

BAYESIAN DECONVOLUTION OF FUNCTIONS IN RKHS USING MCMC TECHNIQUES

Gianluigi Pillonetto

*Department of Information Engineering, University of Padova, Via Gradenigo, 6/B,
Padova, Italy*

giapi@dei.unipd.it

Bradley M. Bell

Applied Physics Laboratory, University of Washington, Seattle, Washington, USA

brad@apl.washington.edu

Abstract We propose a novel stochastic approach to reconstruct the unknown input of a partly known dynamical system from noisy output data. We assume that the unknown function belongs to a Reproducing Kernel Hilbert Space (RKHS). We then design an algorithm based on the Markov chain Monte Carlo (MCMC) framework which is able to recover the minimum variance estimate of the input given the output data.

Keywords: Bayesian regularization; stochastic processes; stochastic simulation

1. Introduction

Deconvolution is the process of reconstructing the input of a dynamical linear system starting from sparse and noisy output data. This problem is important and encountered in many domains of applied science (see e.g. [2, 7]). It is also often difficult to solve since it is subject to ill-posedness and ill-conditioning [2, 14]. Usually, the system designer does not have sufficient information to overcome these difficulties by restricting the unknown function to a finite dimensional model.¹ The most attractive and employed technique to effectively solve the deconvolution problem is instead nonparametric regularization, which does not restrict the unknown function to a particular parameteric form (see e.g. [9, 14]). Many nonparametric deconvolution algorithms select a function from an infinite dimensional Reproducing Kernel Hilbert Space (RKHS)

[1, 5]. The key feature of such spaces is their capability to approximate arbitrarily well a very rich class of functions [8]. The unknown function is then estimated as the solution of a Tychonov-type variational problem [13], containing a quadratic term related to the adherence of experimental data and another one which penalizes unlikely solutions, i.e. functions whose norm amplitude is large. The resulting estimator has an interpretation in stochastic terms [7, 14]. In fact, under certain Gaussian assumptions, it provides the minimum variance estimate of the unknown function given the data.² In real applications, a Tykonov-type estimator has unknown parameters that must be included in the estimation procedure. For example, the regularization parameter, which is a key one since it establishes the right amount of regularization to include in the estimation process, is almost always unknown [12]. Other unknown variables can be present in the linear relationship between the function and the measurements. If we employ a stochastic framework where we model all these additional unknown parameters as random variables, it turns out that the minimum variance estimate of the function requires the evaluation of analytically intractable integrals (see Section 1 in [10]).

In this paper we present a new approach to face this problem together with an efficient algorithm based on a stochastic simulation technique known in literature as Markov chain Monte Carlo (MCMC). Our technique is able to reconstruct a function belonging to a *generic* RKHS together with all the other unknown parameters present in the problem. In contrast to the approach in [10], our computational scheme avoids any kind of discretization of the domain where the function of interest is defined. The paper is organized as follows. In Section 2 we describe the measurement model and recall some properties of RKHSs (that are used in the other sections). In Section 3 we describe our stochastic deconvolution model and the resulting estimation problem. In Section 4 the algorithm which implements the model introduced in Section 3 is illustrated. The performance of the new approach is then tested in Section 5 by one simulated case study. Conclusions are finally offered in Section 6.

2. Preliminaries

2.1 The Measurement Model

For any vector w , we use w_i to refer to the i -th component of w . Moreover, all the vectors are column vectors. We define our problem in mathematical terms. We are given a vector of measurements $y \in \mathbb{R}^n$. The measurement values depend on an unknown function $f : X \rightarrow$

\mathfrak{R} , where X is a compact domain on \mathfrak{R}^q , and on an unknown random vector $\nu \in \mathfrak{R}^n$. The dependence between the function and the $i - th$ measurement is

$$y_i = L_i(f, \theta) + \nu_i \tag{1}$$

where the mapping $f \rightarrow L_i(f, \theta)$ is a linear and continuous functional between a space containing continuous functions and \mathfrak{R} . Moreover, $\theta \in \mathfrak{R}^d$ is an unknown random vector whose probability density function, prior to making the measurements, is $p_\theta(\theta)$. We are also given a model for the statistics of the measurement noise. To be specific, we assume ν is a zero-mean Gaussian random vector whose positive definite covariance matrix is $\Sigma_\nu(\theta)$. We assume that the functions L , p_θ , and Σ_ν are known. In addition, given θ , ν is independent from f . The problem of estimating θ and f without additional information is ill-posed (ill-posed problems are described on page 7 of [13]).

2.2 Reproducing Kernel Hilbert Spaces

Our approach to this ill-posed inverse problem is to place a Bayesian prior on a special type of function space H called Reproducing Kernel Hilbert Space (RKHS). We briefly sketch some properties of these spaces which are relevant in the context of the present work. We use $L^2(X)$ to denote the classical Lebesgue space of square integrable functions on X , equipped with the inner product $\langle \cdot, \cdot \rangle_2$.

DEFINITION 1 *We say that $M : X \times X \rightarrow \mathfrak{R}$ is positive definite if for all finite sets $\{x_1, x_2, \dots, x_k\} \subset X$ the $k \times k$ matrix whose (i, j) entry is $M(x_i, x_j)$ is positive semi-definite. Moreover, we say that M is a Mercer kernel if it is continuous, symmetric and positive definite.*

The following theorem can be obtained by combining the Spectral Theorem for compact operators and Mercer’s theorem (see [4]).

THEOREM 2 *If M is a Mercer Kernel, there exist a sequence $\{\lambda_j \geq 0 : \lambda_{j+1} \geq \lambda_j, j = 1, \dots, \infty\}$ and a basis in $L^2(X)$ of continuous functions $\{\phi_j : j = 1, \dots, \infty\}$ such that*

$$\begin{aligned} \langle \phi_j, \phi_k \rangle_2 &= \begin{cases} 1 & \text{if } j = k \\ 0 & \text{otherwise} \end{cases} \\ \int_X M(s, t) \phi_j(t) dt &= \lambda_j \phi_j(s) \\ M(s, t) &= \sum_{j=1}^{\infty} \lambda_j \phi_j(s) \phi_j(t) \end{aligned}$$

where the above convergence is uniform in $X \times X$.

The following proposition can be derived from [1, 4].

THEOREM 3 *For each Mercer kernel M there exists a unique Hilbert space H such that*

- for each $x \in X$, $M(x, \cdot) \in H$
- the span of the set $\{M(x, \cdot), x \in X\}$ is dense in H
- for each $f \in H$ and $x \in X$, $f(x) = \langle f(\cdot), M(x, \cdot) \rangle_H$

If $\lambda_j > 0$ for every j , the associated RKHS H has the representation

$$H = \left\{ f \in L^2(X) \mid f = \sum_{j=1}^{\infty} a_j \phi_j \quad \text{where} \quad \sum_{j=1}^{\infty} \frac{a_j^2}{\lambda_j} < \infty \right\}$$

equipped with the inner product $\langle \cdot, \cdot \rangle_H$ where, given $f, g \in H$ with $f = \sum_{j=1}^{\infty} a_j \phi_j$ and $g = \sum_{j=1}^{\infty} b_j \phi_j$, we have

$$\langle f, g \rangle_H = \sum_{j=1}^{\infty} \frac{a_j b_j}{\lambda_j}$$

The theorem above enables us to interpret H as a certain subset of smooth functions in $L^2(X)$ ³ generated by the eigenvectors $\{\phi_j\}$. The smoothness condition does in particular concern the behavior of the generalized Fourier coefficients $\{a_k\}$ and is regulated by the eigenvalues of M

We conclude this section by defining the notation Θ to be a subset of \mathfrak{R}^d and $K : \Theta \times X \times X \rightarrow \mathfrak{R}$ to be a parameterized Mercer Kernel; i.e., for each $\theta \in \Theta$, $K(\theta, \cdot, \cdot)$ is a Mercer Kernel. In addition $\phi(\cdot, \theta)$ and $\lambda_i(\theta)$ are the eigen-functions and eigen-values corresponding to $K(\theta, \cdot, \cdot)$.

DEFINITION 4 *We use $L^N(\theta)$ to denote the $n \times N$ matrix whose (i, j) entry is $L_i(\phi_j, \theta)$. In addition, we denote with $\Lambda^N(\theta)$ the $N \times N$ diagonal matrix whose i -th entry of the diagonal is equal to $\lambda_i(\theta)$.*

DEFINITION 5 *Let \mathbf{N} the set of natural numbers. Given $A \subset \mathbf{N}$, we define the following notation:*

$$K_{-A}(\theta, x, y) = K(\theta, x, y) - \sum_{j \in A} \lambda_j(\theta) \phi_j(x, \theta) \phi_j(y, \theta)$$

2.3 An Example of RKHS

In this section we review some properties of an example RKHS space parameterized by the integer $m > 0$. We define the Green's function G_m and the reproducing kernel K_m on $[0, T] \times [0, T]$, $T \in \mathfrak{R}$ as

$$G_m(x, y) = \begin{cases} 0 & \text{if } x \leq y \\ 1 & \text{if } x > y \text{ and } m = 1 \\ (x - y)^{m-1}/(m - 1)! & \text{otherwise} \end{cases}$$

$$K_m(x, y) = \int_0^T G_m(x, \tau)G_m(y, \tau)d\tau$$

Given a function $f : [0, T] \rightarrow \mathfrak{R}$, we use $f^{(i)}$ to denote the i -th derivative of f . The RKHS associated to K_m is then

$$W_m = \left\{ f : [0, T] \rightarrow \mathfrak{R} \left| \begin{array}{l} f^{(m)} \in L^2[0, T] \\ \text{and for } j = 0, \dots, m - 1, \quad f^{(j)}(0) = 0 \\ \text{and } f^{(j)} \text{ is absolutely continuous} \end{array} \right. \right\}$$

equipped with the inner product

$$\langle f, g \rangle_{W_m} = \langle f^{(m)}, g^{(m)} \rangle_2$$

For the special case $m = 1$, the following closed forms for ϕ_j and λ_j are available (see [15])

$$\begin{aligned} \lambda_{W_1, j} &= T^2 / [(j - 1)\pi + \pi/2]^2 \\ \phi_{W_1, j}(x) &= \sqrt{2/T} \sin [(x/T)(j\pi - \pi/2)] \end{aligned} \tag{2}$$

3. Statement of the Estimation Problem

We now define our estimation problem in a Bayesian framework. We start by defining a Bayesian prior for the unknown function f on the RKHS corresponding to $K(\theta, \cdot, \cdot)$.

Assumption: Given θ , the function f is a random field of the form $f(x) = \sum_{j=1}^{\infty} a_j \phi_j(x)$, where $\{a_i\}$ are Gaussian and independent random variables and the variance of a_i is $\lambda_i(\theta)$. In addition, each a_i is independent of the measurement noise ν .

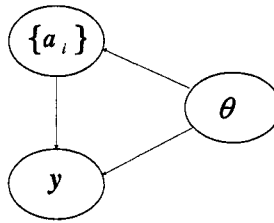


Figure 1. Bayesian network describing the stochastic deconvolution problem

A graphical description of the joint probability density function of y , f and θ is given by the Bayesian network in Figure 1. It is important to note that the node θ is connected via a direct link with both f and y . We also note that given the other two nodes in the network, y depends on f and θ through $\Sigma_\nu(\theta)$ and $L(f, \theta)$. In addition, the probability density function of f given θ depends on θ through the sequence $\{\lambda_i(\theta)\}$. *i.e.*, the smoothness parameters contained in θ parameterize the norm in the RKHS H .

THEOREM 6 *The minimum variance estimate of the function f , given θ and y , is*

$$\hat{f} = \arg \min_{f \in H} [y - L(f, \theta)]^T \Sigma_\nu^{-1}(\theta) [y - L(f, \theta)] + \|f, \theta\|_H^2 \quad (3)$$

where H is the RKHS corresponding to $K(\theta, \cdot, \cdot)$.
(see Appendix for the proof)

Equation (3) shows that if f is the only unknown in the model depicted in Figure 1, its optimal estimate is provided by a Tikhonov-type variational problem. The solution of such problem is linear in the data y and admits a closed form which is well known in literature (see e.g. [14]). However, in real applications θ is seldom completely known. The estimation problem we aim to solve when θ is uncertain is described below.

Problem: Let $p(y|a_i, \theta)$ and $p(a_i|\theta)$ the probability density functions of y given (a_i, θ) and of a_i given θ , respectively. Let also $p(y)$ the marginal probability density function of y . Given the Bayesian network of Figure 1 and known the data y , determine the minimum variance estimate of $f(x)$, *i.e.* compute $\hat{f}(x) = \sum_{i=1}^{\infty} \hat{a}_i \phi_i(x)$ where

$$\hat{a}_i = \frac{\int_{\mathbb{R}^{d+1}} a_i p(y|a_i, \theta) p(a_i|\theta) p(\theta) da_i d\theta}{p(y)}$$

The function \hat{f} takes into account all the possible sources of uncertainty present in the problem and represents our ideal estimate of the random field f . However, its determination turns out difficult since the computation of \hat{a}_i will in general require the solution of an analytically intractable integral. We describe the strategies developed in order to circumvent these problems in the next Section.

4. MCMC Deconvolution Algorithms in RKHS

We solve our stochastic deconvolution problem by reducing it to the reconstruction (in sampled form) of two finite-dimensional probability density functions. The numerical procedure relies on the MCMC framework (for an overview on MCMC theory see e.g. [6]). To simplify our notation below, dependence of some operators on θ is implicit.

4.1 Step 1: Reconstruction of $p(\theta|y)$

The first goal is to reconstruct the probability density function of θ given y after integrating out the unknown random field f from the probabilistic model of Figure 1. For this aim, the following proposition is useful. It can be proved by employing the linearity of the operator L and the fact that ν is independent from f .

THEOREM 7 *We have*

$$p(y|\theta) = \frac{1}{[\det(2\pi\Sigma_y)]^{0.5}} \exp\left(-\frac{1}{2}y^T \Sigma_y^{-1}y\right)$$

where Σ_y is an $n \times n$ matrix, such that

$$\Sigma_y(i, j) = L_i[L_j[K(s, t)]] + \Sigma_\nu(i, j)$$

We then have that $p(\theta|y) \propto p(y|\theta)p(\theta)$ and a MCMC strategy can be employed in order to recover in sampled form this marginal posterior. In particular, we firstly obtain an approximated covariance matrix Σ of the random vector θ given y as the inverse of the Hessian of the minus log of $p(y|\theta)p(\theta)$ computed at its mode (in θ). We then resort to a random-walk Metropolis scheme to reconstruct $p(\theta|y)$. In other words, we use a proposal density which consists of a Gaussian distribution centered at the current point of the Markov chain with covariance matrix proportional to Σ .

4.2 Step 2: Determination of the Minimum Variance Estimate of f given y

THEOREM 8 *Let $A = \{x \in \mathbf{N}; x \leq N\}$. Also, let $\Sigma_{\nu,-A}$ the $n \times n$ matrix such that $\Sigma_{\nu,-A}(i, j) = \Sigma_{\nu}(i, j) + L_i L_j [K_{-A}(x, y)]$. Then, by denoting with $p(y|a^N, \theta)$ the probability density function of y given a^N and θ , we have:*

$$p(y|a^N, \theta) = \frac{\exp\left(-\frac{1}{2}(y - L^N a^N)^T (\Sigma_{\nu,-A})^{-1} (y - L^N a^N)\right)}{[\det(2\pi \Sigma_{\nu,-A})]^{0.5}}$$

Proof: The model of measurements can be rewritten as follows

$$y = L^N a^N + \sum_{j=N+1}^{\infty} L[a_j \phi_j(s)] + \nu \doteq L^N a^N + \xi$$

where ξ is zero-mean normal random vector independent from a^N and having covariance matrix equal to $\Sigma_{\nu,-A}$. This completes the proof.

The following result can be obtained using Bayes formula and Theorem 8.

THEOREM 9 *Given y and θ , the random vector a^N is Gaussian having covariance matrix $\hat{\Sigma}_{a^N}$ and mean $\hat{\mu}_{a^N}$ where:*

$$\hat{\Sigma}_{a^N} = [(\Lambda^N)^{-1} + (L^N)^T \Sigma_{\nu,-A}^{-1} (L^N)]^{-1} \tag{4}$$

$$\hat{\mu}_{a^N} = \hat{\Sigma}_{a^N} (L^N)^T \Sigma_{\nu,-A}^{-1} y \tag{5}$$

We remark that Λ^N, L^N and $\Sigma_{\nu,-A}$ may depend on θ . Thus, $\hat{\Sigma}_{a^N}$ can be computed for some of those values of θ located in high probability regions in accordance with the marginal posterior obtained (in sampled form) at step 1 of the proposed algorithm. This analysis obtains crucial information regarding how the a posteriori probability density function of a component a_i differs from its a priori probability density function. The spectrum of many physical transformations L is located at low frequencies. If one also has that the higher i , the lower is the spectral content of ϕ_i , as e.g. in eq. (2), many of the amplitudes in the set $\{a_i\}$ may be insensitive to the output data. This means that from a certain index i the minimum variance estimate of a_i will be close to the mean of the prior, i.e. close to zero. This makes it possible to find a value of N so that, for every t , $\sum_{i=1}^N \hat{a}_i \phi_i(t) \approx \sum_{i=1}^{\infty} \hat{a}_i \phi_i(t)$.

Finally, after determining the number of amplitudes a_i which is worth reconstructing, the marginal posterior of a^N given y can be recovered.

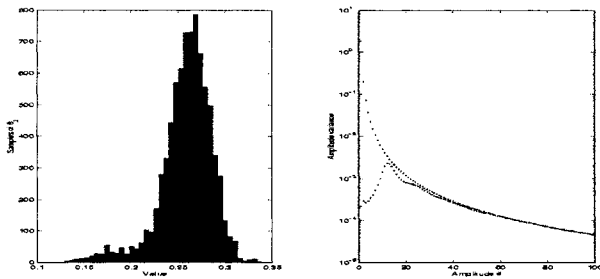


Figure 2. Simulation *Left* A posteriori probability density function of θ_2 obtained in sampled form by MCMC. *Right* A posteriori and a priori variance of a_i as function of i having set θ_1 and θ_2 to their minimum variance estimates

For this aim, let θ_k be a sample drawn from the distribution of θ given y as obtained in step 1. Then, it suffices drawing samples from a Gaussian distribution having mean $\hat{\mu}_{a^N}(\theta_k)$ and covariance matrix $\hat{\Sigma}_{a^N}(\theta_k)$ by using a sufficiently large set of realizations θ_k .

5. Numerical Experiments

We consider a semi-blind deconvolution problem, i.e. a deconvolution problem where the relationship between the unknown input and the output data is only partly known. Let

$$\beta_{pq}(t) = \frac{\Gamma(p + q)}{\Gamma(p)\Gamma(q)} t^{p-1} (1 - t)^{q-1}$$

Then, the simulated function f to reconstruct, taken from [3], is $f(x) = \sum_{j=1}^2 w_j \beta_{p_j, q_j}(x)$ where $0 \leq x \leq 1$ and $w_1 = 0.3, w_2 = 0.6, p_1 = 12, p_2 = 4, q_1 = 7, q_2 = 11$. f is modeled as the unknown input of a shift-invariant linear system with impulse response equal to $\chi([0, \theta_2])$, where $\chi(A)$ is the indicator function of a set A and θ_2 is equal to 0.27, a value drawn from a uniform random variable between 0 and 1. The function has to be reconstructed from 50 output observations, collected by using a uniform sampling grid and corrupted by a white Gaussian process with a constant CV% equal to 10. We model the unknown function as $\sum_{i=1}^{\infty} a_i \phi_{W_1, i}$ where a_i are independent Gaussian random variables of variance $\lambda_i = \theta_1 \lambda_{W_1, i}$ (note that θ_1 represents the regularization parameter). Moreover, we model θ_1 and θ_2 as uniform and independent random variables on $[0, \rho]$, with $\rho \rightarrow +\infty$, and $[0, 1]$, respectively. In Figure 2 (left panel) we report the a posteriori probability density function of θ_2 as reconstructed in sampled form in the first step of our algorithm by a MCMC run where 5500 samples were generated.⁴ The minimum vari-

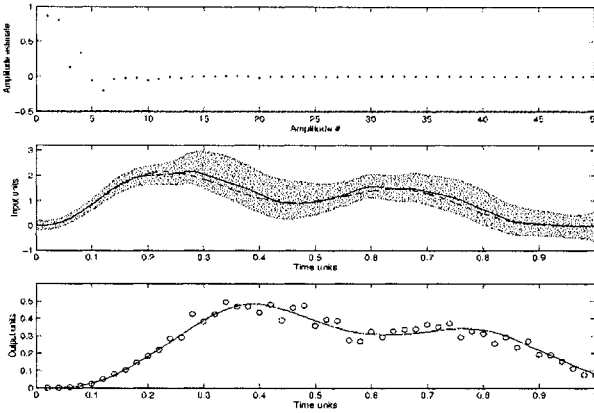


Figure 3. Simulation *Top* Minimum variance estimate of a_i as function of i . *Middle* minimum variance estimate of f (continuous line) with 95% confidence interval (shaded area) and true function (dashed line). *Bottom panel* Reconvolution (continuous line) vs noisy samples (bullets)

ance estimate of this parameter turns out to be 0.258, a value close to truth. Such estimate together with that of θ_1 has been then employed in order to compute the covariance matrix of eq.(4) when N is set to 100. In Figure 2 (right panel) we plot the diagonal elements of such matrix and of the matrix $\Lambda^N(\hat{\theta}_1)$. This plot suggest that only few components a_i are identifiable from the data. The second step of the algorithm has been then performed by setting N to 50 and generating 1500 samples of a^N . In Figure 3 we plot the minimum variance estimates of the components of a^N (top panel) and of f (solid line, middle panel), which appears close to the true function (dashed line, middle panel), together with the 95% confidence interval (shaded area, middle panel). Finally, in the bottom panel of Figure 3 the reconvolution against the noisy samples is depicted.

6. Conclusions

We have proposed a new Bayesian deconvolution algorithm based on the MCMC framework. The technique we have introduced improves on the existing deconvolution algorithms proposed in literature in some important aspects. In particular, differently from the approach developed in [10], our approach is well suited for the reconstruction of a function belonging to a generic RKHS and avoids any kind of discretization of the domain X where the unknown function is defined.

Future developments of this work could consist in extending the esti-

mation technique here presented for reconstructing functions from non-linearly related output data.

Appendix: Proof of Theorem 6

In the sequel, the dependence of Σ_ν, L and $\{\lambda_j\}$ on θ will be implicit. Given a function $f \in H$, where $f = \sum_{j=1}^\infty a_j \phi_j(x)$, let a^N the vector containing the first N components of $\{a_j\}$. Let also $f_a^N(x) = \sum_{j=1}^N a_j^N \phi_j(x)$, where $x \in X$. We define the following prior distribution for a^N

$$\frac{1}{(2\pi)^{N/2} \sqrt{\lambda_1 \cdots \lambda_N}} \exp\left(-\frac{1}{2} \|f_a^N(x)\|_H^2\right)$$

The conditional density for y given a^N and θ is

$$p^N(y|a^N, \theta) = \frac{1}{\sqrt{\det[2\pi\Sigma_\nu]}} \exp\left(-\frac{1}{2}(y - L[f_a^N])^T \Sigma_\nu^{-1} (y - L[f_a^N])\right)$$

The corresponding negative log of the likelihood for a certain $y \in \mathfrak{R}^n$ and a certain a^N is

$$\begin{aligned} l^N(y, a^N|\theta) &= \frac{\|f_a^N\|_H^2}{2} + \frac{1}{2} \sum_{j=1}^N \log(2\pi\lambda_j) + \frac{1}{2} \log \det[2\pi\Sigma_\nu] \\ &+ \frac{1}{2} (y - L[f_a^N])^T \Sigma_\nu^{-1} (y - L[f_a^N]) \end{aligned} \tag{1}$$

We point out that H , being a RKHS, is a subset of the space of continuous functions and convergence in the topology induced by $\|\cdot\|_H$ implies uniform convergence (see [4]). Then, as $N \rightarrow \infty$, the following pointwise convergence holds

$$l^N(y, f_a^N|\theta) - \frac{1}{2} \sum_{j=1}^N \log(2\pi\lambda_j) \rightarrow l(y, f|\theta)$$

where $l(y, f|\theta)$ is defined by

$$l(y, f|\theta) = \frac{1}{2} \|f\|_H^2 + \frac{1}{2} \log \det[2\pi\Sigma_\nu] + \frac{1}{2} (y - L[f])^T \Sigma_\nu^{-1} (y - L[f])$$

Thus, given the model of Figure 1, maximizing $l(y, f|\theta)$ with respect of f corresponds to recovering the maximum a posteriori estimate of f given y and θ . This, combined with the linearity of L and the gaussianity of f , completes the proof.

Notes

1. Otherwise, this would reduce deconvolution to a standard parametric estimation problem, easily solvable by traditional methods as nonlinear least squares
2. There is a strong relationship between RKHSs and Gaussian processes, see e.g. Section 1.4 in [14].
3. all the functions in H are in fact continuous, see Proposition 3 on pag.36 of [4].
4. Results displayed in the sequel have been obtained by assessing the convergence of the generated Markov chains through the binary control of Raftery and Lewis [11]. In particular, we have always required to estimate quantiles 0.025, 0.25, 0.5, 0.75, 0.975 of all the unknown parameters with precision respectively 0.005, 0.01, 0.01, 0.01, 0.005 and with probability 0.95

References

- [1] N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68:337–404, 1950.
- [2] M. Bertero. Linear inverse and ill-posed problems. *Advances in Electronics and Electron Physics*, 75:1–120, 1989.
- [3] P. Craven and G. Wahba. Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische Mathematik*, 31:377–403, 1979.
- [4] F. Cucker and S. Smale. On the mathematical foundations of learning. *Bulletin of the American mathematical society*, 39:1–49, 2001.
- [5] T. Evgeniou, M. Pontil, and T. Poggio. A unified framework for regularization networks and support vector machines. CBCL Paper 171, Massachusetts Institute of Technology, Cambridge, MA, March 1999.
- [6] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter. *Markov chain Monte Carlo in Practice*. London: Chapman and Hall, 1996.
- [7] G. De Nicolao, G. Sparacino, and C. Cobelli. Nonparametric input estimation in physiological systems: problems, methods and case studies. *Automatica*, 33:851–870, 1997.
- [8] G. De Nicolao and G. Ferrari Trecate. Consistent identification of narx models via regularization networks. *IEEE Transactions on Automatic Control*.
- [9] G. De Nicolao and G. Ferrari Trecate. Regularization networks: fast weight calculation via kalman filtering. *IEEE Transactions on Neural Networks*, 12:228–235, 2001.
- [10] P.Magni, R.Bellazzi, and G.De Nicolao. Bayesian function learning using mcmc methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:1219–1331, 1998.
- [11] A.E. Raftery and S.M. Lewis. *Implementing MCMC*, pages 115–130. Markov Chain Monte Carlo in Practice. W.R. Gilks, S.Richardson, and D.J. Spiegelhalter, eds. London: Chapman and Hall, 1996.
- [12] J.A. Rice. Choice of smoothing parameter in deconvolution problems. *Contemp. Math.*, 59:137–151, 1986.
- [13] A.N. Tychonov and V.Y. Arsenin. *Solutions of Ill-Posed Problems*. Washington, D.C.: Winston/Wiley, 1977.
- [14] G. Wahba. *Spline models for observational data*. SIAM, Philadelphia, 1990.
- [15] A.M. Yaglom. *Correlation theory of stationary and related random functions*, volume 1. Springer-Verlag, New York, 1987.