition 1, the part of condition **b)** concerning k seen as a function of two arguments may be reformulated in terms of the $\{k(\cdot, \mathbf{u}), \mathbf{u} \in D\}$ functions. In particular, in cases where Mercer's theorem [24, 25] applies for some measure η on D , k may be expanded as

$$k = \sum_{i=1}^{+\infty} \lambda_i \phi_i \otimes \phi_i : (\mathbf{x}, \mathbf{x}') \in D \times D \longrightarrow \sum_{i=1}^{+\infty} \lambda_i \phi_i(\mathbf{x}) \phi_i(\mathbf{x}') \quad (6)$$

where the convergence is absolute and uniform on compact subsets of D , and $(\lambda_i, \phi_i(\cdot)) \in [0, +\infty) \times L^2(\eta)$ are pairs of eigenvalues and eigenfunctions associated with the Fredholm operator on $L^2(\eta)$ with kernel k . Equation 4 then guarantees that for any $\mathbf{x} \in D$

$$\int \sum_{i=1}^{+\infty} \lambda_i |\phi_i(\mathbf{u})| |\phi_i(\mathbf{v})| d(\nu_{\mathbf{x}} \otimes \nu_{\mathbf{x}})(\mathbf{u}, \mathbf{v}) < +\infty, \quad (7)$$

which is sufficient for $\ell_{\nu_{\mathbf{x}}} \otimes \ell_{\nu_{\mathbf{x}}} k = \sum_{i=1}^{+\infty} \lambda_i \ell_{\nu_{\mathbf{x}}}(\phi_i)^2$ to hold. The kernel part of **b)** is then equivalent to the condition ($\ell_{\nu_{\mathbf{x}}}(\phi_i) = 0$ for all $i : \lambda_i \neq 0$) and hence to $(\forall \mathbf{x}' \in D \ T(k(\cdot, \mathbf{x}')) = \mathbf{0})$ (“argumentwise degeneracy” of k).

2.2. The Gaussian framework

When Z is assumed Gaussian, Proposition 1 still applies since Gaussian processes are systematically second order, but further results can be derived regarding degeneracy with respect to more general operator classes.

The results enounced below are based on three closely related but different approaches, each calling for specific assumptions: the Karhunen-Loève expansion, the Loève isometry, and the Gaussian measure approach.

Continuing first with the set-up of Remark 2 above, let us consider the settings of the Karhunen-Loève expansion, in which Z decomposes as

$$Z_{\mathbf{x}} = m(\mathbf{x}) + \sum_{i=1}^{+\infty} \sqrt{\lambda_i} \zeta_i \phi_i(\mathbf{x}), \quad (8)$$

where $\zeta_i \sim \mathcal{N}(0, 1)$ independently and the pairs (λ_i, ϕ_i) satisfy

$$\lambda_i \phi_i(\cdot) = \int k(\cdot, \mathbf{u}) \phi_i(\mathbf{u}) d\eta(\mathbf{u}) \quad (i \geq 1) \quad (9)$$

Such expansion is guaranteed to exist notably in the case where D is compact and k is continuous on $D \times D$ (See [26] but also [27] for recent generalizations).

Proposition 2. *Let η be a measure on \mathcal{D} , Z be a Gaussian random field indexed by D possessing a Karhunen-Loève expansion with respect to η (as defined by Equations 8 and 9), and T be a bounded operator of $L^2(\eta)$. Let furthermore $m \in L^2(\eta)$. Then, the following are equivalent:*

- a) $T(m) = \mathbf{0}$ (η -a.e.) and $\forall i : \lambda_i \neq 0, T(\phi_i) = \mathbf{0}$ (η -a.e.)
- b) $\mathbb{P}(T(Z) = \mathbf{0} \mid \eta - a.e.) = 1$

Proof. See Appendix A.2. □

Another way of looking at things, involving yet a different kind of assumption on T , is to appeal to the Loève isometry such as presented in [22]. Let us review it briefly before stating our associated result.

Assuming that Z is a centred Gaussian random field with covariance kernel k , a crucial state-of-the-art result is that the reproducing kernel Hilbert space \mathcal{H} associated with k is isometric to the Hilbert space generated by Z ,

$$\mathcal{L}(Z) = \overline{\text{span}(Z_{\mathbf{x}}, \mathbf{x} \in D)},$$

where the closure is taken with respect to the usual $L^2(\mathbb{P})$ norm on equivalence classes of square-integrable random variables. More precisely, the map $\Psi : \mathcal{H} \rightarrow \mathcal{L}(Z)$ defined by $\Psi(k(\cdot, \mathbf{x})) = Z_{\mathbf{x}}$ for all $\mathbf{x} \in D$ and extended by linearity and continuity, is a canonical isometry between $\mathcal{L}(Z)$ and \mathcal{H} , often called the *Loève Isometry* [22]. In Proposition 3 below, instead of demanding from T to be bounded on some prescribed normed vector space as in Proposition 2, we rather focus again on a class of T solely fulfilling a marginal property, namely that for any $\mathbf{x} \in D$, $T(Z)_{\mathbf{x}} \in \mathcal{L}(Z)$.

Proposition 3. *Let $T : \mathcal{F} \rightarrow \mathbb{R}^D$ be an operator such that $T(Z)_\mathbf{x} \in \mathcal{L}(Z)$ for any $\mathbf{x} \in D$. Then, there exists a unique operator $\mathcal{T} : \mathcal{H} \rightarrow \mathbb{R}^D$ satisfying*

$$\text{cov}(T(Z)_\mathbf{x}, Z_{\mathbf{x}'}) = \mathcal{T}(k(\cdot, \mathbf{x}'))(\mathbf{x}) \quad (\mathbf{x}, \mathbf{x}' \in D) \quad (10)$$

and such that $\mathcal{T}(h_n)(\mathbf{x}) \rightarrow \mathcal{T}(h)(\mathbf{x})$ for all $\mathbf{x} \in D$ and $h_n \xrightarrow{\mathcal{H}} h$. Furthermore, the following conditions are equivalent:

$$(i) \quad \forall \mathbf{x} \in D \quad T(Z)_\mathbf{x} = \mathbb{E}(T(Z)_\mathbf{x}) = 0 \quad (a.s.)$$

$$(iii) \quad \forall \mathbf{x}' \in D \quad \mathcal{T}(k(\cdot, \mathbf{x}')) = \mathbf{0}$$

$$(iii) \quad \mathcal{T}(\mathcal{H}) = \{0\}$$

Proof. See Appendix A.2. □

Remark 3. *Proposition 3 holds with a non-zero mean such that $T(m) = 0$.*

Finally, let us consider the set-up of Gaussian measures on function spaces. Given a real separable Banach space $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$ and its topological dual \mathcal{F}^* , a probability measure μ on $\mathcal{B}(\mathcal{F})$ is called Gaussian if for any continuous linear form $x^* \in \mathcal{F}^*$, the image measure $x^*_\# \mu$ (or *pushforward of μ by x^** , characterized by $x^*_\# \mu(B) = \mu(x^* \in B)$ for $B \in \mathcal{B}(\mathbb{R})$) is a Gaussian measure on $\mathcal{B}(\mathbb{R})$. The link between Gaussian processes and Gaussian measures has been established for long [28], and it has been notably shown for several standard classical function spaces that Gaussian processes whose paths belong to these spaces induce Gaussian measures supported by them, and conversely.

A key device for working with such Gaussian measures μ is their characteristic functional (see, e.g., [29]) $\hat{\mu} : \mathcal{F}^* \rightarrow \mathbb{C}$, defined for $x^* \in \mathcal{F}^*$ by

$$\hat{\mu}(x^*) = \int_{\mathcal{F}} \exp(ix^*(x)) d\mu(x). \quad (11)$$

In the case where μ is a Gaussian measure on \mathcal{F} , it is known that

$$\hat{\mu}(x^*) = \exp \left(i \langle m_\mu, x^* \rangle - \frac{1}{2} \langle C_\mu x^*, x^* \rangle \right), \quad (12)$$

where $m_\mu \in \mathcal{F}$ is the mean of μ , $C_\mu : \mathcal{F}^* \rightarrow \mathcal{F}$ is its covariance operator, and $\langle x, x^* \rangle := x^*(x)$ ($x \in \mathcal{F}, x^* \in \mathcal{F}^*$) denotes the duality bracket.

A useful fact for our purposes is that, for any bounded operator $T : \mathcal{F} \rightarrow \mathcal{F}$, the image measure $T_\# \mu$ is Gaussian with mean $T(m_\mu)$ and covariance operator $TC_\mu T^*$, where $T^* : \mathcal{F}^* \rightarrow \mathcal{F}^*$ is the adjoint of T . Thanks to this powerful tools, we obtain the following degeneracy result without difficulty.

Proposition 4. *Let μ be a Gaussian measure on \mathcal{F} with mean $m_\mu \in \mathcal{F}$ and covariance operator $C_\mu : \mathcal{F}^* \rightarrow \mathcal{F}$, and $T : \mathcal{F} \rightarrow \mathcal{F}$ be a bounded operator. Then, the following are equivalent:*

- i) $\mu(T = 0_{\mathcal{F}}) = 1$
- ii) $T(m_\mu) = 0_{\mathcal{F}}$ and $TC_\mu T^* = 0_{\mathcal{F}^* \rightarrow \mathcal{F}}$

We now close the section with a result concerning the propagation of degeneracy properties to conditional distributions, that both proves relevant in applications and illustrates the power of the Gaussian measure approach. The following proposition is an adaptation of [29, Theorem 3.11]:

Proposition 5. *Let \mathcal{F}, \mathcal{G} be real separable Banach spaces, μ be a Gaussian measure on $\mathcal{B}(\mathcal{F})$ with mean zero and covariance operator C_μ . Let $T : \mathcal{F} \rightarrow \mathcal{F}$ be a bounded linear operator such that $TC_\mu T^* = 0_{\mathcal{F}^* \rightarrow \mathcal{F}}$. Let $A : \mathcal{F} \rightarrow \mathcal{G}$ be another bounded linear operator and $A_\# \mu$ be the image of μ under A . Then there exist a Borel measurable mapping $m : \mathcal{G} \rightarrow \mathcal{F}$, a Gaussian covariance $R : \mathcal{F}^* \rightarrow \mathcal{F}$ with $R \leq C_\mu$ and a disintegration $(q_y)_{y \in \mathcal{G}}$ of μ on $\mathcal{B}(\mathcal{F})$ with respect to A such that for any fixed $y \in \mathcal{G}$, q_y is a Gaussian measure with mean m and covariance operator R satisfying $T(m) = 0_{\mathcal{F}}$ and $TRT^* = 0_{\mathcal{F}^* \rightarrow \mathcal{F}}$.*

Remark 4. *The results still hold with a non-centred measure μ , provided that $T(m_\mu) = 0_{\mathcal{F}}$. While C_μ is unchanged, the conditional mean then becomes $m_n(y) = m_\mu + \sum_{i=1}^n \langle y - A(m_\mu), y_i^* \rangle C_\mu x_i^*$ and similar arguments apply.*

3. Applications in random field modelling and related

3.1. On processes with paths integrating to zero

Lemma 1 of Appendix A ensures that the almost sure nullity of $\ell_\nu(Z)$ can be characterized through the nullity of $\ell_\nu(m)$ and $\ell_\nu \otimes \ell_\nu(k)$, provided that the function $\mathbf{u} \in D \rightarrow \sqrt{k(\mathbf{u}, \mathbf{u}) + m(\mathbf{u})^2} \in \mathbb{R}$ is ν -integrable. In practice, assuming that this integrability condition is fulfilled —as it is notably the case for compact D and continuous m and k — it suffices to check that $\ell_\nu(m) = 0$ and $\ell_\nu(k(\cdot, \mathbf{x}')) = 0$ for arbitrary $\mathbf{x}' \in D$. Taking for instance the settings of [30] where ν is a finite (positive) measure and a kernel k_0 of the form

$$(\mathbf{x}, \mathbf{x}') \rightarrow k_0(\mathbf{x}, \mathbf{x}') = k(\mathbf{x}, \mathbf{x}') - \frac{\ell_\nu(k(\mathbf{x}, \cdot))\ell_\nu(k(\cdot, \mathbf{x}'))}{\ell_\nu \otimes \ell_\nu(k)}$$

is considered, we can directly check that $\ell_\nu(k_0(\cdot, \mathbf{x}')) = 0$ for arbitrary $\mathbf{x}' \in D$. Hence, any squared integrable centred random field possessing k as covariance kernel is equivalent to a random field with ν -centred paths, be it Gaussian or not. To give a concrete application in the Gaussian framework, Karhunen-Loève expansions of mean-centered Wiener processes were derived in [31]:

$$W_t - \int_0^1 W_u du = \sum_{j=1}^{+\infty} \zeta_j \frac{\sqrt{2} \cos(j\pi t)}{j\pi}, \quad (13)$$

where the unnormalized eigenfunctions $t \rightarrow \cos(j\pi t)$ integrate to zero, which is sufficient for this process to have zero-mean paths by Proposition 2. Alternatively, one can also use the closed form expression $k(s, t) = \min(s, t) - s - t + \frac{1}{2}s^2 + \frac{1}{2}t^2 + \frac{1}{3}$ to check that $\ell_\nu(k(\cdot, t)) = 0$ for arbitrary $t \in [0, 1]$, guaranteeing the zero-mean property following Proposition 1.

3.2. On combination of composition operators and more

Combinations of composition operators T such as recalled in Equation 5 fit into the settings of Proposition 1 since they can be written as

$$T : f \in \mathcal{F} \rightarrow T(f) = \int_D f(\mathbf{u}) d\nu_{\mathbf{x}}(\mathbf{u}), \quad (14)$$

where $\nu_{\mathbf{x}} = \sum_{i=1}^q \alpha_i \delta_{v_i(\mathbf{x})}$. Of course, α_i and v_i should be chosen such that the functions $\mathbf{x} \rightarrow \sum_{i=1}^q \alpha_i f(v_i(\mathbf{x}))$ be in \mathcal{F} . Note that such class may be extended without difficulty (under similar precautions concerning the image space) to functional v_i 's, and also to infinite sums. However, the basic set-up with $q < \infty$ and scalar coefficients suffices to cover several interesting cases.

First, let us consider a finite group G acting measurably on the domain D through some group action $\Phi : (g, \mathbf{x}) \in G \times D \rightarrow D$. A function $f : D \rightarrow \mathbb{R}$ is said *invariant under Φ* if f is constant over the orbits of Φ , i.e.

$$\forall \mathbf{x} \in D \ \forall g \in G \ f(\mathbf{x}) = f(\Phi(g, \mathbf{x})) = f(g.\mathbf{x}), \quad (15)$$

where the notation $g.\mathbf{x} = \Phi(g, \mathbf{x})$ is used for simplicity when there is no reason to confuse several actions of G on D . Such invariance functions can be characterized by the fact that they coincide at each point with the arithmetic mean of their values on the corresponding orbit, or in other words that

$$\forall \mathbf{x} \in D \ f(\mathbf{x}) = \frac{1}{q} \sum_{g \in G} f(g.\mathbf{x}), \quad (16)$$

where $q = \text{Card}(G)$ stands for the order of G . Now, Equation 16 may be reformulated in terms of signed measures: $\forall \mathbf{x} \in D \int f(\mathbf{u}) d\nu_{\mathbf{x}}(\mathbf{u}) = 0$ where $\nu_{\mathbf{x}} = \delta_{\mathbf{x}} - \sum_{g \in G} \frac{1}{q} \delta_{g.\mathbf{x}}$. The function $\mathbf{x} \rightarrow \frac{1}{q} \sum_{g \in G} f(g.\mathbf{x})$ being measurable, the invariance property can be cast as a particular case of Proposition 1. We conclude that a random field with mean m and kernel k has paths equivalent to Φ -invariant functions if and only if m and $k(\cdot, \mathbf{x}')$ are Φ -invariant for all $\mathbf{x}' \in D$. Furthermore, for suitable definitions of \mathcal{F} (such as the space of continuous function over a compact set, endowed with the sup norm), the corresponding operator turns out to be bounded, so that Property 5 can be applied, and almost sure pathwise invariance results concerning Z and its conditional distributions can be obtained without further effort. Note that different options are available for constructing *argumentwise invariant* kernels, such as double averaging over the orbits or projecting on fundamental domains. More on these questions and some numerical experiments involving Gaussian random fields with invariant paths can be found in [11] and [16].

Remark 5. *This characterization of pathwise group invariance relying on Proposition 1 can actually be extended to actions of locally compact topological groups, such as considered in [10]. Denoting by ν_G the Haar measure of such group, the invariance condition writes $f(\mathbf{x}) = \int f(g.\mathbf{x}) d\nu_G(g)$ which can be reformulated as $f(\mathbf{x}) = \int f(\mathbf{u}) d\nu_{\mathbf{x}}(\mathbf{u})$ where $\nu_{\mathbf{x}} = \Phi(\cdot, \mathbf{x})_{\#} \nu_G$.*

Beyond group-invariance, Proposition 1 also has implications about the sparsity of multivariate random field paths, e.g., additivity. By additivity, we mean the property for $f : D \subset \mathbb{R}^d \rightarrow \mathbb{R}$ with $D = \prod_{i=1}^d D_i$ for some $D_i \subset \mathbb{R}$ to decompose as a sum of univariate f_i 's: $\forall \mathbf{x} = (x_1, \dots, x_d) \in D$, $f(\mathbf{x}) = f_1(x_1) + \dots + f_d(x_d)$. The additivity assumption lies at the heart of popular statistical methods (See [32] and references therein) and has been recently researched on in the framework of Gaussian Process models [13], where it was shown that kernels of the form $k(\mathbf{x}, \mathbf{x}') = k_1(x_1, x'_1) + \dots + k_d(x_d, x'_d)$ lead to centred Gaussian random field with additive paths. Proposition 1 enables a characterization of second order random fields possessing such property in terms of their mean function and covariance kernel. The key is to notice that being additive can be encoded as being in the null space of an operator such as required. Given $\mathbf{a} \in D$, one can show that f is additive if and only if

$$f(\mathbf{x}) = f(v_1(\mathbf{x})) + \dots + f(v_d(\mathbf{x})) - (d-1)f(v_{d+1}(\mathbf{x})), \quad (17)$$

where $v_i(\mathbf{x}) := (a_1, \dots, a_{i-1}, \overbrace{x_i}^{\text{ith coordinate}}, a_{i+1}, \dots, a_d)$, and $v_{d+1}(\mathbf{x}) := \mathbf{a}$. Re-

formulating Equation 17 in terms of signed measures, we get

$$\int f d \left(\delta_{\mathbf{x}} - \sum_{i=1}^d \delta_{v_i(\mathbf{x})} + (d-1) \delta_{v_{d+1}(\mathbf{x})} \right) = 0, \quad (18)$$

and so Proposition 1 applies with $\nu_{\mathbf{x}} = \delta_{\mathbf{x}} - \sum_{i=1}^d \delta_{v_i(\mathbf{x})} + (d-1) \delta_{v_{d+1}(\mathbf{x})}$, leading to random fields with additive mean and argumentwise additive kernels. Such argumentwise additive kernels actually coincide with kernels of sums of correlated univariate processes.

Example 1. *Let us consider the following kernel over $\mathbb{R}^d \times \mathbb{R}^d$:*

$$k(\mathbf{x}, \mathbf{y}) = \sum_{i,j=1}^d \int_{\mathbb{R}} \kappa_i(x_i - u) \kappa_j(y_j - u) du \quad (19)$$

where the κ_i are smoothing kernels over \mathbb{R} . [33] ensures that k is a valid covariance kernel. Furthermore, k is argumentwise additive. According to Proposition 1, a random field with such kernel has additive paths, with univariate components exhibiting possible cross-correlations.

3.3. On some differential operators

While Proposition 1 focuses on integrals, many operators appearing in functional analysis involve derivatives. In particular, linear partial differential equations can be written as $Tf = g$ for some differential operator T on \mathcal{F} and some $g \in \mathcal{F}$. Here we discuss examples where degeneracies under a differential operator can be incorporated into a Gaussian field model, be it through Proposition 3 or Proposition 5. This includes the case of Gaussian fields with harmonic paths, as illustrated in Section 4, but also Gaussian vector fields with divergence-free and curl-free paths as introduced in [15].

A first crucial point in connection to Proposition 3 is that mean-square (“m.s.”) derivatives of a random process, whenever they exist, do systematically belong to the Hilbert space generated by this process.

As recalled in detail in [15, Section 2], m.s. derivatives are obtained as $L^2(\mathbb{P})$ limits of increments, and hence in $\mathcal{L}(Z)$. Another important fact concerns the relation between m.s. and pathwise derivatives: even though *m.s. differentiability neither implies nor requires that Z has differentiable sample paths* [15], it is also known that for k -th order m.s. differentiable separable

Gaussian fields, a.s. pathwise k -th order differentiability can be guaranteed (see. e.g., [26], and [20] for non-Gaussian extensions) under the condition that the k -th order m.s. derivatives be themselves a.s. pathwise continuous (see. e.g., Theorem 5.3.16 of [21] for a sufficient condition). Denoting by $\frac{\partial}{\partial \mathbf{e}} Z_{\mathbf{x}}$ the pathwise directional derivative of Z in direction $\mathbf{e} \in \mathbb{R}^d$, we obtain $\frac{\partial}{\partial \mathbf{e}} Z_{\mathbf{x}} \in \mathcal{L}(Z)$. In particular for a centred Gaussian Z with $\mathbf{x} \rightarrow k(\mathbf{x}, \mathbf{x}')$ twice continuously differentiable for all $\mathbf{x}' \in D$ and such that Z has twice continuously differentiable paths a.s., we obtain that $\forall \mathbf{x} \in D \Delta Z_{\mathbf{x}} \in \mathcal{L}(Z)$. Applying Proposition 3 with $T = \Delta$ yields that

$$(\forall \mathbf{x} \in D \Delta Z_{\mathbf{x}} = 0 \text{ a.s.}) \iff (\forall \mathbf{x}' \in D \Delta k(\cdot, \mathbf{x}') = 0).$$

Note that one may also appeal to the Gaussian measure approach since the Laplace operator is bounded from $H^2(\eta)$ to $L^2(\eta)$. Lemma 1 and Proposition 5 then directly yield almost sure harmonicity and its propagation to disintegrations. We give in Section 4.2 below a numerical example of GP modelling for the heat equation, and demonstrate how taking a harmonic kernel drastically improves modelling over a Gaussian covariance kernel.

4. Numerical experiments

4.1. Estimation of a symmetry axis

We now consider the following test function over $D = [0, 1]^2$:

$$f : (x_1, x_2) \in D \rightarrow \cos(\sqrt{2}(x_1 + x_2) + 0.4) + \sin(3(x_1 - x_2))^2 + x_1 - x_2. \quad (20)$$

This function is symmetric with respect to the axis defined by the equation $x_2 = -x_1 - 0.1\sqrt{2}$. Given noisy evaluation results of f at 20 points, and assuming a priori that f is symmetric with respect to some axis Δ (parametrized by its angle with the ordinate axis $\lambda \in [0, \pi)$ and its distance to the origin $\delta \geq 0$), we aim at both reconstructing f and recovering the equation of this axis by using a Gaussian random field approach with a dedicated argumentwise invariant covariance kernel. Hereafter, the evaluation points are given by a Latin hypercube design with maximin criterion [34].

We denote by $\pi_{\lambda, \delta}$ the symmetry with respect to Δ and by Φ the action of $G = \mathbb{Z}/2\mathbb{Z}$ on D defined by $\Phi(\bar{1}, \mathbf{x}) = \pi_{\lambda, \delta}(\mathbf{x})$. Symmetric functions over $[0, 1]^2$ with respect to Δ are characterized by $f(\mathbf{x}) = f(\pi_{\lambda, \delta}(\mathbf{x}))$ for all

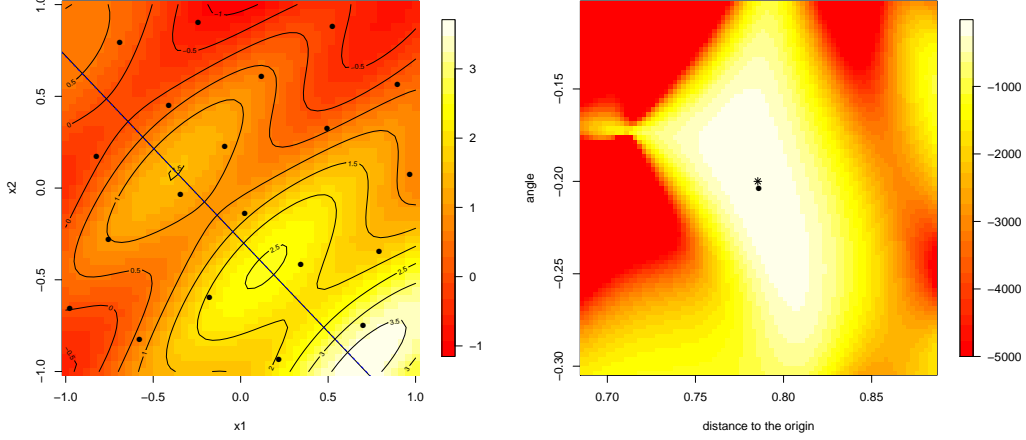


Figure 1: Left. Predicted GP mean with the kernel of Eq. 21 given observations at the solid bullets points. The estimated symmetry axis (solid black line) is indistinguishable from the actual symmetry axis of the test function (dashed blue line). Right. Log-likelihood function in the vicinity of the optimum as a function of the axis parameters. The star and the bullet point respectively stand for the true parameter values and their estimates.

$\mathbf{x} \in D$, i.e. being invariant under Φ or, equivalently, in the null space of $T : f \rightarrow f - f \circ \pi_{\lambda, \delta}$. For any covariance kernel k , it is straightforward that

$$k_{sym} : (\mathbf{x}, \mathbf{x}') \rightarrow k(\mathbf{x}, \mathbf{x}') + k(\mathbf{x}, \pi_{\lambda, \delta}(\mathbf{x}')) + k(\pi_{\lambda, \delta}(\mathbf{x}), \mathbf{x}') + k(\pi_{\lambda, \delta}(\mathbf{x}), \pi_{\lambda, \delta}(\mathbf{x}')) \quad (21)$$

is a valid covariance kernel satisfying $T(k_{sym}(\cdot, \mathbf{x}')) = 0$ for all $\mathbf{x}' \in D$ so that, by Section 3.2, centred random fields with covariance kernel k_{sym} have their paths invariant under Φ almost surely. We will now focus on a case where k is a squared exponential covariance kernel:

$$k(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp \left(-\frac{(x_1 - x'_1)^2}{\theta_1^2} - \frac{(x_2 - x'_2)^2}{\theta_2^2} \right), \quad (22)$$

where σ^2 and the θ_i 's are respectively called variance and length-scale parameters. The kernel k_{sym} thus depends on five parameters ($\sigma^2, \theta_1, \theta_2, \lambda, \delta$) that can be estimated by maximizing their likelihood given the observations. The obtained predicted mean and estimated symmetry axis are shown in the

left panel of Figure 1. Additionally, on the right panel, the log-likelihood is plotted as a function of λ and δ , the other parameters being fixed to their maximum likelihood estimates. The actual symmetry axis of f is here well recovered by maximum likelihood under our proposed GP model with degenerate (argumentwise invariant) covariance kernel. For this and the following example, dedicated R codes were used. Some results were then checked and speeded up with the novel *kergp* R package [35].

4.2. Interpolating a solution to the heat equation and estimating its maximum

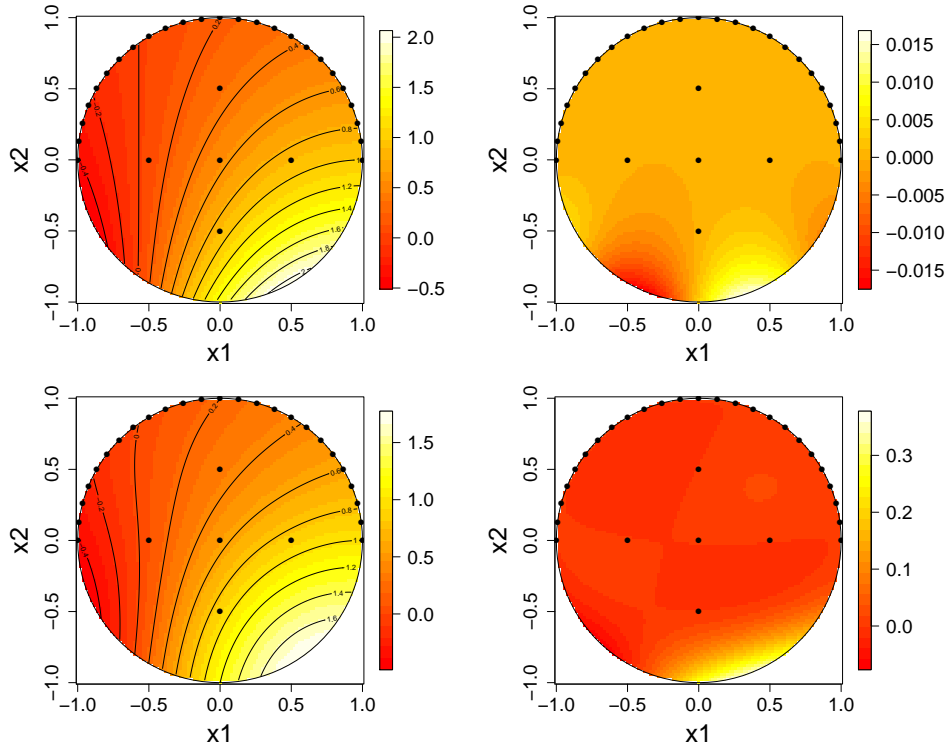


Figure 2: Left: Mean predictor and design points (bullets) for the heat equation interpolation problem of Section 4.2. Right: Prediction errors. Top: with a harmonic kernel. Bottom: with a Gaussian kernel. In both cases, covariance parameters are estimated by maximum likelihood.

We now focus on the heat (or “Laplace”) equation $\Delta f = 0$ over the unit disc $D = \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\| \leq 1\}$. Solutions to this equation are harmonic. As

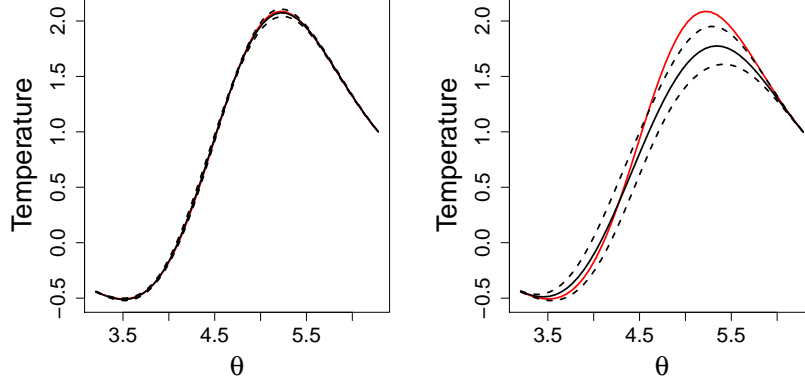


Figure 3: Mean predictor (solid black line) and 95% pointwise prediction intervals (dashed black line) on the lower half boundary (points with coordinates $(\cos(\theta), \sin(\theta))$) for the heat equation interpolation problem of Section 4.2. For the sake of comparison the actual value of f_{test} is shown in solid red. Left: harmonic covariance kernel. Right: Gaussian covariance kernel.

discussed in [36], the symmetric positive definite kernel

$$k(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(\frac{x_1 x'_1 + x_2 x'_2}{\lambda^2}\right) \cos\left(\frac{x_2 x'_1 - x_1 x'_2}{\lambda^2}\right) \quad (23)$$

is harmonic as a function of each of its arguments, the other one being fixed to any arbitrary value: $\forall \mathbf{x}' \in D \nabla k(\cdot, \mathbf{x}') = 0$. Following the discussion held in Section 3, sample paths of centred Gaussian random fields with such a covariance kernel k also satisfy the heat equation with probability 1.

We now consider 30 observations from a harmonic test function $f_{test}(\mathbf{x}) = \cos(1 - x_1) \exp(-x_2)$ and we compare predictions under two GP models based on a) the harmonic kernel of Equation 23 and b) a standard Gaussian kernel. The mean predictors and the prediction errors obtained with the two models are shown in Figure 2: It appears that the prediction errors of the two models differ by one order of magnitude. Besides this, since the test function and the harmonic mean predictor are harmonic, the prediction error of this model is also harmonic, which explains why the extrema of the errors are necessarily located at the domain boundaries. Details of the model predictions on the lower half of the unit circle are highlighted in Figure 3.

Next we investigate how the estimation of a non-linear functional such as the maximum can be improved by injecting structural information into the kernel. Assume to this end that we wish to estimate the maximum of f_{test} relying on our two candidate GP models. For this we appeal to a Monte Carlo approach by using conditional simulations of the two GP models given the 30 observations. For each GP model, 1000 conditional simulations are performed. Two such simulations for each model are illustrated on Figure 4.

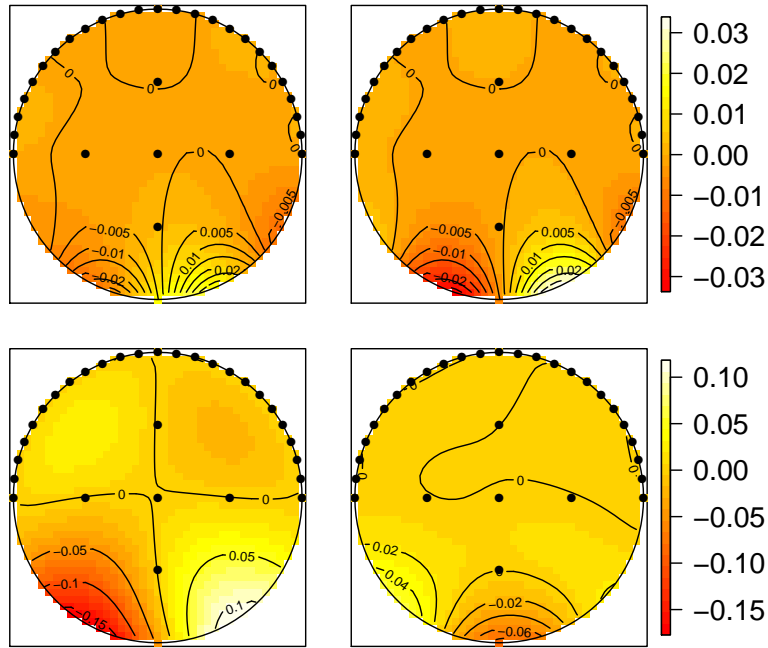


Figure 4: Conditional simulations of solutions to the heat equation problem of Section 4.2. For an improved readability, the mean component of the model has been removed. Top: Harmonic kernel. Bottom: Gaussian kernel.

For each simulated realization, a realization of the maximum is obtained. Histograms of the two 1000-element samples of maxima obtained under both hypotheses on k are illustrated in Figure 5. It can be seen on this example that taking into account the knowledge that the test function is harmonic improve drastically both the mean predictions of the function and the estimation of the maximum obtained by stochastic simulation.

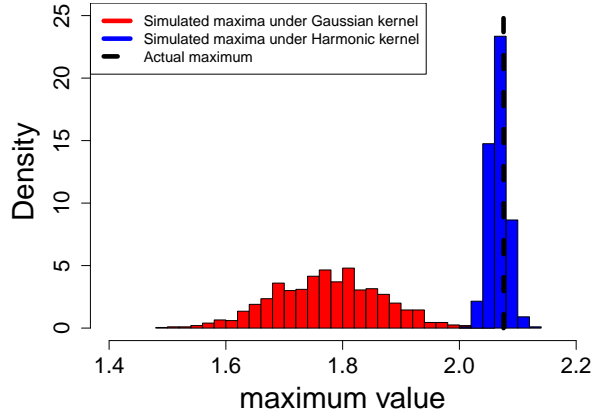


Figure 5: Sample of maxima obtained by Gaussian random field conditional simulations under the two candidate models of Section 4.2. The actual maximum value of the test function is denoted by the straight dashed line.

5. Conclusions and perspectives

This article focuses on the control of pathwise invariances of square-integrable random field through their covariance structure. It is illustrated on various examples how a number of features one may wish to impose on paths such as multivariate sparsity, symmetries, or being solution to a vast class of homogeneous ordinary or partial differential equations may be cast as degeneracy or invariance properties under linear operators.

One of the main results of this work, given in Proposition 1, relates sample path degeneracy to *argumentwise degeneracy* of the covariance kernel, in cases where evaluating the image of a function f by the underlying operator T at some point boils down to integrating f with respect to a signed measure. It turns out that a subclass of these operators, made of combinations of composition operators, suffices to encode various properties such as invariances under finite group actions, or additivity. This allows us to extend recent results from [13] by giving a complete characterization of kernels leading to centred random fields with additive paths, and also to retrieve another result from [11] on kernels leading to random fields with paths invariant under the action of a finite group. Perhaps surprisingly, the obtained results linking sample paths properties to covariance kernels apply to squared-integrable

random fields and do not restrict to the Gaussian case.

Turning then to the particular case of Gaussian random fields, we obtain in Propositions 2 and 3 generalizations of Proposition 1 to broad classes of T 's, that enables constructing Gaussian fields with paths invariant under various integral and differential operators. Proposition 3 is essentially based on the Loève (canonical) isometry between the Hilbert space generated by the field and its reproducing kernel Hilbert space. Finally, Proposition 5 complements the previous results in measure-theoretic settings, with some statement about the propagation of invariances to conditional distributions.

Taking degeneracies and invariances into account in random field modelling is of practical interest, as illustrated in Section 4. Examples involving different kinds of structural priors show how GP prediction may be drastically improved by designing an appropriate kernel. In particular, it was shown how degenerate kernels allow estimating an axis of symmetry by maximum likelihood, but also how it can improve stochastic simulation techniques for estimating non-linear functionals such as the maximum. Perspectives of future work include designing kernels adapted to larger scale interpolation problems where degeneracies or invariances are assumed, but also defining new families of kernels to account for “approximate” invariances, and estimate features or the degree of invariance for predefined classes thereof.

Acknowledgements: the authors would like to thank Dominic Schuhmacher, Fabrice Gamboa, as well as two anonymous referees for useful comments on previous versions of this paper. They are also very grateful to the editor for constructive remarks having led to significant improvements.

References

- [1] G. Matheron, Principles of geostatistics, *Economic Geology* 58 (1963) 1246–1266.
- [2] M. Stein, *Interpolation of Spatial Data, Some Theory for Kriging*, Springer, 1999.
- [3] C. R. Rasmussen, C. K. I. Williams, *Gaussian Processes for Machine Learning*, MIT Press, 2006.
- [4] W. J. Welch, R. J. Buck, J. Sacks, H. P. Wynn, T. J. Mitchell, M. D. Morris, Screening, predicting, and computer experiments, *Technometrics* 34 (1992) 15–25.

- [5] D. Jones, A taxonomy of global optimization methods based on response surfaces, *Journal of Global Optimization* 21 (21) (2001) 345–383.
- [6] A. O’Hagan, Bayesian analysis of computer code outputs: A tutorial, *Reliability Engineering & System Safety* 91 (10-11) (2006) 1290–1300.
- [7] A. Van der Vaart, J. Van Zanten, Rates of contraction of posterior distributions based on Gaussian process priors, *Annals of Statistics* 36 (2008) 1435–1463.
- [8] A. Van der Vaart, J. Van Zanten, Reproducing kernel Hilbert spaces of gaussian priors, in: *Pushing the limits of contemporary statistics: contributions in honor of Jayanta K. Ghosh*, Institute of Mathematical Statistics Collect 3, 2008, pp. 200–222.
- [9] A. Van der Vaart, H. van Zanten, Information rates of nonparametric Gaussian process methods, *Journal of Machine Learning Research* 12 (2011) 2095–2119.
- [10] B. Haasdonk, H. Burkhardt, Invariant kernels for pattern analysis and machine learning, *Machine Learning* 68 (2007) 35–61.
- [11] D. Ginsbourger, X. Bay, O. Roustant, L. Carraro, Argumentwise invariant kernels for the approximation of invariant functions, *Annales de la Faculté des Sciences de Toulouse* 21 (3) (2012) 501–527.
- [12] D. Duvenaud, H. Nickisch, C. E. Rasmussen, Additive Gaussian processes, in: *Advances in Neural Information Processing Systems* 25, Granada, Spain, 2011, pp. 1–8.
- [13] N. Durrande, D. Ginsbourger, O. Roustant, Additive covariance kernels for high-dimensional Gaussian process modeling, *Annales de la faculté des Sciences de Toulouse* 21 (2012) 481 – 499.
- [14] D. Ginsbourger, O. Roustant, D. Schuhmacher, N. Durrande, N. Lenz, On ANOVA decompositions of kernels and Gaussian random field paths, <http://arxiv.org/abs/1409.6008>
- [15] M. Scheuerer, M. Schlather, Covariance models for divergence-free and curl-free random vector fields, *Stochastic Models* 28(3) (2012) 433–451.

- [16] D. Duvenaud, Automatic Model Construction with Gaussian Processes, Ph.D. thesis, University of Cambridge (2014).
- [17] N. Cressie, Statistics for spatial data, Wiley series in probability and mathematical statistics, 1993.
- [18] H. Cramér, M. R. Leadbetter, Stationary and Related Stochastic Processes: Sample Function Properties and Their Applications, Wiley, New York, 1967.
- [19] R. J. Adler, An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes, Vol. 12 of Lecture Notes-Monograph Series, Published by: Institute of Mathematical Statistics, 1990.
- [20] M. Scheuerer, Regularity of the sample paths of a general second order random field, Stochastic Processes and their Applications 120 (2010) 1879–1897.
- [21] M. Scheuerer, A comparison of models and methods for spatial interpolation in statistics and numerical analysis, Ph.D. thesis, University of Göttingen (2009).
- [22] A. Berlinet, C. Thomas-Agnan, Reproducing kernel Hilbert spaces in probability and statistics, Kluwer Academic Publishers, 2004.
- [23] D. Ginsbourger, N. Durrande, O. Roustant, Kernels and designs for modelling invariant functions: From group invariance to additivity, in: D. Uciński, A. Atkinson, C. Patan (Eds.), mODa 10 Advances in Model-Oriented Design and Analysis, Physica-Verlag HD, 2013.
- [24] J. Mercer, Functions of positive and negative type and their connection with the theory of integral equations, Philosophical Transactions of the Royal Society A 209 (1909) 415–446.
- [25] I. Steinwart, C. Scovel, Mercer’s Theorem on General Domains: On the interaction between Measures, Kernels, and RKHSs, Constructive Approximation 35 (2012) 363–417.
- [26] R. Adler, J. Taylor, Random Fields and Geometry, Springer, Boston, 2007.

- [27] I. Steinwart, Convergence types and rates in generic Karhunen-Love expansions with applications to sample path properties, Tech. rep. 2014-007, Fakultät für Mathematik und Physik, Universität Stuttgart (2014).
- [28] B. Rajput, S. Cambanis, Gaussian processes and Gaussian measures, *The Annals of Mathematical Statistics* 43 (6) (1972) 1944–1952.
- [29] V. Tarieladze, N. Vakhania, Disintegration of gaussian measures and average-case optimal algorithms, *Journal of Complexity* 23 (2007) 851–866.
- [30] N. Durrande, D. Ginsbourger, O. Roustant, L. Carraro, ANOVA kernels and RKHS of zero mean functions for model-based sensitivity analysis, *Journal of Multivariate Analysis* 115 (2013) 57–67.
- [31] P. Deheuvels, Karhunen-Love expansions of mean-centered Wiener processes, *Lecture Notes-Monograph Series (High Dimensional Probability)* 51 (2006) 62–76.
- [32] L. Meier, S. Van de Geer, P. Bühlmann, High-dimensional additive modeling, *Annals of Statistics* 37 (2009) 3779–3821.
- [33] T. E. Fricker, J. E. Oakley, N. M. Urban, Multivariate gaussian process emulators with nonseparable covariance structures, *Technometrics* 55(1) (2013) 47–56.
- [34] J. Franco, D. Dupuy, O. Roustant, G. Damblin, B. Iooss, DiceDesign: Designs of Computer Experiments, R package version 1.3 (2013).
- [35] Y. Deville, D. Ginsbourger, O. Roustant, kergp: Gaussian Process Laboratory, R package version 0.1.0 (2015).
- [36] R. Schaback, Solving the Laplace equation by meshless collocation using harmonic kernels, *Advances in Computational Mathematics* 31 (4) (2009) 457–470.
- [37] I. Gihman, A. Skorohod, *The Theory of Stochastic Processes I*, Springer, Berlin, Heidelberg, New York, 1974.

Appendix A. Proof of the main results

Appendix A.1. About the second order framework

Before proving the main result of this section (Proposition 1), that concerns a class of operators acting on second order stochastic processes, let us first focus on a simpler but crucial case involving linear forms.

Let ν be a signed measure on \mathcal{D} , $|\nu| = \nu^+ + \nu^-$ be its variation, and denote by ℓ_ν the linear form over $L^1(\nu) = \{f \in \mathcal{M}(D, \mathbb{R}) : \int |f(\mathbf{u})| d|\nu|(\mathbf{u}) < +\infty\}$ defined by $\ell_\nu(f) = \int f(\mathbf{u}) d\nu(\mathbf{u})$.

Lemma 1. *Let us assume that $m : \mathbf{x} \in D \rightarrow m(\mathbf{x}) = \mathbb{E}[Z_{\mathbf{x}}]$ and $k : (\mathbf{x}, \mathbf{x}') \in D \times D \rightarrow \text{Cov}[Z_{\mathbf{x}}, Z_{\mathbf{x}'}]$ satisfy*

$$\int_D \sqrt{k(\mathbf{u}, \mathbf{u}) + m(\mathbf{u})^2} d|\nu|(\mathbf{u}) < +\infty. \quad (\text{A.1})$$

Then m and k are respectively ν - and $(\nu \otimes \nu)$ -integrable, Z 's trajectories are \mathbb{P} -almost surely ν -integrable, and $\ell_\nu(Z)$ is a squared-integrable real-valued random variable with first and centred second order moments

$$\mathbb{E}[\ell_\nu(Z)] = \ell_\nu(m) = \int m(\mathbf{u}) d\nu(\mathbf{u}), \quad (\text{A.2})$$

$$\text{and } \text{Var}[\ell_\nu(Z)] = \ell_\nu \otimes \ell_\nu(k) = \int_{D \times D} k(\mathbf{u}, \mathbf{v}) d(\nu \otimes \nu)(\mathbf{u}, \mathbf{v}). \quad (\text{A.3})$$

Consequently, the following are equivalent:

- a) $\mathbb{P}(\ell_\nu(Z) = 0) = 1$
- b) $\ell_\nu(m) = 0$ and $\ell_\nu \otimes \ell_\nu(k) = 0$.

Proof of Lemma 1. Equation A.1 entails that m and k are respectively ν - and $(\nu \otimes \nu)$ -integrable as $\int |m(\mathbf{u})| d\nu(\mathbf{u}) \leq \int_D \sqrt{k(\mathbf{u}, \mathbf{u}) + m(\mathbf{u})^2} d\nu(\mathbf{u}) < +\infty$ by Equation A.1, and $\int_{D \times D} |k(\mathbf{u}, \mathbf{v})| d(\nu \otimes \nu)(\mathbf{u}, \mathbf{v}) \leq \left(\int_D \sqrt{k(\mathbf{u}, \mathbf{u})} d\nu(\mathbf{u}) \right)^2 < +\infty$ using $\forall (\mathbf{u}, \mathbf{v}) \in D^2 \quad |k(\mathbf{u}, \mathbf{v})| \leq \sqrt{k(\mathbf{u}, \mathbf{u})} \sqrt{k(\mathbf{v}, \mathbf{v})}$ and Equation A.1 again. Similarly, a successive application of Tonelli's theorem, Cauchy-Schwarz inequality, and Equation A.1 delivers that $\mathbb{E}[\ell_\nu(|Z|)] < +\infty$ whereof $\mathbb{P}(\int |Z_{\mathbf{u}}| d\nu(\mathbf{u}) < +\infty) = 1$, and so Z 's trajectories are \mathbb{P} -almost surely ν -integrable. Equations A.2 and A.3 are similarly obtained by applying Fubini's theorem. Noticing that a) is equivalent to $\mathbb{E}[\ell_\nu(Z)^2] = 0$, we finally get

$$\mathbb{E}[\ell_\nu(Z)^2] = \text{Var}[\ell_\nu(Z)] + \mathbb{E}[\ell_\nu(Z)]^2 = \ell_\nu \otimes \ell_\nu(k) + \ell_\nu(m)^2 \quad (\text{A.4})$$

so we conclude that **a**) holds if and only if $\ell_\nu(m) = 0$ and $\ell_\nu \otimes \ell_\nu(k) = 0$. \square

We are now in position to prove Proposition 1.

Proof of Proposition 1. That **a**) and **b**) are equivalent is a direct consequence of Lemma 1 by noticing that for any given $\mathbf{x} \in D$, $T(Z)_\mathbf{x} = \ell_{\nu_\mathbf{x}}(Z)$, $T(m)(\mathbf{x}) = \ell_{\nu_\mathbf{x}}(m)$, and $(T \otimes T(k))(\mathbf{x}, \mathbf{x}) = \ell_{\nu_\mathbf{x}} \otimes \ell_{\nu_\mathbf{x}}(k)$. The separability assumption (See [37]) on $T(Z)$ finally suffices to deduce **c**) from **a**), while the reciprocal is direct. \square

Appendix A.2. About the Gaussian framework

Proof of Proposition 2. If **a**) holds, then we obtain by boundedness of T and almost sure convergence of the Karhunen-Loève expansion in $L^2(\eta)$ that, with probability 1, $T(Z) = T(m) + T(\sum_{i=1}^{+\infty} \sqrt{\lambda_i} \zeta_i \phi_i) = \mathbf{0} + \sum_{i=1}^{+\infty} \sqrt{\lambda_i} \zeta_i T(\phi_i) = \mathbf{0}$ (η -a.e.). Conversely, **b**) implies that $\mathbb{P}(\|T(Z)\|_{L^2(\eta)}^2 = 0) = 1$ and so $\mathbb{E}(\|T(Z)\|_{L^2(\eta)}^2) = 0$. Now,

$$\begin{aligned} \mathbb{E}(\|T(Z)\|_{L^2(\eta)}^2) &= \mathbb{E} \left(\sum_{i=0}^{+\infty} \sum_{j=0}^{+\infty} \sqrt{\lambda_i} \zeta_i \sqrt{\lambda_j} \zeta_j \langle T(\phi_i), T(\phi_j) \rangle_{L^2(\eta)} \right) \\ &= \|T(m)\|_{L^2(\eta)}^2 + \sum_{i=1}^{+\infty} \lambda_i \|T(\phi_i)\|_{L^2(\eta)}^2, \end{aligned} \quad (\text{A.5})$$

with the convention $\lambda_0 = 1$, $\zeta_0 \sim \delta_1$, and $\phi_0 = m$. From this we get that $\|T(m)\|_{L^2(\eta)} = 0$ and $\forall i \geq 1$, $\lambda_i \|T(\phi_i)\|_{L^2(\eta)}^2 = 0$, so that **a**) is fulfilled. \square

Proof of Proposition 3. First, let $\mathcal{T} : \mathcal{H} \rightarrow \mathbb{R}^D$ be an operator satisfying (10) and the pointwise convergence condition. Since $Z_{\mathbf{x}'} = \Psi(k(\cdot, \mathbf{x}'))$, we have:

$$\mathcal{T}(k(\cdot, \mathbf{x}'))(\mathbf{x}) = \text{cov}(T(Z)_\mathbf{x}, \Psi(k(\cdot, \mathbf{x}'))) \quad (\mathbf{x}, \mathbf{x}' \in D)$$

This is immediately extended in a unique way to \mathcal{H} by linearity and continuity of the isometry Ψ , leading to:

$$\mathcal{T}(h)(\mathbf{x}) = \text{cov}(T(Z)_\mathbf{x}, \Psi(h)) \quad (\mathbf{x} \in D, h \in \mathcal{H}). \quad (\text{A.6})$$

Conversely, using again properties of Ψ , one easily checks that Eq. A.6 defines a linear operator satisfying Eq. 10 and the pointwise convergence condition.

Finally, assuming that for any $\mathbf{x} \in D$, $T(Z)_\mathbf{x} = \mathbb{E}(T(Z)_\mathbf{x})$, we get that $\forall \mathbf{x}' \in D$ $\mathcal{T}(k(\cdot, \mathbf{x}'))(\mathbf{x}) = \text{Cov}(T(Z)_\mathbf{x}, Z_{\mathbf{x}'}) = 0$ and so, by density of the

$k(\cdot, \mathbf{x}')$ functions in \mathcal{H} , that $\mathcal{T}(\mathcal{H}) = \{0\}$. Conversely, assuming that $\mathcal{T}(\mathcal{H}) = \{0\}$ we get that $\text{Cov}(T(Z)_{\mathbf{x}}, \Psi(h)) = 0$ for all $h \in \mathcal{H}$. For the special case $h = \Psi^{-1}(T(Z)_{\mathbf{x}})$, we obtain that $\text{Var}[T(Z)_{\mathbf{x}}] = 0$ and so $T(Z)_{\mathbf{x}} = \mathbb{E}(T(Z)_{\mathbf{x}})$ (a.s.). Since Z is assumed centred and $T(Z)_{\mathbf{x}} \in \mathcal{L}(Z)$, $\mathbb{E}(T(Z)_{\mathbf{x}}) = 0$. \square

Proof of Proposition 4. We know that $T_{\sharp}\mu$ is a Gaussian measure with mean $T(m_{\mu})$ and covariance operator $TC_{\mu}T^*$, so that $T_{\sharp}\mu = \delta_{0_{\mathcal{F}}}$ if and only if $T(m_{\mu}) = 0_{\mathcal{F}}$ and $TC_{\mu}T^* = 0_{\mathcal{F}^* \rightarrow \mathcal{F}}$ (that a Gaussian measure is null if and only if its mean and covariance operator are null can be seen as a consequence of Equations 11 and 12). In other words, $i) \Leftrightarrow ii)$. \square

Proof of Proposition 5. The existence of a Gaussian disintegration $(q_y)_{y \in \mathcal{G}}$ and the definitions of m and R follow from [29, Theorem 3.11]. More precisely, given an infinite $C_{A_{\sharp}\mu}$ -representing sequence $(y_i^*)_{i \in \mathbb{N} \setminus \{0\}}$ and writing $x_i^* = A^*y_i^*$, m is defined as limit when $n \rightarrow +\infty$ of $m_n(y) = \sum_{i=1}^n \langle y, y_i^* \rangle C_{\mu}x_i^*$ for $y \in \mathcal{G}$ such that $m_n(y)$ converges in \mathcal{F} , and 0 else. Similarly, $R = C_{\mu} - R_1$ where $\forall x^* \in \mathcal{F}^* \ R_1x^* = \sum_{i=1}^{+\infty} \langle C_{\mu}x_i^*, x^* \rangle C_{\mu}x_i^*$. Hence there remains to show that $T(m) = 0$ and $TR_1T^* = 0_{\mathcal{F}^* \rightarrow \mathcal{F}}$. For the mean, it is sufficient to show that $T(m_n(y)) = 0$ ($n \geq 1$) as T is bounded and the property passes to the limit for y such that $(m_n(y))_{n \geq 1}$ converges in \mathcal{F} . Now, we have

$$T(m_n(y)) = T \left(\sum_{i=1}^n \langle y, y_i^* \rangle C_{\mu}x_i^* \right) = \sum_{i=1}^n \langle y, y_i^* \rangle TC_{\mu}x_i^* = 0_{\mathcal{F}},$$

as $TC_{\mu} = C_{\mu}T^* = 0_{\mathcal{F}^* \rightarrow \mathcal{F}}$ since $TC_{\mu}T^* = 0_{\mathcal{F}^* \rightarrow \mathcal{F}}$ by assumption. Besides this, for arbitrary $x^* \in \mathcal{F}^*$,

$$TR_1T^*x^* = T \left(\sum_{i=1}^{+\infty} \langle C_{\mu}x_i^*, T^*x^* \rangle C_{\mu}x_i^* \right) = \sum_{i=1}^{+\infty} \langle TC_{\mu}x_i^*, x^* \rangle TC_{\mu}x_i^* = 0_{\mathcal{F}},$$

and so $TR_1T^* = 0_{\mathcal{F}^* \rightarrow \mathcal{F}}$, implying that $TRT^* = 0_{\mathcal{F}^* \rightarrow \mathcal{F}}$. \square