Sobolev regularity of Gaussian random fields

Iain Henderson[a]

[a]Institut de Mathématiques de Toulouse; UMR 5219, Université de Toulouse; CNRS , INSA, F-31077 Toulouse, France, henderso@insa-toulouse.fr

November 25, 2023

Abstract

In this article, we fully characterize the measurable Gaussian processes $(U(x))_{x\in\mathcal{D}}$ whose sample paths lie in the Sobolev space of integer order $W^{m,p}(\mathcal{D})$, $m \in \mathbb{N}_0$, $1 < p < +\infty$, where $\mathcal{D}$ is an arbitrary open set of $\mathbb{R}^d$. The result is phrased in terms of a form of Sobolev regularity of the covariance function on the diagonal. This is then linked to the existence of suitable Mercer or otherwise nuclear decompositions of the integral operators associated to the covariance function and its cross-derivatives. In the Hilbert case $p = 2$, additional links are made w.r.t. the Mercer decompositions of the said integral operators, their trace and the imbedding of the RKHS in $W^{m,2}(\mathcal{D})$. We provide simple examples and partially recover recent results pertaining to the Sobolev regularity of Gaussian processes.

1 Introduction

Sobolev spaces $W^{m,p}(\mathcal{D})$ are central tools in modern mathematics, most notably in the study of partial differential equations (PDEs). These spaces are built upon the notion of weak derivative: v is the weak derivative of u in the direction x_i if for all smooth compactly supported function $\varphi \in C_c^\infty(\mathcal{D})$,

$$\int_{\mathcal{D}} u(x)\frac{\partial \varphi}{\partial x_i}(x)dx = -\int_{\mathcal{D}} v(x)\varphi(x)dx. \tag{1.1}$$

Weak derivatives generalize classical, pointwise defined derivatives. In particular, there are cases where weak derivatives are well defined and pointwise differentiation otherwise fails (see e.g. [23], Examples 3 and 4 p. 260). The popularity of Sobolev spaces is justified by a number of reasons: first, they are separable reflexive Banach spaces when $1 < p < +\infty$, and separable Hilbert spaces when $p = 2$ ([1], Theorem 3.6 p. 61). Through duality, this allows for geometrical interpretations of PDEs which in turn lead to numerous quantitative theoretical results in the study of PDEs [23]. Second, as the Sobolev norm is defined through integrals of powers of the function and its weak derivatives, it is easily interpreted as an energy functional of the said function, which complies with physical interpretations of PDEs. This is a desirable feature as PDEs are generally used for describing physical phenomena. Finally, Sobolev spaces are useful for practical purposes as they are the natural mathematical framework for the celebrated finite element method when seeking numerical solutions to PDEs ([8], Chapter 1).

When a function of interest $u : \mathcal{D} \to \mathbb{R}$ is unknown, it may be relevant to model it as a sample path of a random field $(U(x))_{x\in\mathcal{D}}$, say a Gaussian process, whose realizations lie in a suitable function space. This is e.g. frequent in Bayesian inference of functions [50]. Such suitable spaces can indeed happen to be Sobolev spaces, for example when u describes a physical quantity. The question at hand in this article is thus the following: when do the sample paths of a given Gaussian process lie in some Sobolev space? This question is closely linked to the recent attention that Gaussian processes have drawn for tackling machine learning problems arising from PDE models; see e.g. [30, 35, 38, 51]. Notably (see [11]), Gaussian processes seem to provide a numerically competitive and mathematically tractable alternative to the now widespread "physics informed neural networks" (PINNs, [37]). For the moment though, the machine learning techniques involving Gaussian processes have only been studied within the framework of spaces of functions with classical smoothness : C^0, C^1, etc. As argued before, these spaces are often not as well-suited for studying PDEs as Sobolev spaces.

Relevant literature Though weak differentiability is more general, it is less direct to check than classical differentiability. Weak derivatives are defined implicitly and in the most general case, ensuring Sobolev regularity is not usually done by directly verifying that an integral or a series is finite, as would be the case in L^p spaces; variational or boundedness criteria are used instead (see Lemma 2.1).

In many important cases however, handy characterizations of such regularity do exist, which have effectively been used to bypass the implicit definition of Sobolev regularity and provide results on the sample path regularity of Gaussian processes. When $\mathcal{D} = \mathbb{R}^d$, the space $W^{m,p}(\mathbb{R}^d)$ can be characterized in terms of a sufficient decay of of the Fourier transform ([44], Theorem 3 p. 135; [23], Section 5.8.5; [1], Section 7.63). Still in the case $\mathcal{D} = \mathbb{R}^d$, Sobolev regularity is equivalent to the convergence of its de la Vallée Poussin expansion in a suitable space ([34], Section 8.9). This fact has been the first to be employed for characterizing the Sobolev regularity of stationary Gaussian processes indexed by the unit cube of $\mathbb{R}^d$ in [15, 25], in terms of the spectral measure of its covariance. For some Banach spaces, explicit Schauder bases are known and lying in such spaces can be translated as the convergence of some coordinate series. This has been exploited in [14] for studying the Besov and Besov-Orlicz regularity of one dimensional Gaussian processes (they are natural generalizations of Sobolev regularity, [1]), and the fractional Brownian motion in particular. Note that in this one dimensional framework, those spaces only contain continuous functions ([14], Lemme III.3); a fact which, as we will see, can be quite restrictive. Wavelet analysis is also available for describing Sobolev regularity ([1], Section 7.70) and has been used for studying the smoothness of the Brownian motion [13, 41]. More complex notions such as the existence of an underlying Dirichlet structure have been put to use in [28]. The latter work deals with Besov $B^s_{\infty,\infty}$ regularity, $s > 0$, on compact metric spaces, and relies on a convergence analysis of suitable spectral coefficients, based on the so called Littlewood-Paley decomposition. Note that when $0 < s < 1$ and $\mathcal{D}$ is sufficiently regular, $B^s_{\infty,\infty}(\mathcal{D})$ exactly corresponds to the space of Hölder continuous functions $C^s(\mathcal{D})$ (e.g. [22], p. 26). In [45], Karhunen-Loève expansions are used to study whether or not the sample paths of a general second order random process lie in interpolation spaces between the reproducing kernel Hilbert space (RKHS, Section 4.1 below) of the process and $L^2(\nu)$, where ν is a σ-finite measure. This is then applied to study the $H^s(\mathcal{D}) := W^{s,2}(\mathcal{D})$ regularity of the corresponding sample paths when $s > d/2$ (Corollaries 4.5 and 5.7 in [45]), with applications to Gaussian processes in particular (H^s is a fractional Sobolev space). Note that RKHS are also popular function spaces in the machine learning community [5]. Using the notion of mean square derivatives, [42] shows that the sample paths of a general second order random field lie in $H^m(\mathcal{D})$ under an integrability condition of the symmetric cross derivatives of the kernel over the diagonal ([42], Theorem 1). For the suitable definition and use of the mean square derivatives of the process, [42] additionally requires that the covariance function be continuous over the diagonal as well as its symmetric cross derivatives.

A first study of the H^m regularity of Gaussian processes To make things more explicit, let us apply some of the results described above on two examples, namely centered Gaussian processes $(U(x))_{x\in\mathcal{D}}$ whose covariance function (or "kernel") is either Matérn ([39], pp. 84-85) or finite rank. This will allow us to identify situations in which those previous results may be extended.

Gaussian processes with Matérn kernels are widely used in machine learning for approximating finitely smooth functions, therefore it is quite natural to study this particular case. In the case where the domain $\mathcal{D}$ is a bounded open set whose boundary verifies the strong local Lipschitz condition, it is in fact known that the RKHS of Matérn kernels of real order $\nu > 0$ are exactly $H^{\nu+d/2}(\mathcal{D})$. In this case, Corollary 4.5 from [45] is optimal (see [45], Example 4.8 and Theorem 4.4): the sample paths of the associated Gaussian process lie in the Besov space $B^s_{2,2}(\mathcal{D})$ for all $s < \nu$ and not in $B^s_{2,2}(\mathcal{D})$ for all $s \geqslant \nu$. For such domains, it is also true that $B^s_{2,2}(\mathcal{D}) = H^s(\mathcal{D})$. As stated in [45], p. 370, this result suggests that the sample paths are about $d/2$ less smooth than the functions in the RKHS. A limitation of the result described above is the regularity assumption on $\mathcal{D}$, which is necessary for [45] as its results rely on the existence of suitable extension operators to assert that $B^s_{2,2}(\mathcal{D}) = H^s(\mathcal{D})$ (see [1], p. 230). Without such regularity assumptions, this equality does not hold anymore (although generalizations to less regular open sets exist, [17], Theorem 6.7). Hence the seemingly simple case of Sobolev spaces of integer order defined on arbitrary open sets is left undealt with.

In this regard, it is instructive to investigate the consequences of [42], Theorem 1, which does not make any regularity assumptions over the open set $\mathcal{D}$. Its statement is as follow, given a centered random field $(U(x))_{x\in\mathcal{D}}$ with continuous covariance function k (we refer to Section 2.1.2 for notations). If for all $|\alpha| \leqslant m$, the weak derivative $\partial^{\alpha,\alpha}k$ exists, is continuous on the diagonal of $\mathcal{D}\times\mathcal{D}$ and $\int_{\mathcal{D}} \partial^{\alpha,\alpha}k(x,x)dx < +\infty$, then the sample paths of U lie in $H^m(\mathcal{D})$ almost surely. Matérn kernels are stationary, meaning that they are of the form $k(x,x') = k_S(x-x')$ for some real valued function k_S. For such kernels, the criterion from Theorem 1 in [42] essentially reduces to the condition that the pointwise derivatives $(\partial^{2\alpha}k_S)(0)$ exists for all $|\alpha| \leqslant m$, as well as $\mathcal{D}$ being bounded. In fact, when $m = 0$, we are awkwardly left with the condition that k be continuous and $\mathcal{D}$ be bounded. In hindsight, this criterion is not surprising, as the sample paths of a stationary process "look similar at all locations" ([39], p. 4), and thus expecting them to be square integrable over some unbounded domain is not reasonable. This example suggests that focusing on stationary Gaussian processes somewhat conceals the real nature of accurate Sobolev regularity criteria for the sample paths of a Gaussian process. We thus turn to non stationary kernels.

Perhaps the simplest non stationary Gaussian processes are those with finite rank covariance functions, i.e. processes of the form $U(x)(\omega) = \sum_{i=1}^n \xi_i(\omega)f_i(x)$, where (ξ_i) are standard Gaussian random variables (which we assume independent) and (f_i) are measurable functions. The covariance function of the latter is $k(x,x') = \sum_{i=1}^n f_i(x)f_i(x')$. In this case, it is rather clear that the property that $\mathbb{P}(U \in H^m(\mathcal{D})) = 1$ is equivalent to having $(f_i) \subset H^m(\mathcal{D})$ (see Example 3.3 for a rigorous proof of this statement). It is also true that Sobolev functions may happen to be discontinuous (e.g. they may have local singularities, [23], p. 22 and Example 4 p. 260). Hence, the continuity assumptions over k as well as its derivatives required in [42] seem unnecessarily restrictive with reference to the criterion "$(f_i) \subset H^m(\mathcal{D})$". This also prevents the criterion from [42] to be also necessary. Concerning the results described in [45], the RKHS of k is now equal to $\text{Span}(f_1, ..., f_n)$, which is only embedded in $H^m(\mathcal{D})$ if $(f_i) \subset H^m(\mathcal{D})$. Corollary 4.5 from [45] now only states that the sample paths lie in $H^{m-d/2-\varepsilon}(\mathcal{D})$ for all $\varepsilon > 0$, which is clearly suboptimal (it also requires that $m > d/2$ to be non trivial). In fact, for this process, the sample paths have the *same* regularity as the functions in the RKHS. This shows that the rule of thumb according to which the sample paths are about $d/2$ less smooth than functions in the RKHS (a fact which is tight for Matérn processes) can be quite misleading. It is better understood as a lower bound on the regularity of the realizations of a Gaussian process.

Concerning our previous study of Matérn kernels however, it is noteworthy that the imbedding of $H^s(\mathcal{D})$ in $H^t(\mathcal{D})$ is Hilbert-Schmidt precisely when $s - t > d/2$ (recall that $s - t > d/2$ is the correct criterion for Matérn kernels of order $\nu = t + d/2$; see Example 4.5 below for more details on such embeddings). Likewise, the integrals from [42] exactly correspond to traces of very specific integral operators (see Section 2.1.3 for operator theoretic definitions). These observations suggest the existence of a purely spectral criterion for characterizing the $H^m(\mathcal{D})$ regularity of Gaussian processes, which could encompass the results from both [42] and [45], as well as finite rank Gaussian processes. In fact, we provide such a spectral criterion in Proposition 4.4.

As a final comment, none of the articles previously mentioned except [14,28] deal with spaces of non Hilbert type. General Sobolev spaces $W^{m,p}(\mathcal{D})$, with $1 < p < +\infty$, are particularly useful when studying nonlinear PDEs, e.g. whose nonlinearity is of the form $|u|^r u$ (see [47], Section 3.9 for examples). As such, we will also focus on this general setting.

General assumptions The purpose of this article is to uncover necessary and sufficient characterizations of the Sobolev regularity of nonnegative integer order of a given Gaussian process, in terms of its covariance function. In an attempt to make them as general as possible, we set the following targets and assumptions.

(i) The covariance function of the Gaussian process will only be assumed measurable, as in [45]. This contrasts with some of the previously mentioned works [14, 28, 42], where the covariance function is assumed continuous. As previously observed though, it seems that assuming the continuity of the covariance (and thus more or less that of the sample paths, [3] p. 31) to examine some Sobolev regularity of potentially low order is an unnatural assumption. This is especially true as the dimension of $\mathcal{D}$ increases, since $W^{m,p}(\mathcal{D})$ is embedded in $C_B^0(\mathcal{D})$, the Banach space of continuous and bounded functions over $\mathcal{D}$, only when $m > d/p$ ([1], Theorems 4.12 and 7.34).

(ii) We will not make any regularity or shape assumptions on the open set $\mathcal{D}$. Indeed, Sobolev spaces of integer order are easily defined over arbitrary open sets $\mathcal{D} \subset \mathbb{R}^d$, and thus some results should hold within this general setting. As a result though, we will not deal with fractional Sobolev spaces nor Besov spaces. Indeed, those spaces may have some pathological properties without additional hypotheses on $\mathcal{D}$, namely enjoying a Lipschitz boundary or the cone condition (see e.g. [18], Example 9.1). We will see that elementary characterizations of Sobolev regularity (Lemmas 2.1 and 2.4) will prove to be enough for our purpose.

(iii) Our results should lie outside of the assumption that $m > d/p$, where m, p and d correspond to the notation $W^{m,p}(\mathcal{D})$, $\mathcal{D} \subset \mathbb{R}^d$. Indeed, several previous results concerning the Sobolev regularity of a given Gaussian process concern the spaces $H^m(\mathcal{D})$, $\mathcal{D} \subset \mathbb{R}^d$, only in the case $m > d/2$. This is convenient because it ensures that $H^m(\mathcal{D})$ is continuously embedded in $C_B^0(\mathcal{D})$ when $\mathcal{D}$ is smooth enough, which suppresses the ambiguity of choosing a representative of a function in $H^m(\mathcal{D})$. However, $m > d/2$ excludes the useful spaces $H^1(\mathbb{R}^2)$ and $H^1(\mathbb{R}^3)$, which are central in the study of many important second order PDEs such as the wave equation, the heat equation, Laplace's equation or Schrödinger's equation.

Our characterizations of measurable Gaussian processes with sample paths in $W^{m,p}(\mathcal{D})$ is phrased in terms of a form of Sobolev regularity of the covariance function on the diagonal. It is then linked to the existence of suitable Mercer or otherwise nuclear decompositions of the integral operators associated to the covariance function and its symmetric weak cross-derivatives. In the Hilbert case $p = 2$, additional links are made w.r.t. the Mercer decompositions of the said integral operators, their trace and the Hilbert-Schmidt nature of the imbedding of the RKHS in $H^m(\mathcal{D})$. Our results are strongly reminiscient of those found in [42], where we removed of the continuity assumptions over the covariance in a suitable fashion.

The article is organized as follow. In Section 2, we introduce the necessary notions for properly stating our results as well as some useful lemmas directly related to these notions. In Sections 3 and 4, we state and prove the main results of this article, which treat the general case $p \in (1, +\infty)$ and the special case $p = 2$ respectively. In Section 5, we conclude and provide some further outlooks. We prove the intermediary lemmas used in the main proofs in Section 6.

Notations Given a Banach space X, X^* denotes its topological dual. Given $x \in X$ and $l \in X^*$, we denote the duality bracket as follow: $l(x) = \langle l, x\rangle_{X^*,X}$. $\mathcal{B}(X)$ denotes the Borel σ-algebra of X for its norm topology. Given two linear operators $A : X_1 \to Y_1$ and $B : X_2 \to Y_2$, $A \otimes B : X_1 \otimes X_2 \to Y_1 \otimes Y_2$ denotes their tensor product which verifies $(A \otimes B)(a \otimes b) = (Aa) \otimes (Bb)$. Given two real valued functions f and g, $f \otimes g$ denotes their tensor product defined by $(f \otimes g)(x, y) = f(x)g(y)$. Given $h \in \mathbb{R}^d$, $|h|$ denotes its Euclidean norm. Given $p \in (1, +\infty)$, q will always denote its conjugate: $1/p + 1/q = 1$ i.e. $q = p/(p-1)$. As usual, when $\mathcal{D}$ is an open set of $\mathbb{R}^d$, we identify the dual of $L^p(\mathcal{D})$ with $L^q(\mathcal{D})$. Explicitly, if $f \in L^p(\mathcal{D})$ and $g \in L^q(\mathcal{D})$, we have

$$\langle f, g\rangle_{L^p,L^q} = \int_{\mathcal{D}} f(x)g(x)dx = \langle g, f\rangle_{L^q,L^p}. \tag{1.2}$$

When there is no risk of confusion, we will write $\|f\|_p := \|f\|_{L^p(\mathcal{D})}$. If H is a Hilbert space, $\langle \cdot, \cdot\rangle_H$ denotes its inner product. We denote $\mathbb{N} := \{1, 2, ...\}$ the set of natural numbers and $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. Given an open set $\mathcal{D} \subset \mathbb{R}^d$, we write $\mathcal{D}_0 \Subset \mathcal{D}$ if $\overline{\mathcal{D}_0} \subset \mathcal{D}$ and $\overline{\mathcal{D}_0}$ is compact. $L^1_{loc}(\mathcal{D})$ denotes the space of equivalence classes of locally integrable functions over $\mathcal{D}$, i.e. such that $\int_K |f(x)|dx < +\infty$ for all $K \Subset \mathcal{D}$. Elements of $L^1_{loc}(\mathcal{D})$ are identified when they are equal almost everywhere w.r.t. the Lebesgue measure. Given an equivalence class $f \in L^1_{loc}(\mathcal{D})$, a *representative* of f is a function $\hat{f} : \mathcal{D} \to \mathbb{R}$ such that the equivalence class of $\hat{f}$ in $L^1_{loc}(\mathcal{D})$ is f. We will sometimes denote f and $\hat{f}$ with the same symbol, e.g. f. Given a function k defined over $\mathcal{D} \times \mathcal{D}$, $\mathcal{E}_k$ denotes the associated integral operator (if well defined):

$$(\mathcal{E}_k f)(x) = \int_{\mathcal{D}} k(x, y)f(y)dy. \tag{1.3}$$

The input and output spaces of $\mathcal{E}_k$ will be specified on a case-by-case basis.

2 Background

This section is dedicated to the introduction of the necessary notions required for understanding the main results of the paper, as well as their proofs. It is divided in two parts.

Section 2.1 contains the definitions necessary for understanding the statements of Propositions 3.1, 3.6 and 4.4, which constitute the core results of the article. These definitions concern measurable Gaussian processes, Sobolev spaces as well as certain notions from operator theory.

Section 2.2 describes most of the tools that we will use in Sections 3 and 4. In particular, it contains a first series of propositions and lemmas (either new or already well-known) which all play a central role in the proofs of Propositions 3.1, 3.6 and 4.4. These results are all directly related to the notions introduced in Section 2.1. They concern several characterizations of Sobolev regularity for locally integrable functions, as well as certain facts about integrals of measurable Gaussian processes, Gaussian sequences and measurable Gaussian processes with L^p integrable sample paths. Several characterizations of Gaussian measures over L^p spaces are also given, in terms of measurable Gaussian processes in particular.

2.1 Preliminary definitions

2.1.1 Measurable Gaussian processes

Throughout this article, $(\Omega, \mathcal{F}, \mathbb{P})$ denotes the same probability space. Given $p \in (1, +\infty)$, $L^p(\mathbb{P})$ denotes the space of real valued random variables X such that $\mathbb{E}[|X|^p] < +\infty$.

If $(E, \mathcal{B})$ is a measurable space, the *law* $\mathbb{P}_X$ of a random variable $X : \Omega \to E$ is the push-forward measure of $\mathbb{P}$ through X, which is defined by $\mathbb{P}_X(B) := \mathbb{P}(X^{-1}(B))$ for all measurable set $B \in \mathcal{B}$ ([7], Section 3.7).

A *Gaussian process* ([2], Section 1.2) $(U(x))_{x\in\mathcal{D}}$ is a family of Gaussian random variables defined over $(\Omega, \mathcal{F}, \mathbb{P})$ such that for all $n \in \mathbb{N}$, $(a_1, ..., a_n) \in \mathbb{R}^n$ and $(x_1, ..., x_n) \in \mathcal{D}^n$, $\sum_{i=1}^n a_i U(x_i)$ is a Gaussian random variable. The law it induces over the function space $\mathbb{R}^{\mathcal{D}}$ endowed with its product σ-algebra is uniquely determined by its mean and covariance functions, $m(x) = \mathbb{E}[U(x)]$ and $k(x, x') = \text{Cov}(U(x), U(x'))$ ([26], Section 9.8). We then write $(U(x))_{x\in\mathcal{D}} \sim GP(m, k)$. The covariance function k is positive definite over $\mathcal{D}$, meaning that for all nonnegative integer n and $(x_1, ...x_n) \in \mathcal{D}^n$, the matrix $(k(x_i, x_j))_{1\leqslant i,j\leqslant n}$ is nonnegative definite. Conversely, given a positive definite function over an arbitrary set $\mathcal{D}$, there exists a centered Gaussian process indexed by $\mathcal{D}$ with the this function as its covariance function ([2], p. 11). We will often denote $\sigma(x) := k(x, x)^{1/2}$. Given $\omega \in \Omega$, the corresponding sample path (or realization) of $(U(x))_{x\in\mathcal{D}}$ is the deterministic function $U_\omega : \mathcal{D} \to \mathbb{R}$ defined by $U_\omega(x) := U(x)(\omega)$. A Gaussian process is said to be measurable if the map $(\Omega \times \mathcal{D}, \mathcal{F} \otimes \mathcal{B}(\mathcal{D})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})), (\omega, x) \mapsto U(x)(\omega)$ is measurable. If $(U(x))_{x\in\mathcal{D}}$ is measurable, then from Fubini's theorem the maps of the form $x \mapsto k(x, x'), x \mapsto k(x, x)$, etc, are measurable. If a general random field is continuous in probability, then there exists a measurable modification of this random field ([20], Theorem 2.6 p. 61). This property is e.g. ensured when the covariance function is continuous, from Tchebychev's inequality. Tedious extensions of this result exist ([19], Theorem 2.3). Finally, the measurability property is ensured for processes *defined* by series of the form $U(x)(\omega) := \sum_j \xi_j(\omega)e_j(x)$ where (ξ_j) are random variables and (e_j) are measurable functions such that the series converges in a suitable space. See [45] for similar remarks. We further discuss the measurability property of Gaussian processes in Remark 2.10.

2.1.2 Weak derivatives and Sobolev spaces

Let $\alpha = (\alpha_1, ..., \alpha_d) \in \mathbb{N}_0^d$. We denote $\partial^\alpha = \partial_{x_1}^{\alpha_1}...\partial_{x_d}^{\alpha_d}$ the α^{th} derivative, and $|\alpha| := \sum_{i=1}^d |\alpha_i|$. In this article, the statement "let $|\alpha| \leqslant m$" will mean "let $\alpha = (\alpha_1, ..., \alpha_d) \in \mathbb{N}_0^d$ be such that $|\alpha| \leqslant m$". Given a function k defined on $\mathcal{D} \times \mathcal{D}$, $\partial^{\alpha,\alpha}k$ denotes its symmetric cross derivative: $\partial^{\alpha,\alpha}k(x, y) := \partial_{x_1}^{\alpha_1}...\partial_{x_d}^{\alpha_d}\partial_{y_1}^{\alpha_1}...\partial_{y_d}^{\alpha_d}k(x, y)$ (formally, $\partial^{\alpha,\alpha} = \partial^\alpha \otimes \partial^\alpha$). A function $u \in L^1_{loc}(\mathcal{D})$ has $v \in L^1_{loc}(\mathcal{D})$ for its α^{th} weak derivative if ([1], Section 1.62)

$$\forall \varphi \in C_c^\infty(\mathcal{D}), \quad \int_{\mathcal{D}} u(x)\partial^\alpha \varphi(x)dx = (-1)^{|\alpha|} \int_{\mathcal{D}} v(x)\varphi(x)dx. \tag{2.1}$$

If it exists, v is then unique in $L^1_{loc}(\mathcal{D})$ and is denoted $v = \partial^\alpha u$. Let $p \in [1, +\infty]$. The Sobolev space $W^{m,p}(\mathcal{D})$ is defined as ([1], Section 3.2)

$$W^{m,p}(\mathcal{D}) = \{u \in L^p(\mathcal{D}) : \forall\ |\alpha| \leqslant m, \partial^\alpha u \text{ exists and } \partial^\alpha u \in L^p(\mathcal{D})\}. \tag{2.2}$$

Sobolev spaces are Banach spaces for the norm $\|u\|_{W^{m,p}} := (\sum_{|\alpha|\leqslant m} \|\partial^\alpha u\|_p^p)^{1/p}$; they are separable when $p \neq +\infty$ ([1], Theorem 3.6 p. 61). When $p = 2$, we denote $H^m(\mathcal{D}) := W^{m,2}(\mathcal{D})$. $H^m(\mathcal{D})$ is a Hilbert space for the following inner product

$$\langle u, v\rangle_{H^m(\mathcal{D})} := \sum_{|\alpha|\leqslant m} \langle \partial^\alpha u, \partial^\alpha v\rangle_{L^2(\mathcal{D})}. \tag{2.3}$$

Note that we made no assumptions on the regularity of the open set $\mathcal{D}$.

2.1.3 Notions from operator theory

The following reminders can be found in [6], Section A.2. Let H_1 and H_2 be two Hilbert spaces, and X and Y two Banach spaces.

(i) A linear operator $T : X \to Y$ is bounded if $\|T\| := \sup_{\|x\|_X=1} \|Tx\|_Y < +\infty$. A bounded operator $T : X \to Y$ is compact if $\overline{T(B)}$ is a compact set of Y, where B is the closed unit ball of X. When $X = Y$, the spectrum of a compact operator is purely discrete, and can be reordered as a sequence $(\lambda_n)_{n\in\mathbb{N}}$ which converges to 0.

(ii) If $T : H_1 \to H_2$ is compact, then $T^*T : H_1 \to H_1$ is compact, self-adjoint and nonnegative ($\forall x \in H_1, \langle x, T^*Tx\rangle_{H_1} \geqslant 0$). If H_1 is separable, T^*T can then be diagonalized in an orthonormal basis (e_n) of H_1. Denote (λ_n) the nonnegative eigenvalues of T^*T, their square roots $s_n := \sqrt{\lambda_n}$ are called the singular values of T. If H_1 is separable, T is said to be Hilbert-Schmidt if $\sum_{n\in\mathbb{N}} \|Tf_n\|^2_{H_2} < +\infty$ for one orthonormal basis (f_n) of H_1, in which case the value of this sum does not depend on the orthonormal basis at hand. The Hilbert-Schmidt norm of T, defined as the square root of the sum above, is then also equal to the discrete ℓ^2 norm of its singular values:

$$\|T\|^2_{HS} = \sum_{n\in\mathbb{N}} \|Tf_n\|^2_{H_2} = \sum_{n\in\mathbb{N}} \|Te_n\|^2_{H_2} = \sum_{n\in\mathbb{N}} \langle e_n, T^*Te_n\rangle_{H_1} = \sum_{n\in\mathbb{N}} s_n^2. \tag{2.4}$$

Every Hilbert-Schmidt operator is compact, and every Hilbert-Schmidt operator T acting on $L^2(\mathcal{D})$ can be written in integral form ([6], Lemma A.2.13): there exists a "kernel" $k \in L^2(\mathcal{D} \times \mathcal{D})$ such that for all $f \in L^2(\mathcal{D})$,

$$(Tf)(x) = \int_{\mathcal{D}} k(x, y)f(y)dy = (\mathcal{E}_k f)(x). \tag{2.5}$$

If T is symmetric, nonnegative and Hilbert-Schmidt, there exists an orthonormal basis (ϕ_n) of $L^2(\mathcal{D})$ of eigenvectors of T with nonnegative eigenvalues (λ_n), such that in $L^2(\mathcal{D} \times \mathcal{D})$, we have

$$k(x, y) = \sum_{n\in\mathbb{N}} \lambda_n \phi_n(x)\phi_n(y). \tag{2.6}$$

We will refer to decompositions of f of the form of equation (2.6) as *Mercer decompositions*, in reference to the celebrated Mercer's theorem ([10], Theorem 1.2).

(iii) If H_1 is separable, a general compact operator $T : H_1 \to H_1$ is said to be trace-class (or nuclear) if $\sum_{n\in\mathbb{N}} s_n < +\infty$, where the sequence (s_n) still corresponds to the singular values of T. One can then define its trace as the following linear functional, which is independent of the choice of basis (e_n), and equal to the (absolutely convergent) series of the eigenvalues (μ_n) of T (Lidskii's theorem):

$$\mathrm{Tr}(T) := \sum_{n\in\mathbb{N}} \langle Te_n, e_n\rangle = \sum_{n\in\mathbb{N}} \mu_n. \tag{2.7}$$

Any trace-class operator is Hilbert-Schmidt, and T is Hilbert-Schmidt if and only if either T^*T or TT^* is trace-class, in which case $\mathrm{Tr}(T^*T) = \|T\|_{HS}^2 = \|T^*\|_{HS}^2$. If $H_1 = H_2 = L^2(\mathcal{D})$, if T is trace class with kernel k and if k is sufficiently smooth (say continuous), then the trace of $T = \mathcal{E}_k$ is given by $\mathrm{Tr}(T) = \int_{\mathcal{D}} k(x,x)dx$. Extensions of this formula to general Hilbert-Schmidt kernels $k \in L^2(\mathcal{D} \times \mathcal{D})$ of trace class operators is studied in [10]; see also Proposition 2.9 and Lemma 3.8 below. If $T : H_1 \to H_1$ is bounded, self-adjoint and nonnegative, then we define its trace as the possibly infinite series of nonnegative scalars $\mathrm{Tr}(T) := \sum_{n\in\mathbb{N}} \langle Te_n, e_n\rangle$.

(iv) ([32], p. 160) A bounded operator $T : X \to Y$ is nuclear if there exists sequences $(x_n) \subset X^*$ and $(y_n) \subset Y$ with $\sum_{n=1}^{+\infty} \|x_n\|_{X^*}\|y_n\|_Y < +\infty$ such that $Tx = \sum_{n=1}^{+\infty} \langle x_n, x\rangle_{X^*,X} y_n$ for all $x \in X$. In this case, we write abusively $T = \sum_{n=1}^{+\infty} x_n \otimes y_n$. The nuclear norm of T is then defined as

$$\nu(T) := \inf \left\{ \sum_{n=1}^{+\infty} \|x_n\|_{X^*}\|y_n\|_Y \text{ such that } T = \sum_{n=1}^{+\infty} x_n \otimes y_n \right\}. \tag{2.8}$$

A bounded operator $K : X^* \to X$ is symmetric if for all $x, y \in X^*$, $\langle x, Ry\rangle = \langle y, Rx\rangle$, and nonnegative if $\langle x, Rx\rangle \geqslant 0$. When $X = Y = H$ where H is a separable Hilbert space, the sets of trace class and nuclear operators coincide; moreover, the same can be said for the trace functional (2.7) and the nuclear norm (2.8) if T has a nonnegative spectrum : $\nu(T) = \mathrm{Tr}(T)$.

2.2 Main tools of the article

2.2.1 Characterization of $W^{m,p}$-regularity for locally integrable functions

As for pointwise derivatives, finite difference operators can be used for characterizing Sobolev regularity. Given $y \in \mathbb{R}^d$, introduce the translation operator $(\tau_y u)(x) = u(x+y)$, which is bounded over $L^p(\mathbb{R}^d)$. Introduce the associated finite difference operator:

$$\Delta_y = \tau_y - Id. \tag{2.9}$$

The linear subspace of bounded operators over $L^p(\mathbb{R}^d)$ induced by the translation operators is commutative, as $\tau_{y_1} \circ \tau_{y_2} = \tau_{y_1+y_2} = \tau_{y_2} \circ \tau_{y_1}$. Let $(y_1, ..., y_m) \in (\mathbb{R}^d)^m$, we define the m^{th} order finite difference operator associated to $(y_1, ..., y_m)$ to be $\Delta_{(y_1,...,y_m)} := \prod_{i=1}^m \Delta_{y_i}$ where the product symbol denotes the composition of operators. When $y \in \mathbb{R}^d$, the adjoint of Δ_y is also a finite difference operator, which is computable using the change of variable formula. If $y \in \mathbb{R}^d$, then

$$\Delta_y^* = \tau_{-y} - Id. \tag{2.10}$$

Finally, when $\alpha = (\alpha_1, ..., \alpha_d) \in \mathbb{N}_0^d$ and $h = (h_1, ..., h_d) \in (\mathbb{R}_+^*)^d$, we denote by δ_h^α the finite difference approximation of ∂^α defined by

$$\delta_h^\alpha = \prod_{i=1}^d \left(\frac{\Delta_{h_i e_i}}{h_i}\right)^{\alpha_i} = \left(\frac{\Delta_{h_1 e_1}}{h_1}\right)^{\alpha_1} \cdots \left(\frac{\Delta_{h_d e_d}}{h_d}\right)^{\alpha_d}. \tag{2.11}$$

Above, $(e_1, ..., e_d)$ is the canonical basis of $\mathbb{R}^d$. We use the convention that $(\Delta_{h_i e_i})^{\alpha_i} = Id$ if $\alpha_i = 0$. Depending on which one is the most convenient, we will either use $\Delta_{(y_1,...,y_m)}$ or δ_h^α.

We will use the following characterizations of $W^{m,p}$-regularity; they are "straightforward" generalizations of Proposition 9.3 from [9] to $m \geqslant 2$. These characterizations have the benefit of being valid without any regularity assumptions on the open set $\mathcal{D}$. We prove Lemma 2.1 in Section 6, as we could not find it stated as such in the literature.

Lemma 2.1. *Suppose that $u \in L^1_{loc}(\mathcal{D})$. Let $m \in \mathbb{N}_0$, $p \in (1, +\infty]$ and introduce $q \geqslant 1$ the conjugate of p : $1/p + 1/q = 1$. Then the following statements are equivalent.*

(i) $u \in W^{m,p}(\mathcal{D})$

(ii) (Variational control) for all α such that $|\alpha| \leqslant m$, there exists a constant C_α such that

$$\forall \varphi \in C_c^\infty(\mathcal{D}), \quad \Big| \int_{\mathcal{D}} u(x) \partial^\alpha \varphi(x) dx \Big| \leqslant C_\alpha \|\varphi\|_{L^q(\mathcal{D})}. \tag{2.12}$$

In this case, the L^p norm of $\partial^\alpha u$ is given by

$$\|\partial^\alpha u\|_{L^p(\mathcal{D})} = \sup_{\varphi \in C_c^\infty(\mathcal{D}) \backslash \{0\}} \Big| \int_{\mathcal{D}} u(x) \frac{\partial^\alpha \varphi(x)}{\|\varphi\|_{L^q}} dx \Big|. \tag{2.13}$$

(iii) (Finite difference control) there exists a constant C such that for all open set $\mathcal{D}_0 \Subset \mathcal{D}$, for all $\ell \leqslant m$ and all $(y_1, ..., y_\ell) \in (\mathbb{R}^d)^\ell$ such that $\sum_{i=1}^\ell |y_i| < dist(\mathcal{D}_0, \partial \mathcal{D})$,

$$\|\Delta_{(y_1,...,y_\ell)} u\|_{L^p(\mathcal{D}_0)} \leqslant C |y_1| \times ... \times |y_\ell|. \tag{2.14}$$

Moreover, for all $|\alpha| \leqslant m$, $\|\partial^\alpha u\|_{L^p(\mathcal{D})} \leqslant C$ for any C verifying equation (2.14)*. Finally, one can actually take $C = \|u\|_{W^{m,p}(\mathcal{D})}$ in equation* (2.14)*.*

In Point (iii) above, the assumption that $\sum_{i=1}^\ell |y_i| < \text{dist}(\mathcal{D}_0, \partial\mathcal{D})$ ensures that the quantity $\Delta_{(y_1,...,y_\ell)} u(x)$ makes sense when $x \in \mathcal{D}_0$. A similar criterion to Point (iii) above is given in [31], Theorem 10.55, still in the case $m = 1$ (as well as in [23], Section 5.8.2.a, Theorem 3). This theorem only requires the L^p control of the ratios $\delta_h^\alpha u$ with $|\alpha| \leqslant 1$. As such, we could have also stated a version of Lemma 2.1(iii) solely in terms of the ratios $\delta_h^\alpha u$ with $|\alpha| \leqslant m$.

2.2.2 Sobolev regularity and generalized functions

The theory of generalized functions (or distributions) provides a flexible way of characterizing Sobolev regularity, by building a larger space in which partial derivatives are always defined. Given an open set $\mathcal{D}$, denote $C_c^\infty(\mathcal{D})$ the space of smooth functions with compact support in $\mathcal{D}$. Endow it with its usual LF topology, defined e.g. in [48], Chapter 13. This topology is such that the sequence (φ_n) converges to φ in $C_c^\infty(\mathcal{D})$ if and only if there exists a compact set $K \subset \mathcal{D}$ such that $\text{Supp}(\varphi_n) \subset K$ for all n and

$$\forall \alpha = (\alpha_1, ..., \alpha_d) \in \mathbb{N}_0^d, \quad \sup_{x \in K} |\partial^\alpha \varphi_n(x) - \partial^\alpha \varphi(x)| \longrightarrow 0. \tag{2.15}$$

With $C_c^\infty(\mathcal{D})$ endowed with this topology, the space of generalized functions, or distributions, is then defined as the topological dual of $C_c^\infty(\mathcal{D})$ i.e. the set of all continuous linear forms over $C_c^\infty(\mathcal{D})$. It is traditionally denoted as follow: $\mathscr{D}'(\mathcal{D}) := C_c^\infty(\mathcal{D})'$ ([48], Notation 21.1). A generalized function $T \in \mathscr{D}'(\mathcal{D})$ is said to be regular ([48], p. 224) if it is of the form

$$\forall \varphi \in C_c^\infty(\mathcal{D}), \quad T(\varphi) = \int_{\mathcal{D}} u(x) \varphi(x) dx, \tag{2.16}$$

for some $u \in L^1_{loc}(\mathcal{D})$, in which case one writes $T = T_u$. Given any function $u \in L^1_{loc}(\mathcal{D})$ and $\alpha \in \mathbb{N}_0^d$, its distributional derivative $D^\alpha u$ is *defined* by the following formula ([48], pp. 248-250):

$$D^\alpha u : \varphi \longmapsto (-1)^{|\alpha|} \int_{\mathcal{D}} \partial^\alpha \varphi(x) u(x) dx. \tag{2.17}$$

$D^\alpha u$ then also lies in $\mathscr{D}'(\mathcal{D})$. Sobolev regularity can now be rephrased as follow: u lies in $W^{m,p}(\mathcal{D})$ iff for all $|\alpha| \leqslant m$, the distributional derivative $D^\alpha u$ is in fact a regular generalized function represented by some $v_\alpha \in L^p(\mathcal{D})$ i.e. $D^\alpha u = T_{v_\alpha}$. Then v_α is unique in $L^p(\mathcal{D})$ and $\partial^\alpha u = v_\alpha$ in $L^p(\mathcal{D})$, where $\partial^\alpha u$ is the α^{th} weak derivative of u.

Moreover, the control equation (2.12) shows that $\partial^\alpha u$ exists and lies in $L^p(\mathcal{D})$ if and only if $D^\alpha u : C_c^\infty(\mathcal{D}) \to \mathbb{R}$ can be extended as a continuous linear form over $L^q(\mathcal{D})$. Ensuring the existence of such extensions will thus be of prime interest for us, and is the topic of the next lemma. Specifically, the next result states that given continuous linear or bilinear forms over $C_c^\infty(\mathcal{D})$, the existence of extensions of these maps to $L^q(\mathcal{D})$ can be ensured by obtaining suitable estimates on a well chosen *countable* set $E_q \subset C_c^\infty(\mathcal{D})$. Restricting ourselves to E_q will allow us to eliminate any measurability issues when introducing the supremum of certain random variables indexed by E_q, as a countable supremum of random variables remains a random variable (i.e. a measurable map). Below, we write $\|\cdot\|_q := \|\cdot\|_{L^q(\mathcal{D})}$ for short.

Lemma 2.2 (Extending continuous linear and bilinear forms over $C_c^\infty(\mathcal{D})$ to $L^p(\mathcal{D})$)**.** *Let $p \in (1, +\infty)$. There exists a countable $\mathbb{Q}-$vector space $E_q = \{\Phi_n^q, n \in \mathbb{N}_0\} \subset C_c^\infty(\mathcal{D})$ with the following property.*

(i) A distribution $T \in \mathscr{D}'(\mathcal{D})$ is a regular distribution, $T = T_v$, for some $v \in L^p(\mathcal{D})$ iff it verifies the countable estimate for some constant $C > 0$

$$\forall \varphi \in E_q, \quad |T(\varphi)| \leqslant C\|\varphi\|_q, \tag{2.18}$$

or equivalently, $\sup_{n\in\mathbb{N}} |T(\Phi_n^q)|/\|\Phi_n^q\|_q < +\infty$ (here, setting $\Phi_0^q = 0$ without loss of generality). This is equivalent to T admitting an extension over $L^q(\mathcal{D})$ which is then uniquely given by $T(f) = \int_{\mathcal{D}} f(x)v(x)dx$. Moreover,

$$\sup_{n\in\mathbb{N}} \frac{|T(\Phi_n^q)|}{\|\Phi_n^q\|_q} = \sup_{\varphi \in C_c^\infty(\mathcal{D})} \frac{|T(\varphi)|}{\|\varphi\|_q}, \tag{2.19}$$

whether these quantities are finite or not.

(ii) Let b be a continuous bilinear form over $C_c^\infty(\mathcal{D})$. Then b can be extended to a continuous bilinear form over $L^q(\mathcal{D})$ iff it verifies the countable estimate

$$\forall \varphi, \psi \in E_q, \quad |b(\varphi, \psi)| \leqslant C\|\varphi\|_q\|\psi\|_q. \tag{2.20}$$

In this case, such an extension is unique and there will exist a unique bounded operator $B : L^q(\mathcal{D}) \to L^p(\mathcal{D})$ verifying the following identity

$$\forall \varphi, \psi \in C_c^\infty(\mathcal{D}), \quad b(\varphi, \psi) = \langle B\varphi, \psi \rangle_{L^p, L^q}. \tag{2.21}$$

The proof of this result can be found in Section 6. It relies on Lemma 2.3 below, which is interesting in itself.

Lemma 2.3. *$C_c^\infty(\mathcal{D})$ is sequentially separable, i.e. there exists a countable subset $F \subset C_c^\infty(\mathcal{D})$ such that for all $\varphi \in C_c^\infty(\mathcal{D})$, there exists a sequence $(\varphi_n) \subset F$ such that $\varphi_n \to \varphi$ in $C_c^\infty(\mathcal{D})$ for its LF topology.*[1]

A short proof of this result can be found in [24], Lemma 3.6. Given the set E_q provided by Lemma 2.2, we next define the countable set F_q to be

$$F_q := \{\varphi/\|\varphi\|_q, \varphi \in E_q, \varphi \neq 0\} = \{f_n^q, n \in \mathbb{N}\} \subset S_q(0,1). \tag{2.22}$$

Above, $(f_n^q)_{n\in\mathbb{N}}$ is an enumeration of F_q and $S_q(0,1)$ is the unit sphere of $L^q(\mathcal{D})$. The next lemma is then a direct consequence of Lemmas 2.1 and 2.2.

Lemma 2.4 (Countable characterization of Sobolev regularity)**.** *Let $p \in (1,+\infty)$. For any $u \in L^1_{loc}(\mathcal{D})$, u lies in $W^{m,p}(\mathcal{D})$ iff for all multi index α such that $|\alpha| \leqslant m$, there exists a constant C_α such that*

$$\forall \varphi \in E_q, \quad \left| \int_{\mathcal{D}} u(x)\partial^\alpha \varphi(x)dx \right| \leqslant C_\alpha \|\varphi\|_q, \tag{2.23}$$

or equivalently, in terms of the set F_q defined in equation (2.22)*,*

$$\sup_{\varphi\in F_q} \left| \int_{\mathcal{D}} u(x)\partial^\alpha \varphi(x)dx \right| = \sup_{n\in\mathbb{N}} \left| \int_{\mathcal{D}} u(x)\partial^\alpha f_n^q(x)dx \right| < +\infty. \tag{2.24}$$

Moreover,

$$\sup_{\varphi\in F_q} \left| \int_{\mathcal{D}} u(x)\partial^\alpha \varphi(x)dx \right| = \sup_{\varphi\in C_c^\infty(\mathcal{D})\setminus\{0\}} \left| \int_{\mathcal{D}} u(x)\frac{\partial^\alpha \varphi(x)}{\|\varphi\|_q}dx \right|, \tag{2.25}$$

whether these quantities are finite or not. If one of them is finite, then it is equal to $\|\partial^\alpha u\|_{L^p(\mathcal{D})}$.

This lemma provides us with a somewhat explicit *countable* criteria for Sobolev regularity, which is valid whatever the open set $\mathcal{D}$.

2.2.3 Integrals of measurable Gaussian processes

We will need the following lemma pertaining to the sample path-wise integration of Gaussian processes.

Lemma 2.5. *Let $\mathcal{D} \subset \mathbb{R}^d$ be an open set. Let $(U(x))_{x\in\mathcal{D}} \sim GP(0,k)$ be a measurable centered Gaussian process such that its standard deviation function σ lies in $L^1_{loc}(\mathcal{D})$. Then the sample paths of U lie in $L^1_{loc}(\mathcal{D})$ almost surely and given $\varphi \in C_c^\infty(\mathcal{D})$ and $\alpha \in \mathbb{N}_0^d$, the map defined by*

$$U_\varphi^\alpha : \Omega \ni \omega \longmapsto (-1)^{|\alpha|} \int_{\mathcal{D}} U_\omega(x)\partial^\alpha \varphi(x)dx \tag{2.26}$$

is a Gaussian random variable. Moreover, for all $p \in (1,+\infty)$, $(U_\varphi^\alpha)_{\varphi\in F_q}$ is a centered Gaussian sequence (i.e. a Gaussian process indexed by a countable set), where F_q is defined in equation (2.22)*.*

[1] Recall that a topological space X is *separable* if there exists a countable subset $Y \subset X$ which is dense in X for the topology of X. As for continuity and sequential continuity, separability and sequential separability agree for metric spaces but not for general topological spaces (see e.g. [4] for further discussions).

Following Lemma 2.4, in order to study the $W^{m,p}$ regularity of Gaussian processes, we will be interested in the boundedness of the Gaussian sequence $(U_\varphi^\alpha)_{\varphi \in F_q}$ given $|\alpha| \leqslant m$. As such, the next result pertaining to bounded Gaussian sequences will be quite helpful. It can be seen as a weak form of Fernique's theorem ([6], Theorem 2.8.5, p. 75).

Lemma 2.6 ([2], Theorem 2.1.2)**.** *Let $(U_n)_{n\in\mathbb{N}}$ be a Gaussian sequence and set $|U| := \sup_n |U_n|$. Suppose that $\mathbb{P}(|U| < +\infty) = 1$. Then there exists $\varepsilon > 0$ such that*

$$\mathbb{E}[\exp(\varepsilon |U|^2)] < +\infty. \tag{2.27}$$

In particular, $\mathbb{E}[|U|^p] < +\infty$ for all $p \in \mathbb{N}$.

2.2.4 Gaussian measures over Banach spaces and L^p spaces

This section will be helpful for providing a necessary and sufficient condition according to which $\mathbb{P}(U \in W^{m,p}(\mathcal{D}) = 1)$, in terms of the "spectral" properties of the covariance kernel of U.

A *Gaussian measure* μ ([6], Definition 2.2.1) over a Banach space X is a measure over its Borel σ-algebra such that given any $x^* \in X^*$, the pushforward measure of μ through the functional x^* is a Gaussian measure over $\mathbb{R}$ (see Section 2.1.1 for a definition of the pushforward). Gaussian measures are equipped with a mean vector $a_\mu \in X^{**}$ and a covariance operator $K_\mu : X^* \to X^{**}$, defined in [6], Definition 2.2.7. When X is separable, μ is Radon ([6], p. 125). This implies that a_μ lies in X and that the covariance operator K_μ maps X^* to X ([6], Theorem 3.2.3). The vector a_μ and the covariance operator K_μ are defined by the following formulas

$$\forall x \in X^*,\ \langle a_\mu, x\rangle = \int_X \langle x, z\rangle \mu(dz), \tag{2.28}$$

$$\forall x, y \in X^*, \langle y, K_\mu x\rangle = \int_X \langle x - a_\mu, z\rangle \langle y - a_\mu, z\rangle \mu(dz). \tag{2.29}$$

Any operator $K : X^* \to X^{**}$ which is the covariance operator of a Gaussian measure is called a Gaussian covariance operator. In Propositions 2.7 and 2.8, we present useful characterizations of Gaussian measures μ over two important classes of Banach spaces: spaces of type 2 and cotype 2 respectively. For a definition of spaces of type 2 and cotype 2, see e.g. [12]. In this article, we will only use the fact that $L^p(\mathcal{D})$ is of type 2 when $p \geqslant 2$, and cotype 2 when $1 \leqslant p \leqslant 2$ (see [6], p. 152). Moreover we will restrict ourselves to the case where X is separable. As this implies that μ is Radon, this removes problems pertaining to extensions of measures otherwise considered in [32] and [12].

Proposition 2.7 ([32], Theorem 4 or [6], Remark 3.11.24)**.** *Let X be a separable Banach space of type 2, and let μ be a Gaussian measure over X. Then its covariance operator is symmetric, nonnegative and nuclear. Conversely, given any $a \in X$ and any symmetric, nonnegative and nuclear operator $K : X^* \to X$, there exists a Gaussian measure over X with mean vector a and covariance operator K.*

Denote ℓ^2 the Hilbert space of square summable sequences.

Proposition 2.8 ([12], Theorem 4.1 and Corollary 4.1)**.** *Let X be a separable Banach space of cotype 2, and let μ be a Gaussian measure over X. Then there exists a continuous linear map $A : l^2 \to X$ and a symmetric, nonnegative and trace-class operator $S : l^2 \to l^2$ such that covariance operator of μ is given by ASA^* (in particular, the covariance operator of μ is nuclear). In other words, μ is the pushforward measure of a Gaussian measure μ_0 over ℓ^2*

through some bounded linear map A. Conversely, given any $a \in X$ and any operator of the form ASA^* where $A : \ell^2 \to X$ is a bounded linear map and S a symmetric, nonnegative and trace class operator over ℓ^2, there exists a Gaussian measure over X with mean vector a and covariance operator K.

In practice, we will replace ℓ^2 with $L^2(\mathcal{D})$, which are isomorphic Hilbert spaces. The propositions 2.7 and 2.8 generalize the case where X is a separable Hilbert space, which can be found in [6], Theorem 2.3.1. We finish with the following handy result describing centered Gaussian measures over L^p-spaces.

Proposition 2.9 ([6], Proposition 3.11.15 and Example 2.3.16)**.**

(i) Let μ be a centered Gaussian measure over $L^p(\mathcal{D})$ where $1 \leqslant p < +\infty$ and $\mathcal{D} \subset \mathbb{R}^d$ is an open set. Then there exists a function $k \in L^p(\mathcal{D} \times \mathcal{D})$ such that the covariance operator of μ is $\mathcal{E}_k : L^q(\mathcal{D}) \to L^p(\mathcal{D})$, the integral operator associated to k. Moreover, there exists a centered measurable Gaussian process $(U(x))_{x \in \mathcal{D}}$ whose covariance function $\tilde{k}$ verifies $\tilde{k} = k$ in $L^p(\mathcal{D} \times \mathcal{D})$, and whose sample paths lie in $L^p(\mathcal{D})$ a.s.. Setting $\sigma(x) = \tilde{k}(x,x)^{1/2}$, $\tilde{k}$ verifies

$$\int_{\mathcal{D}} \tilde{k}(x,x)^{p/2} dx = \int_{\mathcal{D}} \sigma(x)^p dx < +\infty. \tag{2.30}$$

Additionally, $\mathbb{P}_U = \mu$, where $\mathbb{P}_U$ is the pushforward of $\mathbb{P}$ through the Borel-measurable map $\omega \mapsto U_\omega \in L^p(\mathcal{D})$. Conversely, given any measurable nonnegative definite function k verifying (2.30)*, the corresponding integral operator $\mathcal{E}_k : L^q(\mathcal{D}) \to L^p(\mathcal{D})$ is the covariance operator of a centered Gaussian measure μ over $L^p(\mathcal{D})$.*

(ii) Given a centered measurable Gaussian process $(U(x))_{x \in \mathcal{D}}$ whose covariance function we denote $\tilde{k}$, the condition (2.30) *is equivalent to $(U(x))_{x \in \mathcal{D}}$ having its sample paths lie in $L^p(\mathcal{D})$ a.s..*

This result is quite strong, as it ensures the existence of a representative in $L^p(\mathcal{D} \times \mathcal{D})$ of the kernel of any Gaussian covariance operator, which is the covariance function of a measurable Gaussian process. This will enable us to remove awkward measurability issues w.r.t. k over its diagonal and equation (2.30). Without the use of an underlying measurable Gaussian process, these issues are not trivial to deal with, see e.g. [10] for an analysis of the Hilbert case $p = 2$.

Remark 2.10. Proposition 2.9 shows that the assumption that a given Gaussian process is measurable is slightly less demanding that it might seem. As observed in Section 2.1.1, the existence of a measurable modification of a general random field is difficult outside of it being continuous in probability. For a Gaussian process $(U(x))_{x \in \mathcal{D}} \sim GP(0, k_u)$ however, Propositions 2.7, 2.8 and 2.9 shows that the measurability of its covariance function over $\mathcal{D} \times \mathcal{D}$ and the integrability of its standard deviation in $L^p(\mathcal{D})$ (or equivalently, suitable nuclear decompositions of its associated integral operator $\mathcal{E}_k$) ensure the existence of a measurable Gaussian process $(V(x))_{x \in \mathcal{D}} \sim GP(0, k_v)$ with the same covariance function in $L^1_{loc}(\mathcal{D} \times \mathcal{D})$. Consequently, $k_u = k_v$ a.e. on $\mathcal{D} \times \mathcal{D}$. Note though that the process V need not be a modification of U. Since $k_u = k_v$ a.e., we only have that U and V have the same finite dimensional marginals "almost everywhere" in the sense of the Lebesgue measure: for all $n \in \mathbb{N}$ and almost every $(x_1, ..., x_n) \in \mathcal{D}^n$, $(U(x_1), ..., U(x_n))$ and $(V(x_1), ..., V(x_n))$ have the same law.

Throughout this article, we will only consider centered Gaussian processes ($\mathbb{E}[U(x)] \equiv 0$) and Gaussian measures ($a_\mu = 0$). Generalizations of the results of this article to non centered Gaussian processes are straightforward.

3 Sobolev regularity for Gaussian processes : the general case, $1 < p < +\infty$

We can now state our first result, which deals with $W^{m,p}(\mathcal{D})$-regularity of Gaussian processes, given any $p \in (1, +\infty)$ and any open set $\mathcal{D} \subset \mathbb{R}^d$.

Proposition 3.1 (Sample path Banach-Sobolev regularity for Gaussian processes). *Let $\mathcal{D} \subset \mathbb{R}^d$ be an open set. Let $(U(x))_{x\in\mathcal{D}} \sim GP(0,k)$ be a measurable centered Gaussian process, defined on a probability set $(\Omega, \mathcal{F}, \mathbb{P})$, such that its standard deviation function σ lies in $L^1_{loc}(\mathcal{D})$. Let $p \in (1, +\infty)$. Then the following statements are equivalent.*

(i) (Sample path regularity) The sample paths of $(U(x))_{x\in\mathcal{D}}$ lie in $W^{m,p}(\mathcal{D})$ almost surely.

(ii) (Integral criteria) For all $|\alpha| \leqslant m$, the distributional derivative $\partial^{\alpha,\alpha}k$ lies in $L^p(\mathcal{D}\times\mathcal{D})$ and admits a representative k_α in $L^p(\mathcal{D}\times\mathcal{D})$ which is the covariance function of a measurable Gaussian process. For all such k_α, denoting $\sigma_\alpha(x) := k_\alpha(x,x)^{1/2}$, we have

$$\int_{\mathcal{D}} \sigma_\alpha(x)^p dx < +\infty. \tag{3.1}$$

(iii) (Covariance structure) For all $|\alpha| \leqslant m$, the distributional derivative $\partial^{\alpha,\alpha}k$ lies in $L^p(\mathcal{D} \times \mathcal{D})$ and the associated integral operator $\mathcal{E}_k^\alpha : L^q(\mathcal{D}) \to L^p(\mathcal{D})$ defined by

$$\mathcal{E}_k^\alpha f(x) = \int_{\mathcal{D}} \partial^{\alpha,\alpha}k(x,y)f(y)dy \tag{3.2}$$

is symmetric, nonnegative and nuclear: there exists $(\lambda_n^\alpha) \subset \mathbb{R}_+$ and $(\psi_n^\alpha) \subset L^p(\mathcal{D})$ such that

$$\sum_{n=0}^{+\infty} \lambda_n^\alpha \|\psi_n^\alpha\|_p^2 < +\infty, \quad \partial^{\alpha,\alpha}k(x,y) = \sum_{n=0}^{+\infty} \lambda_n^\alpha \psi_n^\alpha(x)\psi_n^\alpha(y) \ \text{ in } \ L^p(\mathcal{D}\times\mathcal{D}). \tag{3.3}$$

If $1 \leqslant p \leqslant 2$, then one can choose (λ_n^α) such that $\sum_n \lambda_n^\alpha < +\infty$, and there exists a bounded operator $A_\alpha : L^2(\mathcal{D}) \to L^p(\mathcal{D})$ and an orthonormal basis (ϕ_n^α) of $L^2(\mathcal{D})$ such that $\psi_n^\alpha = A_\alpha \phi_n^\alpha$ for all $n \geqslant 0$ (in particular, we have the uniform bound $\|\psi_n^\alpha\|_p \leqslant \|A_\alpha\|$).

The proposition above shows that a suitable L^p control of the function $\partial^{\alpha,\alpha}k$ over the diagonal is necessary and sufficient for ensuring the Sobolev regularity of the sample paths of the Gaussian process with covariance function k. Formally speaking, the function $(x,y) \mapsto \partial^{\alpha,\alpha}k(x,y)$ is the covariance function of the differentiated process, $(\omega, x) \mapsto \partial^\alpha U_\omega(x)$. This is formal only, as the weak derivative of the sample paths are only defined up to a set of Lebesgue measure zero, and thus there is no obvious way of defining the joint map $(\omega, x) \mapsto \partial^\alpha U_\omega(x)$. Note also that the idea of ensuring a suitable control of this covariance function near its diagonal is not with reminding more standard results pertaining to the differentiability in the *mean square* sense of a random process (see e.g. [2], Section 1.4.2). See [42] for similar remarks on the Sobolev regularity of random fields.

Observe also that there is an asymmetry between Point (ii) and Point (iii) of Proposition 3.1, as one depends on whether p is lower or greater than 2 while the other does not. Moreover, both points rely on the finiteness of some quantity, so explicit bounds should be sought so that Point (ii) controls Point (iii) and conversely. This is the content of Proposition 3.6.

Finally, observe that the integrability criteria (ii) cannot be expected to hold for any positive definite representative $\tilde{k}_\alpha$ of $\partial^{\alpha,\alpha}k$, even if $\tilde{k}_\alpha$ is measurable on its diagonal. For example, set

$\tilde{k}_\alpha(x,y) := k_\alpha(x,y) + \delta_{x,y}$ where $\delta_{x,y}$ is the Kronecker delta, which verifies $\tilde{k}_\alpha = \partial^{\alpha,\alpha}k$ in $L^p(\mathcal{D}\times\mathcal{D})$. But if $\mathcal{D}$ has infinite Lebesgue measure, it is also clear that $\int_\mathcal{D} \tilde{k}_\alpha(x,x)^{p/2}dx \geqslant \int_\mathcal{D} \delta_{x,x}dx = +\infty$. Lemma 3.8 describes a natural set of "admissible" representatives for which Point (ii) holds, in the case $p \geqslant 2$.

Remark 3.2. Under the assumption that $(U(x))_{x\in\mathcal{D}}$ is measurable, the statement that its sample paths lie in some Sobolev space is *not* up to a modification of the process. This is a consequence of Lemmas 2.4, 2.5 and 2.6, which show that the Sobolev regularity of its paths is fully determined by the finite dimensional marginals of the process (see equation (3.6)). This contrasts with more classical results, e.g. pertaining to the continuity of the process ([3], Section 1.4.1). Still, ensuring the measurability of the process is not really straightforward (see Remark 2.10; in fact, this property may happen to be only ensured up to a modification of the initial process).

Example 3.3 (Finite rank covariance functions)**.** Let $p \in (1,+\infty)$, $m \in \mathbb{N}_0$ and $n \in \mathbb{N}$. Consider $f_1,...,f_n \in L^p(\mathcal{D})$ and choose representatives of those functions in $L^p(\mathcal{D})$, also denoted by $f_1,...,f_n$, so that they may be understood as functions in the classical sense. Consider the covariance function $k(x,x') := \sum_{i=1}^n f_i(x)f_i(x')$. Assume that $f_1,...,f_n \in W^{m,p}(\mathcal{D})$, then for all $|\alpha| \leqslant m$, the weak derivative $\partial^{\alpha,\alpha}k$ is obviously given by

$$\partial^{\alpha,\alpha}k(x,x') = \sum_{i=1}^n \partial^\alpha f_i(x)\partial^\alpha f_i(x') \quad \text{in } L^p(\mathcal{D}\times\mathcal{D}), \tag{3.4}$$

and the associated integral operators fulfill the criterion (iii) of Proposition 3.1. Thus the corresponding measurable Gaussian process has its sample paths in $W^{m,p}(\mathcal{D})$ almost surely. Note that this was obvious in the first place, since this Gaussian process can be written as $U(x) = \sum_{i=1}^n \xi_i f_i(x)$ where $\xi_1,...,\xi_n$ are independent standard Gaussian random variables (checking that the covariance function is the right one is trivial). Conversely, assume that the sample paths of the associated measurable Gaussian process lie in $W^{m,p}(\mathcal{D})$ almost surely. Then the function k verifies the criterion (iii) of Proposition 3.1, and in particular $\partial^{\alpha,\alpha}k \in L^p(\mathcal{D}\times\mathcal{D})$. One can then show that $f_1,...,f_n \in W^{m,p}(\mathcal{D})$ (copy the proof of Lemma 4.9, Point (i)).

While this example can easily be studied and solved on its own, we observed in the introduction that (surprisingly) this example fell out of the scope of the previous results pertaining to the Sobolev regularity of Gaussian processes. Indeed, the stationarity assumptions of [15,25] are not met. Since the domain $\mathcal{D}$ is not assumed to be endowed with an underlying Dirichlet structure, the results from [28] pertaining to the $B^s_{\infty,\infty}$ regularity of Gaussian processes do not apply. In fact, in our setting, $\mathcal{D}$ is not even assumed to be compact and k is not assumed continuous, contrarily to [28]. Likewise, the one dimensional framework of [14] is too restrictive for our example. In the case where $p = 2$, the continuity assumptions of the covariance function as well as its cross derivatives over the diagonal required in [42] are also not fulfilled. Still in the case where $p = 2$, the RKHS associated to k (see the upcoming Section 4.1) is equal to $\text{Span}(f_1,...,f_n)$, which is a subspace of $H^m(\mathcal{D})$; without further assumptions on $f_1,...,f_n$, it is not a subspace of $H^{m+\varepsilon}(\mathcal{D})$ for any $\varepsilon > 0$. Thus the results from [45], Corollary 4.5, only ensure the suboptimal fact that the sample paths lie in $H^{m-d/2-\eta}(\mathcal{D})$ for all $\eta > 0$. Moreover, this result only holds under additional regularity assumptions over $\mathcal{D}$.

Proof. (Proposition 3.1) We show $(i) \implies (ii)$ & (iii), $(ii) \implies (i)$ and $(iii) \implies (ii)$.
$\underline{(i) \implies (ii) \ \& \ (iii)}$: Assume (i) and let $|\alpha| \leqslant m$. We first prove that the map $N_\alpha : (\Omega,\mathcal{F},\mathbb{P}) \to (\mathbb{R},\mathcal{B}(\mathbb{R})),\ \omega \mapsto \|\partial^\alpha U_\omega\|_{L^p(\mathcal{D})}$ is measurable. Indeed, given $\varphi \in F_q$ (see equation (2.22) for the definition of F_q), the map

$$U^\alpha_\varphi : \omega \longmapsto \int_\mathcal{D} \partial^\alpha U_\omega(x)\varphi(x)dx = (-1)^{|\alpha|}\int_\mathcal{D} U_\omega(x)\partial^\alpha\varphi(x)dx \tag{3.5}$$

is a real valued random variable (this follows from Lemma 2.5). From Lemma 2.4, one also has

$$\left(\omega \mapsto \|\partial^\alpha U_\omega\|_{L^p(\mathcal{D})}\right) = \sup_{\varphi \in F_q} |U_\varphi^\alpha|. \tag{3.6}$$

The supremum being taken over a countable set, N_α is indeed a measurable map. Given any $u \in L^p(\mathcal{D})$, a slight modification of this proof shows that $\omega \mapsto \|\partial^\alpha U_\omega - u\|_{L^p(\mathcal{D})}$ is also measurable. We can now show the map $T_\alpha : (\Omega, \mathcal{F}, \mathbb{P}) \to (L^p(\mathcal{D}), \mathcal{B}(L^p(\mathcal{D})))$, $\omega \mapsto \partial^\alpha U_\omega$ is measurable. Let $u \in L^p(\mathcal{D}), r > 0$ and $B = B(u, r)$ be an open ball in $L^p(\mathcal{D})$. Then from the measurability of $\omega \mapsto \|\partial^\alpha U_\omega - u\|_{L^p(\mathcal{D})}$,

$$T_\alpha^{-1}(B) = \{\omega \in \Omega : \|\partial^\alpha U_\omega - u\|_{L^p(\mathcal{D})} < r\} \in \mathcal{F}. \tag{3.7}$$

Since $L^p(\mathcal{D})$ is a separable metric space, its Borel σ-algebra is generated by the open balls of $L^p(\mathcal{D})$ (see e.g. [7], Exercise 6.10.28). Thus T_α is Borel-measurable and the pushforward of $\mathbb{P}$ through T_α induces a (centered) probability measure μ_α over the Banach space $L^p(\mathcal{D})$. We show that it is Gaussian. Let $v \in L^q(\mathcal{D})$ and denote T_v the associated linear form over $L^p(\mathcal{D})$. Let $(\phi_n) \subset C_c^\infty(\mathcal{D})$ be such that $\phi_n \to v$ in $L^q(\mathcal{D})$ ([1], Corollary 2.30) and $\omega \in \Omega$ be such that U_ω lies in $L^1_{loc}(\mathcal{D})$:

$$T_f(\partial^\alpha U_\omega) = \int_{\mathcal{D}} \partial^\alpha U_\omega(x) v(x) dx = \lim_{n\to\infty} \int_{\mathcal{D}} \partial^\alpha U_\omega(x)\phi_n(x) dx \tag{3.8}$$

$$= \lim_{n\to\infty} (-1)^{|\alpha|} \int_{\mathcal{D}} U_\omega(x) \partial^\alpha \phi_n(x) dx. \tag{3.9}$$

For each value of n, Lemma 2.5 shows that the map $\omega \mapsto (-1)^{|\alpha|} \int_{\mathcal{D}} U_\omega(x)\partial^\alpha\phi_n(x)dx$ is a Gaussian random variable. Thus $\omega \mapsto T_v(\partial^\alpha U_\omega)$ is a Gaussian random variable as an a.s. limit of Gaussian random variables. This shows that the pushforward of μ_α through T_v is Gaussian (see Section 2.1.1 for the pushforward), since for all Borel set $B \in \mathcal{B}(\mathbb{R})$,

$$\mu_\alpha(T_v^{-1}(B)) = \mu_\alpha(\{g \in L^p(\mathcal{D}) : T_v(g) \in B\}) = \mathbb{P}(\{\omega \in \Omega : T_v(\partial^\alpha U_\omega) \in B\}). \tag{3.10}$$

Hence, μ_α is Gaussian. We next show that $\partial^{\alpha,\alpha} k \in L^p(\mathcal{D} \times \mathcal{D})$ and that the covariance operator of μ_α is the integral operator $\mathcal{E}_k^\alpha : L^q(\mathcal{D}) \to L^p(\mathcal{D})$ with kernel $\partial^{\alpha,\alpha}k$. Let $\mathcal{D}_0 \Subset \mathcal{D} \times \mathcal{D}$ and $K_0 \Subset \mathcal{D}$ be such that $\mathcal{D}_0 \subset K_0 \times K_0$ (for example, set $K_1 := \overline{\{x \in \mathcal{D} : \exists y \in \mathcal{D}, (x, y) \in K\}}$, $K_2 := \overline{\{y \in \mathcal{D} : \exists x \in \mathcal{D}, (x, y) \in K\}}$ which are both compact subsets of $\mathcal{D}$ and $K_0 := K_1 \cup K_2$). Let $h = (h_1, ..., h_d) \in (\mathbb{R}_+^*)^d$ be such that $\sum_{i=1}^d \alpha_i h_i < \text{dist}(K_0, \mathcal{D}_0)$. Use then the bilinearity of the covariance operator:

$$\int_{\mathcal{D}_0} |(\delta_h^\alpha \otimes \delta_h^\alpha)k(x,y)|^p dxdy = \int_{\mathcal{D}_0} |\mathbb{E}[\delta_h^\alpha U(x)\delta_h^\alpha U(y)]|^p dxdy \tag{3.11}$$

$$\leqslant \int_{K_0\times K_0} |\mathbb{E}[\delta_h^\alpha U(x)\delta_h^\alpha U(y)]|^p dxdy \tag{3.12}$$

$$\leqslant \int_{K_0\times K_0} \mathbb{E}[|\delta_h^\alpha U(x)\delta_h^\alpha U(y)|^p] dxdy \tag{3.13}$$

$$\leqslant \mathbb{E}\left[\left(\int_{K_0} |\delta_h^\alpha U(x)|^p dx\right)^2\right] = \mathbb{E}\left[\|\delta_h^\alpha U\|_p^{2p}\right] \tag{3.14}$$

$$\leqslant \mathbb{E}\left[\|U\|_{W^{m,p}(\mathcal{D})}^{2p}\right] =: C^p < +\infty. \tag{3.15}$$

The expectation in equation (3.15) is indeed finite because of the following. Given $|\alpha| \leqslant m$, equation (3.6) shows that the map $\omega \mapsto \|\partial^\alpha U_\omega\|_p$ is the supremum of a Gaussian sequence which is finite a.s. by assumption; Lemma 2.6 then implies that all the moments of this supremum are finite. Writing then $\|U\|_{W^{m,p}}$ in terms of these L^p norms yields equation (3.15). To see that the control (3.15) implies that $\partial^{\alpha,\alpha}k \in L^p(\mathcal{D} \times \mathcal{D})$, we copy below the steps of equations (6.2)-(6.3)-(6.4) in the proof of Lemma 2.1. Let $\varphi \in C_c^\infty(\mathcal{D} \times \mathcal{D})$. Since it is compactly supported in $\mathcal{D} \times \mathcal{D}$, find an open set $\mathcal{D}_0 \Subset \mathcal{D}$ such that $\text{Supp}(\varphi) \subset \mathcal{D}_0$. Use Hölder's inequality and equation (3.15):

$$\left| \int_{\mathcal{D}\times\mathcal{D}} (\delta_h^\alpha \otimes \delta_h^\alpha) k(x,y)\varphi(x,y)dxdy \right| \leqslant \|(\delta_h^\alpha \otimes \delta_h^\alpha)k\|_p \|\varphi\|_q \leqslant C\|\varphi\|_q. \tag{3.16}$$

Next, use the discrete integration by parts formula,

$$\int_{\mathcal{D}\times\mathcal{D}} (\delta_h^\alpha \otimes \delta_h^\alpha) k(x,y)\varphi(x,y)dxdy = \int_{\mathcal{D}} k(x,y)(\delta_h^\alpha \otimes \delta_h^\alpha)^*\varphi(x,y)dxdy. \tag{3.17}$$

When $h \to 0$, observe that $(\delta_h^\alpha \otimes \delta_h^\alpha)^*\varphi(x,y) \to \partial^{\alpha,\alpha}\varphi(x,y)$ pointwise. Use Lebesgue's dominated convergence theorem and equation (3.16) to obtain

$$\left| \int_{\mathcal{D}\times\mathcal{D}} k(x,y)\partial^{\alpha,\alpha}\varphi(x,y)dxdy \right| \leqslant C\|\varphi\|_q, \tag{3.18}$$

which indeed shows that $\partial^{\alpha,\alpha}k \in L^p(\mathcal{D} \times \mathcal{D})$, from Riesz' lemma. We now identify K_α, the covariance operator of μ_α, in terms of $\partial^{\alpha,\alpha}k$. Let $f, g \in L^q(\mathcal{D})$ and using the density of $C_c^\infty(\mathcal{D})$ in $L^q(\mathcal{D})$ ([1], Corollary 2.30), let $(f_n), (g_n) \subset C_c^\infty(\mathcal{D})$ be two sequences such that $f_n \to f$ and $g_n \to g$, both in $L^q(\mathcal{D})$. Then (explanation below),

$$\begin{aligned}
\langle f, K_\alpha g\rangle_{L^q,L^p} &= \lim_{n\to\infty} \langle f_n, K_\alpha g_n\rangle_{L^q,L^p} &(3.19)\\
&= \lim_{n\to\infty} \int_{L^p(\mathcal{D})} \langle f_n, h\rangle_{L^q,L^p} \langle g_n, h\rangle_{L^q,L^p} d\mu_\alpha(h)\\
&= \lim_{n\to\infty} \int_\Omega \langle f_n, \partial^\alpha U_\omega\rangle_{L^q,L^p} \langle g_n, \partial^\alpha U_\omega\rangle_{L^q,L^p} d\mathbb{P}(\omega) &(3.20)\\
&= \lim_{n\to\infty} \int_\Omega \langle \partial^\alpha f_n, U_\omega\rangle_{L^q,L^p} \langle \partial^\alpha g_n, U_\omega\rangle_{L^q,L^p} d\mathbb{P}(\omega)\\
&= \lim_{n\to\infty} \int_{\mathcal{D}\times\mathcal{D}} \partial^\alpha f_n(x)\partial^\alpha g_n(y)k(x,y)dxdy &(3.21)\\
&= \lim_{n\to\infty} \int_{\mathcal{D}\times\mathcal{D}} f_n(x)g_n(y)\partial^{\alpha,\alpha}k(x,y)dxdy\\
&= \int_{\mathcal{D}\times\mathcal{D}} f(x)g(y)\partial^{\alpha,\alpha}k(x,y)dxdy = \langle f, \mathcal{E}_k^\alpha g\rangle_{L^q,L^p} &(3.22)
\end{aligned}$$

We used the sequential continuity of K_α in equation (3.19), the transfer theorem for pushforward measure integration ([7], Theorem 3.6.1) in equation (3.20) and Fubini's theorem in equation (3.21). Thus $K_\alpha = \mathcal{E}_k^\alpha$. According to Proposition 2.9, since μ_α is a Gaussian measure over $L^p(\mathcal{D})$, there exists a representative k_α of $\partial^{\alpha,\alpha}k$ in $L^p(\mathcal{D} \times \mathcal{D})$ which is the covariance function of a measurable Gaussian process. Note $\sigma_\alpha(x) = k_\alpha(x,x)^{1/2}$, then the same proposition shows that

$$\int_{\mathcal{D}} \sigma_\alpha(x)^p dx < +\infty, \tag{3.23}$$

which shows (ii). By Corollary 3.5.11 from [6], $\mathcal{E}_k^\alpha$ is nuclear and admits a symmetric non-negative representation as the one in equation (3.3). If $1 \leqslant p \leqslant 2$, then $L^p(\mathcal{D})$ is of cotype 2 and since we have shown that $\mathcal{E}_k^\alpha$ is a Gaussian covariance operator, from Proposition 2.8 there exists a bounded operator $A_\alpha : L^2(\mathcal{D}) \to L^p(\mathcal{D})$ and a trace class operator $S_\alpha : L^2(\mathcal{D}) \to L^2(\mathcal{D})$ such that $\mathcal{E}_k^\alpha = A_\alpha S_\alpha A_\alpha^*$. Introduce a Mercer decomposition of S_α (equation (2.6)): $S_\alpha = \sum_n \lambda_n^\alpha \phi_n^\alpha \otimes \phi_n^\alpha$. Use the continuity of A_α and A_α^* to obtain that $\partial^{\alpha,\alpha} k(x,y) = \sum_n \lambda_n^\alpha (A_\alpha \phi_n^\alpha)(x)(A_\alpha \phi_n^\alpha)(y)$ in $L^p(\mathcal{D} \times \mathcal{D})$, which finishes to prove (iii).
$(ii) \implies (i)$: from Proposition 2.9, let (V^α) be a centered measurable Gaussian process with covariance function k_α. Then its sample paths lie in $L^p(\mathcal{D})$ a.s. and the Gaussian measure it induces over $L^p(\mathcal{D})$ through the map $\omega \mapsto V_\omega^\alpha \in L^p(\mathcal{D})$ is the centered Gaussian measure with covariance operator $\mathcal{E}_k^\alpha$. Given $\varphi \in C_c^\infty(\mathcal{D})$, denote V_φ^α the following random variable

$$V_\varphi^\alpha : \omega \mapsto \int_{\mathcal{D}} V_\omega^\alpha(x)\varphi(x)dx. \tag{3.24}$$

From Lemma 2.5, $(V_\varphi^\alpha)_{\varphi \in F_q}$ is a Gaussian sequence. It is also centered and using Fubini's theorem to permute $\mathbb{E}$ and $\int$, we have that

$$\begin{aligned}\mathbb{E}[V_\varphi^\alpha V_\psi^\alpha] &= \int_{\mathcal{D}\times\mathcal{D}} \varphi(y)\psi(x)k_\alpha(x,y)dxdy = \int_{\mathcal{D}\times\mathcal{D}} \varphi(y)\psi(x)\partial^{\alpha,\alpha}k(x,y)dxdy \\ &= \int_{\mathcal{D}\times\mathcal{D}} \partial^\alpha\varphi(y)\partial^\alpha\psi(x)k(x,y)dxdy. \end{aligned}\tag{3.25}$$

$$\mathbb{E}[U_\varphi^\alpha U_\psi^\alpha] = \int_{\mathcal{D}\times\mathcal{D}} \partial^\alpha\varphi(y)\partial^\alpha\psi(x)k(x,y)dxdy. \tag{3.26}$$

Having the same mean and covariance, the two Gaussian sequences $(V_\varphi^\alpha)_{\varphi\in F_q}$ and $(U_\varphi^\alpha)_{\varphi\in F_q}$ have the same finite dimensional marginals. One checks in an elementary fashion that their countable suprema over F_q then have the same probability law (e.g. by showing that they have the same cumulative distribution function). Recalling from Lemma 2.4 that $\|V_\omega^\alpha\|_p = \sup_{\varphi\in F_q} |V_\varphi^\alpha(\omega)|$, we obtain that

$$1 = \mathbb{P}(\|V_\omega^\alpha\|_p < +\infty) = \mathbb{P}(\sup_{\varphi\in F_q} |V_\varphi^\alpha| < +\infty) = \mathbb{P}(\sup_{\varphi\in F_q} |U_\varphi^\alpha| < +\infty), \tag{3.27}$$

which shows that $\partial^\alpha U \in L^p(\mathcal{D})$ almost surely. This is true for all $|\alpha| \leqslant m$, which shows (i).
$(iii) \implies (ii)$: if (iii), then from either Proposition 2.7 or 2.8 depending on whether $p \leqslant 2$ or $p \geqslant 2$, there exists a Gaussian measure over $L^p(\mathcal{D})$ whose covariance operator is $\mathcal{E}_k^\alpha$ as defined in equation (3.2). Proposition 2.9 yields (ii). □

Remark 3.4 (Distributing derivatives over nuclear decompositions). In Point (iii) of Proposition 3.1, it is very tempting to distribute the cross derivative $\partial^{\alpha,\alpha}$ over the nuclear decomposition of k, thus setting $\lambda_n^\alpha = \lambda_n^0$ and $\psi_n^\alpha = \partial^\alpha \psi_n^0$. While we can show that $\partial^\alpha \psi_n \in L^p(\mathcal{D})$ (copy the proof of Lemma 4.9(i)), it is not clear whether the obtained decomposition converges in $L^p(\mathcal{D} \times \mathcal{D})$, or that it corresponds to a nuclear one, i.e. $\sum_n \lambda_n \|\partial^\alpha \psi_n\|_p^2 < +\infty$ (it is not even clear in what sense this derivative can be distributed, apart from the distributional sense). When $p = 2$, it turns out that this is true (see the upcoming Proposition 4.4): distributing derivatives on nuclear decompositions yield nuclear decompositions, as soon as nuclear decompositions of the differentiated kernel exist. The following formal computation shows that we should expect this property to hold also when $1 < p < 2$. Assume formally that the derivative can be distributed pointwise, and introduce the functions $v_\alpha(x) := \sum_n \lambda_n \partial^\alpha \psi_n(x)^2$ and $\sigma_\alpha(x) := v_\alpha(x)^{1/2}$. From

Proposition 3.1, we expect that $\|\sigma_\alpha\|_p < +\infty$. When $1 \leqslant p < 2$, the *reverse* Minkowski inequality in $L^{p/2}(\mathcal{D})$ (see [1], Theorem 2.13 p. 28) then yields

$$\sum_{n=0}^{+\infty} \lambda_n \|\partial^\alpha \psi_n\|_p^2 = \sum_{n=0}^{+\infty} \lambda_n \|\partial^\alpha \psi_n^2\|_{p/2} \leqslant \left\| \sum_{n=0}^{+\infty} \lambda_n \partial^\alpha \psi_n^2 \right\|_{p/2} = \|v_\alpha\|_{p/2} = \|\sigma_\alpha\|_p^2 < +\infty, \tag{3.28}$$

so that the series $\sum_n \lambda_n \|\partial^\alpha \psi_n\|_p^2$ converges. From this, it is then readily checked that the equality $\partial^{\alpha,\alpha} k = \sum_n \lambda_n \partial^\alpha \psi_n \otimes \partial^\alpha \psi_n$ holds in $L^p(\mathcal{D} \times \mathcal{D})$, which is then a nuclear decomposition of $\partial^{\alpha,\alpha} k$. Recall however that in spaces of cotype 2, it is not sufficient to require that a self-adjoint, nonnegative operator be nuclear for it to be a Gaussian covariance operator (Proposition 2.8). When $p > 2$ though, the following proposition shows that this property does not hold anymore: distributing derivatives on nuclear decompositions does not, in general, yield nuclear decompositions, even if nuclear decompositions of the differentiated kernel exist.

Proposition 3.5. *Let $p > 2$ and $\mathcal{D} = (0,1)$. There exists an explicit covariance function $k : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ of a measurable Gaussian process with the following properties.*

(i) There exists a countable set $\mathcal{N} \subset \mathcal{D} \times \mathcal{D}$ such that for all $(x,y) \in (\mathcal{D} \times \mathcal{D}) \backslash \mathcal{N}$, the classical derivative $\partial_x \partial_y k(x,y)$ exists. Moreover, the map $(x,y) \mapsto (\partial_x \partial_y k)(x,y)$ can be extended over $\mathcal{D} \times \mathcal{D}$ as a function (still denoted $\partial_x \partial_y k$) which is the covariance function of a measurable Gaussian process.

(ii) $\int_0^1 k(x,x)^{p/2} dx < +\infty$ and $\int_0^1 (\partial_x \partial_y k)(x,x)^{p/2} dx < +\infty$, hence nuclear decompositions of both k and $\partial_x \partial_y k$ exist in $L^p(\mathcal{D} \times \mathcal{D})$, from Propositions 2.7 and 2.9 (note that those two integral equations imply that both k and $\partial_x \partial_y k$ lie in $L^p(\mathcal{D} \times \mathcal{D})$, from the inequality $|h(x,y)|^p \leqslant h(x,x)^{p/2} h(y,y)^{p/2}$ if h is positive definite, so that Proposition 2.9 indeed applies).

(iii) There exists $(\psi_n) \subset W^{1,p}(\mathcal{D})$ such that $k = \sum_{n=0}^{\infty} \psi_n \otimes \psi_n$ in $L^p(\mathcal{D} \times \mathcal{D})$, $\sum_{n=0}^{+\infty} \|\psi_n\|_p^2 < +\infty$ but $\sum_{n=0}^{+\infty} \|\psi_n'\|_p^2 = +\infty$. Hence (ψ_n) provides a nuclear decomposition of k, while (ψ_n') does not for $\partial_x \partial_y k$ (even though such decompositions exist!).

The proof of Proposition 3.5 is deferred to Section 6. Note that a general procedure for obtaining nuclear decompositions of Gaussian covariance operators (hence of suitably integrable covariance functions in the case of L^p spaces), based on a gradual "orthonormalization" procedure, can be found in the proof of Theorem 3.5.10 of [6].

The following proposition deals with the apparent asymmetry in p between Points (*ii*) and (*iii*) of Proposition 3.1. We recall that the nuclear norm $\nu(T)$ is defined in equation (2.8). Contrarily to Proposition 3.1, we do not exclude $p = 1$.

Proposition 3.6. *Let μ be a centered Gaussian measure over $L^p(\mathcal{D})$, where $1 \leqslant p < +\infty$. Let $k \in L^p(\mathcal{D} \times \mathcal{D})$ be the kernel of its covariance operator ($K_\mu = \mathcal{E}_k$), chosen such that k is also the covariance function of a measurable Gaussian process $(U(x))_{x \in \mathcal{D}}$, from Proposition 2.9. Define $\sigma(x) = k(x,x)^{1/2}$ and set $C_p = 2^{p/2} \Gamma((p+1)/2)/\sqrt{\pi}$ ($C_p = \mathbb{E}[|X|^p]$ where $X \sim \mathcal{N}(0,1)$). Then the following bounds hold.*

- *if $1 \leqslant p < 2$, there exists a symmetric, nonnegative and trace class operator S over $L^2(\mathcal{D})$ and a bounded operator $A : L^2(\mathcal{D}) \to L^p(\mathcal{D})$ such that $\mathcal{E}_k = ASA^*$. Moreover,*

$$\nu(\mathcal{E}_k) \leqslant \inf_{\substack{A,S \ s.t. \\ \mathcal{E}_k = ASA^*}} \|A\|^2 \nu(S) \leqslant \|\sigma\|_p^2 \leqslant C_p^{-2/p} \inf_{\substack{A,S \ s.t. \\ \mathcal{E}_k = ASA^*}} \|A\|^2 \nu(S). \tag{3.29}$$

- *if* $2 \leqslant p < +\infty$*, then* $\mathcal{E}_k$ *is symmetric, nonnegative and nuclear, and*

$$C_p^{-2/p}\nu(\mathcal{E}_k) \leqslant \|\sigma\|_p^2 \leqslant \nu(\mathcal{E}_k). \tag{3.30}$$

Observe that if $p = 2$ then $C_2 = 1$ and equation (3.30) yields $\|\sigma\|_2^2 = \nu(\mathcal{E}_k) = \mathrm{Tr}(\mathcal{E}_k)$. It is expected that the nuclear norm of $\mathcal{E}_k$ cannot directly appear on the right hand side of equation (3.29), as not all nuclear operators are Gaussian covariance operators when $1 \leqslant p < 2$ (Proposition 2.8). Proposition 3.6 in fact suggests that for general Banach spaces X of cotype 2, the following map defined over the set of Gaussian covariance operators $B : X^* \to X$,

$$B \mapsto \inf_{\substack{A,S \ s.t. \\ B=ASA^*}} \|A\|^2\nu(S), \tag{3.31}$$

is the natural measurement of the "size" of such operators. When X is of type 2, this would be the case for the nuclear norm $B \mapsto \nu(B)$.

Remark 3.7. Proposition 3.6 is interesting from an application point of view because it states that the operator norms appearing in this proposition, as well as the L^p norm of the standard deviation function σ, are suitable quantities for quantitatively controlling the L^p norm of the sample paths of the underlying Gaussian process. Explicitly, we have the following L^p control in expectation: $\mathbb{E}[\|U\|_p^p] = C_p\|\sigma\|_p^p$ (see equation (3.33)). Applying this fact recursively, we obtain that the $W^{m,p}$-Sobolev norm of the sample paths of the Gaussian process in question is controlled as follow, denoting $\sigma_\alpha(x) = \partial^{\alpha,\alpha}k(x,x)^{1/2}$ (choosing a representative of $\partial^{\alpha,\alpha}k$ which is the covariance of a measurable Gaussian process)

$$\mathbb{E}\big[\|U\|_{W^{m,p}}^p\big] = C_p \sum_{|\alpha| \leqslant m} \|\sigma_\alpha\|_p^p. \tag{3.32}$$

If such a control cannot be obtained, then it means that the sample paths of U do not lie in $W^{m,p}(\mathcal{D})$ in the first place. Finally, we have the following asymptotic behaviour of the constant when $p \to +\infty : C_p^{-2/p} \sim \exp(1)/(p-1)$.

Proof. (Proposition 3.6) We begin with the following general fact concerning the measurable Gaussian process $(U(x))_{x\in\mathcal{D}}$, observing from Fubini's theorem that

$$\mathbb{E}[\|U\|_p^p] = \mathbb{E}\left[\int_{\mathcal{D}} |U(x)|^p dx\right] = \int_{\mathcal{D}} \mathbb{E}[|U(x)|^p]dx = \int_{\mathcal{D}} C_p\sigma(x)^p dx = C_p\|\sigma\|_p^p, \tag{3.33}$$

where $C_p = 2^{p/2}\Gamma((p+1)/2)/\sqrt{\pi}$. Indeed, given $X \sim \mathcal{N}(0,\sigma^2)$, then $\mathbb{E}[|X|^p] = C_p\sigma^p$.

Suppose now that $1 \leqslant p < 2$. Let μ_0 be a Gaussian measure on $L^2(\mathcal{D})$ and $A : L^2(\mathcal{D}) \to L^p(\mathcal{D})$ a bounded operator such that $\mu = {\mu_0}_A$ (pushforward of μ_0 through A, see Section 2.1.1) and S the trace class covariance operator associated to μ_0 (see Proposition 2.8). Recall also that from Proposition 2.9, $\mu = \mathbb{P}_U$. Then (explanation below),

$$C_p\|\sigma\|_p^p = \mathbb{E}[\|U\|_p^p] = \int_\Omega \|U_\omega\|_p^p \mathbb{P}(d\omega) = \int_{L^p(\mathcal{D})} \|f\|_p^p \mu(df) \tag{3.34}$$

$$= \int_{L^2(\mathcal{D})} \|Ag\|_p^p \mu_0(dg) \leqslant \|A\|^p \int_{L^2(\mathcal{D})} \|g\|_2^p \mu_0(dg) \tag{3.35}$$

$$\leqslant \|A\|^p \int_{L^2(\mathcal{D})} \langle g,g\rangle_{L^2}^{p/2} \mu_0(dg) \leqslant \|A\|^p \left(\int_{L^2(\mathcal{D})} \langle g,g\rangle_{L^2} \mu_0(dg)\right)^{p/2} \tag{3.36}$$

$$\leqslant \|A\|^p \ \mathrm{Tr}(S)^{p/2} = \|A\|^p\nu(S)^{p/2}. \tag{3.37}$$

In equation (3.34), we used equation (3.33) and pushforward integration to write the integral w.r.t. $\mathbb{P}$ as an integral w.r.t. $\mu = \mathbb{P}_U$. Likewise in equation (3.35) where we write the integral w.r.t. μ as an integral w.r.t. μ_0 using the pushforward identity $\mu = {\mu_0}_A$. In equation (3.36), we used Jensen's inequality for concave functions ($0 < p/2 < 1$). In equation (3.37), we used the trace identity for Gaussian measures over Hilbert spaces from [6], equation 2.3.2 and the one following p. 49. Moreover, from the nuclear norm estimate of [48], Proposition 47.1 pp. 479-480,

$$\nu(\mathcal{E}_k) = \nu(ASA^*) \leqslant \|A\|\nu(S)\|A^*\| \leqslant \|A\|^2\nu(S) \tag{3.38}$$

. In equations (3.37) and (3.38), taking the infimum over all representations $\mathcal{E}_k = ASA^*$ yields

$$\nu(\mathcal{E}_k) \leqslant \inf_{\substack{A,S\ s.t.\\ \mathcal{E}_k=ASA^*}} \|A\|^2\nu(S), \quad \|\sigma\|_p^2 \leqslant C_p^{-\frac{2}{p}} \inf_{\substack{A,S\ s.t.\\ \mathcal{E}_k=ASA^*}} \|A\|^2\nu(S). \tag{3.39}$$

To prove the remaining inequality ($\inf_{\mathcal{E}_k=ASA^*} \|A\|^2\nu(S) \leqslant \|\sigma\|_p^2$), we use an explicit decomposition $\mathcal{E}_k = ASA^*$ by first setting

$$Af(x) = f(x)\sigma(x)^{1-p/2}. \tag{3.40}$$

Using Hölder's inequality with $a = 2/p,\ 1/a + 1/b = 1$ (notice that $a > 1$), we obtain

$$\|Af\|_p^p = \int_{\mathcal{D}} |f(x)|^p\sigma(x)^{p(1-p/2)}dx \tag{3.41}$$

$$\leqslant \left(\int_{\mathcal{D}} |f(x)|^2dx\right)^{p/2}\left(\int_{\mathcal{D}} \sigma(x)^{bp(1-p/2)}dx\right)^{1/b}. \tag{3.42}$$

But $b = \frac{a}{a-1} = \frac{2/p}{2/p-1} = \frac{1}{1-p/2}$ and $b(1-p/2) = 1$, which together with equation (3.41) yields

$$\|Af\|_p^p \leqslant \|f\|_2^p\|\sigma\|_p^{p(1-p/2)}. \tag{3.43}$$

Thus $A : L^2(\mathcal{D}) \to L^p(\mathcal{D})$ is bounded and $\|A\| \leqslant \|\sigma\|_p^{1-p/2}$. One also verifies that $A^* : L^q(\mathcal{D}) \to L^2(\mathcal{D})$ is given by $A^*f(x) = f(x)\sigma(x)^{1-p/2}$, with $\|A\| = \|A^*\|$. Introduce the functions $k_0(x,y) := k(x,y)\sigma(x)^{p/2-1}\sigma(y)^{p/2-1}$, and $\sigma_0(x) = k_0(x,x)^{1/2}$; k_0 is the covariance function of the measurable Gaussian process $V(x) := \sigma(x)^{p/2-1}U(x)$, and verifies

$$\|\sigma_0\|_2^2 = \int_{\mathcal{D}} \sigma_0(x)^2dx = \int_{\mathcal{D}} k_0(x,x)dx = \int_{\mathcal{D}} \sigma(x)^pdx = \|\sigma\|_p^p < +\infty. \tag{3.44}$$

Therefore $\mathcal{E}_{k_0}$, the integral operator over $L^2(\mathcal{D})$ associated to k_0, is trace class (Proposition 3.1(ii)). Observe also that $k = (A \otimes A)k_0$ which also yields that $\mathcal{E}_k = A\mathcal{E}_{k_0}A^*$. Thus,

$$\inf_{\substack{A,S\ s.t.\\ \mathcal{E}_k=ASA^*}} \|A\|^2\nu(S) \leqslant \|A\|^2\nu(\mathcal{E}_{k_0}) \leqslant \|\sigma\|_p^{2-p}\|\sigma\|_p^p = \|\sigma\|_p^2. \tag{3.45}$$

Combining equations (3.39) and (3.45) yields the desired result of equation (3.29).

Suppose now that $p \geqslant 2$. Recall that $\mu = \mathbb{P}_U$. We successively use the transfer theorem for pushforward measure integration, Jensen's inequality for probability measures ($p/2 \geqslant 1$) and the nuclear norm estimate from [32], Theorem 3:

$$\mathbb{E}[\|U\|_p^p] = \int_{\Omega} \|U_\omega\|_p^p\mathbb{P}(d\omega) = \int_{L^p(\mathcal{D})} \|f\|_p^p\mu(df) = \int_{L^p(\mathcal{D})} \|f\|_p^{2\times p/2}\mu(df) \tag{3.46}$$

$$\geqslant \left(\int_{L^p(\mathcal{D})} \|f\|_p^2\mu(df)\right)^{p/2} \geqslant \nu(\mathcal{E}_k)^{p/2}, \tag{3.47}$$

which together with equation (3.33) yields $\|\sigma\|_p^2 \geqslant C_p^{-2/p}\nu(\mathcal{E}_k)$. We now prove the last remaining inequality, i.e. $\|\sigma\|_p^2 \leqslant \nu(\mathcal{E}_k)$. For this, consider $k(x,y) = \sum_n \mu_n\psi_n(x)\phi_n(y)$, a nuclear representation of k in $L^p(\mathcal{D}\times\mathcal{D})$, with $\|\psi_n\|_p = \|\phi_n\|_p = 1$ and $\sum_n |\mu_n| < +\infty$. Denote by v the function $v : x \mapsto \sum_{n=0}^{+\infty}\mu_n\psi_n(x)\phi_n(x)$. Minkowski's inequality in $L^{p/2}(\mathcal{D})$ shows that $x \mapsto \sum_{n=0}^{+\infty}|\mu_n\psi_n(x)\phi_n(x)|$ is finite a.e. and in fact that $v \in L^{p/2}(\mathcal{D})$:

$$\begin{aligned}\|v\|_{p/2} = \left\|\sum_{n=0}^{+\infty}\mu_n\psi_n\phi_n\right\|_{p/2} &\leqslant \sum_{n=0}^{+\infty}|\mu_n| \times \left\|\psi_n\phi_n\right\|_{p/2} \\ &\leqslant \sum_{n=0}^{+\infty}|\mu_n| \times \|\psi_n\|_p\|\phi_n\|_p = \sum_{n=0}^{+\infty}|\mu_n| < +\infty. \end{aligned} \tag{3.48}$$

In equation (3.48) above, we used used the Cauchy-Schwarz inequality on $\|\phi_n\psi_n\|_{p/2}$. From the nuclear decomposition of k, it is very tempting to write $\|\sigma\|_p^2 = \|v\|_{p/2}$, but unfortunately the diagonal of $\mathcal{D}\times\mathcal{D}$ has a null Lebesgue measure. This equality turns out to be true but this fact is non trivial and deferred to Lemma 3.8 below. From this lemma and equation (3.48) which holds whatever the nuclear decomposition of $\mathcal{E}_k$, taking the infimum over all nuclear representations of $\mathcal{E}_k$ in equation (3.48) yields $\|\sigma\|_p^2 \leqslant \nu(\mathcal{E}_k)$. This finishes the proof. $\square$

The next lemma, which was key in the proof of equation (3.30), states that evaluating the $L^{p/2}$-norm of the diagonal of a nuclear representation of a Gaussian covariance operator K in $L^p(\mathcal{D}), p \geqslant 2$, yields the same result as evaluating $L^{p/2}$-norm of the diagonal of the covariance function k of any measurable Gaussian process $(U(x))_{x\in\mathcal{D}}$ such that $\mathcal{E}_k = K$. This fact is not obvious at all, as the diagonal of $\mathcal{D}\times\mathcal{D}$ has null Lebesgue measure and different representatives of k in $L^p(\mathcal{D}\times\mathcal{D})$ have no reason a priori to agree on sets of null measure. However, the assumptions that the representation is nuclear and that U is measurable turn out to be strong enough to yield the desired conclusion. The proof ideas for this result should largely be credited to [10]; we generalized them in a straightforward fashion from $L^2(\mathcal{D})$ to $L^p(\mathcal{D})$ and applied them to the Gaussian process $(U(x))_{x\in\mathcal{D}}$ of Proposition 3.6. They are based on the Hardy-Littlewood maximal inequality.

Lemma 3.8. *Let $2 \leqslant p < +\infty$, $\mathcal{D} \subset \mathbb{R}^d$ be an open set and $(U(x))_{x\in\mathcal{D}} \sim GP(0,k)$ be a measurable Gaussian process whose sample paths lie in $L^p(\mathcal{D})$ a.s.. From Propositions 2.9 and 2.7, $\mathcal{E}_k : L^q(\mathcal{D}) \to L^p(\mathcal{D})$ is nuclear and there exists sequences $(\mu_n) \subset \mathbb{R}, (\psi_n), (\phi_n) \subset L^p(\mathcal{D})$ such that $k = \sum_n \mu_n\psi_n \otimes \phi_n$ in $L^p(\mathcal{D}\times\mathcal{D})$, with $\|\psi_n\|_p = \|\phi_n\|_p = 1$ and $\sum_n|\mu_n| < +\infty$. Then $x \mapsto \sum_{n=0}^{\infty}|\mu_n\psi_n(x)\phi_n(x)|$ is finite a.e. and $v : x \mapsto \sum_{n=0}^{\infty}\mu_n\psi_n(x)\phi_n(x)$ is nonnegative a.e.. Moreover,*

$$\|\sigma\|_p^p = \int_{\mathcal{D}} k(x,x)^{p/2}dx = \int_{\mathcal{D}}\left(\sum_{n=0}^{+\infty}\mu_n\psi_n(x)\phi_n(x)\right)^{p/2}dx = \|v\|_{p/2}^{p/2}. \tag{3.49}$$

A remarkable consequence of this result is that the $L^{p/2}$-norm of the diagonal of a nuclear representation of $\mathcal{E}_k = \sum_n \mu_n\psi_n\otimes\phi_n$ is *invariant* w.r.t. said nuclear decomposition, while its finiteness fully characterizes the nuclearity of $\mathcal{E}_k$ (Proposition 3.6(*ii*)); the same invariance property does not hold for $\sum_n|\mu_n|$, hence the need to define the nuclear norm of $\mathcal{E}_k$ as the infimum over such quantities.

Proof of Lemma 3.8. We first prove the statement when $\mathcal{D} = \mathbb{R}^d$. We begin with some definitions and observations. For $r > 0$, denote $C_r := [-r,r]^d$ and $C_r(x) := x + C_r$. For $f \in L^p(\mathbb{R}^d)$

(resp. $g \in L^p(\mathbb{R}^d \times \mathbb{R}^d)$), denote its average over $C_r(x)$ (resp. $C_r(x) \times C_r(x)$) as

$$A_r^{(d)} f(x) := \frac{1}{|C_r|} \int_{C_r(x)} f(t)dt, \quad A_r^{(2d)} g(x) := \frac{1}{|C_r|^2} \int_{C_r(x)} \int_{C_r(x)} g(s,t) ds dt \tag{3.50}$$

The functions $A_r^{(d)} f$ and $A_r^{(2d)} g$ are defined pointwise and continuous. The point of averaging over cubes rather than balls is that we have $A_r^{(2d)} = A_r^{(d)} \otimes A_r^{(d)}$. One then introduces the Hardy-Littlewood maximal functions of f and g, as

$$M^{(d)} f(x) := \sup_{r>0} \frac{1}{|C_r|} \int_{C_r(x)} |f(t)|dt, \ M^{(2d)} g(x,y) := \sup_{r>0} \frac{1}{|C_r|^2} \int_{C_r(x)} \int_{C_r(x)} |g(s,t)| ds dt.$$

$M^{(d)} f$ (resp. $M^{(2d)} g$) is measurable, nonnegative and defined pointwise over $\mathbb{R}^d$ (resp. $\mathbb{R}^d \times \mathbb{R}^d$). For all $x \in \mathbb{R}^d$, we obviously have the pointwise majoration

$$|A_r^{(d)} f(x)| \leqslant M^{(d)} f(x), \tag{3.51}$$

and likewise for $M^{(2d)} g$. A key point for us will be the Hardy-Littlewood maximal theorem ([44], Theorem 1 p. 5), which states that there exists a constant $S_p > 0$ such that for all $f \in L^p(\mathbb{R}^d)$,

$$\|M^{(d)} f\|_p \leqslant S_p \|f\|_p. \tag{3.52}$$

This theorem allows a first general observation, given $f \in L^p(\mathbb{R}^d)$. Indeed, the Lebesgue differentiation theorem ([44], Corollary 1 p. 5) states that $A_r^{(d)} f(x) \to f(x)$ a.e.; but we also have the pointwise domination

$$|A_r^{(d)} f(x) - f(x)| \leqslant |A_r^{(d)} f(x)| + |f(x)| \leqslant M^{(d)} f(x) + |f(x)| \quad \text{a.e..} \tag{3.53}$$

From equation (3.52), the function on the right-hand side of equation (3.53) lies in $L^p(\mathbb{R}^d)$ and Lebesgue's dominated convergence theorem in $L^p(\mathcal{D})$ yields that we also have the convergence

$$\|A_r^{(d)} f - f\|_{L^p(\mathbb{R}^d)} \xrightarrow[r \to 0]{} 0. \tag{3.54}$$

We will also use that the nonlinear operator M is submutliplicative and subadditive:

$$M^{(2d)}(\psi \otimes \varphi)(x,y) \leqslant M^{(d)} \psi(x) M^{(d)} \varphi(y), \tag{3.55}$$

$$M^{(d)}(\psi + \varphi)(x) \leqslant M^{(d)} \psi(x) + M^{(d)} \varphi(x). \tag{3.56}$$

With equations (3.54), (3.55) and (3.56), we now prove the desired result. We first focus on the decomposition $k = \sum_n \mu_n \psi_n \otimes \phi_n$, for which the following pointwise equality holds ([10], Corollary 2.2 and Lemma 2.3, or equation 3.6 from [10])

$$A_r^{(2d)} k(x,y) = \sum_{n=0}^{+\infty} \mu_n A_r^{(d)} \psi_n(x) A_r^{(d)} \phi_n(y), \quad \forall (x,y) \in \mathbb{R}^d \times \mathbb{R}^d. \tag{3.57}$$

We now prove that from this decomposition, we can deduce a first important fact, which is

$$\lim_{r \to 0} A_r^{(2d)} k(x,x) = \sum_{n=0}^{+\infty} \mu_n \psi_n(x) \phi_n(x) \quad \text{a.e..} \tag{3.58}$$

For this, first observe that for all $x \in \mathbb{R}^d$ and $n \in \mathbb{N}_0$, the following domination holds:

$$|\mu_n| \times |A_r^{(d)}\psi_n(x)A_r^{(d)}\phi_n(x)| \leqslant |\mu_n| \times M^{(d)}\psi_n(x)M^{(d)}\phi_n(x). \tag{3.59}$$

But the series obtained by summing the right-hand side term of equation (3.59) is an a.e. finite function of x, as Minkowski's inequality in $L^{p/2}(\mathbb{R}^d)$ and equation (3.52) yield:

$$\begin{aligned}\Big\| \sum_{n=0}^{+\infty} |\mu_n| M^{(d)}\psi_n M^{(d)}\phi_n \Big\|_{p/2} &\leqslant \sum_{n=0}^{+\infty} |\mu_n| \times \big\| M^{(d)}\psi_n M^{(d)}\phi_n \big\|_{p/2} \\ &\leqslant \sum_{n=0}^{+\infty} |\mu_n| \times \big\| M^{(d)}\psi_n \big\|_p \big\| M^{(d)}\phi_n \big\|_p \qquad (3.60)\\ &\leqslant \sum_{n=0}^{+\infty} |\mu_n| \times S_p^2 \|\psi_n\|_p \|\phi_n\|_p = S_p^2 \sum_{n=0}^{+\infty} |\mu_n| < +\infty. \qquad (3.61)\end{aligned}$$

We used the Cauchy-Schwarz inequality in equation (3.60). Choose now a conull set $T \subset \mathbb{R}^d$, on which the Lebesgue differentiation theorem applies for all ψ_n and ϕ_n, and on which $x \mapsto \sum_n |\mu_n| M^{(d)}\psi_n(x)M^{(d)}\phi_n(x)$ is finite (such a set exists from the finiteness of its $L^{p/2}$-norm). For all $x \in T$, the Lebesgue dominated convergence theorem for the discrete measure $\sum_{n\in\mathbb{N}_0} \delta_n$ (using the domination (3.59)) yields the equality (3.58).

We now focus on the Gaussian process $(U(x))_{x\in\mathbb{R}^d}$. Since its sample paths U_ω lie in $L^p(\mathbb{R}^d)$ almost surely, equation (3.54) yields that for almost every $\omega \in \Omega$,

$$\|A_r^{(d)}U_\omega - U_\omega\|_p^p \xrightarrow[r\to 0]{} 0. \tag{3.62}$$

We also have that for every such $\omega \in \Omega$ and $r > 0$,

$$\|A_r^{(d)}U_\omega - U_\omega\|_p \leqslant \|A_r^{(d)}U_\omega\|_p + \|U_\omega\|_p \leqslant \|M^{(d)}U_\omega\|_p + \|U_\omega\|_p \leqslant (S_p+1)\|U_\omega\|_p, \tag{3.63}$$

and from Fubini's theorem, the right-hand side of equation (3.63) lies in $L^p(\mathbb{P})$:

$$\mathbb{E}[\omega \mapsto \|U_\omega\|_p^p] = \mathbb{E}[\|U\|_p^p] = \int_{\mathbb{R}^d} \mathbb{E}[|U(x)|^p]dx = C_p\|\sigma\|_p^p < +\infty. \tag{3.64}$$

Thus, from equations (3.62), (3.63), (3.64) and Lebesgue's dominated convergence in $L^p(\mathbb{P})$,

$$\mathbb{E}[\|A_r^{(d)}U - U\|_p^p] \xrightarrow[r\to 0]{} 0. \tag{3.65}$$

In particular, using the reverse triangle inequality on the norm $V \mapsto \mathbb{E}[\|V\|_p^p]^{1/p}$, we have

$$\mathbb{E}\big[\|A_r^{(d)}U\|_p^p\big] \xrightarrow[r\to 0]{} \mathbb{E}\big[\|U\|_p^p\big] = C_p\|\sigma\|_p^p. \tag{3.66}$$

We then wish to use equations (3.66) and (3.58) to prove the desired result. For this, observe that from the linearity of the operator $A_r^{(d)}$, $(A_r^{(d)}U(x))_{x\in\mathcal{D}}$ is a centered measurable Gaussian process whose covariance function is given by

$$\mathrm{Cov}(A_r^{(d)}U(x), A_r^{(d)}U(y)) = \big(A_r^{(d)} \otimes A_r^{(d)}\big)k(x,y) = A_r^{(2d)}k(x,y), \quad \forall (x,y) \in \mathbb{R}^d \times \mathbb{R}^d \tag{3.67}$$

(Note then that $A_r^{(2d)}k(x,x) = \mathrm{Var}(A_r^{(d)}U(x)) \geqslant 0$, which also shows that the limit in equation (3.58) is nonnegative a.e.) The proof of the Gaussianity of $(A_r^{(d)}U(x))_{x\in\mathcal{D}}$ is carried out similarly as for Lemma 2.5, and the expression of its covariance function follows from the measurability of U and Fubini's theorem. Fubini's theorem and the fact that $\mathbb{E}[|X|^p] = C_p s^p$ if $X \sim \mathcal{N}(0, s^2)$ then lead to

$$\mathbb{E}[\|A_r^{(d)} U\|_p^p] = \int_{\mathbb{R}^d} \mathbb{E}[|A_r^{(d)} U(x)|^p] dx = C_p \int_{\mathbb{R}^d} \left(A_r^{(2d)} k(x,x)\right)^{p/2} dx. \tag{3.68}$$

We will finally apply Lebesgue's dominated convergence theorem on equation (3.68) when r goes to zero, using the limit given in equation (3.58). For this, observe that equation (3.51) together with the sublinear properties of $M^{(d)}$ (equations (3.55) and (3.56)) lead to the domination

$$|A_r^{(2d)} k(x,x)| \leqslant M^{(2d)} k(x,x) \leqslant \sum_{n=0}^{+\infty} |\mu_n| M^{(d)} \psi_n(x) M^{(d)} \phi_n(x) \quad \forall x \in \mathbb{R}^d, \tag{3.69}$$

and the right-hand side of equation (3.69) indeed lies in $L^{p/2}(\mathcal{D})$, from equation (3.61). We finally conclude from Lebesgue's dominated convergence theorem that

$$\lim_{r \to 0} \mathbb{E}[\|A_r^{(d)} U\|_p^p] = C_p \int_{\mathbb{R}^d} \lim_{r \to 0} \left(A_r^{(2d)} k(x,x)\right)^{p/2} dx = C_p \int_{\mathbb{R}^d} \left(\sum_{n=0}^{+\infty} \mu_n \psi_n(x) \phi_n(x)\right)^{p/2} dx,$$

which, together with equation (3.66), finishes the proof.

To deal with the general case where $\mathcal{D}$ is only an open subset of $\mathbb{R}^d$, extend any function $f \in L^p(\mathcal{D})$ to a function $\tilde{f} \in L^p(\mathbb{R}^d)$ by setting $\tilde{f}(x) = f(x)$ if $x \in \mathcal{D}$, $\tilde{f}(x) = 0$ elsewhere. $\tilde{f}$ remains measurable, and all the arguments and results stated above are preserved. □

4 Sobolev regularity for Gaussian processes : the Hilbert space case, $p = 2$

In the case $p = 2$, we provide an alternative proof of the integral and spectral criteria of Proposition 3.1, based on the study of the "ellipsoids" of Hilbert spaces (see Section 4.2). These geometrical objects are well understood in relation with Gaussian processes (see [21] or [46], Section 2.5). Compared with the general case $p \in (1, +\infty)$, we draw additional links between the different Mercer decompositions of the kernels $\partial^{\alpha,\alpha} k$, the evaluation of the trace of $\mathcal{E}_k^\alpha$ and the Hilbert-Schmidt nature of the imbedding of the reproducing kernel Hilbert space (see Section 4.1 below) associated to k in $H^m(\mathcal{D})$.

4.1 Reproducing Kernel Hilbert Spaces (RKHS, [5])

Consider a general set $\mathcal{D}$ and a positive definite function $k : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$, i.e. such that given any $n \in \mathbb{N}$ and $(x_1, ..., x_n) \in \mathcal{D}^n$, the matrix $(k(x_i, x_j))_{1 \leqslant i,j \leqslant n}$ is nonnegative definite. One can then build a Hilbert space H_k of functions defined over $\mathcal{D}$ which contains the functions $k(x, \cdot), x \in \mathcal{D}$ and verifies the *reproducing* identities

$$\langle k(x,\cdot), k(x',\cdot) \rangle_{H_k} = k(x,x') \qquad \forall x, x' \in \mathcal{D}, \tag{4.1}$$

$$\langle k(x,\cdot), f \rangle_{H_k} = f(x) \qquad \forall x \in \mathcal{D},\ \forall f \in H_k. \tag{4.2}$$

H_k is the RKHS of k. This space is exactly the set of functions of the form $f(x) = \sum_{i=1}^{+\infty} a_i k(x_i, x)$ such that $\|f\|_{H_k}^2 = \sum_{i,j=1}^{+\infty} a_i a_j k(x_i, x_j) < +\infty$. If for all $x \in \mathcal{D}$, $k(x, \cdot)$ is measurable, then H_k only contains measurable functions. One may then consider imbedding H_k in some Sobolev space $H^m(\mathcal{D})$. Recall that in $H^m(\mathcal{D})$, functions are equal up to a set of Lebesgue measure zero. If such an imbedding $i : H_k \to H^m(\mathcal{D})$ is well-defined (i.e. if $f \in H_k$ then its weak derivatives $\partial^\alpha f$ exist and lie in $L^2(\mathcal{D})$ for all $|\alpha| \leqslant m$), we will sometimes use the same notation for $f \in H_k$ and its equivalence class $f \in H^m(\mathcal{D})$; strictly speaking, the latter should be denoted $i(f)$. It may then happen that i is not injective, as with the RKHS associated to the Kronecker delta $k(x,x') = \delta_{x,x'}$ (in this case, we even have $i(H_k) = \{0\}$).

Remark 4.1. In Proposition 4.4, we will be interested in the Hilbert-Schmidt nature of the imbedding i. However, it may happen that H_k is not separable, such as with the RKHS associated to the Kronecker delta $\delta_{x,x'}$. This results in additional care required for defining the notion of Hilbert Schmidt operators, as the definition from Section 2.1.3(ii) cannot hold. Still, this case is dealt with in Proposition 4.4(iv). See [36] and [6], Remark 3.2.9 p. 103 for discussions on non separable RKHS.

4.2 Ellipsoids of Hilbert spaces and canonical Gaussian processes [21]

Let $(H; \langle , \rangle_H)$ be a separable Hilbert space. We introduce $(V_x)_{x\in H}$ the canonical Gaussian process of H, defined as the centered Gaussian process whose covariance function is the inner product of H :

$$\mathbb{E}[V_x V_y] = \langle x, y \rangle_H. \tag{4.3}$$

A subset K of H is said to be Gaussian bounded (GB) if

$$\mathbb{P}(\sup_{x\in K} |V_x| < +\infty) = 1. \tag{4.4}$$

The GB property was first introduced for studying the compact sets of Hilbert spaces, see [21] on that topic. In equation (4.4), the random variable is defined as $\sup_{x\in K} |V_x| := \sup_{x\in A} |V_x|$ where A is any countable subset of K, dense in K. Different choices of A only modify $\sup_{x\in K} |V_x|$ on a set of probability 0 ([21], p. 291), which leaves equation (4.3) unchanged. We will use the two following results below, taken from [21].

Proposition 4.2 ([21], p. 293 and [21], Proposition 3.4)**.** *We have the two following facts.*

(i) If K is a GB-set, then its closed, convex, symmetric hull is a GB-set.

(ii) The closure of a GB-set is compact.

Given a self-adjoint compact operator $T : H \to H$, introduce a basis of eigenvectors x_n and its real eigenvalues λ_n, $\lambda_n \to 0$. The image of the closed unit ball of H, $B = B_H(0,1)$ is the following "ellipsoid" ([21], p. 312)

$$T(B) = \Big\{ \sum_{\lambda_n>0} a_n x_n \ s.t. \sum_{\lambda_n>0} a_n^2/\lambda_n^2 \leqslant 1 \Big\}. \tag{4.5}$$

The main result we will use is the following.

Proposition 4.3 ([21], Proposition 6.3)**.** *Suppose that T is compact and self-adjoint. Then $T(B)$ is a GB-set if and only if $\sum_{n\in\mathbb{N}} \lambda_n^2 < \infty$, i.e. $T(B)$ is a "Schmidt ellipsoid".*

We can now state our result pertaining to the $H^m(\mathcal{D})$-regularity of Gaussian processes, given an arbitrary open set $\mathcal{D} \subset \mathbb{R}^d$.

Proposition 4.4 (Sample path Hilbert-Sobolev regularity for Gaussian processes)**.** *Let $\mathcal{D} \subset \mathbb{R}^d$ be an open set. Let $(U(x))_{x\in\mathcal{D}} \sim GP(0,k)$ be a measurable centered Gaussian process, defined on a probability set $(\Omega, \mathcal{F}, \mathbb{P})$, such that its standard deviation function σ lies in $L^1_{loc}(\mathcal{D})$. The following statements are equivalent:*

(i) (Sample path regularity) The sample paths of U lie in $H^m(\mathcal{D})$ almost surely.

(ii) (Spectral structure) For all $|\alpha| \leqslant m$, the distributional derivative $\partial^{\alpha,\alpha}k$ lies in $L^2(\mathcal{D}\times\mathcal{D})$ and the associated integral operator

$$\mathcal{E}_k^\alpha f(x) = \int_{\mathcal{D}} \partial^{\alpha,\alpha}k(x,y)f(y)dy \tag{4.6}$$

is trace class. Equivalently, there exists a representative k_α of $\partial^{\alpha,\alpha}k$ in $L^2(\mathcal{D}\times\mathcal{D})$ which is the covariance function of a measurable Gaussian process. For all such k_α, denoting $\sigma_\alpha(x) := k_\alpha(x,x)^{1/2}$, we have

$$Tr(\mathcal{E}_k^\alpha) = \int_{\mathcal{D}} k_\alpha(x,x)dx < +\infty. \tag{4.7}$$

(iii) (Mercer decomposition) The kernel k has the following Mercer decomposition

$$k(x,y) = \sum_{n=0}^{+\infty} \lambda_n \phi_n(x)\phi_n(y) \quad \text{ in } L^2(\mathcal{D}\times\mathcal{D}), \tag{4.8}$$

where (λ_n) is a nonnegative sequence and (ϕ_n) is an orthonormal basis of $L^2(\mathcal{D})$. Moreover, for all $|\alpha| \leqslant m$ and for all $n \in \mathbb{N}_0$ such that $\lambda_n \neq 0$, $\partial^\alpha\phi_n \in L^2(\mathcal{D})$, $\partial^{\alpha,\alpha}k \in L^2(\mathcal{D}\times\mathcal{D})$ and the following facts hold:

$$\sum_{n=0}^{+\infty} \lambda_n \|\partial^\alpha \phi_n\|_2^2 < +\infty, \quad \partial^{\alpha,\alpha}k(x,y) = \sum_{n=0}^{+\infty} \lambda_n \partial^\alpha\phi_n(x)\partial^\alpha\phi_n(y) \quad \text{in} \quad L^2(\mathcal{D}\times\mathcal{D}). \tag{4.9}$$

In this case, $\mathcal{E}_k^\alpha$ in equation (4.6) is well-defined and trace class, with $Tr(\mathcal{E}_k^\alpha) = \sum_{n=0}^{+\infty} \lambda_n \|\partial^\alpha\phi_n\|_2^2$.

(iv) (imbedding of the RKHS) $H_k \subset H^m(\mathcal{D})$, the corresponding natural imbedding $i : H_k \to H^m(\mathcal{D})$ is continuous and $ii^* : H^m(\mathcal{D}) \to H^m(\mathcal{D})$ is trace class. Equivalently, $\ker(i)^\perp$ endowed with the topology of H_k is a separable Hilbert space and $j := i_{|\ker(i)^\perp} : \ker(i)^\perp \to H^m(\mathcal{D})$ is Hilbert-Schmidt. Moreover, the Hilbert-Schmidt norm of j (see Section 2.1.3(ii) and (iii)) is given by

$$\|j\|_{HS}^2 = Tr(ii^*) = \sum_{|\alpha|\leqslant m} Tr(\mathcal{E}_k^\alpha). \tag{4.10}$$

Note that Point (*iv*) above agrees with the definition of Hilbert-Schmidt operators on general, non necessarily separable Hilbert spaces ([6], p. 367). Before proving this result, we discuss Proposition 4.4 in relation with previous results from the literature. First, point (*iv*) is not without reminding Driscoll's theorem ([27], Theorem 4.9) which is well-known in the machine learning/RKHS community; this theorem states the following. Let k and r be two positive definite functions defined over $\mathcal{D}$, and let $U \sim GP(0,k)$. Suppose that $H_k \subset H_r$ with a Hilbert-Schmidt imbedding, then the sample paths of U lie in H_r almost surely.

Second, Proposition 4.4 and equation (4.7) in particular, is a generalization of Theorem 1 from [42] in the case of Gaussian processes; By removing the assumption in [42] that the covariance function be continuous on its diagonal as well as its symmetric cross derivatives, the sufficient condition derived in [42] becomes also necessary. Finally, Proposition 4.4 shows that if $p = 2$, then in the nuclear decomposition of $\mathcal{E}_k^\alpha$ (see Proposition 3.1(iii)) one can choose $\lambda_n^\alpha = \lambda_n$ and $\psi_n^\alpha = \partial^\alpha \psi_n$. It is not obvious that this should hold when $p < 2$ and it in fact fails when $p > 2$ (see Remark 3.4 and Proposition 3.5).

Example 4.5 (Hilbert-Schmidt imbeddings of Sobolev spaces)**.** Proposition 4.4 can be compared with the results found in [45] and its Corollary 4.5 in particular. This corollary states that if $\mathcal{D} \subset \mathbb{R}^d$ is sufficiently smooth, if $H_k \subset H^t(\mathcal{D})$ with a continuous imbedding and if $t > d/2$, then the sample paths of the centered Gaussian process with covariance function k lie in $H^m(\mathcal{D})$ for all real number $m \in [0, t - d/2)$. For example, this holds when k is a Matérn covariance function of order $t - d/2$; its RKHS is then exactly $H^t(\mathcal{D})$ ([45], Example 4.8).

In the particular case where in addition m is an integer, we recover this result from Proposition 4.4. Indeed, it is known that when $m \in (0, t-d/2)$, the imbedding of $H^t(\mathcal{D})$ in $H^m(\mathcal{D})$ is Hilbert-Schmidt. When the involved indexes are nonnegative integers, this is known as Maurin's theorem ([1], Theorem 6.61, p. 202). Maurin's theorem is generalized to fractional exponents in [49], Folgerung 1 p. 310 (in German) or [29], Proposition 7.1 (in French). If $H_k \subset H^t(\mathcal{D})$ with a continuous imbedding, then the inclusion map of H_k in $H^m(\mathcal{D})$ is Hilbert-Schmidt for all $m \in [0, t - d/2) \cap \mathbb{N}_0$. From Proposition 4.4$(iv)$, we obtain that the sample paths of the corresponding Gaussian process indeed lie in $H^m(\mathcal{D})$.

However, not all RKHS that are subspaces of $H^m(\mathcal{D})$ with a Hilbert-Schmidt imbedding are contained in some $H^t(\mathcal{D})$ with $t > m + d/2$, as the following trivial example shows. Fix any $\varepsilon > 0$ and consider the rank one kernel $k(x, x') = f(x)f(x')$ where f is chosen such that $f \in H^m(\mathcal{D})$ and $f \notin H^{m+\varepsilon}(\mathcal{D})$ (choose a representative of f in $L^2(\mathcal{D})$ so that f is a function in the classical sense). Then $H_k = \text{Span}(f)$ and the imbedding of H_k in $H^m(\mathcal{D})$ is Hilbert-Schmidt since it is rank one; but $H_k \not\subset H^{m+\varepsilon}(\mathcal{D})$. Proposition 4.4 yields that the associated trivial Gaussian process $U(x)(\omega) = \xi(\omega)f(x)$ where $\xi \sim \mathcal{N}(0, 1)$ has its sample paths in $H^m(\mathcal{D})$ (it was obvious in the first place).

Example 4.6 (One dimensional case)**.** We build a covariance function which is not pointwise differentiable at any $(q, q') \in \mathbb{Q} \times \mathbb{Q}$, and such that the corresponding Gaussian process has its sample paths in $H^1(\mathbb{R})$. Let $h_a(x) := \max(0, 1 - |x - a|)$ be the hat function centered around $a \in \mathbb{R}$. It lies in $H^1(\mathbb{R})$ but it is not differentiable at $x = a, a - 1$ and $a + 1$. Let (q_n) be an enumeration of $\mathbb{Q}$. Then the following positive definite function over $\mathbb{R}$

$$k(x, x') := \sum_{n=0}^{+\infty} \frac{1}{2^n} h_{q_n}(x) h_{q_n}(x'), \tag{4.11}$$

is not differentiable in the classical sense at each point (x, x') of the form (q_n, q_m), but the map ii^*, with $i : H_k \to H^1(\mathbb{R})$ the canonical imbedding, is trace-class (use equations (4.9) and (4.10)):

$$\begin{aligned} \text{Tr}(ii^*) &= \text{Tr}(\mathcal{E}_k) + \text{Tr}(\mathcal{E}_k^1) && (4.12)\\ &\leqslant \sum_{n=0}^{+\infty} \frac{1}{2^n} \|h_{q_n}\|_2^2 + \sum_{n=0}^{+\infty} \frac{1}{2^n} \|h'_{q_n}\|_2^2 && (4.13)\\ &\leqslant \sum_{n=0}^{+\infty} \frac{1}{2^n} + \sum_{n=0}^{+\infty} \frac{1}{2^n} \times 2^2 = 10. && (4.14) \end{aligned}$$

Before proving Proposition 4.4, we will require a number of lemmas concerning the Mercer decomposition of Hilbert-Schmidt operators over $L^2(\mathcal{D})$. They are proved in Section 6.

Lemma 4.7. *Let k be a measurable positive definite function defined on an open set $\mathcal{D}$. Suppose that $\sigma \in L^1_{loc}(\mathcal{D})$. Then $k \in L^1_{loc}(\mathcal{D} \times \mathcal{D})$. Given a multi-index α, its distributional derivative $D^{\alpha,\alpha}k$ exists and we can introduce the associated continuous bilinear form over $C_c^\infty(\mathcal{D})$,*

$$b_\alpha(\varphi, \psi) := D^{\alpha,\alpha}k(\varphi \otimes \psi) = \int_{\mathcal{D}\times\mathcal{D}} k(x,y)\partial^\alpha\varphi(x)\partial^\alpha\psi(y)dxdy. \tag{4.15}$$

Suppose that it verifies the estimate

$$\forall \varphi, \psi \in E_2, \quad |b_\alpha(\varphi,\psi)| \leqslant C_\alpha \|\varphi\|_2 \|\psi\|_2, \tag{4.16}$$

where E_2 is the set given in Lemma 2.2. Then b_α can be extended to a continuous bilinear form over $L^2(\mathcal{D})$ and there exists a unique bounded, self-adjoint and nonnegative operator $\mathcal{F}_k^\alpha : L^2(\mathcal{D}) \longrightarrow L^2(\mathcal{D})$ such that

$$\forall \varphi, \psi \in C_c^\infty(\mathcal{D}), \quad b_\alpha(\varphi,\psi) = \langle \mathcal{F}_k^\alpha \varphi, \psi \rangle_{L^2(\mathcal{D})}. \tag{4.17}$$

Lemma 4.8. *Let $k \in L^2(\mathcal{D} \times \mathcal{D})$ be a positive definite function and α a multi-index. Suppose that the weak derivative $\partial^{\alpha,\alpha}k$ exists and lies in $L^2(\mathcal{D} \times \mathcal{D})$. Then the bilinear form b_α from equation (4.15) verifies the estimate (4.16) with $C_\alpha = \|\partial^{\alpha,\alpha}k\|_2$. Introduce $\mathcal{F}_k^\alpha$, the bounded operator from Lemma 4.7. Introduce also $\mathcal{E}_k^\alpha$, the integral operator defined on $L^2(\mathcal{D})$ associated to $\partial^{\alpha,\alpha}k$,*

$$(\mathcal{E}_k^\alpha f)(x) = \int_{\mathcal{D}} \partial^{\alpha,\alpha}k(x,y)f(y)dy. \tag{4.18}$$

Then $\mathcal{E}_k^\alpha = \mathcal{F}_k^\alpha$ and $\mathcal{E}_k^\alpha$ is self-adjoint and nonnegative.

Lemma 4.9. *Let $k \in L^2(\mathcal{D} \times \mathcal{D})$ be a positive definite function and $\mathcal{E}_k$ be its associated nonnegative definite Hilbert-Schmidt operator. Let*

$$k(x,y) = \sum_{i=0}^{+\infty} \lambda_i \phi_i(x)\phi_i(y) \tag{4.19}$$

be a symmetric, nonnegative expansion of k in $L^2(\mathcal{D}\times\mathcal{D})$ where (λ_i) is a nonnegative sequence decreasing to 0; it may or may not be its Mercer expansion (i.e. (ϕ_i) may or may not be an orthonormal basis of $L^2(\mathcal{D})$; they are still assumed to be elements of $L^2(\mathcal{D})$ though).

(i) If the partial mixed weak derivative $\partial^{\alpha,\alpha}k$ exists and lies in $L^2(\mathcal{D} \times \mathcal{D})$, then for all $i \in \mathbb{N}_0$ such that $\lambda_i \neq 0, \partial^\alpha \phi_i \in L^2(\mathcal{D})$.

(ii) Assume that for all $i \in \mathbb{N}_0$ such that $\lambda_i \neq 0, \partial^\alpha\phi_i \in L^2(\mathcal{D})$, and that the bilinear form b_α from equation (4.15) verifies the estimate (4.16). Let $\mathcal{F}_k^\alpha$ be the bounded operator from Lemma 4.7. Then

$$Tr(\mathcal{F}_k^\alpha) = \sum_{i=0}^{+\infty} \lambda_i \|\partial^\alpha \phi_i\|^2_{L^2(\mathcal{D})}, \tag{4.20}$$

whether these quantities are finite or not. If in equation (4.20), either one of them is finite, then the series of functions $\sum_{i\in\mathbb{N}_0} \lambda_i \partial^\alpha\phi_i(x)\partial^\alpha\phi_i(y)$ is norm convergent in $L^2(\mathcal{D}\times\mathcal{D})$ (i.e. $\sum_{i\in\mathbb{N}_0} \lambda_i \|\partial^\alpha\phi_i \otimes \partial^\alpha\phi_i\|_{L^2} < +\infty$), $\partial^{\alpha,\alpha}k$ lies in $L^2(\mathcal{D}\times\mathcal{D})$ and we have the following equality:

$$\partial^{\alpha,\alpha}k(x,y) = \sum_{i=0}^{+\infty} \lambda_i \partial^\alpha\phi_i(x)\partial^\alpha\phi_i(y) \quad \text{in} \quad L^2(\mathcal{D}\times\mathcal{D}). \tag{4.21}$$

Moreover, $\mathcal{F}_k^\alpha$ is the (Hilbert-Schmidt) integral operator with kernel $\partial^{\alpha,\alpha}k$, i.e. $\mathcal{F}_k^\alpha = \mathcal{E}_k^\alpha = \mathcal{E}_{\partial^{\alpha,\alpha}k}$. Finally, equation (4.21) then holds for asymmetric derivatives, as for all $|\alpha|, |\beta| \leqslant m$, we also have $\sum_{i\in\mathbb{N}} \lambda_i \|\partial^\beta \phi_i \otimes \partial^\alpha \phi_i\|_{L^2} < +\infty$.

We can now prove Proposition 4.4.

Proof. (Proposition 4.4) We successively prove $(ii) \implies (i)$, $(i) \implies (ii)$, $(ii) \implies (iii)$, $(iii) \implies (ii)$, $(iii) \implies (iv)$ and $(iv) \implies (iii)$.
Before all things, the assumptions and Lemma 2.5 show that the sample paths of U lie in $L^1_{loc}(\mathcal{D})$, that the random variable given by the formula

$$U_\varphi^\alpha : \Omega \ni \omega \longmapsto (-1)^{|\alpha|} \int_{\mathcal{D}} U(x)(\omega) \partial^\alpha \varphi(x) dx \tag{4.22}$$

is well defined and that $(U_\varphi^\alpha)_{\varphi\in F_2}$ is a Gaussian sequence (see equation (2.22) for the definition of F_2).
$\underline{(ii) \implies (i)}$: From Lemma 4.8, $\mathcal{E}_k^\alpha$ is a self-adjoint, nonnegative Hilbert-Schmidt operator; it is actually trace-class by assumption. We can thus define $A_\alpha := \sqrt{\mathcal{E}_k^\alpha}$, which is a Hilbert-Schmidt, self-adjoint, nonnegative operator. From Proposition 4.3, $A_\alpha(B)$ is a GB-set (B is the closed unit ball of $L^2(\mathcal{D})$). Therefore, using the canonical Gaussian process of $L^2(\mathcal{D})$,

$$\mathbb{P}(\sup_{\psi\in A_\alpha(B)} |V_\psi| < +\infty) = 1, \tag{4.23}$$

which, since $F_2 \subset B$, yields in particular that

$$\mathbb{P}(\sup_{\varphi\in F_2} |V_{A_\alpha(\varphi)}| < +\infty) = 1. \tag{4.24}$$

We now observe that the two Gaussian sequences $(V_{A_\alpha(\varphi)})_{\varphi\in F_2}$ and $(U_\varphi^\alpha)_{\varphi\in F_2}$ have the same finite dimensional marginals. Indeed, they are both centered Gaussian sequences with the same covariance:

$$\mathbb{E}[V_{A_\alpha(\varphi)} V_{A_\alpha(\psi)}] = \langle A_\alpha(\varphi), A_\alpha(\psi)\rangle_{L^2} = \langle A_\alpha^2(\varphi), \psi\rangle_{L^2} = \langle \mathcal{E}_k^\alpha \varphi, \psi\rangle_{L^2}. \tag{4.25}$$

$$\begin{aligned} \mathbb{E}[U_\varphi^\alpha U_\psi^\alpha] &= \mathbb{E}\left[\int_{\mathcal{D}} U(x)\partial^\alpha\varphi(x)dx \int_{\mathcal{D}} U(y)\partial^\alpha\psi(y)dy\right] \\ &= \int_{\mathcal{D}\times\mathcal{D}} k(x,y)\partial^\alpha\varphi(x)\partial^\alpha\psi(y)dxdy \\ &= \int_{\mathcal{D}\times\mathcal{D}} \partial^{\alpha,\alpha}k(x,y)\varphi(x)\psi(y)dxdy = \langle \mathcal{E}_k^\alpha\varphi, \psi\rangle_{L^2}. \end{aligned} \tag{4.26}$$

As in the proof of Proposition 3.1 (e.g. equation (3.27)), we deduce that the two random variables $\sup_{\varphi\in F_2} |U_\varphi^\alpha|$ and $\sup_{\varphi\in F_2} |V_{A_\alpha(\varphi)}|$ have the same law, and from equation (4.24), we obtain that

$$\mathbb{P}(\sup_{\varphi\in F_2} |U_\varphi^\alpha| < +\infty) = \mathbb{P}(\sup_{\varphi\in F_2} |V_{A_\alpha(\varphi)}| < +\infty) = 1. \tag{4.27}$$

Since equation (4.27) holds for all $|\alpha| \leqslant m$, this provides a set of probability 1 on which all the sample paths of U lie in $H^m(\mathcal{D})$, which proves (i).
$\underline{(i) \implies (ii)}$: From Lemma 2.4 and the assumption from (i),

$$\mathbb{P}(\sup_{\varphi\in F_2} |U_\varphi^\alpha| < +\infty) = 1. \tag{4.28}$$

From Proposition 2.6, we have that

$$C_\alpha := \mathbb{E}\big[\sup_{\varphi\in F_2} |U^\alpha_\varphi|^2\big] < +\infty. \tag{4.29}$$

Introduce b_α, the continuous bilinear form over $C^\infty_c(\mathcal{D})$ given by

$$b_\alpha(\varphi,\psi) = \int_{\mathcal{D}\times\mathcal{D}} k(x,y)\partial^\alpha\varphi(x)\partial^\alpha\psi(y)dxdy. \tag{4.30}$$

Consider now φ and ψ in F_2. Then,

$$\begin{aligned} |b_\alpha(\varphi,\psi)| &= \left|\int_{\mathcal{D}\times\mathcal{D}} k(x,y)\partial^\alpha\varphi(x)\partial^\alpha\psi(y)dxdy\right| = |\mathbb{E}[U^\alpha_\varphi U^\alpha_\psi]| \\ &\leqslant \mathbb{E}[|U^\alpha_{\varphi_0}U^\alpha_{\psi_0}|] \leqslant \frac{1}{2}\mathbb{E}\big[(U^\alpha_{\varphi_0})^2 + (U^\alpha_{\psi_0})^2\big] \leqslant \mathbb{E}\big[\sup_{\varphi_0\in F}(U^\alpha_{\varphi_0})^2\big] = C_\alpha. \end{aligned} \tag{4.31}$$

From Lemma 4.7, b_α can be extended to a continuous bilinear form over $L^2(\mathcal{D})$ and there exists a unique bounded, self-adjoint and nonnegative operator $\mathcal{E}^\alpha_k$ which verifies

$$\forall\varphi,\psi\in C^\infty_c(\mathcal{D}),\ \int_{\mathcal{D}\times\mathcal{D}} k(x,y)\partial^\alpha\varphi(x)\partial^\alpha\psi(y)dxdy = b_\alpha(\varphi,\psi) = \langle\mathcal{E}^\alpha_k\varphi,\psi\rangle_{L^2}. \tag{4.32}$$

Since $\mathcal{E}^\alpha_k$ is self-adjoint and nonnegative, we can introduce its square root $A_\alpha := \sqrt{\mathcal{E}^\alpha_k}$, which is also a bounded, self-adjoint and nonnegative operator. As in equation (4.27), we can introduce $(V_{A_\alpha(\varphi)})_{\varphi\in F_2}$ and observe that $(V_{A_\alpha(\varphi)})_{\varphi\in F_2}$ and $(U^\alpha_\varphi)_{\varphi\in F_2}$ have the same law. Thus,

$$\mathbb{P}(\sup_{\varphi\in F_2}|V_{A_\alpha(\varphi)}| < +\infty) = \mathbb{P}(\sup_{\varphi\in F_2}|U^\alpha_\varphi| < +\infty) = 1. \tag{4.33}$$

Therefore, $A_\alpha(F_2)$ is a GB-set. From Proposition 4.2(ii), $\overline{\text{Conv}(A_\alpha(F_2))}$ is compact. One then checks by elementary considerations that $\overline{\text{Conv}(A_\alpha(F_2))} = \overline{A_\alpha(B)}$, where B is the unit ball of $L^2(\mathcal{D})$. This shows that A_α is a compact operator. But from Proposition 4.2(i), $\overline{A_\alpha(B)} = \overline{\text{Conv}(A_\alpha(F_2))}$ is also a GB-set. From Proposition 4.3, A_α is Hilbert-Schmidt and $\mathcal{E}^\alpha_k$ is trace-class. In particular, $\mathcal{E}^\alpha_k$ is a Hilbert-Schmidt operator with a kernel k_α that lies in $L^2(\mathcal{D}\times\mathcal{D})$. Moreover,

$$\begin{aligned} \forall\varphi,\psi\in C^\infty_c(\mathcal{D}),\ D^{\alpha,\alpha}k(\varphi\otimes\psi) &= \int_{\mathcal{D}\times\mathcal{D}} k(x,y)\partial^\alpha\varphi(x)\partial^\alpha\psi(y)dxdy \qquad (4.34)\\ &= \int_{\mathcal{D}\times\mathcal{D}} k_\alpha(x,y)\varphi(x)\psi(y)dxdy = T_{k_\alpha}(\varphi\otimes\psi). \qquad (4.35) \end{aligned}$$

Equation (4.35) shows that the distributional derivative $D^{\alpha,\alpha}k$ and the regular distribution T_{k_α} coincide on the set $\mathscr{D}(\mathcal{D})\otimes\mathscr{D}(\mathcal{D})$. From the Schwartz kernel theorem ([48], Theorem 51.7), $D^{\alpha,\alpha}k = T_{k_\alpha}$ in $\mathscr{D}'(\mathcal{D}\times\mathcal{D})$, which shows that $\partial^{\alpha,\alpha}k$ exists in $L^2(\mathcal{D}\times\mathcal{D})$ and that $\partial^{\alpha,\alpha}k = k_\alpha$. For the existence of a representative k_α with the desired properties, we refer to the previous Proposition 3.1(*ii*). Finally, the equality $\text{Tr}(\mathcal{E}_k) = \int k(x,x)dx$ when k is also the covariance function of a measurable Gaussian process is e.g. given in the proof of Proposition 3.11.15 of [6], p. 150. This finishes to prove (*ii*).

(*ii*) $\implies$ (*iii*): If (*ii*), then from Lemma 4.8 (using the notations from Lemmas 4.7 to 4.9), $\mathcal{F}^\alpha_k = \mathcal{E}^\alpha_k$. From Lemma 4.9(*i*), the functions (ϕ_n) lie in $H^m(\mathcal{D})$ and from Lemma 4.9(*ii*), since $\mathcal{E}^\alpha_k = \mathcal{F}^\alpha_k$ is assumed trace class, equation (4.9) holds, as well as the trace formula.

(iii) $\implies$ (ii): if $\sum_{n=0}^{+\infty} \lambda_n \|\partial^\alpha \phi_n\|_2^2 < +\infty$, then from Lemma 4.9(ii), $\mathcal{F}_k^\alpha = \mathcal{E}_k^\alpha$, and still from Lemma 4.9(ii), $\mathrm{Tr}(\mathcal{E}_k^\alpha) < +\infty$, i.e. $\mathcal{E}_k^\alpha$ is trace class. As previously, for the existence of a representative k_α with the desired properties, we refer to the equivalence between Points (ii) and (iii) from Proposition 3.1. As previously, the trace formula is given in [6], p. 150.
(iii) $\implies$ (iv): we first study how finite difference operators behave on elements of H_k in order to use Lemma 2.1(iii). First, using the reproducing formula (4.2), observe that for suitable x and $y \in \mathcal{D}$,

$$\Delta_y f(x) = f(x+y) - f(x) = \langle f, k(x+y,\cdot) - k(x,\cdot)\rangle_{H_k} = \langle f, \Delta_y k(x,\cdot)\rangle_{H_k}. \tag{4.36}$$

More generally, for any finite difference operator $\Delta_{(y_1,...,y_\ell)}$ of order $\ell \leqslant m$, and any open set $\mathcal{D}_0 \Subset \mathcal{D}$ such that $\sum_{i=1}^{\ell} |y_i| < \mathrm{dist}(\mathcal{D}_0, \partial\mathcal{D})$,

$$\Delta_{(y_1,...,y_\ell)} f(x) = \langle f, \Delta_{(y_1,...,y_\ell)} k(x,\cdot)\rangle_{H_k}. \tag{4.37}$$

The Cauchy-Schwarz inequality in H_k yields

$$\Delta_{(y_1,...,y_\ell)} f(x)^2 \leqslant \|f\|_{H_k}^2 \|\Delta_{(y_1,...,y_\ell)} k(x,\cdot)\|_{H_k}^2. \tag{4.38}$$

Furthermore, using the bilinearity of $\langle \cdot,\cdot\rangle_{H_k}$, we have that

$$\|\Delta_{(y_1,...,y_\ell)} k(x,\cdot)\|_{H_k}^2 = [(\Delta_{(y_1,...,y_\ell)} \otimes \Delta_{(y_1,...,y_\ell)})k](x,x). \tag{4.39}$$

We then deduce that (explanation below)

$$\begin{aligned}
\forall f \in H_k,\ \|\Delta_{(y_1,...,y_\ell)} f\|_{L^2(\mathcal{D}_0)}^2 &= \int_{\mathcal{D}_0} \big(\Delta_{(y_1,...,y_\ell)} f\big)(x)^2 dx \\
&\leqslant \|f\|_{H_k}^2 \int_{\mathcal{D}_0} [(\Delta_{(y_1,...,y_\ell)} \otimes \Delta_{(y_1,...,y_\ell)})k](x,x) dx && (4.40)\\
&\leqslant \|f\|_{H_k}^2 \sum_{i=1}^{+\infty} \lambda_i \int_{\mathcal{D}_0} (\Delta_{(y_1,...,y_\ell)} \phi_i)(x)^2 dx && (4.41)\\
&\leqslant \|f\|_{H_k}^2 \sum_{i=1}^{+\infty} \lambda_i \Big(\|\phi_i\|_{H^m}^2 |y_1|^2 \cdots |y_\ell|^2\Big) && (4.42)\\
&\leqslant \|f\|_{H_k}^2 \Bigg(\sum_{|\alpha|\leqslant m} \mathrm{Tr}(\mathcal{E}_k^\alpha)\Bigg)\big(|y_1|^2 \cdots |y_\ell|^2\big). && (4.43)
\end{aligned}$$

We used equations (4.38) and (4.39) to obtain equation (4.40). In equation (4.41), we distributed $\Delta_{(y_1,...,y_\ell)} \otimes \Delta_{(y_1,...,y_\ell)}$ over the Mercer decomposition of k (which exists by the assumption (iii)). In equation (4.42), we used the fact that $\phi_i \in H^m(\mathcal{D})$ (see Lemma 4.9(i)) conjointly with the finite difference control of Lemma 2.1(iii). In equation (4.43), the we used the trace equality from Lemma 4.9(ii). From equation (4.43) and Lemma 2.1(iii) again, we obtain that f lies in $H^m(\mathcal{D})$. Consider now any open set $\mathcal{D}_0 \Subset \mathcal{D}$. Equation (4.43) applied to δ_h^α, the finite difference approximation of ∂^α from equation (2.11) with $h = (h_1, ..., h_d) \in (\mathbb{R}_+^*)^d$ and $|\alpha| \leqslant m$ such that $\sum_{i=1}^{d} \alpha_i h_i < \mathrm{dist}(\mathcal{D}_0, \partial\mathcal{D})$, yields that

$$\forall f \in H_k,\ \|\delta_h^\alpha f\|_{L^2(\mathcal{D}_0)}^2 \leqslant \|f\|_{H_k}^2 \Bigg(\sum_{|\alpha|\leqslant m} \mathrm{Tr}(\mathcal{E}_k^\alpha)\Bigg). \tag{4.44}$$

From the constant estimate "$\|\partial^\alpha f\|_2 \leqslant C$" from Lemma 2.1$(iii)$, we then obtain that

$$\forall f \in H_k,\ \|\partial^\alpha f\|^2_{L^2(\mathcal{D})} \leqslant \|f\|^2_{H_k}\bigg(\sum_{|\alpha|\leqslant m} \mathrm{Tr}(\mathcal{E}^\alpha_k)\bigg). \tag{4.45}$$

Summing the inequality (4.45) for all $|\alpha| \leqslant m$, we obtain that

$$\|f\|_{H^m} \leqslant C\|f\|_{H_k}, \tag{4.46}$$

with $C = \left(N\sum_{|\alpha|\leqslant m}\mathrm{Tr}(\mathcal{E}^\alpha_k)\right)^{1/2}$ and N is the number of multi-indexes α such that $|\alpha| \leqslant m$. Therefore $H_k \subset H^m(\mathcal{D})$ and the corresponding imbedding $i : H_k \to H^m(\mathcal{D})$ is continuous. Using the reproducing formula (4.2), its transpose $i^* : H^m(\mathcal{D}) \to H_k$ is given by

$$i^*(f)(x) = \langle i^*(f), k_x\rangle_{H_k} = \langle f, i(k_x)\rangle_{H^m} = \sum_{|\alpha|\leqslant m}\int_{\mathcal{D}} \partial^\alpha_y k(x,y)\partial^\alpha f(y)dy. \tag{4.47}$$

Above, ∂^α_y denotes differentation w.r.t. the y coordinate (note that $i^*(f)$ is indeed defined pointwise, since $i^*(f) \in H_k$). Let (ψ_j) be an orthonormal basis of $H^m(\mathcal{D})$ and $k = \sum_i \lambda_i\psi_i\otimes\psi_i$ be the Mercer decomposition of k provided by the assumption (iii). The trace of the nonnegative self-adjoint operator ii^* is given by (explanation below)

$$\begin{aligned}
\mathrm{Tr}(ii^*) &= \sum_j \langle\psi_j, ii^*(\psi_j)\rangle_{H^m} = \sum_j\sum_{|\beta|\leqslant m}\langle\partial^\beta\psi_j, \partial^\beta ii^*(\psi_j)\rangle_{L^2}\\
&= \sum_j\sum_{|\beta|\leqslant m}\int_{\mathcal{D}}\partial^\beta\psi_j(x)\partial^\beta ii^*(\psi_j)(x)dx\\
&= \sum_j\sum_{|\beta|\leqslant m}\int_{\mathcal{D}}\partial^\beta\psi_j(x)\partial^\beta_x\sum_{|\alpha|\leqslant m}\int_{\mathcal{D}}\partial^\alpha_y k(x,y)\partial^\alpha\psi_j(y)dydx \qquad (4.48)\\
&= \sum_j\sum_i\lambda_i\sum_{|\alpha|\leqslant m}\sum_{|\beta|\leqslant m}\int_{\mathcal{D}\times\mathcal{D}}\partial^\beta\phi_i(x)\partial^\alpha\phi_i(y)\partial^\beta\psi_j(x)\partial^\alpha\psi_j(y)dydx \qquad (4.49)\\
&= \sum_j\sum_i\lambda_i\bigg(\sum_{|\alpha|\leqslant m}\int_{\mathcal{D}}\partial^\alpha\phi_i(x)\partial^\alpha\psi_j(x)dx\bigg)^2 = \sum_j\sum_i\lambda_i\bigg(\sum_{|\alpha|\leqslant m}\langle\partial^\alpha\phi_i, \partial^\alpha\psi_j\rangle_{L^2}\bigg)^2\\
&= \sum_i\lambda_i\sum_j\langle\phi_i,\psi_j\rangle^2_{H^m} = \sum_i\lambda_i\|\phi_i\|^2_{H^m} = \sum_{|\alpha|\leqslant m}\sum_i\lambda_i\|\partial^\alpha\phi_i\|^2_{L^2} = \sum_{|\alpha|\leqslant m}\mathrm{Tr}(\mathcal{E}^\alpha_k). \qquad (4.50)
\end{aligned}$$

In equation (4.48), we used the fact that $i^*(\psi_j)$ given by equation (4.47) is a representative of $ii^*(\psi_j)$ in $H^m(\mathcal{D})$. In equation (4.49), we used the fact that the series of functions $\sum_i\lambda_i\partial^\beta\phi_i\otimes\partial^\alpha\phi_i$ is norm convergent (Lemma 4.9(ii)) to distribute the partial derivatives over to the Mercer decomposition of k. We also used Fubini's and Tonelli's theorems ad libitum, as all the series $\sum_i\lambda_i\partial^\beta\phi_i\otimes\partial^\alpha\phi_i$ are norm convergent. Since $\sum_{|\alpha|\leqslant m}\mathrm{Tr}(\mathcal{E}^\alpha_k)$ is finite by assumption, equation (4.50) finishes to prove (iv) when H_k is separable.

When H_k is not separable, observe that $\ker(i)$ is closed in H_k since i is continuous. Therefore $H_k = \ker(i)\bigoplus\ker(i)^\perp$ and $\ker(i)^\perp$ endowed with the topology of H_k is a Hilbert space. Moreover, $i^* : H^m(\mathcal{D}) \to H_k$ is compact since ii^* is trace class. Thus its closed range $\overline{\mathrm{im}(i^*)}$ is separable ([16], Exercise 3 p. 176). Finally, observe that $\overline{\mathrm{im}(i^*)} = \ker(i)^\perp$ ([16], Theorem 4.12) so that $\ker(i)^\perp$ is a separable Hilbert space. Consider now $j := i_{|\ker(i)^\perp}$, the restriction of i to $\ker(i)^\perp$. Then $ii^* = jj^*$, so that equation (4.50) indeed yields that j is Hilbert-Schmidt.

$\underline{(iv) \implies (iii)}$: by assumption, ii^* is a compact self-adjoint nonnegative operator acting on the Hilbert space $H^m(\mathcal{D})$. There exists a decreasing nonnegative sequence $(\mu_j)_{j\in\mathbb{N}}$ and an orthonormal basis of eigenvectors of ii^*, $(\psi_j)_{j\in\mathbb{N}}$ such that for all $f \in H^m(\mathcal{D})$,

$$ii^*(f) = \sum_{j=1}^{+\infty} \mu_j \langle \psi_j, f\rangle_{H^m} \psi_j \quad \text{in } H^m(\mathcal{D}). \tag{4.51}$$

Since ii^* is assumed trace class,

$$\sum_{|\alpha|\leqslant m} \sum_{j=1}^{+\infty} \mu_j \|\partial^\alpha \psi_j\|_{L^2}^2 = \sum_{j=1}^{+\infty} \mu_j \|\psi_j\|_{H^m}^2 = \sum_{j=1}^{+\infty} \mu_j < +\infty. \tag{4.52}$$

We now show that the following equality holds in $L^2(\mathcal{D}\times\mathcal{D})$:

$$k(x,y) = \sum_{j=1}^{+\infty} \mu_j \psi_j(x)\psi_j(y). \tag{4.53}$$

In conjunction with equation (4.52), this equation will allow us to use Lemma 4.9(ii), which will imply the point (iii). First, one easily shows that $\sum_{j=1}^{+\infty} \mu_j \psi_j \otimes \psi_j$, the right-hand side of equation (4.53), is indeed in $L^2(\mathcal{D}\times\mathcal{D})$ (e.g. use that $\sum_j \mu_j < +\infty$). The upcoming equation (4.63) will then show that k is indeed in $L^2(\mathcal{D}\times\mathcal{D})$. Now, decompose $i(k_x) \in H^m(\mathcal{D})$ on the basis $(\psi_j)_{j\in\mathbb{N}}$, given any $x \in \mathcal{D}$:

$$i(k_x) = \sum_{j=1}^{+\infty} \langle \psi_j, i(k_x)\rangle_{H^m} \psi_j \quad \text{in } H^m(\mathcal{D}). \tag{4.54}$$

In equation (4.54), the scalar $\langle \psi_j, i(k_x)\rangle_{H^m}$ is obtained through the reproducing formula (4.2):

$$\langle \psi_j, i(k_x)\rangle_{H^m} = \langle i^*(\psi_j), k_x\rangle_{H_k} = i^*(\psi_j)(x). \tag{4.55}$$

Moreover, ψ_j is an eigenvector of ii^*: $\mu_j\psi_j = ii^*(\psi_j)$ in $H^m(\mathcal{D})$. In particular,

$$\|\mu_j\psi_j - ii^*(\psi_j)\|_{L^2(\mathcal{D})} = 0. \tag{4.56}$$

But the pointwise defined function $i^*(\psi_j)$ is a representative of $ii^*(\psi_j)$ in $H^m(\mathcal{D})$, since i is the imbedding of H_k in $H^m(\mathcal{D})$. Setting $S = \sum_j \mu_j = \mathrm{Tr}(ii^*)$, one then has (explanation below)

$$\Big\| k - \sum_{j=1}^{+\infty} \mu_j \psi_j \otimes \psi_j \Big\|_{L^2(\mathcal{D}\times\mathcal{D})}^2 = \int_{\mathcal{D}\times\mathcal{D}} \Big(k(x,y) - \sum_{j=1}^{+\infty} \mu_j \psi_j(x)\psi_j(y)\Big)^2 dxdy \tag{4.57}$$

$$= \int_{\mathcal{D}}\int_{\mathcal{D}} \Big(k_x(y) - \sum_{j=1}^{+\infty} \mu_j \psi_j(x)\psi_j(y)\Big)^2 dydx \tag{4.58}$$

$$= \int_{\mathcal{D}}\int_{\mathcal{D}} \Big(i(k_x)(y) - \sum_{j=1}^{+\infty} \mu_j \psi_j(x)\psi_j(y)\Big)^2 dydx \tag{4.59}$$

$$= \int_{\mathcal{D}}\int_{\mathcal{D}} \Big(\sum_{j=1}^{+\infty} \mu_j \psi_j(y)\big({\mu_j}^{-1} i^*(\psi_j) - \psi_j(x)\big)\Big)^2 dydx \tag{4.60}$$

$$\leqslant \int_{\mathcal{D}} \int_{\mathcal{D}} S \sum_{j=1}^{+\infty} \mu_j \psi_j(y)^2 \big({\mu_j}^{-1} i^*(\psi_j) - \psi_j(x)\big)^2 dy dx \tag{4.61}$$

$$\leqslant S \sum_{j=1}^{+\infty} \mu_j \int_{\mathcal{D}} \psi_j(y)^2 dy \int_{\mathcal{D}} \big({\mu_j}^{-1} ii^*(\psi_j) - \psi_j(x)\big)^2 dx \tag{4.62}$$

$$\leqslant S \sum_{j=1}^{+\infty} \mu_j \|\psi_j\|^2_{L^2(\mathcal{D})} \|{\mu_j}^{-1} ii^*(\psi_j) - \psi_j\|^2_{L^2(\mathcal{D})} = 0 \tag{4.63}$$

Above, we used Tonelli's theorem in equation (4.58). We imbedded k_x in $H^m(\mathcal{D})$ in equation (4.59). We used equations (4.54) and (4.55) in equation (4.60). We used Jensen's discrete inequality on the squaring function $(\cdot)^2$ with the weights μ_j/S ($\mu_j/S \geqslant 0, \sum_j \mu_j/S = 1$) in equation (4.61). We imbedded $i^*(\psi_j)$ in $H^m(\mathcal{D})$ and used Tonelli's theorem in equation (4.62). We used equation (4.56) in equation (4.63).

Therefore we have proved that equation (4.53) holds. Since, for $|\alpha| \leqslant m$, $\sum_j \mu_j \|\partial^\alpha \psi_j\|_2^2 < +\infty$, we have $\partial^{\alpha,\alpha} k = \sum_j \mu_j \partial^\alpha \psi_j \otimes \partial^\alpha \psi_j$ in $L^2(\mathcal{D} \times \mathcal{D})$ and $\mathcal{E}_k^\alpha = \mathcal{F}_k^\alpha$ from Lemma 4.8. By the assumption that ii^* is trace class and using the trace equalities from Lemma 4.9(*ii*) ($\mathrm{Tr}(\mathcal{E}_k^\alpha) = \mathrm{Tr}(\mathcal{F}_k^\alpha) = \sum_j \mu_j \|\partial^\alpha \psi_j\|_2^2$),

$$\sum_{|\alpha| \leqslant m} \mathrm{Tr}(\mathcal{E}_k^\alpha) = \sum_{|\alpha| \leqslant m} \sum_{j=1}^{+\infty} \mu_j \|\partial^\alpha \psi_j\|^2_{L^2} = \sum_{j=1}^{+\infty} \mu_j \|\psi_j\|^2_{H^m} = \sum_i \mu_j = \mathrm{Tr}(ii^*) < +\infty. \tag{4.64}$$

Therefore, Lemma 4.9(*ii*) implies that every $\mathcal{E}_k^\alpha$ is indeed trace-class, which shows (*iii*). □

5 Concluding remarks and perspectives

Given $p \in (1, +\infty)$ and $m \in \mathbb{N}_0$, we showed that the $W^{m,p}$-Sobolev regularity of integer order of a measurable Gaussian process $((U(x))_{x \in \mathcal{D}} \sim GP(0, k)$ is fully equivalent to the fact that $\partial^{\alpha,\alpha} k$ lies in $L^p(\mathcal{D} \times \mathcal{D})$ combined with the integrability in $L^p(\mathcal{D})$ of the standard deviation associated to $\partial^{\alpha,\alpha} k$, provided we use a suitable representative of $\partial^{\alpha,\alpha} k$ in $L^p(\mathcal{D} \times \mathcal{D})$. Using general results on Gaussian measures over Banach spaces of type 2 and cotype 2, we translated this criteria as the existence of suitable nuclear decompositions of the covariance. These can be understood as generalizations to Banach spaces of the eigenfunction expansion of symmetric, nonnegative and trace class operators. In the Hilbert space case $p = 2$, we linked this property with the Hilbert-Schmidt nature of the imbedding of the RKHS in $H^m(\mathcal{D})$, and gave explicit formulas for the traces of the involved integral operators in terms of the Mercer decomposition of the kernel.

The results presented in this article provide a theoretical background w.r.t. the use of Gaussian processes for solving physics-related machine learning problems, in particular when modeling solutions of PDEs as sample paths of some Gaussian process. These results also come along with suitable quantities for controlling the Sobolev norm of the corresponding sample paths (see Remark 3.7). The application of the Gaussian process principles identified here to PDE-related machine learning problems, e.g. following the approach of [11], is certainly an interesting continuation of the results of this article. Controlling the small ball probability (see e.g. [33] for further details) of the Sobolev norm of a Gaussian process, perhaps in terms of some nuclear norm, is also a relevant question for further applications of Gaussian process techniques in such machine learning problems. Finally, the following question (which was implicit in this article) is interesting for probability theory: are all Gaussian measures over $W^{m,p}(\mathcal{D})$ induced by some Gaussian process? Proposition 2.9 states that this is true for $m = 0$, i.e. $L^p(\mathcal{D})$.

The following directions are interesting for generalizing the results presented here. First, similar spectral/integral criteria should be obtained for fractional Sobolev and Besov spaces. Second, similar results should be sought to tackle the limit cases $p = 1$ and $p = +\infty$. Linked to the case $p = 1$, results should be sought for the space of functions of bounded variations ([9], p. 269), which are important in many problems related to physics. In particular, those spaces are adapted to the study of nonlinear hyperbolic PDEs, where shocks (discontinuities in the solution) may appear and solutions may only be understood in the weak or distributional sense ([43], Lemma 2.2.1 and Proposition 2.3.6).

Ackowledgements This research was supported by SHOM (Service Hydrographique et Océanographique de la Marine) project "Machine Learning Methods in Oceanography" no-20CP07. The author warmly thanks Pascal Noble, Olivier Roustant and Rémy Baraille for fruitful discussions, as well as the referees for their constructive feedbacks, which led to an improved version of the manuscript.

6 Proofs of intermediary results and lemmas

Proof. (Lemma 2.1) This proof follows exactly the lines of the proof of Proposition 9.3 from [9]. $(i) \iff (ii)$: suppose that $u \in W^{m,p}(\mathcal{D})$, use the fact that the distributional derivative $D^\alpha u$ is a regular distribution represented by a function that lies in $L^p(\mathcal{D})$, denoted by $\partial^\alpha u$:

$$\forall \varphi \in C_c^\infty(\mathcal{D}), \quad \int_{\mathcal{D}} u(x)\partial^\alpha \varphi(x)dx = (-1)^{|\alpha|}\int_{\mathcal{D}} \partial^\alpha u(x)\varphi(x)dx. \tag{6.1}$$

Hölder's inequality yields (2.12) with $C_\alpha = \|\partial^\alpha u\|_{L^p}$. Conversely, suppose that (2.12) holds and consider any $|\alpha| \leqslant m$. Since $C_c^\infty(\mathcal{D})$ is dense in $L^q(\mathcal{D})$ (whatever the open set $\mathcal{D}$, [1], Section 2.30), equation (2.12) shows that the linear form $L_\alpha : \varphi \longmapsto (-1)^{|\alpha|}\int_{\mathcal{D}} u(x)\partial^\alpha\varphi(x)dx, \varphi \in C_c^\infty(\mathcal{D})$, can be extended to a continuous linear form over $L^q(\mathcal{D})$. From Riesz' representation lemma, there exists $v_\alpha \in L^p(\mathcal{D})$ such that $L_\alpha(\varphi) = \langle v_\alpha, \varphi\rangle_{L^p,L^q}$ for all $\varphi \in L^q(\mathcal{D})$. In particular, this is valid for all $\varphi \in C_c^\infty(\mathcal{D})$, which shows that for all $|\alpha| \leqslant m, \partial^\alpha u$ exists and is equal to v_α. Thus $u \in W^{m,p}(\mathcal{D})$. Finally, Hölder's inequality and the density of $C_c^\infty(\mathcal{D})$ in $L^q(\mathcal{D})$ yield

$$\|\partial^\alpha u\|_{L^p(\mathcal{D})} = \sup_{\varphi \in C_c^\infty(\mathcal{D})\backslash\{0\}} \left|\int_{\mathcal{D}} \partial^\alpha u(x)\frac{\varphi(x)}{\|\varphi\|_{L^q(\mathcal{D})}}dx\right| = \sup_{\varphi \in C_c^\infty(\mathcal{D})\backslash\{0\}} \left|\int_{\mathcal{D}} u(x)\frac{\partial^\alpha\varphi(x)}{\|\varphi\|_{L^q(\mathcal{D})}}dx\right|.$$

$(iii) \implies (ii)$: suppose (iii), let us show (ii). Let $|\alpha| \leqslant m$ and let $\varphi \in C_c^\infty(\mathcal{D})$. Note $K :=$ Supp(φ) its compact support and consider an open set $\mathcal{D}_0$ such that $K \subset \mathcal{D}_0 \Subset \mathcal{D}$. Let $\alpha \in \mathbb{N}_0^d$ and $h = (h_1, ..., h_d) \in (\mathbb{R}_+^*)^d$ be such that $\sum_{i=1}^d \alpha_i h_i < \text{dist}(\mathcal{D}_0, \partial\mathcal{D})$. Recall that δ_h^α from equation (2.11) is a finite difference approximation of ∂^α and from (iii),

$$\left|\int_{\mathcal{D}} \delta_h^\alpha u(x)\varphi(x)dx\right| \leqslant \|\varphi\|_{L^q(\mathcal{D}_0)}\|\delta_h^\alpha u\|_{L^p(\mathcal{D}_0)} \leqslant C\|\varphi\|_{L^q(\mathcal{D})}. \tag{6.2}$$

Note also that we have the discrete integration by parts formula since h is suitably chosen:

$$\int_{\mathcal{D}} \delta_h^\alpha u(x)\varphi(x)dx = \int_{\mathcal{D}} u(x)(\delta_h^\alpha)^*\varphi(x)dx. \tag{6.3}$$

Therefore,

$$\left|\int_{\mathcal{D}} u(x)(\delta_h^\alpha)^*\varphi(x)dx\right| \leqslant C\|\varphi\|_{L^q(\mathcal{D})}. \tag{6.4}$$

The Lebesgue dominated convergence theorem yields that the left hand side converges to $\left|\int_{\mathcal{D}} u(x)\partial^\alpha\varphi(x)dx\right|$. We therefore have (ii).
$(i) \implies (iii)$: We will use recursively the fact that if $\mathcal{O} \subset \mathbb{R}^d$ is an open set and if $f \in W^{1,p}(\mathcal{O}) \cap C^\infty(\mathcal{O})$, then for all open set $\mathcal{O}_0 \subset \mathcal{O}$ and $y \in \mathbb{R}^d$ such that $\mathcal{O}_0 + ty \subset \mathcal{O}$ for all $t \in [0,1]$, we have

$$\|\Delta_y f\|^p_{L^p(\mathcal{O}_0)} = \|\tau_y f - f\|^p_{L^p(\mathcal{O}_0)} \leqslant |y|^p \|\nabla f\|^p_{L^p(\mathcal{O})} = |y|^p \sum_{j=1}^d \|\partial_{x_j} f\|^p_{L^p(\mathcal{O})}. \tag{6.5}$$

This is a slight generalization of equation (4) p. 268 in [9], found in the proof of Proposition 9.3 in [9]. The proof of equation (6.5) is an exact copy of the proof of equation (4) p. 268 in [9], which relies on writing the quantity $u(x+h) - u(x)$ as the integral of the derivative of the map $t \mapsto u(x+th), t \in [0,1]$.

We first show equation (2.14) under the assumption that $u \in W^{m,p}(\mathcal{D}) \cap C^\infty(\mathcal{D})$. The Meyers-Serrin theorem ([1], Theorem 3.17), which asserts the density of $W^{m,p}(\mathcal{D}) \cap C^\infty(\mathcal{D})$ in $W^{m,p}(\mathcal{D})$ for *arbitrary* open sets $\mathcal{D} \subset \mathbb{R}^d$, will imply that equation (2.14) holds for general $u \in W^{m,p}(\mathcal{D})$. Let $\mathcal{D}_0 \Subset \mathcal{D}$ be an open set, $\ell \leqslant m$ and $(y_1, ..., y_\ell) \in (\mathbb{R}^d)^\ell$ be such that $\sum_{i=1}^\ell |y_i| < \mathrm{dist}(\mathcal{D}_0, \partial\mathcal{D})$. We begin by constructing a sequence of open sets $(\mathcal{D}_k)_{0 \leqslant k \leqslant \ell}$ starting from $\mathcal{D}_0$ such that $\mathcal{D}_0 \subset \mathcal{D}_1 \subset ... \subset \mathcal{D}_\ell \Subset \mathcal{D}$, which additionally verify $\mathcal{D}_{k-1} + ty_k \subset \mathcal{D}_k$ for all $t \in [0,1]$ and all $k \in \{1, ..., \ell\}$, with $\sum_{i=k}^\ell |y_i| < \mathrm{dist}(\mathcal{D}_{k-1}, \partial\mathcal{D})$ for all $k \in \{1, ..., \ell\}$. This will enable us to use equation (6.5) recursively over $k \in \{1, ..., \ell\}$, to obtain the desired finite difference control (2.14). We detail below this construction. We define this sequence recursively as follow, for all $k \in \{1, ..., \ell\}$ and starting from the open set $\mathcal{D}_0$,

$$\mathcal{D}_k := \bigcup_{t \in [0,1]} (\mathcal{D}_{k-1} + ty_k). \tag{6.6}$$

This sequence clearly verifies $\mathcal{D}_{k-1} + ty_k \subset \mathcal{D}_k$ for all $t \in [0,1]$. If $\mathcal{D}_{k-1}$ is open, then $\mathcal{D}_{k-1} + ty_k$ is also open, hence $\mathcal{D}_k$ is open. We now check that the property that $\sum_{i=k}^\ell |y_i| < \mathrm{dist}(\mathcal{D}_{k-1}, \partial\mathcal{D})$ is inherited recursively (it is true for $k = 1$). With the assumption that $\sum_{i=k}^\ell |y_i| < \mathrm{dist}(\mathcal{D}_{k-1}, \partial\mathcal{D})$, we indeed have that

$$\begin{aligned}
\mathrm{dist}(\mathcal{D}_k, \partial\mathcal{D}) &= \inf_{x \in \mathcal{D}_k, y \in \partial\mathcal{D}} |x - y| = \inf_{\substack{z \in \mathcal{D}_{k-1}, t \in [0,1] \\ y \in \partial\mathcal{D}}} |z + ty_k - y| \\
&\geqslant \inf_{\substack{z \in \mathcal{D}_{k-1}, t \in [0,1] \\ y \in \partial\mathcal{D}}} \Big(|z-y| - t|y_k|\Big) = \inf_{z \in \mathcal{D}_{k-1}, y \in \partial\mathcal{D}} \inf_{t \in [0,1]} \Big(|z-y| - t|y_k|\Big) \qquad (6.7) \\
&\geqslant \inf_{z \in \mathcal{D}_{k-1}, y \in \partial\mathcal{D}} |z-y| - |y_k| = \mathrm{dist}(\mathcal{D}_{k-1}, \partial\mathcal{D}) - |y_k| > \sum_{i=k+1}^\ell |y_i|. \qquad (6.8)
\end{aligned}$$

In particular, $\mathrm{dist}(\mathcal{D}_\ell, \partial\mathcal{D}) > 0$. In equation (6.8), we used the assumption that $\sum_{i=k}^\ell |y_i| < \mathrm{dist}(\mathcal{D}_{k-1}, \partial\mathcal{D})$. One also checks that $\mathcal{D}_k \Subset \mathcal{D}$, as follow. Given $x \in \mathcal{D}_k$, write $x = z + ty_k$ for some $z \in \mathcal{D}_{k-1}$ and $t \in [0,1]$. This yields

$$|x - z| \leqslant |y_k| \leqslant \sum_{i=k}^\ell |y_i| =: r_k. \tag{6.9}$$

Hence, $x \in \overline{B(z, r_k)}$. Moreover, $\overline{B(z, r_k)} \subset \mathcal{D}$, since $z \in \mathcal{D}_{k-1}$ and $r_k < \mathrm{dist}(\mathcal{D}_{k-1}, \partial\mathcal{D})$ (indeed, one checks that $\overline{B(z, r)} \subset \mathcal{D}$ for all $z \in \mathcal{D}$ and $r > 0$ such that $r < \mathrm{dist}(z, \partial\mathcal{D})$, e.g. by observing that $\partial\mathcal{D} \subset \{y : |z-y| > r\}$, thus (taking the complement) $\overline{B(z,r)} \subset \mathcal{D} \cup \mathrm{Int}(\mathcal{D}^c)$, that $z \in \mathcal{D}$ and that $\overline{B(z,r)}$ is path connected). From equation (6.9), we obtain that $\mathcal{D}_k \subset \bigcup_{x\in\mathcal{D}_{k-1}} \overline{B(x,r_k)} \subset \mathcal{D}$, with $\mathrm{dist}(\bigcup_{x\in\mathcal{D}_{k-1}} \overline{B(x,r_k)}, \partial\mathcal{D}) \geqslant \mathrm{dist}(\mathcal{D}_{k-1}, \partial\mathcal{D}) - r_k > 0$. This finally yields that $\mathcal{D}_k \Subset \mathcal{D}$. In particular, $\mathcal{D}_\ell \Subset \mathcal{D}$.

Given the sequence $(\mathcal{D}_k)_{0\leqslant k\leqslant \ell}$, we can now prove the finite difference control (2.14). First, one easily checks that classical partial derivatives and finite difference operators all commute together, as long as both are well-defined. Recall that

$$\Delta_{(y_1,\ldots,y_\ell)} = \Delta_{y_1} \circ \Delta_{(y_2,\ldots,y_\ell)}. \tag{6.10}$$

Note now that $\Delta_{(y_2,\ldots,y_\ell)} u$ lies in $W^{1,p}(\mathcal{D}_1) \cap C^\infty(\mathcal{D}_1)$, since $\sum_{i=2}^{\ell} |y_i| < \mathrm{dist}(\mathcal{D}_1, \partial\mathcal{D})$. Using equation (6.5) with $\mathcal{O} = \mathcal{D}_1$, $\mathcal{O}_0 = \mathcal{D}_0$ and $f = \Delta_{(y_2,\ldots,y_\ell)} u$,

$$\begin{aligned}
\|\Delta_{(y_1,\ldots,y_\ell)} u\|^p_{L^p(\mathcal{D}_0)} &= \left\|\Delta_{y_1}(\Delta_{(y_2,\ldots,y_\ell)} u)\right\|^p_{L^p(\mathcal{D}_0)} \leqslant |y_1|^p \left\|\nabla(\Delta_{(y_2,\ldots,y_\ell)} u)\right\|^p_{L^p(\mathcal{D}_1)} &(6.11)\\
&\leqslant |y_1|^p \sum_{j=1}^{d} \left\|\partial_{x_j}(\Delta_{(y_2,\ldots,y_\ell)} u)\right\|^p_{L^p(\mathcal{D}_1)} \\
&\leqslant |y_1|^p \sum_{j=1}^{d} \left\|\Delta_{(y_2,\ldots,y_\ell)} \big(\partial_{x_j} u\big)\right\|^p_{L^p(\mathcal{D}_1)}. &(6.12)
\end{aligned}$$

We used equation (6.5) in equation (6.11). We also commuted finite difference operators and partial derivatives in equation (6.12). If $m = 1$, then we have proved equation (2.14) for $u \in W^{1,p}(\mathcal{D}) \cap C^\infty(\mathcal{D})$. If $m \geqslant 2$, note that for all j, $\partial_{x_j} u \in W^{m-1,p}(\mathcal{D}) \cap C^\infty(\mathcal{D})$. In particular, $\partial_{x_j} u \in W^{1,p}(\mathcal{D}) \cap C^\infty(\mathcal{D})$. One can then proceed by induction and perform the above step sequentially over $k \in \{1, \ldots, \ell\}$ (recall that $\ell \leqslant m$), successively using equation (6.5) for $\mathcal{O} = \mathcal{D}_k$, $\mathcal{O}_0 = \mathcal{D}_{k-1}$ and $f = \Delta_{(y_{k+1},\ldots,y_\ell)} u \in W^{1,p}(\mathcal{D}_k) \cap C^\infty(\mathcal{D}_k)$ (the latter holds because $\sum_{i=k+1}^{\ell} |y_i| < \mathrm{dist}(\mathcal{D}_k, \partial\mathcal{D})$). This yields

$$\begin{aligned}
\|\Delta_{(y_1,\ldots,y_\ell)} u\|^p_{L^p(\mathcal{D}_0)} &\leqslant |y_1|^p \times \ldots \times |y_k|^p \sum_{|\beta|=k} \|\Delta_{(y_{k+1},\ldots,y_\ell)}(\partial^\beta u)\|^p_{L^p(\mathcal{D}_k)} \\
&\leqslant |y_1|^p \times \ldots \times |y_\ell|^p \sum_{|\beta|=\ell} \|\partial^\beta u\|^p_{L^p(\mathcal{D}_\ell)} \\
&\leqslant |y_1|^p \times \ldots \times |y_\ell|^p \|u\|^p_{W^{\ell,p}(\mathcal{D})} \leqslant |y_1|^p \times \ldots \times |y_\ell|^p \|u\|^p_{W^{m,p}(\mathcal{D})}. &(6.13)
\end{aligned}$$

This shows equation (2.14) with $C = \|u\|_{W^{m,p}(\mathcal{D})}$ (in fact, the equations above show that finite differences $\Delta_{(y_1,\ldots,y_\ell)} u$ of order $\ell \leqslant m$ are more accurately controlled by derivatives of order ℓ, taking $C_\ell = (\sum_{|\beta|=\ell} \|\partial^\beta u\|^p_{L^p(\mathcal{D})})^{1/p}$). The general case where $u \in W^{m,p}(\mathcal{D})$ is settled by equation (6.13) conjoined with the Meyers-Serrin theorem. We finally show that $\|\partial^\alpha u\|_{L^p(\mathcal{D})} \leqslant C$, given any C which verifies equation (2.14). For this, copy the previous steps of $(iii) \implies (ii)$, which prove that for all $\varphi \in C_c^\infty(\mathcal{D})$, the control from equation (2.12) holds for this C. Using the extremal equality case of Hölder's inequality in equation (2.12) indeed yields

$$\|\partial^\alpha u\|_{L^p(\mathcal{D})} = \sup_{\varphi\in C_c^\infty(\mathcal{D})\setminus\{0\}} \left| \int_{\mathcal{D}} u(x) \frac{\partial^\alpha \varphi(x)}{\|\varphi\|_{L^p(\mathcal{D})}} dx \right| \leqslant C. \tag{6.14}$$

This finishes the proof. □

Proof. (Lemma 2.4) We begin by explicitly constructing the family (Φ_n^q). First, use the fact that $L^q(\mathcal{D})$ is a separable Banach space ([1], Theorem 2.21) : let $(f_n)_{n\in\mathbb{N}} \subset L^q(\mathcal{D})$ be a dense countable subset of $L^q(\mathcal{D})$. For all $n \in \mathbb{N}$, let $(\phi_{nm})_{m\in\mathbb{N}} \subset C_c^\infty(\mathcal{D})$ be such that $\phi_{nm} \longrightarrow f_n$ for the $L^q(\mathcal{D})$ topology (recall that $C_c^\infty(\mathcal{D})$ is dense in $L^q(\mathcal{D})$, [1], Corollary 2.30). We relabel the countable family $(\phi_{nm})_{n,m\in\mathbb{N}}$ as $(\varphi_n)_{n\in\mathbb{N}}$, which is thus dense in $L^q(\mathcal{D})$. Second, let $(h_n)_{n\in\mathbb{N}} \subset C_c^\infty(\mathcal{D})$ be a dense subset of $C_c^\infty(\mathcal{D})$ for its LF-space topology (see Lemma 2.3). We then define E_q to be the set of all finite linear combinations of elements of (φ_n) and (h_n) with rational coefficients :

$$E_q = \mathrm{Span}_{\mathbb{Q}}\{\varphi_n, n \in \mathbb{N}\} + \mathrm{Span}_{\mathbb{Q}}\{h_m, m \in \mathbb{N}\} \tag{6.15}$$

$$= \bigcup_{n,m\in\mathbb{N}} \Big\{ \sum_{i=1}^{n} q_i\varphi_i + \sum_{j=1}^{m} r_j h_j, (q_1, ..., q_n, r_1, ..., r_m) \in \mathbb{Q}^{n+m} \Big\}. \tag{6.16}$$

Note that E_q is countable, as a countable union of countable sets. We then define the family (Φ_n^q) to be an enumeration of E_q : $E_q = \{\Phi_n^q, n \in \mathbb{N}\}$.

Proof of (i): Suppose that $T = T_v$ for some $v \in L^p(\mathcal{D})$. Then the control (2.18) is obviously true. Now, suppose that this countable control holds : let us show that $T = T_v$ for some $v \in L^p(\mathcal{D})$.

We begin by showing that the map $T_{|E_q}$ (restriction of T to the set E_q) can be uniquely extended to a continuous linear form $\tilde{T}$ over $L^q(\mathcal{D})$. Begin with the fact that for all $f, g \in E_q$, then $f - g \in E_q$ and from equation (2.23),

$$|T(f) - T(g)| = |T(f - g)| \leqslant C\|f - g\|_q. \tag{6.17}$$

Equation (6.17) shows that $T_{|E_q}$ is Lipschitz over E_q and therefore uniformly continuous on E_q. Since $\mathbb{R}$ is complete and E_q is dense in $L^q(\mathcal{D})$, $T_{|E_q}$ can be uniquely extended by a map $\tilde{T}$ defined over $L^q(\mathcal{D})$, which is itself uniformly continuous ([40], Problem 44, p. 196). We briefly recall the construction procedure of $\tilde{T}$ over $L^q(\mathcal{D})$. Given $f \in L^q(\mathcal{D})$ and $(f_n) \subset E_q$ *any* sequence such that $\|f_n - f\|_{L^q} \to 0$, one shows that the sequence $(T(f_n))_{n\in\mathbb{N}}$ is Cauchy, thus convergent and one sets $\tilde{T}(f) := \lim_n T(f_n)$. One proves that the value $\tilde{T}(f)$ does not depend on the sequence (f_n), which implies that $\tilde{T}$ is well defined and coincides with T on E_q.

We now check that $\tilde{T}$ remains linear. Let $f, g \in L^q(\mathcal{D})$ and $\lambda \in \mathbb{R}$. Let $(f_n), (g_n) \subset E_q$ and $(\lambda_n) \subset \mathbb{Q}$ be sequences such that $f_n \to f, g_n \to g$ both in $L^q(\mathcal{D})$ and $\lambda_n \to \lambda$. Then $\lambda_n f_n + g_n \to \lambda f + g$ in $L^q(\mathcal{D})$, and the sequence $(\lambda_n f_n + g_n)$ is contained in E_q. Since $\tilde{T}$ is well defined, we have that

$$\tilde{T}(\lambda f + g) = \lim_{n\to\infty} T(\lambda_n f_n + g_n) = \lim_{n\to\infty} \lambda_n T(f_n) + T(g_n) = \lambda\tilde{T}(f) + \tilde{T}(g). \tag{6.18}$$

Thus, $\tilde{T}$ is a (uniformly) continuous linear form over $L^q(\mathcal{D})$. Riesz' representation lemma yields a function $v \in L^p(\mathcal{D})$ such that

$$\forall f \in L^q(\mathcal{D}), \quad \tilde{T}(f) = \int_{\mathcal{D}} f(x)v(x)dx. \tag{6.19}$$

We now need to check that in fact $\tilde{T}(\varphi) = T(\varphi)$ if $\varphi \in C_c^\infty(\mathcal{D})$, to show that $\tilde{T}$ is indeed an extension of T. For this, notice that T and $\tilde{T}$ both define continuous linear forms over $C_c^\infty(\mathcal{D})$, w.r.t. its LF-topology (v lies in $L^1_{loc}(\mathcal{D})$). Note also that T and $\tilde{T}$ coincide on E_q, by construction of $\tilde{T}$:

$$\forall n \in \mathbb{N}, \quad T(\Phi_n) - \tilde{T}(\Phi_n) = 0. \tag{6.20}$$

But E_q is chosen so that it contains (h_n), which is a dense subset of $C_c^\infty(\mathcal{D})$. Given $\varphi \in C_c^\infty(\mathcal{D})$, consider (j_n) a subsequence of (h_n) such that $j_n \longrightarrow \varphi$ for the topology of $C_c^\infty(\mathcal{D})$. Then,

$$(T - \tilde{T})(\varphi) = \lim_{n\to\infty} (T - \tilde{T})(j_n) = \lim_{n\to\infty} 0 = 0, \tag{6.21}$$

which shows that in fact, $\tilde{T}(\varphi) = T(\varphi)$.
Proof of (ii): if b can be extended to a continuous linear form over $L^q(\mathcal{D})$, then the estimate (2.20) is obviously true, by continuity over $L^q(\mathcal{D})$ of the said extension. Suppose now that (2.20) holds. Let $\varphi \in E_q$. Then L_φ, the continuous linear form over $C_c^\infty(\mathcal{D})$ defined by

$$\forall \psi \in C_c^\infty(\mathcal{D}), \quad L_\varphi(\psi) = b(\varphi, \psi), \tag{6.22}$$

verifies

$$\forall \psi \in E_q, \quad |L_\varphi(\psi)| \leqslant C\|\varphi\|_q \|\psi\|_q. \tag{6.23}$$

From the point (i), L_φ is a regular distribution with a representer $v_\varphi \in L^p(\mathcal{D})$ which is unique in $L^p(\mathcal{D})$. Define the map $B : E_q \to L^p(\mathcal{D})$ by $B\varphi = v_\varphi$. Then B verifies

$$\forall \varphi \in E_q, \forall \psi \in L^q(\mathcal{D}), \quad |\langle B\varphi, \psi\rangle_{L^p,L^q}| = |L_\varphi(\psi)| \leqslant C\|\varphi\|_q \|\psi\|_q. \tag{6.24}$$

Taking the supremum w.r.t. $\psi \in L^q(\mathcal{D})$ yields

$$\forall \varphi \in E_q, \quad \|B\varphi\|_p \leqslant C\|\varphi\|_q. \tag{6.25}$$

Observe now that the bilinearity of b yields $B(\varphi + \lambda\psi) = B\varphi + \lambda B\psi$ if $\varphi, \psi \in E_q$ and $\lambda \in \mathbb{Q}$. Taking the exact same steps as for the proof of point (i) and using equation (6.25), $B : E_q \to L^p(\mathcal{D})$ is Lipschitz continuous over E_q, and can thus be uniquely extended as a uniformly continuous map $\tilde{B} : L^q(\mathcal{D}) \to L^p(\mathcal{D})$. This relies on the fact that E_q is dense in $L^q(\mathcal{D})$ and that $L^q(\mathcal{D})$ is complete. As previously, one checks that $\tilde{B}$ is linear. Being uniformly continuous, it is then a bounded operator from $L^q(\mathcal{D})$ to $L^p(\mathcal{D})$ (its adjoint $\tilde{B}^*$ is then automatically bounded). Denote by $\tilde{b}$ the continuous bilinear form over $L^q(\mathcal{D})$ defined by

$$\tilde{b}(f, g) = \langle \tilde{B}f, g\rangle_{L^p,L^q}, \quad \forall f, g \in L^q(\mathcal{D}). \tag{6.26}$$

We now need to check that $\tilde{b}$ indeed coincides with b over $C_c^\infty(\mathcal{D})$, so that it is indeed an extension of b. For this, let $\varphi, \psi \in C_c^\infty(\mathcal{D})$ and $(\varphi_n), (\psi_n)$ two sequences of elements of E_q that converge to φ and ψ respectively, in the LF topology. Then b and $\tilde{b}$ coincide on E_q:

$$b(\varphi_n, \psi_m) = \tilde{b}(\varphi_n, \psi_m). \tag{6.27}$$

Observe that the following chain of equalities holds. It relies on the sequential continuity (for the LF topology of $C_c^\infty(\mathcal{D})$) of the linear forms $\varphi \mapsto b(\varphi, \psi), \psi \mapsto b(\varphi, \psi)$ and $T_v : \varphi \mapsto T_v(\varphi) = \langle v, \varphi\rangle_{L^q,L^p}$ for any $v \in L^q(\mathcal{D})$, as well equation (6.27).

$$\begin{aligned} b(\varphi, \psi) &= \lim_{n\to\infty} b(\varphi_n, \psi) = \lim_{n\to\infty}\lim_{m\to\infty} b(\varphi_n, \psi_m) = \lim_{n\to\infty}\lim_{m\to\infty} \tilde{b}(\varphi_n, \psi_m) \\ &= \lim_{n\to\infty}\lim_{m\to\infty} \langle \tilde{B}\varphi_n, \psi_m\rangle_{L^p,L^q} = \lim_{n\to\infty}\lim_{m\to\infty} T_{\tilde{B}\varphi_n}(\psi_m) = \lim_{n\to\infty} T_{\tilde{B}\varphi_n}(\psi) \\ &= \lim_{n\to\infty} \langle \tilde{B}\varphi_n, \psi\rangle_{L^p,L^q} = \lim_{n\to\infty} \langle \varphi_n, \tilde{B}^*\psi\rangle_{L^q,L^p} = \lim_{n\to\infty} T_{\tilde{B}*\psi}(\varphi_n) = T_{\tilde{B}*\psi}(\varphi) \\ &= \langle \varphi, \tilde{B}^*\psi\rangle_{L^q,L^p} = \langle \tilde{B}\varphi, \psi\rangle_{L^p,L^q} = \tilde{b}(\varphi, \psi). \end{aligned} \tag{6.28}$$

The uniqueness of b follows from the uniqueness of $\tilde{B}$ as an extension of B. □

Proof. (Lemma 2.5) Let (K_n) be an increasing sequence of compact subsets of $\mathcal{D}$ such that $\bigcup_n K_n = \mathcal{D}$. From the measurability of U and Tonelli's theorem, $\omega \mapsto \int_{K_n} |U_\omega(x)|dx$ is measurable and we have that

$$\mathbb{E}\Big[\int_{K_n} |U(x)|dx\Big] = \int_{K_n} \mathbb{E}[|U(x)|]dx = \sqrt{\frac{2}{\pi}}\int_{K_n} \sigma(x)dx < +\infty. \tag{6.29}$$

From equation (6.29), $\omega \mapsto \int_{K_n} |U_\omega(x)|dx$ is finite almost surely. Since the family (K_n) is countable, one obtains a set $\Omega_0 \subset \Omega$ of probability one such that for all $\omega \in \Omega_0$ and for all $n \in \mathbb{N}, \int_{K_n} |U_\omega(x)|dx < +\infty$. Given now any compact subset K of $\mathcal{D}$, there exists $N \in \mathbb{N}$ such that $K \subset K_N$ and thus for all $\omega \in \Omega_0$, $\int_K |U_\omega(x)|dx < +\infty$. Therefore, the sample paths of U lie in $L^1_{loc}(\mathcal{D})$ almost surely. From this fact and Fubini's theorem, we next obtain that given any $\varphi \in C_c^\infty$ and $|\alpha| \leqslant m$, the following map

$$U_\varphi^\alpha : \Omega \ni \omega \longmapsto \int_{\mathcal{D}} U_\omega(x)\partial^\alpha \varphi(x)dx \tag{6.30}$$

is a well defined random variable (i.e. it is measurable; see e.g. [20], Theorem 2.7, p. 62). Moreover, one can show that it is a limit in probability of suitably chosen Riemann sums of the integrand ([20], Theorem 2.8, p. 65). But here, those Riemann sums are all Gaussian random variables because U is a Gaussian process. Thus U_φ^α is a Gaussian random variable. a a limit in probability of Gaussian random variables. This also shows that $\{U_\varphi^\alpha, \varphi \in C_c^\infty(\mathcal{D})\}$ is in fact a Gaussian process, since the linearity of ∂^α yields

$$\sum_{i=1}^n a_i U_{\varphi_i}^\alpha = U^\alpha_{(\sum_{i=1}^n a_i\varphi_i)}, \tag{6.31}$$

and thus $\sum_{i=1}^n a_i U_{\varphi_i}^\alpha$ is a Gaussian random variable. An alternative proof of the Gaussianity of U_φ^α is found in [6], Example 2.3.16. p. 58-59. □

Proof. (Proposition 3.5) Let $\gamma > 0$, set $s_n := n^{-\gamma}$, $m_n := (s_n + s_{n+1})/2$ and consider the functions defined on $(0,1)$ by

$$\phi_n(x) := \mathbb{1}_{[s_{n+1},m_n]}(x) - \mathbb{1}_{[m_n,s_n)}(x), \quad \psi_n(x) = \int_0^x \phi_n(s)ds. \tag{6.32}$$

The functions ψ_n are nonnegative hat functions supported on $[s_{n+1}, s_n]$, with slope ± 1. Consider now the following covariance functions defined for *all* $(x,y) \in (0,1)^2$ by

$$k(x,y) = \sum_{n=1}^{+\infty} \psi_n(x)\psi_n(y), \quad g(x,y) = \sum_{n=1}^{+\infty} \phi_n(x)\phi_n(y). \tag{6.33}$$

The infinite sums above have in fact only one non zero term given any fixed $(x,y) \in (0,1)^2$, hence the functions g and k are well-defined. We now prove the announced properties on k.
<u>(i)</u> : since for fixed $(x,y) \in \mathcal{D} \times \mathcal{D}$, the sums in equation (6.33) only involve one basis function at a time, it is clear that for all $x, y \neq s_n$ or m_n, $\partial_x\partial_y k(x,y) = g(x,y)$. It is also clear that g and k are the covariance functions of the measurable Gaussian processes given by $U(x) = \sum_{n=1}^{+\infty} \xi_i\psi_n(x)$ and $V(x) = \sum_{n=1}^{+\infty} \xi_i\phi_n(x)$ respectively, where (ξ_i) is a sequence of independent standard Gaussian random variables (again, these Gaussian processes are well-defined as only one basis function is activated at a time, given $x \in (0,1)$).

$\underline{(ii)}$: observe that $g(x,x) = 1$ for all $x \in (0,1)$ (except possibly for $x = s_n$ or m_n for some $n \in \mathbb{N}$). Likewise,

$$k(x,x) = \int_0^x \int_0^x g(s,t)dsdt \leqslant \int_0^x \int_0^x |g(s,t)|dsdt \leqslant \left(\int_0^x g(s,s)^{1/2} ds \right)^2 = x^2. \tag{6.34}$$

Hence,

$$\int_0^1 g(x,x)^{p/2} dx = \int_0^1 dx < +\infty, \qquad \int_0^1 k(x,x)^{p/2} dx \leqslant \int_0^1 x^p dx < +\infty. \tag{6.35}$$

$\underline{(iii)}$: we have that

$$\|\psi_n\|_p^p = 2\int_0^{s_n - m_n} x^p dx = 2\frac{(s_n - m_n)^{p+1}}{p+1}, \quad \|\psi_n\|_p^2 = \left(\frac{2}{p+1}\right)^{2/p} (s_n - m_n)^{2+2/p}. \tag{6.36}$$

Hence, since $|s_n - m_n| \leqslant C/n^{\gamma+1}$ for some $C > 0$, $(s_n - m_n)^{2+2/p} \leqslant C'/n^{(2+2/p)(\gamma+1)}$ for some $C' > 0$ and $\sum_n \|\psi_n\|_p^2 < +\infty$. Next,

$$\|\psi_n'\|_p^p = \|\phi_n\|_p^p = \int_{s_{n+1}}^{s_n} dx = s_n - s_{n+1} \sim C/n^{\gamma+1}, \tag{6.37}$$

for some $C > 0$. Thus, $\|\phi_n\|_p^2 \sim C'/n^{2(\gamma+1)/p}$ and $\sum_n \|\phi_n\|_p^2$ converges only if $\gamma > p/2 - 1$. Therefore, our conterexample is found by taking any $\gamma \in (0, p/2-1]$ (observe that when $p \leqslant 2$, this interval becomes empty!) □

Proof. (Lemma 4.7) First, the map k is measurable over $\mathcal{D} \times \mathcal{D}$. Then, given a compact set $K \subset \mathcal{D} \times \mathcal{D}$, there exists a compact set $K_0 \subset \mathcal{D}$ such that $K \subset K_0 \times K_0$ (see e.g. the text before equation (3.11)). Then, using the Cauchy-Schwarz inequality for k,

$$\int_K |k(x,y)|dxdy \leqslant \int_{K_0 \times K_0} \sigma(x)\sigma(y)dxdy = \left(\int_{K_0} \sigma(x)dx \right)^2 < +\infty. \tag{6.38}$$

Therefore, $k \in L^1_{loc}(\mathcal{D} \times \mathcal{D})$ and for all mutli-index α, b_α is a bilinear continuous form over $C_c^\infty(\mathcal{D})$. From Lemma 2.2, b_α can be uniquely extended to a continuous bilinear form over $L^2(\mathcal{D})$. Denote by $\mathcal{F}_k^\alpha$ the associated bounded operator over $L^2(\mathcal{D})$. We now need to show that $\mathcal{F}_k^\alpha$ is self-adjoint and nonnegative. First note that for all $\varphi, \psi \in C_c^\infty(\mathcal{D})$,

$$\langle \mathcal{F}_k^\alpha \varphi, \psi \rangle_{L^2} = \int_{\mathcal{D} \times \mathcal{D}} k(x,y) \partial^\alpha \varphi(x) \partial^\alpha \psi(y) dydx = \langle \varphi, \mathcal{F}_k^\alpha \psi \rangle_{L^2}. \tag{6.39}$$

Equation (6.39), conjoined with the density of $C_c^\infty(\mathcal{D})$ in $L^2(\mathcal{D})$ and the continuity of the bilinear form $(f,g) \mapsto \langle \mathcal{F}_k^\alpha f, g \rangle_{L^2}$ yields that $\langle \mathcal{F}_k^\alpha f, g \rangle_{L^2} = \langle f, \mathcal{F}_k^\alpha g \rangle_{L^2}$ for all $f, g \in L^2(\mathcal{D})$. Therefore $\mathcal{F}_k^\alpha$ is self-adjoint. For the positivity, consider again $\varphi \in C_c^\infty(\mathcal{D})$. Then from Fubini's theorem (justified below),

$$\begin{aligned} \langle \mathcal{F}_k^\alpha \varphi, \varphi \rangle &= \int_{\mathcal{D} \times \mathcal{D}} k(x,y) \partial^\alpha \varphi(x) \partial^\alpha \varphi(y) dydx = \int_{\mathcal{D} \times \mathcal{D}} \mathbb{E}[U(x)U(y)] \partial^\alpha \varphi(x) \partial^\alpha \varphi(y) dydx \\ &= \mathbb{E}\left[\left(\int_{\mathcal{D}} U(x) \partial^\alpha \varphi(x) dx \right)^2 \right] \geqslant 0. \end{aligned} \tag{6.40}$$

Indeed the following integrability condition holds, setting $K = \mathrm{Supp}(\varphi)$:

$$\begin{aligned}\mathbb{E}\Big[\int_{\mathcal{D}\times\mathcal{D}} |\partial^\alpha\varphi(x)\partial^\alpha\varphi(y)U(x)U(y)|dxdy\Big] &= \int_{K\times K} |\partial^\alpha\varphi(x)\partial^\alpha\varphi(y)|\mathbb{E}[|U(x)U(y)|]dxdy \\ \leqslant \int_{K\times K} |\partial^\alpha\varphi(x)\partial^\alpha\varphi(y)|\sigma(x)\sigma(y)dxdy &= \Big(\int_K |\partial^\alpha\varphi(x)|\sigma(x)dx\Big)^2 \\ \leqslant \sup_{x\in K} |\partial^\alpha\varphi(x)|^2\Big(\int_K \sigma(x)dx\Big)^2 &< +\infty. \end{aligned} \tag{6.41}$$

Equation (6.41), conjoined with the density of $C_c^\infty(\mathcal{D})$ in $L^2(\mathcal{D})$ and the continuity of the quadratic form $f \mapsto \langle \mathcal{F}_k^\alpha f, f\rangle_{L^2}$ yields that $\langle \mathcal{F}_k^\alpha f, f\rangle_{L^2} \geqslant 0$ for all $f \in L^2(\mathcal{D})$. Therefore $\mathcal{F}_k^\alpha$ is nonnegative. □

Proof. (Lemma 4.8) From the definition of b_α over $C_c^\infty(\mathcal{D})$,

$$\begin{aligned} b_\alpha(\varphi,\psi) &= \int_{\mathcal{D}\times\mathcal{D}} k(x,y)\partial^\alpha\varphi(x)\partial^\alpha\psi(y)dxdy = \int_{\mathcal{D}\times\mathcal{D}} \partial^{\alpha,\alpha}k(x,y)\varphi(x)\psi(y)dxdy \\ &= \langle \mathcal{E}_k^\alpha\varphi,\psi\rangle_{L^2}. \end{aligned} \tag{6.42}$$

From Cauchy-Schwarz's inequality, it verifies

$$\forall\varphi,\psi \in C_c^\infty(\mathcal{D}),\ |b_\alpha(\varphi,\psi)| \leqslant \|\partial^{\alpha,\alpha}k\|_2\|\varphi\|_2\|\psi\|_2 \tag{6.43}$$

From Lemma 4.7, there exists a unique bounded, self-adjoint and nonnegative operator $\mathcal{F}_k^\alpha$ over $L^2(\mathcal{D})$ such that $b_\alpha(\varphi,\psi) = \langle \mathcal{F}_k^\alpha\varphi,\psi\rangle_{L^2}$ for all $\varphi,\psi \in C_c^\infty(\mathcal{D})$. The uniqueness of $\mathcal{F}_k^\alpha$ and equation (6.42) yield $\mathcal{F}_k^\alpha = \mathcal{E}_k^\alpha$, and thus $\mathcal{E}_k^\alpha$ is self-adjoint and nonnegative. □

Proof. (Lemma 4.9) $\underline{(i)}$: Let $n \in \mathbb{N}_0$ be such that $\lambda_n \neq 0$. Let $\varphi \in C_c^\infty(\mathcal{D})$. Then

$$\begin{aligned}\lambda_n\Big(\int_{\mathcal{D}} \phi_n(x)\partial^\alpha\varphi(x)dx\Big)^2 &\leqslant \sum_{i=1}^{+\infty}\lambda_i\Big(\int_{\mathcal{D}} \phi_i(x)\partial^\alpha\varphi(x)dx\Big)^2 \\ &\leqslant \sum_{i=1}^{+\infty}\lambda_i\int_{\mathcal{D}\times\mathcal{D}} \phi_i(x)\phi_i(y)\partial^\alpha\varphi(x)\partial^\alpha\varphi(y)dxdy \\ &\leqslant \int_{\mathcal{D}\times\mathcal{D}} k(x,y)\partial^\alpha\varphi(x)\partial^\alpha\varphi(y)dxdy \\ &\leqslant \int_{\mathcal{D}\times\mathcal{D}} \partial^{\alpha,\alpha}k(x,y)\varphi(x)\varphi(y)dxdy \\ &\leqslant \|\partial^{\alpha,\alpha}k\|_{L^2(\mathcal{D}\times\mathcal{D})}\|\varphi\|^2_{L^2(\mathcal{D})}. \end{aligned} \tag{6.44}$$

Therefore, from Lemma 2.1, $\partial^\alpha\phi_n \in L^2(\mathcal{D})$.
$\underline{(ii)}$: introduce the finite rank kernel k_n defined by

$$k_n(x,y) = \sum_{i=1}^{n}\lambda_i\phi_i(x)\phi_i(y). \tag{6.45}$$

Then its mixed derivative $\partial^{\alpha,\alpha}k_n(x,y)$ is equal to $\sum_{i=1}^n \lambda_i \partial^\alpha\phi_i(x)\partial^\alpha\phi_i(y)$ in $L^2(\mathcal{D}\times\mathcal{D})$ and the associated operator $\mathcal{E}_{k_n}^\alpha$ is trace class, with

$$\mathrm{Tr}(\mathcal{E}_{k_n}^\alpha) = \sum_{j=1}^{+\infty}\langle \mathcal{E}_{k_n}^\alpha \phi_j, \phi_j\rangle_{L^2} = \sum_{j=1}^{+\infty}\sum_{i=1}^{n}\lambda_i\langle \partial^\alpha\phi_i, \phi_j\rangle_{L^2}^2 \tag{6.46}$$

$$= \sum_{i=1}^{n}\lambda_i\sum_{j=1}^{+\infty}\langle \partial^\alpha\phi_i, \phi_j\rangle_{L^2}^2 = \sum_{i=1}^{n}\lambda_i\|\partial^\alpha\phi_i\|_{L^2}^2. \tag{6.47}$$

Now, observe that $\mathcal{E}_{k_n}^\alpha \leqslant \mathcal{F}_k^\alpha$ in the sense of the Loewner order. Indeed, let first $\varphi \in C_c^\infty(\mathcal{D})$:

$$\langle(\mathcal{F}_k^\alpha - \mathcal{E}_{k_n}^\alpha)\varphi,\varphi\rangle_{L^2} = \langle(\mathcal{E}_k - \mathcal{E}_{k_n})\partial^\alpha\varphi,\partial^\alpha\varphi\rangle_{L^2} = \sum_{i=n+1}^{+\infty}\lambda_i\langle\phi_i,\partial^\alpha\varphi\rangle_{L^2}^2 \geqslant 0. \tag{6.48}$$

The density of $C_c^\infty(\mathcal{D})$ in $L^2(\mathcal{D})$ and the continuity of the quadratic form $f \mapsto \langle(\mathcal{F}_k^\alpha - \mathcal{E}_{k_n}^\alpha)f, f\rangle_{L^2}$ over $L^2(\mathcal{D})$ yields indeed that $\mathcal{E}_{k_n}^\alpha \leqslant \mathcal{F}_k^\alpha$. Taking the trace :

$$\sum_{i=1}^{n}\lambda_i\|\partial^\alpha\phi_i\|_{L^2}^2 = \mathrm{Tr}(\mathcal{E}_{k_n}^\alpha) = \sum_{j=1}^{+\infty}\langle \mathcal{E}_{k_n}^\alpha\phi_j,\phi_j\rangle_{L^2} \leqslant \sum_{j=1}^{+\infty}\langle \mathcal{F}_k^\alpha\phi_j,\phi_j\rangle_{L^2} = \mathrm{Tr}(\mathcal{F}_k^\alpha). \tag{6.49}$$

Taking the limit when n goes to infinity yields $\sum_{i=1}^{+\infty}\lambda_i\|\partial^\alpha\phi_i\|_{L^2}^2 \leqslant \mathrm{Tr}(\mathcal{F}_k^\alpha)$. This shows that if $\sum_{i=1}^{+\infty}\lambda_i\|\partial^\alpha\phi_i\|_{L^2}^2 = +\infty$, then $\mathrm{Tr}(\mathcal{F}_k^\alpha) = +\infty$. Suppose now that $\mathrm{Tr}(\mathcal{F}_k^\alpha) < +\infty$. Equation (6.49) shows that the series of functions $\sum_i \lambda_i\partial^\alpha\phi_i \otimes \partial^\alpha\phi_i$ converges in norm in $L^2(\mathcal{D}\times\mathcal{D})$. Moreover, we check that it is equal to $\partial^{\alpha,\alpha}k$: taking $\varphi \in C_c^\infty(\mathcal{D}\times\mathcal{D})$, then

$$\int_{\mathcal{D}\times\mathcal{D}} k(x,y)\partial^{\alpha,\alpha}\varphi(x,y)dxdy = \sum_i\lambda_i\int_{\mathcal{D}\times\mathcal{D}}\phi_i(x)\phi_i(y)\partial^{\alpha,\alpha}\varphi(x,y)dxdy \tag{6.50}$$

$$= \sum_i\lambda_i\int_{\mathcal{D}\times\mathcal{D}}\partial^\alpha\phi_i(x)\partial^\alpha\phi_i(y)\varphi(x,y)dxdy \tag{6.51}$$

$$= \int_{\mathcal{D}\times\mathcal{D}}\left(\sum_i\lambda_i\partial^\alpha\phi_i(x)\partial^\alpha\phi_i(y)\right)\varphi(x,y)dxdy. \tag{6.52}$$

Moreover, since we have shown that $\partial^{\alpha,\alpha}k \in L^2(\mathcal{D}\times\mathcal{D})$, Lemma 4.8 implies that $\mathcal{F}_k^\alpha = \mathcal{E}_k^\alpha$. We can then write, following the steps of equation (6.46),

$$\mathrm{Tr}(\mathcal{F}_k^\alpha) = \mathrm{Tr}(\mathcal{E}_k^\alpha) = \sum_{j=1}^{+\infty}\langle\mathcal{E}_k^\alpha\phi_j,\phi_j\rangle_{L^2} = \sum_{j=1}^{+\infty}\sum_i\lambda_i\langle\partial^\alpha\phi_i,\phi_j\rangle_{L^2}^2 = \sum_{i=1}^{+\infty}\lambda_i\|\partial^\alpha\phi_i\|_{L^2}^2. \tag{6.53}$$

Suppose now that $\sum_{i=1}^{+\infty}\lambda_i\|\partial^\alpha\phi_i\|_{L^2}^2 < +\infty$. Then as observed before, the series of functions $\sum_i\lambda_i\partial^\alpha\phi_i\otimes\partial^\alpha\phi_i$ converges in norm in $L^2(\mathcal{D}\times\mathcal{D})$, one verifies that $\partial^{\alpha,\alpha}k$ exists in $L^2(\mathcal{D})$ and is in fact given by

$$\partial^{\alpha,\alpha}k = \sum_i\lambda_i\partial^\alpha\phi_i\otimes\partial^\alpha\phi_i \quad \text{in} \quad L^2(\mathcal{D}\times\mathcal{D}). \tag{6.54}$$

Finally, since $\partial^{\alpha,\alpha}k \in L^2(\mathcal{D}\times\mathcal{D})$, $\mathcal{E}_k^\alpha$ is bounded over $L^2(\mathcal{D})$ and from equation (6.54),

$$\sum_{i=1}^{+\infty}\lambda_i\|\partial^\alpha\phi_i\|_{L^2}^2 = \sum_{i=1}^{+\infty}\lambda_i\sum_j\langle\partial^\alpha\phi_i,\phi_j\rangle_{L^2}^2 \tag{6.55}$$

$$= \sum_j\sum_{i=1}^{+\infty}\lambda_i\left(\int_{\mathcal{D}}\partial^\alpha\phi_i(x)\phi_j(x)dx\right)^2 \tag{6.56}$$

$$= \sum_j\int_{\mathcal{D}\times\mathcal{D}}\sum_i\lambda_i\partial^\alpha\phi_i(x)\partial^\alpha\phi_i(y)\phi_j(x)\phi_j(y)dxdy \tag{6.57}$$

$$= \sum_j\langle\mathcal{E}_k^\alpha\phi_j,\phi_j\rangle_{L^2} = \mathrm{Tr}(\mathcal{E}_k^\alpha). \tag{6.58}$$

Therefore $\mathcal{E}_k^\alpha$ is trace class and $\mathrm{Tr}(\mathcal{E}_k^\alpha) = \sum_{i=1}^{+\infty}\lambda_i\|\partial^\alpha\phi_i\|_{L^2}^2$. Moreover, from Lemma 4.8, $\mathcal{E}_k^\alpha = \mathcal{F}_k^\alpha$. To see that this also finishes to prove equation (4.20) in the infinite case, observe that if $\mathrm{Tr}(\mathcal{F}_k^\alpha) = +\infty$, then the previous computation implies that the series $\sum_i\lambda_i\|\partial^\alpha\phi_i\|_2^2 = +\infty$: if this were not the case, Lemma 4.8 would apply again and we would have $\mathcal{E}_k^\alpha = \mathcal{F}_k^\alpha$, which would then be trace class. For asymmetric derivatives, simply observe that for all $|\alpha|,|\beta| \leqslant m$,

$$\|\partial^\alpha\phi_i\otimes\partial^\beta\phi_i\|_2 = \|\partial^\alpha\phi_i\|_2\|\partial^\beta\phi_i\|_2 \leqslant \frac{\|\partial^\alpha\phi_i\|_2^2+\|\partial^\beta\phi_i\|_2^2}{2}. \tag{6.59}$$

Therefore the norm convergence of the series $\sum_{i\in\mathbb{N}}\lambda_i\|\partial^\alpha\phi_i\otimes\partial^\alpha\phi_i\|_{L^2}$ for all $|\alpha|\leqslant m$ implies that of all the series of the form $\sum_{i\in\mathbb{N}}\lambda_i\|\partial^\alpha\phi_i\otimes\partial^\beta\phi_i\|_{L^2}$ converge, provided that $|\alpha|\leqslant m$ and $|\beta|\leqslant m$. As previously, one then deduces that $\partial^{\alpha,\beta}k = \sum_{i=0}^{\infty}\lambda_i\partial^\alpha\phi_i\otimes\partial^\beta\phi_i$. □

References

[1] R. A. Adams and J. J. Fournier. *Sobolev spaces.* Pure and Applied Mathematics. Academic Press, 2 edition, 2003.

[2] R. J. Adler and J. E. Taylor. *Random Fields and Geometry.* Springer-Verlag New York, NY, 2007.

[3] J.-M. Azaïs and M. Wschebor. *Level sets and extrema of random processes and fields.* Wiley & Sons, 2009.

[4] A. Bella, M. Bonanzinga, and M. Matveev. Sequential + separable vs sequentially separable and another variation on selective separability. *Cent. Eur. J. Math.*, 11(3):530–538, 2013.

[5] A. Berlinet and C. Thomas-Agnan. *Reproducing Kernel Hilbert Spaces in Probability and Statistics.* Springer US, 2004.

[6] V. I. Bogachev. *Gaussian measures.* Number 62 in Mathematical Surveys and Monographs. American Mathematical Soc., 1998.

[7] V. I. Bogachev and M. A. S. Ruas. *Measure theory*, volume 1. Springer, 2007.

[8] S. C. Brenner, L. R. Scott, and L. R. Scott. *The mathematical theory of finite element methods*, volume 3. Springer, 2008.

[9] H. Brézis. *Functional analysis, Sobolev spaces and partial differential equations.* Springer New York, NY, 2011.

[10] C. Brislawn. Kernels of trace class operators. *Proc. Amer. Math. Soc.*, 104(4):1181–1190, 1988.

[11] Y. Chen, B. Hosseini, H. Owhadi, and A. M. Stuart. Solving and learning nonlinear PDEs with Gaussian processes. *J. Comput. Phys.*, 447:Paper No. 110668, 29, 2021.

[12] S. A. Chobanjan and V. I. Tarieladze. Gaussian characterizations of certain Banach spaces. *J. Multivariate Anal.*, 7(1):183–203, 1977.

[13] Z. Ciesielski. Modulus of smoothness of the Brownian paths in the L^p norm. *Constructive theory of functions (Varna, Bulgaria, 1991)*, pages 71–75, 1991.

[14] Z. Ciesielski, G. Kerkyacharian, and B. Roynette. Quelques espaces fonctionnels associés à des processus gaussiens. *Studia Math.*, 107(2):171–204, 1993.

[15] B. S. Cirel'son, I. A. Ibragimov, and V. N. Sudakov. Norms of Gaussian sample functions. In *Proceedings of the Third Japan—USSR Symposium on Probability Theory*, pages 20–41. Springer, 1976.

[16] J. B. Conway. *A course in functional analysis*, volume 96. Springer, 2019.

[17] R. A. DeVore and R. C. Sharpley. Besov spaces on domains in $\mathbf{R}^d$. *Trans. Amer. Math. Soc.*, 335(2):843–864, 1993.

[18] E. Di Nezza, G. Palatucci, and E. Valdinoci. Hitchhiker's guide to the fractional Sobolev spaces. *Bull. Sci. Math.*, 136(5):521–573, 2012.

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

[20] J. L. Doob. *Stochastic processes.* Wiley Classics Library. Wiley-Interscience, 1990.

[21] R. M. Dudley. The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. *J. Functional Analysis*, 1:290–330, 1967.

[22] D. E. Edmunds and H. Triebel. *Function spaces, entropy numbers, differential operators*, volume 120 of *Cambridge Tracts in Mathematics.* Cambridge University Press, Cambridge, 1996.

[23] L. Evans. *Partial Differential Equations.* Graduate studies in mathematics. American Mathematical Society, 1998.

[24] I. Henderson, P. Noble, and O. Roustant. Characterization of the second order random fields subject to linear distributional pde constraints. *Bernoulli*, 29(4):3396–3422, 2023.

[25] I. A. Ibragimov. Conditions for Gaussian homogeneous fields to belong to classes H_p^r. *Journal of Mathematical Sciences*, 68(4):484–497, 1994.

[26] J. Jørgensen. *Probability With a View Towards Statistics, Volume II.* Chapman and Hall/CRC, 1994.

[27] M. Kanagawa, P. Hennig, D. Sejdinovic, and B. K. Sriperumbudur. Gaussian processes and kernel methods: A review on connections and equivalences. *arXiv preprint arXiv:1807.02582*, 2018.

[28] G. Kerkyacharian, S. Ogawa, P. Petrushev, and D. Picard. Regularity of Gaussian processes on Dirichlet spaces. *Constr. Approx.*, 47(2):277–320, 2018.

[29] P. T. Lai. Noyaux d'Agmon. *Séminaire Jean Leray*, 277(4):1–37, 1973-1974.

[30] M. Lange-Hegermann. Algorithmic linearly constrained Gaussian processes. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

[31] G. Leoni. *A first course in Sobolev spaces*, volume 105 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2009.

[32] V. Linde, V. I. Tarieladze, and S. A. Chobanyan. Characterization of certain classes of Banach spaces by properties of Gaussian measures. *Theory of Probability & Its Applications*, 25(1):159–164, 1980.

[33] A. Mas. Representation of small ball probabilities in hilbert space and lower bound in regression for functional data. *Electronic Journal of Statistics*, 6:1745–1778, 2012.

[34] S. M. Nikol'skii. *Approximation of functions of several variables and imbedding theorems*, volume 205. Springer Science & Business Media, 2012.

[35] H. Owhadi. Bayesian numerical homogenization. *Multiscale Modeling & Simulation*, 13(3):812–828, 2015.

[36] H. Owhadi and C. Scovel. Separability of reproducing kernel spaces. *Proceedings of the American Mathematical Society*, 145(5):2131–2138, 2017.

[37] M. Raissi, P. Perdikaris, and G. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

[38] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Machine learning of linear differential equations using Gaussian processes. *Journal of Computational Physics*, 348:683–693, 2017.

[39] C. E. Rasmussen and C. Williams. *Gaussian Processes for Machine Learning*. the MIT Press, 2006.

[40] H. L. Royden and P. Fitzpatrick. *Real analysis*, volume 32. Macmillan New York, 1988.

[41] B. Roynette. Mouvement Brownien et espaces de Besov. *Stochastics: An International Journal of Probability and Stochastic Processes*, 43(3-4):221–260, 1993.

[42] M. Scheuerer. Regularity of the sample paths of a general second order random field. *Stochastic Processes and their Applications*, 120(10):1879–1897, 2010.

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

[44] E. M. Stein. *Singular integrals and differentiability properties of functions*, volume 2. Princeton university press, 1970.

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

[46] M. Talagrand. *Upper and lower bounds for stochastic processes*, volume 60. Springer, 2014.

[47] T. Tao. *Nonlinear dispersive equations: local and global analysis*. American Mathematical Soc., 2006.

[48] F. Treves. *Topological Vector Spaces, Distributions and Kernels*. Dover books on mathematics. Dover Publications, 2006.

[49] H. Triebel. Über die Approximationszahlen der Einbettungsoperatoren $J(B^r_{p,q}(\Omega) \to B^{r'}_{p',q'}(S))$. *Archiv der Mathematik*, 19(3):305–312, 1968.

[50] A. Van Der Vaart and H. Van Zanten. Information rates of nonparametric Gaussian process methods. *Journal of Machine Learning Research*, 12(6), 2011.

[51] N. Wahlstrom, M. Kok, T. B. Schön, and F. Gustafsson. Modeling magnetic fields using Gaussian processes. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3522–3526, 2013.