Paths of Stochastic Processes: a Sudden Turnaround

Robert Schaback and Emilio Porcu

Abstract. Paths and their properties (*e.g.*, regularity) have a fundamental importance to many theoretical and applied disciplines. Yet, there is a pitfall in the same definition of paths that has been accepted by the literature with no major criticism, nor an attempt to fix it. The commonly accepted definition of paths starts from a random field but ignores the problem of setting joint distributions of infinitely many random variables for defining paths properly afterwards. This paper provides a *turnaround* that starts with a given covariance function, then defines paths and *finally* a random field. We show how this approach retains essentially the same properties for Gaussian fields while allowing to construct random fields whose finite dimensional distributions are not Gaussian. Specifically, we start with a kernel C and the associated Reproducing Kernel Hilbert Space $\mathscr{H}(C)$, and then assign standardized random values to a deterministic orthonormal expansion in $\mathscr{H}(C)$. This yields paths as random functions with an explicit representation formula. Using Loéve isometry, we prove that pointwise regularity notions like continuity or differentiability hold on functions of $\mathscr{H}(C)$, paths, and the random field R_C in precisely the same way. Yet, norms of paths as functions behave differently, as we prove that paths are *a.e.* not in $\mathscr{H}(C)$, but in certain *larger* spaces that can partially be characterized. In case of Matérn kernels generating Sobolev space $W_2^m(\mathbb{R}^d)$, paths lie almost surely in all $W_2^p(\mathbb{R}^d)$ for $p < m - d/2$, but almost surely not in $W_2^{m-d/2}$. This regularity gap between function and paths is explained easily by square summability of expansion coefficients of functions, not of paths. The required orthonormal expansions, well-known in the probabilistic and the deterministic literature, are analyzed and compared with respect to convergence rates.

Key words and phrases: Kernel, Loève Isometry, Regularity, RKHS, Sobolev Spaces, Smoothness.

1. INTRODUCTION

1.1 The Problem

The literature about *paths* of random fields is ubiquitous. Reconstruction of the plethora of contributions from mathematics, statistics, machine learning and numerical analysis is beyond the scope of this paper. Nonetheless, this paper is about *paths*, their definition and their regularity properties.

To define paths, it is customary to start with a *stochastic process* or *random field* R on a set Ω, as a mapping that assigns a random variable $R(x)$ to each $x \in \Omega$. The random variables have finite variance, with zero mean, and we stick to these assumptions throughout the paper. The set Ω will here be a bounded domain in $\mathbb{R}^d$.

The fact that the random field has finite variance ensures the *covariance function* $C : \Omega \times \Omega \to \mathbb{R}$ being equal to

$$C(x,y) := \mathrm{cov}(R(x),R(y)) = \mathbb{E}(R(x) \cdot R(y)), \quad (1)$$

for all $x, y \in \Omega$, to be well defined. Yet, it is hardly ever explained how the expectation $\mathbb{E}$ above is to be taken - see the attempt in Chapter I of Christakos (2013). In probability theory, such an expectation is taken over the joint probability space for the pair $(R(x),R(y))$, but in case of

Robert Schaback is Professor Emeritus at Department of Mathematics and Computer Science, Georg-August-Universität Göttingen, Germany (e-mail: schaback@math.uni-goettingen.de).
Emilio Porcu is Professor at Department of Mathematics, Khalifa University, and a Senior Fellow at ADIA Lab, Abu Dhabi, The United Arab Emirates (e-mail: emilio.porcu@ku.ac.ae).

a random field over Ω it should be the marginalization of a joint probability space for the full set $\{R(x) : x \in \Omega\}$. In general, the above covariance can depend on all other $R(z)$, $z \in \Omega$, and then (1) does not make sense, notationally. This is where problems start for the general situation.

Things get even worse when *paths* are to be defined. By general consent, they are simultaneous realizations of all $R(x)$ at the same instant, but how does this work? Nature is able to provide a temperature or wind velocity field simultaneously at all points of Earth and all given times, but what is the mathematical technology to simulate that? How can one throw uncountably many strongly dependent dice at the same time?

Here, the general consent is to say that the map $x \mapsto R(x,\xi)$ is a *path*, where ξ is a real value from the state space $\mathbb{R}$ of the identical probability space of all $R(x)$. This would mean that paths are functions on $\Omega \times \mathbb{R}$ and conceals that there is a draw from the joint probability space in the background that yields itself a function of x. This notational nonsense even made it into the wikipedia (Wikipedia contributors, 2024b), following plenty of books and papers that are not cited here. For each fixed realization ξ, now from the joint probability space, whatever it may be, the deterministic function $x \mapsto R(x,\xi(x))$ should be called a path R_ξ of R. This should be the base to study, *e.g.*, the regularity properties of sample paths. To make an example, imagine to take derivatives of it with respect to x.

Because standard probability distributions do not exist on infinite dimensional spaces (Oxtoby, 1946), both steps work directly on finite sets Ω only, and require a workaround for infinite sets. This is the main goal of this paper.

1.2 The Turnaround

To define paths, the standard approach performs the steps

Random Field R $\Rightarrow$ Covariance C $\Rightarrow$ Paths p

where the first two steps require a nonexisting joint distribution. In contrast to this, our turnaround is

Covariance C $\Rightarrow$ Paths p $\Rightarrow$ Random Field R

with two intermediate steps to be described below. It will avoid joint distributions, and it models Nature better, because it starts from the global picture. It should be the standard modelling technique. Before details of the Turnaround are provided in Section 2, the standard workaround should be presented.

1.3 The Gaussian Workaround

Gaussian processes are extremely popular in all disciplines, including approximation theory, statistics, and in particular machine learning (Seeger, 2004, Williams and Rasmusser 1995, 2006). Their main feature is that the joint distribution of finitely many random variables is multivariate Gaussian, and all marginal and conditional distributions are Gaussian again.

This leads to a standard workaround, exploiting the fact that when the Gaussian process is restricted to a finite point set $X \subset \Omega$, the covariance function is restricted to $X \times X$ and inserted as a matrix into the multivariate Gaussian distribution. If x and y are elements of X, the covariance of $R(x)$ and $R(y)$ is then given by $C(x,y)$ and is independent of X. Therefore, one can work bottom-up on large finite sets in the Gaussian case (Billingsley, 1995). The smart trick is to define paths through limits on countable dense subsets of Ω, avoiding the notion of a joint Gaussian distribution on Ω. Hence, certain properties of paths can be defined through continuation to Ω. The strategy now is

Covariance C $\Rightarrow$ joint distribution

$\Rightarrow$ Random Field R

$\Rightarrow$ Paths p

but restricted to finite sets for the first two steps and a continuity argument for the last.

Now the definition of paths needs a transition from finite to infinite point sets. And since regularity properties are based on limits, this transition needs special care because it involves a double limit. However, the Gaussian workaround is successful and presented widely in the literature.

In the course of the paper, it will be shown that the Turnaround is fully compatible with the Gaussian workaround, while simplifying all arguments considerably. In particular, our path definition leads to an explicit formula (3), and Theorem 6 will handle the path continuation problem.

There is a highly technical and abstract workaround towards abstract Wiener space that we do not follow here. It is well described in Stroock (2023).

1.4 Roadmap

The next section will provide the details of the proposed Turnaround. In particular, two intervening steps are needed: the Hilbert space $\mathscr{H}(C)$ where the covariance kernel C is reproducing, and the orthonormal expansions therein, which will be the basis for defining paths first and then the random field R_C.

Based on this, pointwise regularity theorems for random fields R_C and functions from the native space $\mathscr{H}(C)$ are treated simultaneously in Section 5. A basic tool is the Loéve isometry from Section 5.1 that comfortably relates the deterministic Hilbert space $\mathscr{H}(C)$ to an isometric Hilbert space $\mathscr{H}(R_C)$ of second-order zero-mean random functions. Deterministic and probabilistic results turn out to be in total agreement, if interpreted appropriately, but they do not address paths.

Then Section 6 turns to paths, pointwise, as functions, and with norms of paths in function spaces as random variables. Pointwise regularity of paths follows the previous section and is proven to agree with pointwise regularity of functions in $\mathscr{H}(C)$ and regularity of the random field R_C. Yet, a gap appears between regularity of paths as functions and functions, generated by the same covariance kernel C. The crucial point is that the coefficients of paths (3) are not square summable almost surely, while the corresponding expansions for functions in $\mathscr{H}(C)$ are. Exploiting convergence results for orthonormal expansions, this allows to characterise the larger spaces in which the paths live.

As an interlude, Section 3.3 recalls the standard orthonormal expansions in Hilbert spaces and of random fields. They come either eigensystem-based as Mercer or Karhunen-Loéve expansions, or point-based as Newton expansions or Cholesky decompositions of kernel matrices. The former have optimal convergence properties, while the latter depend on how well points are chosen. But if points are chosen by the P-greedy strategy known from Müller and Schaback (2009), Pazouki and Schaback (2011), the Newton/Cholesky decompositions have an asymptotically equivalent convergence rate (Santin and Haasdonk, 2017, Santin and Schaback, 2016).

2. THE TURNAROUND

Our Turnaround consists of the following steps:

Deterministic Phase: D1 We start with a continuous, symmetric and positive definite kernel $C : \Omega \times \Omega \to \mathbb{R}$. This is done in Section 3.1;

D2 Kernels are reproducing in a Hilbert space of functions on Ω, denoted $\mathscr{H}(C)$ throughout. This is explained through Section 3.2;

D3 We consider orthonormal systems (ONS) $W = \{w_n\}_{n\in\mathbb{N}}$ in $\mathscr{H}(C)$ in Section 3.3. They have an explicit connection with C and $\mathscr{H}(C)$ in points 1 and 2 through

$$\sum_{n\in\mathbb{N}} w_n^2(x) = C(x,x),\ x \in \Omega. \tag{2}$$

Transition to Randomness: The second phase will allow for a proper definition of paths that covers a central role in this manuscript. Specifically,

R1 In Section 4, we select a univariate standard probability space over $\mathbb{R}$ with Lebesgue measure. We then take independent samples, denoted S throughout, and being sequences $S = \{s_n\}_{n\in\mathbb{N}}$ in $\mathbb{R}^{\mathbb{N}}$ where each component is an independent realization of a random variable following a zero-mean variance-one distribution $\mathscr{R}$.

R2 This allows to define $(S,W,\mathscr{R})$-*Expansion Paths*, denoted $p_{S,W,\mathscr{R}}$, and specified through the identity

$$p_{S,W,\mathscr{R}} = \sum_{n\in\mathbb{N}} s_n w_n. \tag{3}$$

R3 This implies, for every x, the existence of a random variable $R_C(x)$ whose samples are

$$R_{C,W,\mathscr{R}}(x) = p_{S,W,\mathscr{R}}(x) = \sum_{n\in\mathbb{N}} s_n w_n(x), \tag{4}$$

and it will turn out that

$$\operatorname{cov}(R_{C,W,\mathscr{R}}(x), R_{C,W,\mathscr{R}}(y)) = C(x,y)$$

holds for all $x,y \in \Omega$.

Note that this proceeds from C to paths first, and then to random fields. In more detail:

$$C \underset{\text{repro}}{\overset{D2}{\Rightarrow}} \mathscr{H}(C) \underset{\text{ONS}}{\overset{D3}{\Rightarrow}} W \underset{\text{sample}}{\overset{R1}{\Rightarrow}} S \underset{\text{path}}{\overset{R2}{\Rightarrow}} p_{S,W,\mathscr{R}} \underset{\text{RV}}{\overset{R3}{\Rightarrow}} R_C(x)$$

It works in general because (2) implies that $p_{S,W,\mathscr{R}}(x)$ is finite almost surely. In contrast to the standard approach, the realizations of the random variables $R_C(x)$ are *exactly* the path values at x. Expansions are used widely in the literature, but not as starting points for defining paths and random variables.

The special case of a Gaussian random field is attained by setting $\mathscr{R} = \mathscr{N}(0,1)$ without needing a detour via limits of point sets. These are hidden in the construction of the orthonormal system, when applying either the standard Cholesky-based sampling procedure or a Mercer expansion of the kernel, see Section 3.3. In the latter case, the result is a Karhunen-Loéve expansion of the random field.

Expansions are abundantly used in the literature when it comes to investigate path properties, e.g. in Steinwart (2019), but they are not used for the definition of paths. When starting from a random field R, expansions lead to a random field $\tilde{R}$ that has to be related to R, e.g. by being called a "modification". Starting from C and expansions eliminates this problem.

The turnaround will be shown to work flawlessly, lose nothing compared to the standard workaround via Gaussian Processes, but allowing more general distributions and an easy access to regularization results concerning paths.

3. DETERMINISTIC PHASE IN THE WAY TO PATHS

As announced above, we proceed here from covariance functions C as kernels to the Hilbert spaces $\mathscr{H}(C)$ they generate, including their orthonormal expansions. In Section 4, randomness comes into play and leads to paths and random fields.

3.1 D1: Kernels

Covariance functions are called *kernels* in the deterministic literature (Aronszajn, 1950, Buhmann, 2003, Fasshauer and McCourt, 2015, Wendland, 2004) and are real-valued mappings

$$C : \Omega \times \Omega \to \mathbb{R}.$$

They are *symmetric*, i.e. $C(x,y) = C(y,x)$, $x,y \in \Omega$ and in most cases *positive definite*. This means that for each finite subset $X_n = \{x_1,\ldots,x_n\}$ of Ω the symmetric $n \times n$ *kernel matrix* $\mathbf{C}_{X_n,X_n}$ with entries $C(x_j,x_k)$, $1 \le j,k \le n$ is positive definite. A kernel C on Ω is called *stationary* or *translation-invariant* if it is a function of the *lag* $x-y$. We shall be sloppy for writing C again for a kernel that is translation invariant, and we always assume positive definiteness.

A very common additionally assumption for stationary kernels is *radial symmetry* or *isotropy*:

$$C(x-y) = \sigma^2 \varphi(\|x-y\|), \tag{5}$$

for $x,y \in \mathbb{R}^d$ and $\|\cdot\|_2$ denoting the Euclidean distance. For convenience, the function φ is continuous and normalized, so that $\varphi(0) = 1$. The parameter σ^2 then is the constant variance of R. In many cases, the scalar $r = \|x-y\|$ is used for the shortcut $C(r) = \sigma^2\varphi(r)$. Whenever no confusion can arise, the word isotropy will be omitted whenever it is apparent from the context. We use the words *covariance function* or *covariance kernel* if there is a probabilistic background, and *kernel* elsewhere, keeping the notation C.

We now describe the rôle of kernels as covariance functions, ignoring their importance for machine learning (Schölkopf and Smola, 2002, Sutton et al., 2012) and meshless methods for solving partial differential equations (Schaback and Wendland, 2006).

Any isotropic covariance in $\mathbb{R}^d$ has a representation as a nonnegative mixture of the type (Schoenberg, 1938, Theorem 1):

$$\varphi(r) = 2^{\frac{d}{2}-1}\Gamma\left(\frac{d}{2}\right) r^{1-\frac{d}{2}} \int_0^\infty z^{1-\frac{d}{2}} J_{\frac{d}{2}-1}(zr)\mathrm{d}F_d(z),$$

$r > 0$, where F_d is a nondecreasing bounded measure on $(0,+\infty)$. We follow Daley and Porcu (2014) to call F_d a d-Schoenberg measure throughout. When φ is absolutely integrable, then F_d is absolutely continuous with respect to the Lebesgue measure, with radial spectral density $\widehat{\varphi}$ that is attained through

$$\widehat{\varphi}(z) = \frac{1}{(2\pi)^{\frac{d}{2}}} z^{1-\frac{d}{2}} \int_0^{+\infty} r^{\frac{d}{2}} J_{\frac{d}{2}-1}(zr)\varphi(r)\mathrm{d}r, \qquad z > 0.$$

An very important special case of random fields has isotropic Matérn covariance functions

$$\mathscr{M}_{\nu,\alpha}(r) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{r}{\alpha}\right)^\nu \mathscr{K}_\nu\left(\frac{r}{\alpha}\right), \qquad r \ge 0, \tag{6}$$

where $\alpha > 0$ is a scale parameter, $\nu > 0$ is called the *smoothness* parameter, and $\mathscr{K}_\nu$ is a modified Bessel function of the second kind of order ν. The isotropic Matérn spectral density or radial Fourier transform is (Abramowitz and Stegun, 1970, 11.4.44)

$$\widehat{\mathscr{M}_{\nu,\alpha}}(z) = \frac{\Gamma(\nu+d/2)}{\pi^{d/2}\Gamma(\nu)} \frac{\alpha^d}{(1+\alpha^2 z^2)^{\nu+d/2}}, \quad z \ge 0, \tag{7}$$

where $z = \|\omega\|$ is the scalar variable in Fourier space corresponding to r. See Porcu et al. (2023) for an overview of random fields with Matérn covariance functions, among other cases with similar spectral behaviour.

The *Generalized Wendland* (Gneiting, 2002, Zastavnyi, 2006) kernel is defined for $\kappa,\beta > 0$, as

$$\mathscr{W}_{\mu,\kappa,\beta}(r) := \frac{1}{B(2\kappa,\mu+1)} \int_{r/\beta}^1 u(u^2-(r/\beta)^2)^{\kappa-1}(1-u)^\mu \,\mathrm{d}u, \tag{8}$$

when $r/\beta < 1$, and 0 otherwise. with B denoting the beta function. The univariate kernel $\mathscr{W}_{\mu,\kappa,\beta}(r)$ is positive definite in $\mathbb{R}^d$ if and only if $\mu \ge (d+1)/2+\kappa$ (Zastavnyi, 2006). The special case $\kappa = 0$ is known as the Askey family (Askey, 1973)

$$\mathscr{A}_{\mu,\beta}(r) := \begin{cases} (1-r/\beta)^\mu, & 0 \le r/\beta < 1, \\ 0, & r/\beta \ge 1. \end{cases}$$

We define $\mathscr{W}_{\mu,0,\beta} := \mathscr{A}_{\mu,\beta}$.

Algebraically closed form solutions for (8) are available when $\kappa = k$, a positive integer. We have $\mathscr{W}_{\mu,k,\beta}(r) = \mathscr{A}_{\mu+k,\beta}(r)P_k(r)$, with P_k a polynomial of order k. Such a case is termed *original* Wendland functions after the tour de force by Wendland (1995). Then Schaback (2011) proved that other closed form solutions can be obtained when $\kappa = k+0.5$. These were called *missing Wendland* functions. Arguments in Theorem 1(3) of Bevilacqua et al. (2019) prove that the Fourier transform of the generalized Wendland function, $\widehat{\mathscr{W}_{\mu,\kappa,\beta}}$ is not as simple as (7), but has a similar asymptotic order of $z^{-2\lambda}$ for $z \to \infty$ provided $\mu \ge \lambda := (d+1)/2+\kappa$.

Matérn and Generalized Wendland spectral densities can be parameterized in such a way that their tails behave similarly. Specifically, Theorem 1 in Bevilacqua et al. (2019) shows that $2\nu = 2\kappa+1$ is needed. The scale factors α and β in the two models have no influence on the

spectral asymptotics. Yet, they play a key role, in concert with smoothness and variance, to determine conditions for equivalence of Gaussian measures associated to random fields with these classes of covariance functions (Bevilacqua et al., 2019). In turn, such conditions are the crux to provide sufficient condition for best optimal linear prediction under a misspecified covariance function (Stein, 1999).

3.2 D2: Hilbert Spaces

We now come back to general kernels and state their main properties as needed for the paper. The space $\mathscr{H}(C)$ spanned by all *kernel translates* $C(x,\cdot)$, $x\in\Omega$ is a Hilbert space of functions on Ω that has an inner product $(.,.)_{\mathscr{H}(C)}$ with the properties

(9)
$$f(x)=(f,C(x,\cdot))_{\mathscr{H}(C)},\qquad f\in\mathscr{H}(C),\ x\in\Omega,$$
$$C(x,y)=(C(x,\cdot),C(y,\cdot))_{\mathscr{H}(C)},\ x,y\in\Omega,$$
$$\lambda^x\mu^y C(x,y)=(\lambda,\mu)_{\mathscr{H}(C)^*},\qquad \lambda,\mu\in\mathscr{H}(C)^*$$

where $\mathscr{H}(C)^*$ denotes the (topological) dual of $\mathscr{H}(C)$ containing all continuous (i.e. bounded) linear functionals on $\mathscr{H}(C)$ and λ^x means evaluation of λ with respect to the variable x. In particular, all point evaluation functionals $\delta_x : f\mapsto f(x)$ are continuous with norm $\sqrt{C(x,x)}$. Details are in monographs by Abdelaziz and Hamouine (2008), Fasshauer and McCourt (2015), Schaback (1997), Wendland (2004).

For Sobolev spaces, we work on $\mathbb{R}^d$ and use the definition

$$W_2^m(\mathbb{R}^d):=\left\{f:\mathbb{R}^d\to\mathbb{R},\int_{\mathbb{R}^d}|\hat{f}(\omega)|^2(1+\|\omega\|_2)^m d\omega<\infty\right\}$$

as subspaces of $L_2(\mathbb{R}^d)=W_2^0(\mathbb{R}^d)$. They are separable Hilbert spaces under the inner product

$$(f,g)_{W_2^m(\mathbb{R}^d)}=\int_{\mathbb{R}^d}\hat{f}(\omega)\overline{\hat{g}(\omega)}(1+\|\omega\|_2)^m d\omega,$$

for $f,g\in W_2^m(\mathbb{R}^d)$, and there is a reproduction formula

$$f(x)=(f,\mathscr{M}_{m-d/2,1}(\|x-\cdot\|_2)_{W_2^m(\mathbb{R}^d)},$$

for $f\in W_2^m(\mathbb{R}^d), x\in\mathbb{R}^d$, by using a special version of the general isotropic Matérn kernel from (6). This links the most important function spaces of Real Analysis and Partial Differential Equations to the most important covariance functions in Spatial Statistics.

In turn, Theorem 5 in Bevilacqua et al. (2019) ensures that the above reproducing property is guaranteed provided $m-d/2=\kappa+1/2$, that is when $\kappa=m-(d-1)/2$.

Summarizing, each positive definite covariance function C is a kernel that leads to a *native* Hilbert space $\mathscr{H}(C)$ of deterministic functions on Ω, and Sobolev spaces arise from Matérn kernels.

3.3 D3: Orthonormal Expansions

Like in all separable Hilbert spaces, there are orthonormal systems $\{w_n\}_{n\in\mathbb{N}}$ in $\mathscr{H}(C)$ that allow each function $f\in\mathscr{H}(C)$ to be written as a series

$$f=\sum_{n\in\mathbb{N}}(f,w_n)_{\mathscr{H}(C)}w_n$$

with

$$\|f\|^2_{\mathscr{H}(C)}=\sum_{n\in\mathbb{N}}(f,w_n)^2_{\mathscr{H}(C)}$$

and convergence in norm. In particular,

$$C(x,y)=\sum_{n\in\mathbb{N}}w_n(x)w_n(y)$$

holds for all $x,y\in\Omega$, and there is a pointwise bound

$$\sum_{n\in\mathbb{N}}w_n(x)^2=C(x,x)\text{ for all }x\in\Omega \tag{10}$$

for the functions w_n that ensures $w_n(x)\to 0$ for $n\to\infty$ and

$$\sum_{n\in\mathbb{N}}\|w_n\|^2_{L_2(\Omega)}=\int_\Omega C(x,x)dx.$$

We now devote some attention to two special cases of orthonormal expansions.

The standard way of generating path approximations for random fields is by a stepwise Cholesky decomposition of kernel matrices. We show here that it yields a special case of an orthonormal expansion.

Consider a countable dense set X_∞ of points in Ω. Then the Cholesky decomposition procedure for infinite kernel matrices $\mathbf{C}_{X_\infty,X_\infty}$ can be written as

$$C(x_j,x_k)=\sum_{n=1}^{\infty}w_{n,j}w_{n,k},\ 1\le j,k<\infty$$

with an infinite triangular matrix $\mathbf{W}$ with entries $w_{n,j}$ that satisfy $w_{n,j}=0$, $1\le j\le n-1$ and $w_{nn}=:\sigma_n>0$, $n\ge 1$. The standard stepwise Cholesky procedure generates exactly those matrices for $n\to\infty$, and the result is $\mathbf{C}_{X_\infty,X_\infty}=\mathbf{W}^T\mathbf{W}$.

From a more general viewpoint, this is an expansion

$$C(x,y)=\sum_{n=1}^{\infty}w_n(x)w_n(y)\text{ for all }x,y\in\mathbb{R}^d \tag{11}$$

of the kernel itself into functions w_n, $n\ge 1$. The functions w_n have the properties

$$\begin{aligned}w_n(x_j)&=0, && 1\le j<n\\ w_n(x_n)&=:\sigma_n>0,\\ w_n&\in span\{C(x_1,\cdot),\dots,C(x_n,\cdot)\},\ n\ge 1.\end{aligned}$$

In the deterministic literature (Müller and Schaback, 2009, Pazouki and Schaback, 2011) this is the well-known expansion into a *Newton basis*, and it is orthonormal in

$\mathscr{H}(C)$. The recursion formulas in terms of full functions are

$$
\begin{aligned}
w_j^2(x_j) &= C(x_j,x_j) - \sum_{m=1}^{j-1} w_m(x_j^2), \\
w_j(x)w_j(x_j) &= C(x_j,x) - \sum_{m=1}^{j-1} w_m(x_j)w_m(x).
\end{aligned} \tag{12}
$$

Note that expansions (11) replace a stationary kernel by a sum of products, and stationarity of is lost for partial sums.

This ends the deterministic phase. We now have all tools needed for paths (3) and random variables (4).

4. THE STOCHASTIC PHASE: EXPANSION PATHS AND RANDOM FIELDS

After starting from kernels C and providing their deterministic properties, it is now time to introduce randomness. This will turn the kernel into a covariance function for a random field, without using a joint distribution. We shall use the orthonormal expansions of Section 3.3 for this purpose.

Step R1 introduces a zero mean and unit variance distribution, $\mathscr{R}$, from which an independent and identically distributed sequence, $S = \{s_n\}_n$ is drawn. The background is a completely different random field, now on $\mathbb{N}$ and denoted by $\mathbb{N}^{\mathscr{R}}$, with a trivial joint distribution induced by independent $\mathscr{R}$-distributed random variables $S(n)$ for all $n \in \mathbb{N}$. Its paths $S = \{s_n\}_n$ are well-defined due to the independence, and are the sequences S used here.

Step R2 couples the sample $S = \{s_n\}_n$ with an orthonormal expansion W in $\mathscr{H}(C)$ to arrive at (3). There, we called the function $p_{S,W,\mathscr{R}}$ an $(S,W,\mathscr{R})$-*expansion path*.

Finally, step R3 looks locally at (3) on x, to get a random variable $R_{C,W,\mathscr{R}}(x)$ whose realizations are in (4) with realizations S from $\mathbb{N}^{\mathscr{R}}$. Note that these random variables cannot be sampled individually. A sample S from $\mathbb{N}^{\mathscr{R}}$ provides samples for all $R_{C,W,\mathscr{R}}(x)$ simultaneously. As long as we keep C,W, and $\mathscr{R}$ fixed, we shall reduce the notation to $R_C(x)$ for reasons to be apparent below.

4.1 Properties of Paths and Random Variables

Because we first defined paths and then pointwise random variables, both $R_C(x)$ and $R_C(y)$ have to use the same samples S, and then

$$
\begin{aligned}
\operatorname{cov}(R_C(x),R_C(y)) &= \mathbb{E}\left(\sum_{n\in\mathbb{N}} s_n w_n(x)\right)\left(\sum_{m\in\mathbb{N}} s_m w_m(y)\right) \\
&= \sum_{n\in\mathbb{N}} w_n(x)w_n(y) \\
&= C(x,y)
\end{aligned} \tag{13}
$$

for all $x,y \in \Omega$.

The above approach makes it easy to construct non-Gaussian random fields with a prescribed covariance function, C. Users still have the orthonormal system W and the standardized probability distribution $\mathscr{R}$ at their disposal.

It is interesting to study the set of all paths (3) for fixed C, W, and $\mathscr{R}$. By the following theorem, it is independent of W as a set, but different W and $\mathscr{R}$ will prefer certain subsets of paths over others when producing repeated paths.

THEOREM 1. *The set of all possible expansion paths* (3) *is independent of the orthonormal expansion W, if C and $\mathscr{R}$ are fixed.*

PROOF. If there are two orthonormal systems W and U with

$$w_n = \sum_{m\in\mathbb{N}} d_{nm}u_m,\ n \in \mathbb{N},$$

the infinite transformation matrix D is orthogonal, i.e.

$$DD^T = D^T D = Id,$$

letting the series be absolutely convergent. Now path representations $p_{S,W,\mathscr{R}}$ and $p_{T,U,\mathscr{R}}$ can be related by

$$t_m = \sum_{n\in\mathbb{N}} s_n d_{nm},\ m \in \mathbb{N}$$

to show that

$$
\begin{aligned}
p_{S,W,\mathscr{R}} &= \sum_{n\in\mathbb{N}} s_n w_n \\
&= \sum_{n\in\mathbb{N}} s_n \sum_{m\in\mathbb{N}} d_{nm}u_m \\
&= \sum_{m\in\mathbb{N}} u_m \sum_{n\in\mathbb{N}} s_n d_{nm} \\
&= p_{T,U,\mathscr{R}}.
\end{aligned}
$$

□

4.2 Statistical Equivalence

Assume two orthonormal systems $v = \{v_n\}$ and $w = \{w_n\}$ and keep the same univariate distribution $\mathscr{R}$ and the same covariance function C. The two systems could be called *statistically equivalent*, if for all measurable sets $M \subset \Omega \times \mathbb{R}$ of finite Lebesgue measure, the probability of paths based on v and w to have their graphs in M should be the same.

A simplification replaces M by *cylinder sets*

$$\mathscr{C}(X_n,\mathbf{a},\mathbf{b}) := \{f : \Omega \to \mathbb{R} : a_j \le f(x_j) \le b_j,\ 1 \le j \le n\}$$

for $\mathbf{a} < \mathbf{b} \in \mathbb{R}^n$, and this defines a probability

$$\operatorname{Prob}\left(a_j \le \sum_{m\in\mathbb{N}} s_m w_m(x_j) \le b_j,\ 1 \le j \le n\right)$$

that a path based on w hits $\mathscr{C}(X_n,\mathbf{a},\mathbf{b})$. This is

$$\operatorname{Prob}(a_j \le R_C(x_j) \le b_j,\ 1 \le j \le n) \tag{14}$$

and is independent of the expansion whenever the joint distribution of all $R_C(x_j), 1 \le j \le n$ is.

THEOREM 2. *Under Gaussianity, there is statistical equivalence of all expansion paths on all cylinder sets.* □

Probabilities (14) on cylinder sets are well-defined in case of Gaussianity, and the probability stays the same if X_n is contained in a point-based expansion for an infinite dense set X_∞. This why users can consider the standard constructions of Gaussian paths using point sets X_∞ to be reliable and independent of expansions. Cylinder sets arise in Billingsley (1995) and various approaches to abstract Wiener spaces, used to define so-called *cylinder set measures* (Wikipedia contributors, 2024a) which usually are no measures.

4.3 Paths for Point-Based Expansions

If the standard Cholesky decomposition of the kernel matrix with entries $C(x_j,x_k)$, $1 \le j,k \le N$ is executed only for a finite point set X_N, the path approximations are the partial sums of (3)treated in Theorem 6 below, but usually only calculated on X_N. The recursion formula (12) extends the result to be a function on all of Ω, and it can easily be proven that the extension is the Kriging interpolant using the finite path vector on X_N. Theorem 6 will show how far this is from a calculation on Ω itself. The *residual* kernel

$$C_{N+1}(x,y) := C(x,y) - \sum_{n=1}^{N} w_n(x)w_n(y) \tag{15}$$

is the kernel C conditioned to X_N, and $C_{N+1}(x,x)$ is the variance for Kriging on x given values on X_N, because it coincides with the *Power Function* that has this property, (see, *e.g.*, Fasshauer and McCourt, 2015, p. 98). The early paper (Wu and Schaback, 1993) called it *Kriging function*. This is the probabilistic interpretation of such a point-based expansion. The presentation here is in terms of functions, not finite path vectors. The limit $N \to \infty$ causes no problems if points are reasonably selected to become dense in the limit. (Müller and Schaback, 2009).

4.4 Paths for Mercer Expansions

For a Mercer (Mercer, 1909) expansion of the kernel, the basic equations are

$$\int_\Omega C(x,y)v_n(x)dx = \lambda_n v_n(x),\ x \in \Omega \tag{16}$$

with positive λ_n decaying to zero and orthonormality of the v_n in $L_2(\Omega)$ paired with orthogonality $(v_n,v_m)_{\mathscr{H}(C)} = \delta_{nm}\lambda_n^{-1}$, $n,m \ge 1$ in the native space $\mathscr{H}(C)$. Then

$$C(x,y) = \sum_{n=1}^{\infty} \lambda_n v_n(x)v_n(y)$$

is an expansion in $L_2(\Omega)$. We get (11) and orthonormality in $\mathscr{H}(C)$ by setting $w_n = \sqrt{\lambda_n}v_n$. In the probabilistic context, this turns into a Karhunen-Loève expansion. Paths have the form (3), in particular

$$p_{S,W,\mathscr{R}} = \sum_{n\in\mathbb{N}} s_n w_n = \sum_{n\in\mathbb{N}} s_n \sqrt{\lambda_n} v_n$$

with norm

$$\|p_{S,W,\mathscr{R}}\|^2_{L_2(\Omega)} = \sum_{n\in\mathbb{N}} s_n^2 \lambda_n.$$

A Mercer/Karhunen-Loève expansion is not point-based, and therefore it lacks the above probabilistic interpretation. But it is optimal in the sense that the decay of the λ_n cannot be improved given C and the space $\mathscr{H}(C)$, leading to optimal convergence of Mercer/Karhunen-Loève paths (Santin and Schaback, 2015).

If points in a point-based expansion are chosen by the *P-greedy method* (Müller and Schaback, 2009), picking x_{N+1} as

$$x_{N+1} = \arg\max\{C_{N+1}(x,x),\ x \in \Omega\},$$

the convergence of path approximations is asymptotically optimal as well (Santin and Haasdonk, 2017). This is computationally much cheaper than a Mercer/Karhunen-Loève expansion.

5. REGULARITY OF RANDOM FIELDS AFTER THE TURNAROUND

5.1 A tool: the Loève Isometry

By (13), the linear space

$$\mathscr{S}(R_C) = span\{R_C(x)\ :\ x \in \Omega\}. \tag{17}$$

of zero-mean random variables is isometrically isomorphic to the span of kernel translates $C(x,\cdot)$ by the *Loéve* map

$$\mathscr{L}_C(R_C(x)) := C(x,\cdot) \text{ for all } x \in \Omega$$

on the generators. We denote the completion of $\mathscr{S}(R_C)$ by $\mathscr{H}(R_C)$ to distinguish it from its isometric counterpart $\mathscr{H}(C)$. If one does not start from C, and without Gaussianity, the definition of the Loéve map is problematic due to the definition of the covariance via marginalization of the joint distribution of all $R(\cdot)$ to the joint distribution of pairs $(R(x),R(y))$.

We know the Hilbert space completion $\mathscr{H}(C)$ of the translates $C(x,\cdot)$, $x \in \Omega$, and then each deterministic function $f \in \mathscr{H}(C)$ defines a zero-mean second-order random variable $S_f = \mathscr{L}_C^{-1} \in \mathscr{H}(R_C)$. It has the property

$$\text{cov}(S_f,R_C(x)) = f(x),\ x \in \Omega$$

that can be called the *random reproduction formula* that complements the standard deterministic reproduction formula from (9), obtained by applying the Loéve isometry.

But more important are the duals of the two Hilbert spaces. Let $\mathscr{H}(C)^*$ and $\mathscr{H}(R_C)^*$ denote the dual spaces associated with, respectively, $\mathscr{H}(C)$ and $\mathscr{H}(R_C)$. In view of the Loève isometry, we have that functionals on the native space $\mathscr{H}(C)$ should now correspond to functionals on $\mathscr{H}(R_C)$. If we start from $\lambda \in \mathscr{H}(C)^*$, we can define a functional $\lambda^* \in \mathscr{H}(R_C)^*$ on $\mathscr{H}(R_C)$ by

$$\lambda(\mathscr{L}_C(S)) =: \lambda^*(S) \text{ for all } S \in \mathscr{H}(R_C).$$

This is the standard duality map $\mathscr{L}_C^*$ from $\mathscr{H}(C)^*$ to $\mathscr{H}(R_C)^*$. By the Riesz representer theorem, there is a second-order zero-mean random variable S_λ with

$$\mathbb{E}(S_\lambda S) = \lambda(\mathscr{L}_C(S)) =: \lambda^*(S) \text{ for all } S \in \mathscr{S}(R_C),$$

and in particular

$$\begin{aligned} \mathbb{E}(S_\lambda R_C(x)) &= \lambda(C(x,\cdot)) \quad \text{for all } x \in \Omega, \\ \mathbb{E}(S_\lambda S_\mu) &= (\lambda,\mu)_{\mathscr{H}^*(C)} = \lambda^x \mu^y C(x,y) \end{aligned} \tag{18}$$

for all $\lambda, \mu \in \mathscr{H}^*(C)$.Here, the superscript x on a functional λ denotes action with respect to the variable x.

The consequence is that all functionals $\lambda \in \mathscr{H}^*(C)$ on $\mathscr{H}(C)$ lead to valid second-order mean-zero random variables $S_\lambda \in \mathscr{H}(R_C)$. Under Gaussianity, all second-order random variables S_λ of this type will be Gaussian again. And the admissible and bounded functionals on both Hilbert spaces are comparable via the dual of the Loéve map. This argument will be crucial when investigating pointwise regularity notions for both functions in $\mathscr{H}(C)$ and random fields R_C in the next section.

In particular, a classical pointwise derivative is a functional λ on $\mathscr{H}(C)$, and then S_λ is a random variable with the same norm that describes the corresponding derivative of the random field R_C in the mean-square sense. We shall describe this in more detail.

5.2 Pointwise differentiability

Pointwise derivative functionals $\delta_x^\alpha(f) := (D^\alpha f)(x)$ that are admissible in the native space $\mathscr{H}(C)$ are those having a finite norm $\|\delta_x^\alpha\|^2_{\mathscr{H}^*(C)}$ that is defined through

$$\|\delta_x^\alpha\|^2_{\mathscr{H}^*(C)} = \delta_x^{\alpha,u}\delta_x^{\alpha,v}C(u,v) < \infty, \tag{19}$$

where the upper indices u and v indicate the variables the functionals use. This induces derivative kernels

$$C^{\alpha,\alpha}(x,y) \;:\; (x,y) \mapsto \delta_x^{\alpha,u}\delta_y^{\alpha,v}C(u,v) \text{ for all } x,y \in \Omega$$

that are symmetric and positive semidefinite. Boundedness of pointwise derivative functionals is reduced to existence of "twin" derivatives of the kernel.

For each admissible derivative order $\alpha \in \mathbb{N}^d$ there is a second-order zero-mean random variable $S_{\delta_x^\alpha}$ with the variance in (19), and the map $x \mapsto S_{\delta_x^\alpha}$ is a zero-mean second order random field with covariance function $C^{\alpha,\alpha}$. Starting from $x \mapsto S_{\delta_x} = R(x)$ for $\delta_x = \delta_x^{(0)}$ we can approximate all higher pointwise derivatives by linear combinations of point evaluations, and then the random field $x \mapsto S_{\delta_x^\alpha} =: D_x^\alpha(R)$ coincides with the pointwise mean-square derivative of the random field R. In this sense we have

THEOREM 3. *Pointwise mean-square differentiability properties of the random field R_C coincide with pointwise differentiability properties of functions in the native space $\mathscr{H}(R_C)$ of the covariance kernel C.* □

A simple illustration follows. We take $d = 1$ for sake of simplicity. A random variable R'_x is the mean-square derivative of R at x, if

$$\mathbb{E}(R'_x - (R(x+h) - R(x))/h)^2 \to 0 \text{ for } h \to 0.$$

The above approach defines R'_x via $\mathbb{E}(R'_x R(z)) = D_x^{1,u}C(u,z)$, for all z and $E(R'_x R'_x) = D_x^{1,u}D_x^{1,v}C(u,v)$. Then the above expression is

$$\begin{aligned} &\mathbb{E}(R'_x - (R(x+h) - R(x))/h)^2 \\ &= \|\delta_x^1 - \tfrac{1}{h}(\delta_{x+h} - \delta_x)\|^2_{\mathscr{H}(C)^*} \end{aligned}$$

and this converges to zero if and only if $\frac{1}{h}(\delta_{x+h} - \delta_x)$ converges to δ'_x in norm in $\mathscr{H}(C)^*$.

5.3 Continuity

Like continuity of functions, this is not a notion that works with a single point or a single functional. One way to define it for a random field R is to say that R is *pointwise mean-square continuous* at x if

$$\lim_{h \to 0 \in \mathbb{R}^d} \mathbb{E}(R(x+h) - R(x))^2 = 0, \tag{20}$$

and because this is (Stein, 1999)

$$\lim_{h \to 0 \in \mathbb{R}^d} (C(x+h,x+h) + C(x.x) - 2C(x+h,x)) = 0, \tag{21}$$

it is satisfied if C is continuous. When combined with the previous section, we get

THEOREM 4. *Assume a kernel C has continuous derivatives $C^{\alpha,\alpha}(x,y) = D_x^{\alpha,u}D_y^{\alpha,v}C(u,v)$. Then the second-order zero mean random field $D^\alpha R_C$ is pointwise mean-square continuous almost surely and has the above covariance function.* □

Again, the slight difference between regularity of random fields and functions from their their native spaces lies in the difference of the two regularity notions, here for continuity.

6. POINTWISE CONVERGENCE OF PATHS

We now consider pointwise regularity of paths, using our explicit representation (3). We have by definition

THEOREM 5. *The value of expansion paths $p_{S,W,\mathscr{R}}$ at a point x is a random variable $R_C(x)$ that has mean zero and variance $\sum_{n=1}^{\infty} w_n(x)^2 = C(x,x)$. It is finite almost surely.*

The proof uses the fact that series of second-order zero-mean random variables converge almost surely, if the variances are summable (Durrett, 2019, Theorem 2.5.6, page 84).

A simple argument allows us to control the error that is committed when taking only partial sums in (3) up to some N.

THEOREM 6. *Path approximations $p_{N,S,W,\mathscr{R}}$ by truncated sums*

$$p_{N,S,W,\mathscr{R}}(x) = \sum_{n=1}^{N} w_n(x) s_n$$

always lie in the native Hilbert space generated by the covariance kernel, and the error $p_{S,W,\mathscr{R}}(x) - p_{N,S,W,\mathscr{R}}(x)$ is a random variable with variance

$$C(x,x) - \sum_{n=1}^{N} w_n(x)^2 = \sum_{n=N+1}^{\infty} w_n(x)^2 \tag{22}$$

for all $x \in \Omega$. Going from N to $N+1$ decreases the variance by $w_{N+1}(x)^2$ for all x. □

In a nutshell, this is what allows to go over to infinite point sets in the Gaussian case to get paths on all of Ω. Limits of point sets as a bottom-up or local-to-global process are replaced here by series truncations as a top-down or global-to-local process. Nothing is lost, because the standard finite-to-infinite path construction for Gaussian processes goes via an orthonormal system anyway, by Section 3.3.

The random variable $D^{\alpha}(p_{S,W,\mathscr{R}})(x)$ on expansion paths $p_{S,W,\mathscr{R}}$ from (3) has the representation

$$D^{\alpha}(p_{S,W,\mathscr{R}})(x) = \sum_{n=1}^{\infty} D^{\alpha}(w_n)(x) s_n.$$

It is a random variable with variance zero and variance

$$\sum_{n=1}^{\infty} (D^{\alpha}(w_n)(x))^2 = D^{\alpha,\alpha}C(x,x) = \|\delta^{\alpha}\|^2_{\mathscr{H}(C)^*} \text{ for all } x \in \Omega.$$

Therefore Theorem 3 on almost sure mean-square pointwise differentiability also holds for pointwise values of expansion paths, in the sense that pointwise derivatives have finite values almost surely.

Similarly, we examine pointwise mean-square continuity via (20) and get the condition

$$\lim_{h \to 0 \in \mathbb{R}^d} \mathbb{E}\left(\sum_{n=1}^{\infty} s_n (w_n(x+h) - w_n(x)) \right)^2 = 0,$$

which coincides with (21). Then Theorem 4 on almost sure pointwise mean-square continuity extends to expansion paths.

THEOREM 7. *Pointwise regularity and smoothness notions for paths of a random field and functions in the native space of its covariance function coincide. The former are to be understood as random variables with bounded variance.* □

7. NORM CONVERGENCE OF PATHS

We recall Section 4 for $(S,W,\mathscr{R})$-expansion paths (3). Once sampled, these are functions like any other deterministic function, and we can investigate their regularity with deterministic techniques.

Each partial sum of (3) is a function in the native space whose squared norm is the sum of squares of the coefficients. For $n \to \infty$, the problem requires instruments from probability theory. When S in (3) is an independent sequence, then $S^2 := \{S_n^2\}_n$ is an independent sequence as well. Hence, we can invoke the celebrated Kolmogorov's *Three series theorem*, for which the series $\sum_n s_n^2$ converges if and only if

$$\text{Prob}\left(S_n^2 > K\right) < \infty \qquad \text{and} \qquad \sum_n \mathbb{E} S_n^2 < \infty,$$

where the third condition is not needed in our case because we are working with zero-mean random variables. This provides an immediate implication for the case of IID (independently and identically distributed) sequences, which is largely used in machine learning as explained by Schölkopf (2022). Although the implication is straightforward, we formalize it below for the convenience of the reader.

THEOREM 8. *Let the sequence S in (3) be* IID. *Then, expansion paths (3) do not lie, almost surely, in the native Hilbert space of the covariance kernel.*

Some commments are in order. The assumption of ID can be relaxed to provide situations where the expansion paths (3) lie, almost surely, in their Native space. Yet, these are more mathematical artifacts rather than real situations. For example, one might set $S_n^2 \sim \mathscr{N}(0, 1/n^2)$ and retrieve S_n by backwards transformation. Yet, this has no sense for practical applications. Hence, the general message is that expansion paths based on independent sequences fall almost surely outside their original Native space.

We note that the assumption of independence as per Kolmogorov can be relaxed at the price of very technical conditions (Brown, 1971). Yet, this case does not apply to our context.

Note this serious difference between paths of random fields and functions from their native spaces. Furthermore, we now consider the random norm of a random function, not random values at certain single points. This is the difference to Section 5.

By Theorem 8, native space norm convergence in (3) will fail in general, but there may be weaker norms like

$L_2(\Omega)$ that admit convergence of paths in norm. We postpone this to the next section.

By Theorem 8 we know that IID paths do not lie in the native space almost surely, but are they almost surely in larger spaces? The answer requires a *scale* of spaces with different regularity properties. The easiest and more general approach to scaled spaces is via weighted expansions, and we use it for the scale of Sobolev spaces later.

To stay close to Fourier series and Mercer/Karhunen-Loève expansions, we build a scale of Sobolev-type functions over $L_2(\Omega)$ using a fixed basis of orthonormal functions v_n in $L_2(\Omega)$. We penalize the expansion coefficients by positive sequences $\boldsymbol{\rho} = \{\rho_n\}_n$ and define inner products

$$(f,g)_{L_{2,\boldsymbol{\rho}}(\Omega)} = \sum_{n=1}^{\infty} (f,v_n)_{L_2(\Omega)} (g,v_n)_{L_2(\Omega)} \rho_n^{-1}.$$

We mimic the Mercer/Karhunen-Loève case by assuming $w_n = \sqrt{\lambda_n} v_n$ being orthonormal in $\mathscr{H}(C)$ for a covariance function C on Ω, see Section 3.3 for details. By comparison of expansions, this implies $(f,v_n)_{L_2(\Omega)} = \sqrt{\lambda_n}(f,w_n)_{\mathscr{H}(C)}$, and the native Hilbert space for C then consists of all functions $f \in L_2(\Omega)$ with

$$\sum_{n=1}^{\infty} (f,w_n)^2_{\mathscr{H}(\Omega)} = \sum_{n=1}^{\infty} \lambda_n^{-1} (f,v_n)^2_{L_2(\Omega)} < \infty.$$

i.e. it is $L_{2,\boldsymbol{\lambda}}(\Omega)$. If p is a path in that space, it is in $L_{2,\boldsymbol{\rho}}(\Omega)$ iff

$$\sum_{n=1}^{\infty} (f,v_n)^2_{L_2(\Omega)} \rho_n^{-1} = \sum_{n=1}^{\infty} \omega_n^2 \lambda_n \rho_n^{-1} < \infty.$$

THEOREM 9. *For a scale of spaces based on weighted expansions in $L_2(\Omega)$ under the above assumptions, the native space $L_{2,\boldsymbol{\lambda}}(\Omega)$ lies in $L_{2,\boldsymbol{\rho}}$ if $\sum_{n=1}^{\infty} \lambda_n \rho_n^{-1} < \infty$.* □

We now connect this to the scale of Sobolev spaces $W_2^m(\Omega)$ on compact domains $\Omega \subset \mathbb{R}^d$. The variances λ_n^2 for Mercer/Karhunen-Loève expansions (16) behave like $n^{-2m/d}$ (Santin and Schaback, 2016) for $n \to \infty$. To simplify notation, we define an ID path the expansion path (3) for which S is additionally ID. Hence, we have

THEOREM 10. *If a covariance function generates the native space $W_2^m(\mathbb{R}^d)$, and if all other Sobolev spaces are defined via scaling, its ID paths lie almost surely in all Sobolev spaces $W_2^p(\mathbb{R}^d)$ for $p < m - d/2$, but almost surely not in $W_2^{m-d/2}(\mathbb{R}^d)$.* □

On the Sobolev scale, there is a *smoothness gap* of order $d/2$ between functions in the native space and paths. This goes back to Scheuerer (2009), but we provided an explicit and constructive proof for expansion paths, based on (3) and a revised definition of paths, but limited to Mercer/Karhunen-Loève expansions. Roughly, if $W_2^m(\mathbb{R}^d)$ is the native space for a covariance kernel, the borderline space for paths is $W_2^{m-d/2}(\mathbb{R}^d)$ in the above sense.

Expansion paths via infinite point sets do not work in the proof, because we used that the v_n are independent of the smoothness order.

It should be explained why Theorems 5, 8 and 10 are not contradictory. Imagine a random function generator that produces large numbers of random paths. When considering only a single point x, one gets a random variable over these paths that has bounded variance. But something like the $L_2(\Omega)$ norm of single paths is another random variable over paths that needs a proof for guaranteeing bounded variance.

Theorem 10 has a positive computational aspect. If users want to produce cheap random sample paths from $W_2^m(\Omega)$ without any excessive smoothness, they should not run anything for $n \to \infty$ using that kernel. For roughly the same effect, one can take cases for finite n in $W_2^{m-d/2}(\Omega)$.

The paper (Steinwart, 2019) draws much more detailed conclusions for expansion paths (3) with respect to various function spaces, but it still starts in the standard way from Random Fields.

Remarks

The regularity differences between deterministic interpolation, non-deterministic Kriging, and path construction for random fields should be explained in some more detail, illustrating the above results.

1. **Deterministic interpolation**
   There is an apparent philosophical discrepancy between the statistics and the numerical analysis approach. For the former, data are realizations from something *random*, while for the latter data are the ground truth modulo some additive noise. We start by focusing on the latter: in this case, data are the true values of a function f from the native space of C. For finitely many observations from f, the smoothness of the (Kriging) interpolant is *double* with respect to the smoothness of functions in the native space, because the interpolant is generated from kernel translates. These have the smoothness of the kernel C, but native space functions only have the smoothness of the kernel C_0 with $C = C_0 * C_0$, i.e., $\hat{C} = (\hat{C}_0)^2$. When the number of observations of f tends to infinity, the double smoothness is lost by going to the limit in the native space. This well-known *smoothness gap* is deterministic and arises when going from translates of C to their Hilbert space limits in $\mathscr{H}(C)$.

2. **Kriging**
   The variance of BLUP (Kriging) prediction at some target point x is the square of the Power Function, and the latter is the kernel conditioned to the data locations. So far, this is only point-dependent, not data-dependent. The Kriging functions, when extended to all x, using data of some f at the nodes, coincide with deterministic interpolation. Yet, the assumption that *ground truth* means that data come from a function in the native space is a questionable addition. The "escape scenario" (Karvonen, 2023) results of Numerical Analysis show that if data come from a function in a larger Sobolev space, the Kriging solution is still convergent to the data function, albeit in the weaker Sobolev norm (Narcowich et al., 2005, 2006). This is often called "misspecification" (Stein, 1999), because the C model is different from the f model. For a finite number of observations, the Kriging solution has the smoothness of C-translates. It is roughly twice the $\mathscr{H}(C)$ smoothness, as pointed out above.
3. **Paths of Random Fields**
   Here, there is no given function f and no given values at points. Using, *e.g.*, the Cholesky decomposition of the kernel matrix and a partial sum of (3), one can generate n random data values at each point x of the domain. By (22), the variance for random selection of a new value will get smaller when increasing n. After n steps, one has n random data values that will be the final path values. If these values are taken as "ground truth" like Interpolation or Kriging, they have a Kriging interpolant that lies in the native space and even has the excess smoothness of kernel translates. But there is no f that supplied them, and the "noise-generating" process of adding new random sample values, though using smaller variances from step to step, induces more smoothness loss than the deterministic loss described above. It is exactly this additional smoothness loss by randomness that forms the "gap" described by Theorem 10.
4. Summarizing, it can be seen as a miracle that the randomness of the new samples is limited by C in such a way (namely by (22)) that one safely arrives in larger Sobolev spaces instead of at what Numerical Analysts would call noise, i.e. a function with hardly any regularity. Yet, the *exact* Sobolev limit space cannot be reached, by Theorem 10.

8. DISCUSSION AND OPEN PROBLEMS

This section serves both as a rejoinder as well as an introduction to open problems. We sketch them below.

- Turning the standard approach to paths upside down avoids all problems with joint probability spaces on infinite sets, and it is closer to *Nature* because it goes from global to local.
- We consider the approach proposed in this paper more transparent with respect to earlier literature: the random variables $R_C(x)$ defined via C are just the values of the paths at points x. This is possible because paths are defined before random variables are defined.
- It simplifies the analysis of paths by the explicit form (3) and the detour via the Hilbert space $\mathscr{H}(C)$.
- Since the bottom-up path construction in the Gaussian case is a special case, nothing is lost in that case.
- There is a substantial novelty in that now other choices for introducing randomness after picking a covariance kernel are possible.
- The standard sampling algorithms based on infinite dense point sets are fast if C is smooth, because the residual variance $C_{N+1}(x,x)$ from (15) decays quickly, leading to a fast convergence in $L_2(\Omega)$ of partial sums (3) of paths. These are in the native space, i.e. somewhat too smooth, unless a kernel for a larger native space is used.
  Therefore the literature has various techniques with better numerical complexity for low smoothness, A widely used method based on (Lindgren et al., 2011) uses weak solutions of stochastic differential equations. Another strategy, starting from Vecchia (1988) exploits that inverses of kernel matrices have good approximations by sparse matrices. These techniques calculate approximations to paths, but not paths themselves, and it should be checked if our approach to paths makes the error analysis easier.
- It needs further work to study the statistical differences of path calculations with different univariate probability distributions $\mathscr{R}$. Similarly, the aforementioned methods for calculating path approximations should be compared.
- Since this paper proceeds via the Hilbert space $\mathscr{H}(C)$, an extension to the vector-valued case should be possible using the approach to paths used here.

REFERENCES

Abdelaziz, Y. and Hamouine, A. (2008). A survey of the extended finite element. *Computers and Structures*, 86:1141–1151.

Abramowitz, M. and Stegun, I. (1970). *Handbook of Mathematical Functions*. Dover, New York.

Aronszajn, N. (1950). Theory of reproducing kernels. *Trans. Amer. Math. Soc.*, 68:337–404.

Askey, R. (1973). Radial characteristic functions. *Technical report, Research Center, University of Wisconsin*.

Bevilacqua, M., Faouzi, T., Furrer, R., and Porcu, E. (2019). Estimation and prediction using generalized wendland covariance functions under fixed domain asymptotics. *The Annals of Statistics*, 47(2):828–856.

Billingsley, P. (1995). *Probability and Measure*. Wiley Series in Probability and Statistics. Wiley.

Brown, B. (1971). A general three-series theorem. *Proceedings of the American Mathematical Society*, 28(2):573–577.

Buhmann, M. (2003). *Radial Basis Functions, Theory and Implementations*. Cambridge University Press, Cambridge,UK.

Christakos, G. (2013). *Random field models in earth sciences*. Elsevier.

Daley, D. J. and Porcu, E. (2014). Dimension walks and Schoenberg spectral measures. *Proc. Amer. Math. Soc.*, 142(5):1813–1824.

Durrett, R. (2019). *Probability: Theory and Examples*. Cambridge University Press.

Fasshauer, G. and McCourt, M. (2015). *Kernel-based Approximation Methods using MATLAB*, volume 19 of *Interdisciplinary Mathematical Sciences*. World Scientific, Singapore.

Gneiting, T. (2002). Stationary covariance functions for space-time data. *Journal of the American Statistical Association*, 97:590–600.

Karvonen, T. (2023). Asymptotic bounds for smoothness parameter estimates in gaussian process interpolation. *SIAM/ASA Journal on Uncertainty Quantification*, 11(4):1225–1257.

Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B*, 73:423–498.

Mercer, J. (1909). Functions of positive and negative type and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society A*, 209:441–458.

Müller, S. and Schaback, R. (2009). A Newton basis for kernel spaces. *Journal of Approximation Theory*, 161:645–655.

Narcowich, F., Ward, J., and Wendland, H. (2005). Sobolev bounds on functions with scattered zeros, with applications to radial basis function surface fitting. *Mathematics of Computation*, 74:743–763.

Narcowich, F., Ward, J., and Wendland, H. (2006). Sobolev error estimates and a Bernstein inequality for scattered data interpolation via radial basis functions. *Constructive Approximation*, 24:175–186.

Oxtoby, J. C. (1946). Invariant measures in groups which are not locally compact. *Trans. Amer. Math. Soc.*, 60:215–237.

Pazouki, M. and Schaback, R. (2011). Bases for kernel-based spaces. *Computational and Applied Mathematics*, 236:575–588.

Porcu, E., Bevilacqua, M., Schaback, R., and Oates, C. J. (2023). The mat\'ern model: A journey through statistics, numerical analysis and machine learning. *arXiv preprint arXiv:2303.02759*.

Santin, G. and Haasdonk, B. (2017). Convergence rate of the data-independent P-greedy algorithm in kernel-based approximation. *Dolomites Res. Notes Approx.*, 10(Special Issue):68–78.

Santin, G. and Schaback, R. (2015). Approximation of eigenfunctions in kernel-based spaces. in revision.

Santin, G. and Schaback, R. (2016). Approximation of eigenfunctions in kernel-based spaces. *Adv. Comput. Math.*, 42(4):973–993.

Schaback, R. (1997). Reconstruction of multivariate functions from scattered data. Manuscript, available via http://webvm.num.math.uni-goettingen.de/schaback/teaching/rbfbook_2.pdf.

Schaback, R. (2011). The missing Wendland functions. *Advances in Computational Mathematics*, 34(1):67–81.

Schaback, R. and Wendland, H. (2006). Kernel techniques: from machine learning to meshless methods. *Acta Numerica*, 15:543–639.

Scheuerer, M. (2009). *A Comparison of Models and Methods for Spatial Interpolation in Statistics and Numerical Analysis*. PhD thesis, Georg-August-Universität Göttingen. http://dx.doi.org/10.53846/goediss-2461.

Schoenberg, I. J. (1938). Metric spaces and positive definite functions. *Trans. Amer. Math. Soc.*, 44:522–536.

Schölkopf, B. (2022). Causality for machine learning. In *Probabilistic and causal inference: The works of Judea Pearl*, pages 765–804.

Schölkopf, B. and Smola, A. (2002). *Learning with Kernels*. MIT Press, Cambridge.

Seeger, M. (2004). Gaussian processes for machine learning. *International journal of neural systems*, 14(02):69–106.

Stein, M. (1999). *Statistical Interpolation of Spatial Data: Some Theory for Kriging*. Springer Series in Statistics. Springer.

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

Stroock, D. (2023). *Gaussian Measures in Finite and Infinite Dimensions*. Universitext. Springer International Publishing.

Sutton, C., McCallum, A., et al. (2012). An introduction to conditional random fields. *Foundations and Trends® in Machine Learning*, 4(4):267–373.

Vecchia, A. V. (1988). Estimation and Model Identification for Continuous Spatial Processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 297–312.

Wendland, H. (1995). Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Advances in Computational Mathematics*, 4:389–396.

Wendland, H. (2004). *Scattered data approximation*, volume 17. Cambridge university press.

Wikipedia contributors (2024a). Cylinder set measure — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Cylinder_set_measure. [Online; accessed 23-September-2024].

Wikipedia contributors (2024b). Stochastic process — Wikipedia, the free encyclopedia. [Online; accessed 23-September-2024].

Williams, C. and Rasmussen, C. (1995). Gaussian processes for regression. *Advances in neural information processing systems*, 8.

Williams, C. K. and Rasmussen, C. E. (2006). *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA.

Wu, Z. M. and Schaback, R. (1993). Local error estimates for radial basis function interpolation of scattered data. *IMA J. Numer. Anal.*, 13(1):13–27.

Zastavnyi, V. P. (2006). On some properties of Buhmann functions. *Ukrainian Mathematical Journal*, 58(08):1045–1067.