HAL
open science

Invariances of random fields paths, with applications in Gaussian Process Regression

David Ginsbourger, Olivier Roustant, Nicolas Durrande

To cite this version:

David Ginsbourger, Olivier Roustant, Nicolas Durrande. Invariances of random fields paths, with applications in Gaussian Process Regression. 2013. hal-00850436

HAL Id: hal-00850436

https://hal.science/hal-00850436v1

Preprint submitted on 6 Aug 2013

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.

Invariances of random fields paths, with applications in Gaussian Process Regression

David Ginsbourger[a,*], Olivier Roustant[b], Nicolas Durrande[c]

[a] *University of Bern, Department of Mathematics and Statistics, Alpeneggstrasse, 22 CH-3012 Bern, Switzerland*
[b] *Ecole Nationale Supérieure des Mines de Saint-Etienne, FAYOL-EMSE, LSTI, F-42023 Saint-Etienne, France*
[c] *The University of Sheffield, Department of Computer Science, Regent Court, 211 Portobello, Sheffield S1 4DP, UK*

Abstract

We study pathwise invariances of centred random fields that can be controlled through the covariance. A result involving composition operators is obtained in second-order settings, and we show that various path properties including additivity boil down to invariances of the covariance kernel. These results are extended to a broader class of operators in the Gaussian case, via the Loève isometry. Several covariance-driven pathwise invariances are illustrated, including 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 regression.

Keywords: Covariance kernels, Composition operators, RKHS, Bayesian function learning, Structural priors.
2000 MSC: 60G60, 60G17, 62J02.

1. Introduction

Whether for function approximation, classification, or density estimation, probabilistic models relying on random fields have been increasingly used

*Corresponding author
Email addresses: ginsbourger@stat.unibe.ch (David Ginsbourger), roustant@emse.fr (Olivier Roustant), n.durrande@sheffield.ac.uk (Nicolas Durrande)

in recent works from various research communities. Finding their 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 for time-consuming simulations in science and engineering [4, 5, 6, 7]) to theoretical Bayesian statistics in function spaces (See [8, 9, 10] and references therein). Often, a Gaussian random field model is assumed for the function of interest, and so all prior assumptions on this function are incorporated through the corresponding mean function and covariance kernel. Here we focus on random field models for which the covariance kernel exists, and we discuss some mathematical properties of associated realisations (or *paths*) depending on the kernel, both in the Gaussian case and in a more general second-order framework.

A number of well-known random field properties driven by the covariance kernel (say in the centred case) are in the mean square sense [11], e.g. L^2 continuity and differentiability [12]. Such results are quite general in the sense that they hold in a variety of cases (Gaussian or not), but they generally aren't informative about the pathwise beahviour of underlying random fields. In the Gaussian case, however, much can be said about path regularity properties of stationary random field paths (Cf. classical results in [13] and subsequent works) based on the behaviour of the covariance kernel in the neighbourhood of the origin. Likewise, for non-stationary Gaussian fields, results connecting path regularity to kernel properties can be found in [14]. More recently, [15] studied path regularity of second-order random fields, and could draw conclusions about a.s. continuous differentiability in non-Gaussian settings. Also, we refer to the thesis [16] for an enlightening exposition of state-of-the-art results concerning regularity properties of random field sample paths in various frameworks.

In a different settings, links between invariances of kernels under operations like translations and rotations (that is to say, the notions of *stationarity* and *isotropy* [11]) and invariances in distribution of the corresponding random fields have been covered extensively in spatial statistics and throughout the literature of probability theory [17]. However, such properties are to be understood *in distribution*, and do not directly concern random field paths. Our main focus in the present work is on *pathwise* algebraic and geometric properties of random fields, such as invariances under group actions or sparse

function decompositions of multivariate paths.

We first establish in a quite general framework, that for a centred random field $(Z_x)_{x\in D}$ possessing a covariance (i.e. such that the variance is finite at any location in the index space D), Z has paths invariant under T with probability 1 if and only if $\forall \mathbf{x} \in D\ T(k(\cdot, \mathbf{x})) = k(\cdot, \mathbf{x})$, where T belongs to the class of linear combinations of composition operators. The presented results generalise [18], where random fields with paths invariant under the action of a finite group were studied. Here we also extend recent works on additive kernels for high-dimensional kriging [19, 20], and we provide a simple characterization of the class of kernels leading to squared-integrable centred random fields with additive paths. Furthermore, in the particular case of Gaussian random fields, a more general class of invariances can be covered through the link between operators on the paths and operators on the *reproducing kernel Hilbert space* [21] associated with the random field.

Section 2 presents a general result characterizing path invariance in terms of argumentwise invariance of covariance kernels, in the case of combinations of composition operators. In Section 3, we discuss how the Gaussianity assumption enables extending the results of Section 2 to more general operators. The obtained results are applied to Gaussian process regression in Section 4 where the potential of argumentwise invariant kernels is demonstrated through various examples. Section 5 is dedicated to an overall conclusion of the article, and a discussion on some research perspectives.

2. Invariance under combinations of composition operators

2.1. Motivations

Designing kernels imposing some structural constraints on the associated random field models is of interest in various situations.

One of those situations is the high-dimensional function approximation framework, where simplifying assumptions are needed in order to guarantee a reasonable inference despite the curse of dimensionality. Following its successful use in multidimensional nonparametric smoothing [22, 23], the *additivity assumption* has become a very popular simplifying assumption for dealing with high-dimensional problems, and has recently inspiring further work in mathematical statistics [24, 25, 26].

A class of kernels leading to random fields with additive paths, in the sense detailed below (See also [27]), has recently been considered in [19].

Calling a function $f \in \mathbb{R}^D$ (with multidimensional source space $D = \prod_i^d D_i$ where $D_i \subset \mathbb{R}$) additive when there exists $f_i \in \mathbb{R}^{D_i}$ $(1 \leq i \leq d)$ such that $\forall \mathbf{x} = (x_1, \ldots, x_d) \in D,\ f(\mathbf{x}) = \sum_{i=1}^d f_i(x_i)$, it was indeed shown in [19] that

Proposition 1. *If a random field Z possesses a kernel of the form*

$$k(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^{d} k_i(x_i, x_i') \tag{1}$$

where the k_i's are arbitrary positive definite kernels over the D_i's, then Z is additive up to a modification, i.e. there exists a random field A which paths are additive functions such that $\forall \mathbf{x} \in D\ \mathbb{P}(Z_{\mathbf{x}} = A_{\mathbf{x}}) = 1$.

One may wonder whether kernels of the form $k(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^d k_i(x_i, x_i')$ are the only ones giving birth to centred random fields with additive paths. The answer to this question turns out to be negative, as will be established in Corollary 5.

In a similar fashion, [18] gives a characterization of kernels which associated centred random fields have their paths invariant under a finite group action. Let G be a finite group acting on D via a measurable action

$$\Phi : (\mathbf{x}, g) \in D \times G \longrightarrow \Phi(\mathbf{x}, g) = g.\mathbf{x} \in D$$

Proposition 2. *Z has invariant paths under Φ (up to a modification) if and only if k is argumentwise invariant: $\forall (\mathbf{x}, g) \in D \times G,\ k(g.\mathbf{x}, \cdot) = k(\mathbf{x}, \cdot)$.*

We show in Proposition 3 that both Propositions 1 and 2 are sub-cases of a general result on square-integrable random fields invariant under the class of *combination of composition operators* (see Definition 2). A characterization of kernels leading to random fields possessing additive paths is given in Corollary 5, and it is then shown that having the form of Eq. 1 is not necessary. Another by-product of Proposition 3 is a new proof of Proposition 2 relying on a particular class of combination of composition operators, as illustrated in Example 1. Let us now introduce the set up of composition operators (See, e.g., [28]) and their combinations.

2.2. Composition operators and their combinations

Definition 1. *Let us consider an arbitrary function $v : \mathbf{x} \in D \longrightarrow v(\mathbf{x}) \in D$. The composition operator T_v with symbol v is defined as follows:*

$$T_v : f \in \mathbb{R}^D \longrightarrow T_v(f) = f \circ v \in \mathbb{R}^D$$

Remark 1. *Such operators can be naturally extended to random fields indexed by D:*

$$\forall \mathbf{x} \in D,\ T_v(Z)_{\mathbf{x}} = Z_{v(\mathbf{x})}$$

Definition 2. *We call combination of composition operators with symbols $v_i \in D^D$ and weights $\alpha_i \in \mathbb{R}$ $(1 \leq i \leq q, q \in \mathbb{N}\backslash\{0\})$ the operator*

$$T = \sum_{i=1}^{q} \alpha_i T_{v_i}$$

2.3. Invariance under a combination of composition operators

Proposition 3. *Let Z be a square-integrable centred random field with covariance kernel k. Then Z equals $T(Z)$ up to a modification, i.e.*

$$\forall \mathbf{x} \in D,\ \mathbb{P}\left(Z_{\mathbf{x}} = T(Z)_{\mathbf{x}}\right) = 1$$

if and only if k is T-invariant, i.e.

$$\forall \mathbf{x}' \in D,\ T(k(\cdot, \mathbf{x}')) = k(\cdot, \mathbf{x}'). \tag{2}$$

Proof. $\Rightarrow$: let us fix arbitrary $\mathbf{x},\ \mathbf{x}' \in D$. Since $Z_{\mathbf{x}}$ is a modification of $T(Z)_{\mathbf{x}}$, we have $\mathrm{cov}(Z_{\mathbf{x}}, Z_{\mathbf{x}'}) = \mathrm{cov}(T(Z)_{\mathbf{x}}, Z_{\mathbf{x}'}) = \mathrm{cov}(\sum_{i=1}^{q} \alpha_i Z_{v_i(\mathbf{x})}, Z_{\mathbf{x}'})$, and so:

$$k(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^{q} \alpha_i k(v_i(\mathbf{x}), \mathbf{x}') = T(k(\cdot, \mathbf{x}'))(\mathbf{x}).$$

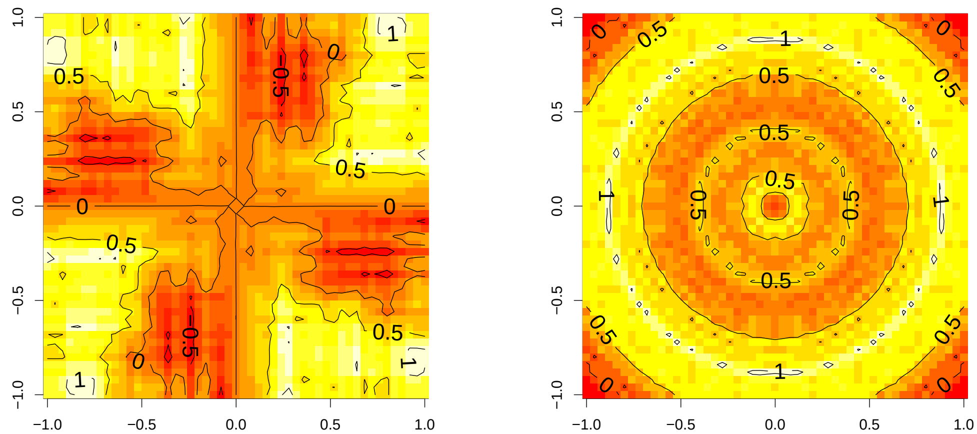
$\Leftarrow$: Using $\forall \mathbf{x}' \in D\ T(k(\cdot, \mathbf{x}')) = k(\cdot, \mathbf{x}')$, we get $\mathrm{var}(T(Z)_{\mathbf{x}}) = \mathrm{cov}(Z_{\mathbf{x}}, T(Z)_{\mathbf{x}}) = \mathrm{var}(Z_{\mathbf{x}})$, so $\mathrm{var}(Z_{\mathbf{x}} - T(Z)_{\mathbf{x}}) = 0$. Since Z is centred, so is $T(Z)$, and hence $Z_{\mathbf{x}} \overset{a.s.}{=} T(Z)_{\mathbf{x}}$. □

Remark 2. *As noted in [29], two processes modifications of each other that are almost surely continuous are indistinguishable. Almost sure results may then directly be obtained for processes with almost surely continuous paths.*

Example 1 (Case of group-invariance)**.** *Prop. 2 now appears as a special case of Prop. 3 with $T(f) = \sum_{i=1}^{\#G} \frac{1}{\#G} f(v_i(\cdot))$ where $v_i(\mathbf{x}) := g_i.\mathbf{x}$ $(1 \leq i \leq \#G)$.*

For instance, let us consider the following functions over $[-1,1]^2 \times [-1,1]^2$:

$$\begin{aligned} k_1(\mathbf{x}, \mathbf{y}) &= min(\rho_{\mathbf{x}} \cos(\tilde{\theta}_{\mathbf{x}}), \rho_{\mathbf{y}} \cos(\tilde{\theta}_{\mathbf{y}})) \times min(\rho_{\mathbf{x}} \sin(\tilde{\theta}_{\mathbf{x}}), \rho_{\mathbf{y}} \sin(\tilde{\theta}_{\mathbf{y}})) \\ k_2(\mathbf{x}, \mathbf{y}) &= min(\rho_{\mathbf{x}}, \rho_{\mathbf{y}}) \end{aligned} \tag{3}$$



(a) Path of a GRF with kernel k_1 (b) Path of a GRF with kernel k_2

Figure 1: Sample paths of centred (Gaussian) random fields with two kinds of argumentwise rotation invariant kernels. The fact that the contour lines are not exactly rotation invariant on the right panel can be explained by the 2-dimensional mesh used for plotting.

where $\rho_{\mathbf{x}}, \theta_{\mathbf{x}}$ *(resp.* $\rho_{\mathbf{y}}, \theta_{\mathbf{y}}$*) are the polar coordinates of* $\mathbf{x}$ *(resp.* $\mathbf{y}$*) and where* $\tilde{\theta} = \theta \mod \pi/2$*.* k_1 *and* k_2 *are positive definite kernels (in the loose sense) as admissible covariances (respectively those of the Brownian Sheet and the Brownian Motion) composed with a change of index, i.e.* $k_i(\cdot,\cdot) = k(h(\cdot), h(\cdot))$ *with* h *from* $[-1,1]^2$ *onto* $[0,1]^2$ *(resp. from* $\mathbb{R}^2$ *onto* $\mathbb{R}_+$*). By construction* k_1 *and* k_2 *are argumentwise invariant with respect to rotations around the origin (with angles multiple of* $\pi/2$ *for* k_1*). As illustrated in Figure 1, Proposition 3 ensures that the sample paths of centred (Gaussian or non-Gaussian) random fields with these kernels inherit their invariance properties. Note that* k_1 *belongs to the class of kernels argumentwise invariant under the action of a finite group treated above, while the argumentwise invariance of* k_2 *under the action of an infinite group can actually be seen as invariance under any composition operator with symbol of the form* $v(\mathbf{x}) = \rho_{\mathbf{x}}\mathbf{u}$ *where* $\mathbf{u}$ *is an arbitrary point on the unit circle.*

Beyond group-invariance, Proposition 3 also has implications concerning the sparsity of multivariate random field paths, as detailed in the following section on additivity, leading to a generalization of Proposition 1.

2.4. Kernels of centred random fields with additive paths

Let us first show how the additivity property boils down to an invariance property under some specific class of combination of composition operators.

Remark 3. *Assuming* $\mathbf{a} \in D$*, a function* $f : D \to \mathbb{R}$ *is additive if and only if* f *is invariant under the following combination of composition operators:*

$$T(f)(\mathbf{x}) = \sum_{i=1}^{d} f(v_i(\mathbf{x})) - (d-1) f(v_{d+1}(\mathbf{x})) \qquad (\mathbf{x} \in D) \tag{4}$$

where $v_i(\mathbf{x}) := (a_1, \ldots, a_{i-1}, \underbrace{x_i}_{i\text{th coordinate}}, a_{i+1}, \ldots, a_d)$*, and* $v_{d+1}(\mathbf{x}) := \mathbf{a}$*.*

Corollary 1. *A centred random field* Z *possessing a covariance kernel* k *has additive paths (up to a modification) if and only if* k *is a positive definite kernel of the form*

$$k(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^{d} \sum_{j=1}^{d} k_{ij}(x_i, x'_j) \tag{5}$$

Proof. If Z has additive paths up to a modification, there exists a random field $(A_{\mathbf{x}})_{\mathbf{x} \in D}$ with additive paths such that $\forall \mathbf{x} \in D$ $\mathbb{P}(Z_{\mathbf{x}} = A_{\mathbf{x}}) = 1$, and so Z and A have the same covariance kernel. Now, A having additive paths, Remark 3 implies that $A_{\mathbf{x}} = \sum_{i=1}^{d} A_{v_i(\mathbf{x})} - (d-1) A_{v_{d+1}(\mathbf{x})} = \sum_{i=1}^{d} A^i_{\mathbf{x}_i}$, where $A^i_{\mathbf{x}_i} := A_{v_i(\mathbf{x})} - \frac{(d-1)}{d} A_{v_{d+1}(\mathbf{x})}$, so Equation 5 holds with $k_{ij}(\mathbf{x}_i, \mathbf{x}'_j) := \mathrm{cov}(A^i_{\mathbf{x}_i}, A^j_{\mathbf{x}'_j})$. Reciprocally, from Proposition 3, we know that it suffices for Z to have additive paths that $k(\cdot, \mathbf{x}')$ is additive $\forall \mathbf{x}' \in D$. For a kernel k such as in Eq. 5 and an arbitrary $\mathbf{x}' \in D$, setting

$$\forall \mathbf{x}_i \in D_i,\ \widetilde{k}_i(\mathbf{x}_i, \mathbf{x}') := \sum_{j=1}^{d} k_{ij}(\mathbf{x}_i, \mathbf{x}'_j) \quad (1 \leq i \leq d)$$

we get $k(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^{d} \widetilde{k}_i(\mathbf{x}_i, \mathbf{x}')$ $(\mathbf{x} \in D)$, so $k(\cdot, \mathbf{x}')$ is additive. □

Example 2. *Let us consider the following kernel over* $\mathbb{R}^d \times \mathbb{R}^d$*:*

$$k(\mathbf{x}, \mathbf{y}) = \sum_{i,j=1}^{d} \int_{\mathbb{R}} \kappa_i(x_i - u) \kappa_j(y_j - u) \mathrm{d}u \tag{6}$$

where the κ_i are smoothing kernels over $\mathbb{R}$. Previous results on vector-valued random fields ensure that k is a valid covariance function [30]. Furthermore, the structure of k corresponds to an additive kernel in the sense of Equation 5. According to Corollary 5, a random field with such kernel has additive paths (up to a modification), with univariate marginals exhibiting possible cross-correlations.

3. Extension to further operators. Focus on the Gaussian case

Composition operators constitute a remarkable class of linear maps since they can be defined on function spaces without any restriction. In particular, they similarly apply to random field paths or to kernel functions (with one argument fixed), so that taking out of a covariance a (combination of) composition(s) applied to a random field and turning it into a (combination of) composition(s) on the covariance kernel appears as a natural operation.

For more general classes of operators, however, operators on paths and operators on kernels are two different mathematical objects: It is a priori not obvious how to transform operators on paths into operators on the kernel, and even less straightforward to know when and how it is possible to define an operator on paths corresponding to a given operator on the kernel space.

Given a linear operator $T : \mathbb{R}^D \to \mathbb{R}^D$ and a second-order centred process Z such that $T(Z)$ is second order, generalizing the approach that lead to Prop. 3 enables us to characterize pathwise invariances of Z by T relying on second-order properties of the joint process $(Z_{\mathbf{x}}, T(Z)_{\mathbf{x}})_{\mathbf{x}\in D}$, without any additional assumption concerning Z's probability distribution:

Proposition 4. *$Z = T(Z)$ up to a modification if and only if*

$$k(\mathbf{x}, \mathbf{x}) = 2\,\mathrm{cov}(T(Z)_{\mathbf{x}}, Z_{\mathbf{x}}) - \mathrm{var}(T(Z)_{\mathbf{x}}) \qquad (\mathbf{x} \in D). \tag{7}$$

Proof. Under the square-integrability and zero-mean hypotheses on Z and $T(Z)$, $\mathrm{var}(T(Z)_{\mathbf{x}} - Z_{\mathbf{x}}) = 0$ is equivalent to $\mathbb{P}(T(Z)_{\mathbf{x}} = Z_{\mathbf{x}}) = 1$. □

In the particular case of combinations of composition operators covered by Prop. 3, it was possible to take T out of the covariance and variance in the right hand side of Eq. 7, leading to a further equivalence between pathwise invariance of Z and invariance of k under T. In greater generality, however, it is not straightforward how T can be taken out of terms such as $\mathrm{cov}(T(Z)_{\mathbf{x}}, Z_{\mathbf{x}})$.

We show in Section 3.3 below that in case Z is Gaussian and T satisfies some technical condition with respect to Z, there exists an operator $\mathcal{T}$ defined over the Reproducing Kernel Hilbert Space associated with Z, such that $\mathrm{cov}(T(Z)_{\mathbf{x}}, Z_{\mathbf{x}'}) = \mathcal{T}(\mathrm{cov}(Z_{\cdot}, Z_{\mathbf{x}'}))(\mathbf{x})$, for all $\mathbf{x}, \mathbf{x}' \in D$.

This construction based on the celebrated *Loève isometry* [21] then enables us extending Prop. 3 to a broader class of operators.

Numerical examples are presented throughout the current section, including simulated paths of Gaussian random fields with argumentwise invariant kernels under various (integral and differential) operators, that subsequently serve as a basis to original applications in Gaussian Process regression, presented in Section 4.

3.1. Operating on the kernel via operators on paths

We focus here on a centred Gaussian random field $(Z_{\mathbf{x}})_{\mathbf{x}\in D}$ defined over a compact set $D \subset \mathbb{R}^d$, with covariance kernel $k : D \times D \longrightarrow \mathbb{R}$. k is here assumed continuous, so that the paths of $(Z_{\mathbf{x}})_{\mathbf{x}\in D}$ belong to some subspace of the space $\mathcal{B}$ of continuous functions over D, and are in particular square-integrable (with respect to Lebesgue's measure on D, say) by compacity of D. Let us further consider a linear map $T : \mathbb{R}^D \to \mathbb{R}^D$ acting on the paths of Z and such that $T(Z)_{\mathbf{x}}$ is centred and square-integrable for all $\mathbf{x} \in D$.

In Proposition 6 below, the so-called *Loève isometry* [21] allows us to define an operator, derived from T, acting on the Reproducing Kernel Hilbert Space (RKHS) associated with k. Let us first recall some useful definitions and the isometry in question. The RKHS $\mathcal{H}$ associated with k [31] can be defined as functional completion of the function space spanned by the $k(\cdot, \mathbf{x})$'s:

$$\mathcal{H} = \overline{\mathrm{span}(k(\cdot, \mathbf{x}), \mathbf{x} \in D)}$$

equipped with the scalar product defined by $\langle k(\cdot, \mathbf{x}), k(\cdot, \mathbf{x}')\rangle_{\mathcal{H}} = k(\mathbf{x}, \mathbf{x}')$. A crucial state-of-the-art result is that $\mathcal{H}$ is isometric to the Hilbert space generated by the random field Z [21]:

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

where the adherence is taken with respect to the usual $L^2(\mathbb{P})$ topology on the space of (equivalence classes of) square-integrable random variables.

Proposition 5. *(Loève isometry) The map $\Psi : \mathcal{H} \to \mathcal{L}(Z)$ defined by:*

$$k(\cdot, \mathbf{x}) \to Z_{\mathbf{x}}$$

for all $\mathbf{x} \in D$ *and extended by linearity and continuity, is an isometry from* $(\mathcal{L}(Z), \langle \cdot, \cdot \rangle_{L^2})$ *to* $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$.

As shown below, the Loève isometry allows us to link operators on the paths of Z to corresponding operators on the RKHS, provided that the random variables $T(Z)_{\mathbf{x}}$ ($\mathbf{x} \in D$) belong to $\mathcal{L}(Z)$:

Proposition 6. *Let* $T : \mathcal{B} \to \mathbb{R}^D$ *be such that for any* $\mathbf{x} \in D$, $T(Z)_{\mathbf{x}} \in \mathcal{L}(Z)$. *Then, there exists a unique operator* $\mathcal{T} : \mathcal{H} \to \mathbb{R}^D$ *satisfying*

$$\mathrm{cov}(T(Z)_{\mathbf{x}}, Z_{\mathbf{x}'}) = \mathcal{T}(k(., \mathbf{x}'))(\mathbf{x}) \quad (\mathbf{x}, \mathbf{x}' \in D) \tag{8}$$

and such that $\mathcal{T}(h_n)(\mathbf{x}) \longrightarrow \mathcal{T}(h)(\mathbf{x})$ *for all* $\mathbf{x} \in D$ *and* $h_n \xrightarrow{\mathcal{H}} h$.

Proof. Let $\mathcal{T} : \mathcal{H} \to \mathbb{R}^D$ be an operator satisfying (8) and the pointwise convergence condition. Since $Z_{\mathbf{x}'} = \Psi(k(., \mathbf{x}'))$, we have:

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

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

$$\mathcal{T}(h)(\mathbf{x}) = \mathrm{cov}(T(Z)_{\mathbf{x}}, \Psi(h)) \quad (\mathbf{x} \in D, h \in \mathcal{H}). \tag{9}$$

Conversely, using again properties of Ψ, one easily checks that (9) defines a linear map satisfying (8) and the pointwise convergence condition. □

The construction proposed above will serve as basis for an invariance result, given in Prop. 7. Before stating it, let us examine and discuss in more detail the assumptions made in Prop. 6 and the relation between T and $\mathcal{T}$, both through examples and analytical considerations.

Example 3. *Let* T *be a linear combination of composition operators,* $T = \sum_{i=1}^{q} \alpha_i T_{v_i}$, *such as introduced in Def. 2. Recall that* T *similarly applies to random field paths or to kernel functions, with* $T(Z)_{\cdot} = \sum_{i=1}^{q} \alpha_i Z_{v_i(\cdot)}$ *and* $T(k(\cdot, \mathbf{x}')) = \sum_{i=1}^{q} \alpha_i k(v_i(\cdot), \mathbf{x}')$. *In particular, we directly obtain that the condition* $T(Z)_{\mathbf{x}} \in \mathcal{L}(Z)$ *is fulfilled, so that Prop. 6 can be applied.*

It is then easy to check that $\mathcal{T}(k(\cdot, \mathbf{x}')) = T(k(\cdot, \mathbf{x}'))$, *implying that* $\mathcal{T}$ *is the unique representer of* T *on* $\mathcal{H}$ *satisfying the pointwise convergence condition of Prop. 6. In other words, here* $\mathcal{T} = T_{|\mathcal{H}}$.

Note that this example also illustrates that $\mathcal{T}(\mathcal{H})$ may differ from $\mathcal{H}$. Indeed, choosing a composition operator $T = T_v$ and fixing $\mathbf{x}' \in D$, we have

$$\mathcal{T}(k(.,\mathbf{x}'))(\mathbf{x}) = cov(Z_v(\mathbf{x}), Z_{\mathbf{x}'}) = k(v(\mathbf{x}), \mathbf{x}'),$$

and so $\mathcal{T}(k(.,\mathbf{x}')) = k(v(.),\mathbf{x}')$. Taking for instance the 1-dimensional RKHS $\mathcal{H}$ spanned by the 1st order polynomial $e_1(t) = t$ on $D = [0,1]$ (with kernel $k(t,t') = e_1(t)e_1(t') = tt'$) and choosing $v(t) = t^2$, we see that $\mathcal{T}(k(.,t')) : t \to t^2t'$ is a second order polynomial, and thus not in $\mathcal{H}$.

Example 4. *Let us now consider a measure ν on D such that*

$$\int_D \sqrt{k(\mathbf{u},\mathbf{u})}\mathrm{d}\nu(\mathbf{u}) < +\infty,$$

and define $T(Z)_{\mathbf{x}} = \int_D Z_{\mathbf{u}} d\nu(\mathbf{u})$ for all $\mathbf{x} \in D$. Then, relying on the Fubini-Tonelli theorem, $\mathcal{T}(h) = \int_D h(\mathbf{u})d\nu(\mathbf{u})$. In other words, $\mathcal{T} = T_{|\mathcal{H}}$ again.

3.2. A detour through the Karhunen-Loève expansion

In both Examples 3 and 4, we found out that $\mathcal{T} = T_{|\mathcal{H}}$. From this, we may wonder under which circumstances $\mathcal{T}$ is a restriction of T. The spectral framework, and more specifically the *Karhunen-Loève (KL) expansion*, is a suitable setting to investigate such question. As a preliminary to a sufficient condition for $\mathcal{T} = T_{|\mathcal{H}}$ to hold, let us recall some useful basics concerning the KL expansion.

In a nutshell, the starting point of KL is the *Mercer decomposition* (See [32], with generalizations in [33, 34]) of the continuous covariance kernel k: Given any finite measure ν on the Borel algebra of D whose support is D (typically the Lebesgue measure λ), there exists an orthonormal basis $(\varphi_n)_{n\geq 1}$ of $L^2(\nu)$ and a sequence of non-negative real numbers $(\gamma_n)_{n\geq 1} \downarrow 0$ such that:

$$k(\mathbf{x},\mathbf{x}') = \sum_{n=1}^{+\infty} \gamma_n \varphi_n(\mathbf{x})\varphi_n(\mathbf{x}'), \tag{10}$$

where the convergence is absolute and uniform on D. Note that the finite trace hypothesis $\int_D k(\mathbf{u},\mathbf{u})\mathrm{d}\nu(\mathbf{u}) < +\infty$ often given as prerequisite of the Mercer theorem is automatically fulfilled here, considering the assumptions made on k.

Relying on Eq. 10, it is well-known (See, e.g., [21]) that the RKHS $\mathcal{H}$ can then be represented as a subspace of $L^2(\nu)$, in the following way:

$$\mathcal{H} = \left\{ \mathbf{x} \to f(\mathbf{x}) = \sum_{n=1}^{+\infty} \alpha_n \varphi_n(\mathbf{x}), \quad s.t. \quad \sum_{n=1}^{+\infty} \frac{\alpha_n^2}{\gamma_n} < +\infty \right\}. \tag{11}$$

Furthermore, relying on the Loève isometry (Cf. Prop. 5), the random field Z itself can be expanded with respect to the φ_n's, leading to the KL expansion:

$$Z_{\mathbf{x}} = \sum_{n=1}^{+\infty} \sqrt{\gamma_n} \zeta_n \varphi_n(\mathbf{x}) \tag{12}$$

where the ζ_n's are independent standard Gaussian random variables, and the series is uniformly convergent with probability 1 [35]. In particular, noting that $\mathbb{E}\left[\int_D Z_{\mathbf{u}}^2 \mathrm{d}\nu(\mathbf{u})\right] = \int_D k(\mathbf{u},\mathbf{u}) \mathrm{d}\nu(\mathbf{u}) < +\infty$, we get (with probability 1) both that the paths of Z are in $L^2(\nu)$ and that the series of Eq. 12 converges normally. Consequently, in case of a bounded operator T from $L^2(\nu)$ to itself,

$$T(Z)_{\cdot} = \sum_{n=1}^{+\infty} \sqrt{\gamma_n} \zeta_n T(\varphi_n)(\cdot) \tag{13}$$

with probability 1, where the convergence is normal. Note that in cases such as the one of the differentiation operator (See, e.g., for a differentiation of the KL expansion), T is not bounded with respect to the usual $L^2(\nu)$ norm, but similar normal convergence results may be obtained by considering a source space of differentiable elements equipped with an *ad hoc* topology (e.g., Sobolev spaces).

Concerning our question on operators on paths vs on the RKHS, we obtain by substituting Z and $T(Z)$ by their respective expansions in $\mathcal{T}$'s definition:

$$\begin{aligned} \mathcal{T}(k(\cdot,\mathbf{x}')) &= \mathrm{cov}(T(Z)_{\cdot}, Z_{\mathbf{x}'}) \\ &= \mathrm{cov}\left(\sum_{n=1}^{+\infty} \sqrt{\gamma_n} \zeta_n T(\varphi_n)(\cdot), \sum_{m=1}^{+\infty} \sqrt{\gamma_m} \zeta_m \varphi_m(\mathbf{x}') \right) \\ &= \sum_{n=1}^{+\infty} \gamma_n \varphi_n(\mathbf{x}') T(\varphi_n)(\cdot). \end{aligned} \tag{14}$$

Now, using the Mercer decomposition (10) of k and the boundedness of T, $T(k(\cdot,\mathbf{x}')) = \sum_{n=1}^{+\infty} \gamma_n \varphi_n(\mathbf{x}') T(\varphi_n)(\cdot)$, so we conclude that $\mathcal{T} = T_{|\mathcal{H}}$. Besides, on may also notice that $\zeta_n \in \mathcal{L}(Z)$ since $\zeta_n = \frac{1}{\sqrt{\gamma_n}} \langle Z_{\cdot}, \varphi_n \rangle_{L^2(\nu)}$. Using that $\mathbb{E}[||T(Z)_{\cdot}||^2_{L^2(\nu)}] < +\infty$, we finally also get that $T(Z)_{\mathbf{x}} \in \mathcal{L}(Z)$ ν-a.e.

3.3. Invariances and Gaussian random fields

Coming back to invariances, we now give a characterization result, that generalizes those of Section 2 in the particular case of Gaussian fields:

Proposition 7. *Under the assumptions of Prop. 6, the three following conditions are equivalent:*

(i) $Z = T(Z)$ *up to a modification*

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

(iii) $\mathcal{T} = Id_{\mathcal{H}}$

Proof. By the pointwise convergence condition on $\mathcal{T}$, (ii) and (iii) are equivalent. Now, let us prove the equivalence between (i) and (ii). Since for any arbitrary $\mathbf{x} \in D$, $Z_{\mathbf{x}} - T(Z)_{\mathbf{x}} \in \mathcal{L}(Z)$, we have by duality:

$$\begin{aligned} Z_{\mathbf{x}} \overset{a.s.}{=} T(Z)_{\mathbf{x}} &\iff Z_{\mathbf{x}} - T(Z)_{\mathbf{x}} \overset{a.s.}{=} 0 \\ &\iff \mathrm{cov}(Z_{\mathbf{x}} - T(Z)_{\mathbf{x}}, Z_{\mathbf{x}'}) = 0 \ \forall \mathbf{x}' \in D \\ &\iff k(\mathbf{x}, \mathbf{x}') = \mathcal{T}(k(\cdot, \mathbf{x}'))(\mathbf{x}) \ \forall \mathbf{x}' \in D \end{aligned}$$

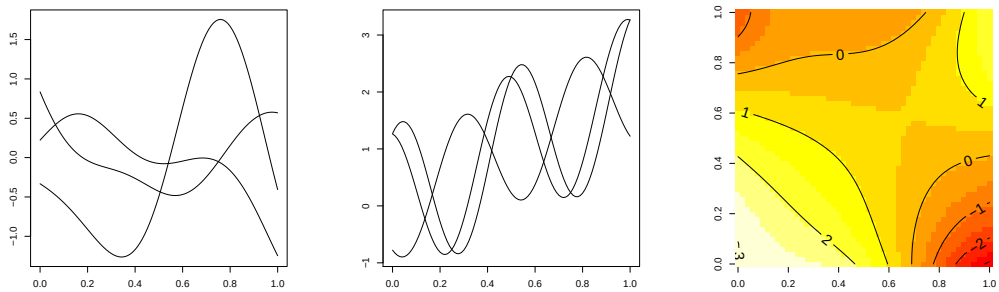
□

Proposition 7 can be used to define families of centred Gaussian field models satisfying linear-type properties, simply by looking at their kernel. This includes for instance the case of Gaussian random fields with centred paths (or *mean-centered* fields, to use the terminology of [37]) and Gaussian random fields whose paths are solutions of linear differential equations, as illustrated below (See also [38] for recent results on vector fields with divergence-free and curl-free paths).

Example 5 (Gaussian random fields with centred paths)**.** *Let ν be a probability measure on $D \subset \mathbb{R}^d$ such that $\int_D \sqrt{k(\mathbf{u}, \mathbf{u})} d\nu(\mathbf{u}) < +\infty$. Then Z has centered paths – i.e. $\int_D Z_{\mathbf{u}} \mathrm{d}\nu(\mathbf{u}) = 0$ – if and only if $\int_D k(\mathbf{x}, \mathbf{u}) \mathrm{d}\nu(\mathbf{u}) = 0, \forall \mathbf{x} \in D$. Indeed, define T by $T(Z)_{\mathbf{x}} = Z_{\mathbf{x}} - \int_D Z_{\mathbf{u}} \mathrm{d}\nu(\mathbf{u})$ for all $\mathbf{x} \in D$. Following Example 4, we have $\mathcal{T}(h) = h - \int_D h(\mathbf{u}) \mathrm{d}\nu(\mathbf{u})$, and the result comes from Proposition 7.*

For instance, the kernel k_0 defined by

$$\begin{aligned} k_0(\mathbf{x}, \mathbf{y}) = k(\mathbf{x}, \mathbf{y}) &- \int k(\mathbf{x}, \mathbf{u}) \mathrm{d}\nu(\mathbf{u}) \\ &- \int k(\mathbf{y}, \mathbf{u}) \mathrm{d}\nu(\mathbf{u}) + \int k(\mathbf{u}, \mathbf{v}) \mathrm{d}\nu(\mathbf{u}) \mathrm{d}\nu(\mathbf{v}) \end{aligned} \tag{15}$$



(a) Centred sample paths, based on k_0.

(b) Random solutions to Eq. 16, based on k_{ode}.

(c) Harmonic sample path, based on k_{harm}.

Figure 2: Examples of sample paths invariant under various operators.

satisfies the above condition. Figure 2 (a) shows some sample paths of a centred Gaussian random process based possessing a kernel of that form.

Example 6. *We illustrate here the case where the sample paths of a Gaussian process are solution to the differential equation:*

$$y''(t) + y(t) = 2t. \tag{16}$$

The solutions of the associated homogeneous equation are the functions satisfying $y = -y''$ so they correspond to invariant functions with respect to $T: \ f \to -f''$. The solutions of the homogeneous equation are well-known to be in $\mathrm{span}(\cos, \sin)$ *which can be endowed with the following kernel*

$$k_{ode}(s,t) = (\cos(s) \ \ \sin(t)) \ \Sigma \ (\cos(s) \ \ \sin(t))^t \tag{17}$$

where Σ is a symmetric positive semi-definite 2×2 matrix. $k_{ode}(s, \cdot)$ is solution to the homogeneous equation (i.e. $k_{ode}(s, \cdot)$ is T-invariant) for all s so the sample paths of a centred Gaussian process with such kernel inherit this property. Let Y be a Gaussian process with mean $\mu(t) = 2t$ and covariance k. Since μ is a particular solution of Eq. 16, Y has sample paths satisfying this differential equation. This is illustrated in Figure 2.b.

Example 7. *In the previous example, the solutions of the ODE belong to a 2-dimensional space. We consider here another ODE, the Laplace equation $\Delta f = 0$, for which the space of solutions is infinite dimensional. The*

solutions to this equation are called harmonic functions and we will call harmonic kernels any positive definite function satisfying the ODE argument-wise: $(\Delta k(\cdot, \mathbf{x}')) = 0$ $(\mathbf{x}' \in D)$. *Examples of such harmonic kernels can be found in the recent literature (See respectively [39, 40] for 2D and 3D input spaces). We will focus here on the following kernel over* $\mathbb{R}^2 \times \mathbb{R}^2$*:*

$$k_{harm}(\mathbf{x}, \mathbf{y}) = \exp\left(\frac{x_1 y_1 + x_2 y_2}{\theta^2}\right) \cos\left(\frac{x_2 y_1 - x_1 y_2}{\theta^2}\right). \tag{18}$$

Proposition 7 can be applied to the operator $f \to f + \Delta f$ *so the sample paths obtained with* k_{harm} *also are harmomic functions. This can be seen in the right panel of Figure 2 where the sample path shows some special features of harmonic functions such as the absence of local minimum.*

4. Applications in Gaussian process regression

The aim of this section is to discuss and illustrate the use of argument-wise invariant kernels in Gaussian process regression (GPR). The main idea behind this approach is to incorporate invariance assumptions within GPR. As we will see, using such kernels can significantly improve the predictivity of GPR in cases where *structural priors* on the function to approximate, involving invariances under bounded linear operators, are available.

GPR gives a very convenient stochastic framework for modelling a function $f : D \to \mathbb{R}$ based on a finite set of observations $y_i = f(\mathbf{x}^{(i)})$, $1 \leq i \leq n$ and a Gaussian process prior on f. The literature of GPR and related methods is scattered over several fields including statistics and geostatistics [2], machine learning [3] and functional analysis [21]. Depending on the scientific community, the predictor of f is either defined as best linear unbiased predictor of a square-integrable (or intrinsic) random field, conditional expectation of a Gaussian Process, or interpolator with minimal norm in RKHS settings. One striking fact is that, given any positive definite kernel k, the approaches end up with the same expressions for the best predictor and for the "conditional" kernel describing the remaining uncertainty on f:

$$\begin{aligned} m(\mathbf{x}) &= \mathbf{k}(\mathbf{x})^t K^{-1} \mathbf{Y} \\ c(\mathbf{x}, \mathbf{x}') &= k(\mathbf{x}, \mathbf{x}') - \mathbf{k}(\mathbf{x})^t K^{-1} \mathbf{k}(\mathbf{x}') \end{aligned} \tag{19}$$

where $\mathbf{k}(\mathbf{x}) = (k(\mathbf{x}, \mathbf{x}^{(i)}))_{1 \leq i \leq n}$ and $K = (k(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}))_{1 \leq i,j \leq n}$. We will discuss in the next section the influence of using invariant kernel in such models.

4.1. Gaussian Process Regression with invariant kernels

We consider here a bounded operator T on the paths as defined in Section 3, and we use for convenience the same letter to denote T's restriction to the RKHS $\mathcal{H}$. Assuming that $k(\mathbf{x}, .) = T(k(\mathbf{x}, .))$ ($\mathbf{x} \in D$), it was already established that the paths of a centred Gaussian random field with kernel k are invariant under T. We now establish further that both the GPR predictor and the conditional distribution of such random field knowing response values at a finite set of points are invariant as well.

Proposition 8. *Let Z be a centred Gaussian field with argumentwise T-invariant kernel k and $y_i = Z(\mathbf{x}^{(i)})$ $(1 \leq i \leq n)$ be a finite set of observations. Then,*

(i) The GPR predictor m is T-invariant

(ii) The "conditional covariance kernel" c is argumentwise T-invariant

(iii) Z conditioned on the evaluation results is T-invariant, up to a modification. Consequently, conditional simulations of Z are T-invariant.

Proof. The properties (i) and (ii) are a direct consequence of the linearity of T. For example, we have for (ii):

$$\begin{aligned} T(c(\mathbf{x}, .))(\mathbf{x}') &= T(k(\mathbf{x}, .) - \mathbf{k}(\mathbf{x})^t K^{-1}\mathbf{k}(.))(\mathbf{x}') \\ &= T(k(\mathbf{x}, .))(\mathbf{x}') - \mathbf{k}(\mathbf{x})^t K^{-1} T(\mathbf{k}(.))(\mathbf{x}') \\ &= k(\mathbf{x}, \mathbf{x}') - \mathbf{k}(\mathbf{x})^t K^{-1}\mathbf{k}(\mathbf{x}'). \end{aligned} \tag{20}$$

Furthermore the conditional distribution of Z knowing evaluation results simplifies as $\mathcal{L}(Z|Z(\mathbf{X}) = \mathbf{Y}) \equiv m + \mathcal{GRF}(0, c)$, where $\mathcal{GRF}(0, c)$ stands for the distribution of a centred Gaussian random field with covariance kernel c. According to Proposition 7, a random field with distribution $\mathcal{GRF}(0, c)$ is T-invariant up to a modification. (iii) follows using the linearity of T. □

4.2. Illustration on examples

We now consider invariant kernels introduced in the examples of the previous sections and study associated GPR models and predictions. More precisely, we focus on 3 case studies involving various priors: zero-mean functions, solutions to $y''(x) + y(x) = 2x$ and solutions to $\Delta y(\mathbf{x}) = 0$.

GPR with centred paths. Here we assume that $D = [-\pi, \pi]$ and that the function to approximate is $f(x) = \cos(x) + \cos(2x) + \cos(3x) + \sin(x/2)$. Assuming that for some reason, it is known *a priori* that the integral of f over D is zero, it is of particular interest to incorporate this knowledge into the model. To get an insight of how GPR is improved by incoporating this structural prior, we compare predictions based on the following kernels:

$$\begin{aligned} k(t,t') &= \exp\left(-4(t-t')^2\right) \\ k_{inv}(t,t') &= k(t,t') - \int k(t,u)\mathrm{d}u - \int k(t',u)\mathrm{d}u + \iint k(u,v)\mathrm{d}u\mathrm{d}v. \end{aligned} \tag{21}$$

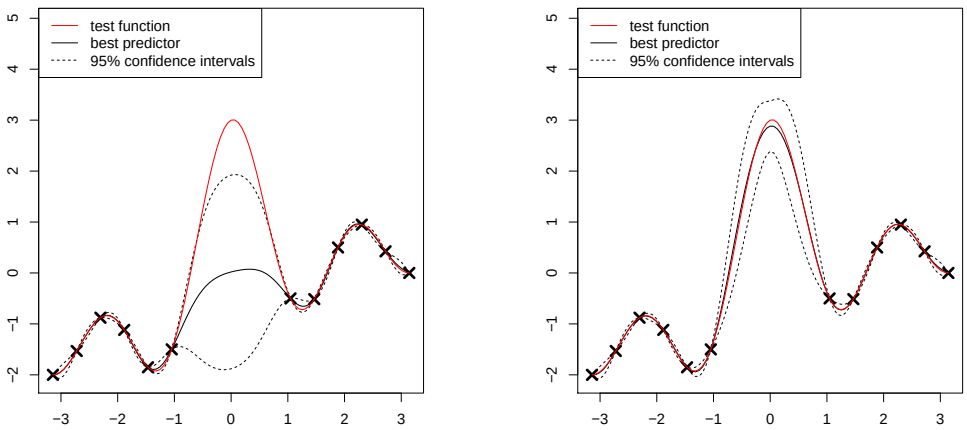
The integral of k_{inv} with respect to any of its variables is zero, so the paths of the associated centred Gaussian Process will inherit this property. As a consequence, choosing a kernel such as k_{inv} allows incorporating the prior information $\int f(u)\mathrm{d}u = 0$ in GPR modeling, as illustrated in Figure 3 (a).

In a second time, we assume that evaluation results $y_i = f(x^{(i)})$ at 12 distinct points $x^{(i)} \in D$ are available, and we compare Gaussian Process conditional distributions based on both kernels. As seen in Figure 3 the use of k_{inv} improves considerably GPR predictions since m recovers the large peak in the center of the domain. This is reflected by root integrated squared errors, with values of 0.04 and 1.06 for k_{inv} and k, respectively.

GPR of a solution to a univariate linear ODE. We saw in Section 3 that a Gaussian Process Y with mean $\mu(x) = 2x$ and kernel given by Eq. 17 is equivalent to a process with paths satisfying the ODE $y''(x) + y(x) = 2x$. Figure 4 shows the conditional distribution of Y given evaluations at one or two points. It can be seen on the right panel that the prediction uncertainty collapses as soon as Y is evaluated at two distinct points. With this behaviour, the model reflects the unicity of the solution to such ODE under two equality conditions.

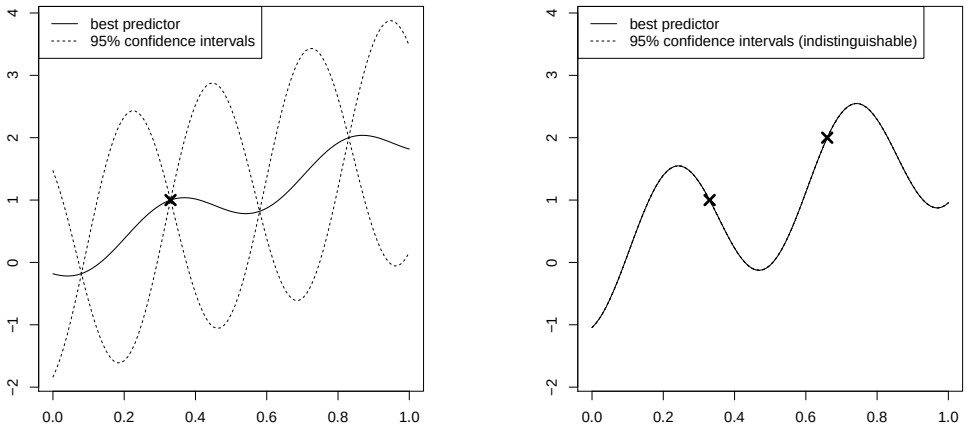
GPR of an harmonic function. We now consider the function $f(\mathbf{x}) = \cos(1 - x_1)\exp(x_2)$, a solution to $\Delta y = 0$. As seen previously, the kernel given in Eq. 18 satisfies this equation argumentwise so it allows to incorporate a structural prior of harmonicity in the GPR model. Figure 5 shows the resulting predictions based on four observations, and the associated prediction error.

Since the best predictor m and f are harmonic, so is also the prediction error $m - f$. It implies that the maximum error is located on the boundary



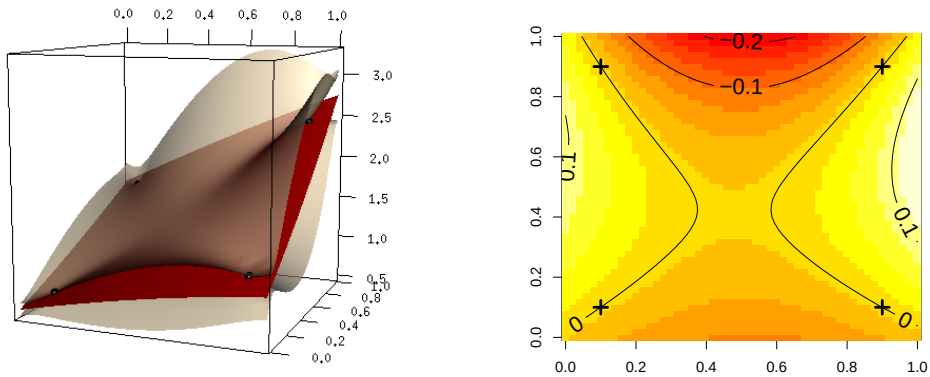
(a) GPR with kernel k (b) GPR with kernel k_{inv}

Figure 3: Comparison of two GPR models. On the left panel, the model is based on an usual squared exponential kernel whereas on the right one it takes into account the zero-mean property of the function to approximate.



(a) GPR with one data-point (b) GPR with two data-points

Figure 4: Examples of Gaussian process models based on a GP Y satisfying $y''(x) + y(x) = 2x$. In this example, the matrix Σ of Eq. 17 is set to identity.



(a) Mean predictor and 95% confidence intervals

(b) prediction error

Figure 5: Example of GPR model based on an argumentwise harmonic kernel.

of the domain (See Figure 5.b). This property may be of interest for the construction of design of experiments for learning harmonic functions.

Sparse ANOVA kernels. Given a product probability measure $\nu = \otimes_{1 \leq i \leq d} \nu_i$ over $\mathbb{R}^d$, High Dimensional Model Representation (HDMR, See, e.g., [41, 42]) corresponds to the decomposition of any $f \in L^2(\nu)$ as the sum of a constant, univariate effects, and interactions terms with increasing orders:

$$f(\mathbf{x}) = f_0 + \sum_{i=1}^{d} f_i(x_i) + \sum_{i<j} f_{i,j}(x_i, x_j) + \cdots + f_{1,\ldots,d}(\mathbf{x}) \qquad (\mathbf{x} \in D) \tag{22}$$

where the f_I's ($I \subset \{1, \ldots, d\}$) satisfy $\int_I f(x_i) \mathrm{d}\nu_i(x_i) = 0$ for all $i \in I$ [42]. This decomposition is of great interest for defining *variance-based* global sensitivity indices (usually referred to as *Sobol' indices*) quantifying the influence of each variable or group of variables on the response:

$$S_I = \frac{\mathrm{Var}[f_I(\mathbf{X}_I)]}{\mathrm{Var}[f(\mathbf{X})]} \tag{23}$$

where $\mathbf{X}$ is a random vector with probability distribution ν. Sparsity of f, in the sense of having many terms equal to zero in Eq. 22, can be interpreted as invariance with respect to operators of the form $T(f) = f - P_I(f)$, where P_I

denotes the projection operator mapping f to f_I. We will now illustrate on a popular function from the sensitivity analysis literature how taking prior knowledge of such sparsity into account may highly benefit prediction.

Sobol's g-function [43] is defined on $[0,1]^d$ as

$$g(\mathbf{x}) = \prod_{i=1}^{d} \frac{|4x_i - 2| + a_i}{1 + a_i} \tag{24}$$

where the a_i's are arbitrary positive parameters. Beyond the convenient fact that the dimensionality d is tunable, one particular feature of g is that the global sensitivity indices have a closed form expression [43]:

$$S_I = \frac{\prod_{i \in I} \beta_i}{\prod_{i=1}^{d}(1+\beta_i) - 1} \qquad \text{with } \beta_i = \frac{1}{3}(1+a_i)^{-2}. \tag{25}$$

Prior knowledge of Sobol' indices allows defining a subset $\mathcal{S}$ of main effects and interactions with a significant influence on the output. We will consider hereafter that terms explaining less than 1e-3 % of variance are non significant.

We consider the g-function in ten dimensions ($d = 10$) with parameter values $\mathbf{a} = (0,0,0,2,2,2,4,4,4,8)$. In such settings, $\mathcal{S}$ is a set of 23 subsets of indices including all the main effects except the one of x_{10}, and some first and second order interaction terms. Let us now compare prediction performances obtained with four GPR models respectively based on the following kernels:

$$\begin{aligned}
k_{add}(\mathbf{x},\mathbf{y}) &= \sigma_0^2 + \sum_{i=1}^{d} k_i^0(x_i, y_i) \\
k_{spa}(\mathbf{x},\mathbf{y}) &= \sigma_0^2 + \sum_{I \in \mathcal{S}} \prod_{i \in I} k_i^0(x_i, y_i) \\
k_{anova}(\mathbf{x},\mathbf{y}) &= \sigma^2 \prod_{i=1}^{d} (k_i^0(x_i, y_i) + 1) \\
k_{gauss}(\mathbf{x},\mathbf{y}) &= \sigma^2 \prod_{i=1}^{d} \exp\left(-\frac{(x_i - y_i)^2}{\theta_i^2}\right).
\end{aligned} \tag{26}$$

where the k_i^0 correspond to argumentwise centred Gaussian kernels as in Eq. 15, parameterized by variances σ_i^2 and lengthscales θ_i. The kernels k_{add},

k_{spa}, k_{anova} and k_{gauss} are respectively parametrized by 21, 19, 21 and 11 parameters. Since the product in the expression of k_{anova} can be expanded as a sum of 2^d kernels with increasing interaction orders, k_{add} and k_{spa} can be seen as sparse variations of k_{anova} where most of the terms are set to zero.

The learning set is made of 100 uniformly distributed points over the input space. The parameters σ_i^2 and θ_i as well as the observation noise τ^2 are estimated by maximum likelihood. Furthermore, a test set of 1000 uniformly distributed points is considered for assessing the model accuracies. The obtained results are summarized in Table 1. It can be seen that the model based on k_{gauss} performs rather poorly on this example. This can be explained by the fact that k_{gauss} does not include any bias term (i.e. a constant in the kernel expression). As a consequence, the associated model tends to come back to 0 when the prediction point is far from the training points. On the other hand, models based on k_{add}, k_{spa} and k_{anova} perform significantly better since they explain at least half of the variance of g. The sum of sensitivity indices associated to the main effects shows that 66% of the variance of g is explained by its additive part so the additive structure assumed by k_{add}, though not completely unrealistic, is a strong assumption that disadvantages the model. Conversely, the structures of k_{spa} and k_{anova} are well-suited to the problem at hand and the associated models give the best results. This is particularly true for k_{spa} which only includes the relevant terms for approximating g.

kernel	k_{add}	k_{spa}	k_{anova}	k_{gauss}
Log-likelihood	-32.36	-12.65	**-1.48**	-45.73
RMSE	0.98	**0.74**	0.86	1.17
Q^2	0.49	**0.71**	0.62	0.28

Table 1: Comparison of the predictivity of models based on various kernels.

In this example, the best model has been obtained by using a sparse kernel obtained from the knowledge of sensitivity indices. Since the latter are usually not available, the issue of automatic sparsity detection is of great importance in practice. Various methods based on a trade-off between a L^2-norm and a L^1-norm can be found in the literature (see for example [44, 45]).

5. Concluding remarks and perspectives

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

One of the main results of this work, given in Proposition 3, relates sample path invariances to the *argumentwise invariance* of the covariance kernel, in cases where T is a combination of composition operators. Although conceptually simple, such class of operators suffices to describe various mathematical properties on functions such as invariances under finite group actions, or additivity (i.e., being sum of univariate functions). This result allows us in particular to extend recent results from [19] by giving a complete characterization of kernels leading to centred random fields with additive paths, and also to retrieve another result from [18] 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 the covariance 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 Proposition 7 a generalization of Proposition 3 to a broader class of operators, that enables constructing Gaussian fields with paths invariant under various integral and differential operators. The core results essentially base on the Loève isometry between the Hilbert space generated by the field and its Reproducing Kernel Hilbert Space. Perspectives include revisiting those invariance results in measure-theoretic settings.

Taking invariances into account in random field modelling and prediction is of huge practical interest, as illustrated in Section 4. Various examples involving different kinds of structural priors show how Gaussian process regression models may be drastically improved by designing an appropriate invariant kernel. One striking fact is that invariances assumptions may increase the accuracy of the model even if the function to approximate is not perfectly invariant. This can be seen on the last example where the assumed sparsity allows to improve the model by avoiding the curse of dimensionality.

Acknowledgements: The authors would like to thank Fabrice Gamboa for his advice regarding the present article.

[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] A. O'Hagan, Bayesian analysis of computer code outputs: A tutorial, Reliability Engineering & System Safety 91 (10-11) (2006) 1290–1300.

[6] D. Jones, A taxonomy of global optimization methods based on response surfaces, Journal of Global Optimization 21 (21) (2001) 345–383.

[7] T. Santner, B. Williams, W. Notz, The design and analysis of computer experiments, Springer Verlag, 2003.

[8] 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.

[9] 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.

[10] A. Van der Vaart, H. van Zanten, Information rates of nonparametric gaussian process methods, Journal of Machine Learning Research 12 (2011) 2095–2119.

[11] N. Cressie, Statistics for spatial data, Wiley series in probability and mathematical statistics, 1993.

[12] P. Abrahamsen, A review of gaussian random fields and correlation functions, second edition, Tech. rep., Norwegian Computing Center (1997).

[13] H. Cramér, M. R. Leadbetter, Stationary and Related Stochastic Processes: Sample Function Properties and Their Applications, Wiley, New York, 1967.

[14] 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.

[15] M. Scheuerer, Regularity of the sample paths of a general second order random field, Stochastic Processes and their Applications 120 (2010) 1879–1897.

[16] M. Scheuerer, A comparison of models and methods for spatial interpolation in statistics and numerical analysis, Ph.D. thesis, University of Göttingen (2009).

[17] K. Parthasarathy, K. Schmidt, Positive Definite Kernels, Continuous Tensor Products, and Central Limit Theorems of Probability Theory, Lecture Notes in Mathematics, Springer, 1972.

[18] 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.

[19] 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.

[20] D. Duvenaud, H. Nickisch, C. E. Rasmussen, Additive Gaussian processes, in: Advances in Neural Information Processing Systems 25, Granada, Spain, 2011, pp. 1–8.

[21] A. Berlinet, C. Thomas-Agnan, Reproducing kernel Hilbert spaces in probability and statistics, Kluwer Academic Publishers, 2004.

[22] C. Stone, Additive regression and other nonparametric models, The Annals of Statistics (1985) 689–705.

[23] T. Hastie, R. Tibshirani, Generalized additive models, Chapman & Hall/CRC, 1990.

[24] L. Meier, S. Van de Geer, P. Bühlmann, High-dimensional additive modeling, Annals of Statistics 37 (2009) 3779–3821.

[25] G. Raskutti, M. Wainwright, B. Yu, Minimax-optimal rates for sparse additive models over kernel classes via convex programming (2011), http://arxiv.org/abs/1008.3654.

[26] G. Gayraud, Y. Ingster, Detection of sparse additive functions, Electron. J. Statist. 6 (2012) 1409–1448.

[27] J. V. Liu, Karhunen-Loève expansion for additive brownian motions, Stochastic Processes and their Applications 123 (2013) 4090–4110.

[28] R. Singh, J. Manhas, Composition Operators on Function Spaces, Elsevier, 1993.

[29] D. Revuz, M. Yor, Continuous Martingales and Brownian motion, Springer, 1991.

[30] T. E. Fricker, J. E. Oakley, N. M. Urban, Multivariate gaussian process emulators with nonseparable covariance structures, Technometrics (just-accepted).

[31] N. Aronszajn, Theory of reproducing kernels, Transaction of the American Mathematical Society 68 (3) (1950) 337 – 404.

[32] 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. doi:10.1098/rsta.1909.0016.

[33] H. König, Eigenvalue distribution of compact operators, Birkhäuser Verlag, 1986.

[34] I. Steinwart, C. Scovel, Mercers theorem on general domains: On the interaction between measures, kernels, and rkhss, Constructive Approximation 35 (3) (2012) 363–417.

[35] J. Kuelbs, Expansions of vectors in a Banach space related to Gaussian measures, Proceedings of the American Mathematical Society 27(2) (1971) 364–370.

[36] T. Kadota, Differentiation of Karhunen-Loève expansion and application to optimum reception of sure signals in noise, IEEE Transactions on information theory 13(2) (1967) 255–260.

[37] P. Deheuvels, Karhunen-Loève expansion for a mean-centered brownian bridge, Statistics and Probability Letters 77 (2007) 1190–1200.

[38] M. Scheuerer, M. Schlather, Covariance models for divergence-free and curl-free random vector fields, Stochastic Models 28(3) (2012) 433–451.

[39] R. Schaback, Solving the Laplace equation by meshless collocation using harmonic kernels, Advances in Computational Mathematics 31 (4) (2009) 457–470.

[40] Y. Hon, R. Schaback, Solving the 3d Laplace equation by meshless collocation via harmonic kernels, Advances in Computational Mathematics 38 (1) (2013) 1–19.

[41] I. Sobol, Theorems and examples on high dimensional model representation, Reliability Engineering & System Safety 79 (2) (2003) 187–193.

[42] F. Y. Kuo, I. H. Sloan, G. W. Wasilkowski, H. Wozniakowski, On decompositions of multivariate functions, Mathematics of Computation 79 (2010) 953–966.

[43] I. Sobol, Theorems and examples on high dimensional model representation, Reliability Engineering & System Safety 79 (2) (2003) 187–193.

[44] F. Bach, High-dimensional non-linear variable selection through hierarchical kernel learning, Tech. rep., INRIA - WILLOW Project-Team. Laboratoire d'Informatique de l'Ecole Normale Supérieure (2009).

[45] S. Gunn, M. Brown, SUPANOVA: A sparse, transparent modelling approach, in: Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop, IEEE, 1999, pp. 21–30.