Nonuniform Interpolation of Noisy Time Series Using Support Vector Machines

J.L. Rojo-Álvarez, *Member, IEEE*, C. Figuera,
C.E. Martínez-Cruz, *Student Member, IEEE*, G. Camps-Valls, *Member, IEEE*,
F. Alonso-Atienza, *Student Member, IEEE*, M. Martínez-Ramón, *Senior Member, IEEE*

***Abstract*—The problem of time series interpolation has been intensively studied in the Information Theory literature, in conditions such as bandlimited signals, nonuniform sampling, and presence of noise. During the last decade, Support Vector Machines (SVM) have been widely used for approximation problems, including function and time series interpolation. However, the time series structure has not always been taken into account in SVM interpolation. We propose the statement of two SVM algorithms for time series interpolation, specifically, primal and dual signal model based algorithms. Shift-invariant Mercer's kernels are used as building blocks, according to the requirement of bandlimited signal. The sinc kernel, which has received little attention in the SVM literature, is used for bandlimited reconstruction. Well-known properties of general SVM algorithms (sparseness of the solution, robustness, and regularization) are explored with simulation examples, with improved results with respect standard algorithms, and revealing good characteristics in nonuniform interpolation of noisy signals.**

***Index Terms*— Nonuniform sampling, support vector machine, sinc kernel, Mercer's kernel, primal signal model, dual signal model, time series, interpolation, bandlimited.**

I. Introduction

The interpolation of time series is a widely studied research area [1–4]. The interpolation in the information and communication era has its roots on Sampling Theory, and specifically, on the Whittaker-Shannon-Kotel'nikov (WSK) equation, also known as Shannon's sampling theorem [5,6], which states that a bandlimited, noise free signal can be reconstructed from an uniformly sampled sequence of its values, assumed that the sampling period is properly chosen according to the signal bandwidth. The nonuniform sampling of a bandlimited signal can also be addressed whenever the average sampling period still fulfills Shannon's sampling theorem, and it is used in a number of applications, such as approximation of geophysical potential fields, tomography, and synthetic aperture radar [7–9]. Given that noise can often be present, the reconstruction of a bandlimited and noise corrupted signal from its nonuniformly sampled observations becomes a hard problem. According to [10], two strategies have been mainly followed: (1) consideration of shift-invariant spaces, similar to the case of uniform sampling; and (2) definition of new basis functions (or new spaces) that are better suited to the nonuniform structure of the problem. The first one has been studied the most, following the work developed in the late fifties by Yen [11] using the sinc function as an interpolation kernel. Although, in the theory, Yen's interpolator is optimal in the least squares (LS) sense, ill-posing appears when computing the interpolated values numerically [8]. To overcome this limitation, numerical regularization has been widely used [12]. Alternatively, a number of iterative methods have been proposed, including alternating mapping, projections onto convex sets, and conjugate gradient [7, 13–15]. Other authors have used non-iterative methods, such as filter banks, either to reconstruct the continuous time signal, or to interpolate to uniformly spaced samples [3, 16, 17], but none of these methods is optimal in a LS sense, and many approximate forms of the Yen's interpolator have been developed [18, 19]. The previously mentioned methods have addressed the reconstruction of bandlimited signals, but the question of whether a not bandlimited signal can be recovered from its samples has emerged. Specifically, a finite set of samples from a continuous-time function can be seen as a duration-limited discrete-time signal in practice, and then it cannot be bandlimited. In this case, the reconstruction of a signal from its samples depends on the *a priori* information that we have, and the classical sinc kernel has been replaced by more general kernels that are not necessarily bandlimited [2, 3]. This issue has been mostly studied in problems of uniformly sampled time series.

As a summary, the following main elements are (either implicitly or explicitly) considered by time series interpolation algorithms: the kind of sampling (uniform or nonuniform), the noise (present or not), the spectral content (bandlimited or not), and the use (or not) of numerical regularization. But, in spite of the great amount of work developed to the date, the search for new efficient interpolation procedures is still an active research area.

In this paper, we propose the use of Support Vector Machines (SVM) as a time series interpolator in the presence of noise. SVM were originally stated for classification and regression problems, but its formulation has been extended to a number of digital signal processing problems [20–22]. SVM jointly use the Stuctural Risk Minimization (SRM) Principle and the rather old kernel trick [23]. The SRM Principle states that a better solution (in terms of generalization capabilities) can be found by minimizing an upper bound of the generalization error. This minimization constitutes a Tikhonov regularization [24] that, in

JLRA, CEMC, FAA, and MMR, are with Departamento de Teoría de la Señal y Comunicaciones, Universidad Carlos III de Madrid, Av Universidad 30, 28911 Leganés, Madrid, Spain ({jlrojo,carlosmc,falonso,manel}@tsc.uc3m.es). CF is with Departamento de Teoría de la Señal y Comunicaciones, Universidad Rey Juan Carlos, Camino del Molino s/n, 28943 Fuenlabrada, Madrid (Spain). GCV is with Grup de Processament Digital de Senyals, Departament de Enginyeria Electrònica, Universitat de València, C/ Dr Moliner 50, 46100 Burjassot, València, Spain (gcamps@uv.es).

turn, yields the least possible complexity to the resulting machine. As a result, SVM exhibit less overfitting than machines that are optimized with classical procedures. Also, SVM are robust against outliers and impulse noise, due to its cost function of the residuals [21]. Interestingly, such a procedure produces sparse solutions, which can dramatically reduce the computational burden of the solution in its application stage.

On the one hand, SVM have been previously used for interpolation applications, but key and basic concepts from Information Theory, such as the bandwidth or the kind of sampling, have not been taken into account, so that little connection has been established with the wide existing work in time series interpolation. On the other hand, sparseness and robustness would be extremely useful in the hard problem of interpolation of noisy, possibly nonuniformly sampled, time series. Additionally, the bandlimited nature of the SVM interpolation can be readily controlled by the Mercer kernel that is being used. The Gaussian (or Radial Basis Function, RBF) Mercer's kernel is not a bandlimited function, and hence, it is appropriate for interpolation of not bandlimited time series. Alternatively, it has been proven that the sinc kernel, when adequately expressed, lies in a RKHS, and therefore, it can be used as a Mercer's kernel in SVM interpolation of bandlimited time series [25, 26]. The RBF kernel has been widely studied in the SVM literature, but this is not the case of the sinc kernel, which has received little attention in this setting. As a result, the study and introduction of SVM methods in this context is well motivated and founded.

In particular, we present here two different kinds of SVM interpolation algorithms. The first kind uses a *primal signal model* formulation of the problem, according to the SVM linear framework for digital signal processing presented in [20,22], in which a robust estimation of the model coefficients is indirectly obtained from the SVM Lagrange multipliers. The second kind, or *dual signal model* formulation, uses a nonlinear regression in a RKHS of the time instant corresponding to each observed sample, in such a way that when the solution is expressed in terms of dot products in the RKHS, it can readily be expressed by means of a Mercer's kernel. For the purpose of comparison with precedent Information Theory based formulations, we develop the SVM algorithms in terms of the sinc kernel, the extension of the notation to RBF kernel being straightforward.

The scheme of the paper is as follows. In the next section, a brief introduction to Yen's algorithm, along with some proposed versions, and Munson minimax optimal interpolators, is made. In Section III, the two SVM kinds of algorithms for time series interpolation are presented. Section IV presents simulation results. Finally, in Section V, discussion and conclusions are given.

II. Nonuniform Interpolation

As mentioned in the introduction, a wide variety of methods have been proposed for time series interpolation. Among all the available algorithms for reconstruction of nonuniformly sampled time series, the sinc kernel has received special attention, and accordingly, this implies that a bandlimited nature of the time series is assumed. We limit ourselves to briefly present here the basic's of Yen's algorithm and some improved versions [11, 27].

Let $x(t)$ be a bandlimited, Gaussian noise corrupted signal, and let $\{x_n = x(t_n), n = 1, \ldots, N\}$ be a set of N nonuniformly sampled observations. Given $\{t_n, x_n; n = 1, \ldots, N\}$, the interpolation problem consists of finding an approximating function $x^N(t)$ that fits the data as follows:

$$x(t) = x^N(t) + e(t) = \sum_{i=1}^{N} a_i \text{sinc}(\sigma_0(t - t_i)) + e(t) \quad (1)$$

where $\text{sinc}(t) = \frac{\sin(t)}{t}$, $\sigma_0 = \frac{\pi}{T_0}$ is the bandwidth of the interpolating sinc units, and $e(t)$ represents the noise. The previous continuous time model, after nonuniform sampling, is expressed as the following discrete time model:

$$x_n = x_n^N + e_n = \sum_{i=1}^{N} a_i \text{sinc}(\sigma_0(t_n - t_i)) + e_n \quad (2)$$

An optimal bandlimited interpolation algorithm, in the least squares (LS) sense, was first proposed by Yen [11]. The problem can be expressed as the minimization of the quadratic loss function, given by

$$\frac{1}{2} \sum_{n=1}^{N} \left(x_n - \sum_{i=1}^{N} a_i \text{sinc}(\sigma_0(t_n - t_i)) \right)^2 \quad (3)$$

which, in matrix notation, consists of minimizing

$$\frac{1}{2} \|\boldsymbol{x} - \boldsymbol{S}\boldsymbol{a}\|^2 \quad (4)$$

where $\boldsymbol{a} = [a_1, \ldots, a_N]^T$ is the vector of model coefficients, $\boldsymbol{x} = [x_1, \ldots, x_N]^T$, and $\boldsymbol{S}$ is a squared matrix whose elements are

$$\boldsymbol{S}(n, m) = \text{sinc}(\sigma_0(t_n - t_m)) \quad (5)$$

It is straightforward to see that the solution vector is:

$$\boldsymbol{a} = \boldsymbol{S}^{-1}\boldsymbol{x} \quad (6)$$

Note that we have as many free parameters as observations, which induces an ill-posed problem [12]. In fact, in the presence of even a low level of noise, small perturbations on the coefficient estimations lead to large interpolation errors outside the observed samples. To overcome this limitation, the regularization of the quadratic loss has been proposed, thus leading to a different problem that consists of minimizing

$$\frac{1}{2} \|\boldsymbol{x} - \boldsymbol{S}\boldsymbol{a}\|^2 + \frac{\delta}{2} \|\boldsymbol{a}\|^2 \quad (7)$$

where δ is a regularization parameter, which represents the trade-off between losses and smoothness of the solution. The regularized solution is:

$$\boldsymbol{a} = \left(\boldsymbol{S}^2 + \delta \boldsymbol{I}\right)^{-1} \boldsymbol{S}\boldsymbol{x} \quad (8)$$

where $\boldsymbol{I}$ is the $N \times N$ identity matrix. Note that bandwidth (in both approaches), as well as the trade-off parameter (in the second approach), must be previously fixed.

Many other non-optimal algorithms have been proposed. For instance, the Jacobian weighting [27] uses the following direct interpolation equation:

$$x(t) = x^J(t) + e(t) = \sum_{i=1}^{N} b_i x_i \text{sinc}(\sigma_0(t - t_i)) + e(t) \quad (9)$$

where coefficients b_i are chosen to be the sample spacings, i.e., $b_i = t_{i+1} - t_i$, which corresponds to a Riemann sum approximation to the following integral identity:

$$x(t) = \int_{-\infty}^{+\infty} x(\tau)\text{sinc}(\sigma_0(t - \tau))d\tau \quad (10)$$

if we assume that σ_0 is the true bandwidth of $x(t)$. This algorithm has the drawback of poor performance interpolation, and the advantage of an extremely reduced computational burden.

Another suboptimal (but rather improved) approach was proposed in [8], where a generalization of the sinc kernel interpolator is presented. The model relies on a minimax optimality criterion as an approximate design strategy, and it yields the following expression for the coefficients:

$$a_i = \frac{\pi}{\sigma_0} \left(\sum_{n=1}^{N} \text{sinc}^2(\sigma_0(t_i - t_n)) \right)^{-1} \quad (11)$$

Both the performance and the computational burden of this approach are intermediate between Yen's and Jacobian sinc kernel interpolators. However, all these approaches exhibit some limitations, such as poor performance in low signal-to-noise scenarios, or in the presence of non Gaussian noise, as a direct consequence of the use of quadratic loss function. In addition, these methods result in non sparse solutions. These limitations can be alleviated by accommodating the SVM formulation to the nonuniform sampling problem. In the next section, two versatile SVM-based algorithms for nonuniformly sampled, noisy time series interpolation, are introduced.

Finally, note that instead of using the sinc function as the interpolation kernel for bandlimited interpolation, this formulation can be extended to other non bandlimited basis functions, such as the Gaussian kernel, given by

$$g(t_k - t_n) = \exp\left(-|t_k - t_n|^2/(2\sigma^2)\right) \quad (12)$$

where σ is the kernel free parameter. Also, polynomial functions can be readily used, (see e.g., [28, 29]), which have been subsequently analysed in terms of Information Theory principles. An excellent review can be found in [4].

III. SVM FOR NONUNIFORM INTERPOLATION

In this section, we propose to use several SVM approaches for estimating efficiently coefficients $\{a_i\}$ in signal model (2). In the SVM framework for digital signal processing [21], the optimality criterion is a regularized and constrained version of the regularized LS criterion. Residuals $\{e_n\}$ account for the effect of both noise and model approximation errors. In general, SVM algorithms minimize a regularized cost function of the residuals, usually the Vapnik's ε-insensitivity cost function [23]. Alternatively, we introduce in the formulation the ε-Huber robust cost function [21], which is given by

$$\mathcal{L}(e_n) = \begin{cases} 0, & |e_n| \le \varepsilon \\ \frac{1}{2\gamma}(|e_n| - \varepsilon)^2, & \varepsilon \le |e_n| \le e_C \\ C(|e_n| - \varepsilon) - \frac{1}{2}\gamma C^2, & |e_n| \ge e_C \end{cases} \quad (13)$$

Here, parameter ε is a nonnegative scalar that represents the insensitivity to a low noise level, but more relevant for us, it can provide with a sparse solution, which can be a highly desirable property in the model. Parameters γ and C represent the relevance of the residuals that are in the quadratic and in the linear cost zone, respectively. It can be easily seen that $e_C = \gamma C$ for a residual cost function with continuous first derivative. By an adequate choice of free parameters ε, γ, C, the ε-Huber cost function can be adapted to different kinds of noise while allowing sparse solutions. The function to be minimized by SVM regression consists of a residual cost term plus a regularization term, given by the L_2 norm of the model parameters [23].

For the signal model in (2), there are two possible SVM formulations, which are described next. The first one consists of using signal model (2) as the primal problem in the SVM formulation. The second one consists of considering a generic nonlinear SVM regression, obtaining the SVM dual and solution equations for a generic Mercer's kernel, and finally introducing the sinc kernel as a Mercer's kernel for obtaining signal model (2) in the dual solution.

A. Primal signal model formulation

The nonuniform signal model in (2) can be used within the SVM linear framework [21]. In this case, the signal model to be considered is given in (2). But instead of following the LS criterion, we consider the minimization of ε-Huber robust cost in (13), together with the quadratic norm of model coefficients $\{a_i\}$, which can be seen as a regularization term. As usual in SVM linear framework, the optimization of this regularized robust cost can be achieved [21] by minimizing

$$\frac{1}{2}\sum_{k=1}^{N} a_k^2 + \frac{1}{2\gamma}\sum_{n \in I_1}(\xi_n^2 + \xi_n^{*2}) + C\sum_{n \in I_2}(\xi_n + \xi_n^*) - \sum_{n \in I_2}\frac{\gamma C^2}{2} \quad (14)$$

constrained to

$$x_n - \sum_{i=1}^{N} a_i \text{sinc}(\sigma_0(t_n - t_i)) \le \varepsilon + \xi_n \quad (15)$$

$$-x_n + \sum_{i=1}^{N} a_i \text{sinc}(\sigma_0(t_n - t_i)) \le \varepsilon + \xi_n^* \quad (16)$$

$$\xi_n, \xi_n^* \ge 0 \quad (17)$$

for $n = 1, \ldots, N$, and where ξ_n,ξ_n^*, are slack variables or losses, and I_1 (I_2) are the indices of residuals that are in the quadratic (linear) cost zone. The solution to this optimization problem is given by the saddle point of the following Lagrangian function

$$\begin{aligned}&\frac{1}{2}\sum_{k=1}^{N}a_k^2+\frac{1}{2\gamma}\sum_{n\in I_1}(\xi_n^2+\xi_n^{*2})+C\sum_{n\in I_2}(\xi_n+\xi_n^*)\\&-\sum_{n=1}^{N}\alpha_n\left(-x_n+\sum_{i=1}^{N}a_i\text{sinc}(\sigma_0(t_n-t_i))+\varepsilon+\xi_n\right)-\\&-\sum_{n=1}^{N}\alpha_n^*\left(x_n-\sum_{i=1}^{N}a_i\text{sinc}(\sigma_0(t_n-t_i))+\varepsilon+\xi_n^*\right)-\\&-\sum_{n=1}^{N}(\beta_n\xi_n+\beta_n^*\xi_n^*)-\sum_{n\in I_2}\frac{\gamma C^2}{2}\end{aligned} \tag{18}$$

Lagrangian duality enables the primal problem to be transformed into its dual one, by taking the derivative of (18) with respect to the primal variables. It is easy to show that if we denote

$$\boldsymbol{T}(k,m)=\sum_{n=1}^{N}\text{sinc}(\sigma_0(t_n-t_k))\text{sinc}(\sigma_0(t_n-t_m)) \tag{19}$$

then the dual problem consists of maximizing

$$\begin{aligned}&-\frac{1}{2}(\boldsymbol{\alpha}-\boldsymbol{\alpha}^*)^T\boldsymbol{T}(\boldsymbol{\alpha}-\boldsymbol{\alpha}^*)+(\boldsymbol{\alpha}-\boldsymbol{\alpha}^*)^T\boldsymbol{x}-\\&-\varepsilon\boldsymbol{1}^T(\boldsymbol{\alpha}+\boldsymbol{\alpha}^*)-\frac{\gamma}{2}\left(\boldsymbol{\alpha}^T\boldsymbol{I}\boldsymbol{\alpha}+\boldsymbol{\alpha}^*\boldsymbol{I}\boldsymbol{\alpha}^*\right)\end{aligned} \tag{20}$$

constrained to $0\leq\alpha_n^{(*)}\leq C$, and where $\boldsymbol{\alpha}^{(*)}=[\alpha_1^{(*)},\ldots,\alpha_N^{(*)}]$. This minimization problem can be solved with quadratic programming techniques [30]. Once dual coefficients $\{\alpha_i^{(*)}\}$ are obtained, primal coefficients $\{a_j\}$ are given by

$$a_j=\sum_{i=1}^{N}(\alpha_i-\alpha_i^*)\text{sinc}(\sigma_0(t_j-t_i)) \tag{21}$$

Note that coefficients are proportional to the empirical cross correlation of the Lagrange multipliers and a set of sinc basis functions, each centered in time instant t_n (see [21] for a related discussion on generic primal signal models).

B. Dual signal model formulation

A second SVM-based version of the interpolation function can be obtained by starting with a conventional SVM nonlinear regression [23]. In this setting, given observations $\{x_n\}$ at time instants $\{t_n\}$, we map these time instants to a higher dimensional (H, possibly infinity) feature space $\mathcal{H}$ by using a nonlinear transformation $\boldsymbol{\phi}$, this is, we consider $\boldsymbol{\phi}:\mathbb{R}\rightarrow\mathcal{H}$ that maps $t\in\mathbb{R}\rightarrow\boldsymbol{\phi}(t)\in\mathcal{H}$, where a linear approximation to the data can properly fit the observations as follows

$$x_n=x_n^N+e_n=\langle\boldsymbol{w},\boldsymbol{\phi}(t_n)\rangle+e_n \tag{22}$$

for $n=1,\ldots,N$. The primal problem consists now of minimizing

$$\frac{1}{2}\sum_{k=1}^{H}w_k^2+\frac{1}{2\gamma}\sum_{n\in I_1}(\xi_n^2+\xi_n^{*2})+C\sum_{n\in I_2}(\xi_n+\xi_n^*)-\sum_{n\in I_2}\frac{\gamma C^2}{2} \tag{23}$$

constrained to

$$\begin{aligned}x_n-\langle\boldsymbol{w},\boldsymbol{\phi}(t_n)\rangle&\leq\varepsilon+\xi_n &(24)\\-x_n+\langle\boldsymbol{w},\boldsymbol{\phi}(t_n)\rangle&\leq\varepsilon+\xi_n^* &(25)\\\xi_n,\xi_n^*&\geq0 &(26)\end{aligned}$$

Again, Lagrange functional can be stated by following a similar procedure to the precedent section. In brief, by taking the gradient, we now obtain:

$$\boldsymbol{w}=\sum_{n=1}^{N}(\alpha_n-\alpha_n^*)\boldsymbol{\phi}(t_n) \tag{27}$$

After substitution of $\boldsymbol{w}$ into the Lagrangian and some simple manipulations, the following Gramm matrix can be identified:

$$\boldsymbol{G}(k,m)=\langle\boldsymbol{\phi}(t_k),\boldsymbol{\phi}(t_m)\rangle=K(t_k,t_m) \tag{28}$$

where $K(t_k,t_m)$ is a Mercer's kernel, which allows to obviate the explicit knowledge of nonlinear mapping $\boldsymbol{\phi}(\cdot)$ [23]. The dual problem consists now of maximizing

$$\begin{aligned}&-\frac{1}{2}(\boldsymbol{\alpha}-\boldsymbol{\alpha}^*)^T\boldsymbol{G}(\boldsymbol{\alpha}-\boldsymbol{\alpha}^*)+(\boldsymbol{\alpha}-\boldsymbol{\alpha}^*)^T\boldsymbol{x}-\\&-\varepsilon\boldsymbol{1}^T(\boldsymbol{\alpha}+\boldsymbol{\alpha}^*)-\frac{\gamma}{2}\left(\boldsymbol{\alpha}^T\boldsymbol{I}\boldsymbol{\alpha}+\boldsymbol{\alpha}^*\boldsymbol{I}\boldsymbol{\alpha}^*\right)\end{aligned} \tag{29}$$

constrained to $0\leq\boldsymbol{\alpha}^{(*)}\leq C$. Note the similarity with the dual problem of the primal signal model formulation in (**??**), in which $\boldsymbol{T}$ is replaced with the kernel matrix $\boldsymbol{G}$. The final solution is expressed as

$$x_n^N=\sum_{i=1}^{N}(\alpha_i-\alpha_i^*)K(t_i,t_n) \tag{30}$$

As usual in the SVM framework, by setting $\varepsilon>0$, we obtain that only a subset of the Lagrange multipliers will be nonzero, thus providing with a sparse solution. Moreover, we can define at this point

$$K(t_k,t_n)=\text{sinc}(\sigma_0(t_k-t_n)) \tag{31}$$

as it is possible to show that K is a Mercer's kernel [31]. We will call this choice the *sinc Mercer's kernel*. Therefore, when using the sinc Mercer's kernel, (30) is the nonuniform interpolation model given in (2) with $a_j=\alpha_j-\alpha_j^*$.

Note that other Mercer's kernels could be easily used with this approach. Kernel (12) is a valid Mercer one, so we can define

$$K(t_k,t_n)=\exp\left(-|t_k-t_n|^2/(2\sigma^2)\right) \tag{32}$$

thus giving a non bandlimited Gaussian interpolator.

C. Comparison between primal and dual signal models

In order to qualitatively compare the sinc kernel SVM primal and dual signal models for nonuniform interpolation, note the

	Gaussian Noise					Impulse Noise				
Method	**No noise**	**40dB**	**30dB**	**20dB**	**10dB**	**15dB**	**10dB**	**5dB**	**0dB**	**-5dB**
Y1	47.5±4.1	38.7±1.7	29.6±1.3	19.9±1.2	9.9±1.1	18.9±1.6	17.4±2.3	14.9±3.2	11.4±4.1	7.5±4.8
Y2	53.4±1.9	39.5±1.4	29.8±1.3	20.1±1.2	10.6±1.2	19.2±1.6	17.8±2.2	15.3±3.1	12.1±3.9	8.5±4.3
S1	-0.5±0.7	-0.5±0.7	-0.5±0.7	-0.6±0.8	-1.3±1.0	-0.6±0.8	-0.7±0.8	-0.8±0.9	-1.2±1.2	-2.1±1.8
S2	15.9±3.0	15.9±3.0	15.7±2.9	14.5±2.2	9.8±1.3	14.5±2.2	13.9±2.0	12.6±2.2	10.5±2.8	7.6±3.6
S3	16.9±2.9	16.9±2.9	16.7±2.8	15.4±2.2	10.2±1.3	15.3±2.0	14.7±2.0	13.3±2.3	11.0±3.0	7.9±3.7
SVM-P	49.1±4.3	39.1±1.3	29.9±1.2	20.5±1.1	12.3±1.6	19.8±1.5	18.4±2.0	16.1±2.9	13.3±3.6	11.0±3.9
SVM-D	50.2±3.2	39.5±1.3	29.9±1.2	20.4±1.1	12.4±1.6	19.7±1.5	18.4±2.0	16.0±2.9	13.0±3.5	10.8±3.9
SVM-R	49.8±1.4	39.6±1.1	30.0±1.2	21.4±1.3	13.5±2.0	20.7±1.8	19.2±2.4	16.7±3.2	13.9±3.7	11.8±3.9

TABLE I

S/E RATIOS (MEAN ± STD) FOR GAUSSIAN AND IMPULSE NOISE.

following expansion of the solution for the primal signal model approach

$$x_n = \sum_{i=1}^{N} a_i \text{sinc}(\sigma_0(t_n - t_i)) = \\ = \sum_{i=1}^{N} \left(\sum_{r=1}^{N} (\alpha_r - \alpha_r^*) \text{sinc}(\sigma_0(t_i - t_r) \right) \text{sinc}(\sigma_0(t_n - t_i)) \quad (33)$$

Comparison between (33) and (30) reveals that these are different approaches using SVM for solving a similar signal processing problem. For the primal signal model formulation, and but according to (21), limiting the value of C will prevent these coefficients from an uncontrolled growing (regularization effect). For the dual signal model formulation, the SRM principle implicit in the SVM formalism [23] will lead to a reduced number of nonzero coefficients, thus providing with a desirable sparse solution. Also, in this case, it is easy to see that coefficients are bounded.

IV. SIMULATIONS AND RESULTS

In this section, we evaluate the performance of the three proposed SVM-based signal interpolators, and they are compared with four standard interpolation techniques. Specifically, the following signal interpolators are considered:

1) Yen's interpolator without regularization (Y1).
2) Yen's interpolator with regularization (Y2).
3) Sinc interpolator with uniform weighting (S1).
4) Sinc interpolator with Jacobian weighting (S2).
5) Sinc interpolator with minimax weighting (S3).
6) Primal signal model SVM with sinc kernel (SVM-P).
7) Dual signal model SVM with sinc kernel (SVM-D).
8) Dual signal model SVM with RBF kernel (SVM-R).

Training and test signals. In order to compare the methods, simulations with known solution were conducted. Our experimental setup is adapted from [8], where a set of signals with stochastic bandlimited spectra were generated, but here we used a signal with deterministic bandlimited spectra instead. In particular, the recovery of a bandlimited signal with relatively lower energy on its high frequency components of the spectrum was chosen, aiming to explore the effect that regularization could have on the high-frequency components. The set of signals consisted of the sum of two squared sinc functions, one of them being a lower level, amplitude modulated version of the baseband component, i.e.,

$$x(t) = \text{sinc}^2(\pi t) \left(1 + \frac{1}{2} \sin(2\pi f t) \right) + e(t) \quad (34)$$

where $f = 0.4$ Hz and $e(t)$ is additive noise.

A set of L samples was used with averaged sampling interval T s. The sampling instants were obtained by adding uniform noise, in the range $[-0.1T, 0.1T]$, to equally spaced time points $\{t_k = kT,\ k = 1, 2, \ldots, L\}$. Different values of L were taken, changing accordingly averaged sampling interval T, i.e, when $L = m \times 32$ samples were considered, the averaged sampling interval were changed to $T = 0.5/m$ s, with $m = 1, 2, 3, 4$. Different signal to noise ratios (SNR) were explored (no noise, 40dB, 30dB, 20dB, and 10dB). Sampling intervals falling outside $[0, LT]$ were wrapped inside. A total of 100 realizations were generated for each set of experiments.

The performance of the interpolators was measured by building a test set consisting of a noise-free, uniformly sampled version of the output signal with sampling interval $T/16$, as an approximation to the continuous time signal, and then comparing it with the predicted interpolator estimations at the same time instants. The signal to error (S/E) ratio was computed in decibels as

$$\left(\frac{S}{E} \right)_{dB} = 10 log_{10} \left(\frac{E\{(x_n^N)^2\}}{E\{e_n^2\}} \right) \quad (35)$$

in the train set. Means and standard deviations of S/E were averaged over 100 realizations.

Tuning the free parameters. Four free parameters have to be tuned in SVM algorithms, which are cost function parameters (ε, C, γ), and kernel parameter σ_0 (or equivalently, time duration T_0). These free parameters need to be *a priori* fixed, either by theoretical considerations or by cross-validation search with an additional validation data set. In this paper, and for each developed interpolator, the optimal free parameters were searched according to the reconstruction on the test set. For SVM interpolators, cost function parameters and kernel parameter were optimally adjusted. For the other algorithms, the best kernel width was obtained, and for Y2, the best regularization parameter was determined by using the available test set.

Gaussian noise and SNR. Left part of Table I shows the performance of the algorithms, in the presence of additive, Gaussian noise, as a function of SNR. The poorest performance

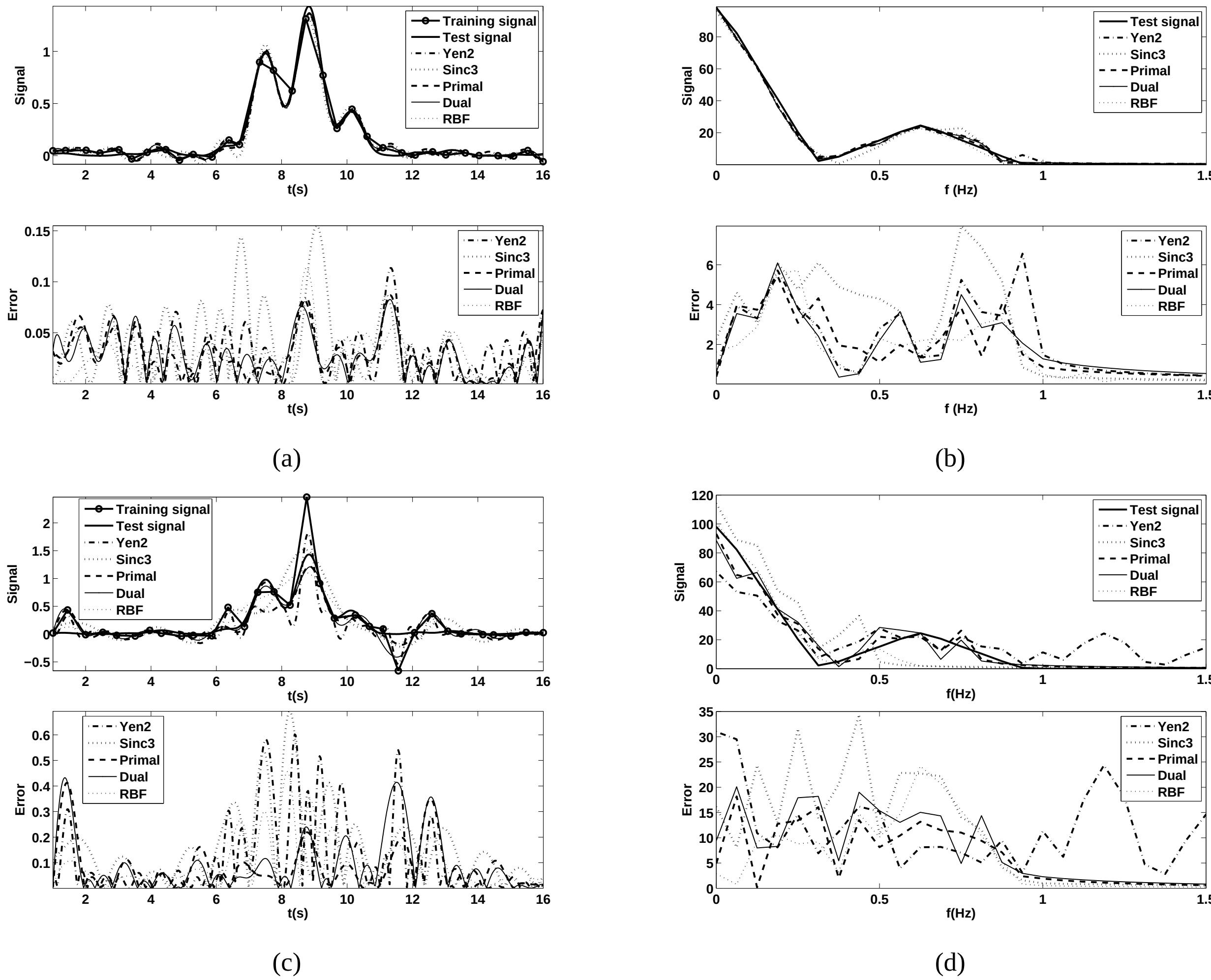


Fig. 1. Examples of interpolation in the time (a,c) and frequency (b,d) domains, for Gaussian noise (SNR = 20dB, L = 32 samples) (a,b) and BG noise (SNR= 20 dB, SIR = 0dB, L = 32 samples) (c,d).

is noticeably exhibited by S1, and some improvement is observed with S2 and S3. Y1 yields a good performance only for low noise levels, whereas Y2 shows a good performance for all noise levels, according to its theoretical optimality (from a Maximum Likelihood point of view) for Gaussian noise. All the SVM approaches remain close to this optimum for high and medium SNR, and even improve around 2dB for very low SNR.

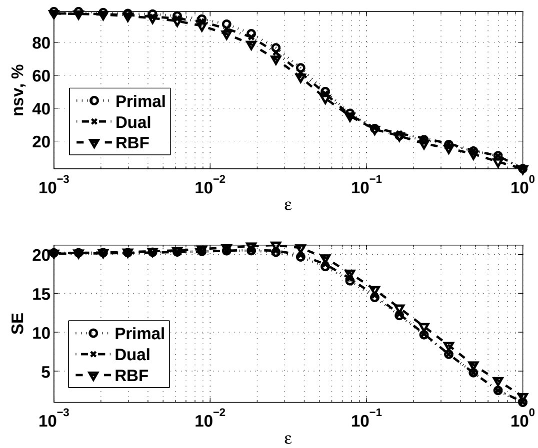
Top panels in Figure 1 show a representative example of a modulated sinc signal with SNR=20 dB, a moderated yet more realistic noise level. The interpolation in the time domain is shown to provide a better approximation to the test signal for Y2 (theoretical optimum) and for the SVM-based methods. Hence, SVM interpolators can yield close-to-optimum performance in the presence of Gaussian noise.

Number of support vectors. Sparseness in SVM-based interpolators was studied as a function of both SNR and signal length, as shown in Tables II and III. All the SVM methods clearly tend to yield more sparse solutions (in average) with decreasing SNR and with increasing signal lengths. For almost all the situations, the most and the least sparse solutions are provided by SVM-R (up to 40% of SV) and SVM-P, respectively. This is an interesting property, which in general is not yielded by conventional Information Theory interpolators or by other previous kernel interpolators.

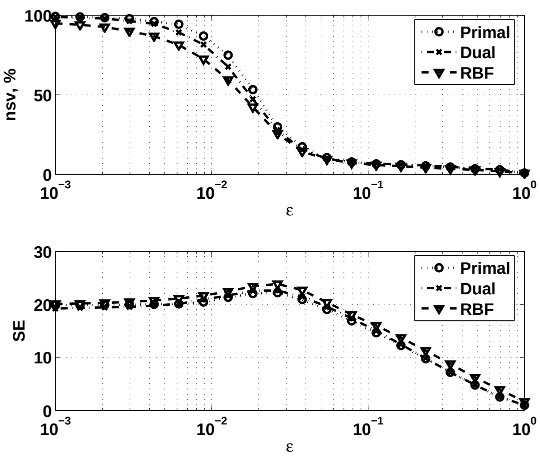
Figure 2 shows the sparseness (number of support vectors, NSV, in %) and the S/E obtained as a function of ε (SNR = 20dB). With low number of samples (L = 32), there is a range of values of ε for which the sparseness can be reduced without significantly modifying the S/E. With increased number of samples (L = 128), there is a clear optimum value of ε for SVM-D and SVM-R (dual formulations), for which a notably sparse solution is obtained. Figure 3 shows two examples (for L = 32 and L = 128) of the obtained coefficients for the optimum ε. Interestingly, for SVM-P dual coefficients the sparseness is lower, but coefficients a_i, which are obtained by means of dual

Method	**No noise**	**40dB**	**30dB**	**20dB**	**10dB**
SVM-P	96.5±10.8	96.0±14.3	99.3±5.0	92.8±18.8	64.1±18.6
SVM-D	93.9±17.8	95.3±14.4	95.9±14.2	89.1±22.2	59.3±17.3
SVM-R	100.0±0.3	99.9±0.7	99.3±2.5	71.8±16.8	49.5±14.9

TABLE II
RATE (%) OF SV (MEAN ± STD) WITH SNR.

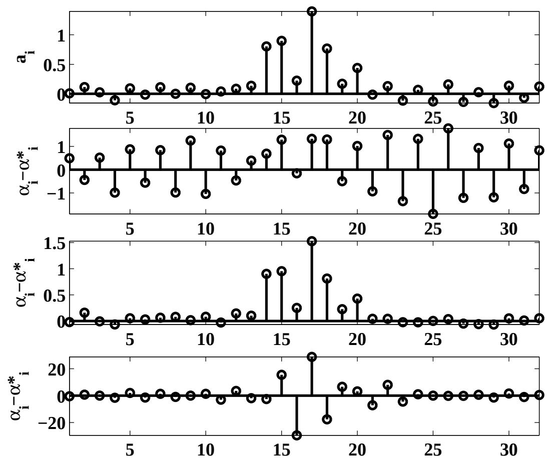


(a)

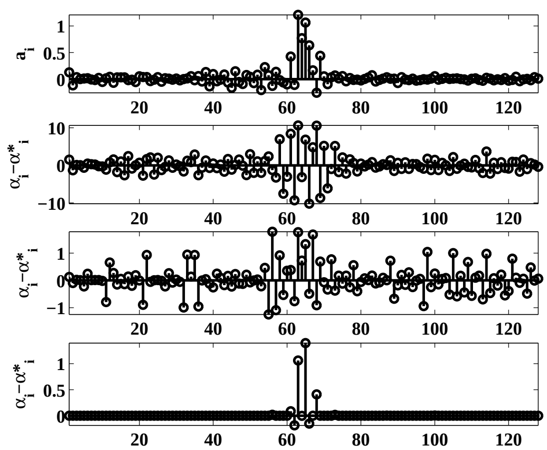


(b)

Fig. 2. Sparseness (upper pannels) and S/E (lower pannels) as a function of ε, for signals with (a) 32 samples, and (b) 128 samples.



(a)



(b)

Fig. 3. (a) Example of sparseness in SVM coefficients for a signal with L = 32 samples. From top to bottom: a_i of SVM-P, $\alpha_i - \alpha_i^*$ of SVM-P, SVM-D and SVM-R. (b) The same for an example with L = 128 samples.

coefficients, trend to be more sparse than dual coefficients *per se*. The number of SV for L = 128 samples is dramatically reduced for SVM-R when compared to SVM-D. The high values of standard deviations was due to the presence of bimodality in the distribution of the number of SV across experiments, specially present in SVM-P and SVM-D (not shown).

Analysis of γ and C for SVM models. A relevant stage when using SVM algorithms is the selection of the free parameters of the cost function. We studied the effect of changing γ and C for $L = 32$ samples, and the results are shown in Fig. 4. In average, values of C in $(10, 1000)$ yield good performance, but higher values produce high variance in the SE, and γ in $(10^{-5}, 10^{-2})$ are appropriate, while lower values may produce numerical problems due to the lack of regularization.

Method	16	32	64	128
SVM-P	97.2±7.8	90.2±21.2	88.6±26.0	76.6±38.8
SVM-D	93.9±12.9	85.1±25.5	81.7±33.0	72.3±40.9
SVM-R	88.7±13.8	69.7±17.5	48.6±12.0	29.6±11.8

TABLE III
RATE (%) OF SV (MEAN ± STD) WITH NUMBER OF SAMPLES.

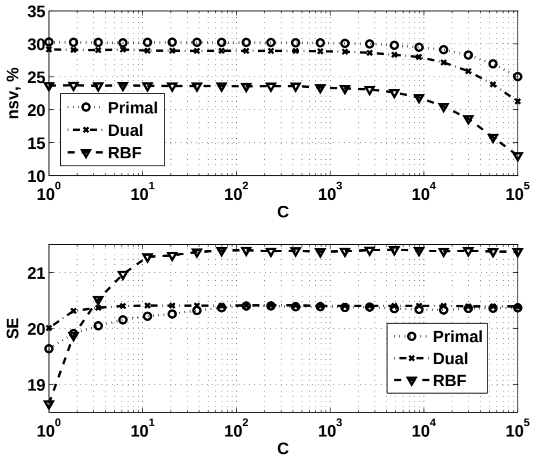
Impulse noise. In order to test the robustness against impulse noise, similar experiments were conducted. Impulse noise was generated with the Bernouilli-Gaussian (BG) function, $n_n^{BG} = v_n \lambda_n$, where v_n is a random process with Gaussian distribution and power σ_{BG}^2, and where λ_n is a random process with probability

$$P_r(\lambda_n) = \begin{cases} p, & \lambda = 1 \\ 1-p, & \lambda = 0 \end{cases} \quad (36)$$

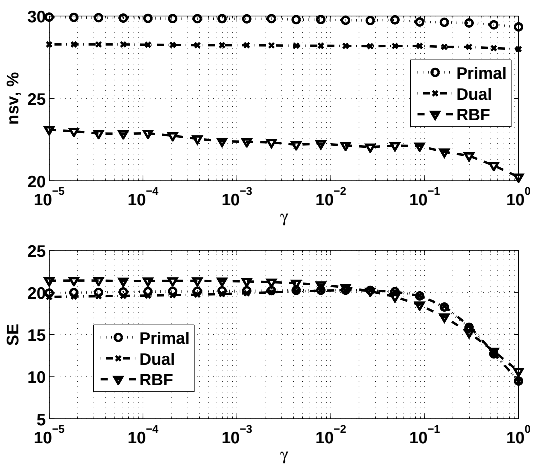
Accordingly, BG noise with $p = 0.1$ was added to the train signals, with different rates of Signal to Impulse Ratio (SIR), defined as

$$SIR_{dB} = 10 \log_{10} \left(\frac{E\{|x_n - n_n^G - n_n^{BG}|^2\}}{\sigma_{BG}^2} \right) \quad (37)$$

where n_n^G is the added Gaussian noise.



(a)



(b)

Fig. 4. Sparseness (up) and S/E (down), for $L = 32$ samples, as a function of C and γ.

Right part of Table I shows the comparison for all the methods in impulse noise. As expected, the SVM algorithms outperform the Y2 algorithm for significantly low SIR values, given that Y2 is no longer the optimum with this noise. Though a fair comparison should take into account M-estimates versions of Y2 algorithm, this was not addressed because these can be considered a particular case of SVM estimates with ε-Huber cost function [21]. Tables IV and V show the sparseness as a function of SIR and of the number of samples. The behavior of SVM algorithms in terms of sparseness in the presence of BG noise was similar than in Gaussian noise, but a slightly reduction in sparseness could be observed in terms of a comparable number of samples. Bottom panel of Figure 1 shows an example of signal reconstruction in BG noise, both in the time and in the frequency domain. Note that an improved reconstruction is obtained by SVM algorithms at the high-amplitude spike locations, but a distorting effect in the band close to the Nyquist frequency of the signal can be observed. On the other hand, Y2 algorithm uses a width such that it often invades the whole spectrum in this example.

V. Conclusions

A new approach to the problem of interpolation of nonuniformly sampled time series has been presented, based on the SVM signal processing framework. Not only the sinc kernel, but also the RBF kernel, popular in the SVM literature, have been shown to be close to Yen's optimal in the presence of Gaussian noise, and robust to low number of available samples, outperforming other precedent approximations. Besides, sparse solutions can be obtained. The robustness of SVM algorithms when impulse noise is present has also been explored, with promising results.

Although generated from the same subjacent signal model of a weighted sum of sincs basis centered at each sample time, the approaches coming from Yen's optimal, and the SVM primal and dual signal model formulations, have different nature, and thus their particular properties can be exploited in different ways. On the one hand, the SVM algorithm obtained from a primal signal model provides with a robust, yet indirect, estimation of the coefficients, and it is closer in its nature to Yen's optimal formulation. It can be shown that primal model SVM solution is Y2 for $\varepsilon = 0$ and $C \rightarrow +\infty$. For finite values of the parameters and for Gaussian noise, the SVM is biased, while Yen's solution is the optimal unbiased solution through the use of the Maximum Likelihood cost function if the sample size is large enough. If the sample size is small, Yen's solution needs to be regularized, thus becoming biased. On the other hand, dual SVM algorithms, arising from dual signal model formulations, are in fact another form of nonlinear SVM-based regression, and hence, the coefficients are directly obtained as the Lagrange multipliers from the SRM principle.

In conclusion, by adequately choosing parameter ε, dual SVM algorithms provide sparse solutions, which can be an attractive property for time series interpolation purposes. The sparseness of SVM-R has been shown to be significantly better than SVM-D. However, which algorithm to use (primal signal model or dual signal model, and RBF or sinc kernel) should be decided according to the application requirements.

Method	**15dB**	**10dB**	**5dB**	**0dB**	**-5dB**
SVM-P	87.2±22.3	79.5±26.0	75.6±25.6	78.9±24.6	85.2±23.7
SVM-D	79.6±27.0	74.5±27.4	67.1±26.9	67.5±27.2	81.3±27.1
SVM-R	67.5±17.1	63.7±17.4	63.2±20.1	71.3±25.3	81.7±26.1

TABLE IV
Rate (%) of SV (mean ± std) with SIR in BG noise.

Method	**16**	**32**	**64**	**128**
SVM-P	92.3±13.1	81.1±24.7	64.1±35.5	44.7±41.1
SVM-D	92.1±14.8	68.7±28.8	55.2±36.8	34.2±37.6
SVM-R	85.7±14.6	63.5±19.7	42.8±16.7	24.0±11.0

TABLE V
Rate (%) of SV (mean ± std) with L in BG noise (SIR = 0dB).

REFERENCES

[1] A. Jerri, "The Shannon sampling theorem-its various extensions and applications: A tutorial review," *Proc IEEE*, vol. 65, no. 11, pp. 1565–96, 1977.

[2] P. Butzer and R. Stens, "Sampling theory for not necessarily band-limited functions: A historical overview," *SIAM Review*, vol. 34, no. 1, pp. 40–53, 1992.

[3] P. Vaidyanathan, "Generalizations of the sampling theorem: Seven decades after Nyquist," *IEEE Trans Circ Syst I*, vol. 48, no. 9, pp. 1094–1109, 2001.

[4] E. Meijering, "A chronology of interpolation: From ancient astronomy to modern signal and image processing," *Proc IEEE*, vol. 90, no. 3, pp. 319–42, 2002.

[5] C. Shannon, "Classic paper: Communication in the presence of noise," *Proc IEEE*, vol. 86, no. 2, pp. 447–57, 1998.

[6] A. García, "Orthogonal sampling formulas: A unified approach," *SIAM Rev.*, vol. 42, no. 3, pp. 499–512, 2000.

[7] T. Strohmer, "Numerical analysis of the non-uniform sampling problem," *J Comput Appl Math*, vol. 122, no. 1-2, pp. 297–316, 2000.

[8] H. Choi and D. Munson, "Direct-Fourier reconstruction in tomography and synthetic aperture radar," *Int J Imag Syst Technol*, vol. 9, no. 1, pp. 1–13, 1998.

[9] J. Jackson, C. Meyer, D. Nishimura, and A. Macovski, "Selection of a convolution function for Fourier inversion using gridding," *IEEE Trans Med Imag*, vol. 10, no. 3, pp. 473–478, 1991.

[10] M. Unser, "Sampling-50 years after Shannon," *Proc IEEE*, vol. 88, no. 4, pp. 569–87, 2000.

[11] J. Yen, "On nonuniform sampling of bandwidth-limited signals," *IRE Trans Circ Theory*, vol. 3, no. 4, pp. 251–57, 1956.

[12] E. J. Diethron and D. Munson Jr, "A linear time-varying system framework for noniterative discrete-time band-limited signal extrapolation," *IEEE Trans Sig Proc*, vol. 39, no. 1, pp. 55–68, 1991.

[13] S. Yeh and H. Stark, "Iterative and one-step reconstruction from nonuniform samples by convex projections," *J Opt Soc Amer*, vol. 7, no. 3, pp. 491–99, 1990.

[14] C. Cenker, H. Feichtinger, and H. Steier, "Fast iterative and non-iterative reconstruction of band-limited functions from irregular sampling values," in *Proc ICASSP'91*, vol. 3, (Toronto, Canada), pp. 1773–76, 1991.

[15] F. Marvasti, M. Analoui, and M. Gamshadzahi, "Recovery of signals from nonuniform samples using iterative methods," *IEEE Trans Sig Proc*, vol. 39, no. 4, pp. 872 – 878, 1991.

[16] M. Unser and J. Zerubia, "A generalized sampling theory without band-limiting constraints," *IEEE Trans Circ Syst II*, vol. 45, no. 8, pp. 959–69, 1998.

[17] Y. Eldar and A. Oppenheim, "Filterbank reconstruction of bandlimited signals from nonuniform and generalized samples," *IEEE Trans Sig Proc*, vol. 47, no. 10, pp. 2768–82, 1999.

[18] G. Kakazu and D. Munson Jr., "A frequency-domain characterization of interpolation from nonuniformly spaced data," in *IEEE Int Symp Circ Syst*, (Portland, Oregon), pp. 288–91, 1989.

[19] G. Calvagno and D. C. Munson Jr., "New results on Yen's approach to interpolation from nonuniformly spaced samples," in *Proc ICASSP'90*, pp. 1535–38, 1990.

[20] J. L. Rojo-Álvarez, M. Martínez-Ramón, A. R. Figueiras-Vidal, M. de-Prado Cumplido, and A. Artés-Rodríguez, "Support vector method for ARMA system identification," *IEEE Trans Sig Proc*, vol. 52, no. 1, pp. 155–64, 2004.

[21] J. L. Rojo-Álvarez, G. Camps-Valls, M. Martínez-Ramón, E. Soria-Olivas, A. Navia Vázquez, and A. R. Figueiras-Vidal, "Support vector machines framework for linear signal processing," *Sig Proc*, vol. 85, no. 12, pp. 2316–26, 2005.

[22] G. Camps-Valls, J. L. Rojo-Álvarez, and M. Martínez-Ramón, *Kernel Methods in Bioengineering, Signal and Image Processing*. Hershey, PA (USA): Idea Group Inc., 2006.

[23] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer–Verlag, 1995.

[24] A. Tikhonov and V. Arsenen, *Solution to Ill-Posed Problems*. Washington DC: V.H. Winston & Sons, 1977.

[25] K. Yao, "Applications of reproducing kernel Hilbert spaces-bandlimited signal models," *Information and Control*, vol. 11, pp. 429–44, 1967.

[26] B. Schölkopf and A. Smola, *Learning with Kernels*. 1 ed.

[27] H. Choi and D. C. Munson Jr., "Analysis and design of minimax-optimal interpolators," *IEEE Trans Sig Proc*, vol. 46, pp. 1571–79, 1998.

[28] M. Wand and M. Jones, *Kernel Smoothing*. London: Chapman & Hall, 1995.

[29] J. Fan and I. Gijbels, *Local Polynomial Modelling and Its Applications*. London: Chapman & Hall, 1995.

[30] J. Platt, *Advances in Kernel Methods: Support Vector Learning (B. Schölkopf, C. J. C. Burgues and A. J. Smola, Eds.)*, ch. Fast Training of Support Vector Machines Using Sequential Minimal Optimization, pp. 185–208. MIT Press, 1999.

[31] L. Zhang, Z. Weida, and L. Jiao, "Wavelet support vector machine," *IEEE Trans Sys Man Cybern B*, vol. 34, pp. 34–39, Feb. 2004.