Periodic Splines and Gaussian Processes for the Resolution of Linear Inverse Problems

Anaïs Badoual*, Julien Fageot*, and Michael Unser

***Abstract*—This paper deals with the resolution of inverse problems in a periodic setting or, in other terms, the reconstruction of periodic continuous-domain signals from their noisy measurements. We focus on two reconstruction paradigms: variational and statistical. In the variational approach, the reconstructed signal is solution to an optimization problem that establishes a tradeoff between fidelity to the data and smoothness conditions via a quadratic regularization associated to a linear operator. In the statistical approach, the signal is modeled as a stationary random process defined from a Gaussian white noise and a whitening operator; one then looks for the optimal estimator in the mean-square sense. We give a generic form of the reconstructed signals for both approaches, allowing for a rigorous comparison of the two. We fully characterize the conditions under which the two formulations yield the same solution, which is a periodic spline in the case of sampling measurements. We also show that this equivalence between the two approaches remains valid on simulations for a broad class of problems. This extends the practical range of applicability of the variational method.**

***Index Terms*—Periodic signals, variational methods, representer theorem, Gaussian processes, MMSE estimators, splines.**

I. Introduction

This paper deals with inverse problems: one aims at recovering an unknown signal from its corrupted measurements. To be more specific, the motivation of this work is the reconstruction of an unknown *continuous-domain* and *periodic* signal f from its M noisy measurements $y_m \approx \langle \nu_m, f \rangle = \int_0^1 \nu_m(t) f(t) \mathrm{d}t$ for $m = 1 \dots M$, where the ν_m are measurement functions. The goal is then to build an output signal f_{opt} that is as close as possible to f.

A. Inverse Problems in the Continuous Domain

Inverse problems are often formulated in the discrete domain [1]–[5]. This is motivated by the need of manipulating digital data on computers. Nevertheless, many naturally occurring signals depend on continuous variables (*e.g.*, time or position). This leads us to attempt recovering a signal $f_{\mathrm{opt}}(t)$ that depends on the continuous variable $t \in [0, 1]$. In contrast with the classical discrete setting, our search space for this reconstructed signal is thus infinite-dimensional [6]. Moreover, we choose a regularization based on true derivatives (as opposed to finite differences) to impose some smoothness on the reconstructed signal, a concept that is absent in the discrete setting.

When considering continuous-domain reconstruction methods, a majority of works, typically in machine learning, deal with sampling measurements. The goal is then to recover f from its (possibly noisy) values $y_m \approx f(t_m)$ at fixed location t_m. In order to investigate a more general version of inverse problems, we shall consider generalized measurements [7], [8]. They largely exceed the sampling case and include Fourier sampling or convolution (*e.g.*, MRI, x-ray tomography [9], [10]). Our only requirement is that the measurements y_m depend linearly on, and evolve continuously with, the unknown signal f up to some additive noise, so that $y_m \approx \langle \nu_m, f \rangle$.

B. Variational vs. Statistical Methods

In the discrete domain, two standard strategies are used to reconstruct an input signal $\mathbf{x}$ from its noisy measurements $\mathbf{y} \approx \mathbf{H}\mathbf{x}$, where $\mathbf{H}$ models the acquisition process [5]. The first approach is deterministic and can be tracked back to the '60s with Tikhonov's seminal work [11]. The ill-posedness of the problem usually imposes the addition of a regularizer. By contrast, Wiener filtering is based on the stochastic modelization of the signals of interest and the optimal estimation of the targeted signal $\mathbf{x}$. This paper generalizes these ideas for the reconstruction of *continuous* signals from their *discrete* measurements.

In the variational setting, the reconstructed signal is a solution to an optimization problem that imposes some smoothness conditions [12]. More precisely, the optimization problem may take the form

$$f_{\mathrm{opt}} = \arg\min_f \left(\sum_{m=1}^{M} \left(y_m - \langle \nu_m, f \rangle \right)^2 + \lambda \| \mathrm{L} f \|_{L_2}^2 \right), \quad (1)$$

where L is a linear operator. The first term in (1) controls the data fidelity. The regularization term $\| \mathrm{L} f \|_{L_2}^2$ constrains the function to satisfy certain smoothness properties (for this reason, the variational approach is sometimes called a smoothing approach). The parameter λ in (1) quantifies the tradeoff between the fidelity to the data and the regularization constraint.

In the statistical setting, the signal is modeled as a random process and is optimally reconstructed using estimation theory [13]. More precisely, one assumes that the continuous-domain signal is the realization of a stochastic process s and that the samples are given by $y_m = \langle \nu_m, s \rangle + \epsilon_m$, where ϵ_m is a random perturbation and ν_m a linear measurement function. In

Copyright (c) 2015 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

The authors are with the Biomedical Imaging Group, École polytechnique fédérale de Lausanne (EPFL), Switzerland. This work was funded by the Swiss National Science Foundation under Grant 200020-162343 and by the ERC grant agreement No 692726 - GlobalBioIm.

* These authors contributed equally to this work.

this case, one specifies the reconstructed signal as the optimal statistical estimator in the mean-square sense

$$f_{\text{opt}} = \arg\min_{\tilde{s}} \mathbb{E}\left[\|s - \tilde{s}(\cdot|\mathbf{y})\|_{L_2}^2\right], \tag{2}$$

where the estimators $t \mapsto \tilde{s}(t|\mathbf{y})$ are computed from the generalized samples y_m. The solution depends on the measurement function ν_m and the stochastic models specified for s and ϵ_m. In our case, the random process s is characterized by a linear operator L that is assumed to have a whitening effect (it transforms s into a periodic Gaussian white noise), while the perturbation is i.i.d. Gaussian.

C. Periodic and General Setting

The variational and statistical approaches have been extensively studied for continuous-domain signals defined on the infinitely supported real line. However, it is often assumed in practice that the input signals are periodic. In fact, a standard computational approach to signal processing is to extend by periodization the signals of otherwise bounded support. Periodic signals arise also naturally in applications such as the parametric representation of closed curves [14]–[16]. This has motivated the development of signal-processing tools and techniques specialized to periodic signals in sampling theory, error analysis, wavelets, stochastic modelization, or curve representation [17]–[23].

In this paper, we develop the theory of the variational and statistical approaches for periodic continuous-domain signals in a very general context, including the following aspects:

- We consider a broad class of measurement functions, with the only assumptions that they are linear and continuous.
- Both methods refer to an underlying linear operator L that affects the smoothness properties of the reconstruction. We deal with a very broad class of linear operators acting on periodic functions.
- We consider possibly non-quadratic data fidelity terms in the smoothing approach.

D. Related Works

The topics investigated in this paper have already received some attention in the literature, mostly in the non-periodic setting.

a) Reconstruction over the Real Line: Optimization problems of the form (1) appear in many fields and receive different names, including inverse problems in image processing [5], representer theorems in machine learning [24], or sometimes interpolation elsewhere. Schoenberg was the first to show the connection between (1) and spline theory [25]. Since then, this has been extended to other operators [26], or to the interpolation of the derivative of the signal [27], [28]. Many recent methods are dealing with non-quadratic regularization, especially the ones interested in the reconstruction of sparse discrete [29], [30] or continuous signals [6], [31]–[33]. We discuss this aspect more extensively in Section VI-B.

A statistical framework requires the specification of the noise and of the signal stochastic model. The signal is then estimated from its measurements. A classical measure of the quality of an estimator is the mean-square error. This criterion is minimized by the minimum mean-square error (MMSE) estimator [13], [34]. The theory has been developed mostly for Gaussian processes and in the context of sampling measurements [35]. We are especially interested in innovation models, for which one assumes that the signal can be whitened (*i.e.*, transformed into a white noise) by the application of a linear operator [36], [37]. Non-periodic models have been studied in many situations, including the random processes associated with differential [38], [39] or fractional operators [40]. Extensions to non-Gaussian models are extensively studied by Unser and Tafti [41].

The statistical and variational frameworks are deeply connected. It is remarkable that the solution of either problem can be expressed as spline functions in relation with the linear operator L involved in regularization (variational approach) or whitening (statistical approach). Wahba has shown that the two approaches are strictly equivalent in the case of stationary Gaussian models [42]. This equivalence has also been recognized by several authors since then, as shown by Berlinet and Thomas-Agnan [35], and Unser and Blu [43]. In the non-stationary case, this equivalence is not valid any more and the existence of connections has received less attention.

b) Reconstruction of Periodic Signals: Some strong practical concerns have motivated the need for an adaptation of the theory to the periodic setting. Important contributions in that direction have been proposed. Periodic splines are constructed and applied to sampling problems by Schoenberg [44] and Golomb [45]. The smoothing spline approach is studied in the periodic setting by Wahba [42] for derivative operators of any order. Although the periodic extension of the classical theory is briefly mentioned by several authors [35], [42], [46], we are not aware of a global treatment. Providing a general analysis in the periodic setting is precisely what we propose in this paper.

E. Outline and Main Contributions

Section II contains the main notations and tools for periodic functions and operators. In Section III, we state the periodic representer theorem (Theorem 1). It fully specifies the form of the solution in the variational approach in a very general setting. For the specific case of sampling measurements, we show that this solution is a periodic spline (Proposition 5). Section IV is dedicated to the statistical approach. We introduce a class of periodic stationary processes (the *Gaussian bridges*) for which we specify the MMSE estimator in the case of generalized linear measurements (Theorem 2). We also provide a theoretical comparison between the variational and statistical approaches by reformulating the MMSE estimation as the solution of a new optimization problem (Proposition 7). This highlights the strict equivalence of the two approaches for invertible operators and extends known results from sampling to generalized linear measurements. For non-invertible operators, we complete our analysis with simulations in Section V. In particular, we give empirical evidence of the practical relevance of the variational approach for the reconstruction of periodic stationary signals. We provide in Section VI a

comparison between our results in the periodic setting and the known results over the real line. Finally, we conclude in Section VII. All the proofs have been postponed to the Appendix sections.

II. Mathematical Background for Periodic Signals

Throughout the paper, we consider periodic functions and random processes. Without loss of generality, the period can always be normalized to one. Moreover, we identify a periodic function over $\mathbb{R}$ with its restriction to a single period, chosen to be $\mathbb{T} = [0, 1)$. We use the symbols f, s, and $\tilde{s}$ to specify a function, a random process, and an estimator of s, respectively.

We call $\mathcal{S}(\mathbb{T})$ the space of 1-periodic functions that are infinitely differentiable, $\mathcal{S}'(\mathbb{T})$ the space of 1-periodic generalized functions (dual of $\mathcal{S}(\mathbb{T})$), and $L_2(\mathbb{T})$ the Hilbert space of square integrable 1-periodic functions associated with the norm $\|f\|_{L_2} = (\int_0^1 |f(t)|^2 \mathrm{d}t)^{1/2}$. Working with $\mathcal{S}'(\mathbb{T})$ allows us to deal with functions with no pointwise interpretation, such as the Dirac comb defined by

$$\text{Ш} = \sum_{k\in\mathbb{Z}} \delta(\cdot - k), \tag{3}$$

where δ is the Dirac impulse. The duality product between an element $f \in \mathcal{S}'(\mathbb{T})$ and a smooth function $g \in \mathcal{S}(\mathbb{T})$ is denoted by $\langle f, g\rangle$. For instance, $\langle \text{Ш}, g\rangle = g(0)$ for every g. When the two real functions are in $L_2(\mathbb{T})$, we simply have the usual scalar product $\langle f, g\rangle = \int_0^1 f(t)g(t)\mathrm{d}t$. All these concepts are extended to complex-valued functions in the usual manner with the convention that $\langle f, g\rangle = \int_0^1 f(t)\overline{g(t)}\mathrm{d}t$ for square-integrable functions. The complex sinusoids are denoted by $e_k(t) = \mathrm{e}^{\mathrm{j}2\pi kt}$ for any $k \in \mathbb{Z}$ and $t \in \mathbb{T}$. Any periodic generalized function $f \in \mathcal{S}'(\mathbb{T})$ can be expanded as

$$f(t) = \sum_{k\in\mathbb{Z}} \widehat{f}[k]\mathrm{e}^{\mathrm{j}2\pi kt} = \sum_{k\in\mathbb{Z}} \widehat{f}[k]e_k(t), \tag{4}$$

where the $\widehat{f}[k]$ are the Fourier coefficients of f, given by $\widehat{f}[k] = \langle f, e_k\rangle$. Finally, the convolution between two periodic functions f and g is given by

$$(f * g)(t) = \langle f, g(t - \cdot)\rangle. \tag{5}$$

If $f, g \in L_2(\mathbb{T})$, we have that $(f*g)(t) = \int_0^1 f(\tau)g(t-\tau)\mathrm{d}\tau$.

A. Linear and Shift-Invariant Operators

Let L be a linear, shift-invariant (LSI), and continuous operator from $\mathcal{S}(\mathbb{T})$ to $\mathcal{S}'(\mathbb{T})$. The shift invariance implies the existence of $\widehat{L}[k] \in \mathbb{C}$ such that

$$\mathrm{L}e_k = \widehat{L}[k]e_k, \tag{6}$$

for any $k \in \mathbb{Z}$. We call $\widehat{L}[k]$ the frequency response of the operator L; it is also given by

$$\widehat{L}[k] = \langle \mathrm{L}\{\text{Ш}\}, e_k\rangle = \int_0^1 \mathrm{L}\{\text{Ш}\}(t)\mathrm{e}^{-\mathrm{j}2\pi kt}\mathrm{d}t. \tag{7}$$

The sequence $(\widehat{L}[k])$ is the Fourier series of the periodic generalized function $\mathrm{L}\{\text{Ш}\}$, and is therefore of slow growth [47, Chapter VII]. This implies that L, a priori from $\mathcal{S}(\mathbb{T})$ to $\mathcal{S}'(\mathbb{T})$, actually continuously maps $\mathcal{S}(\mathbb{T})$ into itself. This is a significant difference with the non-periodic setting — we discuss this point in the conclusion in Section VII. Therefore, one can extend it by duality from $\mathcal{S}'(\mathbb{T})$ to $\mathcal{S}'(\mathbb{T})$. Then, for every $f \in \mathcal{S}'(\mathbb{T})$, we easily obtain from (6) that

$$\mathrm{L}f(t) = \sum_{k\in\mathbb{Z}} \widehat{(\mathrm{L}f)}[k]e_k(t), \text{ where } \widehat{(\mathrm{L}f)}[k] = \widehat{f}[k]\widehat{L}[k]. \tag{8}$$

The null space of L is $\mathcal{N}_\mathrm{L} = \{f \in \mathcal{S}'(\mathbb{T}) \mid \mathrm{L}f = 0\}$. We shall only consider operators whose null space is finite-dimensional, in which case $\mathcal{N}_\mathrm{L}$ can only be made of linear combinations of sinusoids at frequencies that are annihilated by L. We state this fact in Proposition 1 and prove it in Appendix A.

Proposition 1. *Let* L *be a continuous LSI operator. If* L *has a finite-dimensional null space $\mathcal{N}_\mathrm{L}$ of dimension N_0, then the null space is of the form*

$$\mathcal{N}_\mathrm{L} = \mathrm{span}\{e_{k_n}\}_{n=1}^{N_0}, \tag{9}$$

where the $k_n \in \mathbb{Z}$ are distinct.

From (6) and (9), we deduce that $\widehat{L}[k] = 0$ if and only if $k = k_n$ for some $n \in [1 \dots N_0]$. In the following, we consider real-valued operators. In that case, we have the Hermitian symmetry $\overline{\widehat{L}[-k]} = \widehat{L}[k]$. Moreover, $e_{k_n} \in \mathcal{N}_\mathrm{L}$ if and only if $e_{-k_n} \in \mathcal{N}_\mathrm{L}$. The orthogonal projection of f on the null space $\mathcal{N}_\mathrm{L}$ is given by

$$\mathrm{Proj}_{\mathcal{N}_\mathrm{L}}\{f\} = \sum_{n=1}^{N_0} \widehat{f}[k_n]e_{k_n}. \tag{10}$$

Let $\mathcal{K}_\mathrm{L} = \mathbb{Z}\backslash\{k_n\}_{n\in\{1\dots N_0\}}$. Then, (4) can be re-expressed as $f = \mathrm{Proj}_{\mathcal{N}_\mathrm{L}}\{f\} + \sum_{k\in\mathcal{K}_\mathrm{L}} \widehat{f}[k]e_k$ and we have that $\mathrm{L}f(t) = \sum_{k\in\mathcal{K}_\mathrm{L}} \widehat{f}[k]\widehat{L}[k]e_k(t)$, which yields the Parseval relation

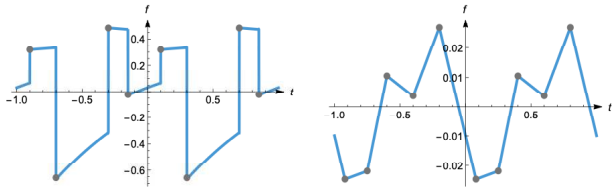
$$\int_0^1 |\mathrm{L}f(t)|^2\mathrm{d}t = \sum_{k\in\mathcal{K}_\mathrm{L}} \left|\widehat{f}[k]\right|^2\left|\widehat{L}[k]\right|^2. \tag{11}$$

B. Periodic L-Splines

Historically, splines are functions defined to be piecewise polynomials [48]. A spline is hence naturally associated to the derivative operator of a given order [49] in the sense that, for a fixed $N \geq 1$, a spline function $f : \mathbb{R} \to \mathbb{R}$ satisfies $\mathrm{L}f(t) = \sum a_m\delta(t - t_m)$ with $\mathrm{L} = \mathrm{D}^N$ the Nth derivative. Splines have been extended to differential [50]–[53], fractional [26], [54] or, more generally, spline-admissible operators [41]. We adapt here this notion to the periodic setting, where the Dirac impulse δ is replaced by the Dirac comb Ш.

Definition 1. *Consider an LSI operator* L *with finite-dimensional null space. We say that a function f is a* periodic L-spline *if*

$$\mathrm{L}f(t) = \sum_{m=1}^{M} a_m\text{Ш}(t - t_m) \tag{12}$$



(a) $\mathrm{L} = \mathrm{D} + \mathrm{I}$ and $M = 4$ (b) $\mathrm{L} = \mathrm{D}^2$ and $M = 5$

Fig. 1. Illustrations of periodic L-splines. Dots: nodes $(t_m, f(t_m))$. The spline in (a) corresponds to the periodization of an exponential B-spline (see Figure 1 in [52]).

for some integer $M \geq 1$, weights $a_m \in \mathbb{R}$, and knot locations $t_m \in \mathbb{T}$.

Periodic L-splines play a crucial role in the variational and statistical approaches for the resolution of inverse problems in the periodic setting. We represent some periodic splines associated to different operators in Figure 1.

III. Periodic Representer Theorem

We now consider a continuous LSI operator L with finite-dimensional null space $\mathcal{N}_\mathrm{L}$. Let $\boldsymbol{\nu}$ be the vector of the linear measurement functions $\nu_1, \ldots, \nu_M$. They usually are of the form $\nu_m = \delta(\cdot - t_m)$ for time-domain sampling problems. Here, we consider general linear measurements to include any kind of inverse problems. In this section, our goal is to recover a function f from observed data $\mathbf{y} = (y_1, \ldots, y_M)$ such that $y_m \simeq \langle \nu_m, f \rangle$. To do so, we consider the variational problem

$$\min_f \left(F(\mathbf{y}, \boldsymbol{\nu}(f)) + \lambda \|\mathrm{L}f\|_{L_2}^2 \right), \tag{13}$$

where $F : \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}^+$ is a strictly convex and continuous function called the *cost function*. This function controls the fidelity to data. A special attention will be given to the quadratic data fidelity of the form

$$F(\mathbf{y}, \boldsymbol{\nu}(f)) = \sum_{m=1}^{M} (y_m - \langle \nu_m, f \rangle)^2. \tag{14}$$

We give the solution of (13) for the space of 1-periodic functions in Theorem 1. To derive this solution, we first introduce and characterize the space of functions on which (13) is well-defined.

A. Search Space

The optimization problem (13) deals with functions such that $\mathrm{L}f$ is square-integrable, which leads us to introduce $\mathcal{H}_\mathrm{L} = \{f \in \mathcal{S}'(\mathbb{T}) \mid \mathrm{L}f \in L_2(\mathbb{T})\}$. Due to (11), we have that

$$\mathcal{H}_\mathrm{L} = \{f \in \mathcal{S}'(\mathbb{T}) \mid \sum_{k \in \mathcal{K}_\mathrm{L}} |\widehat{f}[k]|^2 |\widehat{L}[k]|^2 < +\infty\}. \tag{15}$$

Similar constructions have been developed for functions over $\mathbb{R}$ or for sequences by Unser *et al.* [32], [55]. We now identify a natural Hilbertian structure on $\mathcal{H}_\mathrm{L}$. If $\mathrm{L} : \mathcal{H}_\mathrm{L} \to L_2(\mathbb{T})$ is invertible, then $\mathcal{H}_\mathrm{L}$ inherits the Hilbert-space structure of L_2 via the norm $\|\mathrm{L}f\|_{L_2}$. However, when L has a nontrivial null space, $\|\mathrm{L}f\|_{L_2}$ is only a semi-norm, in which case there exists $f \neq 0$ (any element of the null space of L) such that $\|\mathrm{L}f\|_{L_2} = 0$. To obtain a *bona fide* norm, we complete the semi-norm with a special treatment for the null-space components in Proposition 2.

Proposition 2. *Let* L *be a continuous LSI operator whose finite-dimensional null space is defined by* $\mathcal{N}_\mathrm{L} = \mathrm{span}\{e_{k_n}\}_{n=1}^{N_0}$. *We fix* $\gamma^2 > 0$. *Then,* $\mathcal{H}_\mathrm{L}$ *is a Hilbert space for the inner product*

$$\langle f, g \rangle_{\mathcal{H}_\mathrm{L}} = \langle \mathrm{L}f, \mathrm{L}g \rangle + \gamma^2 \sum_{n=1}^{N_0} \widehat{f}[k_n] \overline{\widehat{g}[k_n]}. \tag{16}$$

The proof is given in Appendix B. We have that $\|f\|_{\mathcal{H}_\mathrm{L}}^2 = \|\mathrm{L}f\|_{L_2}^2 + \gamma^2 \|\mathrm{Proj}_{\mathcal{N}_\mathrm{L}}\{f\}\|_{L_2}^2$, where $\mathrm{Proj}_{\mathcal{N}_\mathrm{L}}\{f\}$ is given by (10). The coefficient γ^2 balances the contribution of both terms.

B. Periodic Reproducing-Kernel Hilbert Space

Reproducing-kernel Hilbert spaces (RKHS) are Hilbert spaces on which the evaluation maps $f \mapsto f(t)$ are well-defined, linear, and continuous. In this section, we answer the question of when the Hilbert space $\mathcal{H}_\mathrm{L}$ associated to an LSI operator L with finite-dimensional null space is a RKHS. This property is relevant to us because periodic function spaces that are RKHS are precisely the ones for which one can use measurement functions of the form $\nu_m = \mathrm{III}(\cdot - t_m)$ in (13).

Definition 2. *Let* $\mathcal{H} \subseteq \mathcal{S}'(\mathbb{T})$ *be a Hilbert space of* 1*-periodic functions and* $\mathcal{H}'$ *be its dual. Then, we say that* $\mathcal{H}$ *is a* RKHS *if the shifted Dirac comb* $\mathrm{III}(\cdot - t_0) \in \mathcal{H}'$ *for any* $t_0 \in \mathbb{T}$.

This implies that any element f of a RKHS has a pointwise interpretation as a function $t \to f(t)$. As is well known, for any RKHS there exists a unique function $h : \mathbb{T} \times \mathbb{T} \to \mathbb{R}$ such that $h(\cdot, t_0) \in \mathcal{H}'$ and $\langle f, h(\cdot, t_0) \rangle = f(t_0)$, for every $t_0 \in \mathbb{T}$ and $f \in \mathcal{H}$. We call h the *reproducing kernel* of $\mathcal{H}$.

Proposition 3. *Let* L *be a continuous LSI operator with finite-dimensional null space. The Hilbert space* $\mathcal{H}_\mathrm{L}$ *(see* (15)*) is a RKHS if and only if*

$$\sum_{k \in \mathcal{K}_\mathrm{L}} \frac{1}{|\widehat{L}[k]|^2} < +\infty. \tag{17}$$

Then, the reproducing kernel for the scalar product (16) *is given by* $h(t, \tau) = h_\gamma(t - \tau)$, *where* $h_\gamma \in \mathcal{S}'(\mathbb{T})$ *is*

$$h_\gamma(t) = \sum_{n=1}^{N_0} \frac{e_{k_n}(t)}{\gamma^2} + \sum_{k \in \mathcal{K}_\mathrm{L}} \frac{e_k(t)}{|\widehat{L}[k]|^2}. \tag{18}$$

The proof is given in Appendix C. Note that the reproducing kernel only depends on the difference $(t - \tau)$.

C. Periodic Representer Theorem

Now that we have defined the search space of the optimization problem (13), we derive the representer theorem that gives the explicit form of its unique periodic solution.

Theorem 1. *We consider the optimization problem*

$$\min_{f\in\mathcal{H}_{\mathrm{L}}}\left(F(\mathbf{y},\boldsymbol{\nu}(f))+\lambda\|\mathrm{L}f\|_{L_2}^2\right), \tag{19}$$

where

- $F:\mathbb{R}^M\times\mathbb{R}^M\to\mathbb{R}^+$ *is strictly convex and continuous;*
- L *is an LSI operator with finite-dimensional null space;*
- $\boldsymbol{\nu}=(\nu_1,\dots,\nu_M)\in(\mathcal{H}'_{\mathrm{L}})^M$ *such that* $\mathcal{N}_{\mathrm{L}}\cap\mathcal{N}_{\boldsymbol{\nu}}=\{0\}$*;*
- $\mathbf{y}=(y_1,\dots,y_M)\in\mathbb{R}^M$ *are the observed data; and*
- $\lambda>0$ *is a tuning parameter.*

Then, (19) *has a unique solution of the form*

$$f_{\mathrm{RT}}(t)=\sum_{m=1}^{M}a_m\varphi_m(t)+\sum_{n=1}^{N_0}b_n e_{k_n}(t), \tag{20}$$

where $a_m,b_n\in\mathbb{R}$, $\varphi_m=h_\gamma*\nu_m$, *and* h_γ *is given by* (18)*. Moreover, the vector* $\mathbf{a}=(a_1,\dots,a_M)$ *satisfies the relation* $\mathbf{P}^{\mathsf{T}}\mathbf{a}=\mathbf{0}$*, with* $\mathbf{P}$ *the* $(M\times N_0)$ *matrix with entries* $[\mathbf{P}]_{m,n}=\langle e_{k_n},\nu_m\rangle$.

The proof of Theorem 1 is given in Appendix D. The optimal solution depends on $(M+N_0)$ coefficients, but the condition $\mathbf{P}^{\mathsf{T}}\mathbf{a}=\mathbf{0}$ implies that there are only $(M+N_0-N_0)=M$ degrees of freedom. In the case when F is quadratic of the form (14), the solution is made explicit in Proposition 4.

Proposition 4. *Under the conditions of Theorem 1, if* F *is given by* (14)*, then the vectors* $\mathbf{a}$ *and* $\mathbf{b}$ *satisfy the linear system*

$$\begin{pmatrix}\mathbf{a}\\ \mathbf{b}\end{pmatrix}=\begin{pmatrix}\mathbf{G}+\lambda\mathbf{I} & \mathbf{P}\\ \mathbf{P}^{\mathsf{T}} & \mathbf{0}\end{pmatrix}^{-1}\begin{pmatrix}\mathbf{y}\\ \mathbf{0}\end{pmatrix}, \tag{21}$$

where $\mathbf{P}\in\mathbb{C}^{M\times N_0}$ *is defined by* $[\mathbf{P}]_{m,n}=\langle e_{k_n},\nu_m\rangle$ *and* $\mathbf{G}\in\mathbb{R}^{M\times M}$ *is a Gram matrix such that*

$$[\mathbf{G}]_{m_1,m_2}=\int_0^1\int_0^1\nu_{m_1}(t)h_\gamma(t-\tau)\nu_{m_2}(\tau)\mathrm{d}t\mathrm{d}\tau. \tag{22}$$

The proof is given in Appendix E. In the case of sampling measurements, we show moreover in Proposition 5 that the optimal solution is a periodic spline in the sense of Definition 1. We recall that such measurements are valid as soon as the search space $\mathcal{H}_{\mathrm{L}}$ is a RKHS, a situation that has been fully characterized in Proposition 3.

Proposition 5. *Under the conditions of Proposition 4, if* L *satisfies* (17) *and if the measurements are of the form* $\nu_m=\text{Ш}(\cdot-t_m)$, $t_m\in\mathbb{T}$*, then the unique solution of* (19) *is a periodic* $(\mathrm{L}^*\mathrm{L})$*-spline with weights* a_m *and knots* t_m.

The proof is given in Appendix F.

IV. Periodic Processes and MMSE

In this section, we change perspective and consider the following statistical problem: given noisy measurements of a zero-mean and real periodic Gaussian process, we are looking for the optimal estimator (for the mean-square error) of the complete process over $\mathbb{T}$.

A. Non-Periodic Setting

In a non-periodic setting, it is usual to consider stochastic models where the random process s is a solution to the stochastic differential equation [41]

$$\mathrm{L}s=w, \tag{23}$$

where L is a linear differential operator and w a continuous domain (non-periodic) Gaussian white noise. When the null space of the operator is nontrivial, it is necessary to add boundary conditions such that the law of the process s is uniquely defined.

B. Gaussian Bridges

In the periodic setting, the construction of periodic Gaussian processes has to be adapted. We first introduce the notion of periodic Gaussian white noise, exploiting the fact that the law of a zero-mean periodic Gaussian process s is fully characterized by its covariance function $r_s(t,\tau)$ such that

$$\mathbb{E}[\langle s,f\rangle\langle s,g\rangle]=\int_0^1\int_0^1 f(t)r_s(t,\tau)\overline{g(\tau)}\mathrm{d}t\mathrm{d}\tau. \tag{24}$$

Definition 3. *A* periodic Gaussian white noise[1] *is a Gaussian random process* w *whose covariance is* $r_w(t,\tau)=\text{Ш}(t-\tau)$.

For any periodic real function f, the random variable $\langle w,f\rangle$ is therefore Gaussian with mean 0 and variance $\|f\|_{L_2}^2$. Moreover, $\langle w,f\rangle$ and $\langle w,g\rangle$ are independent if and only if $\langle f,g\rangle=0$. Hence, the Fourier coefficients $\widehat{w}[k]=\langle w,\mathrm{e}_k\rangle$ of the periodic Gaussian white noise satisfy the following properties:

- $\widehat{w}[k]=\Re(\widehat{w}[k])+\mathrm{j}\,\Im(\widehat{w}[k])$;
- $\overline{\widehat{w}[-k]}=\widehat{w}[k]$;
- $\Re(\widehat{w}[k])$, $\Im(\widehat{w}[k])\sim\mathcal{N}(0,\frac{1}{2})$, $\forall k>0$;
- $\widehat{w}[0]\in\mathbb{R}$ and $\widehat{w}[0]\sim\mathcal{N}(0,1)$;
- $\Re(\widehat{w}[k])$, $\Im(\widehat{w}[k])$, and $\widehat{w}[0]$ are independent.

Put differently, for any nonzero frequency k, $\mathbb{E}[\widehat{w}[k]^2]=0$ and $\mathbb{E}[\widehat{w}[k]\overline{\widehat{w}[k]}]=1$. This means that $\widehat{w}[k]$, $k\neq0$, follows a complex normal distribution with mean 0, covariance 1, and pseudo-covariance 0 [56].

When L has a nontrivial null space, there is no hope to construct a periodic process s solution of (23) with w a periodic Gaussian white noise. Indeed, the operator L kills the null-space frequencies, which contradicts that $\widehat{w}[k_n]\neq0$ almost surely for $n=1\dots N_0$. One should adapt (23) accordingly by giving special treatment to the null-space frequencies. We propose here to consider a new class of periodic Gaussian processes: the *Gaussian bridges*. Given some operator L and $\gamma_0>0$, we set

$$\mathrm{L}_{\gamma_0}=\mathrm{L}+\gamma_0\mathrm{Proj}_{\mathcal{N}_{\mathrm{L}}}, \tag{25}$$

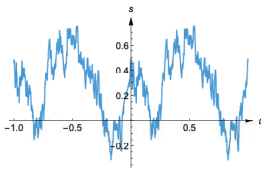
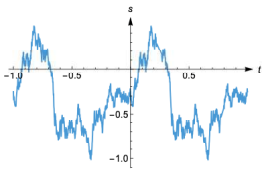
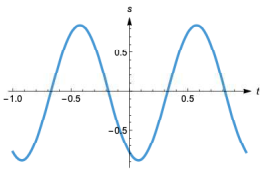
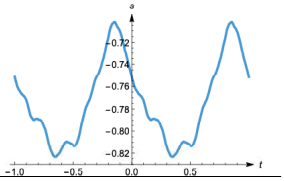
where $\mathrm{Proj}_{\mathcal{N}_{\mathrm{L}}}$ is given by (10). Note that $\mathrm{L}_{\gamma_0}=\mathrm{L}$ for any γ_0 when the null space of L is trivial. Moreover, we remark that

$$\|\mathrm{L}_{\gamma_0}f\|_{L_2}^2=\|\mathrm{L}f\|_{L_2}^2+\gamma_0^2\|\mathrm{Proj}_{\mathcal{N}_{\mathrm{L}}}\{f\}\|_{L_2}^2=\|f\|_{\mathcal{H}_{\mathrm{L}}}^2, \tag{26}$$

where $\|f\|_{\mathcal{H}_{\mathrm{L}}}^2=\langle f,f\rangle_{\mathcal{H}_{\mathrm{L}}}$ is given in (16) (with $\gamma=\gamma_0$).

[1]Without loss of generality, we only consider Gaussian white noise with zero-mean and variance 1.

TABLE I
GAUSSIAN BRIDGES FOR SEVERAL OPERATORS.

	$\mathrm{D}+\mathrm{I}$	D	$\mathrm{D}^2+4\pi^2\mathrm{I}$	D^2
$\widehat{L}[k]$	$\mathrm{j}2\pi k+1$	$\mathrm{j}2\pi k$	$4\pi^2(1-k^2)$	$-4\pi^2k^2$
$\mathcal{N}_{\mathrm{L}}$	$\mathrm{span}\{0\}$	$\mathrm{span}\{e_0\}$	$\mathrm{span}\{e_1, e_{-1}\}$	$\mathrm{span}\{e_0\}$
Gaussian bridges $\gamma_0^2=1$				

Definition 4. *A* Gaussian bridge *is a periodic Gaussian process s, solution to the stochastic differential equation*

$$\mathrm{L}_{\gamma_0} s = w, \tag{27}$$

with w a periodic Gaussian white noise and L_{γ_0} given by (25) *for some LSI operator L with finite-dimensional null space and $\gamma_0 > 0$. We summarize this situation with the notation $s \sim \mathcal{GB}(\mathrm{L}, \gamma_0^2)$. When the null space is trivial, in which case the parameter γ_0^2 is immaterial, we write $s \sim \mathcal{GB}(\mathrm{L})$.*

The Gaussian-bridge terminology is inspired by the Brownian bridge, the periodic version of the Brownian motion[2]. Several realizations of our Gaussian bridges for various operators are shown in Table I for $\gamma_0^2 = 1$. The influence of the parameter γ_0^2 is illustrated in Figure 2.

Proposition 6. *The covariance function of the Gaussian bridge $s \sim \mathcal{GB}(\mathrm{L}, \gamma_0^2)$ is*

$$r_s(t,\tau) = h_{\gamma_0}(t-\tau), \tag{28}$$

where h_{γ_0} is defined in (18)*. It implies that*

$$\mathbb{E}[\langle s, f\rangle\langle s, g\rangle] = \langle h_{\gamma_0} * f, g\rangle. \tag{29}$$

In particular, we have that

$$\mathbb{E}[|\widehat{s}[k]|^2] = \widehat{h}_{\gamma_0}[k]. \tag{30}$$

The proof of Proposition 6 is given in Appendix G. An important consequence is that a Gaussian bridge is stationary since its covariance function only depends on the difference $(t-\tau)$.

C. Measurement Model and MMSE Estimator

For this section, we restrict ourselves to operators L for which the native space $\mathcal{H}_{\mathrm{L}}$ is a RKHS. In that case, using (30) and (18), the Gaussian bridge s satisfies

$$\mathbb{E}[\|s\|_{L_2}^2] = \sum_{k\in\mathbb{Z}} \mathbb{E}[|\widehat{s}[k]|^2] = \sum_{k\in\mathcal{K}_{\mathrm{L}}} \frac{1}{|\widehat{L}[k]|^2} + \sum_{n=1}^{N_0} \frac{1}{\gamma_0^2}, \tag{31}$$

which is finite according to (17). Therefore, the Gaussian bridge s is (almost surely) square-integrable.

[2]Our definition differs from the classical one, in which the Brownian bridge is zero at the origin instead of being zero-mean [57].

The observed data $\mathbf{y}$ are assumed to be generated as

$$\mathbf{y} = \langle \boldsymbol{\nu}, s\rangle + \boldsymbol{\epsilon}, \tag{32}$$

where $s \sim \mathcal{GB}(\mathrm{L}, \gamma_0^2)$ is a Gaussian bridge (see Definition 4), $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_M)$ is a vector of M linear measurement functions, and $\boldsymbol{\epsilon}$ are independent random perturbations such that $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma_0^2\mathbf{I})$. Given $\mathbf{y}$ in (32), we want to find the estimator $\tilde{s}$ of the Gaussian bridge s, imposing that it minimizes the quantity $\mathbb{E}[\|s-\tilde{s}\|_2^2]$.

Theorem 2. *Let $\mathbf{y} = (y_1, \ldots, y_M)$ be the noisy measurement vector* (32) *of the Gaussian bridge $s \sim \mathcal{GB}(\mathrm{L}, \gamma_0^2)$, with measurement functions $\nu_m \in \mathcal{H}'_{\mathrm{L}}$, $m = 1 \ldots M$. Then, the MMSE estimator of s given the samples $\{y_m\}_{m\in[1\ldots M]}$ is*

$$\tilde{s}_{\mathrm{MMSE}}(t) = \sum_{m=1}^{M} d_m \varphi_m(t), \tag{33}$$

where $\varphi_m = h_{\gamma_0} * \nu_m$ with $\nu_m \in \mathcal{H}'_{\mathrm{L}}$, $\mathbf{d} = (d_1, \ldots, d_M) = (\mathbf{G} + \sigma_0^2\mathbf{I})^{-1}\mathbf{y}$, and $\mathbf{G}$ is the Gram matrix defined in (21)*.*

The proof is given in Appendix H. Theorem 2 can be seen as a generalization of the classical Wiener filtering, designed for discrete signals, to the hybrid case where the input signal is in a (periodic) continuous-domain and the (finite-dimensional) measurements are discrete. A leading theme of this paper is that the form of the MMSE estimator $\tilde{s}_{\mathrm{MMSE}}$ is very close to the one of the solution of the representer theorem f_{RT} with $\lambda = \sigma_0^2$ and for a quadratic cost function. This connection is exploited in Section IV-D.

D. MMSE Estimation as a Representer Theorem

The MMSE estimator given in Theorem 2 can be interpreted as the solution of the optimization problem described in Proposition 7.

Proposition 7. *Consider an LSI operator L with finite-dimensional null space, $\gamma > 0$, and $\nu_m \in \mathcal{H}'_{\mathrm{L}}$ for $m = 1 \ldots M$. We set L_γ as in* (25)*. Then, the solution of the optimization problem*

$$\min_{f\in\mathcal{H}_{\mathrm{L}}} \left(\sum_{m=1}^{M} (y_m - \langle f, \nu_m\rangle)^2 + \lambda \|\mathrm{L}_\gamma f\|_{L_2}^2 \right) \tag{34}$$

exists, is unique, and given by

$$f_{\text{opt}}(t) = \sum_{m=1}^{M} d_m \varphi_m(t), \tag{35}$$

where $\varphi_m = h_\gamma * \nu_m$ *and* $\mathbf{d} = (d_1, \dots, d_M) = (\mathbf{G} + \lambda \mathbf{I})^{-1}\mathbf{y}$. *In particular, the unique minimizer of* (34) *is the MMSE estimator given in Theorem 2 for* $\lambda = \sigma_0^2$ *and* $\gamma = \gamma_0$.

The proof of Proposition 7 follows the same steps as the ones of Theorem 1 (form of the minimizer for the periodic representer theorem) and Proposition 4 (explicit formulas in terms of system matrix for the vectors $\mathbf{a}$ and $\mathbf{b}$), with significant simplifications that are detailed in Appendix I. Proposition 7 has obvious similarities with Theorem 1, but it also adds new elements.

- Proposition 7 gives an interpretation of the MMSE estimator of a Gaussian bridge given its measurements as the solution to an optimization problem. This problem is very close to the periodic representer theorem (Theorem 1) for a quadratic cost function. However, (34) differs from (19) because the regularization also penalizes null-space frequencies.
- If the null space $\mathcal{N}_{\text{L}}$ is trivial, then

$$f_{\text{RT}} = \tilde{s}_{\text{MMSE}} \tag{36}$$

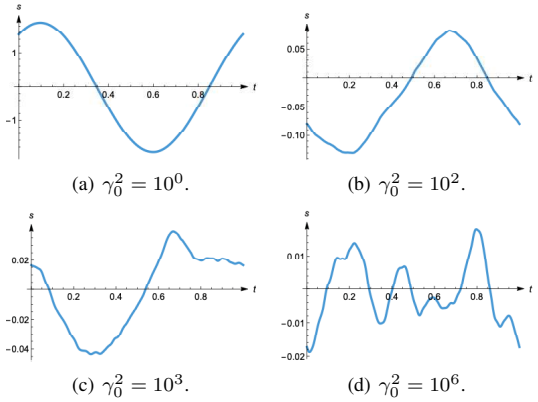
  for $\lambda = \sigma_0^2$. This means that Theorem 1 (smoothing approach) and 2 (statistical approach) correspond to the same reconstruction method. This equivalence is well-known for stationary processes on $\mathbb{R}$ in the case of time-domain sampling measurements [42]. Our results extend this to the periodic setting and to the case of generalized linear measurements.
- If the null space is nontrivial, then Theorem 1 and Proposition 7 yield different reconstructions. In particular, this implies that one cannot interpret the optimizer f_{RT} in Theorem 1 as the MMSE estimator of a Gaussian bridge. Yet, the solutions get closer and closer as $\gamma_0 \to 0$. In Section V, we investigate more deeply this situation.

V. Quality of the Estimators on Simulations

We consider $\tilde{s}_{\gamma,\lambda}(t|\mathbf{y}) = \sum_{m=1}^{M} d_m \varphi_m(t)$ as the linear estimator of s given $\mathbf{y}$, where $\varphi_m = h_\gamma * \nu_m$, $\mathbf{d} = (\mathbf{G}+\lambda\mathbf{I})^{-1}\mathbf{y}$, and $\mathbf{G}$ is defined in Proposition 4. To simplify notations, we shall omit $\mathbf{y}$ when considering $\tilde{s}_{\gamma,\lambda}(\cdot|\mathbf{y}) = \tilde{s}_{\gamma,\lambda}$. Each pair (λ, γ) gives an estimator. In particular, if s is a Gaussian bridge, then $\tilde{s}_{\text{MMSE}} = \tilde{s}_{\gamma_0,\sigma_0^2}$, according to Theorem 2. The mean-square error (MSE) of $\tilde{s}_{\gamma,\lambda}$ over N experiments is computed as $\text{MSE} = \frac{1}{N}\sum_{n=1}^{N} \|s_n - \left(\tilde{s}_{\gamma,\lambda}\right)_n\|_{L_2}^2$, where the s_n are independent realizations of s that yield a new noisy measurement $\mathbf{y}_n$ and $\left(\tilde{s}_{\gamma,\lambda}\right)_n = \tilde{s}_{\gamma,\lambda}(\cdot|\mathbf{y}_n)$ is the estimator based on $\mathbf{y}_n$. We define the normalized mean-square error (NMSE) by

$$\text{NMSE} = \frac{\text{MSE}}{\frac{1}{N}\sum_{n=1}^{N}\|s_n\|_{L_2}^2} \approx \frac{\mathbb{E}[\|s - \tilde{s}_{\gamma,\lambda}\|_{L_2}^2]}{\mathbb{E}[\|s\|_{L_2}^2]}. \tag{37}$$

In this section, we first detail the generation of Gaussian bridges (Section V-A). We then investigate the role of the parameters λ (Section V-B) and γ^2 (Section V-C) on the quality of the estimator $\tilde{s}_{\gamma,\lambda}$. We primarily focus on time-domain sampling measurements with $\langle \boldsymbol{\nu}, s\rangle = (s(t_1), \dots, s(t_M))^{\mathsf{T}}$, where the t_m are in $\mathbb{T}$.



(a) $\gamma_0^2 = 10^0$. (b) $\gamma_0^2 = 10^2$. (c) $\gamma_0^2 = 10^3$. (d) $\gamma_0^2 = 10^6$.

Fig. 2. Illustration of $s \sim \mathcal{GB}(\mathrm{D}^2 + 4\pi^2\mathrm{I}, \gamma_0^2)$ for different values of γ_0^2.

A. Generation of Gaussian Bridges

We first fix the operator L with null space $\mathcal{N}_{\text{L}}$ of dimension N_0 and $\gamma_0 > 0$. Then, we generate $(2N_{\text{coef}} + 1)$ Fourier coefficients $\{\widehat{w}[k]\}_{k\in[-N_{\text{coef}}\dots N_{\text{coef}}]}$ of a Gaussian white noise according to Definition 3. Finally, we compute the Gaussian bridge s as

$$s(t) = \sum_{\substack{k\in\mathcal{K}_{\text{L}} \\ |k|\le N_{\text{coef}}}} \frac{\widehat{w}[k]}{\widehat{L}[k]} e_k(t) + \sum_{n=1}^{N_0} \frac{\widehat{w}[k_n]}{\gamma_0} e_{k_n}(t). \tag{38}$$

Since $N_0 < \infty$, (38) provides a mere approximation of the Gaussian bridge. However, the approximation error can be made arbitrarily small by taking N_{coef} large enough. In Figure 2, we generate $s \sim \mathcal{GB}(\mathrm{D}^2 + 4\pi^2\mathrm{I}, \gamma_0^2)$ for four values of γ_0^2. For small values of γ_0^2, the null-space component dominates, which corresponds in this case to the frequency $|k| = 1$. When γ_0^2 increases, the null-space component has a weaker influence.

B. Influence of λ

We evaluate the influence of the parameter λ for the case of the invertible operator $\mathrm{L} = \mathrm{D} + \mathrm{I}$. In this case we have that $\text{Proj}_{\mathcal{N}_{\text{L}}} = 0$ (since $\mathcal{N}_{\text{L}} = \{0\}$), which simplifies (25). Hence, the parameter γ_0^2 is immaterial and we denote by $\tilde{s}_\lambda$ the estimator associated to $\lambda > 0$. We consider $s \sim \mathcal{GB}(\mathrm{D}+\mathrm{I})$ and $\sigma_0^2 = 10^{-2}$.

Time-Domain Sampling Measurements. We generated $N = 500$ realizations of s. From each one, we extracted $M = 30$ noisy measurements. We then computed 30 estimators $\{\left(\tilde{s}_\lambda\right)_n\}_{\lambda\in\mathcal{L}_1}$, where $\mathcal{L}_1$ is the set of values obtained by uniform sampling of the interval $[0.001, 0.03]$. The plot of the NMSE (approximated according to (37)) as a function of λ is given in Figure 3 (a). The minimum error is obtained for $\lambda \simeq 0.01$, which corresponds to σ_0^2. This result validates the

theory presented in Theorem 2. Actually, when λ is small, the estimator interpolates the noisy measurements while, for a large λ, the estimator tends to oversmooth the curve. The MMSE estimator makes an optimal tradeoff between fitting the data and smoothing the curve. These observations about λ retain their validity for other operators, including noninvertible ones.

Fourier-Domain Sampling Measurements. We consider complex exponential measurement functionals, inducing $\langle \boldsymbol{\nu}, s \rangle = (\widehat{s}[k_1], \dots, \widehat{s}[k_M])^{\mathsf{T}}$, where the k_m are in $\mathbb{Z}$. We define $\mathcal{N}_{\boldsymbol{\nu}} = \{k_m\}_{m=1\dots M}$, such that $(-k_m) \in \mathcal{N}_{\boldsymbol{\nu}}$ for every $k_m \in \mathcal{N}_{\boldsymbol{\nu}}$. We consider the measurements $\boldsymbol{\nu} = (e_{k_1}, \dots, e_{k_M})$. Note that these measurement functionals are complex, which calls for a slight adaptation of the framework presented so far[3]. The noise $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_M)$ is then also complex and satisfies the properties:

- $\epsilon_m = \Re(\epsilon_m) + \mathrm{j}\, \Im(\epsilon_m)$;
- $\epsilon_{m_1} = \overline{\epsilon_{m_2}}$, $k_{m_1} = -k_{m_2}$;
- $\Re(\epsilon_m),\ \Im(\epsilon_m) \sim \mathcal{N}(0, \frac{\sigma_0^2}{2})$, $\forall k_m \neq 0$;
- $\epsilon_m \in \mathbb{R}$ and $\epsilon_m \sim \mathcal{N}(0, \sigma_0^2)$, $k_m = 0$;
- $\Re(\epsilon_m),\ \Im(\epsilon_m)$ and ϵ_{m_1}, $k_{m_1} = 0$, are independent.

This means that $\mathbb{E}[|\epsilon_m|^2] = \sigma_0^2$ for every m.

We repeated the experiment done with the time-domain sampling using exactly the same procedure and parameters, and $\mathcal{N}_{\boldsymbol{\nu}} = \{-2, -1, 0, 1, 2\}$. The experimental curve of the evolution of the NMSE with λ is given in Figure 3 (b). Again, the minimum is obtained for $\lambda \simeq 0.01 = \sigma_0^2$. We now want to compare this curve to the theoretical one.

For the Fourier-sampling case, we were also able to derive the corresponding closed-form formulas for the NMSE (37).

Proposition 8. *Let s be a Gaussian bridge associated with an invertible operator L, and $y_m = \widehat{s}[k_m] + \epsilon_m$, $m = 1 \dots M$, with $k_m \in \mathcal{N}_{\boldsymbol{\nu}}$ the sampled frequencies and ϵ a complex Gaussian noise with variance σ_0^2 as above. Then, the MSE of the estimator $\tilde{s}_\lambda = \tilde{s}_\lambda(\cdot|\mathbf{y})$ is given by*

$$\mathbb{E}\left[\|s - \tilde{s}_\lambda\|_{L_2}^2\right] = \sum_{m=1}^{M} \frac{\widehat{h}[k_m](\lambda^2 + \widehat{h}[k_m]\sigma_0^2)}{(\widehat{h}[k_m] + \lambda)^2} + \sum_{k \notin \mathcal{N}_{\boldsymbol{\nu}}} \widehat{h}[k], \tag{39}$$

where h is the reproducing kernel of $\mathcal{H}_{\mathrm{L}}$.

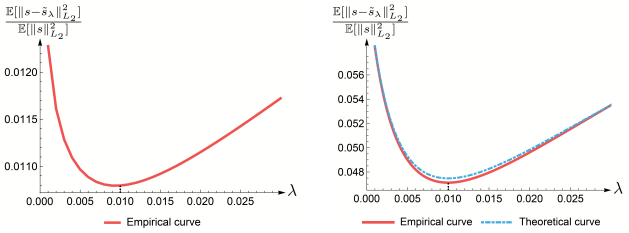
The proof is given in Appendix J. Note that $\widehat{h}[k] = 1/|\widehat{L}[k]|^2$ is real-valued and strictly positive for every k. From (39), we also recover the property that the optimum is reached for $\lambda = \sigma_0^2$ since each of the M terms that appear in the first sum is minimized for this value of λ.

The theoretical curve for $\mathcal{N}_{\boldsymbol{\nu}} = \{-2, -1, 0, 1, 2\}$ is given in Figure 3 (b) and is in good agreement with the experimental curve. We explain the slight variation (0.15% for the L_2-norm over $\lambda \in [0.001, 0.03]$) by the fact that (37) is only an estimation of the theoretical NMSE.

C. Influence of γ^2

In this section, we only consider noninvertible operators since invertibility has already been addressed in Section IV-D

[3]One could equivalently consider cosine and sine measurements, to the cost of heavier formulas.



(a) Time-domain sampling. (b) Fourier-domain sampling.

Fig. 3. Evolution of the NMSE in terms of λ for $s \sim \mathcal{GB}(\mathrm{D} + \mathrm{I})$ for time and Fourier-domain sampling measurements.

(see (36)). In order to evaluate the specific influence of γ, we set $\lambda = \sigma_0^2$. Hence, $\tilde{s}_{\gamma,\sigma_0^2} = \tilde{s}_\gamma$. We generated $N = 500$ realizations of a Gaussian bridge s, and from each one, we extracted $M = 30$ noisy measurements. We repeated this for several operators L and values of γ_0^2 and σ_0^2. For each case, we compared $\tilde{s}_{\mathrm{MMSE}}$ to $\tilde{s}_{\gamma\to 0}$, $\tilde{s}_{\gamma\to\infty}$, and f_{RT} in (20), seen here as an additional estimator. The corresponding NMSEs (see (37)) are given in Table II. We make four observations.

1) In each case, the best result is obtained with $\tilde{s}_{\mathrm{MMSE}}$, as expected. We see, moreover, that $\lim_{\gamma\to 0} \mathbb{E}[\|s - \tilde{s}_\gamma\|_{L_2}^2] \simeq \mathbb{E}[\|s - f_{\mathrm{RT}}\|_{L_2}^2]$. This is in line with the fact that the functional (19) to minimize in Theorem 1 corresponds to (34) with $\gamma = 0$.

2) For small values of γ_0^2 (*i.e.*, 10^{-3} or 10^0), we see that $\mathbb{E}[\|s - f_{\mathrm{RT}}\|_{L_2}^2] \simeq \mathbb{E}[\|s - \tilde{s}_{\mathrm{MMSE}}\|_{L_2}^2]$. This means that the performances of $\tilde{s}_{\mathrm{MMSE}}$ and f_{RT} are very similar. This is illustrated in Figure 4 (a), where $\tilde{s}_{\mathrm{MMSE}}$ and f_{RT} do coincide. Meanwhile, we see that $\lim_{\gamma\to\infty} \mathbb{E}[\|s - \tilde{s}_\gamma\|_{L_2}^2] \gg \mathbb{E}[\|s - \tilde{s}_{\mathrm{MMSE}}\|_{L_2}^2]$. This is also illustrated in Figure 4 (a) for $\mathrm{L} = \mathrm{D}$. The reconstruction for $\gamma \to +\infty$ significantly fails to recover the original signal s, as the corresponding estimator tends to have zero-mean.

3) For intermediate values of γ_0^2 (*i.e.*, $\gamma_0^2 = 10^3$ or 10^6 according to σ_0 and the order of the operator), the minimal NMSE is obtained for $\tilde{s}_{\mathrm{MMSE}}$ only. We also observe that $\mathbb{E}[\|s - f_{\mathrm{RT}}\|_{L_2}^2] < \lim_{\gamma\to\infty} \mathbb{E}[\|s - \tilde{s}_\gamma\|_{L_2}^2]$. This is illustrated in Figure 4 (b) for $\mathrm{L} = \mathrm{D}^2 + 4\pi^2\mathrm{I}$, $\gamma_0^2 = 10^6$ and $\sigma_0^2 = 10^{-4}$, where we can distinguish $\tilde{s}_{\mathrm{MMSE}}$, $\tilde{s}_{\gamma\to\infty}$, and f_{RT}.

4) For large values of γ_0^2 (*i.e.*, $\gamma_0^2 = 10^9$), we observe that $\lim_{\gamma\to\infty} \mathbb{E}[\|s - \tilde{s}_\gamma\|_{L_2}^2] \simeq \mathbb{E}[\|s - \tilde{s}_{\mathrm{MMSE}}\|_{L_2}^2]$ and $\mathbb{E}[\|s - f_{\mathrm{RT}}\|_{L_2}^2] > \mathbb{E}[\|s - \tilde{s}_{\mathrm{MMSE}}\|_{L_2}^2]$. In fact, for large γ_0^2, the Gaussian bridge tends to have vanishing null-space frequencies (with (38), we have that $\widehat{s}[k_n] = \widehat{w}[k_n]/\gamma_0$ for $n = 1 \dots N_0$). Meanwhile, the reconstructed signal f_{RT} is not constrained to attenuate null-space frequencies. The null-space part in (20) is mainly responsible for a higher error compared to $\tilde{s}_{\mathrm{MMSE}}$. This is highlighted in Figure 4 (c).

Observations 2), 3), and 4) suggest the existence of three regimes. For further investigation, we present in Figure 5 the evolution of NMSE as a function of $\log \gamma^2$ for $\mathrm{L} = \mathrm{D}$ and $\gamma_0^2 = 10^0, 10^3$, and 10^6. The minimal error is always obtained for $\gamma^2 \simeq \gamma_0^2$, as predicted by the theory. For the three cases, we observe two plateaus: one for $\gamma^2 \in (0, v_1)$ and the other for $\gamma^2 \in (v_2, \infty)$, where v_1, $v_2 > 0$. It means

that, for each value of γ_0^2, the estimators $\tilde{s}_\gamma$ with $\gamma^2 \in (0, v_1)$ ((v_2, ∞), respectively) are very similar and the reconstruction algorithms are practically indistinguishable. The values of v_1 and v_2 depend on γ_0^2. When $\gamma_0^2 = 10^0$ (10^6, respectively), we have that $\gamma_0^2 \in (0, v_1)$ ((v_2, ∞), respectively). However, $\gamma_0^2 = 10^3 \in [v_1, v_2]$ belongs to none of the plateaus.

Two main conclusions can be drawn from our experiments. First, we have strong empirical evidence that

$$\tilde{s}_\gamma \underset{\gamma \to 0}{\longrightarrow} f_{\mathrm{RT}}, \tag{40}$$

which we conjecture to be true for any Gaussian-bridge model. This is remarkable because it presents the reconstruction based on the periodic representer theorem as a limit case of the statistical approach. Second, we empirically see that, for reasonably small values of γ_0^2, the estimators corresponding to $\gamma^2 \leq \gamma_0^2$ are practically indistinguishable from the MMSE estimator. This is in particular valid for the representer-theorem reconstruction, for which we then have that

$$f_{\mathrm{RT}} \approx \tilde{s}_{\mathrm{MMSE}}. \tag{41}$$

The variational method is *theoretically* suboptimal to reconstruct Gaussian bridges. However, based on our experiments, it is reasonable to consider this method as *practically* optimal for small values of γ_0^2 and $\lambda = \sigma_0^2$.

VI. Discussion

A. Comparison with Inverse Problems on the Real Line

It is worth noting that the periodic setting has important differences as compared to reconstruction methods over the complete real line, which motivated and played an important role in this paper.

- The role of the Dirac impulse δ is played by the Dirac comb $\mathrm{Ш}$ in the periodic setting. It is indeed the neutral element of the periodic convolution (5) and appears in the definition of the periodic L-splines (Definition 1) and RKHS (Definition 2).
- In the real-line setting, in addition to smoothness properties, functions are also characterized by their property of decay at infinity [58]. For periodic functions, we only consider the smoothness properties, which brings substantial simplifications.
- In general, a continuous LSI operator does not preserve the asymptotic behavior of the input function. For instance, a test function in the space $\mathcal{S}(\mathbb{R})$ of smooth and rapidly decaying functions is not necessarily mapped to a rapidly decaying function. In contrast, any continuous LSI operator maps the space of periodic test functions $\mathcal{S}(\mathbb{T})$ onto itself (see Section II-A). This greatly simplifies the study of operators that act on periodic functions.
- The null space of a continuous LSI operator can differ for the two cases. In particular, when acting on periodic functions, the null space of the nth derivative D^n is reduced to constant functions for every $n \geq 1$. This is crucial due to the role of the null space in Theorems 1 and 2.
- In Proposition 3, we give a necessary and sufficient condition for a continuous LSI operator of finite-dimensional null space to specify a RKHS in the sense of Definition 2. This is significantly more complicated over the real line, for which only partial results are known [32].
- We have seen that it is not always possible to find a periodic solution s to the equation $\mathrm{L}s = w$, where w is a periodic Gaussian white noise. This lead us to modify the stochastic differential equation (see (27)) and to introduce the family of Gaussian bridges.
- In Theorem 2, we give the MMSE estimator of the *complete* process s, not only for the estimation of $s(t_0)$ at a fixed time t_0. In the non-periodic setting, however, solutions of stochastic differential equations are generally not square-integrable. For instance, if s is a nontrivial stationary Gaussian process, then

$$\begin{aligned}\mathbb{E}[\|s\|^2_{L_2(\mathbb{R})}] &= \sum_{k \in \mathbb{Z}} \mathbb{E}[\|1_{[k,k+1)} \cdot s\|^2_{L_2(\mathbb{R})}] \\ &\overset{(i)}{=} \sum_{k \in \mathbb{Z}} \mathbb{E}[\|1_{[0,1)} \cdot s\|^2_{L_2(\mathbb{R})}] = \infty, \end{aligned} \tag{42}$$

  where $1_{[a,b)}$ is the indicator function on $[a, b)$ and (i) exploits stationarity. Another example is the Brownian motion, whose supremum over $[0, t]$ grows faster than t^p for any $p < 1/2$ (almost surely) when t goes to infinity [59], hence being of infinite energy. As a consequence, it is irrelevant to consider the MMSE estimator of the complete process and one ought to, for instance, restrict to MMSE estimators of local values $s(t_0)$ of the process.

B. Comparison with TV Regularization

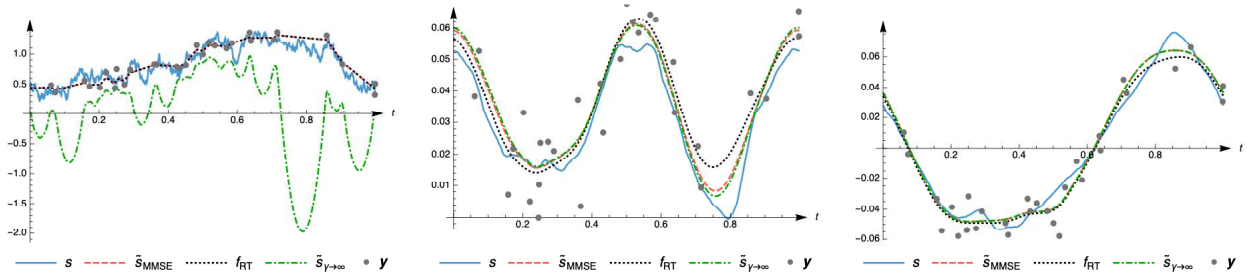
A recent tendency in the field of signal reconstruction is to rely on sparsity-promoting regularization, motivated by the fact that many real-world signals are sparse in some adequate transform domain [41], [60], [61].

The vast majority of works focuses on the finite-dimensional setting via ℓ_1-type regularization. However, some authors have recently promoted the reconstruction of infinite-dimensional sparse signals [6], [62]. The adaptation of discrete ℓ_1 methods to the continuous domain is based on the total-variation (TV) regularization norm, for which it is possible to derive representer theorems (see [32, Theorem 1]). A comparison between Tikhonov and TV variational techniques is proposed in Gupta *et al.* [33] for non-periodic signals. In brief, at identical measurements and regularization operator L, Tikhonov regularization favors smooth solutions restricted to a finite-dimensional space, while TV regularization allows for adaptive and more compressible solutions. In [33, Table I], it was shown on simulations that Tikhonov methods perform better on fractal-type signals, while TV methods are better suited to sparse signals. We expect similar behaviors for the periodic setting.

At the heart of the present paper is the connection between L_2-regularization and the statistical formalism of MMSE estimation of Gaussian processes. A theoretical link between deterministic and stochastic frameworks is much harder to provide for sparsity-inducing priors. There is strong empirical evidence that sparse stochastic models are intimately linked

TABLE II
COMPARISON OF NMSE FOR $\tilde{s}_{\gamma\to 0}$, f_{RT}, $\tilde{s}_{\mathrm{MMSE}}$, AND $\tilde{s}_{\gamma\to\infty}$ OVER $N = 500$ ITERATIONS. **BOLD**: OPTIMAL RESULT.

		$\sigma_0 = 10^{-1}$				$\sigma_0 = 10^{-2}$			
L	γ_0^2	$\tilde{s}_{\gamma\to 0}$	f_{RT}	$\tilde{s}_{\mathrm{MMSE}}$	$\tilde{s}_{\gamma\to\infty}$	$\tilde{s}_{\gamma\to 0}$	f_{RT}	$\tilde{s}_{\mathrm{MMSE}}$	$\tilde{s}_{\gamma\to\infty}$
D	10^{-3}	$\mathbf{1.37\times 10^{-5}}$	$\mathbf{1.37\times 10^{-5}}$	$\mathbf{1.37\times 10^{-5}}$	1.78	$\mathbf{8.40\times 10^{-6}}$	$\mathbf{8.40\times 10^{-6}}$	$\mathbf{8.40\times 10^{-6}}$	2.94
	10^{0}	$\mathbf{1.17\times 10^{-2}}$	$\mathbf{1.17\times 10^{-2}}$	$\mathbf{1.17\times 10^{-2}}$	1.66	$\mathbf{8.44\times 10^{-3}}$	$\mathbf{8.44\times 10^{-3}}$	$\mathbf{8.44\times 10^{-3}}$	2.72
	10^{3}	1.59×10^{-1}	1.56×10^{-1}	$\mathbf{1.49\times 10^{-1}}$	1.58×10^{-1}	1.05×10^{-1}	1.05×10^{-1}	$\mathbf{9.96\times 10^{-2}}$	1.21×10^{-1}
	10^{6}	1.61×10^{-1}	1.60×10^{-1}	$\mathbf{1.43\times 10^{-1}}$	$\mathbf{1.43\times 10^{-1}}$	1.07×10^{-1}	1.07×10^{-1}	$\mathbf{9.11\times 10^{-2}}$	$\mathbf{9.11\times 10^{-2}}$
	10^{9}	1.66×10^{-1}	1.66×10^{-1}	$\mathbf{1.47\times 10^{-1}}$	$\mathbf{1.47\times 10^{-1}}$	1.10×10^{-1}	1.10×10^{-1}	$\mathbf{9.34\times 10^{-2}}$	$\mathbf{9.34\times 10^{-2}}$
D^2	10^{-3}	$\mathbf{8.43\times 10^{-7}}$	$\mathbf{8.43\times 10^{-7}}$	$\mathbf{8.43\times 10^{-7}}$	1.07	3.12×10^{-8}	$\mathbf{3.11\times 10^{-8}}$	$\mathbf{3.11\times 10^{-8}}$	1.34
	10^{0}	9.06×10^{-4}	9.06×10^{-4}	$\mathbf{9.05\times 10^{-4}}$	1.07	$\mathbf{3.34\times 10^{-5}}$	$\mathbf{3.34\times 10^{-5}}$	$\mathbf{3.34\times 10^{-5}}$	1.33
	10^{3}	4.04×10^{-1}	4.04×10^{-1}	$\mathbf{3.61\times 10^{-1}}$	7.1×10^{-1}	$\mathbf{1.46\times 10^{-2}}$	$\mathbf{1.46\times 10^{-2}}$	$\mathbf{1.46\times 10^{-2}}$	5.78×10^{-1}
	10^{6}	6.53×10^{-1}	6.53×10^{-1}	$\mathbf{3.66\times 10^{-1}}$	$\mathbf{3.66\times 10^{-1}}$	2.63×10^{-2}	2.63×10^{-2}	$\mathbf{2.26\times 10^{-2}}$	2.29×10^{-2}
	10^{9}	6.62×10^{-1}	6.62×10^{-1}	$\mathbf{3.86\times 10^{-1}}$	$\mathbf{3.86\times 10^{-1}}$	2.65×10^{-2}	2.65×10^{-2}	$\mathbf{2.16\times 10^{-2}}$	$\mathbf{2.16\times 10^{-2}}$
$\mathrm{D}^2+4\mathrm{I}$	10^{-3}	$\mathbf{5.53\times 10^{-7}}$	$\mathbf{5.53\times 10^{-7}}$	$\mathbf{5.53\times 10^{-7}}$	1.03	$\mathbf{1.71\times 10^{-8}}$	$\mathbf{1.71\times 10^{-8}}$	$\mathbf{1.71\times 10^{-8}}$	1.22
	10^{0}	5.56×10^{-4}	5.56×10^{-4}	$\mathbf{5.55\times 10^{-4}}$	1.04	$\mathbf{1.77\times 10^{-5}}$	$\mathbf{1.77\times 10^{-5}}$	$\mathbf{1.77\times 10^{-5}}$	1.24
	10^{3}	3.67×10^{-1}	3.67×10^{-1}	$\mathbf{3.04\times 10^{-1}}$	8.79×10^{-1}	1.21×10^{-2}	1.21×10^{-2}	$\mathbf{1.20\times 10^{-2}}$	8.71×10^{-1}
	10^{6}	1.52	1.52	$\mathbf{4.63\times 10^{-1}}$	$\mathbf{4.63\times 10^{-1}}$	3.94×10^{-2}	3.94×10^{-2}	$\mathbf{2.98\times 10^{-2}}$	3.04×10^{-2}
	10^{9}	1.47	1.47	$\mathbf{4.87\times 10^{-1}}$	$\mathbf{4.87\times 10^{-1}}$	4.67×10^{-2}	4.67×10^{-2}	$\mathbf{3.18\times 10^{-2}}$	$\mathbf{3.18\times 10^{-2}}$



(a) L = D, $\gamma_0^2 = 1$, and $\sigma_0^2 = 10^{-2}$. (b) L = $\mathrm{D}^2 + 4\pi^2\mathrm{I}$, $\gamma_0^2 = 10^6$, and $\sigma_0^2 = 10^{-4}$. (c) L = D^2, $\gamma_0^2 = 10^9$, and $\sigma_0^2 = 10^{-4}$.

Fig. 4. Illustrations of $s \sim \mathcal{GB}(\mathrm{L}, \gamma_0^2)$, $\tilde{s}_{\mathrm{MMSE}}$, f_{RT}, and $\tilde{s}_{\gamma\to\infty}$ for several operators and values of γ_0^2 and σ_0^2 . We used $M = 30$ noisy measurements $\mathbf{y} = (y_1, \ldots, y_M)$.

to TV-based methods [41], but the extent to which such estimators approach the MMSE solution is still unknown.

VII. CONCLUSION

We have presented two approaches for the reconstruction of periodic continuous-domain signals from their corrupted discrete measurements. The first approach is based on optimization theory and culminates with the specification of a periodic representer theorem (Theorem 1). In the second approach, a signal is modeled as a stationary periodic random process and the reconstruction problem is transformed into an estimation problem. Theorem 2 then gives the optimal estimator (in the mean-square sense) for Gaussian bridges.

We have also provided theoretical and experimental comparisons of the two approaches and identified two main findings. First, for invertible operators, the statistical and variational approaches are equivalent and correspond to an identical reconstruction scheme. For noninvertible operators, however, this equivalence is not valid anymore, but the variational method corresponds to the statistical reconstruction when the parameter γ vanishes. More importantly, for small values of γ_0^2, the variational method is practically equivalent to the optimal statistical reconstruction. This demonstrates the efficiency of the representer theorem for reconstructing Gaussian bridges, even for noninvertible operators.

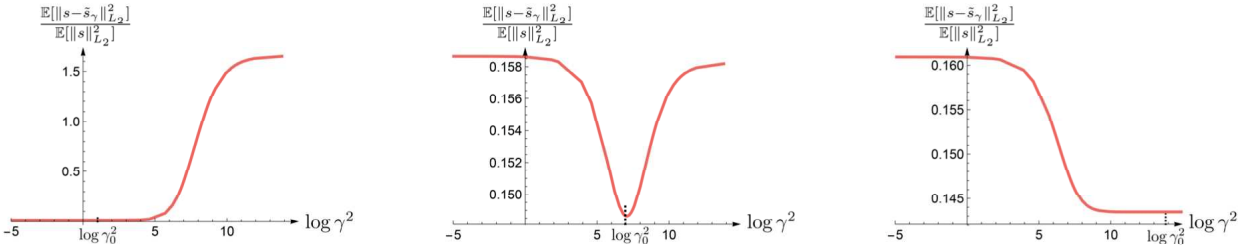
APPENDIX

A. Proof of Proposition 1

The main argument is very classical in the non-periodic setting. We detail it for the sake of completeness and adapt it to the periodic case.

Let p be a function of $\mathcal{N}_{\mathrm{L}}$. As L is shift-invariant, $p(\cdot - t_0) \in \mathcal{N}_{\mathrm{L}}$ for every $t_0 \in \mathbb{T}$. Moreover, $\mathcal{N}_{\mathrm{L}}$ is closed in $\mathcal{S}'(\mathbb{T})$ (as any finite-dimensional linear subspace), thus the first derivative $p' = p^{(1)}$ of p is in $\mathcal{N}_{\mathrm{L}}$ as the limit of the function $\frac{1}{t_0}(p(\cdot - t_0) - p) \in \mathcal{N}_{\mathrm{L}}$ when $t_0 \to 0$. We propagate this property to all the derivatives of p.

We now have that $\mathcal{N}_{\mathrm{L}}$ is a finite-dimensional space of dimension N_0 and $p^{(k)} \in \mathcal{N}_{\mathrm{L}}$, $\forall k \in [1 \ldots N_0]$. Hence, the family of $(N_0 + 1)$ functions $p, p^{(1)}, \ldots, p^{(N_0)}$ satisfies an equation of the form $a_{N_0}p^{(N_0)} + \cdots + a_0 p = 0$, where $a_k \in \mathbb{C}$ and $(a_0, \ldots, a_{N_0}) \neq \mathbf{0}$. This implies that p, as solution of



(a) $\gamma_0^2 = 1$. (b) $\gamma_0^2 = 10^3$. (c) $\gamma_0^2 = 10^6$.

Fig. 5. Evolution of NMSE according to γ for $s \sim \mathcal{GB}(\mathrm{D}, \gamma_0^2)$.

a differential equation with constant coefficients, is a sum of functions of the form $q(t)e^{\mu t}$ with q a polynomial and $\mu \in \mathbb{C}$.

Finally, since we deal with 1-periodic functions, this constrains q to be a constant function and $\mu = 2\pi \mathrm{j} k$ with $k \in \mathbb{Z}$. This concludes the proof.

B. Proof of Proposition 2

The linearity, Hermitian symmetry, and non-negativity are easily obtained. We only need to verify that $\|f\|_{\mathcal{H}_\mathrm{L}} = \langle f, f\rangle^{\frac{1}{2}}_{\mathcal{H}_\mathrm{L}} = 0 \Leftrightarrow f = 0$. For this, we observe that

$$\begin{aligned}\langle f, f\rangle_{\mathcal{H}_\mathrm{L}} = 0 &\Leftrightarrow \int_0^1 |\mathrm{L}f(t)|^2 \mathrm{d}t + \gamma^2 \sum_{n=1}^{N_0} |\widehat{f}[k_n]|^2 = 0 \\ &\Leftrightarrow \sum_{k\in\mathcal{K}_\mathrm{L}} \left|\widehat{f}[k]\right|^2 \underbrace{\left|\widehat{L}[k]\right|^2}_{\neq 0} + \gamma^2 \sum_{n=1}^{N_0} |\widehat{f}[k_n]|^2 = 0,\end{aligned} \tag{43}$$

which implies that $\widehat{f}[k] = 0$ for all $k \in \mathbb{Z}$. Hence, $\langle f, f\rangle_{\mathcal{H}_\mathrm{L}} = 0 \Leftrightarrow f = 0$.

C. Proof of Proposition 3

For the proof, we set $A = \sum_{k\in\mathcal{K}_\mathrm{L}} \frac{1}{|\widehat{L}[k]|^2}$. The Hilbert space $\mathcal{H}_\mathrm{L}$ is a RKHS if and only if $\mathrm{III} \in \mathcal{H}'_\mathrm{L}$ or, equivalently, if there exists $C > 0$ such that

$$\forall f \in \mathcal{S}(\mathbb{T}), \quad |\langle \mathrm{III}, f\rangle| \leq C\|f\|_{\mathcal{H}_\mathrm{L}}. \tag{44}$$

Assume that $A < +\infty$. Let c be the sequence such that $c[k] = 1/\widehat{L}[k]$ if $k \in \mathcal{K}_\mathrm{L}$ and $c[k] = 1/\gamma$ otherwise. Using the Cauchy-Schwarz inequality, we have, for every $f \in \mathcal{S}(\mathbb{T})$, that

$$\begin{aligned}\langle \mathrm{III}, f\rangle^2 = \left(\sum \widehat{f}[k]\right)^2 &\leq \left(\sum |c[k]|^2\right)\left(\sum \left|\frac{\widehat{f}[k]}{c[k]}\right|^2\right) \\ &= (N_0/\gamma^2 + A)\|f\|^2_{\mathcal{H}_\mathrm{L}}.\end{aligned} \tag{45}$$

Hence, (44) is satisfied for $C = (N_0/\gamma^2 + A)^{1/2} > 0$. For the converse, we define $f_m \in \mathcal{S}(\mathbb{T})$ such that

$$\widehat{f_m}[k] = \begin{cases} 0, & \text{if } |k| > m \text{ or } k = k_n, n \in [1 \dots N_0] \\ \frac{1}{|\widehat{L}[k]|^2}, & \text{otherwise.}\end{cases}$$

Then, we readily observe that $\lim_{m\to+\infty} \frac{|\langle \mathrm{III}, f_m\rangle|}{\|f_m\|_{\mathcal{H}_\mathrm{L}}} = \sqrt{A}$. Therefore, as soon as $A = +\infty$, $\langle \mathrm{III}, f\rangle / \|f\|_{\mathcal{H}_\mathrm{L}}$ is not bounded in $\mathcal{S}(\mathbb{T})$ and $\mathcal{H}_\mathrm{L}$ is not a RKHS.

The reproducing kernel is characterized by the relation $f(\tau) = \langle h(\cdot, \tau), f\rangle_{\mathcal{H}_\mathrm{L}}$ for every $f \in \mathcal{H}_\mathrm{L}$. Let R be the operator, often called the Riesz map, such that $\langle \mathrm{R}g, f\rangle_{\mathcal{H}_\mathrm{L}} = \langle g, f\rangle$ for any $f \in \mathcal{H}_\mathrm{L}$ and $g \in \mathcal{H}'_\mathrm{L}$. Then, $h(\cdot, \tau) = \mathrm{R}\{\mathrm{III}(\cdot - \tau)\}$. Moreover, we have that $\langle \mathrm{R}e_k, e_m\rangle_{\mathcal{H}_\mathrm{L}} = \delta[k-m]$. In addition,

$$\begin{aligned}\langle \mathrm{R}e_k, e_m\rangle_{\mathcal{H}_\mathrm{L}} &= \langle \mathrm{LR}e_k, \mathrm{L}e_m\rangle + \gamma^2 \sum_{n=1}^{N_0} \widehat{\mathrm{R}e_k}[k_n]\overline{\widehat{e_m}[k_n]} \\ &= \langle \mathrm{R}e_k, \mathrm{L}^*\mathrm{L}e_m\rangle + \gamma^2 \sum_{n=1}^{N_0} \widehat{\mathrm{R}e_k}[k_n]\delta[m-k_n] \\ &= |\widehat{L}[m]|^2 \widehat{\mathrm{R}e_k}[m] + \gamma^2 \sum_{n=1}^{N_0} \widehat{\mathrm{R}e_k}[k_n]\delta[m-k_n].\end{aligned} \tag{46}$$

Hence, R is characterized for $k, m \in \mathbb{Z}$ by the relation

$$|\widehat{L}[m]|^2 \widehat{\mathrm{R}e_k}[m] + \gamma^2 \sum_{n=1}^{N_0} \widehat{\mathrm{R}e_k}[k_n]\delta[m-k_n] = \delta[k-m]. \tag{47}$$

For $k \in \mathcal{K}_\mathrm{L}$, we deduce from (47) that $\widehat{\mathrm{R}e_k}[m] = 1/|\widehat{L}[k]|^2$ if $m = k$ and 0 otherwise. We also deduce that, for $k = k_n$, $\widehat{\mathrm{R}e_{k_n}}[m] = 1/\gamma^2$ if $m = k_n$ and 0 otherwise. Thus, R is shift-invariant ($\widehat{\mathrm{R}e_k}[m] = 0$ for every $m \neq k$), meaning that $h(t, \tau)$ depends only on $(t-\tau)$. Moreover, the Fourier multiplier of R, which is also the discrete Fourier transform of $h_\gamma(t) = h(t, 0)$, is $\widehat{R}[k] = 1/|\widehat{L}[k]|^2$ if $k \in \mathcal{K}_\mathrm{L}$ and $1/\gamma^2$ if $k = k_n$. This is equivalent to (18) and concludes the proof.

D. Proof of Theorem 1

To prove Theorem 1, we first show that the optimization problem (19) has a unique solution by convex-optimization arguments. Then, we connect this solution to the abstract representer theorem (see for instance [63, Theorem 16.1]) to deduce the form of the solution. We start with some preliminary results for the first part.

Lemma 1. *Under the condition of Theorem 1, the functional* $\phi : \mathcal{H}_\mathrm{L} \to \mathbb{R}^+$ *defined by* $\phi(f) = F(\mathbf{y}, \langle \boldsymbol{\nu}, f\rangle) + \lambda\|\mathrm{L}f\|^2_{L_2}$ *is strictly convex and coercive, meaning that* $\phi(f) \to \infty$ *when* $\|f\|_{\mathcal{H}_\mathrm{L}} \to \infty$.

Proof. Strict convexity: ϕ is convex as a sum of two convex functions. For the strict convexity, we fix $\mu \in (0, 1)$ and $f, g \in \mathcal{H}_\mathrm{L}$. It is then sufficient to show that the equality

$$\phi(\mu f + (1-\mu)g) = \mu\phi(f) + (1-\mu)\phi(g) \tag{48}$$

implies that $f = g$. The functions $F(\mathbf{y}, \boldsymbol{\nu}\{\cdot\})$ and $\|\mathrm{L}\cdot\|_{L_2}$ are convex, therefore (48) together with the linearity of both $\boldsymbol{\nu}$ and L implies the two relations

$$F(\mathbf{y}, \mu\nu(f) + (1-\mu)\nu(g)) = \mu F(\mathbf{y}, \nu(f)) + (1-\mu)F(\mathbf{y}, \nu(g))$$
$$\|\mu \mathrm{L}f + (1-\mu)\mathrm{L}g\|_{L_2}^2 = \mu\|\mathrm{L}f\|_{L_2}^2 + (1-\mu)\|\mathrm{L}g\|_{L_2}^2. \quad (49)$$

Now, taking advantage of the strict convexity of $F(\mathbf{y}, \cdot)$ and $\|\cdot\|_{L_2}^2$, we deduce that $\boldsymbol{\nu}(f) = \boldsymbol{\nu}(g)$ and $\mathrm{L}f = \mathrm{L}g$. This means, in particular, that $(f-g)$ is in the intersection of the null spaces of $\boldsymbol{\nu}$ and L, assumed to be trivial. Finally, $f = g$ as expected.

Coercivity: The measurement functional $\boldsymbol{\nu}$ is linear and continuous, hence there exists $A > 0$ such that $|\langle \nu, f\rangle|^2 \leq A\|f\|_{\mathcal{H}_\mathrm{L}}^2$ for any $f \in \mathcal{H}_\mathrm{L}$. Moreover, since $\boldsymbol{\nu}$ is injective and linear when restricted to the finite-dimensional null space $\mathcal{N}_\mathrm{L}$, there exists $B > 0$ such that $\|\langle \nu, p\rangle\|^2 \geq B\|p\|_{\mathcal{H}_\mathrm{L}}^2$ for any $p \in \mathcal{N}_\mathrm{L}$. Any $f \in \mathcal{H}_\mathrm{L}$ can be decomposed uniquely as

$$f = \sum_{k \in \mathcal{K}_\mathrm{L}} \widehat{f}[k] e_k + \sum_{n=1}^{N_0} \widehat{f}[k_n] e_{k_n} = g + p. \quad (50)$$

In that case, we easily see that $\|g\|_{\mathcal{H}_\mathrm{L}} = \|\mathrm{L}f\|_{L_2}$. In particular, we deduce that

$$\begin{aligned}\|f\|_{\mathcal{H}_\mathrm{L}}^2 &= \|g\|_{\mathcal{H}_\mathrm{L}}^2 + \|p\|_{\mathcal{H}_\mathrm{L}}^2 \leq \|\mathrm{L}f\|_{L_2}^2 + \frac{1}{B}\|\langle \boldsymbol{\nu}, p\rangle\|^2 \\ &\leq \|\mathrm{L}f\|_{L_2}^2 + \frac{1}{B}\left(\|\langle \boldsymbol{\nu}, f\rangle\| + \|\langle \boldsymbol{\nu}, g\rangle\|\right)^2 \\ &\leq \|\mathrm{L}f\|_{L_2}^2 + \frac{1}{B}\left(\|\langle \boldsymbol{\nu}, f\rangle\| + A^{1/2}\|\mathrm{L}f\|_{L_2}\right)^2 \\ &\leq C\left(\|\mathrm{L}f\|_{L_2}^2 + \|\langle \boldsymbol{\nu}, f\rangle\|^2\right) \quad (51)\end{aligned}$$

for $C > 0$ large enough. Now, consider a sequence of functions $f_m \in \mathcal{H}_\mathrm{L}$ such that $\|f\|_{\mathcal{H}_\mathrm{L}} \to \infty$. We want to show that, for m large enough, $\phi(f_m)$ is arbitrarily large. Due to (51), for m large enough, $\|\mathrm{L}f_m\|_{L_2}$ or $\|\langle \boldsymbol{\nu}, f_m\rangle\|$ are arbitrarily large. The former implies obviously that $\phi(f_m)$ can be made as large as we want. It is also true for the latter because $\phi(f_m) \geq F(\mathbf{y}, \langle \boldsymbol{\nu}, f_m\rangle)$ and F is coercive. This means that $\phi(f_m)$ goes to infinity when $m \to \infty$, hence ϕ is coercive. □

As ϕ is a strictly convex and coercive functional (Lemma 1), the optimization problem (19) has the unique solution f_RT. We denote $z_0 = \langle \boldsymbol{\nu}, f_\mathrm{RT}\rangle$. The function f_RT can be uniquely decomposed as

$$f_\mathrm{RT} = \sum_{k \in \mathcal{K}_\mathrm{L}} \widehat{f}_\mathrm{RT}[k] e_k + \sum_{n=1}^{N_0} \widehat{f}_\mathrm{RT}[k_n] e_{k_n} = g_\mathrm{RT} + p_\mathrm{RT}. \quad (52)$$

We recall the abstract representer theorem. This result can be found in [33, Theorem 8] with a formulation close to ours.

Proposition 9. *Let $\mathcal{H}$ be a Hilbert space, $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_M)$ be a vector of M linear and continuous measurement functionals over $\mathcal{H}$, and $\mathbf{y}_0 \in \mathbb{R}^M$. There exists a unique minimizer of the optimization problem*

$$\min_{f \in \mathcal{H}} \|f\|_\mathcal{H} \text{ s.t. } \boldsymbol{\nu} = \mathbf{y}_0, \quad (53)$$

which is of the form $f_\mathrm{opt} = \sum_{m=1}^{M} a_m \mathrm{R}\nu_m$, where $a_m \in \mathbb{R}$ and $\mathrm{R} : \mathcal{H}' \to \mathcal{H}$ is the Riesz map of $\mathcal{H}$.

We consider the Hilbert space $\widetilde{\mathcal{H}}_\mathrm{L} = \{f \in \mathcal{H}_\mathrm{L},\ \mathrm{Proj}_{\mathcal{N}_\mathrm{L}}\{f\} = 0\}$, on which $\|\mathrm{L}f\|_{L_2}$ is a Hilbertian norm. The linear measurements ν_m are in the dual space $\widetilde{\mathcal{H}}_\mathrm{L}'$, once restricted as linear functionals on $\widetilde{\mathcal{H}}_\mathrm{L}$. The interpolation constraint is chosen as $\mathbf{y}_0 = \mathbf{z}_0 - \boldsymbol{\nu}(p_\mathrm{RT})$. Applying Proposition 9 to this case, we deduce that there exists a unique minimizer

$$h_\mathrm{opt} = \underset{h \in \widetilde{\mathcal{H}}_\mathrm{L}, \boldsymbol{\nu}(h) = \mathbf{y}_0}{\arg\min} \|\mathrm{L}h\|_{L_2} \quad (54)$$

which is of the form $h_\mathrm{opt} = \sum_{m=1}^{M} a_m \mathrm{R}\nu_m$, R being the Riesz map between $\widetilde{\mathcal{H}}_\mathrm{L}'$ and $\widetilde{\mathcal{H}}_\mathrm{L}$. In our case, the function $\mathrm{R}\nu_m$ is given by $\mathrm{R}\nu_m = \sum_{k \in \mathcal{K}_\mathrm{L}} \frac{\widehat{\nu}_m[k]}{|\widehat{L}[k]|^2} e_k$. In particular, one easily sees from the expression of φ_m that it satisfies

$$\mathrm{R}\nu_m = \varphi_m - \gamma^2 \mathrm{Proj}_{\mathcal{N}_\mathrm{L}}\{\nu_m\}. \quad (55)$$

Moreover, we have that $h_\mathrm{opt} = g_\mathrm{RT}$. Indeed, g_RT is clearly among the functions h over which one minimizes and one cannot have that $\|\mathrm{L}h_\mathrm{opt}\|_{L_2} < \|\mathrm{L}g_\mathrm{RT}\|_{L_2}$ (otherwise, the function $f = h_\mathrm{opt} + p_\mathrm{RT}$ would be a minimizer of (19) different from f_RT, which is impossible). Putting things together, we get that

$$\begin{aligned}f_\mathrm{RT} = g_\mathrm{RT} + p_\mathrm{RT} &= \sum_{m=1}^{M} a_m \mathrm{R}\nu_m + p_\mathrm{RT} \\ &= \sum_{m=1}^{M} a_m \varphi_m - \gamma^2 \sum_{m=1}^{M} a_m \mathrm{Proj}_{\mathcal{N}_\mathrm{L}}\{\nu_m\} + p_\mathrm{RT}. \quad (56)\end{aligned}$$

Since $(-\gamma^2 \sum_{m=1}^{M} a_m \mathrm{Proj}_{\mathcal{N}_\mathrm{L}}\{\nu_m\} + p_\mathrm{RT})$ is in the null space of L, it can be developed as $\sum_{n=1}^{N_0} b_n e_{k_n}$, giving (20).

The last ingredient is to remark that a_m satisfies $\mathbf{P}^\mathsf{T}\mathbf{a} = \mathbf{0}$. This comes from the fact that, by construction, $\sum a_m \mathrm{R}\nu_m \in \widetilde{\mathcal{H}}_\mathrm{L}'$ and, by applying the Riesz map, $\sum a_m \nu_m \in \widetilde{\mathcal{H}}_\mathrm{L}$, meaning that the projection of this element into the null space is zero. This is precisely equivalent with the expected condition.

E. Proof of Proposition 4

We compute (19) for F the quadratic cost function. We have that $f_\mathrm{RT} = \sum_{m=1}^{M} a_m \varphi_m + \sum_{n=1}^{N_0} b_n e_{k_n}$, as given by (20). It then suffices to find the optimal vectors $\mathbf{a}$ and $\mathbf{b}$. We therefore rewrite (19) in terms of these two vectors.

From simple computations, we have, with the notations of Proposition 4, that $\langle \boldsymbol{\nu}, \sum_{n=1}^{N_0} b_n e_{k_n}\rangle = \mathbf{Pb}$ and $\langle \boldsymbol{\nu}, \sum_{m=1}^{M} a_m \varphi_m\rangle = \mathbf{Ga}$, where we used for the latter that $\mathbf{G}_{m_1,m_2} = \langle \nu_{m_1}, h_\gamma * \nu_{m_2}\rangle = \langle \nu_{m_1}, \varphi_{m_2}\rangle$. Hence,

$$\|\mathbf{y} - \langle \boldsymbol{\nu}, f\rangle\|^2 = \|\mathbf{y} - \mathbf{Ga} - \mathbf{Pb}\|^2. \quad (57)$$

From the definition of h_γ in (18), we see that $(\mathrm{L}^*\mathrm{L}h_\gamma) * f = f$ for every f whose Fourier coefficients $\widehat{f}[k_n]$ do vanish for every $n = 1 \ldots N_0$. Now, the relation $\overline{\mathbf{P}}^\mathsf{T}\mathbf{a} = \mathbf{0}$ in Theorem 1 shows precisely that $\sum_{n=1}^{M} a_m \nu_m$ satisfies this property. In particular, we deduce that

$$\mathrm{L}^*\mathrm{L}\left\{\sum_{m=1}^{M} a_m \varphi_m\right\} = (\mathrm{L}^*\mathrm{L}h_\gamma) * \sum_{m=1}^{M} a_m \nu_m = \sum_{m=1}^{M} a_m \nu_m. \quad (58)$$

As a consequence, we have that

$$\|\mathrm{L}f_{\mathrm{RT}}\|_{L_2}^2 = \langle \mathrm{L}^*\mathrm{L} \sum_{m_1=1}^{M} a_{m_1}\varphi_{m_1}, \sum_{m_2=1}^{M} a_{m_2}\varphi_{m_2}\rangle = \sum_{m_1=1}^{M}\sum_{m_2=1}^{M} a_{m_1}\mathbf{G}_{m_1,m_2}a_{m_2} = (\mathbf{G}\mathbf{a})^\mathsf{T}\mathbf{a}. \quad (59)$$

Finally, one has that

$$\|\mathbf{y}-\langle \boldsymbol{\nu}, f_{\mathrm{RT}}\rangle\|^2+\lambda\|\mathrm{L}f_{\mathrm{RT}}\|_{L_2}^2 = \|\mathbf{y}-\mathbf{G}\mathbf{a}-\mathbf{P}\mathbf{b}\|^2+\lambda(\mathbf{G}\mathbf{a})^\mathsf{T}\mathbf{a}. \quad (60)$$

By computing the partial derivatives, we find that the vectors $\mathbf{a}$ and $\mathbf{b}$ are given by (21).

F. Proof of Proposition 5

Since $\nu_m = \mathrm{III}(\cdot - t_m)$, the form of the solution (20) is $f_{\mathrm{RT}}(t) = \sum_{m=1}^{M} a_m h_\gamma(t-t_m) + \sum_{n=1}^{N_0} b_n e_{k_n}(t)$. We have moreover that $\mathbf{P}^\mathsf{T}\mathbf{a} = \mathbf{0}$, where $[\mathbf{P}]_{m,n} = \mathrm{e}^{\mathrm{j}2\pi k_n t_m}$. From (18), we then deduce that $\mathrm{L}^*\mathrm{L}\{h_\gamma\}(t) = \sum_{k\in\mathcal{K}_\mathrm{L}} |\widehat{L}[k]|^2 \frac{e_k(t)}{|\widehat{L}[k]|^2} = \Big(\mathrm{III}(t) - \mathrm{Proj}_{\mathcal{N}_\mathrm{L}}\{\mathrm{III}\}(t)\Big)$. By linearity, we get that

$$\begin{aligned}
\mathrm{L}^*\mathrm{L}\{f_{\mathrm{RT}}\}(t) &= \sum_{m=1}^{M} a_m \mathrm{L}^*\mathrm{L}\{h_\gamma\}(t-t_m)\\
&= \sum_{m=1}^{M} a_m \mathrm{III}(t-t_m) - \sum_{m=1}^{M} a_m \mathrm{Proj}_{\mathcal{N}_\mathrm{L}}\{\mathrm{III}(\cdot-t_m)\}(t)\\
&= \sum_{m=1}^{M} a_m \mathrm{III}(t-t_m) - \sum_{n=1}^{N_0}\sum_{m=1}^{M} a_m \mathrm{e}^{-\mathrm{j}2\pi k t_m} e_{k_n}\\
&= \sum_{m=1}^{M} a_m \mathrm{III}(t-t_m) - \sum_{n=1}^{N_0} [\overline{\mathbf{P}}^\mathsf{T}\mathbf{a}]_n e_{k_n} \quad (61)\\
&= \sum_{m=1}^{M} a_m \mathrm{III}(t-t_m), \quad (62)
\end{aligned}$$

where we used that $[\overline{\mathbf{P}}]_{m,n} = \mathrm{e}^{-\mathrm{j}2\pi k t_m}$ in (61) and that $\overline{\mathbf{P}}^\mathsf{T}\mathbf{a} = \overline{\mathbf{P}^\mathsf{T}\mathbf{a}} = \mathbf{0}$ in (62). Finally, f_{RT} is a periodic $(\mathrm{L}^*\mathrm{L})$-spline with weights a_m and knots t_m.

G. Proof of Proposition 6

We start from

$$s = \sum_{k\in\mathcal{K}_\mathrm{L}} \frac{\widehat{w}[k]}{\widehat{L}[k]} e_k + \sum_{n=1}^{N_0} \frac{\widehat{w}[k_n]}{\gamma_0} e_{k_n}. \quad (63)$$

Our goal is to compute $r_s(t,\tau) = \mathbb{E}[s(t)s(\tau)]$. We do so by replacing $s(t)$ and $s(\tau)$ with (63). We develop the product and use the relations $\mathbb{E}[\widehat{w}[k]\widehat{w}[\ell]] = \mathbb{E}[\widehat{w}[k]^2] = 0$, $\mathbb{E}[|\widehat{w}[k]|^2] = 1$ for every $k,\ell \in \mathbb{Z}$, $k \neq \ell$ to deduce that

$$r_s(t,\tau) = \left(\sum_{k\in\mathcal{K}_\mathrm{L}} \frac{e_k(t)e_{-k}(\tau)}{|\widehat{L}[k]|^2} + \frac{1}{\gamma_0^2}\sum_{n=1}^{N_0} e_{k_n}(t)e_{-k_n}(\tau)\right). \quad (64)$$

Since $e_k(t)e_{-k}(\tau) = e_k(t-\tau)$, we have shown that $r_s(t,\tau) = h_\gamma(t-\tau)$, as expected. Then, we obtain (29) by injecting (28) into (24). Finally, we obtain (30) by particularizing (29) with $\nu_m = e_k$.

H. Proof of Theorem 2

We fix a time $t_0 \in \mathbb{T}$. We first obtain the MMSE estimator for $s(t_0)$ (estimation of s at time t_0). (Note that $s(t_0) = \langle s, \mathrm{III}(\cdot - t_0)\rangle$ is well defined because $\mathrm{III}(\cdot - t_0) \in \mathcal{H}_\mathrm{L}$ by assumption).

The linear MMSE estimator of $s(t_0)$ based on $\mathbf{y}$ is of the form $\tilde{s}_{t_0} = \sum_{m=1}^{M} u_m y_m$. Because s and ϵ are Gaussian, the linear MMSE estimator coincides with the MMSE estimator [13]. The orthogonality principle [Section 3.2] [13] then implies that

$$\mathbb{E}[y_m(s(t_0) - \tilde{s}_{t_0})] = 0, \ \forall m = 1 \dots M. \quad (65)$$

We know from Proposition 6 that $\mathbb{E}[\langle s,f\rangle\langle s,g\rangle] = \langle h_{\gamma_0} * f, g\rangle$. We use this relation to develop the different terms of (65). First, we have that

$$\begin{aligned}
\mathbb{E}[y_m s(t_0)] &= \mathbb{E}[\langle \nu_m, s\rangle s(t_0)] + \mathbb{E}[\epsilon_m s(t_0)]\\
&= \mathbb{E}[\langle \nu_m, s\rangle \langle s, \mathrm{III}(\cdot - t_0)\rangle] + \underbrace{\mathbb{E}[\epsilon_m]}_{0}\mathbb{E}[s(t_0)]\\
&= (h_{\gamma_0} * \nu_m)(t_0). \quad (66)
\end{aligned}$$

As the estimator is of the form $\tilde{s}_{t_0} = \sum_{m=1}^{M} u_m y_m$ and exploiting that ϵ and s are independent, we have that

$$\begin{aligned}
\mathbb{E}[\langle \nu_m, s\rangle y_k] &= \mathbb{E}[\langle \nu_m, s\rangle \langle \nu_k, s\rangle] + \mathbb{E}[\langle \nu_m, s\rangle \epsilon_k] = \langle h_{\gamma_0} * \nu_m, \nu_k\rangle\\
\mathbb{E}[\epsilon_m y_k] &= \mathbb{E}[\epsilon_m \langle \nu_k, s\rangle] + \mathbb{E}[\epsilon_m \epsilon_k] = \sigma^2\delta[m-k] \quad (67)
\end{aligned}$$

We have therefore that

$$\begin{aligned}
\mathbb{E}[y_m \tilde{s}_{t_0}] &= \mathbb{E}[\langle \nu_m, s\rangle \tilde{s}_{t_0}] + \mathbb{E}[\epsilon_m \tilde{s}_{t_0}]\\
&= \sum_{k=1}^{M} u_k \mathbb{E}[\langle \nu_m, s\rangle y_k] + \sum_{k=1}^{M} u_k \mathbb{E}[\epsilon_m y_k]\\
&= \sum_{k=1}^{M} u_k \langle h_{\gamma_0} * \nu_m, \nu_k\rangle + u_m \sigma_0^2. \quad (68)
\end{aligned}$$

We remark that $\langle h_{\gamma_0} * \nu_m, \nu_k\rangle = [\mathbf{G}]_{m_1,m_2}$ given in (22). Injecting (66) and (68) into (65), we have for $m = 1 \dots M$ that $(h_{\gamma_0} * \nu_m)(t_0) = \sum_{k=1}^{M} u_k [\mathbf{G}]_{m_1,m_2} + u_m\sigma_0^2$. Hence, $\mathbf{u} = (\mathbf{G} + \sigma_0^2\mathbf{I})^{-1}\mathbf{c}$, where $\mathbf{c} = (h_{\gamma_0} * \boldsymbol{\nu})(t_0)$. As $\tilde{s}_{t_0} = \mathbf{u}^\mathsf{T}\mathbf{y}$, we finally have that $\tilde{s}_{t_0} = \sum_{m=1}^{M} d_m (h_{\gamma_0} * \nu_m)(t_0)$, where $\mathbf{d} = (d_1, \dots, d_M) = (\mathbf{G} + \sigma_0^2\mathbf{I})^{-1}\mathbf{y}$.

We have now obtained the form of the MMSE estimator $\tilde{s}_{t_0}$ for $s(t_0)$ at a fixed time t_0. We then deduce the MMSE estimator of the complete continuous random process $s : \mathbb{T} \to \mathbb{R}$ that minimizes $\mathbb{E}[\|s - \tilde{s}\|_{L_2}^2]$ among all the estimators $\tilde{s}$ based on $\mathbf{y}$. We fix an estimator $\tilde{s}$. We have that

$$\begin{aligned}
\mathbb{E}[\|s - \tilde{s}\|_{L_2}^2] &= \mathbb{E}[\int_0^1 (s(t) - \tilde{s}(t))^2 \mathrm{d}t] = \int_0^1 \mathbb{E}[(s(t) - \tilde{s}(t))^2]\mathrm{d}t\\
&\geq \int_0^1 \mathbb{E}[(s(t) - \tilde{s}_t)^2]\mathrm{d}t = \mathbb{E}[\|s - \tilde{s}_{\mathrm{MMSE}}\|_{L_2}^2]. \quad (69)
\end{aligned}$$

Hence, the function $\tilde{s}_{\mathrm{MMSE}} : t \to \tilde{s}_t$ is the MMSE estimator of the complete process $s(t)$.

I. Proof of Proposition 7

The proof is obtained by following the arguments of Theorem 1 (for existence, unicity, and form of the solution) and

Proposition 4 (for the explicit formula of the coefficients d_m in (35)) with the following simplifications:
First, the existence and unicity of a solution is now direct. Indeed, the functional to minimize is $\|\mathbf{y} - \nu(f)\|_2^2 + \lambda \|f\|_{\mathcal{H}_{\mathrm{L}}}^2$. It is clearly coercive and strictly convex because $\|\cdot\|_{\mathcal{H}_{\mathrm{L}}}$ is. Second, the abstract representer theorem can now be applied directly to the Hilbert space $\mathcal{H}_{\mathrm{L}}$. The form of the solution is then directly deduced. Third, the coefficients d_m are found with the arguments of Appendix E, except that there is no term for the null-space component (coefficients b_n) in that case, hence the system matrix is simpler.

J. Proof of Proposition 8

We know the expression of $\tilde{s}_\lambda$ from Proposition 7. For Fourier sampling, the φ_m are complex exponential themselves, given by $\varphi_m = h * e_{k_m} = \widehat{h}[k_m] e_{k_m}$, while the Gram matrix $\mathbf{G}$ is diagonal since $\mathbf{G}_{m_1,m_2} = \langle h * e_{k_{m_1}}, e_{k_{m_2}} \rangle = \widehat{h}[k_{m_1}] \delta[k_{m_1} - k_{m_2}]$. Hence, (35) gives that

$$\tilde{s}_\lambda = \sum_{m=1}^{M} \frac{(\widehat{s}[k_m] + \epsilon_m)\widehat{h}[k_m]}{\widehat{h}[k_m] + \lambda} e_{k_m}. \tag{70}$$

After simplification, we have that

$$s - \tilde{s}_\lambda = \sum_{m=1}^{M} \left(\frac{\lambda \widehat{s}[k_m]}{\widehat{h}[k_m] + \lambda} - \frac{\widehat{h}[k_m]\epsilon_m}{\widehat{h}[k_m] + \lambda} \right) e_{k_m} + \sum_{k \notin \mathcal{N}_\nu} \widehat{s}[k] e_k. \tag{71}$$

Exploiting the Fourier-domain independence, we deduce that

$$\begin{aligned} \mathbb{E}\left[\|s - \tilde{s}_\lambda\|_{L_2}^2 \right] = \sum_{m=1}^{M} & \frac{\lambda^2}{(\widehat{h}[k_m] + \lambda)^2} \mathbb{E}\left[|\widehat{s}[k_m]|^2 \right] \\ & + \frac{\widehat{h}[k_m]^2}{(\widehat{h}[k_m] + \lambda)^2} \mathbb{E}\left[|\epsilon_m|^2 \right] \\ & + \sum_{k \notin \mathcal{N}_\nu} \mathbb{E}\left[|\widehat{s}[k]|^2 \right]. \end{aligned} \tag{72}$$

From the relations $\mathbb{E}\left[|\widehat{s}[k]|^2 \right] = \widehat{h}[k]$ (see (30)) and $\mathbb{E}\left[|\epsilon_m|^2 \right] = \sigma_0^2$, we finally obtain (39).

REFERENCES

[1] M. Banham and A. Katsaggelos, "Digital image restoration," *IEEE Signal Processing Magazine*, vol. 14, no. 2, pp. 24–41, March 1997.

[2] N. Karayiannis and A. Venetsanopoulos, "Regularization theory in image restoration—The stabilizing functional approach," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 7, pp. 1155–1179, July 1990.

[3] M. Figueiredo and R. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Transactions on Image Processing*, vol. 12, no. 8, pp. 906–916, August 2003.

[4] M. Afonso, J. Bioucas-Dias, and M. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Transactions on Image Processing*, vol. 20, no. 3, pp. 681–695, March 2011.

[5] M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging*. CRC press, 1998.

[6] B. Adcock and A. Hansen, "Generalized sampling and infinite-dimensional compressed sensing," *Foundations of Computational Mathematics*, vol. 16, no. 5, pp. 1263–1323, October 2016.

[7] A. Papoulis, "Generalized sampling expansion," *IEEE Transactions on Circuits and Systems*, vol. 24, no. 11, pp. 652–654, November 1977.

[8] Y. Eldar and T. Dvorkind, "A minimum squared-error framework for generalized sampling," *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 2155–2167, June 2006.

[9] E. Piccolomini, F. Zama, G. Zanghirati, and A. Formiconi, "Regularization methods in dynamic MRI," *Applied Mathematics and Computation*, vol. 132, no. 2, pp. 325–339, November 2002.

[10] E. Bostan, U. Kamilov, M. Nilchian, and M. Unser, "Sparse stochastic processes and discretization of linear inverse problems," *IEEE Transactions on Image Processing*, vol. 22, no. 7, pp. 2699–2710, July 2013.

[11] A. Tikhonov, "Solution of incorrectly formulated problems and the regularization method," *Soviet Mathematics Doklady*, vol. 4, pp. 1035–1038, 1963.

[12] K. Cassel, *Variational Methods with Applications in Science and Engineering*. Cambridge University Press, 2013.

[13] T. Moon and W. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. Prentice Hall Upper Saddle River, NJ, 2000, vol. 1.

[14] F. S. Cohen and J.-Y. Wang, "Part I: Modeling image curves using invariant 3-D object curve models—A path to 3-D recognition and shape estimation from image contours," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 1, pp. 1–12, January 1994.

[15] R. Delgado-Gonzalo, P. Thévenaz, C. Seelamantula, and M. Unser, "Snakes with an ellipse-reproducing property," *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 1258–1271, March 2012.

[16] A. Badoual, D. Schmitter, V. Uhlmann, and M. Unser, "Multiresolution subdivision snakes," *IEEE Transactions on Image Processing*, vol. 26, no. 3, pp. 1188–1201, March 2017.

[17] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Transactions on Signal Processing*, vol. 50, no. 6, pp. 1417–1428, June 2002.

[18] I. Maravic and M. Vetterli, "Sampling and reconstruction of signals with finite rate of innovation in the presence of noise," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 2788–2805, August 2005.

[19] T. Blu, P.-L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, "Sparse sampling of signal innovations," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 31–40, March 2008.

[20] M. Jacob, T. Blu, and M. Unser, "Sampling of periodic signals: A quantitative error analysis," *IEEE Transactions on Signal Processing*, vol. 50, no. 5, pp. 1153–1159, May 2002.

[21] H. Triebel, *Function Spaces and Wavelets on Domains*, ser. EMS Tracts in Mathematics. European Mathematical Society (EMS), Zürich, 2008, vol. 7.

[22] J. Fageot, M. Unser, and J. P. Ward, "On the Besov regularity of periodic Lévy noises," *Applied and Computational Harmonic Analysis*, vol. 42, no. 1, pp. 21–36, January 2017.

[23] A. Badoual, D. Schmitter, and M. Unser, "An inner-product calculus for periodic functions and curves," *IEEE Signal Processing Letters*, vol. 23, no. 6, pp. 878–882, June 2016.

[24] B. Schölkopf, R. Herbrich, and A. Smola, "A generalized representer theorem," in *Computational Learning Theory*. Springer, 2001, pp. 416–426.

[25] I. Schoenberg, "Spline functions and the problem of graduation," *Proceedings of the National Academy of Sciences*, vol. 52, no. 4, pp. 947–950, August 1964.

[26] M. Unser and T. Blu, "Self-similarity: Part I—Splines and operators," *IEEE Transactions on Signal Processing*, vol. 55, no. 4, pp. 1352–1363, April 2007.

[27] V. Uhlmann, J. Fageot, and M. Unser, "Hermite snakes with control of tangents," *IEEE Transactions on Image Processing*, vol. 25, no. 6, pp. 2803–2816, June 2016.

[28] L. Condat and T. Möller, "Quantitative error analysis for the reconstruction of derivatives," *IEEE Transactions on Signal Processing*, vol. 59, no. 6, pp. 2965–2969, June 2011.

[29] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, February 2006.

[30] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[31] Q. Denoyelle, V. Duval, and G. Peyré, "Support recovery for sparse deconvolution of positive measures," *arXiv preprint arXiv:1506.08264*, 2015.

[32] M. Unser, J. Fageot, and J. P. Ward, "Splines are universal solutions of linear inverse problems with generalized TV regularization," *SIAM Review*, vol. 59, no. 4, pp. 769–793, December 2017.

[33] H. Gupta, J. Fageot, and M. Unser, "Continuous-domain solutions of linear inverse problems with Tikhonov *vs.* generalized TV regularization,"

IEEE Transactions on Signal Processing, vol. 66, no. 17, pp. 4670–4684, 2018.
[34] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, 2005.
[35] A. Berlinet and C. Thomas-Agnan, *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Springer Science & Business Media, 2011.
[36] T. Kailath, "An innovations approach to least-squares estimation–Part I: Linear filtering in additive white noise," *IEEE Transactions on Automatic Control*, vol. 13, no. 6, pp. 646–655, December 1968.
[37] M. Unser, P. D. Tafti, and Q. Sun, "A unified formulation of Gaussian versus sparse stochastic processes—Part I: Continuous-domain theory," *IEEE Transactions on Information Theory*, vol. 60, no. 3, pp. 1945–1962, March 2014.
[38] G. Kimeldorf and G. Wahba, "Spline functions and stochastic processes," *Sankhyā: The Indian Journal of Statistics, Series A*, vol. 32, no. 2, pp. 173–180, June 1970.
[39] V. Uhlmann, J. Fageot, H. Gupta, and M. Unser, "Statistical optimality of Hermite splines," in *Proceedings of the Eleventh International Workshop on Sampling Theory and Applications (SampTA'15)*, Washington DC, USA, May 25-29, 2015, pp. 226–230.
[40] T. Blu and M. Unser, "Self-similarity: Part II—Optimal estimation of fractal processes," *IEEE Transactions on Signal Processing*, vol. 55, no. 4, pp. 1364–1378, April 2007.
[41] M. Unser and P. D. Tafti, *An Introduction to Sparse Stochastic Processes*. Cambridge University Press, 2014.
[42] G. Wahba, *Spline Models for Observational Data*. SIAM, 1990.
[43] M. Unser and T. Blu, "Generalized smoothing splines and the optimal discretization of the Wiener filter," *IEEE Transactions on Signal Processing*, vol. 53, no. 6, pp. 2146–2159, June 2005.
[44] I. Schoenberg, "On trigonometric spline interpolation," *Journal of Mathematics and Mechanics*, vol. 13, no. 5, pp. 795–825, 1964.
[45] M. Golomb, "Approximation by periodic spline interpolants on uniform meshes," *Journal of Approximation Theory*, vol. 1, no. 1, pp. 26–65, June 1968.
[46] C. D. Boor, *A Practical Guide to Splines*. Springer-Verlag New York, 1978, vol. 27.
[47] L. Schwartz, *Théorie des distributions*. Hermann, 1966.
[48] I. Schoenberg, *Cardinal Spline Interpolation*. Philadelphia, PA: SIAM, 1973.
[49] M. Unser, "Splines: A perfect fit for signal and image processing," *IEEE Signal Processing Magazine*, vol. 16, no. 6, pp. 22–38, November 1999.
[50] L. Schumaker, *Spline Functions: Basic Theory*. Cambridge University Press, 2007.
[51] M. Schultz and R. Varga, "L-splines," *Numerische Mathematik*, vol. 10, no. 4, pp. 345–369, November 1967.
[52] M. Unser and T. Blu, "Cardinal exponential splines: Part I—Theory and filtering algorithms," *IEEE Transactions on Signal Processing*, vol. 53, no. 4, pp. 1425–1438, April 2005.
[53] M. Unser, "Cardinal exponential splines: Part II—Think analog, act digital," *IEEE Transactions on Signal Processing*, vol. 53, no. 4, pp. 1439–1449, April 2005.
[54] R. Panda and M. Dash, "Fractional generalized splines and signal processing," *Signal Processing*, vol. 86, no. 9, pp. 2340–2350, September 2006.
[55] M. Unser, J. Fageot, and H. Gupta, "Representer theorems for sparsity-promoting ℓ_1 regularization," *IEEE Transactions on Information Theory*, vol. 62, no. 9, pp. 5167–5180, September 2016.
[56] N. R. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction)," *The Annals of Mathematical Statistics*, vol. 34, no. 1, pp. 152–177, March 1963.
[57] D. Revuz and M. Yor, *Continuous Martingales and Brownian Motion*. Springer Science & Business Media, 2013, vol. 293.
[58] B. Simon, "Distributions and their Hermite expansions," *Journal of Mathematical Physics*, vol. 12, no. 1, pp. 140–148, October 2003.
[59] I. Karatzas and S. Shreve, *Brownian Motion and Stochastic Calculus*. Springer Science & Business Media, 2012, vol. 113.
[60] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer, 2010.
[61] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*. Academic press, 2008.
[62] E. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, June 2014.
[63] H. Wendland, *Scattered Data Approximation*. Cambridge university press, 2004, vol. 17.

Anaïs Badoual received the M.Sc. degree in information technology and communication for health from Télécom Physique Strasbourg, France, and the M.Sc. degree in image processing from the University of Strasbourg, in 2014. She is currently pursuing the Ph.D. degree with the Biomedical Imaging Group, École Polytechnique Fédérale de Lausanne, Switzerland, under the direction of Michael Unser. Her research interests include subdivision theory, deep learning, and segmentation problems.

Julien Fageot was a student at the École Normale Supérieure, Paris, France, where he received M.Sc. degrees in mathematics and in imaging science in 2009 and 2011, respectively. He completed his Ph.D. degree with the Biomedical Imaging Group under the supervision of Michael Unser at École Polytechnique Fédérale de Lausanne (EPFL), Switzerland, 2017. He is currently a post-doctoral researcher, working jointly at the Biomedical Imaging Group and the Chair of Statistical Field Theory at EPFL. His primary area of investigation is the mathematical modelling of sparse signals, with special emphases on stochastic and variational approaches.

Michael Unser (M'89—SM'94—F'99) is professor and director of EPFL's Biomedical Imaging Group, Lausanne, Switzerland. His primary area of investigation is biomedical image processing. He is internationally recognized for his research contributions to sampling theory, wavelets, the use of splines for image processing, stochastic processes, and computational bioimaging. He has published over 300 journal papers on those topics. He is the author with P. Tafti of the book titled *An introduction to Sparse Stochastic Processes* (Cambridge University Press, 2014).

From 1985 to 1997, he was with the Biomedical Engineering and Instrumentation Program, National Institutes of Health, Bethesda USA, conducting research on bioimaging.

Dr. Unser has held the position of associate Editor-in-Chief (2003-2005) for the IEEE Transactions on Medical Imaging. He is currently member of the editorial boards of SIAM J. Imaging Sciences, IEEE J. Selected Topics in Signal Processing, and Foundations and Trends in Signal Processing. He is the founding chair of the technical committee on Bio Imaging and Signal Processing (BISP) of the IEEE Signal Processing Society.

Prof. Unser is a fellow of the IEEE (1999), an EURASIP fellow (2009), and a member of the Swiss Academy of Engineering Sciences. He is the recipient of several international prizes including four IEEE-SPS Best Paper Awards and two Technical Achievement Awards from the IEEE (2008 SPS and EMBS 2010).