

Fast computation of smoothing splines subject to equality constraints

Gianluigi Pillonetto^a, Alessandro Chiuso^b

^a *Department of Information Engineering
University of Padova, Padova (Italy)*

^b *Dipartimento di Tecnica e Gestione dei Sistemi Industriali, Vicenza
University of Padova, Vicenza (Italy)*

Abstract

The issue of constructing periodic smoothing splines has been recently formulated as a controlled two point boundary value problem which admits a state-space description. In the context of minimum norm problems in Hilbert spaces, it has been shown that the solution is the sum of a finite number of basis functions and can be obtained with a number of operations which scales with the cube of the sum of the number of measurements and boundary constraints. In this paper we consider a more general class of variational problems subject to equality constraints which contains the periodic smoothing spline problem as a special case. Using the theory of reproducing kernel Hilbert spaces we derive a solution to the problem which has the same computational complexity as that recently proposed. Next, assuming that the problem admits a state-space representation, we obtain an algorithm whose complexity is linear in the number of measurements. We also show that the solution of the problem admits the structure of a particular regularization network whose weights can be computed in linear time. Closed form expressions for the basis functions associated with the periodic cubic smoothing spline problem are finally derived.

Key words: regularization; nonparametric identification; inverse problems; equality constrained regularization network

1 Introduction

Smoothing splines are widely used for nonparametric estimation of functions starting from noisy data [1]. Applications can be found also in the context of optimal control and state-feedback linearization, see e.g. [2–4]. When the unknown signal has to satisfy periodic boundary constraints, a periodic smoothing spline problem arises. In this context, applications regard e.g. biomedical data analysis, differential geometry and robot path planning where closed planar curves need to be generated, see e.g. [5–7].

In absence of constraints on the spline to reconstruct, many algorithms whose complexity is linear in the number of measurements are available in the literature, see e.g. [8,9]. For what concerns the periodic case, it admits a particularly simple solution when the data points are uniformly sampled [1,7]. In a recent work, the task of constructing periodic smoothing splines, using

arbitrary sampling grids, has been faced in terms of a constrained optimal control problem in state-space [10]. In the context of minimum norm problems in Hilbert spaces and linear systems theory, an elegant expression for the solution has been worked out. It has been shown that the optimal control is the sum of a finite number of basis functions (see eq. (6) in [10]), whose expansion coefficients can be computed with $O((n+l)^3)$ operations, where n and l are the number of output values and boundary constraints, respectively (see eq. (14) in [10]). In this paper, we analyze a wider class of variational problems, subject to equality constraints, under the framework of regularization theory in Reproducing kernel Hilbert spaces (RKHS) and Kalman smoothing [11–14]. Our contribution is twofold. In the first part of the paper we define a constrained function estimation problem in an RKHS without resorting to state-space formulations. It includes the optimal control problem considered in [10] as a special case. Then, by resorting to RKHS theory [15] we derive a solution for our more general problem which has the same computational complexity of that derived in [10]. In the second part of the paper, we assume that the estimation problem admits a state-space description and obtain an $O(n)$ algorithm

Email addresses: giapi@dei.unipd.it (Gianluigi Pillonetto), chiuso@dei.unipd.it (Alessandro Chiuso).

¹ This paper was not presented at any IFAC meeting. Corresponding author Gianluigi Pillonetto Ph. +390498277607

by the Bayesian interpretation of regularization. In our framework, a more general class of boundary value problems than that treated in [16] can be handled and more compact derivations of the solutions are obtained. The paper is organized as follows: in Section 2 basic concepts on RKHS are recalled. In Section 3, we define an equality constrained Tikhonov variational problem in an RKHS and discuss its relationship with the optimal control problem of [10]. Next, we show that the solution of the problem has the structure of a particular regularization network (RN), which we call *equality constrained RN*. In Section 4 we assume that the problem admits a state-space formulation and derive an $O(n)$ algorithm, able also to determine efficiently possibly unknown hyperparameters present in the problem, e.g. the regularization parameter. In addition, we show that the weights of the equality constrained RN can be obtained in linear time, also obtaining closed form expressions for the periodic splines basis functions. In Section 6, a numerical example involving periodic cubic smoothing splines is reported. Some conclusions end the paper.

2 Preliminaries

Hereafter, vectors are column vectors and, given a vector w , w_i is its i -th component. In addition, X is the interval $[0, T]$ on \mathbb{R} , $\mathcal{C}(X)$ is the space of continuous functions on X while $\mathcal{L}^2(X)$ is the classical Lebesgue space with inner product $\langle \cdot, \cdot \rangle_2$. Further, $\mathbb{E}[\cdot]$ is the expectation operator, given two random vectors q, w , $\Sigma_{qw} := \mathbb{V}[q, w] = \mathbb{E}[(q - \mathbb{E}[q])(w - \mathbb{E}[w])^T]$, $\mathbb{V}[q] := \mathbb{E}[(q - \mathbb{E}[q])(q - \mathbb{E}[q])^T]$ and $\mathbf{N}(\mu, \Sigma)$ is a multivariate normal density of mean μ and covariance Σ .

According to [15], any continuous, symmetric and positive-definite kernel $K : X \times X \mapsto \mathbb{R}$ can be associated with one and only one Hilbert space \mathcal{H} of continuous functions, named Reproducing Kernel Hilbert Space (RKHS), having K as *reproducing kernel*. The inner product on \mathcal{H} is denoted by $\langle \cdot, \cdot \rangle_{\mathcal{H}}$. Define a causal linear operator $L[u] : \mathcal{L}^2(X) \mapsto \mathcal{C}(X)$ with impulse response $h : \mathbb{R}^2 \mapsto \mathbb{R}$, $h(t, \tau) = 0$ for $\tau > t$ and h essentially bounded on $X \times X$, i.e.

$$L[u](t) = \int_0^t h(t, \tau)u(\tau)d\tau \quad t \in X \quad u \in \mathcal{L}^2(X) \quad (1)$$

Define also the kernel

$$K_1(t, s) = \int_X h(t, \tau)h(s, \tau)d\tau \quad (t, s) \in X \times X.$$

The following result characterizes one RKHS associated with K_1 and establishes an isometric isomorphism with \mathcal{L}^2 . The proof is omitted since it can be obtained by using the same arguments reported in Section 1.2 of [1].

Proposition 2.1 *Consider $L : \mathcal{L}^2(X) \rightarrow \mathcal{C}(X)$, with*

inverse L^{-1} assumed to exist². Then, the RKHS \mathcal{H}_1 on X associated with K_1 is $\mathcal{H}_1 = \{L[u] \mid u \in \mathcal{L}^2(X)\}$ with inner product $\langle f, g \rangle_{\mathcal{H}_1} = \langle L^{-1}[f], L^{-1}[g] \rangle_2$

Let \mathcal{H}_2 be a finite-dimensional RKHS with kernel $K_2(t, s) = \sum_{i=1}^m p_i \phi_i(t)\phi_i(s)$ where $\{\phi_i\}_{i=1}^m$ is a set of linearly independent functions. For a proof of the following result, see e.g. [15] or [18].

Proposition 2.2 *Define*

$$R = \text{diag}\{p_1^{-1}, p_2^{-1}, \dots, p_m^{-1}\} \quad (2)$$

where $p_i \in \mathbb{R}^+$, $\forall i$. Then, $\mathcal{H}_2 = \text{span}\{\phi_1, \phi_2, \dots, \phi_m\}$ and, for $g(t) = \sum_{i=1}^m w_i \phi_i(t)$, $\|g\|_{\mathcal{H}_2}^2 = w^T R w$

Example 1 *In the sequel, a combination of \mathcal{H}_1 and \mathcal{H}_2 will define the Hilbert space the unknown function is assumed to belong to. As a notable example, we introduce the hypothesis space associated with the spline regression problem. Let $f^{(i)}(t)$ denote the i -th derivative of f evaluated at t and let also h in (1) be defined by*

$$h(t, \tau) = h(t - \tau) = \frac{(t - \tau)_+^{m-1}}{(m-1)!} \quad (3)$$

where $(t)_+ = t$ for $t \geq 0$ and 0 otherwise. With this choice, \mathcal{H}_1 becomes the following particular type of one-dimensional Sobolev space (see e.g. [19])

$$\{f : f^{(i)}(0) = 0 \text{ and } f^{(i)} \text{ absolutely continuous} \\ \text{for } i = 0, \dots, m-1, f^{(m)} \in \mathcal{L}^2\}$$

It holds that $\|f\|_{\mathcal{H}_1}^2 = \int_0^T (f^{(m)}(t))^2 dt$, a typical example of norm in RKHS used to penalize too complex/irregular functions. However, the conditions $f^{(i)}(0) = 0$ on the unknown function are often too restrictive and one needs to “enlarge” \mathcal{H}_1 . To this aim, it is useful to set

$$\phi_{i+1}(t) = t^i, \quad i = 0, 1, \dots, m-1$$

so as to define an algebraic isomorphism between \mathcal{H}_2 and the m -dimensional space of the unknown initial conditions $\{f^{(i)}(0)\}_{i=0}^{m-1}$. Finally, the spline hypothesis space is defined by the sum of \mathcal{H}_1 and \mathcal{H}_2 (which turns out to be direct in this case).

3 Variational problems subject to linear equality constraints

3.1 Formulation of the problem

Hereafter, \mathcal{H} is the sum of \mathcal{H}_1 and \mathcal{H}_2 , assumed direct, i.e. $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2$. Furthermore, for $f = f_1 + f_2$, with

² Such invertibility hypothesis could be removed by resorting to the concept of generalized inverses in RKHS, see [17].

$f_1 \in \mathcal{H}_1$ and $f_2 \in \mathcal{H}_2$, and $g = g_1 + g_2$, with $g_1 \in \mathcal{H}_1$ and $g_2 \in \mathcal{H}_2$, the inner product is defined as follows

$$\langle f, g \rangle_{\mathcal{H}} = \langle f_1, g_1 \rangle_{\mathcal{H}_1} + \langle f_2, g_2 \rangle_{\mathcal{H}_2}$$

Hence

$$\|f\|_{\mathcal{H}}^2 = \|f_1\|_{\mathcal{H}_1}^2 + \|f_2\|_{\mathcal{H}_2}^2$$

and \mathcal{H}_1 and \mathcal{H}_2 turn out to be perpendicular subspaces. The relations between an unknown function $f \in \mathcal{H}$ and the observable components of a vector $y \in \mathbb{R}^n$ are

$$y_i = f(t_i) + \nu_i, \quad i = 1, 2, \dots, n \quad (4)$$

where $\{t_i\}$ are the ordered time samples (i.e. $0 \leq t_1 < t_2 < \dots < t_n \leq T$) and $\{\nu_i\}$ is white Gaussian noise of variance σ^2 .

Now, we introduce a vector-valued linear operator which will define equality constraints on the function to estimate. In particular, let $C : \mathcal{H} \mapsto \mathbb{R}^l$ where

$$C[f] := [C_1[f] \quad C_2[f] \dots C_l[f]]^T$$

The unknown function is known to satisfy the following condition

$$C[f] = c$$

with $c \in \mathbb{R}^l$. According to Tikhonov regularization theory, the problem of estimating f subject to such constraint can be formulated in terms of the following constrained optimization problem

$$\hat{f} = \arg \min_{f \in \mathcal{H}} \sum_{i=1}^n (y_i - f(t_i))^2 + \gamma \|f\|_{\mathcal{H}}^2 \quad \text{s.t. } C[f] = c \quad (5)$$

where the regularization parameter $\gamma \in \mathbb{R}^+$, which is often unknown, has to correctly balance the relative importance of the values assumed by the norm in \mathcal{H} and by the experimental evidence $\{y_i\}$.

The next proposition points out that the estimation problem (5) is a generalization of the *controlled* two-point boundary value problem of Section 1 of [10]. In particular, a perfect correspondence is obtained by setting

$$C_i[f] = f^{(i-1)}(0) - f^{(i-1)}(T) = 0, \quad \forall f \in \mathcal{H}, \quad i = 1, \dots, l \quad (6)$$

where \mathcal{H} is the spline hypothesis space discussed in Example 1. Its proof can be immediately deduced from definition of \mathcal{H} and by Propositions 2.1 and 2.2. Below, one can now think of y as containing pre-specified output values and the goal is to determine a control function u with low norm amplitude able to drive the system reasonably close to them while satisfying the constraints.

Proposition 3.1 *Let*

$$F[u, w](t) := \int_0^t h(t, \tau) u(\tau) d\tau + \sum_{i=1}^m w_i \phi_i(t)$$

$$F_i[u, w] := F[u, w](t_i)$$

Then, the solution of Problem (5) is $\hat{f} = F[\hat{u}, \hat{w}]$ where

$$\begin{aligned} (\hat{u}, \hat{w}) = \arg \min_{u \in \mathcal{L}^2, w \in \mathbb{R}^m} & \sum_{i=1}^n (y_i - F_i[u, w])^2 + \gamma \int_X u^2(t) dt \\ & + \gamma w^T R w \quad \text{subject to} \quad C[F[u, w]] = c \end{aligned} \quad (7)$$

3.2 Solution of the problem

We derive an algorithm which computes \hat{f} in (5) with $O((n+l)^3)$ operations. In what follows, the notation $C_i[C_j[K(\cdot, \cdot)]]$ means that first the functional C_j is applied to the kernel as a function of one of its two arguments (no ambiguity is present due to the symmetry of the kernel); this leads to a well defined function in \mathcal{H} the functional C_i is applied to.

Proposition 3.2 *Assume that $C[\cdot]$ is linear and bounded in the topology of \mathcal{H} . Define*

$$y^+ = [y^T \quad c^T]^T, \quad K(s, t) = K_1(s, t) + K_2(s, t) \quad (8)$$

Let also δ_{ij} denote the Kronecker delta, $I_1 = [1, \dots, n]$ and $I_2 = [n+1, \dots, n+l]$. Then, \hat{f} in (5) is given by

$$\hat{f}(t) = \sum_{i=1}^n a_i K(t_i, t) + \sum_{i=1}^l a_{i+n} [C_i[K(\cdot, t)]]$$

where $a \in \mathbb{R}^{n+l}$ satisfies the linear system $\Sigma a = y^+$ with

$$\begin{aligned} \Sigma_{ij} &= K(t_i, t_j) + \gamma \delta_{ij} & i \in I_1, j \in I_1 \\ \Sigma_{ij} &= \Sigma_{ji} = C_{j-n}[K(\cdot, t_i)] & i \in I_1, j \in I_2 \\ \Sigma_{ij} &= C_{i-n}[C_{j-n}[K(\cdot, \cdot)]] & i \in I_2, j \in I_2 \end{aligned}$$

Remark 1 *The proof hinges on the representer theorem [20] and can be found in [21]. In view of Proposition 3.1, the result above provides a different but equivalent representation of the solution to the constrained optimal control problem in [10], namely Problem 1. Problem 2 in [10] derives by setting all the entries of R to zero, i.e. $p_i \rightarrow \infty \forall i$ in (2). Under the framework of learning theory [12], this limiting situation corresponds to interpreting H_2 as a bias space so that the projection of f onto it is not assigned any penalty. The solution still belongs to the same subspace reported above and the expansion coefficients still solve a linear system, see e.g. subsection 1.1.2 in [20].*

Example 1 (continued) *The hypothesis space associated with cubic splines is widely used in applications. It is obtained by setting $m = 2$ in (3), thus making f two-fold integration of u . Under these assumptions, one has*

$$K(s, t) = \begin{cases} \frac{s^2}{2} \left(t - \frac{s}{3}\right) + p_1 + p_2 ts & s \leq t \\ \frac{t^2}{2} \left(s - \frac{t}{3}\right) + p_1 + p_2 ts & s > t \end{cases} \quad (9)$$

The periodic cubic smoothing spline problem is then defined by setting $c = 0$ and

$$C[f] := \begin{bmatrix} f(0) - f(T) \\ f^{(1)}(0) - f^{(1)}(T) \end{bmatrix} \quad (10)$$

It is useful to define

$$\dot{K}(s, t) := \frac{\partial K}{\partial s}(s, t) = \begin{cases} st - \frac{s^2}{2} + p_2 t & s \leq t \\ \frac{t^2}{2} + p_2 t & s > t \end{cases} \quad (11)$$

$$\ddot{K}(s, t) := \frac{\partial^2 K}{\partial s \partial t}(s, t) = \min(s, t) + p_2 \quad (12)$$

The vector $C[K(\cdot, t)]^\top$ has the expression

$$\begin{aligned} C[K(\cdot, t)]^\top &= [K(0, t) - K(T, t) \quad \dot{K}(0, t) - \dot{K}(T, t)] \\ &= \begin{bmatrix} \frac{t^2}{2} \left(\frac{t}{3} - T\right) - p_2 T t & -\frac{t^2}{2} \end{bmatrix} \end{aligned} \quad (13)$$

Using Proposition 3.2, the solution \hat{f} of the periodic cubic smoothing spline problem belongs to the subspace spanned by the following $n + 2$ basis functions

$$\varphi_i(t) = \frac{t_i^2}{2} \left(t - \frac{t_i}{3}\right) + p_1 + tp_2 t_i, \quad t \geq t_i \quad (14)$$

$$\varphi_i(t) = \frac{t^2}{2} \left(t_i - \frac{t}{3}\right) + p_1 + tp_2 t_i, \quad t < t_i \quad (15)$$

$$\varphi_{n+1}(t) = \frac{t^2}{2} \left(\frac{t}{3} - T\right) - p_2 T t \quad (16)$$

$$\varphi_{n+2}(t) = -\frac{t^2}{2} \quad (17)$$

where $i = 1, \dots, n$. In addition, one also obtains

$$\Sigma_{ij} = K(t_i, t_j) + \gamma \delta_{ij} \quad i \in I_1, j \in I_1 \quad (18)$$

$$\Sigma_{n+1j} = \Sigma_{jn+1} = K(0, t_j) - K(T, t_j) \quad j \in I_1 \quad (19)$$

$$\Sigma_{n+2j} = \Sigma_{jn+2} = \dot{K}(0, t_j) - \dot{K}(T, t_j) \quad j \in I_1 \quad (20)$$

$$\Sigma_{n+1n+1} = K(0, 0) - 2K(T, 0) + K(T, T) \quad (21)$$

$$\Sigma_{n+2n+2} = \ddot{K}(0, 0) - 2\ddot{K}(T, 0) + \ddot{K}(T, T) \quad (22)$$

$$\begin{aligned} \Sigma_{n+2n+1} &= \Sigma_{n+1n+2} \\ &= \dot{K}(0, 0) - \dot{K}(T, 0) - \dot{K}(0, T) + \dot{K}(T, T) \end{aligned} \quad (23)$$

4 Variational problems admitting state-space descriptions subject to linear equality constraints

4.1 State-space formulation of the problem

As in [10], now we assume that the linear operator $L[\cdot]$ in (1) is time-invariant and finite-dimensional, i.e.

$$\begin{cases} \dot{x}(t) = Fx(t) + Gu(t) \\ f(t) = Hx(t) \end{cases} \quad (24)$$

where $x \in \mathbb{R}^m$ and the system initial condition is denoted by x_0 . From (24), one has

$$f(t) = He^{Ft}x_0 + H \int_0^t e^{F(t-\tau)}Gu(\tau)d\tau$$

Thus, notice that $\{\phi_i\}$ correspond to the ‘‘columns’’ of He^{Ft} , w equals x_0 and $h(t, \tau) = h(t - \tau) = He^{F(t-\tau)}G$. Without loss of generality, we also assume that the system is controllable and observable (considering different descriptions would just increase the computational complexity of the method described below) and that F and G are in control canonical form with $H = [1 \ 0 \dots 0]$. The following result relies upon the duality between RKHS and Gaussian processes. It also exploits the correspondence between Bayesian estimation problems characterized by the fact that the prior for the unknown function admits a state-space representation and fixed-interval Kalman smoothing, see [8,22,13].

Proposition 4.1 *Assume that f is a continuous-time Gaussian process on X independent of $\{\nu_i\}$ with autocovariance $\lambda^2 K(s, t)$. Then, if $\gamma = \sigma^2/\lambda^2$, the solution \hat{f} of (5) is the minimum variance estimate of f conditional on y and the constraint $C[f] = c$. In addition, let*

- *the prior for f admits the state-space description (24)*
- *u be white Gaussian noise of variance λ^2 and independent of $\{\nu_i\}$*
- *$x_0 \sim \mathbf{N}(0, \lambda^2 R^{-1})$ independent of u and $\{\nu_i\}$*

Then, it holds that $\hat{f}(t) = H\mathbb{E}[x(t)|y, C[f]]$, i.e. \hat{f} can be obtained by computing the smoothed estimate of the state vector conditional on y and the constraint $C[Hx] = c$.

4.2 $O(n)$ algorithm via Kalman smoothing

Let the conditions on system (24) stated in Proposition 4.1 hold. Hereafter, $\{x_k\}$ denotes the state sequence

given by $x(t)$ sampled at instants $\{t_k\}$ which satisfies

$$x_{k+1} = e^{F(t_{k+1}-t_k)}x_k + \lambda \int_{t_k}^{t_{k+1}} e^{F(t_{k+1}-t)}Gd\beta(t) \quad (25)$$

$$\mathbb{E}[x_{k+1}|x_k] = e^{F(t_{k+1}-t_k)}x(t_k) \quad (26)$$

$$\mathbb{V}[x_{k+1}|x_k] = \lambda^2 \int_{t_k}^{t_{k+1}} e^{F(t_{k+1}-t)}GG^T e^{F^T(t_{k+1}-t)}dt \quad (27)$$

where $\beta(t)$ is Brownian motion. In addition, in the rest of this subsection we assume that the constraints on the unknown function involve a finite number of state vectors. To be more specific, let $\Omega \subset X$ be a finite set of cardinality q , where $\Omega \subseteq \{t_k\}$ without loss of generality. Define $x_\Omega \in \mathbb{R}^{qm}$ as the column vector obtained by stacking the state vectors at the time instants contained in Ω . Then, it is assumed that $C[f] = c \leftrightarrow Cx_\Omega = c$ where, with some abuse of notation, $C \in \mathbb{R}^{l \times qm}$.

Example 1 (continued) A state-space formulation of the smoothing spline problem is obtained by setting

$$F = \begin{pmatrix} 0 & 1 & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 1 \\ 0 & \dots & 0 & 0 \end{pmatrix} \quad G = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \quad H^T = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (28)$$

Further, by following Section 4 in [23] we have

$$\begin{cases} x_{k+1} = F^k x_k + \omega^k \\ \omega^k \sim \mathbf{N}(0, Q^k) \end{cases} \quad (29)$$

where $\{\omega^k\}$ are mutually independent, F^k is an $m \times m$ lower-triangular Toeplitz matrix whose $(i, 1)$ entry is

$$F^k(i, 1) = \frac{\Delta_k^{i-1}}{(i-1)!}, \quad \Delta_k = t_{k+1} - t_k \quad (30)$$

while Q^k is an $m \times m$ matrix whose (i, l) -entry is given by

$$Q^k(i, l) = \lambda^2 \frac{\Delta_k^{i+l-1}}{(i-1)!(l-1)!(i+l-1)} \quad (31)$$

The periodic smoothing spline problem involves the boundary constraint $x_0 - x_T = 0$, where $x_T := x(T)$. Notice that it is obtained by setting $q = 2$, $\Omega = \{0, T\}$, $x_\Omega = [x_0^T \quad x_T^T]^T$, $c = 0$ and $C = [I_m \quad -I_m]$, with I_m the $m \times m$ identity matrix.

Define $\{\hat{x}_k\} := \mathbb{E}\{x_k|y, C[f]\}$ and $\hat{x}_\Omega := \mathbb{E}[x_\Omega|y, C[f]]$. In view of the Gaussian assumptions, it holds that

$$\{\hat{x}_k\} = \mathbb{E}\{x_k|y, \hat{x}_\Omega\} \quad (32)$$

Eq. (32) points out that, if x_Ω were assumed ‘‘known’’ and set to the estimate \hat{x}_Ω , the solution could be obtained with $O(n)$ operations by a classical discrete-time Kalman smoothing filter including noiseless measurements of some components of the state-sequence. For any t , $\mathbb{E}[x(t)|y, \hat{x}_\Omega]$ could also be obtained by means of a continuous-time filter as e.g. described in [13]. Thus, the overall problem reduces to efficiently computing \hat{x}_Ω . To this aim, using standard results on estimation of jointly Gaussian vectors, see e.g. [14], we obtain

$$\begin{aligned} \hat{x}_\Omega &= \mathbb{E}[x_\Omega|y] + \mathbb{V}(x_\Omega, C[f]|y)\mathbb{V}(C[f]|y)^{-1} \\ &\quad \times (C[f] - \mathbb{E}[C[f]|y]) \\ &= \mathbb{E}[x_\Omega|y] + \mathbb{V}(x_\Omega|y)C^T(C\mathbb{V}(x_\Omega|y)C^T)^{-1} \\ &\quad \times C(x_\Omega - \mathbb{E}[x_\Omega|y]) \end{aligned} \quad (33)$$

In (33), $\mathbb{E}[x_\Omega|y]$ can be obtained by a Kalman smoothing filter while the elements outside the diagonal blocks of $\mathbb{V}(x_\Omega|y)$ are not returned by the standard filter, see e.g. [14,24]. The following result shows that also such quantities can be computed with $O(n)$ operations (see the Appendix for the proof).

Lemma 1 For $0 \leq k \leq k+p \leq q \leq n$, define

$$\check{x}_{k|q} = \mathbb{E}[x_k|\{y_j\}_{j=1}^q] \quad (34)$$

$$\Sigma_{k,k+p|q} = \mathbb{V}(x_k - \check{x}_{k|q}, x_{k+p} - \check{x}_{k+p|q}) \quad (35)$$

$$K_k = \Sigma_{k,k|k}(F^k)^T \Sigma_{k+1,k+1|k}^{-1} \quad (36)$$

Then, one has

$$\mathbb{V}(x_k, x_{k+p}|y) = \Sigma_{k,k+p|n} = \prod_{i=k}^{k+p-1} K_i \Sigma_{k+p,k+p|n} \quad (37)$$

Below we summarize the entire numerical procedure for estimation of stochastic functions which admit state-space descriptions and satisfy the constraints $Cx_\Omega = c$.

Algorithm for computing \hat{f}

The input to this algorithm includes the ordered time instants $\{t_k\}$, the measurements $\{y_k\}$, the state-space model (24), the set Ω , the matrix C and the vector c . The output of this algorithm is the estimate \hat{f} .

- Compute the sampled version of model (24) at $\{t_k\}$ according to (26,27).
- Compute $\mathbb{E}[x_\Omega|y]$ by a Kalman smoothing filter and $\mathbb{V}(x_\Omega|y)$ using (37) for suitable values of k and p which depend on the grid Ω .
- Compute \hat{x}_Ω using (33).
- Compute $\{\hat{x}_k\}$ using (32) while, for any t , $\mathbb{E}[x(t)|y, C[f]]$ is obtained by a continuous-time Kalman smoothing filter with x_Ω assumed known and set to \hat{x}_Ω .
- For any t , set $\hat{f}(t) = H\hat{x}(t)$.

Finally, we consider computation of confidence intervals for the estimate of f at a generic time instant t . To this aim, notice that

$$\mathbb{V}[f(t)|y, C[f]] = H\mathbb{V}[x(t)|y, C[f]]H^T$$

where $\mathbb{V}[x(t)|y, C[f]]$ is

$$\begin{aligned} & \mathbb{V}[x(t)|y] - \mathbb{V}(x(t), C[f]|y)\mathbb{V}(C[f]|y)^{-1}\mathbb{V}(x(t), C[f]|y)^T \\ &= \mathbb{V}[x(t)|y] - \mathbb{V}(x(t), x_\Omega|y)C^T(C\mathbb{V}(x_\Omega|y)C^T)^{-1}C \\ & \times \mathbb{V}(x_\Omega, x(t)|y) \end{aligned}$$

We also notice that $\mathbb{V}(x(t), x_\Omega|y)$ can be obtained with $O(n)$ operations by applying a discrete-time Kalman smoothing filter to the model (24) with “extended” sampling instants $\{t_k^*\} = \{t_k\} \cup t$ and using (37). This thus permits an efficient evaluation of $\mathbb{V}[x(t)|y, C[f]]$, from which any desired confidence interval can be achieved.

4.3 Estimating unknown hyper-parameters

So far, both the prior and measurement noise covariances in the equations above have been assumed known. In real applications, λ^2 (see Proposition 4.1) and σ^2 (see (4)) need to be estimated from data. A possible solution is the so called empirical Bayes approach [25]. First, hyperparameters are estimated via likelihood maximization. Then, to determine f , the maximum likelihood estimates are plugged into the formulas derived in the previous subsections. Letting $\theta = [\lambda^2 \ \sigma^2]$, its estimate is

$$\theta^{ML} = \arg \min_{\theta} J(y^+; \theta), \quad y^+ = [y^T \ c^T]^T$$

$$J(y^+; \theta) := 0.5 \log[\det(\mathbb{V}_\theta[y^+])] + y^{+T} (\mathbb{V}_\theta[y^+])^{-1} y^+$$

where, apart from a constant term, J is equal to the negative of the logarithm of the marginal likelihood $p_\theta(y^+)$ of y^+ , i.e. the total probability of $y, C[f], f$ where f is integrated out. Such an objective function can be efficiently evaluated for any value of θ with $O(n)$ operations. In fact, consider the factorization $p_\theta(y^+) = p_\theta(c|y)p_\theta(y)$. Then, we start noticing that $p_\theta(y)$ is a Gaussian density which can be evaluated for any θ with $O(n)$ operations using a standard Kalman filter as e.g. described in [14] or Section 4 of [23]. The conditional density $p_\theta(c|y)$ is also Gaussian with mean $C\mathbb{E}_\theta[x_\Omega|y]$ and autocovariance $C\mathbb{V}_\theta[x_\Omega|y]C^T$. This latter has to be computed for any θ with $O(n)$ operations as described in Lemma 1. This shows that $J(y^+; \theta)$ may be evaluated for any θ in linear time with respect to output measurements.

4.4 $O(n)$ computation of the equality constrained RN weights

In this subsection, we obtain a RN-type expression for the solution, as in Proposition 3.2, and show that the

RN weights can be computed in linear time.

First of all, let us observe that the matrix Σ in Proposition 3.2 is the autocovariance matrix of the random vector $[y^T \ C[f]^T]^T$ and hence can be partitioned as follows:

$$\Sigma = \begin{bmatrix} \Sigma_{yy} & \Sigma_{yC} \\ \Sigma_{Cy} & \Sigma_{CC} \end{bmatrix}$$

Using Proposition 4.1, it holds that

$$\hat{f}(t) = \mathbb{E}[f(t)|y, C[f]] = \mathbb{E}_{\parallel C[f]}[f(t)|y] + \mathbb{E}_{\parallel y}[f(t)|C[f]] \quad (38)$$

where $\mathbb{E}_{\parallel C[f]}[\cdot|y]$ denotes the oblique projection onto y along $C[f]$ [26]. One has

$$\mathbb{E}_{\parallel C[f]}[f(t)|y] = \mathbb{V}(f(t), y|C[f])\mathbb{V}(y|C[f])^{-1}y \quad (39)$$

$$\mathbb{E}_{\parallel y}[f(t)|C[f]] = \mathbb{V}(f(t), C[f]|y)\mathbb{V}(C[f]|y)^{-1}C[f] \quad (40)$$

We shall show that the oblique projections can be obtained with $O(n)$ operations.

Given three Gaussian vectors a , b and c , the conditional covariance matrix is $\Sigma_{ab|c} := \mathbb{V}(a, b|c) = \Sigma_{ab} - \Sigma_{ac}\Sigma_{cc}^{-1}\Sigma_{cb}$. Let's start considering (39). The term

$$\mathbb{V}(f(t), y|C[f]) = K(t, t_i) - C[K(\cdot, t)]^T \Sigma_{CC}^{-1} C[K(\cdot, t_i)] \quad (41)$$

provides the basis functions for the first projection in the sense that $\mathbb{E}_{\parallel C[f]}[f(t)|y]$ can be written as

$$\mathbb{E}_{\parallel C[f]}[f(t)|y] = \sum_{i=1}^n \psi_i(t) \bar{a}_i \quad (42)$$

where \bar{a}_i is the i -th component of the vector

$$\bar{a} := \mathbb{V}(y|C[f])^{-1}y \quad (43)$$

and $\psi_i(t) := K(t, t_i) - C[K(\cdot, t)]^T \Sigma_{CC}^{-1} C[K(\cdot, t_i)]$. Computation of (41) requires inversion of Σ_{CC} which has complexity $O(l^3)$ while the multiplication have complexity $O(l^2)$. This latter has to be computed n times ($i = 1, \dots, n$) and hence has complexity $O(nl^2)$. We also note that all the basis functions $\psi_i(t)$ satisfy the constraint $C[\psi_i] = 0$. In fact

$$\begin{aligned} C[\psi_i] &= C[K(\cdot, t_i)] - C[C[K(\cdot, \cdot)]^T \Sigma_{CC}^{-1} C[K(\cdot, t_i)]] \\ &= C[K(\cdot, t_i)] - \Sigma_{CC} \Sigma_{CC}^{-1} C[K(\cdot, t_i)] = 0 \end{aligned} \quad (44)$$

This reflects the fact that the oblique projection $\mathbb{E}_{\parallel C[f]}[f(t)|y]$ removes the component which lies in $\text{span}\{C[f]\}$, and, in fact, $C[\mathbb{E}_{\parallel C[f]}[f(t)|y]] = \mathbb{E}_{\parallel C[f]}[C[f]|y] = 0$. Now, consider the term

$$\Sigma_{yy|C}^{-1} := \mathbb{V}(y|C[f])^{-1} = (\Sigma_{yy} - \Sigma_{yC}\Sigma_{CC}^{-1}\Sigma_{Cy})^{-1}$$

and let $LL^\top = \Sigma_{yy}$ be the Cholesky factorization of Σ_{yy} (which, actually, does not need to be computed as shown later). Define $\alpha := L^{-1}\Sigma_{yC}$. Using the matrix inversion lemma [27] it follows that

$$\Sigma_{yy|C}^{-1} = L^{-\top} \left(I - \alpha (\Sigma_{CC} - \alpha^\top \alpha)^{-1} \alpha^\top \right) L^{-1}$$

Let us now define $\bar{y} := L^{-1}y$ so that we have

$$\bar{a} = \Sigma_{yy|C}^{-1}y = L^{-\top} \left[\bar{y} - \alpha (\Sigma_{CC} - \alpha^\top \alpha)^{-1} \alpha^\top \bar{y} \right]$$

Computing $\alpha^\top \alpha$ has complexity $O(nl^2)$ and the matrix inversion $(\Sigma_{CC} - \alpha^\top \alpha)^{-1}$ requires $O(l^3)$ operations. Thus, the vector

$$\bar{\bar{y}} := \left[\bar{y} - \alpha (\Sigma_{CC} - \alpha^\top \alpha)^{-1} \alpha^\top \bar{y} \right] \quad (45)$$

can be computed, given α and \bar{y} , with $O(nl^2)$ computations. It now remains to compute $\bar{a} = L^{-\top} \bar{\bar{y}}$. To this purpose we borrow the algorithm described in [8], based on the observation that L^{-1} is the matrix representation of the whitening filter of the vector y , whose covariance is Σ_{yy} . It thus follows that $\bar{y} := L^{-1}y$, $\alpha := L^{-1}\Sigma_{yC}$ and $\bar{a} := L^{-\top} \bar{\bar{y}}$ ($\bar{\bar{y}}$ in (45)) can be computed by running, respectively, the forward iteration in equations (11), (12) and (13) and the backward iteration in equations (14), (15) and (16) of Theorem 1 in [8] (see also Lemma 4 and 5 in [8]). This requires $O(nl)$ computations ($O(n)$ for each iteration of the filters, the matrix Σ_{yC} has l columns and hence l iterations have to be performed). The parameters used in these recursions are computed as in Lemma 4 of [8] where F_k , G_k and C (see eq. (10) in [8]), are related to the quantities in this paper by, respectively, $F_k := e^{F(t_{k+1}-t_k)}$ (see eq. (25)), $G^k = [Q^k]^{1/2}$ (see eq. (27)) and $C := H$ (see eq. (24)); $\mathbb{V}\{\epsilon\} = \sigma^2$ and the initial covariance matrix is $P = \mathbb{V}(x_0) = \lambda^2 R^{-1}$. This implies that, α , \bar{y} , $\bar{\bar{y}}$, and hence, tracing back the calculations above $\bar{a} = \Sigma_{yy|C}^{-1}y = L^{-\top} \bar{\bar{y}}$, can be computed with $O(nl^2)$ operations. As for the computation of the second oblique projection in (40), see [21].

Example 1 (continued) *Let us now go back to the cubic smoothing spline example, obtained by setting $m = 2$ in (3). Recall that the constraint $C[f] = 0$ takes the form*

$$C[f] = \begin{bmatrix} f(0) - f(T) \\ f^{(1)}(0) - f^{(1)}(T) \end{bmatrix} = 0$$

We have already derived the expression for the elements of the matrix Σ and, in particular, by exploiting (21,22,23), we have

$$\Sigma_{CC} = \begin{bmatrix} \frac{T^3}{3} & \frac{T^2}{2} \\ \frac{T^2}{2} & T \end{bmatrix}, \quad \Sigma_{CC}^{-1} = \begin{bmatrix} \frac{12}{T^3} & -\frac{6}{T^2} \\ -\frac{6}{T^2} & \frac{4}{T} \end{bmatrix}.$$

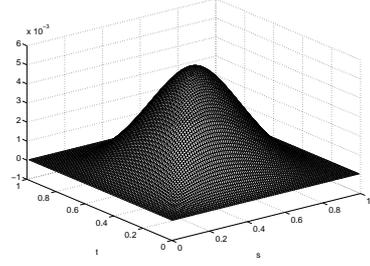


Fig. 1. Periodic cubic smoothing spline kernel ($p_1 = p_2 = 0$).

From the expression of $C[K(\cdot, t)]^\top$ in (13), we obtain that $\psi_i(t)$ (the i -th of the n basis functions in (42)) is

$$\begin{aligned} \psi_i(t) = & \varphi_i(t) - \frac{12}{T^3} \varphi_{n+1}(t) \varphi_{n+1}(t_i) - \frac{4}{T} \varphi_{n+2}(t) \varphi_{n+2}(t_i) \\ & + \frac{6}{T^2} [\varphi_{n+1}(t) \varphi_{n+2}(t_i) + \varphi_{n+1}(t_i) \varphi_{n+2}(t)] \end{aligned} \quad (46)$$

where $\{\varphi_i\}_{i=1}^{n+2}$ are defined in (14). From (42) one obtains that (46) also defines the autocovariance of the cubic splines conditional on $C[f]$, i.e. $\mathbb{V}[f|C[f]]$. It is displayed in Fig. 1 setting $p_1 = p_2 = 0$.

Finally, it is worth stressing that the procedure described in this Section can also be employed, with a few modifications which do not alter the complexity, for computing the conditional variance $\mathbb{V}(f(t)|y, C[f]) = \mathbb{V}(f(t)) - a^\top \Sigma a$, with a and Σ defined as in Proposition 3.2. Last, also the marginal likelihood for estimation of the regularization parameters (see subsection 4.3) can be computed with $O(n)$ operations, see [21].

5 Learning closed curves using a mobile agent moving in an unknown environment

We consider the problem of learning the trajectory of a mobile agent moving along closed curves in an unknown environment. In particular, at the beginning of the experiment the robot is placed at a crossing. Subsequently it moves in order to get information about the environment. The agent will come back to the crossing two times at unknown temporal instants, then proceeding towards other directions to gain new information.

Let $f_1(t)$ and $f_2(t)$ denote the Cartesian coordinates of the robot, as a function of time t . Let

$$f_1(t) = \begin{cases} \cos(t) & 0 \leq t \leq 2\pi \\ 1.5 - 0.5 \cos(t) & 2\pi < t < 4\pi \\ 1 & 4\pi \leq t \leq 4.2\pi \end{cases}$$

$$f_2(t) = \begin{cases} 0.5 \sin(t) & 0 \leq t \leq 2\pi \\ 0.25 \sin(t) & 2\pi < t < 4\pi \\ 0.5 \sin(t) & 4\pi \leq t \leq 4.2\pi \end{cases}$$

The true trajectory is displayed in Fig. 2 (thick line). The starting point is $(1, 0)$ and the agent initially moves in counterclockwise direction. Then, it goes back to the starting point and follows another circle in clockwise direction. It comes back to the initial point again and then follows a vertical path.

The robot path has to be reconstructed from $n = 1300$ noisy measurements of f_1 and f_2 collected at sampling instants $\{t_k\}$ obtained by drawing (and ordering) realizations from a uniform distribution on $[0, 4.2\pi]$. Let y_1 and y_2 be the measurements vectors. Then, it holds that

$$y_{ik} = f_i(t_k) + \nu_{ik} \quad k = 1, \dots, 1300 \quad i = 1, 2$$

where $\{\nu_{1k}\}$ and $\{\nu_{2k}\}$ are mutually independent white Gaussian noises with standard deviation equal to 0.1. Data are displayed in Fig. 2 using points, circles and x-marks for measurements collected at sampling instants in the interval $[0, 2\pi]$, $(2\pi, 4\pi)$ and $[4\pi, 4.2\pi]$, respectively. They are abundant but somewhat noisy.

We model f_1 and f_2 as two-fold integration of two independent white noises ($m = 2$ in (28)) of intensity λ_1^2 and λ_2^2 under periodicity constraints. Thus, for s and t in the interval $[0, 4.2\pi]$ and $i = 1, 2$, our model is defined by

$$\begin{aligned} \mathbb{V}(f_i(s), f_i(t)) &= \begin{cases} \lambda_i^2 \frac{s^2}{2} (t - \frac{s}{3}) + p_1 + p_2 ts & s \leq t \\ \lambda_i^2 \frac{t^2}{2} (s - \frac{t}{3}) + p_1 + p_2 ts & s > t \end{cases} \\ \mathbb{V}(f_1(s), f_2(t)) &= 0, \quad f_i(0) = f_i(T_1) = f_i(T_2), i = 1, 2 \end{aligned}$$

where p_1 and p_2 specify poorly informative priors on the initial conditions. Notice that T_1 and T_2 represent the unknown instants at which the robot goes back to the starting point. Hence, the unknown hyperparameter vector is $\theta = [T_1, T_2, \lambda_1^2, \lambda_2^2]$ and is determined via marginal likelihood maximization. In view of the independence assumptions, it holds that

$$\begin{aligned} \theta^{ML} &= \arg \min_{\theta} J(y_1^+, y_2^+; \theta), \quad y_i^+ = [y_i^T \quad 0]^T, i = 1, 2 \\ J(\{y_i^+\}; \theta) &:= \sum_{i=1}^2 \frac{\log[\det(\mathbb{V}_{\theta}[y_i^+])] + y_i^{+T} (\mathbb{V}_{\theta}[y_i^+])^{-1} y_i^+}{2} \end{aligned}$$

After obtaining the hyperparameters, the equality constrained RN-type estimate is computed (see subsection 4.4). It is shown in Fig. 2 (continuous line). One can notice the effectiveness of the maximum likelihood estimator of the regularization parameters, whose value is known to have a major effect on the quality of the final estimate. As a matter of fact, \hat{f} is close to truth.

6 Conclusions

We have faced the problem of estimating functions subject to linear equality constraints by exploiting regularization and RKHS theory. The class of problems here introduced contains the controlled two-point boundary

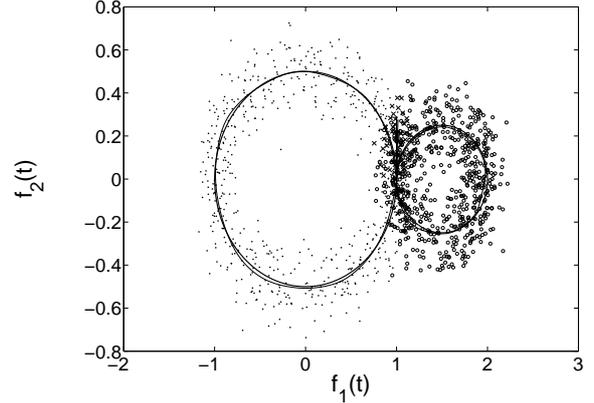


Fig. 2. Robot's path reconstruction. True path f (thick line), noisy measurements (points, circles and x-marks) and estimate of f using periodic cubic smoothing splines (thin line)

value problem described in [10] as a special case. Such connection has been exploited to obtain an algorithm which returns the solution with the same $O((n+l)^3)$ complexity of that reported in [10]. In addition, when the problem admits a state-space representation, as in [10], an $O(n)$ algorithm has been derived.

It is worth stressing that an efficient computational scheme could likely be obtained also by interpreting smoothing as an equality constrained optimization problem. In particular, this could be achieved by exploiting sparsity of the Hessian of the resulting objective, similarly to what is done in the unconstrained scenario in [28,9]. The merit of the present paper is to solve the problem by resorting to geometric concepts, which also permit to easily obtain efficient formulas for computation of confidence intervals and unknown hyperparameters. Finally, we have shown that the solution of the problem admits the structure of a particular RN whose weights are computable in linear time. Results have been specialized to obtain closed form expressions of the basis functions relative to the periodic cubic smoothing spline problem. Simulations illustrate benefits of the new algorithms.

Acknowledgements

This research was partially supported by the PRIN Project entitled 'Tecniche ed applicazioni innovative di identificazione e controllo adattativo'.

Appendix

Proof of Lemma 1

We start noticing that all the quantities displayed in (34,35,36) can be computed with $O(n)$ operations via a Kalman smoothing filter. Then, recall that $\check{x}_{k+p|n} \in \text{span} \langle y_1, \dots, y_n \rangle$ and that the smoothing equation is

$$\check{x}_{k|n} = \check{x}_{k|k} + K_k(\check{x}_{k+1|n} - \check{x}_{k+1|k})$$

(see e.g. Section 2.C in [29]). Then, it holds that

$$\begin{aligned} & \mathbb{V}(x_k - \check{x}_{k|n}, x_{k+p} - \check{x}_{k+p|n}) = \mathbb{V}(x_k - \check{x}_{k|n}, x_{k+p}) \\ & = \mathbb{V}(x_k - \check{x}_{k|k} - K_k(\check{x}_{k+1|n} - \check{x}_{k+1|k}), x_{k+p}) \end{aligned} \quad (47)$$

Adding and subtracting the quantity $K_k x_{k+1}$, we obtain that (47) equals

$$\begin{aligned} & \mathbb{V}(x_k - \check{x}_{k|k} + K_k(\check{x}_{k+1|k} - x_{k+1}), x_{k+p}) \\ & + \mathbb{V}(K_k(x_{k+1} - \check{x}_{k+1|n}), x_{k+p}) \end{aligned} \quad (48)$$

For what concerns the first term in (48), consider state vectors x_k and x_{k+1} conditional on $\{y_1, y_2, \dots, y_k\}$ and notice that

$$K_k = \mathbb{V}(x_k, x_{k+1} | \{y_j\}_{j=1}^k) \mathbb{V}(x_{k+1} | \{y_j\}_{j=1}^k)^{-1}$$

Then, still using known results on estimation of jointly Gaussian vectors, see e.g. [14], it emerges that $z := \check{x}_{k|k} + K_k(x_{k+1} - \check{x}_{k+1|k})$ is the projection of x_k onto $\{y_1, y_2, \dots, y_k, x_{k+1}\}$. Hence, $x_k - z$ is independent of x_{k+1} and so also independent of x_{k+2}, \dots, x_{k+p} . This allows us to conclude that the first term in (48) is equal to zero. As for the second term in (48), we have

$$\begin{aligned} & \mathbb{V}(K_k(x_{k+1} - \check{x}_{k+1|n}), x_{k+p}) \\ & = K_k \mathbb{V}(x_{k+1} - \check{x}_{k+1|n}, x_{k+p} - \check{x}_{k+p|n}) = K_k \Sigma_{k+1, k+p|n} \end{aligned}$$

and this completes the proof. \square

References

- [1] G. Wahba. *Spline models for observational data*. SIAM, Philadelphia, 1990.
- [2] M. Egerstedt and C. Martin. Optimal trajectory planning and smoothing splines. *Automatica*, 37:1057–1064, 2001.
- [3] A. De Luca, L. Lanari, and G. Oriolo. A sensitivity approach to optimal spline robot trajectories. *Automatica*, 27:535–539, 1991.
- [4] S.A. Bortoff. Approximate state-feedback linearization using spline functions. *Automatica*, 33:1449–1458, 1997.
- [5] C. Martin and J. Smith. Approximation, interpolation and sampling. In *Differential Geometry: The Interface between Pure and Applied Mathematics (Contemporary Mathematics Series)*, pages 227–251, Providence, USA, Amer. Math. Soc., 1987.
- [6] C.G. Kaufman, V. Ventura, and R.E. Kass. Spline-based non-parametric regression for periodic functions and its application to directional tuning of neurons. *Statistical medicine*, 24:2255–2265, 2005.
- [7] R. Cogburn and H.T. Davis. Periodic splines and spectral estimation. *The Annals of Statistics*, 2:1108–1126, 1974.
- [8] G. De Nicolao and G. Ferrari Trecate. Regularization networks: fast weight calculation via Kalman filtering. *IEEE Transactions on Neural Networks*, 12:228–235, 2001.
- [9] M.F. Hutchinson and F.R. de Hoog. Smoothing noisy data with spline functions. *Numerische Mathematik*, 47:99–106, 1985.
- [10] H. Kano, M.B. Egerstedt, H. Fujioka, S. Takahashi, and C.F. Martin. Periodic smoothing splines. *Automatica*, 44:185–192, 2008.
- [11] A.N. Tikhonov and V.Y. Arsenin. *Solutions of Ill-Posed Problems*. Washington, D.C.: Winston/Wiley, 1977.
- [12] T. Poggio and F. Girosi. Networks for approximation and learning. *Proceedings of the IEEE*, 78:1481–1497, 1990.
- [13] G.J. Bierman. Fixed interval smoothing with discrete measurements. *Int. J. Control*, 18:65–75, 1973.
- [14] B. D. O. Anderson and J. B. Moore. *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, N.J., USA, 1979.
- [15] N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68:337–404, 1950.
- [16] H.L. Weinert. *Fixed-interval smoothing for state space models*. Kluwer, 2001.
- [17] M.Z. Nashed and G. Wahba. Generalized inverses in reproducing kernel spaces. *SIAM J. Math. Anal.*, 5:974–987, 1974.
- [18] T. Evgeniou, M. Pontil, and T. Poggio. Regularization networks and support vector machines. *Advances in Computational Mathematics*, 13:1–50, 2000.
- [19] R.A. Adams and J. Fournier. *Sobolev Spaces*. Academic Press, 2003.
- [20] G. Wahba. Support vector machines, reproducing kernel Hilbert spaces and randomized GACV. Technical Report 984, Department of Statistics, University of Wisconsin, 1998.
- [21] G. Pillonetto and A. Chiuso. Fast computation of smoothing splines subject to equality constraints. Technical report, University of Padova, 2009. Available at www.dei.unipd.it/~chiuso/DOWNLOAD/Fastpersplines.pdf.
- [22] G. De Nicolao and G. Ferrari Trecate. Regularization networks for inverse problems: a state space approach. *Automatica*, 39:669–676, 2003.
- [23] G. Pillonetto and M.P. Saccomani. Input estimation in nonlinear dynamic systems using differential algebra concepts. *Automatica*, 42:2117–2129, 2006.
- [24] P. Maybeck. *Stochastic Models, Estimation and Control, volume 1*. Academic Press, 1979.
- [25] J. S. Maritz and T. Lwin. *Empirical Bayes Method*. Chapman and Hall, 1989.
- [26] A. Chiuso and G. Picci. On the ill-conditioning of subspace identification with inputs. *Automatica*, 40(4):575–589, 2004.
- [27] A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, New York, 1970.
- [28] B.M. Bell. The marginal likelihood for parameters in a discrete Gauss-Markov process. *IEEE Trans. on Signal Processing*, 48:870–873, 2000.
- [29] G. Kitagawa and W. Gersch. A smoothness priors-time varying AR coefficient modeling of nonstationary covariance time series. *IEEE Transactions on Automatic Control*, 30:48–56, 1985.