Sampling Signals With Finite Rate of Innovation

Martin Vetterli, *Fellow, IEEE*, Pina Marziliano, and Thierry Blu, *Member, IEEE*

***Abstract*—Consider classes of signals that have a finite number of degrees of freedom per unit of time and call this number the rate of innovation. Examples of signals with a finite rate of innovation include streams of Diracs (e.g., the Poisson process), nonuniform splines, and piecewise polynomials.**

Even though these signals are not bandlimited, we show that they can be sampled uniformly at (or above) the rate of innovation using an appropriate kernel and then be perfectly reconstructed. Thus, we prove sampling theorems for classes of signals and kernels that generalize the classic "bandlimited and sinc kernel" case. In particular, we show how to sample and reconstruct periodic and finite-length streams of Diracs, nonuniform splines, and piecewise polynomials using sinc and Gaussian kernels. For infinite-length signals with finite local rate of innovation, we show local sampling and reconstruction based on spline kernels.

The key in all constructions is to identify the innovative part of a signal (e.g., time instants and weights of Diracs) using an annihilating or locator filter: a device well known in spectral analysis and error-correction coding. This leads to standard computational procedures for solving the sampling problem, which we show through experimental results.

Applications of these new sampling results can be found in signal processing, communications systems, and biological systems.

***Index Terms*—Analog-to-digital conversion, annihilating filters, generalized sampling, nonbandlimited signals, nonuniform splines, piecewise polynomials, poisson processes, sampling.**

I. Introduction

Most continuous-time phenomena can only be seen through sampling the continuous-time waveform, and typically, the sampling is uniform. Very often, instead of the waveform itself, one has only access to a smoothed or filtered version of it. This may be due to the physical set up of the measurement or may be by design.

Calling $x(t)$ the original waveform, its filtered version is $y(t) = x(t) * \tilde{h}(t)$, where $\tilde{h}(t) = h(-t)$ is the convolution kernel. Then, uniform sampling with a sampling interval T leads to samples $y(nT)$ given by

$$y(nT) = \langle h(t-nT), x(t)\rangle = \int_{-\infty}^{\infty} h(t-nT)\, x(t)\, dt. \quad (1)$$

Manuscript received May 16, 2001; revised February 11, 2002. The work of P. Marziliano was supported by the Swiss National Science Foundation. The associate editor coordinating the review of this paper and approving it for publication was Dr. Paulo J. S. G. Ferreira.

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

P. Marziliano was with the Laboratory for Audio Visual Communication (LCAV), Swiss Federal Institute of Technology, Lausanne, Switzerland. She is now with Genista Corporation, Genimedia SA, Lausanne, Switzerland (e-mail: pmarziliano@genimedia.com).

T. Blu is with the Biomedical Imaging Group, DMT, Swiss Federal Institute of Technology, Lausanne, Switzerland (e-mail: thierry.blu@epfl.ch).

Publisher Item Identifier S 1053-587X(02)04392-1.

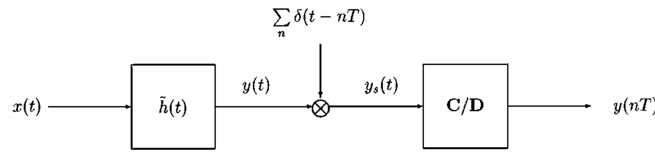

Fig. 1. Sampling setup: $x(t)$ is the continuous-time signal; $\tilde{h}(t) = h(-t)$ is the smoothing kernel; $y(t)$ is the filtered signal; T is the sampling interval; $y_s(t)$ is the sampled version of $y(t)$; and $y(nT)$, $n \in \mathbb{Z}$ are the sample values. The box C/D stands for continuous-to-discrete transformation and corresponds to reading out the sample values $y(nT)$ from $y_s(t)$.

The intermediate signal $y_s(t)$ corresponding to an idealized sampling is given by

$$y_s(t) = \sum_{n\in\mathbb{Z}} y(nT)\, \delta(t-nT). \quad (2)$$

This setup is shown in Fig. 1.

When no smoothing kernel is used, we simply have $y(nT) = x(nT)$, which is equivalent to (1) with $h(t) = \delta(t)$. This simple model for having access to the continuous-time world is typical for acquisition devices in many areas of science and technology, including scientific measurements, medical and biological signal processing, and analog-to-digital converters.

The key question is, of course, if the samples $y(nT)$ are a faithful representation of the original signal $x(t)$. If so, how can we reconstruct $x(t)$ from $y(nT)$, and if not, what approximation do we get based on the samples $y(nT)$? This question is at the heart of signal processing, and the dominant result is the well-known sampling theorem of Whittaker *et al.*, which states that if $x(t)$ is bandlimited, or $X(\omega) = 0, |\omega| > \omega_m$, then samples $x(nT)$ with $T \leq \pi/\omega_m$ are sufficient to reconstruct $x(t)$ [5], [15], [18]. Calling B, which is the bandwidth of $x(t)$ in cycles per second ($B = 2\omega_m/2\pi$), we see that $B = 1/T$ sample/s are a sufficient representation of $x(t)$. The reconstruction formula is given by

$$x(t) = \sum_{n=-\infty}^{\infty} x(nT)\, \mathrm{sinc}\left(\frac{t}{T} - n\right) \quad (3)$$

where $\mathrm{sinc}(t) = \sin(\pi t)/\pi t$. If $x(t)$ is not bandlimited, convolution with $\tilde{h}(t) = \mathrm{sinc}(t/T)$ (an ideal lowpass filter with support $[-\pi/T, \pi/T]$) allows us to apply sampling and reconstruction of $y(t)$, which is the lowpass approximation of $x(t)$. This restriction of $X(\omega)$ to the interval $[-\pi/T, \pi/T]$ provides the best approximation in the least squares sense of $x(t)$ in the sinc space [18].

A possible interpretation of the interpolation formula (3) is the following. Any real bandlimited signal can be seen as having $1/T$ degrees of freedom per unit of time, which is the number of samples per unit of time that specify it. In the present paper, this number of degrees of freedom per unit of time is called the *rate* *of innovation* of a signal[1] and is denoted by ρ. In the bandlimited case above, the rate of innovation is $\rho = 1/T = \omega_m/\pi$.

In the sequel, we are interested in signals that have a *finite rate of innovation*, either on intervals or on average. Take a Poisson process, which generates Diracs with independent and identically distributed (i.i.d.) interarrival times, the distribution being exponential with probability density function $\mu e^{-\mu t}$. The expected interarrival time is given by $1/\mu$. Thus, the rate of innovation is μ since, on average, μ real numbers per unit of time fully describe the process.

Given a signal with a finite rate of innovation, it seems attractive to sample it with a rate of ρ samples per unit of time. We know it will work with bandlimited signals, but will it work with a larger class of signals? Thus, the natural questions to pursue are the following.

1) What classes of signals of finite rate of innovation can be sampled uniquely, in particular, using uniform sampling?
2) What kernels $h(t)$ allow for such sampling schemes?
3) What algorithms allow the reconstruction of the signal based on the samples?

In the present paper, we concentrate on streams of Diracs, nonuniform splines, and piecewise polynomials. These are signal classes for which we are able to derive exact sampling theorems under certain conditions. The kernels involved are the sinc, the Gaussian, and the spline kernels. The algorithms, although they are more complex than the standard sinc sampling of bandlimited signals, are still reasonable (structured linear systems) but also often involve root finding.

As will become apparent, some of the techniques we use in this paper are borrowed from spectral estimation [16] as well as error-correction coding [1]. In particular, we make use of the annihilating filter method (see Appendices A and B), which is standard in high-resolution spectral estimation. The usefulness of this method in a signal processing context has been previously pointed out by Ferreira and Vieira [24] and applied to error detection and correction in bandlimited images [3]. In array signal processing, retrieval of time of arrival has been studied, which is related to one of our problems. In that case, a statistical framework using multiple sensors is derived (see, for example, [14] and [25]) and an estimation problem is solved. In our case, we assume deterministic signals and derive exact sampling formulas.

The outline of the paper is as follows. Section II formally defines signal classes with finite rate of innovation treated in the sequel. Section III considers periodic signals in continuous time and derives sampling theorems for streams of Diracs,[2] nonuniform splines, and piecewise polynomials. These types of signals have a finite number of degrees of freedom, and a sampling rate that is sufficiently high to capture these degrees of freedom, together with appropriate sampling kernels, allows perfect reconstruction. Section IV addresses the sampling of finite-length signals having a finite number of degrees of freedom using infinitely supported kernels like the sinc kernel and the Gaussian kernel. Again, if the critical number of samples is taken, we can derive sampling theorems. Section V concentrates on local reconstruction schemes. Given that the local rate of innovation is bounded, local reconstruction is possible, using, for example, spline kernels. In the Appendix, we review the "annihilating filter" method from spectral analysis and error-correction coding with an extension to multiple zeros. This method will be used in several of the proofs in the paper.

II. Signals With Finite Rate of Innovation

In the Introduction, we informally discussed the intuitive notion of signals with finite rate of innovation. Let us introduce more precisely sets of signals having a finite rate of innovation. Consider a known function $\varphi(t)$ and signals of the form

$$x(t) = \sum_{n\in\mathbb{Z}} c_n \, \varphi\left(\frac{t-nT}{T}\right) \tag{4}$$

of which bandlimited signals [see (3)] are a particular case. Clearly, the rate of innovation is $\rho = 1/T$. There are many examples of such signals, for example, when $\varphi(t)$ is a scaling function in a wavelet multiresolution framework [8], [20] and in approximation theory [17].

A more general case appears when we allow arbitrary shifts or

$$x(t) = \sum_{n\in\mathbb{Z}} c_n \, \varphi\left(\frac{t-t_n}{T}\right). \tag{5}$$

For example, when $c_n = 1$, $\varphi(t) = \delta(t)$, $T = 1/\mu$, and $t_n - t_{n-1}$ are i.i.d. with exponential density, then we have the Poisson process of rate μ.

Allowing a set of functions $\{\varphi_r(t)\}_{r=0,\dots,R}$ and arbitrary shifts, we obtain

$$x(t) = \sum_{n\in\mathbb{Z}} \sum_{r=0}^{R} c_{nr} \, \varphi_r\left(\frac{t-t_n}{T}\right). \tag{6}$$

Finally, we will be considering signals that are built using a noninnovative part, for example, a polynomial and an innovative part such as (5) for nonuniform splines or, more generally, (6) for piecewise polynomials of degree R and arbitrary intervals (in that case, φ and φ_r are one-sided power functions $t_+^R = \max(t,0)^R$).

Assuming the functions $\varphi_r(t)$ are known, it is clear that the only degrees of freedom in a signal $x(t)$ as in (6) are the time instants t_n and the coefficients c_{nr}. Introducing a counting function $C_x(t_a, t_b)$ that counts the number of degrees of freedom in $x(t)$ over the interval $[t_a, t_b]$, we can define the rate of innovation ρ as

$$\rho = \lim_{\tau\to\infty} \frac{1}{\tau} C_x\left(-\frac{\tau}{2}, \frac{\tau}{2}\right). \tag{7}$$

Definition 1: A signal with a finite rate of innovation is a signal whose parametric representation is given in (5) and (6) and with a finite ρ, as defined in (7).

If we consider finite length or periodic signals of length τ, then the number of degrees of freedom is finite, and the rate of innovation is $1/\tau \, C_x(0, \tau)$. Note that in the uniform case given in (4), the instants, being predefined, are not degrees of freedom.

[1]This is defined in Section II and is different from the rate used in rate-distortion theory [2]. Here, rate corresponds to a degree of freedom that is specified by real numbers. In rate-distortion theory, the rate corresponds to bits used in a discrete approximation of a continuously valued signal.

[2]We are using the Dirac distribution and all of the operations, including derivatives, must be understood in the distribution sense.

One can also define a *local* rate of innovation with respect to a moving window of size τ. Given a window of size τ, the local rate of innovation at time t is

$$\rho_\tau(t) = \frac{1}{\tau} C_x \left(t - \frac{\tau}{2}, t + \frac{\tau}{2}\right). \tag{8}$$

In this case, one is often interested in the maximal local rate or $\rho_{max}(\tau)$

$$\rho_{max}(\tau) = \max_{t \in \mathbb{R}} \rho_\tau(t). \tag{9}$$

As $\tau \to \infty$, $\rho_{max}(\tau)$ tends to ρ. To illustrate the differences between ρ and ρ_{max}, consider again the Poisson process with expected interarrival time $1/\mu$. The rate of innovation ρ is given by μ. However, for any finite τ, there is no bound on $\rho_{max}(\tau)$.

The reason for introducing the rate of innovation ρ is that one hopes to be able to "measure" a signal by taking ρ samples per unit of time and be able to reconstruct it. We know this to be true in the uniform case given in (4). The main contribution of this paper is to show that it is also possible for many cases of interest in the more general cases given by (5) and (6).

III. Periodic Continuous-Time Case

In this section, we consider τ–periodic signals that are either made of Diracs or polynomial pieces. The natural representation of such signals is through Fourier series

$$x(t) = \sum_{m \in \mathbb{Z}} X[m]\, e^{i(2\pi m t/\tau)}. \tag{10}$$

We show that such signals can be recovered uniquely from their projection onto a subspace of appropriate dimension. The subspace's dimension equals the number of degrees of freedom of the signal, and we show how to recover the signal from its projection. One choice of the subspace is the lowpass approximation of the signal that leads to a sampling theorem for periodic stream of Diracs, nonuniform splines, derivatives of Diracs, and piecewise polynomials.

A. Stream of Diracs

Consider a stream of K Diracs periodized with period τ, $x(t) = \sum_{m \in \mathbb{Z}} c_n\, \delta(t - t_n)$, where $t_{n+K} = t_n + \tau$, and $c_{n+K} = c_n, \forall n \in \mathbb{Z}$. This signal has $2K$ degrees of freedom per period, and thus, the rate of innovation is

$$\rho = \frac{2K}{\tau}. \tag{11}$$

The periodic stream of Diracs can be rewritten as

$$\begin{aligned} x(t) &= \sum_{k=0}^{K-1} c_k \sum_{n \in \mathbb{Z}} \delta(t - t_k - n\tau) \\ &= \sum_{k=0}^{K-1} c_k \frac{1}{\tau} \sum_{m \in \mathbb{Z}} e^{i(2\pi m (t - t_k)/\tau)} \\ &\quad \text{from Poisson's summation formula} \\ &= \sum_{m \in \mathbb{Z}} \underbrace{\frac{1}{\tau} \left(\sum_{k=0}^{K-1} c_k\, e^{-i(2\pi m t_k/\tau)} \right)}_{X[m]} e^{i(2\pi m t/\tau)}. \end{aligned} \tag{12}$$

The Fourier series coefficients $X[m]$ are thus given by

$$X[m] = \frac{1}{\tau} \sum_{k=0}^{K-1} c_k\, e^{-i(2\pi m t_k/\tau)}, \quad m \in \mathbb{Z} \tag{13}$$

that is, the linear combination of K complex exponentials. Consider now a finite Fourier series $A[m], m = 0, \ldots, K$ with z-transform

$$A(z) = \sum_{m=0}^{K} A[m]\, z^{-m} \tag{14}$$

and having K zeros at $u_k = e^{-i(2\pi t_k/\tau)}$, that is

$$A(z) = \prod_{k=0}^{K-1} \left(1 - e^{-i(2\pi t_k/\tau)}\, z^{-1}\right). \tag{15}$$

Note that $A[m]$ is the convolution of K elementary filters with coefficients $[1, -e^{-i(2\pi t_k/\tau)}], k = 0, \ldots, K-1$ and that the convolution of such a filter with the exponential $\{e^{-i(2\pi m t_k/\tau)}\}_{m \in \mathbb{Z}}$ is zero

$$\begin{aligned} &\left[1, -e^{-i(2\pi t_k/\tau)}\right] \\ &\quad * \left[\ldots, e^{i(2\pi t_k/\tau)}, 1, e^{-i(2\pi t_k/\tau)}, e^{-i(4\pi t_k/\tau)}, \ldots\right] = 0. \end{aligned} \tag{16}$$

Therefore, because $X[m]$ is the sum of K exponentials and each being zeroed out by one of the roots of $A[m]$, it follows that

$$A[m] * X[m] = 0. \tag{17}$$

The filter $A[m]$ is thus called an annihilating filter since it annihilates K exponentials [16]. A more detailed discussion of annihilating filters is given in Appendix A. In error-correction coding, it is called the error locator polynomial [1]. The periodic time-domain function $a(t)$ is obtained by inversion of the Fourier series, or equivalently, by evaluating $A(z)$ for $z = e^{-i(2\pi t/\tau)}$. Thus

$$a(t) = A(z)|_{z = e^{-i(2\pi t/\tau)}} = \prod_{k=0}^{K-1} \left(1 - e^{-i(2\pi (t_k - t)/\tau)}\right) \tag{18}$$

that is, $a(t)$ has zeros at $t = t_k, k \in \mathbb{Z}$. Thus, in the time domain, we have the equivalent of (17) as $a(t) \cdot x(t) = 0$.

Note that $A[m]$ with $K+1$ nonzero coefficients for $m = 0, \ldots, K$ and that $A[0] = 1$ is unique for $x(t)$ given by (12) with $c_k \neq 0, k \in \mathbb{Z}$ and distinct locations $t_k \neq t_\ell, k \neq \ell$. This follows from the uniqueness of the set of roots in (15). Further, define the sinc function of bandwidth $[-B\pi, B\pi]$, where $B \in \mathbb{R}_+$ as $h_B(t)$, that is, $h_B(t) = B\,\mathrm{sinc}(Bt)$. We are now ready to prove a sampling result for periodic streams of Diracs.

Theorem 1: Consider $x(t)$, which is a periodic stream of Diracs of period τ with K Diracs of weight $\{c_k\}_{k=0}^{K-1}$ and at location $\{t_k\}_{k=0}^{K-1}$, as in (12). Take as a sampling kernel $h_B(t) = B\,\mathrm{sinc}(Bt)$, where B is chosen such that it is greater or equal to the rate of innovation ρ given by (11) and sample $(h_B * x)(t)$ at N uniform locations $t = nT, n = 0, \ldots, N-1$, where $N \geq 2M + 1$, and $M = \lfloor B\tau/2 \rfloor$. Then, the samples

$$y_n = \langle h_B(t - nT), x(t) \rangle, \quad n = 0, \ldots, N-1 \tag{19}$$

are a sufficient characterization of $x(t)$.

Proof: We first show how to obtain the Fourier series coefficients $X[m], m = -M, \ldots, M$ from the samples $y_n, n = 0, \ldots, N-1$. Based on these Fourier series coefficients, we then show how to obtain the annihilating filter $A[m]$, which leads to the locations of the Diracs. Finally, the weights of the Diracs are found, thus specifying uniquely $x(t)$.

1) Finding $X[m], |m| \le M$ from $y_n, n = 0, \ldots, N-1$.

Using (10) in (19), we have

$$y_n = \sum_m X[m] \langle h_B(t-nT), e^{i(2\pi mt/\tau)} \rangle \tag{20}$$

$$= \sum_m X[m]\, H_B\left(\frac{2\pi m}{\tau}\right) e^{i(2\pi mnT/\tau)} \tag{21}$$

$$= \sum_{m=-M}^{M} X[m]\, e^{i(2\pi mnT/\tau)} \tag{22}$$

where H_B is the Fourier transform of $h_B(t)$. This system of equations is not invertible when $\tau/T = p/q$ with $p < N$, where $p, q \in \mathbb{N}$. In all other cases, the N equations are of maximal rank $2M+1$. When T is a divisor of $\tau (q=1)$, this is simply the inverse discrete-time Fourier transform (IDTFT) of $X[m]$.

2) Finding the coefficients of the filter $A[m]$ that annihilates $X[m], m \in [-M, M]$.

We need to solve (17) for $A[m], m = 1, \ldots, K$, given $X[m], m = -M, \ldots, M$. For example, pick $K = 3$, $B\tau = 2K$, and $N = 7$ to illustrate. Discarding $X[-3]$, we have that (17) is equivalent to

$$\begin{bmatrix} X[0] & X[-1] & X[-2] \\ X[1] & X[0] & X[-1] \\ X[2] & X[1] & X[0] \end{bmatrix} \cdot \begin{bmatrix} A[1] \\ A[2] \\ A[3] \end{bmatrix} = - \begin{bmatrix} X[1] \\ X[2] \\ X[3] \end{bmatrix}. \tag{23}$$

In general, at critical sampling, we have a system given by

$$\begin{bmatrix} X[0] & X[-1] & \cdots & X[-K+1] \\ X[1] & X[0] & \cdots & X[-K+1] \\ & & \ddots & \\ X[K-1] & X[K-2] & \cdots & X[0] \end{bmatrix} \cdot \begin{bmatrix} A[1] \\ A[2] \\ \vdots \\ A[K] \end{bmatrix} = - \begin{bmatrix} X[1] \\ X[2] \\ \vdots \\ X[K] \end{bmatrix}. \tag{24}$$

This is a classic Yule–Walker system [4], which in our case has a unique solution when there are K distinct Diracs in $x(t)$ because there is a unique annihilating filter.

3) Factoring $A(z)$.

Given the coefficients $1, A[1], \ldots, A[K]$, we factor its $z-$ transform into its roots

$$A(z) = \prod_{k=0}^{K-1} \left(1 - u_k z^{-1}\right) \tag{25}$$

where $u_k = e^{-i(2\pi t_k/\tau)}$, which leads to the K locations $\{t_k\}_{k=0}^{K-1}$.

4) Finding the weights c_k.

Given the locations $\{t_k\}_{k=0}^{K-1}$, we can write K values of $X[k]$ as linear combinations of exponentials following (13). Again, for $K = 3$

$$\begin{bmatrix} X[0] \\ X[1] \\ X[2] \end{bmatrix} = \frac{1}{\tau} \begin{bmatrix} 1 & 1 & 1 \\ e^{-i(2\pi t_0/\tau)} & e^{-i(2\pi t_1/\tau)} & e^{-i(2\pi t_2/\tau)} \\ e^{-i(4\pi t_0/\tau)} & e^{-i(4\pi t_1/\tau)} & e^{-i(4\pi t_2/\tau)} \end{bmatrix} \cdot \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix}. \tag{26}$$

In general, using $u_k = e^{-i(2\pi t_k/\tau)}$, the system of equations is given by

$$\begin{bmatrix} X[0] \\ X[1] \\ \vdots \\ X[K-1] \end{bmatrix} = \frac{1}{\tau} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ u_0 & u_1 & \cdots & u_{K-1} \\ \vdots & \vdots & \cdots & \vdots \\ u_0^{K-1} & u_1^{K-1} & \cdots & u_{K-1}^{K-1} \end{bmatrix} \cdot \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{K-1} \end{bmatrix}. \tag{27}$$

which is a Vandermonde system, which always has a solution when the t_k's are distinct. ■

An interpretation of the above result is the following. Take $x(t)$, and project it onto the lowpass subspace corresponding to its Fourier series coefficients $-K$ to K. This projection, which is denoted by $y(t)$, is a unique representation of $x(t)$. Since it is lowpass with bandwidth $[-K2\pi/\tau, K2\pi/\tau]$, it can be sampled with a sampling period smaller than $\tau/2K$. Note that step 2 in the proof required $2K$ adjacent coefficients of $X[m]$. We chose the ones around the origin for simplicity, but any set will do. In particular, if the set is of the form $[\ell N - K + 1, \ell N + K]$, then one can use bandpass sampling and recover $x(t)$ as well.

Example 1: Consider a periodic stream of $K = 8$ weighted Diracs with period $\tau = 1024$ and sinc sampling kernel with bandwidth $[-\rho\pi, \rho\pi]$, ($\rho = 2K/\tau$), as illustrated in Fig. 2(a) and (b). The lowpass approximation is obtained by filtering the stream of Diracs with the sampling kernel, as illustrated in Fig. 2(c). The reconstruction of $x(t)$ from the samples is exact to machine precision and, thus, is not shown. The annihilating filter is illustrated in Fig. 3, and it can be seen that the locations of the Diracs are exactly the roots of the filter. In the third step of the proof of Theorem 1, there is a factorization step. This can be avoided by a method that is familiar in the coding literature and known as the Berlekamp–Massey algorithm [1]. In our case, it amounts to a spectral extrapolation procedure.

Corollary 1: Given the annihilating filter $A[m], m = 0, \ldots, K$ and the spectrum $X[m], m = -K, \ldots, K$, one can recover the entire spectrum of $x(t)$ by the following recursion:

$$X[m] = -\sum_{k=1}^{K} A[k] X[m-k], \quad m = K+1, K+2, \ldots. \tag{28}$$

For negative m, use $X[-m] = X^*[m]$ since $x(t)$ is real.

Proof: The Toeplitz system specifying the annihilating filter shows that the recursion must be satisfied for all m. This is a Kth-order recursive difference equation. Such a difference equation is uniquely specified by K initial conditions, in our case by $X[1], \ldots, X[K]$, showing that the entire Fourier series is specified by (28). ■

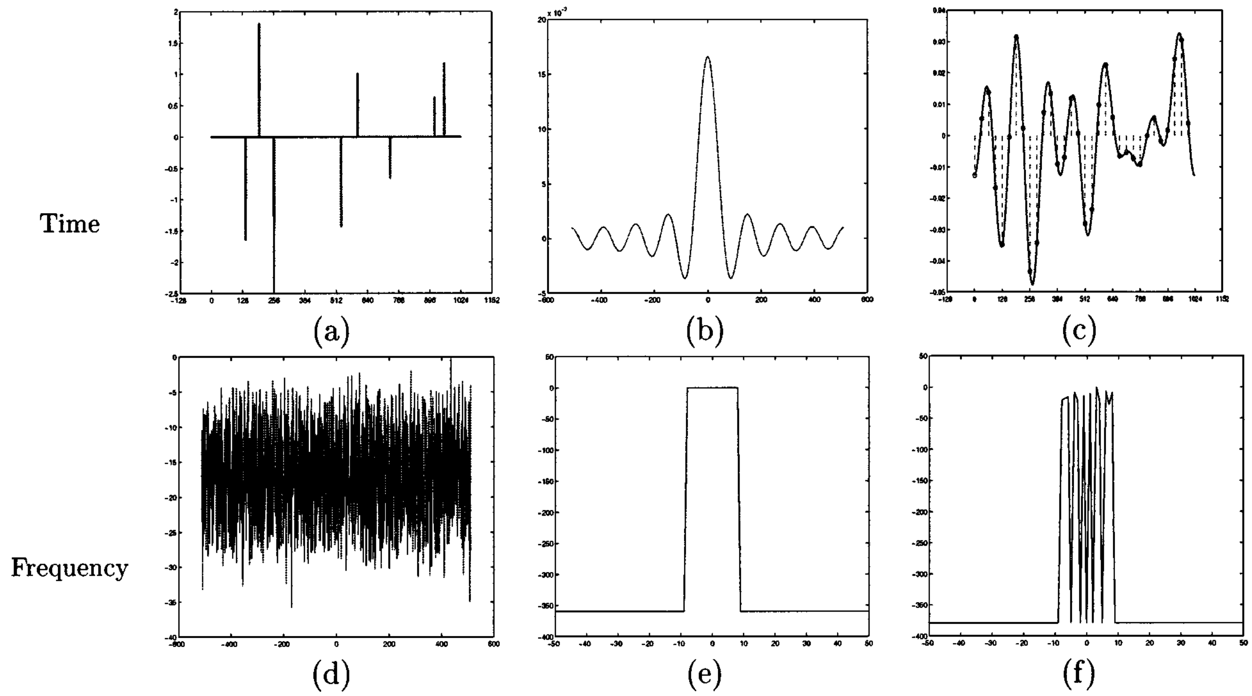

Fig. 2. Stream of Diracs. (a) Periodic stream of $K = 8$ weighted Diracs with period $\tau = 1024$. (b) Sinc sampling kernel $h_B(t)$. (c) Lowpass approximation $y(t)$, dashed lines are samples $y(nT)$, $T = 32$. Sample values $y(nT) = \langle h_B(t - nT), x(t)\rangle$. (d) Fourier series $X[m]$ (dB). (e) Central Fourier series $H_B[m], m = -50, \ldots, 50$ (dB). (f) Central Fourier series of sample values $Y_s[m], m = -50, \ldots, 50$ (dB).

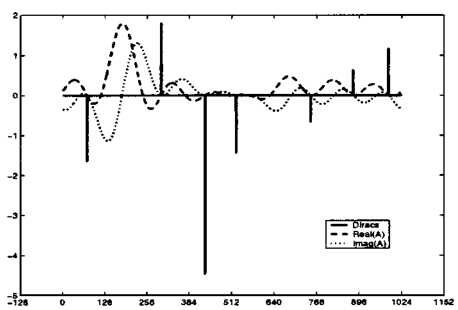

Fig. 3. Real and imaginary parts of the annihilating filter $a(t) = A(z)|_{z=e^{-i(2\pi t/\tau)}} = \prod_{k=0}^{K-1}\left(1 - e^{-i(2\pi(t_k - t)/\tau)}\right)$. The roots of the annihilating filter are exactly the locations of the Diracs.

The above recursion is routinely used in error-correction coding in order to fill in the error spectrum based on the error locator polynomial. In that application, finite fields are involved, and there is no problem with numerical precision. In our case, the real and complex field is involved, and numerical stability is an issue. More precisely, $A(z)$ has zeros on the unit circle, and the recursion (28) corresponds to convolution with $1/A(z)$, that is, a filter with poles on the unit circle. Such a filter is marginally stable and will lead to numerical instability. Thus, Corollary 1 is mostly of theoretical interest.

B. Nonuniform Splines

In this section, we consider periodic nonuniform splines of period τ. A signal $x(t)$ is a periodic nonuniform spline of degree R with knots at $\{t_k\}_{k=0}^{K-1} \in [0, \tau]$ if and only if its $(R+1)$th derivative is a periodic stream of K weighted Diracs $x^{(R+1)}(t) = \sum_{n\in\mathbb{Z}} c_n\,\delta(t - t_n)$, where $t_{n+K} = t_n + \tau$, and $c_{n+K} = c_n, \forall n \in \mathbb{Z}$. Thus, the rate of innovation is

$$\rho = \frac{2K}{\tau}. \tag{29}$$

Using (12), we can state that the Fourier series coefficients of $x^{(R+1)}(t)$ are $X^{(R+1)}[m] = 1/\tau \sum_{k=0}^{K-1} c_k\, e^{-i2\pi m t_k}$. Differentiating (10) $R+1$ times shows that these coefficients are

$$X^{(R+1)}[m] = \left(\frac{i2\pi m}{\tau}\right)^{R+1} X[m], \quad m \in \mathbb{Z}. \tag{30}$$

This shows that $X^{(R+1)}[m]$ can be annihilated by a filter $A[m]$ of length $K+1$. From Theorem 1, we can recover the periodic stream of K Diracs from the Fourier series coefficients $X^{(R+1)}[m], m \in [-K, K]$ and thus follows the periodic nonuniform spline.

Theorem 2: Consider a periodic nonuniform spline $x(t)$ with period τ, containing K pieces of maximum degree R. Take a sinc sampling kernel $h_B(t)$ such that B is greater or equal to the rate of innovation ρ given by (29), and sample $(h_B * x)(t)$ at N uniform locations $t = nT, n = 0, \ldots, N-1$, where $N \geq 2M+1$, and $M = \lfloor B\tau/2 \rfloor$. Then, $x(t)$ is uniquely represented by the samples

$$y_n = \langle h_B(t - nT), x(t)\rangle, \quad n = 0, \ldots, N-1. \tag{31}$$

Proof: The proof is similar to the proof of Theorem 1.

We determine the Fourier series $X[m], |m| \leq M$ from the N samples $y_n, n = 0, \ldots, N-1$ exactly as in Step 1 in the proof of Theorem 1 .

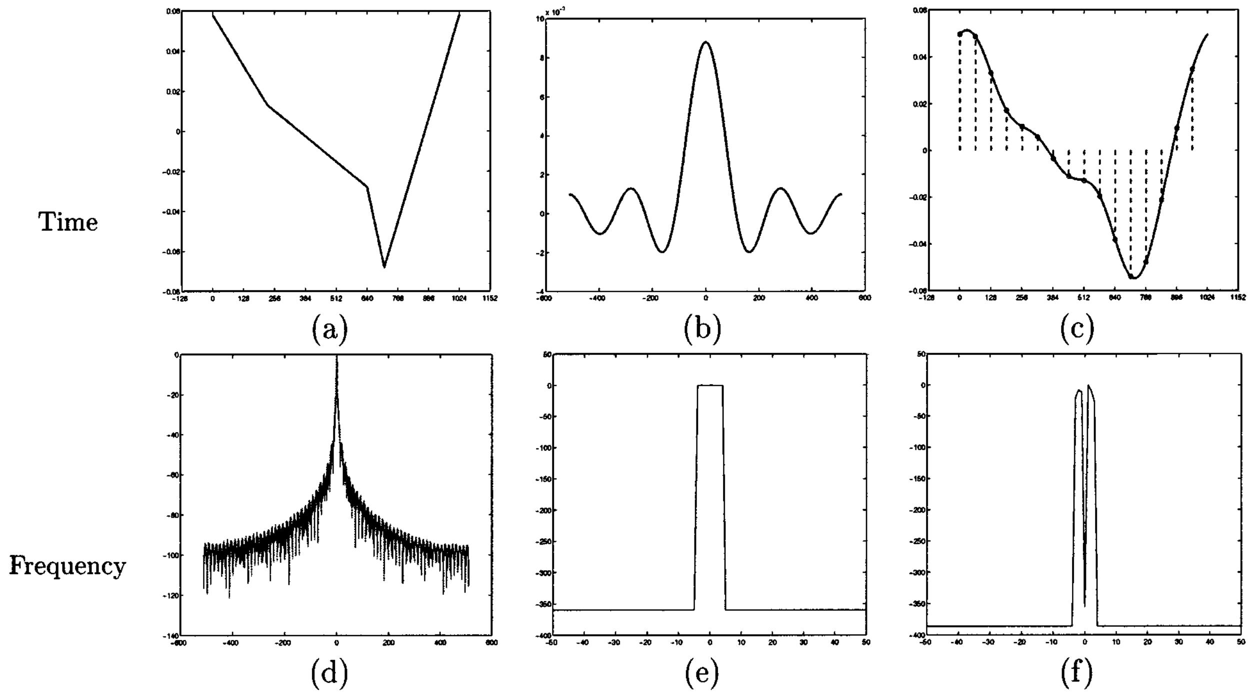

Fig. 4. Nonuniform spline. (a) Periodic nonuniform linear spline with $K = 4$ knots or transitions with period $\tau = 1024$. (b) Sinc sampling kernel $h_B(t)$. (c) Lowpass approximation $y(t)$; dashed lines are samples $y(nT)$, $T = 64$. Sample values $y(nT) = \langle h_B(t - nT), x(t)\rangle$. (d) Fourier series $X[m]$ (dB). (e) Central Fourier series $H_B[m], m = -50, \ldots, 50$ (dB). (f) Central Fourier series $Y_s[m], m = -50, \ldots, 50$ (dB). Reconstruction is within machine precision.

Note that the Fourier series coefficients of the $(R+1)$ differentiated nonuniform spline are given by (30); therefore, the values $X[m], |m| \le M$ provide $X^{(R+1)}[m], m \in [-M, M]$. We substitute $X[m]$ by $X^{(R+1)}[m]$ in Steps 2–4 in the proof of Theorem 1 and thus obtain the stream of K Diracs $x^{(R+1)}(t)$, that is, the K locations t_k and K weights c_k.

The nonuniform spline (see Fig. 4) $x(t)$ is obtained using (30) to get the missing $X[m]$ (that is, $|m| \ge M+1$) and substituting these values in (10).[3]

C. Derivative of Diracs

Derivative of Diracs are considered in this section to set the grounds for the following section on piecewise polynomial signals. The Dirac δ function is a distribution function whose rth derivative has the property $\int f(t)\, \delta^{(r)}(t - t_0)\, dt = (-1)^r f^{(r)}(t_0)$, where $f(t)$ is r times continuously differentiable [9].

Consider a periodic stream of differentiated Diracs

$$x(t) = \sum_{n \in \mathbb{Z}} \sum_{r=0}^{R_n - 1} c_{nr}\, \delta^{(r)}(t - t_n) \tag{32}$$

with the usual periodicity conditions $t_{n+K} = t_n + \tau$ and $c_{n+K,r} = c_{nr}, \forall n \in \mathbb{Z}$.

Note that there are K locations and $\tilde{K} = \sum_{k=0}^{K-1} R_k$ weights that makes at most $K + \tilde{K}$ degrees of freedom per period τ, that is, the rate of innovation is

$$\rho = \frac{K + \tilde{K}}{\tau}. \tag{33}$$

[3]Note that $X[0]$ is obtained directly.

The corresponding Fourier series coefficients are given by

$$X[m] = \frac{1}{\tau} \sum_{k=0}^{K-1} \sum_{r=0}^{R_k - 1} c_{kr} \left(\frac{i2\pi m}{\tau}\right)^r e^{-i(2\pi m t_k/\tau)}, \quad m \in \mathbb{Z}. \tag{34}$$

Let $\tilde{c}_{kr} = (1/\tau) c_{kr} (i2\pi/\tau)^r$ and $u_k = e^{-i(2\pi t_k/\tau)}$; then, the Fourier series simplifies to

$$X[m] = \sum_{k=0}^{K-1} \sum_{r=0}^{R_k - 1} \tilde{c}_{kr}\, m^r u_k^m, \quad m \in \mathbb{Z}. \tag{35}$$

From Proposition 4 in Appendix A, the filter $(1 - u_k z^{-1})^R$ annihilates the exponential $m^r u_k^m$, with $r \le R - 1$, and therefore, the filter defined by

$$A(z) = \prod_{k=0}^{K-1} \left(1 - u_k z^{-1}\right)^{R_k} \tag{36}$$

with R_k poles at $z = u_k$ annihilates $X[m]$. The locations t_k of the differentiated Diracs are obtained by first finding the annihilating filter coefficients $A[m], m = 0, \ldots, K$ and then finding the roots of $A(z)$. The weights $\tilde{c}_{kr}$, on the other hand, are found by solving the system in (35) for $m = 0, \ldots, \tilde{K} - 1$.

Theorem 3: Consider a periodic stream of differentiated Diracs $x(t)$ with period τ, as in (32). Take as a sampling kernel $h_B(t) = B\,\mathrm{sinc}(Bt)$, where B is greater or equal to the rate of innovation ρ given by (33), and sample $(h_B * x)(t)$ at N uniform locations $t = nT, n = 0, \ldots, N-1$, where $N \ge 2M + 1$ and $M = \lfloor B\tau/2 \rfloor$. Then, the samples

$$y_n = \langle h_B(t - nT), x(t)\rangle, \quad n = 0, \ldots, N-1 \tag{37}$$

are a sufficient characterization of $x(t)$.

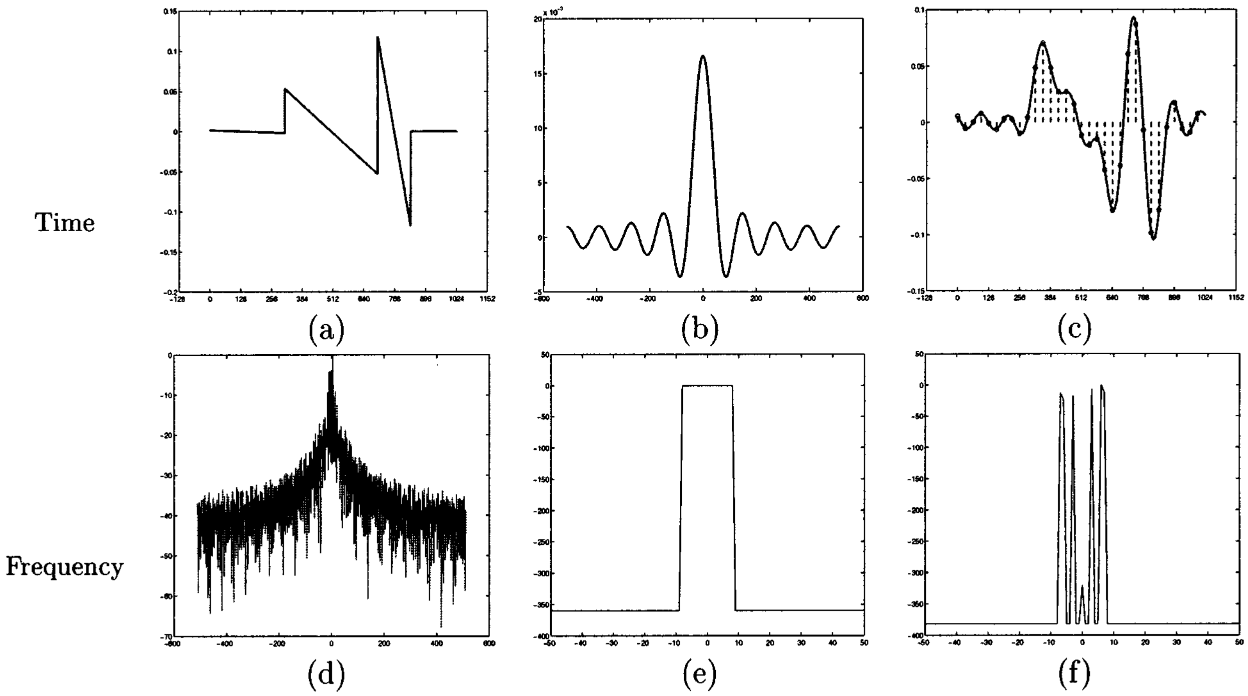

Fig. 5. Piecewise polynomial spline. (a) Periodic piecewise linear spline with $K = 4$ knots or transitions with period $\tau = 1024$. (b) Sinc sampling kernel $h_B(t)$. (c) Lowpass approximation $y(t)$; dashed lines are samples $y(nT), T = 32$. Sample values $y(nT) = \langle h_B(t - nT), x(t)\rangle$. (d) Fourier series $X[m]$ (dB). (e) Central Fourier series $H_B[m], m = -50, \ldots, 50$ (dB). (f) Central Fourier series of sample values $Y_s[m], m = -50, \ldots, 50$ (dB). Reconstruction from the samples is within machine precision.

Proof: We follow the same steps as in the proof of Theorem 1. From the N samples y_n with $N \geq 2M + 1$, we obtain the values $X[m], m \in [M, M]$ from which we determine the annihilating filter coefficients $A[m], m = 0, \ldots, \tilde{K}$. In Step 3, the annihilating filter is factored as in (36), from which we obtain the multiple roots $u_k = e^{-i(2\pi t_k/\tau)}, k = 0, \ldots, K - 1$ and, therefore, the K locations t_k. The $\tilde{K}$ weights c_{kr} are found by solving the generalized Vandermonde system in (35) for $\tilde{c}_{kr}$, from which we obtain $c_{kr} = (\tau^{r+1}/(i2\pi)^r)\,\tilde{c}_{kr}$. Similar to usual Vandermonde matrices, the determinant of the matrix given by the system vanishes only when $u_k = u_\ell$ for some $k \neq \ell$. This is not our case, and thus, the matrix is nonsingular and, therefore, admits a unique solution. ∎

D. Piecewise Polynomials

Similar to the definition of a periodic nonuniform splines, a signal $x(t)$ is a periodic piecewise polynomial with K pieces each of maximum degree R if and only if its $(R+1)$ derivative is a stream of differentiated Diracs, that is, given by $x^{(R+1)}(t) = \sum_{n\in\mathbb{Z}} \sum_{r=0}^{R} c_{nr}\, \delta^{(r)}(t - t_n)$ with the usual periodicity conditions. The degrees of freedom per period are K from the locations and $\tilde{K} = (R+1)K$ from the weights; thus, the rate of innovation is

$$\rho = \frac{(R+2)K}{\tau}. \tag{38}$$

By analogy with nonuniform splines, we have the following

Theorem 4: Consider a periodic piecewise polynomial $x(t)$ with period τ, containing K pieces of maximum degree R. Take a sinc sampling kernel $h_B(t)$ such that B is greater or equal to the rate of innovation ρ given by (38), and sample $(h_B * x)(t)$ at N uniform locations $t = nT, n = 0, \ldots, N-1$, where $N \geq 2M + 1$, and $M = \lfloor B\tau/2 \rfloor$. Then, $x(t)$ is uniquely represented by the samples

$$y_n = \langle h_B(t - nT), x(t)\rangle, \quad n = 0, \ldots, N-1. \tag{39}$$

Proof: The proof follows the same steps as the proof of Theorems 2 and 3.

First, from the N samples y_n, we obtain the Fourier series coefficients of the periodic piecewise polynomial signal $X[m], m \in [-M, M]$. Then, using (30), we obtain $2M + 1$ Fourier series coefficients of the $(R+1)$ times differentiated signal, which is a stream of differentiated Diracs. Using Theorem 3, we are able to recover the stream of differentiated Diracs from $X^{(R+1)}[m], m \in [-M, M]$. Similar to the periodic nonuniform spline, the periodic piecewise polynomial $x(t)$ (see Fig. 5) is obtained using (30) to get the missing $X[m]$ (that is, $|m| \geq M + 1$) and substituting these values in (10). ∎

IV. Finite-Length Signals With Finite Rate of Innovation

A finite-length signal with finite rate of innovation ρ clearly has a finite number of degrees of freedom. The question of interest is as follows: Given a sampling kernel with *infinite support*, is there a *finite set of samples* that uniquely specifies the signal? In the following sections, we will sample signals with finite number of weighted Diracs with infinite support sampling kernels such as the sinc and Gaussian.

A. Sinc Sampling Kernel

Consider a continuous-time signal with a finite number of weighted Diracs

$$x(t) = \sum_{k=0}^{K-1} c_k\, \delta(t - t_k) \tag{40}$$

and the sinc sampling kernel. The sample values are obtained by filtering the signal with a sinc sampling kernel. This is equivalent to taking the inner product between a shifted version of the sinc and the signal, that is, $y_n = \langle h_B(t - nT), x(t)\rangle$, where $h_B(t) = B\, \mathrm{sinc}(Bt)$, with $B = 1/T$. The question that arises is the following: How many of these samples do we need to recover the signal? The signal has $2K$ degrees of freedom: K from the weights and K from the locations of the Diracs; thus, N samples $N \geq 2K$ will be sufficient to recover the signal. Similar to the previous cases, the reconstruction method will require solving two systems of linear equations: one for the locations of the Diracs involving a matrix $\mathbf{V}$ and one for the weights of the Diracs involving a matrix $\mathbf{A}$. These systems admit solutions if the following conditions are satisfied.

C1] Rank$(\mathbf{V}) \leq K$, where $\mathbf{V} \in \mathbb{R}^{(N-K)\times(K+1)}$ is defined by (45).

C2] Rank$(\mathbf{A}) = K$, where $\mathbf{A} \in \mathbb{R}^{K\times K}$ is defined by (43).

Theorem 5: Given a finite stream of K weighted Diracs and a sinc sampling kernel $h_B(t)$ with $B = 1/T$, if conditions C1] and C2] are satisfied, then the $N \geq 2K$ samples

$$y_n = \langle h_B(t - nT), x(t)\rangle, \quad n = 0, \ldots, N-1 \tag{41}$$

are a sufficient representation of the signal.

Proof: The sample values in (41) are equivalent to

$$\begin{aligned} y_n &= \sum_{k=0}^{K-1} c_k\, B\, \mathrm{sinc}\left(\frac{t_k}{T} - n\right) \\ &= \sum_{k=0}^{K-1} \frac{c_k\, B\, \sin\left(\frac{\pi t_k}{T} - \pi n\right)}{\pi\left(\frac{t_k}{T} - n\right)} \\ &= (-1)^n \sum_{k=0}^{K-1} \frac{c_k\, B\, \sin\left(\frac{\pi t_k}{T}\right)}{\pi\left(\frac{t_k}{T} - n\right)} \\ \Leftrightarrow (-1)^n\, y_n &= \frac{1}{\pi} \sum_{k=0}^{K-1} c_k\, B\, \sin\left(\frac{\pi t_k}{T}\right) \cdot \frac{1}{\left(\frac{t_k}{T} - n\right)}. \end{aligned} \tag{42}$$

Let us define the degree $K-1$ Lagrange polynomial $L_k(u) = (P(u)/(u - t_k/T))$, where $P(u) = \prod_{k=0}^{K-1}(u - t_k/T)$. Multiplying both sides of (42) by $P(n)$, we find an expression in terms of the interpolating polynomials.

$$\underbrace{(-1)^{n+1}\, P(n)\, y_n}_{Y_n} = \sum_{k=0}^{K-1} c_k \underbrace{B\, \sin\left(\frac{\pi t_k}{T}\right) \frac{L_k(n)}{\pi}}_{[\mathbf{A}]_{nk}} \tag{43}$$

$$\Leftrightarrow \mathbf{Y} = \mathbf{A} \cdot \mathbf{c}. \tag{44}$$

Since the right-hand side of (43) is a polynomial of degree $K-1$ in the variable n, applying K finite differences makes the left-hand side vanish,[4] that is, $\Delta^K((-1)^n P(n)\, y_n) = 0, n = K, \ldots, N-1$. If we let $P(u) = \sum_k p_k\, u^k$, then this annihilating equation is equivalent to

$$\sum_{k=0}^{K} p_k \underbrace{\Delta^K\left((-1)^n\, n^k\, y_n\right)}_{[\mathbf{V}]_{nk}} = 0 \tag{45}$$

$$\Leftrightarrow \mathbf{V} \cdot \mathbf{p} = 0 \tag{46}$$

where $\mathbf{V}$ is an $(N-K)\times(K+1)$ matrix. The system (46) admits a nontrivial solution when $N - K \geq K$ and the Rank$(\mathbf{V}) \leq K$, that is, condition C1]. Therefore, (45) can be used to find the $K+1$ unknowns p_k, which lead to the K locations t_k since these are the roots of $P(u)$. Once the K locations t_k are determined, the weights of the Diracs c_k are found by solving the system in (44) for $n = 0, \ldots, K-1$. Since $t_k \neq t_l, \forall k \neq l$, the system admits a solution from condition C2]. ■

Note that the result does not depend on T. This of course holds only in theory since in practice, the matrix $\mathbf{V}$ may be ill-conditioned if T is not chosen appropriately. A natural solution to this conditioning problem is to take more than the critical number of samples and solve (46) using a singular value decomposition (SVD). This is also the method of choice when noise is present in the signal. The matrix $\mathbf{A}$, which is used to find the weights of the Diracs, is less sensitive to the value of T and better conditioned on average than $\mathbf{V}$.

[4]Note that the K finite-difference operator plays the same role as the annihilating filter in the previous section.

B. Gaussian Sampling Kernel

Consider sampling the same signal as in (40) but, this time, with a Gaussian sampling kernel $h_\sigma(t) = e^{-t^2/2\sigma^2}$. Similar to the sinc sampling kernel, the samples are obtained by filtering the signal with a Gaussian kernel. Since there are $2K$ unknown variables, we show next that N samples with $N \geq 2K$ are sufficient to represent the signal.

Theorem 6: Given a finite stream of K weighted Dirac and a Gaussian sampling kernel $h_\sigma(t) = e^{-t^2/2\sigma^2}$. If $N \geq 2K$, then the N sample values

$$y_n = \langle h_\sigma(t - nT), x(t)\rangle \tag{47}$$

are sufficient to reconstruct the signal.

Proof: The sample values are given by

$$\begin{aligned} y_n &= \sum_{k=0}^{K-1} c_k\, e^{-(t_k - nT)^2/2\sigma^2} \\ &= \sum_{k=0}^{K-1} \left(c_k\, e^{-t_k^2/2\sigma^2}\right) \cdot e^{n t_k T/\sigma^2} \cdot e^{-n^2T^2/2\sigma^2}. \end{aligned} \tag{48}$$

If we let $S[n] = e^{n^2T^2/2\sigma^2}\, y_n$, $a_k = c_k\, e^{-t_k^2/2\sigma^2}$, and $u_k = e^{t_k T/\sigma^2}$, then (48) is equivalent to

$$S[n] = \sum_{k=0}^{K-1} a_k\, u_k^n, \quad n = 0, \ldots, N-1. \tag{49}$$

Note that we reduced the expression $S[n]$ to a linear combination of real exponentials u_k. Since $N \geq 2K$, the annihilating filter method described in Appendix B allows us to determine c_k and u_k. In addition, note that the Toeplitz system in (68) has real exponential components $S[n] = e^{n^2/2\sigma^2}\, y_n$, and therefore, a solution exists when the number of equations is greater than the number of unknowns, that is, $N - K \geq K$, and the rank of

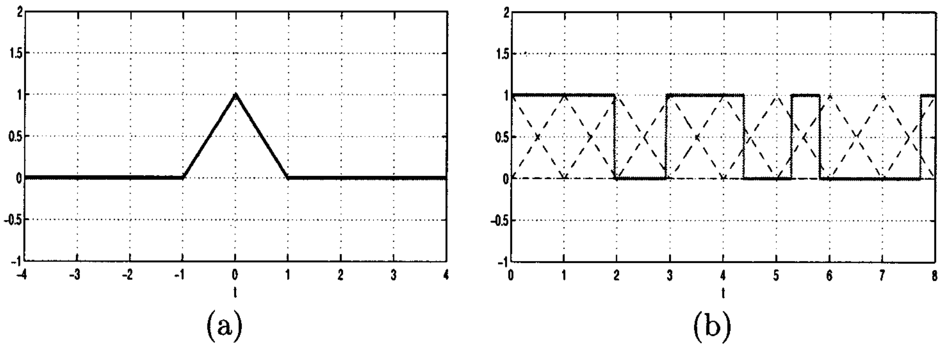

Fig. 6. (a) Hat spline sampling kernel $\beta_1(t/T)$, $T = 1$. (b) Bilevel signal with up to two transitions in an interval $[n, n+1]$ sampled with a hat sampling kernel.

the system is equal to K, which is the case by hypothesis. Furthermore, σ must be carefully chosen; otherwise, the system is ill conditioned. The locations are then given by

$$t_k = \frac{\sigma^2 \ln u_k}{T} \tag{50}$$

and the weights of the Diracs are simply given by

$$c_k = a_k \, e^{t_k^2/2\sigma^2}. \tag{51}$$

■

Here, unlike in the sinc case, we have an almost local reconstruction because of the exponential decay of the Gaussian sampling kernel, which brings us to the next topic.

V. Infinite-Length Bilevel Signals With Finite Local Rate of Innovation

In this section, we consider the dual problem of Section IV, that is, *infinite* length *signals* $x(t), t \in \mathbb{R}^+$ with a finite *local* rate of innovation and sampling kernels with *compact support* [19]. In particular, the B-splines of different degree d are considered [17]

$$\beta_d(t) = (\beta_{d-1} * \beta_0)(t), \quad d \in \mathbb{N}^+ \tag{52}$$

where the box spline is defined by $\beta_0(t) = 1$ for $0 \le t < 1$ and 0 elsewhere.

We develop local reconstruction algorithms that depend on moving intervals equal to the size of the support of the sampling kernel. [5] The advantage of local reconstruction algorithms is that their complexity does not depend on the length of the signal. In this section, we consider bilevel signals but similar to the previous sections the results can be extended to piecewise polynomials.

Let $x(t)$ be an infinite-length continuous-time signal that takes on two values 0 and 1 with initial condition $x(t)\big|_{t=0} = 1$ with a finite local rate of innovation ρ. These are called bilevel signals and are completely represented by their transition values t_k.

Suppose a bilevel signal is sampled with a box spline $\beta_0(t/T)$. The sample values obtained are $y_n = \langle \beta_0(t/T - n), x(t) \rangle$, which correspond to the area occupied by the signal in the interval $[nT, (n+1)T]$. Thus, if there is at most one transition per box, then we can recover the transition from the sample. This leads us to the following proposition.

[5] The size of the support of $\beta_d(t/T)$ is equal to $(d+1)T$.

Proposition 1: A bilevel signal $x(t), t > 0$ with initial condition $x(t)\big|_{t=0} = 1$ is uniquely determined from the samples $y_n = \langle x(t), \beta_0(t/T - n) \rangle$, where $\beta_0(t)$ is the box spline if and only if there is at most one transition t_k in each interval $[nT, (n+1)T]$.

Proof: For simplicity, let $T = 1$, and suppose that $x(n) = 1$. If there are no transitions in the interval $[n, n+1]$, then the sample value is $y_n = 1$. If there is one transition in the interval $[n, n+1]$, then the sample value is equal to $y_n = t_k - n$ from which we uniquely obtain the transition value $t_k = y_n + n$. To show necessity, suppose there are two transitions t_k, t_{k+1} in the interval $[n, n+1]$, then the sample value is equal to $y_n = t_k - t_{k+1} + 1$ and is not sufficient to determine both transitions. Thus, there must be at most one transition in an interval $[n, n+1]$ to uniquely define the signal. ■

Now, consider shifting the bilevel signal by an unknown shift ϵ; then, there can be two transitions in an interval of length T, and one box function will not be sufficient to recover the transitions. Suppose we double the sampling rate; then, the support of the box sampling kernel is doubled, and we have two sample values y_n, y_{n+1} covering the interval $[nT, (n+1)T]$, but these values are identical (see their areas). Therefore, increasing the sampling rate is still insufficient.

This brings us to consider a sampling kernel not only with a larger support but with added information. For example, the hat spline function $\beta_1(t) = 1 - |t|$ for $|t| < 1$ and 0 elsewhere leads to the sample values defined by $y_n = \langle \beta_1(t/T - n), x(t) \rangle$.

Fig. 6 illustrates that there are two sample values covering the interval $[nT, (n+1)T]$ from which we can uniquely determine the signal.

Proposition 2: An infinite-length bilevel signal $x(t)$, with initial condition $x(0) = 1$, is uniquely determined from the samples defined by $y_n = \langle \beta_1(t/T - n), x(t) \rangle$, where $\beta_1(t)$ is the hat sampling kernel if and only if there are at most two transitions $t_k \neq t_j$ in each interval $[nT, (n+2)T]$.

Proof: Again, for simplicity, let $T = 1$, and suppose the signal is known for $t \le n$ and $x(t)\big|_{t=n} = 1$. First, we show sufficiency by showing the existence and uniqueness of a solution. Then, we show necessity by a counterexample.

$\Leftarrow$ Similar to the box sampling kernel, the sample values will depend on the configuration of the transitions in the interval $[n, n+2]$. If there are at most two transitions in the interval $[n, n+2]$, then the possible configurations are (0,0), (0,1), (0,2), (1,0), (1,1), (2,0), where the first and second component indicate the number of transitions in the intervals $[n, n+1]$, $[n+1, n+2]$, respectively. Furthermore, since the hat sampling kernel is of degree one, we obtain for each configuration a quadratic system of equations with variables t_0, t_1

$$y_n = \int_{n-1}^{n} x(t)(1+t-n)\,dt + \int_{n}^{n+1} x(t)(1-(t-n))\,dt \tag{53}$$

$$y_{n+1} = \int_{n}^{n+1} x(t)(1+t-(n+1))\,dt + \int_{n+1}^{n+2} x(t)(1-(t-(n+1)))\,dt \tag{54}$$

which admits a solution in the given interval.

As for uniqueness, if $y_n = 1$ and $y_{n+1} = 1$, then this implies configuration (0,0).

If $y_n = 1$ and $1/2 \leq y_{n+1} \leq 1$, then the possible configurations are (0,1),(0,2). By hypothesis, there are at most two transitions in the interval $[n+1, n+3]$; therefore, if $y_{n+2} \leq 1/2$, then the configuration in the interval $[n, n+2]$ is (0,1); otherwise, if $y_{n+2} \geq 1/2$, then the configuration is (0,2).

If $1/2 \leq y_n \leq 1$ and $1/2 \leq y_{n+1} \leq 1$, then this implies configuration (2,0).

If $1/2 \leq y_n \leq 1$ and $0 \leq y_{n+1} \leq 1/2$, then this implies configuration (1,0).

$\Rightarrow$ Necessity is shown by counterexample.

Consider a bilevel signal with three transitions in the interval [0,2] but with all three in the interval [0,1]. Then, the quadratic system of equations is

$$y_0 = \frac{1}{2} + t_0 - t_1 + t_2 - \frac{t_0^2}{2} + \frac{t_1^2}{2} - \frac{t_2^2}{2} \tag{55}$$

$$y_1 = \frac{t_0^2}{2} - \frac{t_1^2}{2} + \frac{t_2^2}{2} \tag{56}$$

which does not admit a unique but an infinite number of solutions. Thus, there must be at most 2 transitions in an interval [0,2]. ■

The pseudo-code for sampling bilevel signals using the box and hat functions are given in full detail in [10], [22]. A stronger condition than the one in Proposition 2 is to require $\rho_{max}(2T) \leq 1/T$. In that case, we are ensured that on any interval of length $2T$, there are at most two transitions, and therefore, the reconstruction is unique. Based on Propositions 1 and 2, we conjecture that using splines of degree d, a local rate of innovation $\rho_{max}((d+1)T) \leq 1/T$ ensures unicity of the reconstruction.

VI. Conclusion

We considered signals with finite rate of innovation that allow uniform sampling after appropriate smoothing and perfect reconstruction from the samples. For signals like streams of Diracs, nonuniform splines, and piecewise polynomials, we were able to derive exact reconstruction formulas, even though these signals are nonbandlimited.

The methods rely on separating the innovation in a signal from the rest and identifying the innovative part from the samples only. In particular, the annihilating filter method plays a key role in isolating the innovation. Some extensions of these results, like to piecewise bandlimited signals, are presented in [11] and [21], and more details, including a full treatment of the discrete-time case, are available in [10], [22], and [23].

To prove the sampling theorems, we assumed deterministic, noiseless signals. In practice, noise will be present, and this can be dealt with by using oversampling and solving the various systems involved using the singular value decomposition (SVD). Such techniques are standard, for example, in noisy spectral estimation [16]. Initial investigations using these techniques in our sampling problem are promising [13].

It is of also interest to compare our notion of "finite rate of innovation" with the classic Shannon bandwidth [12]. The Shannon bandwidth finds the dimension of the subspace (per unit of time) that allows us to represent the space of signals of interest. For bandpass signals, where Nyquist's rate is too large, Shannon's bandwidth is the correct notion, as it is for certain other spread-spectrum signals. For pulse position modulation (PPM) [7], Shannon's bandwidth is proportional to the number of possible positions, whereas the rate of innovation is fixed per interval.[6] Thus, the rate of innovation coincides with our intuition for degrees of freedom.

This discussion also indicates that an obvious application of our results is in communications systems, like, for example, in ultrawide band communication (UWB). In such a system, a very narrow pulse is generated, and its position is carrying the information. In [6], initial results indicate that a decoder can work at much lower rate than Nyquist's rate by using our sampling results. Finally, filtered streams of Diracs, known as shot noise [7], can also be decoded with low sampling rates while still recovering the exact positions of the Diracs. Therefore, we expect the first applications of our sampling results to appear in wideband communications.

Finally, the results presented so far raise a number of questions for further research. What other signals with finite rate of innovation can be sampled and perfectly reconstructed? What tools other than the annihilating filter can be used to identify innovation, and what other types of innovation can be identified? What are the multidimensional equivalents of the above results, and are they computationally feasible? These topics are currently under investigation. Thus, the connection between sampling theory on the one hand, and spectral analysis and error correction coding on the other hand, is quite fruitful.

Appendix A
Annihilating Filters

Here, we give a brief overview of annihilating filters (see [16] for more details). First, we have the following definition.

Definition 2: A filter $A[n]$ is called an annihilating filter of a signal $S[n]$ when

$$(A * S)[n] = 0 \quad \forall n \in \mathbb{N}. \tag{57}$$

[6]The number of degrees of freedom per interval is 1 or 2, depending if the amplitude is fixed or free.

Next, we give annihilating filters for signals that are linear combinations of exponentials.

Proposition 3: The signal $S[n] = \sum_{k=0}^{K-1} c_k\, u_k^n$, where $c_k \in \mathbb{R}$, $u_k \in \mathbb{C}$, is annihilated by the filter

$$A(z) = \prod_{k=0}^{K-1} (1 - u_k\, z^{-1}) = \sum_{\ell=0}^{K} A[\ell] z^{-\ell}. \tag{58}$$

Proof: Note that

$$(A * S)[n] = \sum_{\ell=0}^{K} A[\ell]\, S[n-\ell] \tag{59}$$

$$= \sum_{\ell=0}^{K} \sum_{k=0}^{K-1} c_k\, A[\ell]\, u_k^{n-\ell} \tag{60}$$

$$= \sum_{k=0}^{K-1} c_k \underbrace{\left(\sum_{\ell=0}^{K} A[\ell] u_k^{-\ell} \right)}_{A(u_k)} u_k^n = 0. \tag{61}$$

Thus, $A[n]$ annihilates $S[n]$. ■

Next, we show that the signal $n^r\, u^n$ is annihilated by a filter with R poles, where $r \leq R-1$.

Proposition 4: The signal $S[n] = n^r\, u^n$ is annihilated by the filter

$$A(z) = (1 - u\, z^{-1})^R = \sum_{\ell=0}^{R} A[\ell] z^{-\ell}. \tag{62}$$

Proof: Note that

$$(A * S)[n] = \sum_{\ell=0}^{K} A[\ell]\, S[n-\ell] \tag{63}$$

$$= \sum_{\ell=0}^{K} A[\ell]\, (n-\ell)^r\, u^{n-\ell}. \tag{64}$$

By differentiating r times $(1 - u\, z^{-1})^R$, we easily see that

$$\sum_{\ell=0}^{R} A[\ell]\, \ell\, (\ell-1) \cdot \cdots \cdot (\ell - r + 1)\, u^{-\ell} = 0. \tag{65}$$

This is true for $r = 0, \ldots, R-1$. Thus, $\sum_{\ell=0}^{R} A[\ell]\, P[\ell]\, u^{-\ell} = 0$ for all polynomials $P(\ell)$ of degree less than or equal to $R-1$, in particular, for $P[\ell] = (n-\ell)^r$. Thus, $A[n]$ annihilates $S[n]$. ■

It follows from Proposition 3 and 4 that the filter $\prod_{k=0}^{K-1} (1 - u_k\, z^{-1})^{R+1}$ annihilates the signal $S[n] = \sum_{k=0}^{K-1} c_k\, n^R\, u_k^n$.

APPENDIX B
ANNIHILATING FILTER METHOD

The annihilating filter method consists of finding the values c_k and u_k in

$$S[n] = \sum_{k=0}^{K-1} c_k\, u_k^n, \quad \forall n \in \mathbb{Z} \tag{66}$$

and is composed of three parts: First, we need to find the annihilating filter that involves solving a linear system of equations; second, we need to find the roots of the z-transform of the annihilating filter, which is a nonlinear function; third, we must solve another linear system of equations to find the weights.

1) Finding the annihilating filter.

The filter coefficients $A[\ell]$ in $A(z) = \sum_{\ell=0}^{K} A[\ell]\, z^{-\ell}$ must be such that (57) is satisfied or

$$\sum_{\ell=0}^{K} A[\ell]\, S[n-\ell] = 0, \quad \forall n \in \mathbb{Z}. \tag{67}$$

In matrix/vector form, the system in (67) is equivalent to

$$\begin{bmatrix} \vdots & \vdots & \cdots & \vdots \\ S[0] & S[-1] & \cdots & S[-K] \\ S[1] & S[0] & \cdots & S[-(K-1)] \\ \vdots & \vdots & \ddots & \\ S[K] & S[K-1] & \cdots & S[0] \\ \vdots & \vdots & \cdots & \vdots \end{bmatrix} \cdot \begin{pmatrix} A[0] \\ A[1] \\ \vdots \\ A[K] \end{pmatrix} = 0. \tag{68}$$

Suppose N values $S[n]$ are available. Since there are $K+1$ unknown filter coefficients, we need at least $K+1$ equations, and therefore, N must be greater or equal to $2K+1$. Define $\mathbf{S}$ the appropriate submatrix; then, the system $\mathbf{S} \cdot \mathbf{A} = 0$ will admit a solution when Rank($\mathbf{S}$) $= K$.

In practice, this system is solved using an SVD where the matrix $\mathbf{S}$ is decomposed into $\mathbf{S} = \mathbf{U}\Sigma\mathbf{V}^*$. We obtain that $\mathbf{A} = \mathbf{V} \cdot \mathbf{e}_{K+1}$, where $\mathbf{e}_{K+1}$ is a vector with 1 on position $K+1$ and 0 elsewhere. The method can be adapted to the noisy case by minimizing $||\mathbf{S}\cdot\mathbf{A}||$, in which case, $\mathbf{A}$ is given by the eigenvector associated with the smallest eigenvalue of $\mathbf{S}^T\mathbf{S}$.

2) Finding the u_k.

Once the the filter coefficients $A[n]$ are found, then the values u_k are simply the roots of the annihilating filter $A(z)$.

3) Finding the c_k.

To determine the weights c_k, it suffices to take K equations in (66) and solve the system for c_k. Let $n = 0, \ldots, K-1$; then, in matrix vector form, we have the following Vandermonde system:

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ u_0 & u_1 & \cdots & u_{K-1} \\ \vdots & \vdots & \cdots & \vdots \\ u_0^{(K-1)} & u_1^{(K-1)} & \cdots & u_{K-1}^{(K-1)} \end{bmatrix} \cdot \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_{K-1} \end{pmatrix} = \begin{pmatrix} S[0] \\ S[1] \\ \vdots \\ S[K-1] \end{pmatrix} \tag{69}$$

and has a unique solution when

$$u_k \neq u_\ell, \forall k \neq \ell. \tag{70}$$

This concludes the annihilating filter method.

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their constructive remarks. Suggestions by Prof. S. Shamai on the Shannon bandwidth and by Prof. E. Telatar and Prof. R. Urbanke on the connection to error correction codes are greatly acknowledged. A special thank you to Prof. P. J. S. G. Ferreira for his invaluable suggestions.

REFERENCES

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

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley Interscience, 1991.

[3] P. J. S. G. Ferreira and J. M. N. Vieira, "Locating and correcting errors in images," in *Proc. IEEE Int. Conf. Image Process.*, vol. I, Santa Barbara, CA, Oct 1997, pp. 691–694.

[4] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD: John Hopkins Univ. Press, 1996.

[5] A. J. Jerri, "The Shannon sampling theorem-its various extensions and applications: A tutorial review," *Proc. IEEE*, vol. 65, pp. 1565–1597, 1977.

[6] J. Kusuma, A. Ridolfi, and M. Vetterli, "Sampling of communication systems with bandwidth expansion," presented at the Int. Contr. Conf., 2002.

[7] E. Lee and D. Messerschmitt, *Digital Communication*. Boston, MA: Kluwer, 1994.

[8] S. Mallat, *A Wavelet Tour of Signal Processing*. San Diego, Ca: Academic, 1998.

[9] J. E. Marsden, *Elementary Classical Analysis*. New York: Springer-Verlag, 1974.

[10] P. Marziliano, "Sampling Innovations," Ph.D. dissertation, Swiss Fed. Inst. Technol., Audiovis. Commun. Lab., Lausanne, Switzerland, 2001.

[11] P. Marziliano, M. Vetterli, and T. Blu, "A sampling theorem for a class of piecewise bandlimited signals," *submitted for publication*.

[12] J. L. Massey, "Toward an information theory of spread-spectrum systems," in *Code Division Multiple Access Communications*, S.G. Glisic and P.A. Leppanen, Eds. Boston, MA: Kluwer, 1995, pp. 29–46.

[13] A. Ridolfi, I. Maravic, J. Kusuma, and M. Vetterli, "Sampling signals with finite rate of innovation: the noisy case," *Proc. ICASSP*, 2002.

[14] T. J. Shan, A. M. Bruckstein, and T. Kailath, "Adaptive resolution of overlapping echoes," in *Proc. IEEE Int. Symp. Circuits Syst.*, 1986.

[15] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, 1948.

[16] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 2000.

[17] M. Unser, "Splines: A perfect fit for signal and image processing," *IEEE Signal Processing Mag.*, vol. 16, pp. 22–38, Nov. 1999.

[18] ——, "Sampling–50 years after Shannon," *Proc. IEEE*, vol. 88, pp. 569–587, Apr. 2000.

[19] M. Vetterli, "Sampling of piecewise polynomial signals," Swiss Fed. Inst. Technol., Lausanne, Switzerland, LCAV, 1999.

[20] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[21] M. Vetterli, P. Marziliano, and T. Blu, "Sampling discrete-time piecewise bandlimited signals," in *Proc. Sampling Theory and Applications Workshop*, Orlando, FL, May 2001, pp. 97–102.

[22] ——, "Sampling signals with finite rate of innovation," Swiss Fed. Inst. Technol., Lausanne, Switzerland, DSC-017, 2001.

[23] ——, "A sampling theorem for periodic piecewise polynomial signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Salt Lake City, UT, May 2001.

[24] J. M. N. Vieira and P. J. S. G. Ferreira, "Interpolation, spectrum analysis, error-control coding, and fault-tolerant computing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 3, 1997, pp. 1831–1834.

[25] A. J. Weiss, A. S. Willsky, and B. C. Levy, "Eigenstructure approach for array-processing with unknown intensity coefficients," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 1613–1617, Oct. 1988.

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 D.Sc. degree from 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 last 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 Science until 1997. He now holds an Adjunct Professor position at Berkeley. Since 1995, he has been a Professor of communication systems at EPFL, where he chaired the Communications Systems Division from 1996 to 1997 and heads the Audio-Visual Communications Laboratory. Since 2001, he has directed the National Competence Center in Research on mobile information and communication systems (www.terminodes.org). He has held visiting positions at ETHZ (1990) and Stanford (1998). He is on the editorial boards of *Annals of Telecommunications, Applied and Computational Harmonic Analysis*, and *The Journal of Fourier Analysis and Applications*. He is the co-author, with J. Kovacević, of the book *Wavelets and Subband Coding* (Englewood Cliffs, NJ: Prentice-Hall, 1995). He has published about 85 journal papers on a variety of topics in signal/image processing and communications and holds seven patents. His research interests include sampling, wavelets, multirate signal processing, computational complexity, signal processing for communications, digital video processing, and joint source/channel coding.

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 from 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 Awards in 1991 and in 1996 (for papers with D. LeGall and K. Ramchandran, respectively). He won 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 (www.swtr.ch). He was a plenary speaker at various conferences (e.g., 1992 IEEE ICASSP).

Pina Marziliano was born in Montréal, QC, Canada, in 1971. She received the B.Sc. degree in applied mathematics in 1994 and the M.Sc. degree in computer science (operations research) in 1996, both from the Université de Montréal. She received the Ph.D. degree in communication systems from the Swiss Federal Institute of Technology, Lausanne, in 2001.

She is presently doing research on perceptual quality metrics for images and video at Genista Corporation, Genimedia SA, Lausanne. Her research interests include Fourier and wavelet analysis, sampling theory applied to audio-visual communications, and perceptual quality metrics.

Thierry Blu (M'96) was born in Orléans, France, in 1964. He received the Diplôme d'ingénieur degree from École Polytechnique, Paris, France, in 1986 and from Télécom Paris (ENST) in 1988. He received the Ph.D degree in electrical engineering in 1996 from ENST for a study on iterated rational filterbanks applied to wideband audio coding.

He is currently with the Biomedical Imaging Group at the Swiss Federal Institute of Technology (EPFL), Lausanne, on leave from France Télécom National Centre for Telecommunication Studies (CNET), Issy-les-Moulineaux, France. His interests include (multi)wavelets, multiresolution analysis, multirate filterbanks, approximation and sampling theory, and psychoacoustics.