Sampling and Reconstruction of Signals With Finite Rate of Innovation in the Presence of Noise

Irena Maravić and Martin Vetterli, *Fellow, IEEE*

***Abstract*—Recently, it was shown that it is possible to develop exact sampling schemes for a large class of parametric nonbandlimited signals, namely certain signals of finite rate of innovation. A common feature of such signals is that they have a finite number of degrees of freedom per unit of time and can be reconstructed from a finite number of uniform samples. In order to prove sampling theorems, Vetterli et al. considered the case of deterministic, noiseless signals and developed algebraic methods that lead to perfect reconstruction. However, when noise is present, many of those schemes can become ill-conditioned. In this paper, we revisit the problem of sampling and reconstruction of signals with finite rate of innovation and propose improved, more robust methods that have better numerical conditioning in the presence of noise and yield more accurate reconstruction. We analyze, in detail, a signal made up of a stream of Diracs and develop algorithmic tools that will be used as a basis in all constructions. While some of the techniques have been already encountered in the spectral estimation framework, we further explore preconditioning methods that lead to improved resolution performance in the case when the signal contains closely spaced components. For classes of periodic signals, such as piecewise polynomials and nonuniform splines, we propose novel algebraic approaches that solve the sampling problem in the Laplace domain, after appropriate windowing. Building on the results for periodic signals, we extend our analysis to finite-length signals and develop schemes based on a Gaussian kernel, which avoid the problem of ill-conditioning by proper weighting of the data matrix. Our methods use structured linear systems and robust algorithmic solutions, which we show through simulation results.**

***Index Terms*—Annihilating filters, generalized sampling, nonbandlimited signals, nonuniform splines, piecewise polynomials, rate of innovation, singular value decomposition.**

I. Introduction

SAMPLING theory has undergone a strong research revival over the past decade, mainly motivated by the intense activity taking place around wavelets, which led to an extension of Shannon's original theory and development of advanced formulations with direct relevance to signal processing and communications [23]. For example, a modern Hilbert-space formulation allows the standard sampling paradigm for the representation of bandlimited functions to be extended to a more general class of shift-invariant subspaces, including uniform splines and wavelets [1], [23]. Such a framework resulted in simpler and more realistic interpolation models, many of which have proved to be useful from an implementation point of view [23].

Recently, it was shown that it is possible to develop exact sampling schemes for classes of signals that are neither bandlimited nor live on a shift-invariant subspace, namely, certain signals of finite rate of innovation [24]. A common feature of such signals is that they have a parametric representation with a finite number of degrees of freedom and can be reconstructed from a finite set of uniform samples. Examples include streams of Diracs, nonuniform splines and piecewise polynomials. The key in all constructions is to identify the innovative part of a signal, such as time instants of Diracs, using an annihilating or locator filter, a well-known tool from spectral estimation [19] or error correction coding [2]. This reduces a nonlinear estimation problem into the simpler problem of estimating the parameters of a linear model.

In [24], only a class of deterministic, noiseless signals was considered. While in such a case the developed schemes lead to perfect reconstruction by sampling the signal at (or above) the rate of innovation, many of those methods involve steps that can result in numerical ill-conditioning in the presence of noise. For example, it was shown that the problem of reconstructing nonuniform splines or piecewise polynomials can be reduced to the problem of reconstructing streams of Diracs by taking a sufficient number of signal derivatives. However, when noise is present, such an approach often results in an ill-conditioned problem, where standard techniques from noisy spectral estimation, including oversampling and solving various systems using the singular value decomposition, are not sufficient for improving the numerical performance. This naturally requires a revision of some of the techniques presented in [24] and development of alternative algebraic approaches that can solve the problem of ill-conditioning in the presence of noise and allow for precise reconstruction.

In this paper, we develop improved, more robust methods that make use of proper preconditioning techniques and achieve good numerical performance, while retaining a linear, model-based flavor of the original sampling schemes. We develop a subspace framework for signal reconstruction [16], [17] which, along with efficient noise suppression via singular value decomposition, provides an elegant and robust solution to the sampling problem. We specifically analyze the case of a stream of weighted Diracs, used as a basic building block in all our schemes, and develop algorithmic tools that yield precise estimates of all relevant signal parameters. Furthermore, we develop techniques that can significantly improve resolution ca-

Manuscript received February 5, 2004; revised July 5, 2004. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Zixiang Xiong.

I. Maravić was with the Swiss Federal Institute of Technology, CH-1015 Lausanne, Switzerland. She is now with the Euorpean Molecular Biology Laboratory, D-69117 Hedelberg, Germany (e-mail: irena.maravic@embl.de).

M. Vetterli is with the Electrical Engineering and Computer Science Department, University of California at Berkeley, Berkeley, CA 94720 USA, and also with the Swiss Federal Institute of Technology, CH-1015 Lausanne, Switzerland (e-mail: martin.vetterli@epfl.ch).

Digital Object Identifier 10.1109/TSP.2005.850321

pabilities of the proposed subspace method in the case when the signal contains closely spaced components. For classes of periodic signals, such as piecewise polynomials and nonuniform splines, we propose novel algebraic approaches that use proper data windowing and solve the problem in the Laplace domain. We also consider the case of finite-length signals, and develop improved schemes based on a Gaussian sampling kernel and weighting of the data matrix.

Some of the techniques we use are already encountered in the context of parametric spectral estimation [8], [17], [19] and model-based signal analysis [16]. In particular, we use algebraic methods that reduce a set of samples into a sum of exponentials; therefore, the reconstruction problem can be broadly considered as the one of estimating the parameters of superimposed exponentials. This problem has been studied extensively in the literature and several classes of high-resolution or subspace methods have already been developed [16], [17], [19]. We specifically focus on state space parameterization of a signal subspace [16], which allows us to use high-resolution techniques based on eigendecomposition of certain well-conditioned matrices [8], [16]. A closely related method, the ESPRIT algorithm, was developed in [17], which can be viewed as an extension of the state space approach applicable to antenna arrays. In [8], several subspace techniques for estimating generalized eigenvalues of matrix pencils are discussed, such as PRO-ESPRIT and TLS-ESPRIT.

The outline of the paper is as follows. In Section II, we present some of the main sampling results for signals of finite rate of innovation from [24], and discuss the problems that arise in the presence of noise. In Section III, we consider the problem of sampling a periodic continuous-time signal made up of a stream of Diracs, embedded in additive white Gaussian noise. We develop a frequency-domain subspace framework for high-resolution parameter estimation [16] and prove a relation between the subspace approach and the polynomial realization of the estimator [8], [24]. We also discuss techniques for improving the resolution performance of the proposed method, without increasing the computational complexity. In Section IV, we extend the results to more complex classes of periodic signals of finite rate of innovation, such as nonuniform splines and piecewise polynomials, and present methods that solve the problem in the Laplace domain. In Section V, we consider the sampling problem for finite-length signals using a Gaussian kernel, and develop schemes that allow for an almost local reconstruction of the signal. In Section VI, we present an analysis of the numerical performance of the subspace estimator and discuss the problem of a model mismatch. Simulation results that illustrate the numerical performance of the developed techniques are shown in Section VII, and finally, in Section VIII, we conclude with a brief summary of the main results.

II. Problem Statement

A class of signals with finite rate of innovation can be informally defined as a class of parametric signals having a finite number of degrees of freedom per unit of time [18], [24]. Even though such signals are generally not bandlimited, it was shown that some of them can be perfectly reconstructed from a set of samples taken at the innovation rate [24]. These results can also be extended to multidimensional signals [12], [13].

Consider the following simple example. Let a signal $x(t)$ be a continuous-time signal made up of K weighted Diracs, i.e.

$$x(t) = \sum_{k=0}^{K-1} a_k \delta(t - t_k). \tag{1}$$

Clearly, such a signal is not bandlimited and the standard Shannon sampling theorem cannot be used in this case. However, note that $x(t)$ has only $2K$ degrees of freedom, that is, time delays $\{t_k\}_{k=0}^{K-1}$ and weighting coefficients $\{a_k\}_{k=0}^{K-1}$. Therefore, it seems intuitive that by taking only $2K$ measurements of the signal, one can perfectly estimate all the unknown parameters. This can be achieved, provided that the signal is sampled with a proper sampling kernel, such as a Gaussian or a sinc sampling kernel [24].

In order to show the main idea behind the approach from [24], assume that the signal is filtered with the Gaussian kernel $h_g(t) = e^{-t^2/2\sigma^2}$ and that $N_t \geq 2K$ samples are taken from a filtered version

$$y_n = \langle h_g(t - nT), x(t) \rangle, \quad n = 1, \ldots, N_t. \tag{2}$$

In this case, the sample values are given by

$$\begin{aligned} y_n &= \sum_{k=0}^{K-1} a_k e^{-(t_k - nT)^2/2\sigma^2} \\ &= \sum_{k=0}^{K-1} a_k e^{-t_k^2/2\sigma^2} \cdot e^{nt_kT/\sigma^2} \cdot e^{-n^2T^2/2\sigma^2}. \end{aligned} \tag{3}$$

If we denote by $u_n = y_n e^{n^2T^2/2\sigma^2}$ and $c_k = a_k e^{-t_k^2/2\sigma^2}$, then (3) is equivalent to

$$u_n = \sum_{k=0}^{K-1} c_k e^{nt_kT/\sigma^2} = \sum_{k=0}^{K-1} c_k z_k^n \tag{4}$$

where $z_k = e^{t_kT/\sigma^2}$. Note that the samples u_n are given by a linear combination of exponentials z_k^n; thus, we can reduce the problem of estimating the unknown parameters $\{t_k\}_{k=0}^{K-1}$ and $\{c_k\}_{k=0}^{K-1}$, into the classical spectral estimation problem, that is, the problem of estimating frequencies and weighting coefficients of superimposed exponentials [8], [16], [19].

In the above example, we assumed a deterministic, noiseless signal, when the presented method yields perfect estimates of all the parameters from only $2K$ samples. Yet, in the presence of noise, such an approach often gives rise to numerical ill-conditioning. To understand the main reason for performance degradation, consider a noisy version of the signal, that is, $\tilde{x}(t) = x(t) + \eta(t)$, where $\eta(t)$ is additive white Gaussian noise, and consider the set of noisy samples $\tilde{y}_n$, taken with the Gaussian kernel. As in the previous case, by denoting $\tilde{u}_n = \tilde{y}_n e^{n^2T^2/2\sigma^2}$ and $c_k = a_k e^{-t_k^2/2\sigma^2}$, the set of samples $\tilde{u}_n$ can be expressed as (3)

$$\tilde{u}_n = \sum_{k=0}^{K-1} c_k z_k^n + \eta_n e^{n^2T^2/2\sigma^2}. \tag{5}$$

In this case, however, the samples of noise η_n become significantly amplified as n increases, due to the weighting of $\tilde{y}_n$s with exponentially increasing terms $e^{n^2T^2/2\sigma^2}$. This obviously

makes the above method for reconstructing the signal from the samples $\tilde{u}_n$ ill-conditioned.

A similar problem occurs with other classes of signals as well, such as piecewise polynomials or nonuniform splines. In [24], it was proved that in the absence of noise, the sampling problem for these signals can be reduced to the problem of sampling streams of Diracs, by taking a sufficient number of signal derivatives. For example, consider a periodic nonuniform spline $x(t)$ of period T and degree R, that is, a signal whose $(R+1)$th derivative $x^{(R+1)}(t)$ is a periodic stream of weighted Diracs. If we denote by $X[m]$ and $X^{(R+1)}[m]$ the Fourier series coefficients of $x(t)$ and $x^{(R+1)}(t)$, respectively, the following relation holds:

$$X^{(R+1)}[m] = (jm\omega_0)^{(R+1)} X[m] = \frac{1}{T}\sum_{k=0}^{K-1} c_k e^{-jm\omega_0 t_k}, \quad \omega_0 = 2\pi/T. \tag{6}$$

Similarly to the argument for numerical ill-conditioning in the case of a set of Diracs, the problem with this approach is derivation. In particular, in order to compute the coefficients $X^{(R+1)}[m]$ and estimate the unknown parameters c_k and t_k, one first has to multiply the coefficients $X[m]$ with $(jm\omega_0)^{(R+1)}$, which amplifies noise as frequency increases. In the case of piecewise polynomials, the method in [24] uses derivation as well, leading to a sum of derivatives of Diracs. In addition to noise amplification, such an approach requires identification of multiple roots of the annihilating filter, a task that is difficult even in the noiseless case.

This obviously calls for an extension of the original results from [24] to solve the problems of ill-conditioning and robustness to noise, and to investigate alternative algebraic approaches that will yield numerically stable and precise reconstruction. In the following, we will show that by exploiting the signal structure properly, one can come up with more general constructions that satisfy all of the above requirements.

III. Periodic Stream of Diracs: Continuous-Time Case

In this section, we consider the problem of sampling and reconstruction of a periodic stream of Diracs. While this signal has a relatively simple parametric representation, it provides a basis for all the constructions that will be discussed later.

A. Frequency Domain Formulation

Consider a periodic signal $x(t)$ of period T, given by a sum of weighted Diracs

$$x(t) = \sum_n \sum_{k=0}^{K-1} c_k \delta(t - t_k - nT). \tag{7}$$

Since the signal is periodic, its Fourier series coefficients $X[m]$ are given by

$$X[m] = \frac{1}{T}\sum_{k=0}^{K-1} c_k e^{-jm\omega_0 t_k} = \frac{1}{T}\sum_{k=0}^{K-1} c_k z_k^m, \quad m \in \mathbb{Z} \tag{8}$$

where $\omega_0 = 2\pi/T$. That is, the coefficients $X[m]$ are given by a sum of K complex exponentials $z_k^m = e^{-jm\omega_0 t_k}$, where z_k are usually referred to as signal poles [16], [19]. In [24], the parameters $\{t_k\}_{k=0}^{K-1}$ and $\{c_k\}_{k=0}^{K-1}$ are estimated from a set of $N_t \geq 2K$ adjacent coefficients $X[m]$, using a method based on annihilating filters, which belongs to the class of model-based parametric methods for harmonic retrieval [16]. In particular, the annihilating filter approach exploits the fact that in the absence of noise, each exponential $\{e^{-jm\omega_0 t_k}\}_{m\in Z}$ can be "nulled out" or annihilated by a first order FIR filter $H_k(z) = (1 - e^{-j\omega_0 t_k} z^{-1})$, that is

$$e^{-jm\omega_0 t_k} * [1, -e^{-j\omega_0 t_k}] = 0.$$

Therefore, one can consider a Kth order FIR filter $H(z) = \sum_{k=0}^{K} H[k] z^{-k}$, with K zeros at $z_k = e^{-j\omega_0 t_k}$

$$H(z) = \prod_{k=1}^{K} (1 - e^{-j\omega_0 t_k} z^{-1}). \tag{9}$$

Note that the filter $H[k]$ is the convolution of K elementary filters with coefficients $[1, -e^{-j\omega_0 t_k}], k = 1, \ldots, K$. Since $X[m]$ is the sum of complex exponentials, each will be annihilated by one of the roots of $H(z)$; thus, we have

$$(H * X)[m] = \sum_{k=0}^{K} H[k] X[m-k] = 0, \quad m \in \mathbb{Z}. \tag{10}$$

Without loss of generality, we can set $H[0] = 1$ and rewrite (10) as

$$X[m] = -\sum_{k=1}^{K} H[k] X[m-k], \quad m \in \mathbb{Z}. \tag{11}$$

Therefore, (11) allows us to solve for the annihilating filter coefficients $H[k], k = 1, \ldots, K$, given a set of $N_t \geq 2K$ adjacent Fourier series coefficients $X[m]$. Once the annihilating filter $H(z)$ has been found, the information about the time instants t_k can be extracted from the roots of the filter, while the corresponding coefficients c_k are then estimated by solving the system of linear equations (8).

Note that the signal $x(t)$ has only $2K$ degrees of freedom, $\{t_k\}_{k=0}^{K-1}$ and $\{c_k\}_{k=0}^{K-1}$. Therefore, in the absence of noise, it suffices to use only $2K$ adjacent Fourier series coefficients $X[m]$ to obtain perfect estimates of all the unknown parameters [24]. For example, one can consider the coefficients $X[m], m = -K+1, \ldots, K$, which can be obtained by sampling a lowpass version of the signal. Namely, if x_n denote the uniform samples taken with a sinc kernel $\varphi(t)$ of bandwidth $B = [-K\omega_0, K\omega_0]$, that is

$$x_n = \langle \varphi(t - nT_s) x(t) \rangle, \quad n = 0, 1, \ldots, N_t$$

then the following relation holds:

$$x_n = \left\langle \varphi(t - nT_s), \sum_m X[m] e^{jm\omega_0 t} \right\rangle = \sum_m X[m] \Psi(m\omega_0) e^{jm\omega_0 nT_s} = \sum_{m=-K}^{K} X[m] e^{jm\omega_0 nT_s} \tag{12}$$

where $\Psi(\omega)$ is the Fourier transform of $\varphi(t)$. If the sampling period T_s is chosen such that $N_t = T/T_s \geq 2K$, then the above system of equations is invertible and will yield a unique solution for $X[m], m \in [-K, K]$. Finally, we should note that in order to solve for the annihilating filter coefficients $H[k]$ from (11), one can also consider the coefficients $X[m]$ corresponding to any other signal subspace, as long as its dimension is greater than or equal to the number of degrees of freedom of the signal [11].

In practice, noise will be present, and in general this can be dealt with by oversampling (i.e., one has to consider a larger signal subspace) and using standard techniques from noisy spectral estimation, such as the singular value decomposition. However, even though such an approach improves numerical accuracy on the estimates of the filter coefficients, it is not sufficient for good overall performance of the algorithm. Specifically, both theoretical analysis and practice have shown that in order to reduce sensitivity of time-delay estimates to noise, the filter order should be chosen according to the length of the data set, rather than the number of unknown signal components [7], [14]. That is, even though the number of components may be relatively low, typically, a high-order filter must be used, which imposes a significant computational burden since it is necessary to find roots of a large size polynomial in order to extract a small number of signal poles [16]. In addition to increased computational complexity, overmodeling gives rise to spurious filter zeros, which can be incorrectly identified as signal poles.

This brings us to a more practical version of the model-based approach, the so-called subspace estimator, which avoids the root finding step and relies only on a proper use of matrix operations. It takes advantage of the so-called shift-invariant subspace property and leads to robust estimates without overmodeling, by properly exploiting the algebraic structure of the signal subspace [7], [16].

B. Subspace-Based Approach

Consider again the set of the Fourier series coefficients $X[m]$, given by (8), and construct a Hankel[1] data matrix $\mathbf{X}$ of size $M \times N$, where $M, N > K$

$$\mathbf{X} = \begin{pmatrix} X[0] & X[1] & \cdots & X[N-1] \\ X[1] & X[2] & \cdots & X[N] \\ \vdots & & & \\ X[M-1] & X[M] & \cdots & X[M+N-2] \end{pmatrix}. \quad (13)$$

For simplicity, we have constructed the data matrix using only the coefficients $X[m]$ with nonnegative indices m. In [11], we showed that the method can be directly extended to the case when the coefficients with negative indices are used as well.

In the absence of noise, $\mathbf{X}$ can be decomposed as $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^H$, where $\mathbf{U}, \mathbf{S}$ and $\mathbf{V}^H$ are given by

$$\mathbf{U} = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ z_0 & z_1 & z_2 & \cdots & z_{K-1} \\ \vdots & & & & \\ z_0^{M-1} & z_1^{M-1} & z_2^{M-1} & \cdots & z_{K-1}^{M-1} \end{pmatrix} \quad (14)$$

$$\mathbf{S} = \frac{1}{T}\mathrm{diag}(c_0 \quad c_1 \quad c_2 \quad \cdots \quad c_{K-1}) \quad (15)$$

[1]A Hankel matrix is a matrix in which the (i, j)th entry depends only on the sum $i + j$.

$$\mathbf{V}^H = \begin{pmatrix} 1 & z_0 & z_0^2 & \cdots & z_0^{N-1} \\ 1 & z_1 & z_1^2 & \cdots & z_1^{N-1} \\ \vdots & & & & \\ 1 & z_{K-1} & z_{K-1}^2 & \cdots & z_{K-1}^{N-1} \end{pmatrix}. \quad (16)$$

At this point, it is important to note that the above factorization is not unique. That is, if $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^H$, then $\mathbf{X} = \mathbf{U}\mathbf{P} \cdot \mathbf{P}^{-1}\mathbf{S}\mathbf{Q} \cdot \mathbf{Q}^{-1}\mathbf{V}^H$ is another possible factorization, for every choice of $K \times K$ nonsingular matrices $\mathbf{P}$ and $\mathbf{Q}$. However, as we will show in the following, any such factorization can be used to estimate the signal parameters.

Consider first the matrix $\mathbf{U}$, given by (14). Since $\mathbf{U}$ has a Vandermonde structure, it can be written in the following, more compact form

$$\mathbf{U} = \begin{pmatrix} \mathbf{b} \\ \mathbf{b} \cdot \mathbf{\Phi} \\ \mathbf{b} \cdot \mathbf{\Phi}^2 \\ \vdots \\ \mathbf{b} \cdot \mathbf{\Phi}^{M-1} \end{pmatrix} \quad (17)$$

where $\mathbf{b}$ is a row vector of length K, given by $\mathbf{b} = [1 \ 1 \ \cdots \ 1]$, while $\mathbf{\Phi}$ is a $K \times K$ diagonal matrix containing the signal poles on the main diagonal, i.e., $\mathbf{\Phi} = \mathrm{diag}(z_k)$. Similarly, due to the Vandermonde structure of the matrix $\mathbf{V}$ (16), it can be written as

$$\mathbf{V} = \begin{pmatrix} \mathbf{b} \\ \mathbf{b} \cdot \mathbf{\Phi}_\mathbf{h} \\ \mathbf{b} \cdot \mathbf{\Phi}_\mathbf{h}^2 \\ \vdots \\ \mathbf{b} \cdot \mathbf{\Phi}_\mathbf{h}^{N-1} \end{pmatrix} \quad (18)$$

where $\mathbf{\Phi}_\mathbf{h} = \mathbf{\Phi}^H$.

The subspace approach takes advantage of two properties of the data matrix $\mathbf{X}$. The first property is that in the case of noiseless data, $\mathbf{X}$ has rank K. This will allow us to reduce the noise level by approximating the noisy data matrix with a rank K matrix. The second property is the shift-invariant subspace property. Namely, if we consider the matrices $\mathbf{U}$ and $\mathbf{V}$, given by (17) and (18), they satisfy the following relations:

$$\bar{\mathbf{U}} = \underline{\mathbf{U}} \cdot \mathbf{\Phi} \quad \text{and} \quad \bar{\mathbf{V}} = \underline{\mathbf{V}} \cdot \mathbf{\Phi}_\mathbf{h} \quad (19)$$

where $\overline{(\cdot)}$ and $\underline{(\cdot)}$ denote the operations of omitting the first and the last row of $(\cdot)$, respectively. Note that the shift-invariance property is satisfied not only by $\mathbf{U}$ and $\mathbf{V}$, but also by all matrices $\mathbf{U}\mathbf{P}$ and $\mathbf{V}\mathbf{Q}$, where, as already mentioned, $\mathbf{P}$ and $\mathbf{Q}$ are any nonsingular $K \times K$ matrices. In order to prove this property, consider, for example, the matrix $\mathbf{U}\mathbf{P}$. This matrix can be expressed as

$$\mathbf{U}\mathbf{P} = \begin{pmatrix} \mathbf{b}\mathbf{P} \\ \mathbf{b}\mathbf{P} \cdot \mathbf{P}^{-1}\mathbf{\Phi}\mathbf{P} \\ \mathbf{b}\mathbf{P} \cdot \mathbf{P}^{-1}\mathbf{\Phi}^2\mathbf{P} \\ \vdots \\ \mathbf{b}\mathbf{P} \cdot \mathbf{P}^{-1}\mathbf{\Phi}^{M-1}\mathbf{P} \end{pmatrix} \quad (20)$$

where we have inserted $\mathbf{P} \cdot \mathbf{P}^{-1}$ between $\mathbf{b}$ and $\mathbf{\Phi}^k, k = 1, \ldots, M-1$. Given that $(\mathbf{P}^{-1}\mathbf{\Phi}\mathbf{P})^k = \mathbf{P}^{-1}\mathbf{\Phi}^k\mathbf{P}$, it becomes

obvious that $\mathbf{UP}$ satisfies the shift-invariance property as well, that is

$$\overline{\mathbf{UP}} = \underline{\mathbf{UP}} \cdot \mathbf{P^{-1}\Phi P}. \tag{21}$$

Since the matrix $\mathbf{P}^{-1}\mathbf{\Phi P}$ in (21) is related to $\mathbf{\Phi}$ by a similarity transformation, it has the same eigenvalues as $\mathbf{\Phi}$, i.e., $\{z_k\}_{k=0}^{K-1}$. Similarly, it can be proved that the matrix $\mathbf{VQ}$, where $\mathbf{Q}$ is any $K \times K$ nonsingular matrix, satisfies the following relation:

$$\overline{\mathbf{VQ}} = \underline{\mathbf{VQ}} \cdot \mathbf{Q^{-1}\Phi_h Q}. \tag{22}$$

In practice, the data matrix $\mathbf{X}$ will be decomposed using the singular value decomposition (SVD) as

$$\mathbf{X} = \mathbf{U_s S_s V_s}^H + \mathbf{U_n S_n V_n}^H \tag{23}$$

where the columns of $\mathbf{U_s}$ and $\mathbf{V_s}$ are K principal left and right singular vectors of $\mathbf{X}$, respectively, while the second term contains remaining nonprincipals. Since the presence of additive white noise has little effect on the principal singular vectors,[2] the singular vectors $\mathbf{U_s}$ and $\mathbf{V_s}$, corresponding to the K dominant singular values, will be good estimates of the singular vectors of the original, noiseless matrix $\mathbf{X}$. Since both $\mathbf{U_s}$ and $\mathbf{V_s}$ are matrices of rank K (as well as $\mathbf{U}$ and $\mathbf{V}$ in (17) and (18)), there will exist $K \times K$ nonsingular matrices $\mathbf{P}$ and $\mathbf{Q}$ such that $\mathbf{U_s} = \mathbf{U} \cdot \mathbf{P}$ and $\mathbf{V_s} = \mathbf{V} \cdot \mathbf{Q}$. As a result, both matrices $\mathbf{U_s}$ and $\mathbf{V_s}$ will satisfy the shift-invariance property, and, therefore, the time instants of Diracs $\{t_k\}_{k=0}^{K-1}$ can be uniquely determined from the eigenvalues λ_k of an operator that maps $\underline{\mathbf{U}}_\mathbf{s}$ onto $\bar{\mathbf{U}}_\mathbf{s}$ (or $\underline{\mathbf{V}}_\mathbf{s}$ onto $\bar{\mathbf{V}}_\mathbf{s}$), that is

$$t_k = -\frac{T\angle\lambda_k}{2\pi}. \tag{24}$$

Finally, we should note that when the signal poles are estimated from the left singular vectors $\mathbf{U_s}$, the minimum required size of the data matrix $\mathbf{X}$ in the noiseless case is $(K+1) \times K$. Alternatively, if the right singular vectors $\mathbf{V_s}$ are used for estimation, the minimum size of $\mathbf{X}$ is $K \times (K+1)$. Once the signal poles have been estimated, the weighting coefficients c_k can be found as a least-squares solution to (8). In the following, we give a summary of the algorithm.

Subspace-Based Algorithm:

1) Given a set of the Fourier series coefficients $X[m]$, construct an $M \times N$ matrix data $\mathbf{X}$ as in (13), where $M, N > K$.
2) Compute the singular value decomposition of $\mathbf{X}$, that is, $\mathbf{X} = \mathbf{USV}^H$. Find the principal left and right singular vectors, $\mathbf{U_s}$ and $\mathbf{V_s}$, as the singular vectors corresponding to the K largest singular values of $\mathbf{X}$.
3) Estimate the signal poles $z_k = e^{-j\omega_0 t_k}$ by computing the eigenvalues of a matrix $\mathbf{Z}$, defined as

$$\mathbf{Z} = \underline{\mathbf{U}_\mathbf{s}}^{+} \cdot \bar{\mathbf{U}}_\mathbf{s}. \tag{25}$$

   Note that if $\mathbf{V_s}$ is used in (25), one would estimate complex conjugates of z_ks, since in the singular value decomposition of $\mathbf{X}$, $\mathbf{V_s}$ is used with the Hermitian transpose.

4) Find the coefficients c_k as a least-squares solution to the Vandermonde system (8), that is

$$X[m] = \sum_{k=0}^{K-1} c_k e^{-jn\omega_0 t_k} + \mathcal{N}[m].$$

[2]This is true under the assumption that the smallest singular value corresponding to the signal is not dominated by noise.

C. Relation Between the Subspace and the Polynomial Estimator

In the above case, where the coefficients $X[m]$ are given by a linear combination of exponentials, it is possible to find a decomposition of the matrix $\mathbf{X} = \mathbf{USV}^H$, where both $\mathbf{U}$ and $\mathbf{V}$ are Vandermonde matrices. This allowed us to exploit the shift-invariance property (19), and estimate the signal poles as the eigenvalues of the operator $\mathbf{Z}$ that maps one signal subspace onto another, "shifted" subspace. However, in the case when the coefficients $X[m]$ have more complex structure (e.g., in the case of piecewise polynomial signals), finding an exact algebraic expression for matrices obtained by any such decomposition becomes a much more involved task, whereas the polynomial parameterization may still allow for an intuitive and relatively simple solution. Therefore, in this section, we prove a general relation between the annihilating filter approach [24] and the subspace estimator, which will allow us to extend the subspace method to other classes of signals. To avoid any confusion about notation, in the following, lowercase bold and uppercase bold will denote, respectively, a column/row vector and a matrix.

Given a set of Fourier series coefficients $X[n]$, we first define the state vector [16] of length K as

$$\mathbf{x}[n] = (X[n-1] \quad X[n-2] \quad X[n-3] \quad \cdots \quad X[n-K])^T. \tag{26}$$

From the system (11), one can see that each coefficient $X[n]$ can be predicted from its K past values; thus, we can write

$$\begin{aligned} X[n] &= (-H[1] \quad -H[2] \quad -H[3] \quad \cdots \quad -H[K])\mathbf{x}[n] \\ &\Leftrightarrow X[n] = \mathbf{h}\,\mathbf{x}[n]. \end{aligned} \tag{27}$$

Combining (26) and (27), we obtain

$$\begin{pmatrix} X[n] \\ X[n-1] \\ X[n-2] \\ \vdots \\ X[n-K+1] \end{pmatrix} = \begin{pmatrix} -H[1] & -H[2] & -H[3] & \cdots & -H[K] \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & & \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} X[n-1] \\ X[n-2] \\ X[n-3] \\ \vdots \\ X[n-K] \end{pmatrix} \tag{28}$$

or in matrix form

$$\mathbf{x}[n+1] = \mathbf{H}\,\mathbf{x}[\mathbf{n}]. \tag{29}$$

Therefore, the state space representation of the polynomial estimator is given by

$$X[n] = \mathbf{h}\,\mathbf{x}[n] \tag{30}$$

$$\mathbf{x}[n+1] = \mathbf{H}\,\mathbf{x}[n]. \tag{31}$$

Starting with (30) and (31), we can write

$$X[n] = \mathbf{h}\,\mathbf{H}^n\,\mathbf{x}[0]. \tag{32}$$

Now let us show how the above relations can be used to find a subspace-based solution to the estimation problem. Similar to the approach from Section III-B, one should first construct a Hankel data matrix $\mathbf{X}$ as in (13), of size $M \times N$, where $M, N \geq K$. Using (31) and (32), the matrix $\mathbf{X}$ in (13) can be factored as follows:

$$\mathbf{X} = \begin{pmatrix} \mathbf{h} \\ \mathbf{h}\,\mathbf{H} \\ \mathbf{h}\,\mathbf{H}^2 \\ \vdots \\ \mathbf{h}\,\mathbf{H}^{M-1} \end{pmatrix} \cdot (\mathbf{x}[0] \quad \mathbf{H}\,\mathbf{x}[0] \quad \mathbf{H}^2\,\mathbf{x}[0] \quad \cdots \quad \mathbf{H}^{N-1}\,\mathbf{x}[0]) = \mathbf{L}\mathbf{R}^T. \tag{33}$$

Thus, one can think of $\mathbf{H}$ as being an operator that maps $\underline{\mathbf{L}}$ onto $\bar{\mathbf{L}}$ (or, alternatively, $\underline{\mathbf{R}}$ onto $\bar{\mathbf{R}}$). Now the key is to observe that the characteristic polynomial of $\mathbf{H}$ is given by

$$\det(\lambda\mathbf{I} - \mathbf{H}) = \lambda^K + \sum_{i=1}^{K} H[i]\lambda^{K-i}. \tag{34}$$

By comparing (34) with an expression for the annihilating filter, that is, $H(z) = 1 + \sum_{i=1}^{K} H[i]z^{-i}$, it follows that the eigenvalues of the operator $\mathbf{H}$ (i.e., the zeros of its characteristic polynomial) are identical to the zeros of the annihilating filter. Note that this relation holds in the general case and allows one to obtain the subspace estimator once the annihilating filter has been determined, and vice versa. As already discussed in Section III-B, there will be no difference whether $\mathbf{H}$ is obtained from the matrix $\mathbf{L}$ in (33), or from another matrix $\mathbf{LQ}$, where $\mathbf{Q}$ is any nonsingular matrix. The corresponding estimates of $\mathbf{H}$ are related by similarity transformation and, thus, have the same eigenvalues. This property also holds in the case when $\mathbf{H}$ is computed from the matrix $\mathbf{R}$ or, alternatively, $\mathbf{QR}$. In practice, since the matrix $\mathbf{X}$ will be decomposed using the singular value decomposition as $\mathbf{X} = \mathbf{U_s}\mathbf{S_s}\mathbf{V_s}^H + \mathbf{U_n}\mathbf{S_n}\mathbf{V_n}^H$, with the first term corresponding to the principal components, the operator $\mathbf{H}$ can be found either from the matrix $\mathbf{U_s}$ or from the matrix $\mathbf{V_s}$. This leads us to the following proposition.

Proposition 1: Consider a Hankel matrix $\mathbf{X}$ with entries $X[m]$ (13), and let $\mathbf{X} = \mathbf{USV}^H$ denote its singular value decomposition. Assume next that $X[m]$ can be annihilated by an FIR filter $H(z) = \sum_{k=0}^{K} H[m]z^{-m}$, that is, $(H * X)[m] = \sum_{k=0}^{K} H[k]X[m-k] = 0, \forall m \in \mathbb{Z}$, where K is chosen such that $H(z)$ is of minimum order. Then, in the noiseless case, the roots of the filter $H(z)$ are identical to the nonzero eigenvalues of a matrix $\mathbf{H} = \underline{\mathbf{U}}^{+} \cdot \bar{\mathbf{U}}$.

D. Estimation of Closely Spaced Diracs: Improving the Resolution

Model-based parameter estimation using subspace methods has received significant attention in the literature [8], [9], [16]. In many problems encountered in practice, such as direction finding, frequency estimation, channel estimation and others, subspace-based methods provide an attractive alternative to a more complex maximum likelihood (ML) estimator, as they yield accurate estimates at a reasonable computational cost. However, the problem encountered in all model-based methods is that their performance typically degrades if the signal contains closely spaced components. This can become critical in certain applications, such as channel estimation in ultra-wideband systems [11], [14], where one has to estimate many closely spaced components in a very low signal-to-noise ratio regime. In the following, we present two different techniques that improve the resolution performance of the developed scheme without increasing the data set used for estimation.

Consider again the data matrix $\mathbf{X}$, defined in (13). In order to estimate the signal poles z_ks, we have exploited the shift-invariant subspace property (19), that is, $\bar{\mathbf{U}} = \underline{\mathbf{U}} \cdot \mathbf{\Phi}$, or alternatively, $\bar{\mathbf{V}} = \underline{\mathbf{V}} \cdot \mathbf{\Phi_h}$, where $\mathbf{\Phi}$ is a diagonal matrix with z_ks along the main diagonal, while $\mathbf{\Phi_h} = \mathbf{\Phi}^H$. However, the Vandermonde structure of $\mathbf{U}$ and $\mathbf{V}$ allows for a more general version of (19), specifically

$$\bar{\mathbf{U}}^p = \underline{\mathbf{U}}_p \cdot \mathbf{\Phi}^p \quad \text{and} \quad \bar{\mathbf{V}}^p = \underline{\mathbf{V}}_p \cdot \mathbf{\Phi}_{\mathbf{h}}^p \tag{35}$$

where $\overline{(\cdot)}^p$ and $\underline{(\cdot)}_p$ denote the operations of omitting the first p rows and last p rows of $(\cdot)$, respectively [22]. In this case, the matrix $\mathbf{\Phi}^p$ has elements $z_k^p = e^{-j\omega_0 p t_k}$ on its main diagonal. Therefore, the advantage of using the values of p larger than $p = 1$, is that the effective separation among the estimated time delays is increased p times, which in turn improves the resolution capabilities of the method [14]. In the sequel, this approach will be referred to as the "subspace-shifting" approach.

Another way to improve the performance in the case of closely spaced components is the following. Instead of constructing the data matrix $\mathbf{X}$ as in (13), one can construct another data matrix $\mathbf{X}_i$ as

$$\mathbf{X}_i = \begin{pmatrix} X[0] & X[1] & \dots & X[N-1] \\ X[p] & X[p+1] & \dots & X[p+N-1] \\ X[2p] & X[2p+1] & \dots & X[2p+N-1] \\ \vdots & \ddots & & \\ X[1] & X[2] & \dots & X[N] \\ X[p+1] & X[p+2] & \dots & X[p+N] \\ X[2p+1] & X[2p+2] & \dots & X[2p+N] \\ \vdots & \ddots & & \\ X[p-1] & X[p] & \dots & X[p+N-2] \\ X[2p-1] & X[2p] & \dots & X[2p+N-2] \\ \vdots & & & \end{pmatrix}. \tag{36}$$

That is, $\mathbf{X}_i$ is obtained by interleaving the rows of the original matrix $\mathbf{X}$ in (13). Following the discussion from Section III-B, in the noiseless case, the matrix $\mathbf{X}_i$ can be decomposed as $\mathbf{X}_i =$

$\mathbf{U}_i\mathbf{S}_i\mathbf{V}_i^H$, where $\mathbf{S}_i$ and $\mathbf{V}_i$ are the matrices given by (15) and (16), respectively, while the matrix $\mathbf{U}_i$ is now given by

$$\mathbf{U}_i = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ z_0^p & z_1^p & z_2^p & \dots & z_{K-1}^p \\ z_0^{2p} & z_1^{2p} & z_2^{2p} & \dots & z_{K-1}^{2p} \\ \vdots & \ddots & & & \\ z_0 & z_1 & z_2 & \dots & z_{K-1} \\ z_0^{p+1} & z_1^{p+1} & z_2^{p+1} & \dots & z_{K-1}^{p+1} \\ z_0^{2p+1} & z_1^{2p+1} & z_2^{2p+1} & \dots & z_{K-1}^{2p+1} \\ \vdots & \ddots & & & \\ z_0^{p-1} & z_1^{p-1} & z_2^{p-1} & \dots & z_{K-1}^{p-1} \\ z_0^{2p-1} & z_1^{2p-1} & z_2^{2p-1} & \dots & z_{K-1}^{2p-1} \\ \vdots & & & & \end{pmatrix}. \quad (37)$$

Note that each column k of the matrix $\mathbf{U}_i$ is made up of p blocks that contain consecutive powers of the signal pole z_k. Therefore, in order to estimate z_ks, one can exploit the following shift-invariance property:

$$\bar{\mathbf{U}}_i^{p,b} = \underline{\mathbf{U}_i}_{p,b} \cdot \mathbf{\Phi}^p \quad (38)$$

where in this case $\overline{(\cdot)}^{p,b}$ and $\underline{(\cdot)}_{p,b}$ denote the operations of omitting the first row and last row in each block of $(\cdot)$, respectively. Note that the matrix $\mathbf{\Phi}^p$ is the same diagonal matrix as before, with elements $z_k^p = e^{-j\omega_0 p t_k}$ on its main diagonal. However, since the matrix $\mathbf{X}_i$ has better conditioning than the original data matrix $\mathbf{X}$, the interleaving approach results in better resolution performance, as we will show in Section VI. Therefore, we can state the following proposition.

Proposition 2: Consider the matrices $\mathbf{X}$ and $\mathbf{X}_i$ defined in (13) and (36), respectively, and let $\mathbf{X} = \mathbf{USV}^H$ and $\mathbf{X}_i = \mathbf{U}_i\mathbf{S}_i\mathbf{V}_i^H$ denote their singular value decompositions. Then, the following holds.

1) If $\{z_k\}_{k=0}^{K-1}$ are eigenvalues of the matrix $\mathbf{Z}_1 = \underline{\mathbf{U}}^+ \cdot \bar{\mathbf{U}}$, then $\{z_k^p\}_{k=0}^{K-1}$ are eigenvalues of the matrix $\mathbf{Z}_2 = \underline{\mathbf{U}}_p{}^+ \cdot \bar{\mathbf{U}}^p$.
2) Matrices $\mathbf{Z}_2 = \underline{\mathbf{U}}_p{}^+ \cdot \bar{\mathbf{U}}^p$ and $\mathbf{Z}_3 = \underline{\mathbf{U}_\mathbf{i}}_{p,b}{}^+ \cdot \bar{\mathbf{U}}_\mathbf{i}^{p,b}$ have identical eigenvalues.
3) If $\{t_k\}_{k=0}^{K-1}$ are the time locations of Diracs estimated from the eigenvalues $\{z_k\}_{k=0}^{K-1}$, then $\{pt_k\}_{k=0}^{K-1}$ are the locations estimated from $\{z_k^p\}_{k=0}^{K-1}$, that is, the separation between each two components is increased p times.

Finally, we would like to note that since we are considering periodic signals, estimates of the time locations t_k obtained from the powers of the signal poles z_k^p are not unique. That is, for each computed eigenvalue z_k^p, there exists a set of p possible corresponding time delays $\tilde{t}_k$, given by $\tilde{t}_k = t_k + nT/p$, where $n = 0, 1, \dots, p-1$. In order to avoid this ambiguity, one can first find an approximate location of the cluster of Diracs, by estimating only one principal component, using the original method from Section III-B, since it is well-known that the largest signal-space singular vector is relatively insensitive to signal separation [10]. This information can be used later to select a proper set of the locations t_k, once the values of z_k^p have been estimated.

IV. Extension to Other Periodic Signals With Finite Rate of Innovation

In this section, we extend our results to more general classes of periodic signals with finite rate of innovation, such as nonuniform splines and piecewise polynomials. In particular, we show how to modify methods from [24] in order to achieve good numerical performance in the noisy case.

A. Periodic Nonuniform Splines

We first consider the sampling problem for periodic nonuniform splines, since this case is a direct extension of the sampling problem for a stream of Diracs, which will also set the grounds for the following section on piecewise polynomials.

A signal $x(t)$ is a periodic nonuniform spline of period T and degree R, with knots at $\{t_k\}_{k=0}^{K-1} \in [0, T]$, if its $(R+1)$th derivative is a periodic stream of K weighted Diracs, that is

$$x^{(R+1)}(t) = \sum_n \sum_{k=0}^{K-1} c_k \delta(t - t_k - nT). \quad (39)$$

The Fourier series coefficients $X^{(R+1)}[m]$ are given by (6), that is, $X^{(R+1)}[m] = (1/T)\sum_{k=0}^{K-1} c_k e^{-jm\omega_0 t_k}$, where $\omega_0 = 2\pi/T$. Therefore, by considering the coefficients that correspond to the $(R+1)$th derivative of the original signal $x(t)$, we can reduce the problem of estimating the unknown signal parameters to the one of estimating the parameters of superimposed complex exponentials. Note that the derivation can be done in the frequency domain, by multiplying the Fourier series coefficients $X[m]$ with $(jm\omega_0)^{(R+1)}$. Once the coefficients $X^{(R+1)}[m]$ have been computed, the signal parameters can be estimated using the method developed in Section III-B. However, in the presence of noise, derivation of $x(t)$ is an ill-conditioned step, as it enhances noise. In the following, we will present a modified version of the method from [24], which yields better performance while retaining the shift-invariant flavor of the original scheme.

Consider again the Fourier series coefficients $X[m]$, given by (6). Note that the term $(jm\omega_0)^{R+1}$, corresponding to the $(R+1)$th derivative operator in the frequency domain, grows as m^{R+1}. Thus, the idea is to weight the coefficients $X^{(R+1)}[m]$ with a multiplicative term $S[m]$, which has an exponential decay, in order to compensate for the polynomial growth of $(jm\omega_0)^{R+1}$. One possible solution is to choose $S[m] = e^{-sm\omega_0}$, where s is a parameter that can be adjusted according to the value of SNR (s is a positive real number) and the size of the data set used for estimation. Consider, thus, an expression for the new, weighted coefficients $X_s[m]$

$$\begin{aligned} X_s[m] &= (j\omega_0 m)^{(R+1)} X[m] e^{-sm\omega_0} \\ &= \sum_{k=0}^{K-1} c_k e^{-jm\omega_0 t_k} e^{-sm\omega_0} \\ &= \sum_{k=0}^{K-1} c_k e^{-m\omega_0(s+jt_k)}. \end{aligned} \quad (40)$$

Note that the $X_s[m]$s are given by a sum of damped exponentials, however, the shift invariance property (19) still holds. Therefore, one can use the method from Section III-B to estimate the unknown parameters $\{c_k\}_{k=0}^{K-1}$ and $\{t_k\}_{k=0}^{K-1}$ from the

coefficients $X_s[m]$. The only difference compared to the original approach is that the estimated eigenvalues z_{ks} are now given by $z_{ks} = e^{-\omega_0 \sigma_k}$, where $\sigma_k = s + jt_k$. Given that the value of the damping factor s is known, the time instants t_ks can be found directly from the eigenvalues z_{ks}, while estimation of the weighting coefficients c_k follows the same procedure as before. We note that this is a general result, which holds for all classes of periodic signals considered in [24]. Namely, all the methods developed for periodic signals use the Fourier series coefficients $X[m]$ to extract the signal poles $z_k = e^{-j\omega_0 t_k}$, and, thus, estimate the unknown locations t_k. Alternatively, one can consider the set of weighted coefficients $X_s[m]$ and use the same techniques to estimate the scaled version of the signal poles, $z_{ks} = e^{-\omega_0(s+jt_k)}$. Therefore, instead of solving the problem in the Fourier domain [24], one can solve the problem in the Laplace transform domain, which allows for the same algorithmic tools, and yet leads to better conditioning for certain classes of signals. Thus, we have the following proposition.

Proposition 3: Consider a periodic nonuniform spline $x(t)$ of period T and degree R, with the $(R+1)$th derivative of the form $x^{(R+1)}(t) = \sum_n \sum_{k=0}^{K-1} c_k \delta(t - t_k - nT)$, and corresponding Fourier series coefficients $X^{(R+1)}[m] = (1/T)\sum_{k=0}^{K-1} c_k e^{-jm\omega_0 t_k}, \omega_0 = 2\pi/T$. Consider a weighted set of coefficients $X_s[m] = X^{(R+1)}[m]e^{-sm\omega_0}, s \geq 0$. If $H(z)$ is the annihilating filter for $X[m]$, then $H(ze^{-s\omega_0})$ will be the annihilating filter for $X_s[m]$, and vice versa. An equivalent statement holds for eigenvalues obtained by the subspace approach (25).

While in the above proposition we made no assumption on the value of the parameter s, in practice, it should be chosen such that in a considered frequency band, the power spectral density (PSD) of noise, weighted by $|(jm\omega_0)^{(R+1)}e^{-sm\omega_0}|^2$, is as uniform as possible (and does not exceed the PSD of a signal). However, we should note that in the general case of nonuniform splines of degree $R > 1$, the method may still be sensitive to noise, despite windowing. Namely, in the case when transitions between adjacent pieces are almost smooth, it is difficult to extract the discontinuities in the differentiated signal, even with proper windowing. In the case of piecewise polynomials, this is not an issue, since the signal itself already contains discontinuities.

B. Piecewise Polynomials

Similar to the definition of periodic nonuniform splines, a signal $x(t)$ is a periodic piecewise polynomial of period T, having K pieces of maximum degree R, if and only if its $(R+1)$th derivative is a periodic stream of differentiated Diracs, that is

$$x^{(R+1)}(t) = \sum_n \sum_{k=0}^{K-1} \sum_{r=0}^{R_k-1} c_{k,r}\delta^{(r)}(t - t_k - nT). \quad (41)$$

The corresponding Fourier series coefficients are now given by

$$X^{(R+1)}[m] = \frac{1}{T}\sum_{k=0}^{K-1} \sum_{r=0}^{R_k-1} c_{k,r}(j\omega_0 m)^r e^{-jm\omega_0 t_k}. \quad (42)$$

By denoting $\tilde{c}_{k,r} = (1/T)c_{k,r}(j\omega_0)^r$, we obtain

$$X^{(R+1)}[m] = \sum_{k=0}^{K-1} \sum_{r=0}^{R_k-1} \tilde{c}_{k,r} m^r e^{-jm\omega_0 t_k}. \quad (43)$$

As already discussed in the previous section, the problem in the presence of noise is derivation. Therefore, instead of considering the set of coefficients corresponding to the $(R+1)$th derivative, the idea is to consider a set of weighted coefficients $X_s^{(R+1)}[m] = X^{(R+1)}[m]e^{-jsm\omega_0}$, given by

$$X_s^{(R+1)}[m] = \sum_{k=0}^{K-1} \sum_{r=0}^{R_k-1} \tilde{c}_{k,r} m^r e^{-m\omega_0(s+jt_k)}. \quad (44)$$

However, in this case, it is no longer obvious that the shift-invariant subspace property (19) can be exploited, since each term $e^{-m\omega_0(s+jt_k)}$ is additionally multiplied by a polynomial $\sum_{r=0}^{R_k-1} m^r$. That is, the coefficients $X_s^{(R+1)}[m]$ are given by a nonlinear combination of complex exponentials. In the following, we will show that one can still obtain a closed-form subspace solution to the problem of parameter estimation from $X_s^{(R+1)}[m]$. Specifically, since the exact solution can be relatively easily obtained using the annihilating filter method [24], we will use Proposition 1, to find a subspace solution to the problem. This will provide a practical approach for solving a more general class of nonlinear estimation problems [14], and will also extend classic high-resolution spectral estimation techniques [8], [16], [19].

Consider the annihilating filter $H(z) = \sum_{m=0}^{N} H[m]z^{-m}$, which satisfies

$$\left(H * X_s^{(R+1)}\right)[n] = 0, \quad \forall n \in \mathbb{Z}. \quad (45)$$

In [24], it was shown that in the case when the coefficients $X_s^{(R+1)}[n]$ are given by (44), the annihilating filter has multiple roots at $z_k = e^{-\omega_0(s+jt_k)}$

$$H(z) = \prod_{k=0}^{K-1}\left(1 - e^{-\omega_0(s+jt_k)}z^{-1}\right)^R = \sum_{m=0}^{\mathrm{RK}} H[m]z^{-m}. \quad (46)$$

Namely, the key is to observe that each component $S_{k,r}^{(R+1)}[m] = m^r e^{-m\omega_0(s+jt_k)}$ in (44) is annihilated by a filter which has $r+1$ zeros at $z_{ks} = e^{-\omega_0(s+jt_k)}$ [24], i.e.,

$$H_{k,r}(z) = \left(1 - e^{-\omega_0(s+jt_k)}z^{-1}\right)^{r+1}. \quad (47)$$

Since the filter $H_{k,R-1}(z)$ annihilates all the components $S_{k,r}^{(R+1)}[m], r = 0, \ldots, R-1$, the annihilating filter for the signal $X_s^{(R+1)}[n]$ is given by

$$H(z) = \prod_{k=1}^{K} H_{k,R-1}(z) = \prod_{k=0}^{K-1}\left(1 - e^{-\omega_0(s+jt_k)}z^{-1}\right)^R. \quad (48)$$

Therefore, the information about the time delays t_k can be extracted from the Rth order roots z_{ks} of the filter $H(z)$, while the corresponding weights $c_{k,r}$ can be then found by solving the system of linear equations (42).

Let us show next how this result can be used to find a subspace solution to the estimation problem. Following the approach from Section III-B, given a set of the coefficients $X_s^{(R+1)}[m]$, one first has to construct a Hankel data matrix $\mathbf{X}$ of size $M \times N$, where $M, N \geq \mathrm{RK}$. The second step is to compute the SVD of $\mathbf{X}$ and extract its RK principal left singular

vectors $\mathbf{U_s}$ (or alternatively, RK principal right singular vectors $\mathbf{V_s}$). Once $\mathbf{U_s}$ has been estimated, one should compute the matrix $\mathbf{Z}$ as in (25), that is, $\mathbf{Z} = \underline{\mathbf{U}}_\mathbf{s}^{+} \cdot \bar{\mathbf{U}}_\mathbf{s}$. From Proposition 1, it follows that the eigenvalues of $\mathbf{Z}$ are identical to the roots of the annihilating filter $H(z)$ (48). Specifically, since the filter has Rth order roots at $z_{ks} = e^{-\omega_0(s+jt_k)}, k = 1,\ldots,K$, the matrix $\mathbf{Z}$ will have K distinct eigenvalues at z_{ks}, each with an algebraic multiplicity R. Thus, an equivalent statement to the one in Proposition 1, holds in this case as well. However, we note that in the presence of noise, it is desirable to estimate the signal poles from those eigenvalues z_{ks} closest to the circle $|z| = e^{-s\omega_0}$ [19]. Finally, we note that since the signal already contains discontinuities, the above method will generally result in better numerical performance than in the case of nonuniform splines.

V. Aperiodic Signals of Finite Rate of Innovation

So far, we have considered the sampling problem for periodic signals of finite rate of innovation and developed methods that exploit the structure of the Fourier series coefficients. The problem becomes more challenging in the aperiodic case, since the frequency domain approach cannot be used. Therefore, we will present alternative methods that solve the problem in the time domain. We will consider schemes where a signal is sampled with a Gaussian kernel, since in such a case, the signal samples have an algebraic structure that can be easily exploited; see (3) and (4).

A. Streams of Diracs

Consider first the basic problem discussed in Section II, that is, the problem of sampling the finite stream of K weighted Diracs, $x(t) = \sum_{k=0}^{K-1} a_k \delta(t - t_k)$, with the Gaussian kernel $h_\sigma(t) = e^{-t^2/2\sigma^2}$. Denote by y_n the samples of the noiseless signal and let $w_n = e^{n^2T^2/2\sigma^2}$. We have shown that the weighted set of samples $u_n = y_n w_n$ can be expressed as a sum of real exponentials (4), i.e.,

$$u_n = \sum_{k=0}^{K-1} c_k e^{nt_kT/\sigma^2} = \sum_{k=0}^{K-1} c_k z_k^n$$

where $z_k = e^{t_kT/\sigma^2}$ and $c_k = a_k w_k$. In practice, the samples u_n will be perturbed by noise, and the noisy measurements $\tilde{u}_n$ can be expressed as

$$\tilde{u}_n = u_n + w_n \eta_n \tag{49}$$

where η_n denote the samples of additive white Gaussian noise. Clearly, the problem in the presence of noise is exponential weighting of the samples, and the Laplace domain formulation of the problem, discussed in Section IV-A, may not always provide a good alternative. We will, thus, present a similar approach to the one developed for periodic signals, which can also be viewed as a generalization of the Laplace domain solution.

Consider again the subspace approach, where the following Hankel matrix is constructed:

$$\tilde{\mathbf{X}} = \begin{pmatrix} \tilde{u}_0 & \tilde{u}_1 & \cdots & \tilde{u}_{N-1} \\ \tilde{u}_1 & \tilde{u}_2 & \cdots & \tilde{u}_N \\ \vdots & \vdots & \ddots & \\ \tilde{u}_{M-1} & \tilde{u}_M & \cdots & \tilde{u}_{M+N-2} \end{pmatrix} \tag{50}$$

with $M, N \geq K$. Denote by $\mathbf{X}$ a matrix constructed from the noiseless samples u_n. Then we can write $\tilde{\mathbf{X}} = \mathbf{X} + \mathbf{W}$, where the matrix $\mathbf{W}$ is made up of weighted samples of noise

$$\mathbf{W} = \begin{pmatrix} w_0\eta_0 & w_1\eta_1 & \cdots & w_{N-1}\eta_{N-1} \\ w_1\eta_1 & w_2\eta_2 & \cdots & w_N\eta_N \\ \vdots & \vdots & \ddots & \\ w_{M-1}\eta_{M-1} & w_M\eta_M & \cdots & w_{M+N-2}\eta_{M+N-2} \end{pmatrix}. \tag{51}$$

For medium to high values of SNR, perturbation on the matrix $\mathbf{X}$ has little effect on the principal singular vectors. In such a case, a rank K approximation of the noiseless data matrix $\mathbf{X}$ can be obtained by computing the singular value decomposition of $\tilde{\mathbf{X}}$ and setting all but the K largest singular values to zero. However, in our case, due to the exponential weighting of noise samples, dominant singular values do not necessarily belong to the signal space, as they may also include those corresponding to noise. In the following, we propose a scaling technique that additionally multiplies the entries of the noisy matrix $\tilde{\mathbf{X}}$ in order to make the noise variance as uniform (and minimal) as possible. This will allow us to use the SVD for noise suppression.

The idea is to replace $\tilde{\mathbf{X}}$ with another matrix

$$\tilde{\mathbf{X}}_s = \mathbf{A}\tilde{\mathbf{X}}\mathbf{B} = \mathbf{AXB} + \mathbf{AWB}.$$

In general, $\mathbf{A}$ and $\mathbf{B}$ can be any invertible matrices, however, the goal is to choose those matrices such that the entries of $\mathbf{W}_s = \mathbf{AWB}$ have a uniform and minimum variance. One possible solution is to make both $\mathbf{A}$ and $\mathbf{B}$ be diagonal matrices, since any linear combination of lines or columns of $\mathbf{W}$ would increase the noise variance. In order to obtain a uniform variance, the elements of $\mathbf{A}$ and $\mathbf{B}$ can be chosen as

$$\mathbf{A}[i,i] = \frac{1}{\frac{1}{N}\sum_{n=0}^{N-1} w_{n+i}}, \quad i = 0,\ldots,M-1 \tag{52}$$

$$\mathbf{B}[j,j] = \frac{1}{\frac{1}{M}\sum_{m=0}^{M-1} A[m,m]w_{m+j}}, \quad j = 0,\ldots,N-1. \tag{53}$$

As a result of such a transformation, one can think of the entries of $\mathbf{W}_s$ as being the samples of additive white noise. Now we can use the singular value decomposition of the noisy matrix $\tilde{\mathbf{X}}_s$, i.e.

$$\tilde{\mathbf{X}}_s = \mathbf{U_s}\mathbf{S_s}\mathbf{V_s}^H + \mathbf{U_n}\mathbf{S_n}\mathbf{V_n}^H$$

where the first term contains K principal components of $\tilde{\mathbf{X}}_s$. However, note that the above transformation destroys the Hankel structure of the original data matrix. Therefore, once the principal components of $\tilde{\mathbf{X}}_s$ have been computed, one can compensate for the effects of $\mathbf{A}$ and $\mathbf{B}$ by constructing a new, denoised data matrix $\mathbf{X}_d$ as $\mathbf{X}_d = \mathbf{A}^{-1}\mathbf{U_s}\mathbf{S_s}\mathbf{V_s}^H\mathbf{B}^{-1}$.

At this point, we should note that there are other whitening transformations that can be used to mitigate the effect of exponential weighting of noisy samples. For example, one can compute a Cholesky decomposition [6] of the noise covariance matrix, that is, $\mathbf{R_w} = \mathbf{C}^T\mathbf{C}$ (assuming that $\mathbf{R_w}$ is a positive definite matrix), and multiply the data matrix by $\mathbf{C}^{-T}$ prior to computing the singular value decomposition. The main advantage of the method we proposed (with two diagonal matrices $\mathbf{A}$ and $\mathbf{B}$)

is its simplicity: it involves only averaging along rows/columns and avoids matrix inversion, while achieving the desired goal.

B. Nonuniform Splines

We will now extend the analysis to the more general class of finite-length signals with finite rate of innovation, such as nonuniform splines. A signal $x(t)$ is a nonuniform spline of degree R if and only if its $(R+1)$th derivative is a stream of weighted Diracs, that is

$$x^{(R+1)}(t) = \sum_{k=0}^{K-1} c_k \delta(t - t_k). \tag{54}$$

Clearly, by sampling the signal $x^{(R+1)}(t)$ with the Gaussian kernel, we can reduce the problem of signal reconstruction to the one of estimating the parameters of superimposed weighted Diracs, which can be solved using the method from Section III-B. However, since one often has no access to the derivatives of the input signal, an equivalent approach would be to sample the input signal with an $(R+1)$th derivative Gaussian kernel. Such a kernel can be expressed as

$$h_\sigma^{(r+1)}(t) = \frac{d^r}{dt^r}\{e^{-t^2/2\sigma^2}\} = P_r(t)e^{-t^2/2\sigma^2} \tag{55}$$

where $P_r(t)$ is a polynomial of degree r, given by the following recurrence relation:

$$P_0(t) = 1; \quad P_r(t) = P'_{r-1}(t) - \frac{t}{\sigma^2}P_{r-1}(t). \tag{56}$$

Note that in the case when $\sigma^2 = 1/2$, the polynomial $P_r(t)$ in (55) and (56) is a Hermite polynomial of degree r [5]. In Section V-D, we will show how one can approximate such kernels with a linear combination of Gaussian functions. Such an approach will lead to a more practical version of the sampling scheme where the signal is sampled with the Gaussian kernel, while all subsequent manipulations are carried out on a set of samples.

C. Piecewise Polynomials

Similar to the periodic case, a signal $x(t)$ is a piecewise polynomial with K pieces, each of maximum degree R, if and only if its $(R+1)$th derivative is a stream of differentiated Diracs, that is

$$x^{(R+1)}(t) = \sum_{k=0}^{K-1}\sum_{r=0}^{R} c_{k,r}\delta^{(r)}(t - t_k). \tag{57}$$

If the signal is sampled with a $(R+1)$th derivative Gaussian kernel, the samples y_n are given by

$$y_n = \sum_{k=0}^{K-1}\sum_{r=0}^{R} c_{k,r}\frac{d^r}{dt^r}\left\{e^{-(t-t_k)^2/2\sigma^2}\right\}\Bigg|_{t=nT}. \tag{58}$$

As already discussed in Section V-A, we can consider a new set of samples u_n, obtained by multiplying y_n with w_n. If we let $P_r(t)$ be an rth order polynomial from (56) and $\tilde{c}_{k,r} = c_{k,r}e^{t_k^2/2\sigma^2}$, then the samples $u_n = y_n w_n$ can be expressed as

$$u_n = \sum_{k=0}^{K-1}\sum_{r=0}^{R} \tilde{c}_{k,r}P_r(nT - t_k)e^{nt_kT/\sigma^2}. \tag{59}$$

Since in the above expression each exponential e^{nt_kT/σ^2} is additionally multiplied by a polynomial in t_k, one can use the subspace approach discussed in Section IV-B to solve for all the unknown parameters. Note that in this case one has to have access to the derivatives of the signal $x(t)$ as well, or alternatively, one should use the $(R+1)$th derivative Gaussian kernel. However, such an approach is not desirable in practice, since typically one cannot choose the sampling kernel arbitrarily. In the following, we present a method that allows the signal to be sampled with the Gaussian kernel, whereas the derivatives of the signal are then computed from a set of samples.

D. Practical Realization of the Gaussian Sampling Schemes

Consider again the Gaussian kernel $h_\sigma(t) = e^{-t^2/2\sigma^2}$. The idea is to express the nth derivative Gaussian kernel as a linear combination of the shifted versions of $h_\sigma(t)$. Let $h_\sigma^{(1)}(t) = h_\sigma(t - \sigma/m) - h_\sigma(t + \sigma/m)$, where σ/m is the corresponding shift. By choosing $m \geq 3$, the function $h_\sigma^{(1)}(t)$ becomes a very good approximation of the first derivative Gaussian function (up to a scaling factor). In order to show this, consider the Fourier transform of $h_\sigma^{(1)}(t)$, i.e.

$$H_\sigma^{(1)}(\omega) = H_\sigma(\omega)e^{j\omega\sigma/m} - H_\sigma(\omega)e^{-j\omega\sigma/m} \tag{60}$$

where $H_\sigma(\omega)$ denotes the Fourier transform of $h_\sigma(t)$. Since $H_\sigma(\omega) = \sqrt{2\pi}\sigma e^{-\omega^2\sigma^2/2}$, we have

$$H_\sigma^{(1)}(\omega) = \sqrt{2\pi}\sigma e^{-\omega^2\sigma^2/2}\cdot(e^{j\omega\sigma/m} - e^{-j\omega\sigma/m}). \tag{61}$$

Using the Taylor series expansion of the second term in (61), we obtain

$$\begin{aligned} H_\sigma^{(1)}(\omega) &= \sqrt{2\pi}\sigma e^{-\omega^2\sigma^2/2}\cdot\left[\left(1 + \frac{j\omega\sigma}{m} + \frac{(j\omega\sigma/m)^2}{2!} + \cdots\right)\right. \\ &\quad \left. - \left(1 - \frac{j\omega\sigma}{m} + \frac{(j\omega\sigma/m)^2}{2!} + \cdots\right)\right] \\ &= \sqrt{2\pi}\sigma e^{-\omega^2\sigma^2/2}\cdot\left[2j\omega\sigma/m + 2\frac{(j\omega\sigma/m)^3}{3!}\right. \\ &\quad \left. + 2\frac{(j\omega\sigma/m)^5}{5!} + \cdots\right]. \end{aligned} \tag{62}$$

The key is to observe that $H_\sigma^{(1)}(\omega)$, which is given by a product of a polynomial in variable $\nu = \omega\sigma$ and an exponentially decaying function of ν^2, contains only one dominant term, that is

$$H_\sigma^{(1)}(\omega) \approx \sqrt{2\pi}\sigma e^{-\omega^2\sigma^2/2}\cdot 2j\omega\frac{\sigma}{m}. \tag{63}$$

This can be seen in Fig. 1(a), where we plot the magnitude of the first three terms in the above sum for $m = 3$. Now one can compare the expression in (63) with an expression for the Fourier transform of the first-derivative Gaussian function, given by

$$\mathcal{F}\left\{\frac{d}{dt}h_\sigma(t)\right\} = \sqrt{2\pi}\sigma e^{-\omega^2\sigma^2/2}\cdot j\omega. \tag{64}$$

Note that there is only a scaling relation between the two expressions; therefore, the above approach gives a good approximation of the first derivative Gaussian function, which is illustrated in Fig. 1(b)–(d).

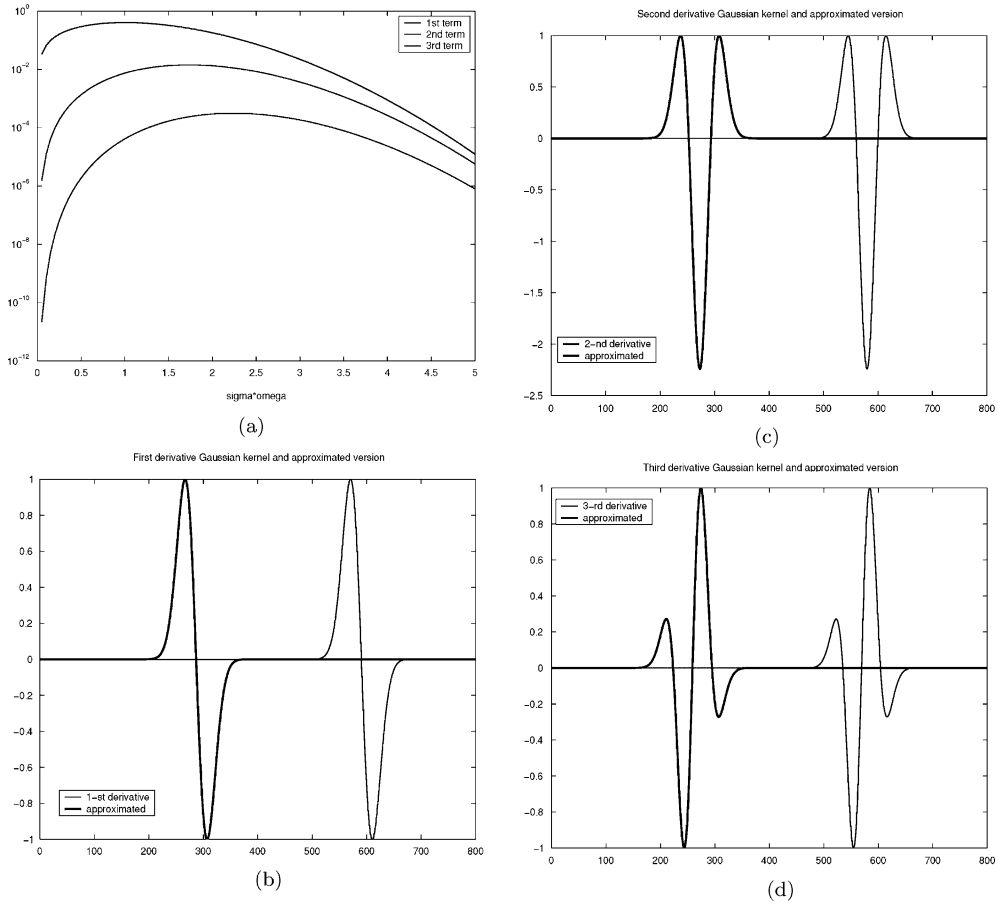

Fig. 1. Approximation of the derivatives of a Gaussian kernel. (a) Magnitude of the first three terms in the expression (62). (b) First derivative Gaussian kernel and its approximated version. (c) Second derivative Gaussian kernel and its approximation. (d) Third derivative Gaussian kernel and its approximation.

The same idea can be used to approximate higher-order derivatives of the Gaussian kernel, using the following recurrence relation:

$$h_\sigma^{(n)}(t) = \frac{m}{2\sigma}\left(h_\sigma^{(n-1)}(t-\sigma/m) - h_\sigma^{(n-1)}(t+\sigma/m)\right) \qquad n = 2, 3, 4, \ldots. \quad (65)$$

Namely, as in the previous case, the key is to note the following relation in the frequency domain:

$$e^{-\omega^2\sigma^2/2}(j\omega)^{n-1}(e^{j\omega\sigma/m} - e^{-j\omega\sigma/m}) \approx \frac{2\sigma}{m}e^{-\omega^2\sigma^2/2}(j\omega)^n. \quad (66)$$

Let us now return to the sampling problem for finite-length signals. For example, consider a piecewise polynomial of maximum degree R. We have seen that in order to reconstruct the signal, one has to take samples with the $(R+1)$th derivative Gaussian kernel at locations kT. Following the procedure described above, these samples can be obtained from the set of samples taken with the Gaussian kernel at sampling instants $\mathrm{kT} - r\sigma/2$, where $r = -R, -R+1, \ldots, R$. This leads to a practical version of the developed algorithm, which we will illustrate in Section VII.

VI. Performance Evaluation

A. Analysis of Noise Sensitivity

The statistical properties of the estimates obtained using high-resolution methods have been studied extensively, mainly in the context of estimating the frequencies of superimposed complex sinusoids from noisy measurements [7], [8], [15], [16]. Expressions for the mean square error (MSE) of the frequency estimates [7] indicate that the numerical performance of such methods is very close to the Cramer–Rao bound (CRB) [20], which represents the lowest achievable MSE by any unbiased estimator. In the following, we present some of the key ideas used in the statistical analysis of the state space approach, while more on this topic can be found in [7], [16].

Consider the data matrix $\mathbf{X}$ and denote by $\mathbf{R_c} = \mathbf{X}\mathbf{X}^H$ the corresponding covariance matrix. We will analyze the subspace approach which uses the eigendecomposition of the covariance matrix $\mathbf{R_c}$. However, note that the same analysis applies to our method, based on the singular value decomposition of the matrix $\mathbf{X}$, since the left singular vectors $(\mathbf{U_s})$ of $\mathbf{X}$ are the eigenvectors of $\mathbf{R}$, that is, $\mathbf{R_c} = \mathbf{U_s}\mathbf{\Lambda_s}\mathbf{U_s}^H$. Let $\tilde{\mathbf{U}}_\mathbf{s} = \mathbf{U_s} + \triangle\mathbf{U_s}$, where $\triangle\mathbf{U_s}$ is the error in the eigenvectors corresponding to the signal subspace. Then, $\triangle\mathbf{U_s}$ can be written as $\triangle\mathbf{U_s} = \mathbf{P_s}\triangle\mathbf{U_s} + \mathbf{P_n}\triangle\mathbf{U_s}$, where $\mathbf{P_s} = \mathbf{U_s}\mathbf{U_s}^H$ is the unique orthogonal projection matrix onto the signal subspace, that is, if $\mathbf{x}$ belongs to the signal subspace, then $\mathbf{P_s}\mathbf{x} = \mathbf{x}$.

Similarly, $\mathbf{P_n} = \mathbf{I} - \mathbf{P_s}$ is a projection operator onto the noise subspace. Now the key is to observe that only $\mathbf{P_n}\triangle\mathbf{U_s}$ contributes to the error in the eigenvalue estimates [16], which simplifies the analysis considerably.

The first step in the analysis is to relate $\triangle\mathbf{R_c}$, the error in the covariance matrix, to $\mathbf{P_n}\triangle\mathbf{U_s}$

$$\mathbf{P_n}\triangle\mathbf{U_s} = \mathbf{P_n}\triangle\mathbf{R_c}\mathbf{U_s}\mathbf{\Lambda_s}^{-1}. \tag{67}$$

Next, we consider how $\mathbf{P_n}\triangle\mathbf{U_s}$ affects $\mathbf{H_s} = \underline{\mathbf{U}}_\mathbf{s}^{+}\bar{\mathbf{U}}_\mathbf{s}$ as well as its eigenvalues. By denoting $\tilde{\mathbf{H}}_s = \mathbf{H_s} + \triangle\mathbf{H_s}$, it can be shown [16] that

$$\triangle\mathbf{H_s} = -\underline{\mathbf{U}}_\mathbf{s}{}^{+}(\triangle\underline{\mathbf{U}}_\mathbf{s}\mathbf{H_s} - \triangle\bar{\mathbf{U}}_\mathbf{s}) \tag{68}$$

while the error in $\mathbf{H_s}$ can be related to the error of its eigenvalue z_k through the following relation:

$$\triangle z_k = \mathbf{p}_k\triangle\mathbf{H_s}\mathbf{q}_k \tag{69}$$

where $\mathbf{p}_k$ and $\mathbf{q}_k$ are left and right eigenvectors of $\mathbf{H_s}$ corresponding to the eigenvalue z_k. Finally, by combining (67)–(69), we obtain

$$\triangle z_k = -z_k\boldsymbol{\varepsilon}_k^H\triangle\mathbf{R_c}\mathbf{U_s}\mathbf{\Lambda_s}^{-1}\mathbf{q}_k \tag{70}$$

where $\boldsymbol{\varepsilon}_k^H = -p_k\underline{\mathbf{U}}_\mathbf{s}{}^{+}(\mathbf{W}_1 - z_k^*\mathbf{W}_2)$, with $\mathbf{W}_1 = [\mathbf{I}_{(K-1)\times(K-1)}, 0]$ and $\mathbf{W}_2 = [0, \mathbf{I}_{(K-1)\times(K-1)}]$.

Therefore, starting from (70) and using the statistics of $\triangle\mathbf{R_c}$, one can compute the mean-square-error (MSE) of $\triangle z_k$. However, due to the complex dependency of $\triangle\mathbf{R_c}$ on the data, the general expressions are quite complex. Thus, we give simplified expressions for the MSE of the frequency estimate in the case of a single exponential, which in our framework corresponds to the estimate of the time delay t_1 of one Dirac impulse.

Consider for simplicity the case of a periodic signal with period T. Let the data matrix $\mathbf{X}$ be of size $M \times N$, and let $N_t = M + N - 1$ be the total number of samples used for estimation. Assuming that the signal and noise are uncorrelated, the optimum performance is achieved when $N = N_t/3$ or $N = 2N_t/3$, resulting in the MSE of time delay estimation [7]

$$E\left\{\Delta t_1^2\right\} \approx \frac{1}{\omega_0^2}\frac{27}{4N_t^3}\frac{1}{\text{SNR}} \tag{71}$$

where $\omega_0 = 2\pi/T$. This is close to the CRB [20], given by

$$\text{CRB} = \frac{1}{\omega_0^2}\frac{6}{N_t^3}\frac{1}{\text{SNR}} \tag{72}$$

which indicates desirable numerical performance of the subspace-based approach. Similar conclusions can be reached for the two-source case [4], [7], while for the general results, we refer to [7] and [21].

B. Model Mismatch

In all the methods presented so far, we assumed that both a signal model and the model order are known *a priori*. This allowed us to select an appropriate reconstruction technique, as well as the size of the data matrix, to ensure that all the signal parameters are reliably estimated. In the presence of noise, a low-rank subspace property is destroyed and such an increased dimension of the solution set must be carefully dealt with.

In the problem of estimating the parameters of superimposed exponentials, the model order can be obtained from the number of dominant singular values of the noisy data matrix, which is a very good estimate, provided that the smallest singular value of the original, noiseless matrix is not dominated by the noise variance. However, for low values of SNR, it is often difficult to discriminate between small singular values corresponding to the signal from extraneous ones due to noise and, typically, only dominant signal components can be reliably estimated [19]. The problem becomes more involved in the case of piecewise polynomials and nonuniform splines, since one has to take a sufficient number of signal derivatives prior to constructing the data matrix. Such an approach obviously raises the following question: how can one reconstruct the signal in the case when neither the signal model nor the model order are known? While at this point we do not have a formal answer to this question, we would still like to point to one possible solution, namely, finding a piecewise constant approximation of the signal.

In order to explain the main idea behind such an approach, consider a periodic signal $x(t)$ of period T, filtered with a lowpass filter and sampled uniformly at a critical rate. In the case when the bandwidth of the filter is relatively low compared to the effective bandwidth of the original signal, one could expect that the Shannon interpolation formula would not yield a good approximation of the signal. Therefore, following the approach used for reconstruction of a stream of Diracs from a lowpass version, one can obtain a piecewise constant approximation $x_a(t)$ of the signal. This would potentially yield a better approximation, particularly for signals with discontinuities.

Assume, thus, that $x_a(t)$ contains M constant pieces. In this case, $x_a(t)$ is uniquely determined by a set of transition instants $\{t_m\}_{m=1}^M$ and the corresponding amplitudes $\{a_m\}_{m=1}^M$. The key is to consider the first derivative $x_d(t)$ of the piecewise constant approximation $x_a(t)$. By denoting $c_m = a_m - a_{m-1}$, the first derivative $x_d(t)$ is given by a periodic sum of weighted impulses

$$x_d(t) = \frac{d}{dt}x_a(t) = \sum_n\sum_{m=0}^{M-1} c_m\delta(t - t_m - nT). \tag{73}$$

The idea is to approximate the first derivative of $x(t)$ with the sum of weighted Diracs, instead of approximating the original signal $x(t)$. This can be done using the method developed in Section III and estimating the values of $\{t_m\}_{m=1}^M$ and $\{c_m\}_{m=1}^M$ from the following system:

$$jn\omega_0 X[n]e^{-sn\omega_0} = \sum_{m=0}^{M-1} c_m e^{-n\omega_0(s+jt_m)} \tag{74}$$

where $jn\omega_0 X[n]$ are the Fourier series coefficients corresponding to the first derivative of the original signal. Note that these coefficients are windowed by exponentially decaying terms $w[n] = e^{-sn\omega_0}$, in order to avoid ill-conditioning of the system (40). The same idea can be used in the case of aperiodic signals, where the samples corresponding to the first derivative can be computed from samples taken with the Gaussian kernel, as described in Section V-D.

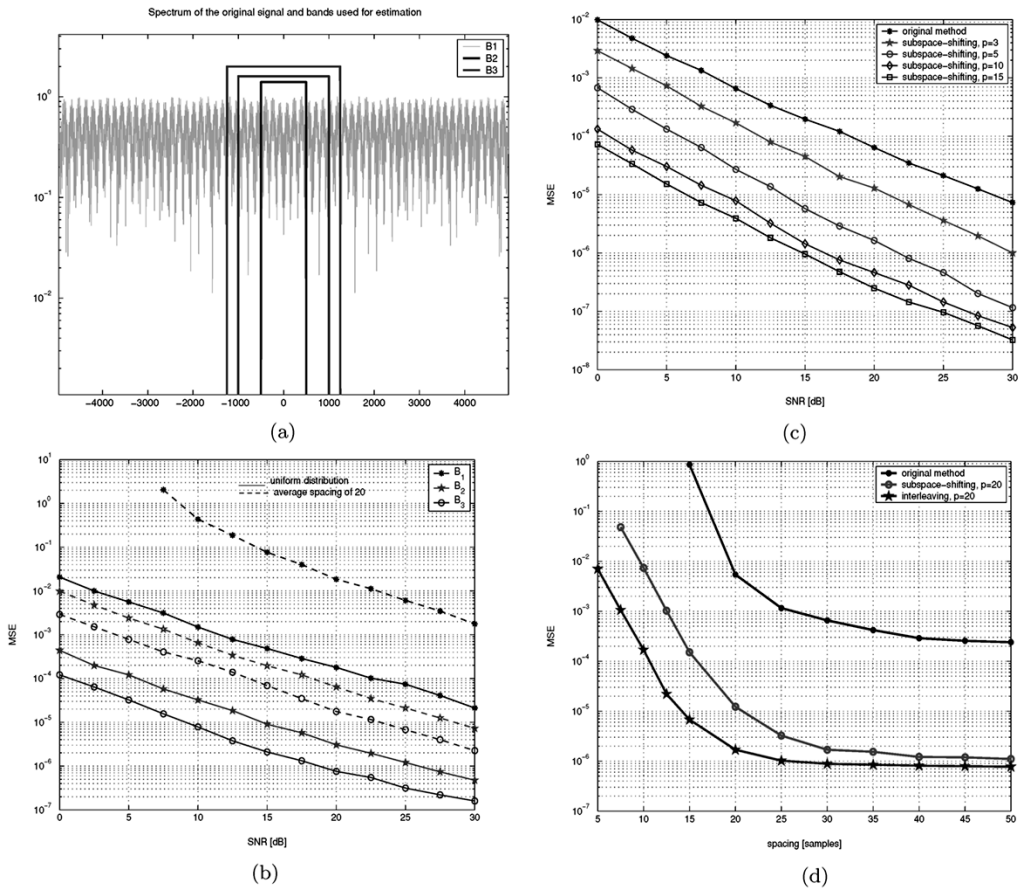

Fig. 2. Periodic stream of Diracs. A periodic signal of length 10000, made up of six weighted Diracs, is passed through a lowpass filter and sampled uniformly at the critical rate, determined by the filter bandwidth. We considered three different filter bandwidths: $B_1 = [-500\,500], B_2 = [-1000, 1000]$, and $B_3 = [-1250, 1250]$. (a) Magnitude of the signal spectrum and bands used for estimation. (b) MSE of location estimates versus SNR. Solid lines correspond to the case when the locations are randomly chosen according to a uniform distribution over the interval [1, 10 000], while dashed lines correspond to the MSE in the case when the average spacing between components is 20 (i.e., 0.2% of the signal period). (c) MSE of location estimates for the original method and the matrix-shifting method versus SNR. For the latter method, the error is plotted for different values of the shift parameter p. (d) MSE of location estimates versus average spacing between the components. The MSE of the original method is compared to the MSEs of the interleaving technique and the subspace-shifting approach.

VII. Simulation Results

In this section, we illustrate the performance of the developed schemes with simulation results. Experiments are done in discrete-time, with a very long block size in oder to simulate continuous time.

A. Case of Periodic Signals

We first consider a length 10 000 signal made up of a periodic stream of $K = 6$ weighted Diracs (Fig. 2), and analyze the following cases: the locations of Diracs are chosen randomly according to a uniform distribution over the interval [1, 10 000]; the first component is chosen randomly over [1, 10 000], while the spacing between the components is a Gaussian random variable with mean $m = 20$ and standard deviation $d = 1$. The signal is filtered with a lowpass filter, having one of the following bandwidths: $B_1 = [-500, 500], B_2 = [-1000, 1000]$, and $B_3 = [-1250, 1250]$, and in each case, a lowpass version of the signal is sampled uniformly at a critical rate. The spectrum of the signal and different frequency bands used for estimation are illustrated in Fig. 2(a). In Fig. 2(b), we plot the mean-square error (MSE) of the position estimates versus signal-to-noise (SNR) ratio. The error is computed as an average MSE over 100 different trials and normalized to the signal period. The results indicate that the performance of the method can be improved by increasing the bandwidth and estimating the parameters from a larger set of samples. Yet, such an improved performance is achieved at the expense of increased computational requirements, since the complexity of the reconstruction scheme using N_t samples is on the order of $\mathcal{O}(N_t^3)$ [6]. More importantly, the performance of the method degrades in the case when the average spacing between the Diracs is small compared to the signal period. For example, when the average spacing is 20 (i.e., 0.2% of the signal period) and SNR < 10 dB, it is no longer possible to reconstruct the signal using only the band B_1, and one should estimate the parameters from a larger signal subspace.

However, a more attractive solution to this problem is the one presented in Section III-D, where the resolution performance is improved by exploiting the shift-invariant subspace property in a different way, rather than by increasing a data set used for estimation. In Fig. 2(c), we compare the MSE obtained with the original method and the subspace-shifting method from Section III-D. We show the estimation performance for different

values of the parameter p, which determines the increase in the effective separation between the estimated components. Clearly, for all considered values of SNR, the latter approach results in much better numerical precision. Note that by increasing the value of p, the estimation accuracy improves, and in this particular case, the value of $p = 10$ already yields very good precision. Also note that such an approach does not increase the overall computational requirements, since the size of the data matrix has not changed. Still, p cannot be chosen arbitrarily large, since one may end up deleting too many rows/columns from the data matrix, which would result in degraded performance. In Fig. 2(d), we compare the estimation performance of the subspace-shifting approach and the interleaving method, also presented in Section III-D, for different values of the average spacing between the Diracs. The value of SNR used in this set of simulations is SNR $= 5$ dB. The results we obtained indicate that the interleaving technique yields better performance as the spacing between the components decreases. Also note that the numerical precision achievable by both techniques is by an order of magnitude better than the precision of the original method.

Next, we consider the case of a periodic nonuniform spline of period $T = 1000$ and degree $R = 1$, embedded in additive white Gaussian noise. The signal is filtered with a lowpass filter of bandwidth $B = [-50, 50]$ and a uniform set of 100 samples is taken from the lowpass version. In order to reconstruct the signal from the set of samples, we used the approach from Section IV-A, where the scaling parameter s is chosen to be $s = 0.015$. In particular, we first estimated the locations of the transition points, and then found the least-squares (LS) linear fit between each two transitions based on a set of samples. In Fig. 3, we show the noisy version (SNR $= 27$ dB) and the reconstructed signal, where the reconstruction error is MSE $= 0.0135$. While in this case we have obtained a good reconstruction of the original signal, in general, for $R > 1$, the method becomes more sensitive to noise, as already discussed in Section IV-A. In the following example, we consider a noisy piecewise linear signal (SNR $= 15$ dB) of length 1000, made up of seven pieces. The signal is passed through a lowpass filter of bandwidth $B = [-100\,100]$, and a set of 200 uniform samples is taken from the lowpass version. We used the approach from Section IV-B to reconstruct the signal, with the scaling parameter $s = 0.01$. As in the previous case, we first extracted the locations and weights of the discontinuities, and then computed the best LS linear fit between each two adjacent discontinuities. In Fig. 3(b), we show the noisy version and the reconstructed signal. The reconstruction error is MSE $= 0.015$; however, we should note that as SNR decreases, the method is less sensitive to noise than it was the case with nonuniform splines.

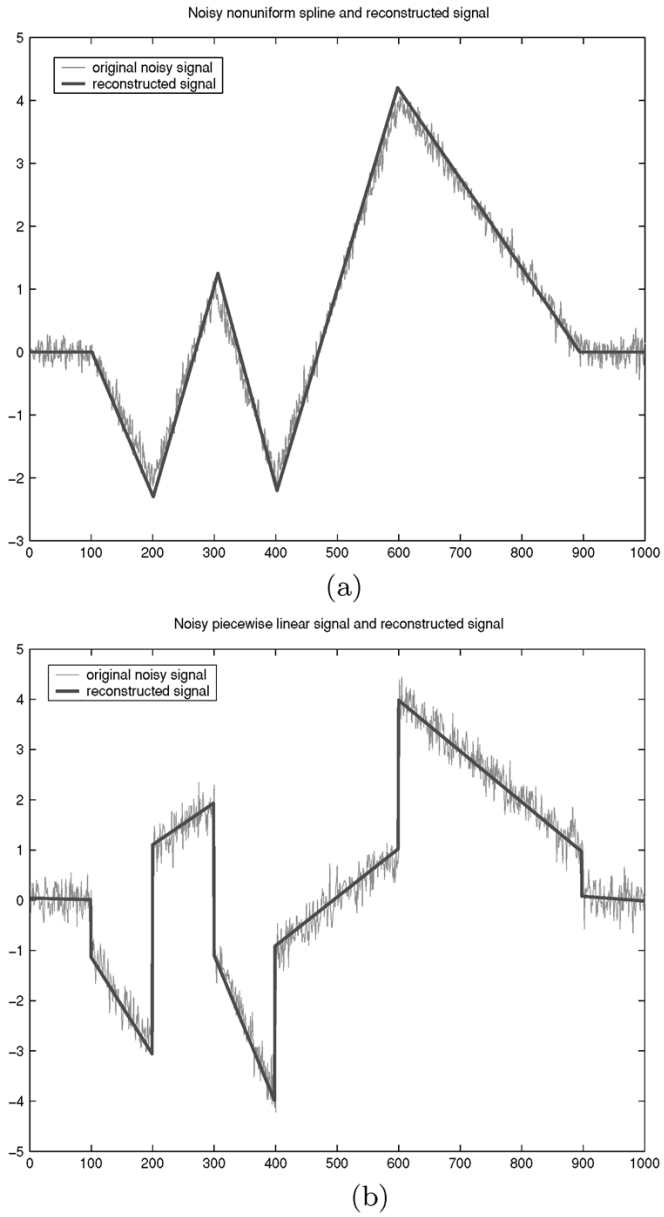

Fig. 3. Periodic nonuniform splines and piecewise polynomials. (a) Noisy nonuniform spline (SNR $= 27$ dB) of degree $R = 1$ and reconstructed signal. The signal is reconstructed with an error of MSE $= 0.0135$. The error is defined as MSE $= E\{(x_{\rm est} - x)^2\}/E\{x^2\}$, where $x_{\rm est}$ and x denote, respectively, the estimated signal and the original signal in one period. (b) Noisy piecewise linear signal (SNR $= 15$ dB) and reconstructed signal. MSE of reconstruction is MSE $= 0.015$.

B. Finite-Length Signals

In this set of simulations, we consider finite-length signals and evaluate the performance of the scheme based on a Gaussian sampling kernel. We first analyze the case of a length 1000 signal, made up of K weighted Diracs, where K takes on values between $K = 2$ and $K = 12$. We assume that the spacing between the components is a Gaussian random variable with mean $t_s = 60$ and standard deviation $d = 3$. The signal is filtered with a Gaussian kernel $h_\sigma(t) = e^{-t^2/2\sigma^2}$. In Fig. 4(a), we show the MSE (normalized to the length of the signal) of position estimates obtained using the method from Section V-A, in the case when $K = 6$ and $\sigma = 35$. The results are compared to the error obtained using the original method (i.e., no weighting of the data matrix). Note that by using the original method, it is possible to reconstruct the signal only for high values of SNR (i.e., SNR > 25 dB), whereas preconditioning of the data matrix $\mathbf{X}$ from (50) allows for significantly better estimation performance. Yet, in order to ensure a good performance of the algorithm, the width of the Gaussian kernel σ must be chosen care-

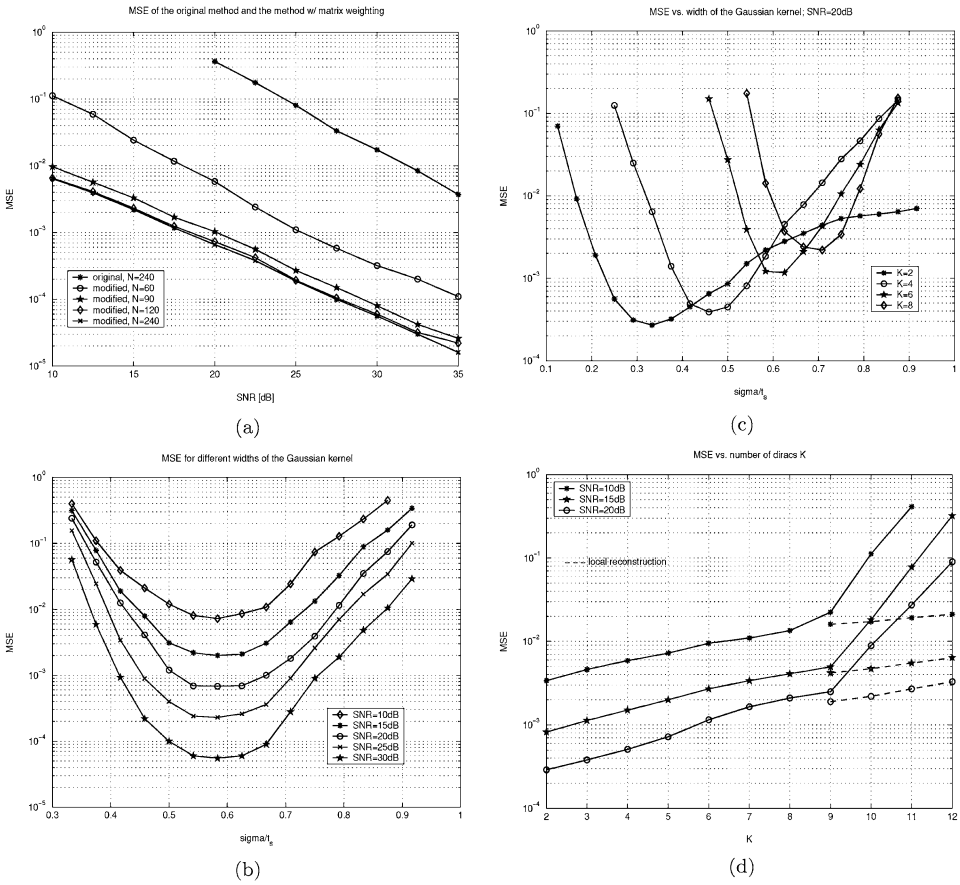

Fig. 4. Aperiodic stream of Diracs: We consider a signal of length 1000, made up of K weighted Diracs (K is varied between 2 and 12) and sampled with the Gaussian kernel $e^{-t^2/2\sigma^2}$. Average spacing between the components is assumed to be $t_s = 60$. (a) MSE of position estimates versus SNR, for $K = 6, \sigma = 35$. The MSE of the original method (i.e., with no weighting of the data matrix) is compared to the MSE obtained by the method from Section V-A. (b) MSE of reconstruction versus the width of the Gaussian kernel. The error is plotted as a function of the parameter σ/t_s (and is shown for different values of SNR), indicating a sensitivity of the method to the choice of the width σ. (c) MSE versus σ/t_s, for different values of K. (d) MSE versus number of Diracs K. For each value of K, we chose the optimal value of the kernel width σ, that is, the one which minimizes a reconstruction error. Dashed lines correspond to the MSE obtained by sampling the signal over two smaller time windows and finding local reconstruction in each window.

fully. This can be seen in Fig. 4(b) and (c), where we plot the reconstruction error as a function of the parameter σ/t_s, where t_s denotes the average spacing between the components. The error is plotted for different values of SNR [Fig. 4(b)], as well as different number of Diracs K [Fig. 4(c)]. The results indicate strong sensitivity to the choice of the parameter σ, specifically as K increases. In Fig. 4(d), we show the reconstruction error versus the number of Diracs K, while the parameter σ is chosen such that the MSE of reconstruction is minimized. Note that for all considered values of SNR, the method yields good performance for $K < 10$; however, when the number of Diracs further increases, the performance degrades significantly. The main reason for such a behavior is the following: as the number of components K increases, the time window where the signal must be sampled increases as well, and its duration is approximately given by Kt_s. Yet, the optimum width σ_{opt} of the sampling kernel is only a fraction of the average spacing t_s, for example, when $K = 8, \sigma_{\text{opt}} \approx 0.7t_s$ [see Fig. 4(c)]. Given the exponential decay of the Gaussian kernel, for large values of K, at each sampling instant we will obtain the information only about a limited number $(K_1 < K)$ of Diracs, which results in bad conditioning of the system. This makes the Gaussian scheme suitable mainly for local reconstruction, where the overall sampling window is adapted to the width of the kernel. This is illustrated in Fig. 4(d) as well, where we plot the MSE for $K \geq 9$, in the case when we sample the signal over two distinct time windows (of approximately equal duration), and in each window we perform local reconstruction. Such an approach clearly improves the performance of the original method.

In Fig. 5(a), we illustrate a noisy piecewise constant signal (SNR $= 25$ dB) of length 1000, made up of seven pieces. The signal is filtered with a Gaussian kernel with $\sigma = 75$, as shown in Fig. 5(b), and a set of 160 samples is taken from the filtered version shown in Fig. 5(c). The first derivative of the signal is computed according to (65), and the method from Section V-A is used for reconstruction. A reconstructed signal is illustrated in Fig. 5(d), where the reconstruction error is MSE $= 6 \cdot 10^{-3}$. In Fig. 5(e), we show the MSE of reconstruction versus SNR, for several different values of the number of samples N. The results indicate that the performance of the method improves as the number of samples increases. In this case, a very good reconstruction can be already obtained for $N = 120$, and by further increasing the number of samples, the performance does not significantly improve.

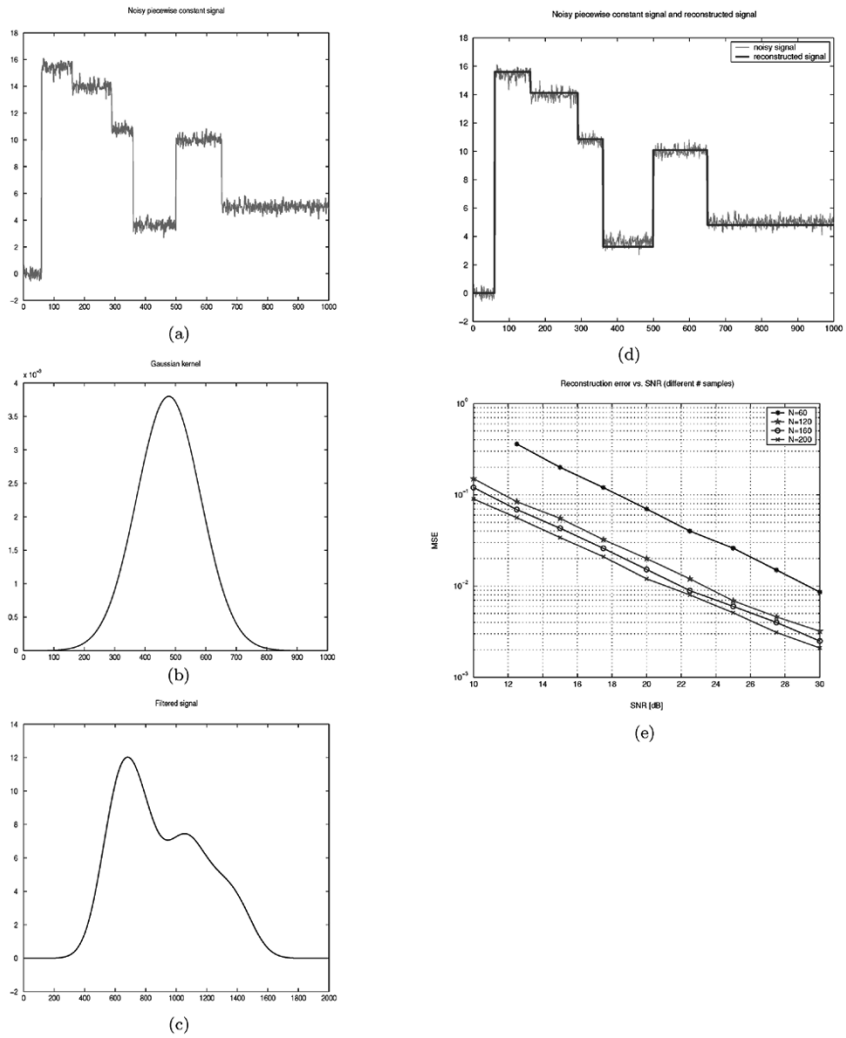

Fig. 5. Aperiodic piecewise constant signal. (a) Noisy piecewise constant signal. (b) Gaussian sampling kernel. (c) Filtered signal. (d) Noisy signal and reconstructed piecewise constant signal. The signal is reconstructed from $N = 160$ samples, with an error of MSE $= 6 \cdot 10^{-3}$. (e) Reconstruction error versus SNR for different number of samples. The error is defined as MSE $= E\{(x_{\text{est}} - x)^2\}/E\{x^2\}$.

C. Model Mismatch

In Figs. 6 and 7, we illustrate robustness of our schemes to model mismatch. We first consider a noisy signal made up of $K = 15$ weighted Diracs, where the locations are randomly chosen according to a uniform distribution over the interval [1, 1000], while the weights are i.i.d. zero mean Gaussian random variables with unit variance. The signal is filtered with a lowpass filter of bandwidth $[-100\ 100]$ and 200 uniform samples are taken from the lowpass version. We used the subspace method from Section III-B, where no prior knowledge of the model order is assumed. That is, the number of components is estimated as the number of dominant singular vectors of the corresponding data matrix. In Fig. 6(b), we show the reconstructed stream of pulses, and obviously, only the dominant pulses have been extracted.

We next consider the case of a periodic piecewise polynomial signal of degree $R = 3$ embedded in noise, where the signal model is not known in advance. The signal is lowpass filtered and a uniform set of 40 samples is taken from the lowpass version. Since we made no assumption on the signal model, we approximated the signal with a piecewise constant function with $M = 16$ pieces, as illustrated in Fig. 7(a). In Fig. 7(b), we show the noisless signal and the reconstructed lowpass version using Shannon interpolation formula, where the frequency band used for reconstruction is the same as in the previous case. Obviously, our approach yields better representation of discontinuities, which points to some robustness of our scheme to model mismatch.

VIII. Conclusion

In this paper, we have considered the sampling problem for signals of finite rate of innovation in the presence of noise. We have revisited some of the results for deterministic, noiseless signals [24] and developed more robust methods that improve conditioning of the original schemes and allow for much better numerical performance. We specifically focused on the case of a stream of Diracs and developed a subspace framework to signal reconstruction [16], [17], which provides an elegant and robust solution to the sampling problem. For classes of periodic signals, such as piecewise polynomials and nonuniform splines, we proposed novel algebraic solutions that use proper windowing

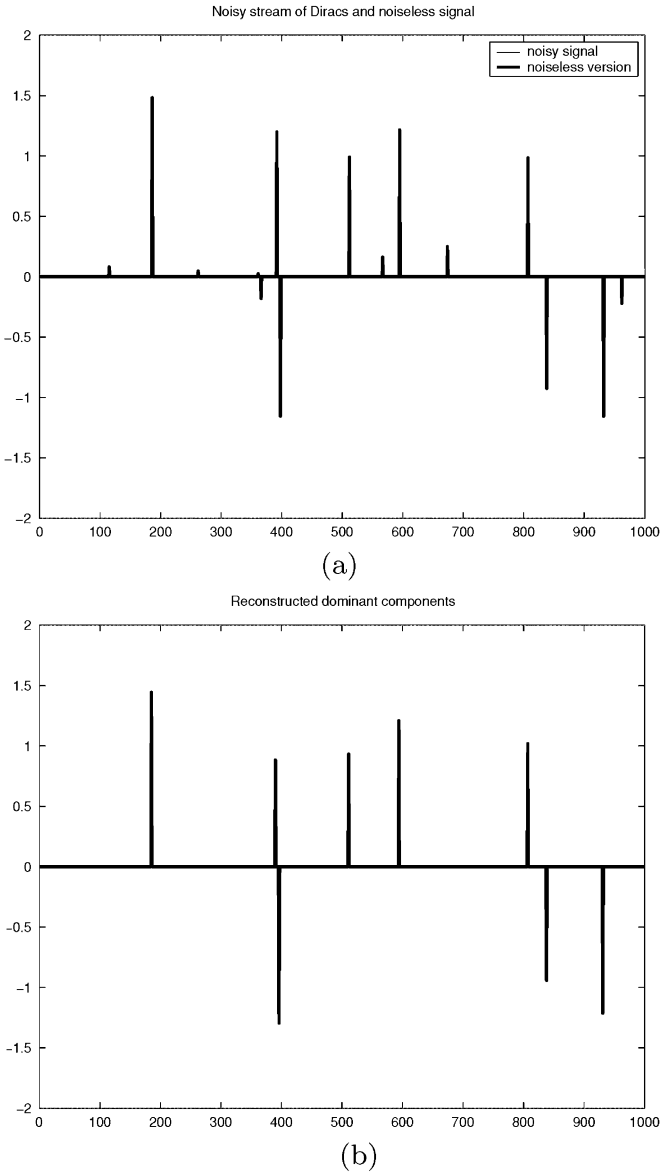

Fig. 6. Model mismatch: Unknown model order. (a) Signal made up of 15 weighted Dirac impulses, with eight impulses being dominant. (b) Reconstructed dominant components.

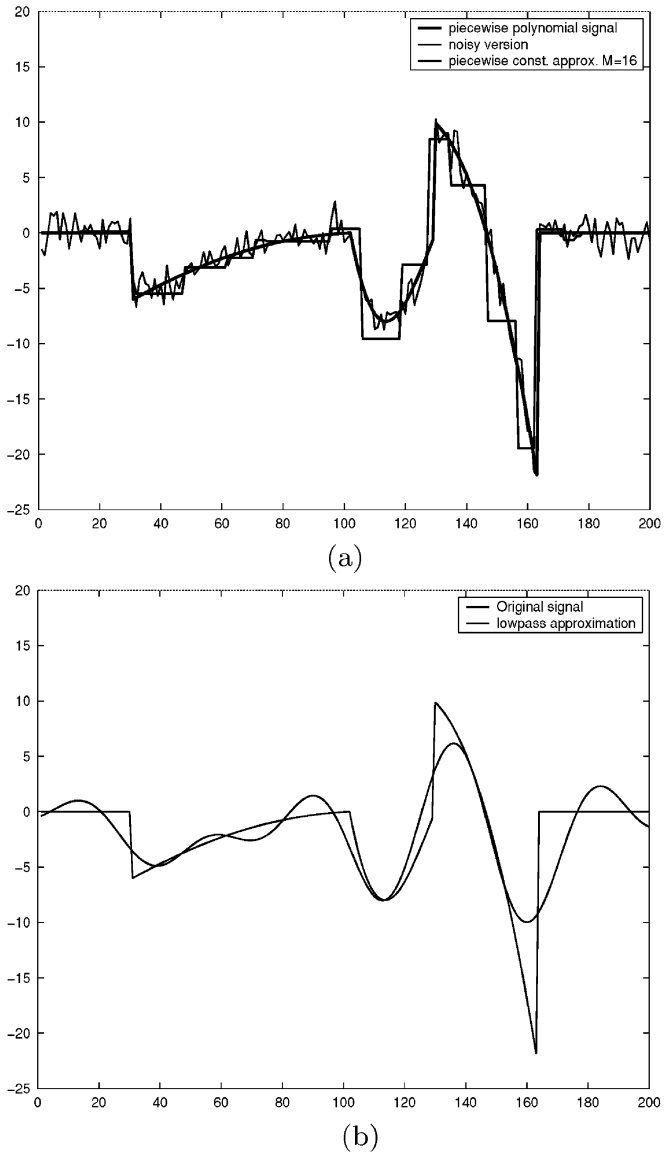

Fig. 7. Model mismatch: Unknown signal model. (a) Piecewise polynomial signal, noisy version, and the piecewise constant approximation. (b) Original signal and reconstructed lowpass version using Shannon's interpolation formula.

and solve the problem in the Laplace domain. While some of the tools we used were borrowed from spectral analysis [19], our framework extends classic spectral estimation techniques and allows for solving more general classes of nonlinear estimation problems. Furthermore, we have developed techniques that improve resolution capabilities of the existing spectral estimation schemes in the case when the signal contains closely spaced components. We have also considered finite-length signals, and proposed improved schemes based on a Gaussian sampling kernel and weighting of the data matrix. Both the numerical analysis and simulation results indicate desirable properties of the proposed methods, particularly for classes of signals that contain discontinuities. Applications of our framework can be found in signal and image processing [3], [12], [13], spectral estimation, and communications [14]. The results presented so far raise several questions for future research. Some of the interesting topics include developing a formal framework for analyzing the problem of model mismatch and investigating local reconstruction schemes for finite-length signals.

ACKNOWLEDGMENT

The authors would like to thank Dr. L. Sbaiz and Prof. U. Mitra for their valuable comments.

REFERENCES

[1] A. Aldroubi and K. Grochenig, "Non-uniform sampling in shift-invariant spaces," *SIAM Rev.*, vol. 43, pp. 585–620, 2001.

[2] R. E. Blahut, *Theory and Practice of Error Control Codes*. Reading, MA: Addison-Wesley, 1983.

[3] A. Chebira, P. L. Dragotti, L. Sbaiz, and M. Vetterli, "Sampling and interpolation of the plenoptic function," in *Proc. IEEE Int. Conf. Image Processing*, vol. 2, Sep. 2003, pp. 917–920.

[4] H. Clergeot, S. Tressens, and A. Ouamri, "Performance of high-resolution frequencies estimation methods compared to the Cramer-Rao bounds," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 11, pp. 1703–1720, Nov. 1989.

[5] P. J. Davis, *Interpolation and Approximation*. New York: Dover, 1975.
[6] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD: Johns Hopkins Univ. Press, 1989.
[7] Y. Hua and T. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 5, pp. 814–824, May 1990.
[8] ——, "On SVD for estimating generalized eigenvalues of singular matrix pencil in noise," *IEEE Trans. Signal Process.*, vol. 39, no. 4, pp. 892–900, Apr. 1991.
[9] M. Kristensson, M. Jansson, and B. Ottersten, "Further results and insights on subspace based sinusoidal frequency estimation," *IEEE Trans. Signal Process.*, vol. 49, no. 12, pp. 2962–2974, Dec. 2001.
[10] H. B. Lee, "Eigenvalues and eigenvectors of covariance matrices for signals closely spaced in frequency," *IEEE Trans. Signal Process.*, vol. 40, no. 10, pp. 2518–2535, Oct. 1992.
[11] I. Maravic, "Sampling methods for parametric non-bandlimited signals: Extensions and applications," Ph.D. dissertation, Dept. Comput. Commun. Sci., Swiss Federal Inst. Technol., Lausanne, Switzerland, 2004.
[12] I. Maravic and M. Vetterli, "Exact sampling results for classes of parametric nonbandlimited 2-D signals," *IEEE Trans. Signal Process.*, vol. 5, no. 1, pp. 175–189, Jan. 2004.
[13] ——, "A sampling theorem for the Radon transform of finite complexity objects," in *Proc. IEEE ICASSP*, vol. 2, May 2002, pp. 1197–1200.
[14] I. Maravic, M. Vetterli, and K. Ramchandran, "Channel estimation and synchronization with sub-Nyquist sampling and application to UWB systems," in *Proc. IEEE ISCAS*, vol. 5, May 2004, pp. 381–384.
[15] B. Rao, "Sensitivity analysis of state space methods in spectral estimation," presented at the IEEE ICASSP, Apr. 1987.
[16] B. D. Rao and K. S. Arun, "Model based processing of signals: A state space approach," *Proc. IEEE*, vol. 80, no. 2, pp. 283–309, Feb. 1992.
[17] R. Roy and T. Kailath, "ESPRIT estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, Jul. 1989.
[18] P. Marziliano, "Sampling innovations,", Dept. Comput. Commun. Sci., Swiss Federal Inst. Technol., Lausanne, Switzerland, 2001.
[19] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Upper Saddle River, NJ: Prentice-Hall, 2000.
[20] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood and Cramer-Rao bound," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 5, pp. 720–741, May 1989.
[21] ——, "Performance comparison of subspace rotation and MUSIC methods for direction estimation," *IEEE Trans. Signal Process.*, vol. 39, no. 2, pp. 446–453, Feb. 1991.
[22] A. L. Swindlehurst, "Time delay and spatial signature estimation using known asynchronous signals," *IEEE Trans. Signal Process.*, vol. 46, no. 2, pp. 449–462, Feb. 1998.
[23] M. Unser, "Sampling-50 years after Shannon," *Proc. IEEE*, vol. 88, no. 4, pp. 569–587, Apr. 2000.
[24] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. Signal Process.*, vol. 50, no. 6, pp. 1417–1428, Jun. 2002.

Irena Maravić received the B.S. degree from the Department of Electrical Engineering, University of Belgrade, Belgrade, Yugoslavia, in 1997, the M.S. degree from the California Institute of Technology (Caltech), Pasadena, in 2000, and the Ph.D. degree from the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, in 2004.

From 1997 to 1999, she was a member of academic staff at the University of Belgrade. From 1999 to 2001, she was with the Optical Information Processing Laboratory, Caltech. In 2001, she joined the Laboratory for Audio-Visual Communications at EPFL, where she worked on her Ph.D. degree in the field of digital signal processing. Currently, she is with the European Molecular Biology Laboratory, Heidelberg, Germany. Her research interests include model-based signal analysis and algorithms, multidimensional signal processing, bioinformatics, biomedical signal processing, and signal processing for communications.

Martin Vetterli (F'95) received the Dipl. El.-Ing. degree from ETH Zürich (ETHZ), Zürich, Switzerland, in 1981, the M.S. degree from Stanford University, Stanford, CA, in 1982 and the Doctorat ès Science degree the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, in 1986.

He was a Research Assistant at Stanford and EPFL and has worked for Siemens and AT&T Bell Laboratories. In 1986, he joined Columbia University, New York, NY, where he was an Associate Professor of electrical engineering and co-director of the Image and Advanced Television Laboratory. In 1993, he joined the University of California, Berkeley, where he was a Professor with the Department of Electrical Engineering and Computer Sciences until 1997 and now holds an Adjunct Professor position. Since 1995, he has been a Professor of communication systems at EPFL, where he chaired the Communications Systems Division from 1996 to 1997 and was head of the Audio-Visual Communications Laboratory. He held visiting positions at ETHZ in 1990 and Stanford in 1998. He is also on the editorial boards of *Annals of Telecommunications*, *Applied and Computational Harmonic Analysis*, and *The Journal of Fourier Analysis and Applications*. His research interests include wavelets, signal processing for communications, computational complexity, image/video processing and compression, and distributed signal processing and communications. He is the co-author, with J. Kovacevic, of the book *Wavelets and Subband Coding* (Englewood Cliffs, NJ: Prentice-Hall, 1995). He has published about 90 journal papers on a variety of topics in signal/image processing and communications and holds seven patents.

Dr. Vetterli is a member of SIAM and was the Area Editor for Speech, Image, Video, and Signal Processing of the IEEE TRANSACTIONS ON COMMUNICATIONS. He received the Best Paper Award of EURASIP in 1984 for his paper on multidimensional subband coding, the Research Prize of the Brown Bovery Corporation (Switzerland) in 1986 for his doctoral thesis, and the IEEE Signal Processing Society's Senior Award in 1991 and in 1996 (for papers with D. LeGall and K. Ramchandran, respectively). He received the Swiss National Latsis Prize in 1996, the SPIE Presidential award in 1999, and the IEEE Signal Processing Technical Achievement Award in 2001. He is a member of the Swiss Council on Science and Technology. Since 2001, he has directed the National Competence Center in Research on mobile information and communication systems (www.terminodes.org). He was a plenary speaker at various conferences (e.g., 1992 IEEE ICASSP).