# Sampling Moments and Reconstructing Signals of Finite Rate of Innovation: Shannon Meets Strang–Fix

Pier Luigi Dragotti, *Member, IEEE*, Martin Vetterli, *Fellow, IEEE*, and Thierry Blu, *Member, IEEE*

***Abstract*—Consider the problem of sampling signals which are not bandlimited, but still have a finite number of degrees of freedom per unit of time, such as, for example, nonuniform splines or piecewise polynomials, and call the number of degrees of freedom per unit of time the rate of innovation. Classical sampling theory does not enable a perfect reconstruction of such signals since they are not bandlimited. Recently, it was shown that, by using an adequate sampling kernel and a sampling rate greater or equal to the rate of innovation, it is possible to reconstruct such signals uniquely [34]. These sampling schemes, however, use kernels with infinite support, and this leads to complex and potentially unstable reconstruction algorithms. In this paper, we show that many signals with a finite rate of innovation can be sampled and perfectly reconstructed using physically realizable kernels of compact support and a local reconstruction algorithm. The class of kernels that we can use is very rich and includes functions satisfying Strang–Fix conditions, exponential splines and functions with rational Fourier transform. This last class of kernels is quite general and includes, for instance, any linear electric circuit. We, thus, show with an example how to estimate a signal of finite rate of innovation at the output of an *RC* circuit. The case of noisy measurements is also analyzed, and we present a novel algorithm that reduces the effect of noise by oversampling.**

***Index Terms*—Analog-to-digital conversion, annihilating filter method, multiresolution approximations, sampling methods, splines, wavelets.**

## I. Introduction

SAMPLING theory plays a central role in modern signal processing and communications and has experienced a recent revival thanks, in part, to the recent advances in wavelet theory [12], [27]. In the typical sampling setup depicted in Fig. 1, the original continuous-time signal $x(t)$ is filtered before being (uniformly) sampled with sampling period $T$. The filtering may be a design choice or may be due to the acquisition device. If we denote with $y(t) = h(t) * x(t)$ the filtered version of $x(t)$, the samples $y_n$ are given by

$$y_n = \langle x(t), \varphi(t/T - n) \rangle = \int_{-\infty}^{\infty} x(t)\varphi(t/T - n)dt$$

where the sampling kernel $\varphi(t)$ is the scaled and time-reversed version of $h(t)$.

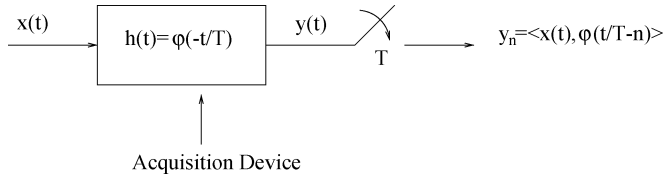


Fig. 1. Sampling setup. Here, $x(t)$ is the continous-time signal, $h(t)$ is the impulse response of the acquisition device, and $T$ is the sampling period. The measured samples are $y_n = \langle x(t), \varphi(t/T - n) \rangle$.

The key problem then is to find the best way to reconstruct $x(t)$ from the given samples, and the key questions are as follows. 1) What classes of signals can be reconstructed? 2) What classes of kernels allow such reconstructions? 3) What kind of reconstruction algorithms are involved? Ideally, we would like to be able to reconstruct large classes of signals, using simple reconstruction algorithms and, most importantly, with general and physically realizable kernels.

The classical answer to the sampling problem is provided by the famous Shannon sampling theorem which states the conditions to reconstruct bandlimited signals from their samples. In this case, the reconstruction process is linear and the kernel is the sinc function. In fact, the whole sampling process can be interpreted as an approximation procedure in which the original signal is projected onto the shift-invariant subspace of bandlimited functions and only this projection can be reconstructed. This subspace interpretation has then been used to extend Shannon's theorem to classes of nonbandlimited signals that belong to shift-invariant subspaces, such as uniform splines [27], [29].

Recently, it was shown that it is possible to develop sampling schemes for classes of signals that are neither bandlimited nor belong to a fixed subspace [34] (see also [16], [18], and [21]). For instance, it was shown that it is possible to sample streams of Diracs or piecewise polynomial signals using a sinc or a Gaussian kernel. The common feature of such signals is that they have a parametric representation with a finite number of degrees of freedom and are, therefore, called signals with finite rate of innovation (FRI) [34]. The reconstruction process

Manuscript received February 22, 2006; revised July 12, 2006. This work was supported in part by the Swiss National Foundation under Grant 200020-103729 and in part by the EPFL's Bernoulli Center during the "Wavelets and Applications" semester. The material in this paper was presented in part at the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP),Montréal, Canada, in May 2004 and Philadelphia, PA, in March 2005. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Gerald Schuller.

P. L. Dragotti is with the Electrical and Electronic Engineering Department, Imperial College London, London SW7 2AZ, U.K. (e-mail: p.dragotti@imperial.ac.uk).

M. Vetterli is with the Institute of Communication Systems, Swiss Federal Institute of Technology, Lausanne (EPFL), CH-1015 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).

T. Blu is with the Institute of Imaging and Applied Optics, Swiss Federal Institute of Technology, Lausanne (EPFL), CH-1015 Lausanne, Switzerland (e-mail: thierry.blu@epfl.ch).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2006.890907

1053-587X/$25.00 © 2007 IEEE

for these schemes is based on the use of a locator or annihilating filter, a tool widely used in spectral estimation [22] and error correction coding [3], but also for sampling, interpolation [14], [35], and shape reconstruction [13], [19].

The fundamental limit of the above sampling methods, as well as of the classical Shannon reconstruction scheme, is that the choice of the sampling kernel is very limited and the required kernels are of infinite support. As a consequence, the reconstruction algorithm is usually physically nonrealizable (e.g., realization of an ideal low-pass filter) or, in the case of FRI signals, becomes complex and unstable. The complexity is in fact influenced by the global rate of innovation of $x(t)$.

In this paper, we show that many signals with a local finite rate of innovation can be sampled and perfectly reconstructed using a wide range of sampling kernels and a local reconstruction algorithm. The reconstruction algorithm proposed in this paper is also based on the annihilating filter method. However, the main property the kernel has to satisfy is to be able to reproduce polynomials or exponentials. Thus, functions satisfying Strang–Fix conditions [24] (e.g., splines and scaling functions), exponential splines [31], and functions with rational Fourier transforms can be used in our formulation. This last family of kernels is of particular importance since most linear devices used in practice have a transfer function which is rational. Despite the fact that kernels with rational Fourier transform have infinite support, we show that the reconstruction algorithm remains local, and, thus, its complexity still depends on the local, rather than global, rate of innovation of $x(t)$.

It is also worth mentioning that the problem of reconstructing signals from sparse (nonuniform) measurements has gained a lot of attention recently (see, for instance, [7] and [11]). These recent works, however, focus mostly on discrete signals that have a sparse representation in a basis or a frame. The focus of our work instead is on exact reconstruction of continuous-time signals from uniform samples.

The paper is organized as follows. In the next section, we review the notion of signals with finite rate of innovation and present the families of sampling kernels that are used in our sampling schemes. Section III presents our main sampling results for the case of kernels reproducing polynomials. In particular, we show how to sample and perfectly reconstruct streams of Diracs, streams of differentiated Diracs and piecewise polynomial signals. The following section focuses on the use of wavelets and scaling functions to sample signals with finite rate of innovation and discusses connections between these sampling results and the problem of increasing the resolution of a given signal. In Sections V and VI, the previous sampling results are extended to the case in which the sampling kernel reproduces exponentials, moreover, as an example, we show how to estimate FRI signals at the output of electric circuits. Section VII deals with the case of noisy measurements, and, finally, conclusions are drawn in Section VIII.

## II. On Signals and Kernels

In the introduction, we have informally discussed the signals and kernels that will be used in our sampling formulation. Let us now introduce more formally the notion of signals with finite rate of innovation [34] and present the families of sampling kernels that will be used in our sampling schemes.

### A. Signals With Finite Rate of Innovation

Consider a signal of the form

$$x(t) = \sum_{k\in\mathbb{Z}} \sum_{r=0}^{R-1} \gamma_{k,r} g_r(t - t_k). \tag{1}$$

Clearly, if the set of functions $\{g_r(t)\}_{r=0,1,\ldots,R-1}$ is known, the only free parameters in the signal $x(t)$ are the coefficients $\gamma_{k,r}$ and the time shifts $t_k$. It is, therefore, natural to introduce a counting function $C_x(t_a, t_b)$ that counts the number of free parameters in $x(t)$ over an interval $[t_a, t_b]$. The rate of innovation of $x(t)$ is then defined as [34]

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

*Definition 1:* A signal with a finite rate of innovation is a signal whose parametric representation is given in (1) and with a finite $\rho$ as defined in (2).

It is of interest to note that shift-invariant signals, including bandlimited signals, fall under Definition 1. For instance, if we call $f_{\max}$ the maximum nonzero frequency in a bandlimited real signal, then $\rho = 2f_{\max}$. Therefore, one possible interpretation is that it is possible to sample bandlimited signals because they have finite rate of innovation (rather than because they are bandlimited).

In some cases, it is more convenient to consider a local rate of innovation with respect to a moving window of size $\tau$. The local rate of innovation at time $t$ is, thus, given by [34]

$$\rho_\tau(t) = \frac{1}{\tau} C_x\left(t - \frac{\tau}{2}, t + \frac{\tau}{2}\right). \tag{3}$$

Clearly, $\rho_\tau(t)$ tends to $\rho$ as $\tau \to \infty$. In our context, as it will become evident later, the notion of local rate of innovation plays a more important role than the global rate of innovation, since our reconstruction schemes are local.

### B. Sampling Kernels

As mentioned in the introduction, the signal $x(t)$ is usually filtered before being sampled. The samples $y_n$ are given by $y_n = \langle x(t), \varphi(t/T - n)\rangle$, where the sampling kernel $\varphi(t)$ is the time reversed version of the filter's impulse response. The impulse response of the filter depends on the physical properties of the acquisition device and, in most cases, is specified a-priori and cannot be modified. It is, therefore, important to develop sampling schemes that do not require the use of very particular or even physically nonrealizable filters. In our formulation, we can use a wide range of different kernels. For the sake of clarity, we divide them into three different families.

1) *Polynomial reproducing kernels*: Any kernel $\varphi(t)$ that together with its shifted versions can reproduce polynomials of maximum degree $N$. That is, any kernel that satisfies

$$\sum_{n\in\mathbb{Z}} c_{m,n}\varphi(t - n) = t^m, \qquad m = 0, 1, \ldots, N \tag{4}$$

for a proper choice of the coefficients $c_{m,n}$.

2) *Exponential reproducing kernels*: Any kernel $\varphi(t)$ that together with its shifted versions can reproduce complex exponentials of the form $e^{\alpha_m t}$ with $\alpha_m = \alpha_0 + m\lambda$ and $m = 0, 1, \ldots, N$. That is, any kernel satisfying

$$\sum_{n\in\mathbb{Z}} c_{m,n}\varphi(t-n) = e^{\alpha_m t} \quad \text{with } \alpha_m = \alpha_0 + m\lambda \text{ and } m = 0, 1, \ldots, N \tag{5}$$

for a proper choice of the coefficients $c_{m,n}$.

3) *Rational kernels*: Any stable kernel $\varphi(t)$ with rational Fourier transform of the form

$$\hat{\varphi}(\omega) = \frac{\prod_{i=0}^{I}(j\omega - b_i)}{\prod_{m=0}^{N}(j\omega - \alpha_m)}, \quad \text{with } I < N \quad \alpha_m = \alpha_0 + m\lambda \text{ and } m = 0, 1, \ldots, N \tag{6}$$

where $\hat{\varphi}(\omega)$ is the Fourier transform of $\varphi(t)$.

In all cases, the choice of $N$ depends on the local rate of innovation of the original signal $x(t)$ as will become clear later on. Since our reconstruction scheme is based on the use of a digital filter (i.e., the annihilating filter), the exponents in (5) and the poles in (6) must be restricted to $\alpha_m = \alpha_0 + m\lambda$, where $\alpha_0$ and $\lambda$ can be chosen arbitrarily but $m$ is an integer. This fact will be more evident in Section V. Finally, the coefficients $c_{m,n}$ in (4) are given by $c_{m,n} = (1/T)\int_{-\infty}^{\infty} t^m \tilde{\varphi}(t/T - n)dt$, where $\tilde{\varphi}(t)$ is chosen to form with $\varphi(t)$ a quasi-biorthonormal set [6]. This includes the particular case where $\tilde{\varphi}(t)$ is the dual of $\varphi(t)$, that is, $\langle \tilde{\varphi}(t-n), \varphi(t-k)\rangle = \delta_{n,k}$. A similar expression applies to the coefficients $c_{m,n}$ in (5).

The first family of kernels includes any function satisfying the so-called Strang–Fix conditions [24]. Namely, $\varphi(t)$ satisfies (4) if and only if

$$\hat{\varphi}(0) \neq 0 \text{ and } \hat{\varphi}^{(m)}(2n\pi) = 0 \text{ for } n \neq 0 \text{ and } m = 0, 1, \ldots, N$$

where $\hat{\varphi}(\omega)$ is again the Fourier transform of $\varphi(t)$ and the superscript $(m)$ stands for the $m$th derivative of $\varphi(t)$. These conditions were originally valid for functions with compact support only, later they have been extended to noncompactly supported functions [6], [8], [10].

One important example of functions satisfying Strang–Fix conditions is given by the family of B-splines [26]. A B-spline $\beta_N(t)$ of order $N$ is obtained from the (N+1)-fold convolution of the box function $\beta_0(t)$, that is

$$\beta_N(t) = \underbrace{\beta_0(t) * \beta_0(t) \ldots * \beta_0(t)}_{N+1 \text{ times}}$$

with $\hat{\beta}_0(\omega) = (1 - e^{-j\omega})/j\omega$. The B-spline of order $N$ can reproduce polynomials of maximum degree $N$ and the size $(N+1)$ of its support is the smallest for a function that can achieve that order of approximation. More important, it is possible to show that any function $\varphi(t)$ that reproduces polynomials of degree $N$ can be decomposed into a B-splines and a distribution $u(t)$ with $\int u(t)dt \neq 0$, that is, $\varphi(t) = u(t) * \beta_N(t)$ [4], [5], [20].

Strang–Fix conditions are used extensively in wavelet theory, as well. In that context, the focus is on the design of wavelets with a certain number of vanishing moments [9], [17], [25], [33]. The interesting point, here, is that a wavelet with $N+1$ vanishing moments is generated by a scaling function that can reproduce polynomials of degree $N$. This means that such a scaling function can be included in our family of sampling kernels. A more detailed discussion of the use of wavelets and scaling functions to sample FRI signals will be given in Section IV.

The theory related to the reproduction of exponentials is somewhat more recent and relies on the notion of exponential splines (E-splines) [31]. A function $\beta_\alpha(t)$ with Fourier transform

$$\hat{\beta}_\alpha(\omega) = \frac{1 - e^{\alpha - j\omega}}{j\omega - \alpha}$$

is called E-spline of first order. Notice that $\alpha$ can be either real or complex. Moreover, notice that $\beta_\alpha(t)$ reduces to the classical zero-order B-spline when $\alpha = 0$. The function $\beta_\alpha(t)$ satisfies several interesting properties, in particular, it is of compact support and a linear combination of shifted versions of $\beta_\alpha(t)$ reproduces $e^{\alpha t}$. As in the classical case, higher order E-splines are obtained by successive convolutions of lower-order ones, or

$$\hat{\beta}_{\vec{\alpha}}(\omega) = \prod_{n=0}^{N} \frac{1 - e^{\alpha_n - j\omega}}{j\omega - \alpha_n}$$

where $\vec{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_N)$. The higher-order spline is again of compact support and it is possible to show that it can reproduce any exponential in the subspace spanned by $\{e^{\alpha_0 t}, e^{\alpha_1 t}, \ldots, e^{\alpha_N t}\}$ [31]. Moreover, since the exponential reproduction formula is preserved through convolution [31], any composite function of the form $\varphi(t) * \beta_{\vec{\alpha}}(t)$ is also able to reproduce exponentials. Therefore, the second group of kernels contains any composite function of the form $\varphi(t) * \beta_{\vec{\alpha}}(t)$ with $\beta_{\vec{\alpha}}(t) = \beta_{\alpha_0}(t) * \beta_{\alpha_1}(t) * \ldots * \beta_{\alpha_N}(t)$, $\alpha_m = \alpha_0 + m\lambda$ and $m = 0, 1, \ldots, N$. Notice that the exponential case reduces to that of reproduction of polynomials when $\alpha_m = 0$ for $m = 0, 1, \ldots, N$. For this reason, we could study our sampling schemes in the exponential case only and then particularize it to the polynomial case. However, we prefer to keep the two cases separate for the sake of pedagogy.

The last group of kernels includes many *linear differential acquisition devices*. That is, linear devices or systems for which the input and output are related by a linear differential equation. This includes most of the commonly used electrical or mechanical systems.

Kernels with rational Fourier transforms can be linked to E-splines through a proper digital filtering as will be shown in VI. The use of E-splines and kernels with rational Fourier transforms will be investigated in Sections V and VI, respectively.

## III. Reconstruction of FRI Signals Using Kernels That Reproduce Polynomials

In this section, we assume that the sampling kernel $\varphi(t)$ satisfies the Strang–Fix conditions [24], that is, a linear combination of shifted versions of $\varphi(t)$ can reproduce polynomials of maximum degree $N$ [see (4)]. We further assume that the sampling kernel is of compact support $L$, that is, $\varphi(t) = 0$

for $t \notin [-L/2, L/2]$ where $L$ is an integer for simplicity.[1] We study the sampling of streams of Diracs, streams of differentiated Diracs and piecewise polynomial signals. Furthermore, possible extensions to any signal with finite rate of innovation are briefly discussed at the end of Section III-B. We present the results for streams of Diracs in detail and derive the other sampling theorems directly from these results.

### A. Streams of Diracs

We split the problem of sampling streams of Diracs into two sub-problems. First, we show how to sample a signal containing at most $K$ Diracs, where $K$ is a finite integer and is known a-priori. Then, we consider the case of signals with an unknown (maybe infinite) number of Diracs but with a finite local rate of innovation, and present a sequential algorithm that can reconstruct such signals from their samples.

*1) Stream of $K$ Diracs:* Consider a stream, $x(t)$, of $K$ Diracs: $x(t) = \sum_{k=0}^{K-1} a_k \delta(t - t_k), t \in \mathbb{R}$. Call $y_n$ the observed samples, that is, $y_n = \langle x(t), \varphi(t-n) \rangle = \sum_{k=0}^{K-1} a_k \varphi(t_k - n)$ where, for simplicity, we have assumed $T = 1$ and assume that the sampling kernel $\varphi(t)$ is able to reproduce polynomials of maximum degree $N \geq 2K-1$. We now show that under these hypotheses, it is possible to retrieve the locations $t_k$ and the amplitudes $a_k$ of $x(t)$ from its samples. The reconstruction algorithm operates in three steps. First, the first $N+1$ moments of the signal $x(t)$ are found. Second, the Diracs' locations are retrieved using an annihilating filter (for a detailed description of the annihilating filter method, we refer to [22] and [34]). Third, the amplitudes $a_k$ are obtained solving a Vandermonde system. The three steps of our scheme can be more precisely described as follows.

1) Retrieve the first $N+1$ moments of the signal $x(t)$.
Call $\tau_m = \sum_n c_{m,n} y_n$, $m = 0, 1, \ldots, N$ the weighted sum of the observed samples, where the weights $c_{m,n}$ are those in (4) that reproduce $t^m$. We have

$$\begin{aligned}\tau_m &= \sum_n c_{m,n} y_n \\ &\overset{(a)}{=} \left\langle x(t), \sum_n c_{m,n} \varphi(t-n) \right\rangle \\ &\overset{(b)}{=} \left\langle \sum_{k=0}^{K-1} a_k \delta(t-t_k), \sum_n c_{m,n} \varphi(t-n) \right\rangle \\ &\overset{(c)}{=} \int_{-\infty}^{\infty} \sum_{k=0}^{K-1} a_k \delta(t-t_k) t^m dt \\ &= \sum_{k=0}^{K-1} a_k t_k^m, \quad m = 0, 1, \ldots, N \end{aligned} \tag{7}$$

where $(a)$ follows from the linearity of the inner product, $(b)$ from the fact that $x(t) = \sum_{k=0}^{K-1} a_k \delta(t-t_k)$, and $(c)$ from the polynomial reproduction formula in (4). The integral in $(c)$ represents precisely the $m$th-order moment of the original signal $x(t)$. Hence, proper linear combinations of the observed samples provide the first $N+1$ moments of the signal. This fact is graphically illustrated in Fig. 2. Since the original signal is a stream of $K$ Diracs, the moments of $x(t)$ have the form given by the last term of (7) which is very often encountered in spectral estimation. It is, therefore, possible to estimate locations and amplitudes of the Diracs from the moments $\tau_m$ using the annihilating filter method which is commonly used in that context.

2) Find the locations $t_k$ of $x(t)$.
Call $h_m$ $m = 0, 1, \ldots, K$ the filter with $z$ transform

$$H(z) = \sum_{m=0}^{K} h_m z^{-m} = \prod_{k=0}^{K-1} (1 - t_k z^{-1}). \tag{8}$$

That is, the roots of $H(z)$ correspond to the locations $t_k$. It clearly follows that:

$$\begin{aligned} h_m * \tau_m &= \sum_{i=0}^{K} h_i \tau_{m-i} = \sum_{i=0}^{K} \sum_{k=0}^{K-1} a_k h_i t_k^{m-i} \\ &= \sum_{k=0}^{K-1} a_k t_k^m \underbrace{\sum_{i=0}^{K} h_i t_k^{-i}}_{0} = 0. \end{aligned} \tag{9}$$

The filter $h_m$ is, thus, called annihilating filter since it annihilates the observed signal $\tau_m$. The zeros of this filter uniquely define the set of locations $t_k$ since the locations are distinct. The filter coefficients $h_m$ are found from the system of equations in (9). Since $h_0 = 1$, the identity in (9) leads to a Yule–Walker system of equations involving at least $2K$ consecutive values of $\tau_m$ and can be written in matrix form, as follows:

$$\begin{bmatrix} \tau_{K-1} & \tau_{K-2} & \cdots & \tau_0 \\ \tau_K & \tau_{K-1} & \cdots & \tau_1 \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{N-1} & \tau_{N-2} & \cdots & \tau_{N-K} \end{bmatrix} \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_K \end{pmatrix} = - \begin{pmatrix} \tau_K \\ \tau_{K+1} \\ \vdots \\ \tau_N \end{pmatrix}. \tag{10}$$

This classic Yule–Walker system has, in this case, a unique solution since $h_m$ is unique for the given signal. Given the filter coefficients $h_m$, the locations of the Diracs are the roots of the polynomial in (8). Notice that, since we need at least $2K$ consecutive values of $\tau_m$ to solve the Yule–Walker system, we need the sampling kernel to be able to reproduce polynomials of maximum degree $N \geq 2K-1$.

3) Find the weight $a_k$.
Given the locations $t_0, t_1, \ldots, t_k$, the weights $a_k$ are obtained by solving, for instance, the first $K$ consecutive equations in (7). These equations can be written in matrix form as follows:

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ t_0 & t_1 & \cdots & t_{K-1} \\ \vdots & \vdots & \ddots & \vdots \\ t_0^{K-1} & t_1^{K-1} & \cdots & t_{K-1}^{K-1} \end{bmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{K-1} \end{pmatrix} = \begin{pmatrix} \tau_0 \\ \tau_1 \\ \vdots \\ \tau_{K-1} \end{pmatrix}.$$

This is a Vandermonde system which yields a unique solution for the weights $a_k$ given that the $t_k$s are distinct.

[1] Recall that functions satisfying Strang–Fix conditions can be of either compact or infinite support. The case of kernels with compact support is more interesting from a practical point of view. Thus, in this paper, we concentrate only on this case.

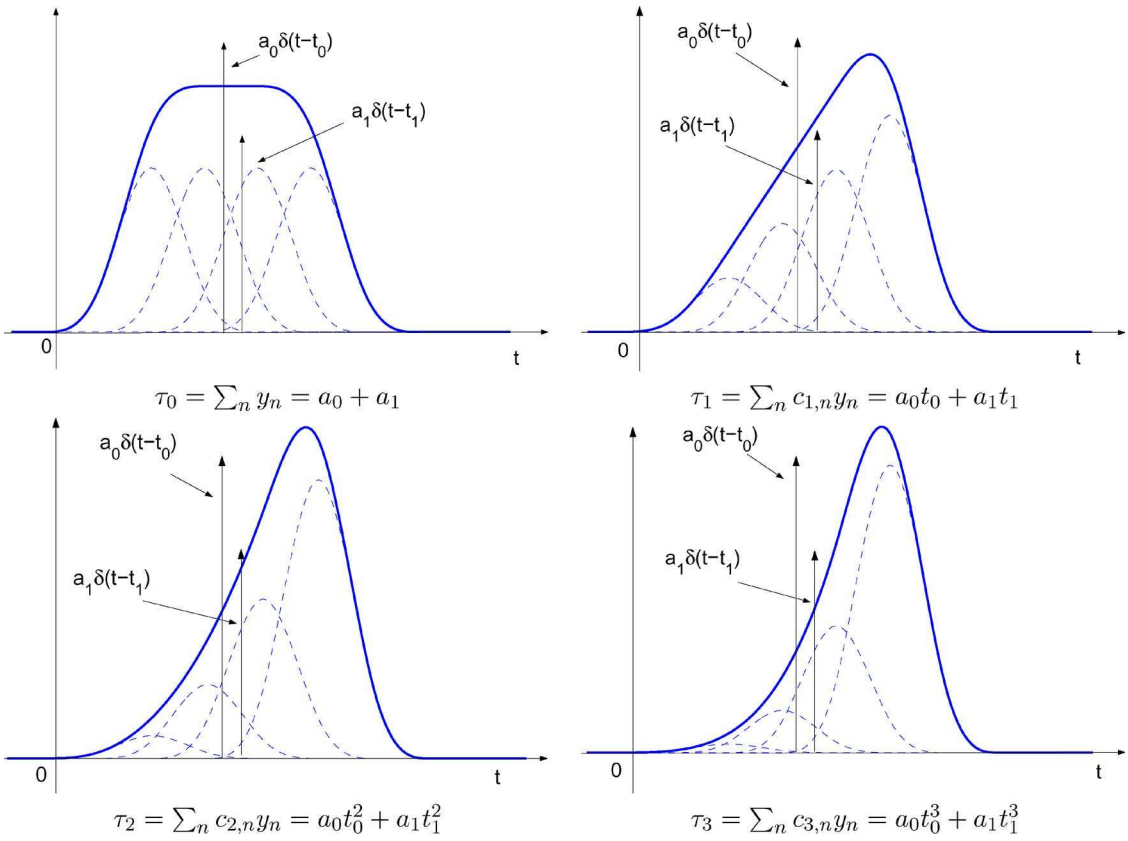


Fig. 2. Illustration of the reproduction of polynomials of maximum degree three using cubic splines. In this example, only four translated versions of the splines overlap the two Diracs. The dashed functions in each plot represent the properly weighted splines overlapping the two Diracs. The four solid-line functions represent the weighted sums of these four splines. Because of the polynomial reproduction formula, the following is true: $c_{m,0}y_0 + c_{m,1}y_1 + c_{m,2}y_2 + c_{m,3}y_3 = a_0t_0^m + a_1t_1^m$ for $m = 0, 1, 2, 3$.

The three previous steps show that it is indeed possible to reconstruct uniquely a stream of $K$ Diracs from its samples. We, thus, have the following result.

*Theorem 1:* Consider a stream $x(t)$ of $K$ Diracs: $x(t) = \sum_{k=0}^{K-1} a_k\delta(t - t_k)$ and a sampling kernel $\varphi(t)$ that can reproduce polynomials of maximum degree $N \geq 2K - 1$. Then the samples defined by $y_n = \langle x(t), \varphi(t/T - n)\rangle$ are sufficient to characterize $x(t)$ uniquely.

Before concluding this section, we would also like to mention that the topic related to the reconstruction of signals from a finite number of moments is an old one and for a comprehensive overview, which is beyond the scope of this paper, we refer to [1] and [15].

*2) A Sequential Local Algorithm for an Infinite Stream of Diracs:* In the previous section, we showed the existence of a basic method for retrieving $K$ Diracs. The problem is that the reconstruction scheme becomes more and more complex and unstable when the number $K$ of Diracs increases. It is, therefore, critical to see if we can take advantage of the locality of the sampling kernel to develop a sequential, local reconstruction algorithm. Intuitively, if we have groups of Diracs separated by empty intervals, then we should be able to separate these groups and reconstruct them sequentially.

Now, the support of the sampling kernel is $L$; thus, a single Dirac can influence at most $L$ consecutive samples and $K$ consecutive Diracs can generate a block of at most $KL$ consecutive nonzero samples. Thus, if two groups of $K$ consecutive Diracs are sufficiently distant, the two blocks of nonzero samples are separated by some zero samples, and by locating these zeros, we can separate the two blocks and apply the reconstruction method of the previous section on each block independently. If we assume that there are at most $K$ Diracs in an interval of size $KL + 1$, we are assured that at least one zero sample will separate two groups of nonzeros.

While, in most cases, the above condition is sufficient, there are situations in which it is not. This can happen, for instance, when a zero sample corresponds to a particular combination of Diracs. To avoid that these rare events prevent the algorithm from working properly, we need to make stronger assumptions. We, therefore, assume that there are at most $K$ Diracs in an interval of size $2KL$. This condition can be relaxed in most situations and, from a practical point of view, the event of having a "false" zero is very unlikely. Still, for simplicity and for the rest of the paper, we will assume that there are at most $K$ Diracs in an interval of size $2KL$. The rationale behind this assumption is that it ensures that for any sequence of consecutive "false" zeros, there is in the same window a longer sequence of consecutive "true" zeros. Therefore, the only thing the algorithm has to do is to search for the longest sequence of zeros in a group of $2KL$ samples. More precisely, the reconstruction algorithm operates as follows (also see Fig. 3). The algorithm starts by looking for the first nonzero sample in the sequence, call it $y_{n_1}$. The algorithm then checks the $2KL$ consecutive samples $y_{n_1}, y_{n_1+1}, \ldots, y_{n_1+2KL-1}$ and looks for the longest sequence

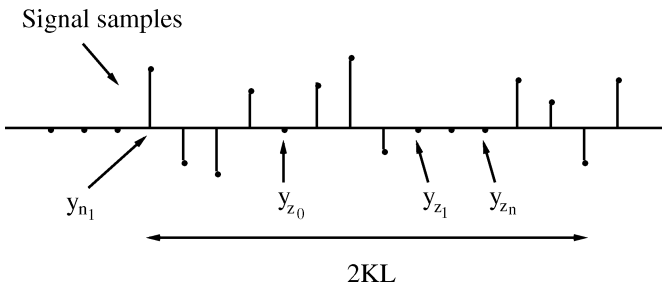


Fig. 3. Sequential reconstruction algorithm starts by looking for the first nonzero sample (in this case, the sample $y_{n_1}$), it then looks for the longest sequence of consecutive zeros in the block of $2KL$ samples $y_{n_1}, y_{n_1+1}, \ldots, y_{n_1+2KL-1}$. In this example, such a sequence start with the sample $y_{z_1}$. It is possible to show that such a sequence must contains "true" zeros and can, therefore, be used to separate two blocks of nonzero samples. Notice that the algorithm disregards the isolated zero sample $y_{z_0}$. That is because we cannot guarantee that such a sample truly indicates absence of Diracs.

of consecutive zeros inside this block. Denoting this sequence as $y_{z_1}, y_{z_1+1}, \ldots, y_{z_n}$, it is easy to show that such a sequence must include 'true' zeros and as such can be used to separate two blocks of nonzero samples. This means that the Diracs that have generated the nonzero samples $y_{n_1}, y_{n_1+1}, \ldots, y_{z_1-1}$ are not influenced by any other Dirac and can, therefore, be reconstructed using the reconstruction scheme presented before. After the reconstruction, the algorithm starts the whole process again from the sample $y_{z_n+1}$ on.

We can, thus, summarize the discussion of this section and the sampling result of the previous section as follows.

*Theorem 2:* Assume a sampling kernel $\varphi(t)$ that can reproduce polynomials of maximum degree $N \geq 2K-1$ and of compact support $L$. An infinite-length stream of Diracs $x(t) = \sum_{n\in\mathbb{Z}} a_n \delta(t-t_n)$ is uniquely determined from the samples defined by $y_n = \langle x(t), \varphi(t/T-n)\rangle$ if there are at most $K$ Diracs in an interval of size $2KLT$.

Using the notation introduced in Section II, the above theorem says that it is possible to sample any stream of Diracs with local rate of innovation $\rho_{2KLT} \leq 1/LT$. This means that there is a fundamental connection between the local complexity of the signal and the complexity of the reconstruction process. For instance, if there is at most one Dirac in an interval of size $2TL$, only two moments need to be retrieved at each iteration and the estimation of the amplitude and location of the Dirac becomes straightforward. In contrast, the reconstruction process becomes more complex and unstable, when the number $K$ of Diracs to retrieve at each iteration becomes very large. This fact is of particular interest in the case of noisy measurements. In that context, in fact, stability of the reconstruction algorithm is of crucial importance.

To conclude this section, we show in Fig. 4 an example of our sampling scheme. In this example, the signal is made of two groups of $K = 4$ Diracs and is shown in Fig. 4(a). The signal is sampled with a B-spline that can reproduce polynomials of degree $2K-1 = 7$ [Fig. 4(b)] and the samples are shown in Fig. 4(c). Since the nonzero samples generated by the two groups of Diracs are separated by a sequence of zero samples, the reconstruction algorithm can operate on the first group of nonzero samples to retrieve the first $K$ Diracs, and then reiterate the process on the following group of nonzero samples to retrieve the remaining $K$ Diracs. The reconstructed signal is shown in Fig. 4(d) and reconstruction is exact to machine precision.

### B. Stream of Differentiated Diracs

Consider now a stream of differentiated Diracs

$$x(t) = \sum_{k=0}^{K-1} \sum_{r=0}^{R_k-1} a_{k,r} \delta^{(r)}(t-t_k).$$

Note that this signal has $K$ Diracs and $\hat{K} = \sum_{k=0}^{K-1} R_k$ weights. Moreover, recall that the $r$th derivative of a Dirac is a distribution that satisfies the property $\int f(t)\delta^{(r)}(t-t_0)dt = (-1)^r f^{(r)}(t_0)$.

Assume that $x(t)$ is sampled with a kernel that can reproduce polynomials of maximum degree $N \geq 2\hat{K}-1$, and call $y_n = \langle x(t), \varphi(t-n)\rangle = \sum_{k=0}^{K-1}\sum_{r=0}^{R_k-1}(-1)^r a_{k,r}\varphi^{(r)}(t_k-n)$ the observed samples. As shown in the previous section, using the polynomial reproduction formula, we can compute the first $N+1$ moments of $x(t)$ from its samples $y_n$

$$\begin{aligned}\tau_m = \sum_n c_{m,n} y_n &= \int_{-\infty}^{\infty} x(t) t^m dt \\ &= \sum_{k=0}^{K-1}\sum_{r=0}^{R_k-1} (-1)^r a_{k,r} \frac{m!}{(m-r)!} t_k^{m-r} \\ & m = 0, 1, \ldots, N \end{aligned} \tag{11}$$

where we have used the fact that $\int t^m \delta^{(r)}(t-t_0)dt = (-1)^r (m!/(m-r)!) t_0^{m-r}$.

We can, thus, say that what we observe is

$$\tau_m = \sum_{k=0}^{K-1}\sum_{r=0}^{R_k-1} (-1)^r a_{k,r} \frac{m!}{(m-r)!} t_k^{m-r}.$$

It can be shown that the filter $(1-t_k z^{-1})^M$ annihilates the signal $m^r t_k^m$, with $r \leq M-1$. Therefore, the filter $h_m$ with $z$ transform

$$H(z) = \prod_{k=0}^{K-1} (1-t_k z^{-1})^{R_k}$$

annihilates $\tau_m$. The $\hat{K}$ unknown coefficients of $h_m$ can be found solving a Yule–Walker system similar to the one in the previous section. We need at least $\hat{K}$ equations to find these coefficients; therefore, we need to know at least $2\hat{K}$ consecutive values of $\tau_m$ (this is why $N \geq 2\hat{K}-1$). From the annihilating filter, we obtain the locations $t_0, t_1, \ldots, t_{K-1}$. We then need to solve the first $\hat{K}$ equations in (11) to obtain the weights $a_{k,r}$. This is a generalized Vandermonde system which has again a unique solution given that the $t_k$s are distinct.

The above analysis can be summarized in the following theorem.

*Theorem 3:* Assume a sampling kernel $\varphi(t)$ that can reproduce polynomials of maximum degree $N \geq 2\hat{K}-1$ and of compact support $L$. An infinite-length stream of differentiated

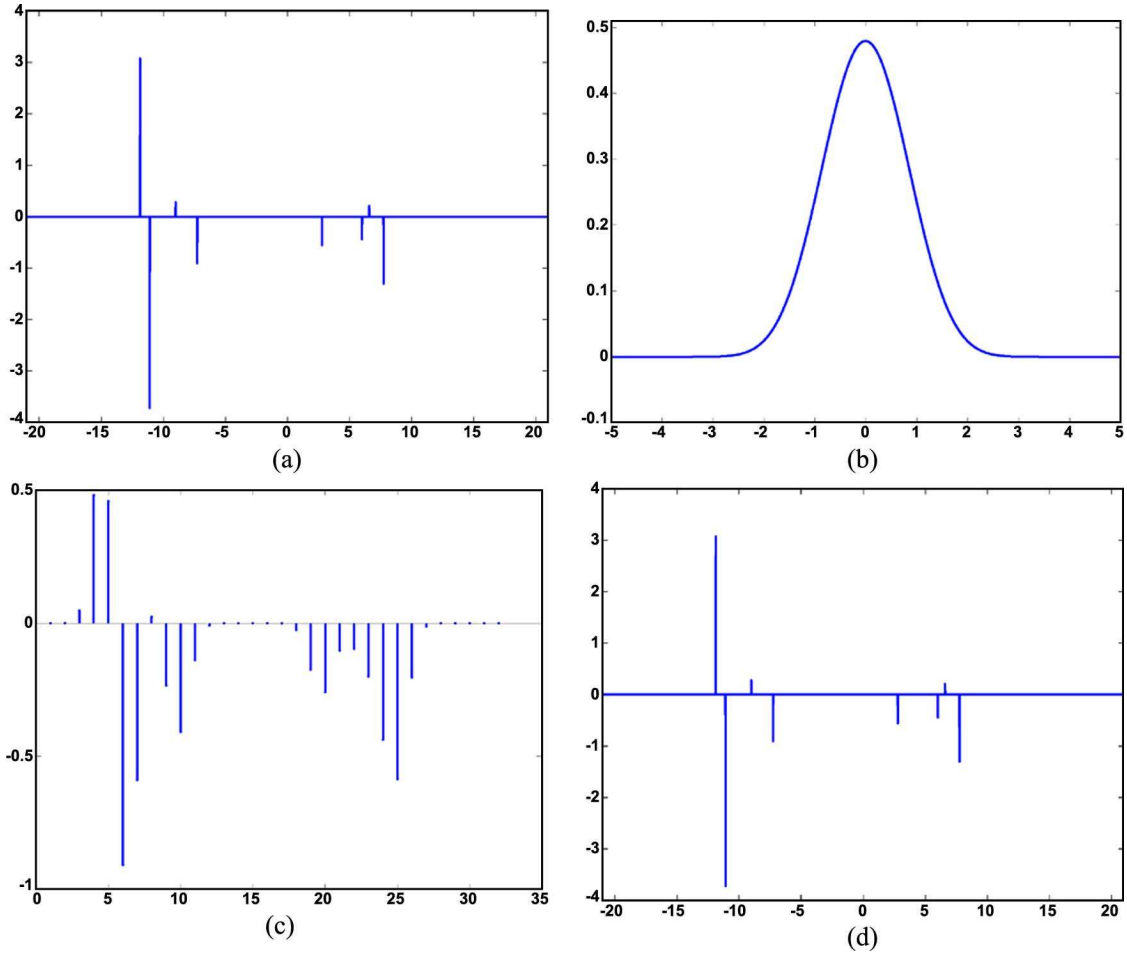


Fig. 4. Sampling of streams of Diracs. In this example, the original signal, shown in (a), is made of two groups of $K = 4$ Diracs. The sampling kernel is shown in (b) and is a B-spline $\beta_7(t)$ that can reproduce polynomials of maximum degree $2K - 1 = 7$. The observed samples are shown (c). Notice that the nonzero samples generated by the two sets of Diracs are separated by a sequence of zero samples. This allows the sampling algorithm to retrieve the two groups of $K$ Diracs sequentially. The reconstructed signal is shown in (d) and the reconstruction is exact to numerical precision.

Diracs $x(t) = \sum_{k\in\mathbb{Z}} \sum_{r=0}^{R_k-1} a_{k,r}\delta^{(r)}(t - t_k)$ is uniquely determined by the samples $y_n = \langle x(t), \varphi(t/T - n)\rangle$ if there are at most $K$ differentiated Diracs with $K$ weights in an interval of size $2KLT$.

Let us now return to the definition of signals with finite rate of innovation given in Section II

$$x(t) = \sum_{k\in\mathbb{Z}} \sum_{r=0}^{R-1} \gamma_{k,r} g_r(t - t_k). \tag{12}$$

The sampling schemes developed so far correspond to the case in which $g_0(t - t_k) = \delta(t - t_k)$ and $g_r(t - t_k) = \delta^{(r)}(t - t_k)$, $r = 1, \ldots, R - 1$. However, further extensions are possible. Assume, for instance, that $g_0(t)$ is of compact support $\hat{L}$ and that $\hat{g}_0(\omega) \neq 0$ for $\omega = 0$. Moreover, assume that $g_r(t) = g_0^{(r)}(t)$, that is, $g_r(t)$ is the $r$th-order derivative of $g_0(t)$. Then, under these conditions, the sampling of $x(t)$ is possible and can be reduced to the sampling of a stream of differentiated Diracs.

First, notice that, under the previous assumptions, $x(t)$ can be written as

$$x(t) = \left[\sum_{k\in\mathbb{Z}} \sum_{r=0}^{R-1} (-1)^r \gamma_{k,r}\delta^{(r)}(t - t_k)\right] * g_0(t).$$

Therefore, if, for simplicity, we assume that $g_0(t) = g_0(-t)$, the observed samples $y_n = \langle x(t), \varphi(t - n)\rangle$ are equivalent to those given by $y_n = \langle \sum_{k\in\mathbb{Z}} \sum_{r=0}^{R-1} (-1)^r \gamma_{k,r}\delta^{(r)}(t - t_k), g_0(t - n) * \varphi(t - n)\rangle$. Now, assume the sampling kernel $\varphi(t)$ can reproduce polynomials of degree $N$ and has compact support $L$, the new kernel $g_0(t-n) * \varphi(t-n)$ has compact support $L + \hat{L}$ and, since $\hat{g}(\omega) \neq 0$ for $\omega = 0$, can still reproduce polynomials of degree $N$ (Strang–Fix conditions are still satisfied). Therefore, if there are no more than $K$ Diracs in an interval of size $2K(L + \hat{L})$ and $N \geq 2KR - 1$, the hypotheses of Theorem 3 are satisfied, and the samples $y_n$ are sufficient to retrieve the weights $\gamma_{k,r}$ and the locations $t_k$. We can formalize this discussion with the following corollary.

*Corollary 1:* Assume a sampling kernel $\varphi(t)$ of compact support $L$ and that can reproduce polynomials of maximum degree $N$. An infinite-length signal $x(t) = \sum_{k\in\mathbb{Z}} \sum_{r=0}^{R-1} \gamma_{k,r} g_r((t - t_k)/T)$, where $g_0(t)$ is of compact support $\hat{L}$ and $\hat{g}_0(\omega) \neq 0$ for $\omega = 0$ and where $g_r(t) = g_0^{(r)}(t)$, is uniquely defined by the samples $y_n = \langle x(t), \varphi(t/T - n)\rangle$ if there are at most $K$ time shifts $t_k$ in an interval of size $2K(L+\hat{L})T$ and $N \geq 2KR-1$.

Finally, another important example of signals with finite rate of innovation, namely, piecewise polynomial signals, will be discussed in the next section.

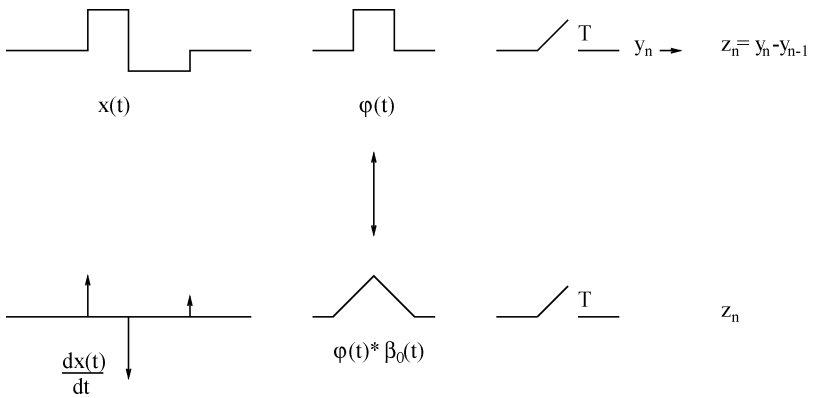


Fig. 5. Observed samples $y_n$ are given by $y_n = \langle x(t), \varphi(t/T - n)\rangle$, where the sampling kernel $\varphi(t)$ is, in this example, the box function and the original signal $x(t)$ is piecewise constant. The finite difference $y_n - y_{n-1}$ leads to the new samples $z_n$ that are equivalent to those obtained by sampling $dx(t)/dt$ with the new kernel $\varphi(t) * \beta_0(t)$ which, in this case, is a linear spline.

### C. Piecewise Polynomial Signals

Consider a piecewise polynomial signal with pieces of maximum degree $R-1 (R > 0)$. That is

$$x(t) = \sum_{n=0}^{\infty} \sum_{r=0}^{R-1} a_{n,r} (t - t_n)_+^r \tag{13}$$

where $t_+^r = \max(t, 0)^r$. Clearly, the derivative of order $R$ of $x(t)$ is a stream of differentiated Diracs given by $x^{(R)}(t) = \sum_{n=0}^{\infty} \sum_{r=0}^{R-1} r! a_{n,r} \delta^{(R-r-1)}(t - t_n)$. This means that, if we are able to relate the samples of $x(t)$ to those of $x^{(R)}(t)$, we can use Theorem 3 to reconstruct $x(t)$. This is, indeed, possible by recalling the link existing between discrete differentiation and derivation in continuous domain. More precisely, consider a function $\varphi(t)$ with Fourier transform $\hat{\varphi}(\omega)$ and consider the following difference: $\varphi(t) - \varphi(t-1)$. The Fourier transform of $\varphi(t) - \varphi(t-1)$ is

$$\varphi(t) - \varphi(t-1) \Longleftrightarrow \hat{\varphi}(\omega)(1 - e^{-j\omega}) = j\omega\hat{\varphi}(\omega)\frac{(1 - e^{-j\omega})}{j\omega} = j\omega\hat{\varphi}(\omega)\hat{\beta}_0(\omega).$$

It, thus, follows that:

$$\varphi(t) - \varphi(t-1) = \frac{d}{dt}\left[\varphi(t) * \beta_0(t)\right]. \tag{14}$$

We now use the above formula in our sampling formulation. Consider the samples $y_n = \langle x(t), \varphi(t - n)\rangle$ where $\varphi(t)$ is a generic sampling kernel. Let $z_n^{(1)}$ denote the finite difference $y_{n+1} - y_n$. It follows that:

$$\begin{aligned} z_n^{(1)} = y_{n+1} - y_n &= \langle x(t), \varphi(t - n - 1) - \varphi(t - n)\rangle \\ &\overset{(a)}{=} \left\langle x(t), -\frac{d}{dt}\left[\varphi(t - n) * \beta_0(t - n)\right]\right\rangle \\ &\overset{(b)}{=} \left\langle \frac{dx(t)}{dt}, \varphi(t - n) * \beta_0(t - n)\right\rangle \end{aligned} \tag{15}$$

where, in $(a)$, we have used (14), and $(b)$ follows from integration by parts. Thus, the samples $z_n^{(1)}$ are equivalent to those given by the inner products between the derivative of $x(t)$ and the new kernel $\varphi(t) * \beta_0(t)$. This equivalence is illustrated graphically in Fig. 5. In the same way, it is straightforward to show that the $R$th finite difference $z_n^{(R)}$ represents the samples obtained by sampling $x^{(R)}(t)$ with the kernel $\varphi(t) * \beta_{R-1}(t)$, where $\beta_{R-1}(t)$ is the B-spline of degree $R-1$.

Now, assume that $\varphi(t)$ is of compact support $L$ and that it can reproduce polynomials of maximum degree $N$. Then $\varphi(t) * \beta_{R-1}(t)$ has support $L + R$ and can reproduce polynomials of maximum degree $N + R$. Thus, if the new kernel satisfies the hypotheses of Theorem 3, the samples $z_n^{(R)}$ are a sufficient representation of $x^{(R)}(t)$ and, therefore, of $x(t)$.[2] This leads to the following theorem.

*Theorem 4:* Assume a sampling kernel $\varphi(t)$ of compact support $L$ and that can reproduce polynomials of maximum degree $N$. An infinite-length piecewise polynomial signal $x(t)$ with pieces of maximum degree $R - 1 (R > 0)$ as defined in (13) is uniquely defined by the samples $y_n = \langle x(t), \varphi(t/T - n)\rangle$ if there are at most $K$ polynomial discontinuities in an interval of size $2K(L + R)T$ and $N + R \geq 2KR - 1$.

*Proof:* Assume again $T = 1$. Given the samples $y_n$, compute the $R$th finite difference $z_n^{(R)}$. As shown before, $z_n^{(R)} = \langle x^{(R)}(t), \varphi(t - n) * \beta_{R-1}(t - n)\rangle$ and $x^{(R)}(t) = \sum_{n=0}^{\infty} \sum_{r=0}^{R-1} r! a_{n,r} \delta^{(R-r-1)}(t - t_n)$. The new kernel $\varphi(t) * \beta_{R-1}(t)$ has support $L + R$ and can reproduce polynomials of maximum degree $N + R$. Since by hypothesis $x(t)$ has at most $K$ polynomial discontinuities in an interval of size $2K(L + R)$, $x^{(R)}(t)$ has at most $K$ Diracs in that interval with a total number of weights $\hat{K} = KR$. Since we are assuming $N + R \geq 2KR - 1$, the hypotheses of Theorem 3 are satisfied, thus, the samples $z_n^{(R)}$ are sufficient to reconstruct $x^{(R)}(t)$ and, therefore, $x(t)$. □

A numerical example is shown in Fig. 6. In this case, the signal is piecewise constant and we assume that the signal can have at most two arbitrarily close discontinuities ($K = 2$). For this reason the sampling kernel must be able to reproduce polynomials of degree two and, in this example, is a quadratic spline $\beta_2(t)$. The observed samples $y_n$ are shown in Fig. 6(b) and the first-order finite difference of $y_n$ results in the samples $z_n$ which are shown in Fig. 6(c). These samples are equivalent to those obtained by sampling $dx(t)/dt$, which is a stream of Diracs, with

[2]Note that the reconstruction of $x(t)$ from $x^{(R)}(t)$ is unique since, by definition, $x(t) = 0$ for $t < t_0$ [see (13)].

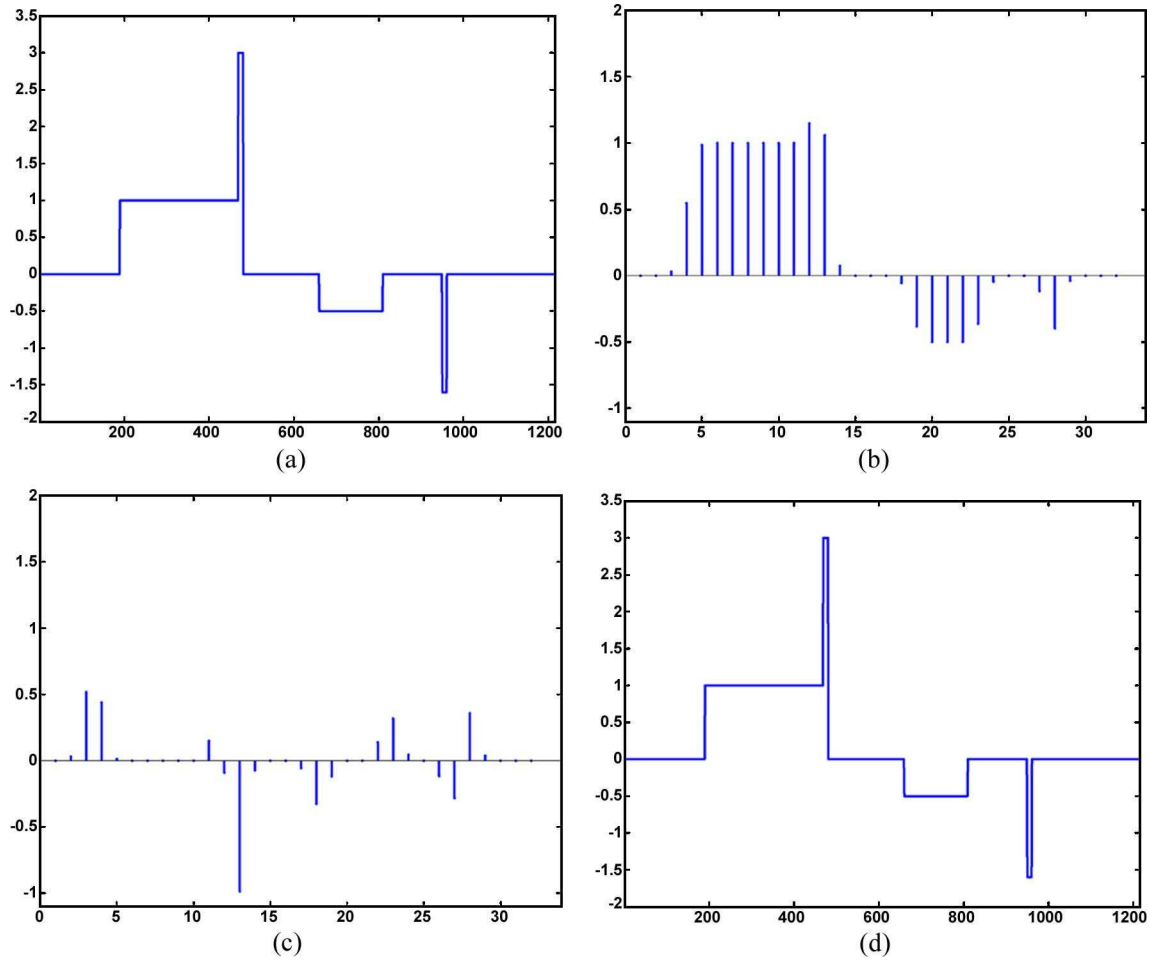


Fig. 6. Sampling of piecewise polynomial signals. In this example, the original signal is piecewise constant and is shown in (a). The signal can have up to two arbitrarily close discontinuities ($K = 2$). The sampling kernel is in this case a quadratic spline and the observed samples $y_n$ are shown in (b). The first-order finite difference of the samples $y_n$ leads to the samples $z_n$ shown in (c). From this sample, it is then possible to reconstruct the original signal exactly and the reconstructed signal is shown in (d).

the new kernel $\beta_3(t) = \beta_2(t) * \beta_0(t)$. Thus, the hypotheses of Theorem 2 and Theorem 4 are satisfied and the samples $z_n$ are sufficient to reconstruct $dx(t)/dt$ and $x(t)$. The reconstructed piecewise constant signal is shown in Fig. 6(d).

## IV. From Coarse Approximations to Infinite Resolutions

It is well known that wavelets play a fundamental role in many signal processing applications, compression being the most visible example [32]. In this section, we explore the use of scaling functions and wavelets to sample FRI signals. Moreover, we will use the multiresolution property of the wavelet transform to make a connection between our sampling results and the problem of increasing the resolution of a given signal. While it is not our aim to provide a detailed and rigorous treatment of wavelets for which we refer to standard texts [9], [17], [25], [33], we want to highlight the properties of wavelets and scaling functions that are of interest in our sampling formulation. To keep the notation as simple as possible, we concentrate only on the case of orthogonal wavelets.

A function $\varphi(t)$ is an admissible scaling function of $L_2(\mathbb{R})$ if and only if it satisfies the following three conditions [27], [30].

1) Riesz basis criterion: $A \leq \sum_{n\in\mathbb{Z}} |\hat{\varphi}(\omega + 2\pi n)|^2 \leq B$.
2) Two-scale relation: $\varphi(t) = \sqrt{2}\sum_{k\in\mathbb{Z}} h_k \varphi(2t - k)$.
3) Partition of unity: $\sum_{k\in\mathbb{Z}} \varphi(t - k) = 1$.

Condition 1 ensures that $\varphi(t)$ generates a basis for the subspace

$$V_0 = \text{span}\,\{\varphi(t-k)\}_{k\in\mathbb{Z}}, \qquad \text{with } V_0 \subset L_2(\mathbb{R}).$$

The two-scale relation guarantees that the subspaces $V_i = \text{span}\{\varphi_{i,k}(t)\}_{k\in\mathbb{Z}}$ generated by the scaled versions of $\varphi(t)$ with the usual notation $\varphi_{i,k}(t) = 2^{-i/2}\varphi(t/2^i - k)$, are embedded and form a multiresolution decomposition of $L_2(\mathbb{R})$. Finally, partition of unity ensures that such a decomposition is dense in $L_2(\mathbb{R})$ [27], [30].

In our context, the partition of unity has a second meaning: It tells us that any valid scaling function is able to reproduce at least constant functions and as such can be used to sample piecewise constant signals with no more than one discontinuity in an interval of size $2LT$.

The two-scale relation can be written in the frequency domain as follows:

$$\hat{\varphi}(\omega) = \frac{1}{\sqrt{2}}\left(\frac{1 + e^{-j\omega/2}}{2}\right)^N \hat{R}_0(e^{j\omega/2})\hat{\varphi}(\omega/2)$$

where $\hat{R}_0(e^{j\omega})$ is a residual with $\hat{R}_0(e^{j\omega}) \neq 0$ for $\omega = 0$, and $N$ must be at least equal to one because of the partition of unity. Usually, one tries to design scaling functions with a large value of $N$ since this normally guarantees that the function has a certain regularity [9]. The term $((1+e^{-j\omega/2})/2)^N$ also indicates that scaling functions satisfy Strang–Fix conditions of order $N-1$.

Given a valid scaling function, there exists a corresponding wavelet $\psi(t)$ that generates a basis of $L_2(\mathbb{R})$. The wavelet is expressed as a linear combination of shifted versions of $\varphi(t)$

$$\psi(t) = \sqrt{2}\sum_{n\in\mathbb{Z}} g_n \varphi(2t-n) \tag{16}$$

and is designed so that the scaled wavelet $\psi_{m,n}$ spans the "detail" subspace $W_i = V_i - V_{i-1}$.

In the orthogonal case, the two-scale relation of (16) can be written in the frequency domain as follows:[3]

$$\hat{\psi}(\omega) = \frac{1}{\sqrt{2}}\left(\frac{1-e^{-j\omega/2}}{2}\right)^N \hat{R}_1(e^{j\omega/2})\hat{\varphi}(\omega/2) \tag{17}$$

where $\hat{R}_1(e^{j\omega}) = -e^{-j\omega}\hat{R}_0(e^{-j(\omega+\pi)})$ and $\hat{R}_1(e^{j\omega}) \neq 0$ for $\omega = 0$. This relation tells us that a wavelet has $N$ zeros at $\omega = 0$ and using the moment property of the Fourier transform, we have

$$\int_{-\infty}^{\infty} t^m \psi(t)dt = 0, \qquad m = 0, 1, \ldots, N-1.$$

In other words, while the scaling function reproduces polynomial of up to degree $N-1$, the corresponding wavelet 'kills' polynomials of the same order. This is, in fact, the much celebrated vanishing moment property of the wavelet transform.

Consider now the representation of a signal $x(t) \in L_2(\mathbb{R})$ in terms of wavelets and scaling functions

$$x(t) = \sum_{n=-\infty}^{\infty} y_{J,n}\varphi_{J,n}(t) + \sum_{m=-\infty}^{J}\sum_{n=-\infty}^{\infty} d_{m,n}\psi_{m,n}(t) \tag{18}$$

and assume that $x(t)$ and $\varphi(t)$ satisfy the hypotheses of the theorems in Section III. That is, $x(t)$ is a stream of Diracs or a piecewise polynomial signal with a local finite rate of innovation, and $\varphi(t)$ is a compact support scaling function that can reproduce polynomials of a certain degree. Then the sampling theorems of Section III ensure that, for a proper choice of $J$ (i.e., $T = 2^J$), the inner products $y_{J,n}$ of (18) are sufficient to characterize $x(t)$ or, in other words, that the finite resolution version $x_J(t) = \sum_{n=-\infty}^{\infty} y_{J,n}\varphi_{J,n}(t)$ is sufficient to reconstruct the signal exactly. This means that by knowing $x(t)$ at a finite resolution, we can infer the value of the wavelet coefficients $d_{m,n}$ with $m \leq J$ and, therefore, arbitrarily increase the resolution of our approximation to eventually recover the original continuous-time signal. For example, in the case of a stream of Diracs, we can state the following corollary to Theorem 2.

*Corollary 2:* Assume a scaling function $\varphi(t)$ of compact support $L$ and that can reproduce polynomials of maximum degree $N \geq 2K-1$. The coarse approximation $x_J(t) = \sum_{n=-\infty}^{\infty} y_{J,n}\varphi_{J,n}(t)$ of a stream of Diracs $x(t)$ is a sufficient representation of $x(t)$ if there are at most $K$ Diracs in an interval of size $2KL2^J$.

An example of the reconstruction of a piecewise polynomial signal from its coarse representation is illustrated in Fig. 7. We consider a discrete-time piecewise linear signal with $N = 128$ points Fig. 7(a). The signal is sampled with a Daubechies scaling function with two vanishing moments and the sampling period is $T = 8$ points ($T = 2^J$). The distance between two consecutive discontinuities is such that the hypotheses of Theorem 4 are satisfied. The 16 sample values are shown in Fig. 7(b). The coarse approximation of the signal (what we have called $x_J(t)$) is shown in Fig. 7(c). The reconstruction of $x(t)$ with the annihilating filter method is instead exact to machine precision and is shown in Fig. 7(d).

Notice that the coarse version $x_J(t)$ of $x(t)$ is the version that we would normally obtain when using a classical sampling scheme where the reconstruction process is linear and the whole sampling operation is equivalent to projecting $x(t)$ onto the subspace $V_J = \operatorname{span}\{\varphi_{J,n}(t)\}_{n\in\mathbb{Z}}$. Thus, this example gives a clear indication of the gain that one obtains by replacing a classical linear reconstruction method with a nonlinear, yet with reasonable complexity, scheme based on the annihilating filter method.

Scaling functions and wavelets are intimately related as indicated in particular by (16). It seems, therefore, natural to imagine that, in some cases, wavelets can be used to sample FRI signals in much the same way as scaling functions. This is, in fact, true in the case of piecewise polynomial signals.

Assume $x(t)$ is a piecewise polynomial signal with pieces of maximum degree $N-1 (N > 0)$ and consider the samples $d_{0,n} = \langle x(t), \psi(t-n)\rangle$, where we are assuming $T = 1$ and $\psi(t)$ is a wavelet with at least $N$ vanishing moments. Denote by $L$ the support of $\psi(t)$ and assume that there is at most one discontinuity in an interval of size $2L$. It follows that:

$$\begin{aligned} d_{0,n} &= \langle x(t), \psi(t-n)\rangle \\ &\overset{(a)}{=} \left\langle x(t), \frac{d^N}{dt^N}\left[\frac{1}{2^{2N}\sqrt{2}}\beta_{N-1}(2t-n)\right.\right. \\ &\qquad \left.\left. *\varphi(2t-n) * R_1(2t-n)\right]\right\rangle \\ &\overset{(b)}{=} (-1)^N \left\langle \frac{d^N x(t)}{dt^N}, \frac{1}{2^{2N}\sqrt{2}}\beta_{N-1}(2t-n)\right. \\ &\qquad \left. *\varphi(2t-n) * R_1(2t-n)\right\rangle \end{aligned}$$

where in $(a)$, we used the fact that

$$\begin{aligned} \hat{\Psi}(\omega) &= \frac{1}{\sqrt{2}}\left(\frac{1-e^{-j\omega/2}}{2}\right)^N \hat{R}_1(e^{j\omega/2})\hat{\varphi}(\omega/2) \\ &= \frac{(j\omega)^N}{2^{2N}\sqrt{2}}\hat{\beta}_{N-1}(\omega/2)\hat{R}_1(e^{j\omega/2})\hat{\varphi}(\omega/2) \end{aligned}$$

and $(b)$ follows from integration by parts. Since $\hat{R}_1(e^{j\omega/2}) \neq 0$ for $\omega = 0$, the new kernel can reproduce polynomials of maximum degree $2N-2$, and it is easy to show that it has the same support as $\psi(t)$. The signal $d^N x(t)/dt^N$ is a stream of differentiated Diracs and by hypothesis there is at most one Dirac in an interval of size $L$ with a total number of weights $\hat{K} = N-1$. This means that the hypotheses of Theorem 4 are satisfied and

[3]A similar relation applies to the biorthogonal case, as well.

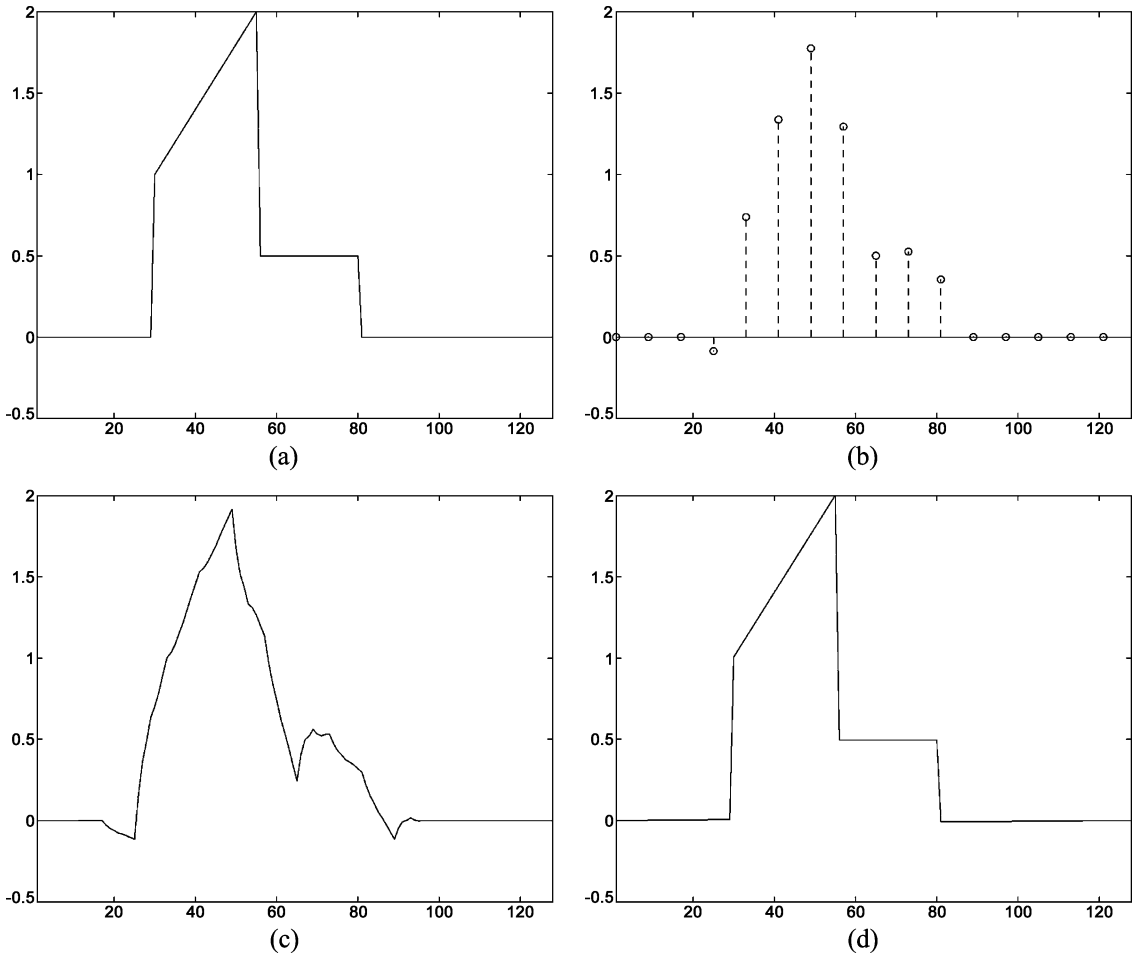


Fig. 7. Illustration of the reconstruction of a piecewise linear signal with the sampling scheme presented in Sections III and IV. (a) Original discrete-time piecewise linear signal. In this case the original signal has 128 points. (b) Sample values obtained with a Daubechies filter with two vanishing moments. In this example, the sampling period is $T = 8$; thus, we are taking only 16 samples. (c) Coarse reconstruction $x_J(t)$ of the signal or projection onto the subspace spanned by $\varphi_{J,n}(t) = 2^{-J/2}\varphi(t/2^J - n)$. (d) Reconstruction with the annihilating filter method using the 16 samples of (b).

$d_{0,n}$ is a sufficient representation of $x(t)$. We can, thus, summarize these findings with the following corollary to Theorem 4.

*Corollary 3:* Assume a wavelet function $\psi(t)$ with $N$ vanishing moments and compact support $L$. An infinite-length piecewise polynomial signal $x(t)$ with pieces of maximum degree $N-1 (N > 0)$ is uniquely defined by the samples $y_n = \langle x(t), \psi(t/T - n)\rangle$ if there is at most one discontinuity in an interval of size $2LT$.

The above corollary, thus, indicates that the knowledge of a piecewise polynomial signal at a single scale is normally sufficient to reconstruct the entire signal exactly.

## V. Exponential Case

We have seen in Sections III and IV that we can use the property that $\varphi(t)$ reproduces polynomials to reduce our sampling problem to that of finding the coefficients $a_k$ and $t_k$ of the discrete signal $\tau_m = \sum_{k=0}^{K-1} a_k t_k^m$, $m = 0, 1, \ldots, N$ and this is achieved using the annihilating filter method. The interesting point is that the annihilating filter method can also be used for an observed signal of the form $s_m = \sum_{k=0}^{K-1} a_k e^{\alpha_m t}$ and $\alpha_m = \alpha_0 + m\lambda$. For this reason, FRI signals can be sampled and reconstructed using kernels that reproduce exponentials. The reconstruction scheme is the same as in the polynomial case. First, the signal $s_m = \sum_{k=0}^{K-1} a_k e^{\alpha_m t}$ is estimated from the samples $y_n$, then locations and amplitudes of the Diracs are retrieved from $s_m$.

Assume that our kernel is of compact support and that it is able to reproduce exponential of the form $e^{\alpha_m t}$ with $\alpha_m = \alpha_0 + m\lambda$ and $m = 0, 1, \ldots, N$. For instance, $\varphi(t)$ is an E-spline $\beta_{\vec{\alpha}}(t)$ with $\vec{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_N)$ and $\alpha_m = \alpha_0 + m\lambda$ or a composite function $\varphi(t) * \beta_{\vec{\alpha}}(t)$. Consider again a stream of $K$ Diracs $x(t) = \sum_{k=0}^{K-1} a_k \delta(t - t_k)$. The samples $y_n$ are then given by $y_n = \langle x(t), \varphi(t - n)\rangle$ and, using (5), it follows that:

$$s_m = \sum_n c_{m,n} y_n = \int_{-\infty}^{\infty} x(t) e^{(\alpha_0 + m\lambda)t} dt$$
$$= \sum_{k=0}^{K-1} a_k e^{\alpha_0 t_k} e^{m\lambda t_k}, \qquad m = 0, 1, \ldots, N.$$

This means that, as in the polynomial case, proper linear combinations of the samples $y_n$ lead to a signal $s_m$ of the form $s_m = \sum_{k=0}^{K-1} \hat{a}_k u_k^m$, where $\hat{a}_k = a_k e^{\alpha_0 t_k}$ and $u_k = e^{\lambda t_k}$. Since $m$ is an integer, the new measurements $s_m$ have a form similar to the measurements $\tau_m$ of the polynomial case and, as

in that case, the exponentials $u_k$ and the amplitudes $\hat{a}_k$ are retrieved using the annihilating filter method. Finally, from the $\hat{a}_k$s and $u_k$s, we retrieve the amplitudes and locations of the original Diracs. Again, since the kernel has compact support, this reconstruction algorithm can be applied to any stream of Diracs with local rate of innovation $\rho_{2KTL}(t) \leq 1/TL$. We can, thus, summarize the above analysis as follows.

*Theorem 5:* Assume a sampling kernel $\varphi(t)$ of compact support $L$ and that can reproduce exponentials $e^{(\alpha_0+m\lambda)t}$ with $m = 0, 1, \ldots, N$ and $N \geq 2K-1$. An infinite-length stream of Diracs $x(t) = \sum_{n\in\mathbb{Z}} a_n \delta(t-t_n)$ is uniquely determined from the samples defined by $y_n = \langle x(t), \varphi(t/T - n)\rangle$ if there are at most $K$ Diracs in an interval of length $2KTL$.

Notice that this theorem reduces to Theorem 2 when $\alpha_0 = \lambda = 0$. Moreover, notice that, by using the same procedure indicated in the previous sections, it is possible to extend the above results to the case of stream of differentiated Diracs and of piecewise polynomial signals. We omit the proofs of these extensions since they are straightforward.

The exponent $\alpha_m$ in the exponential $e^{\alpha_m t}$ can be either real or complex. An interesting case is when $\alpha_m$ is purely imaginary and is given by $\alpha_m = j\omega_0 m$, $m = 0, 1, \ldots, N$. In this case, in fact, we have

$$s_m = \sum_n c_{m,n} y_n = \int_{-\infty}^{\infty} x(t) e^{jm\omega_0 t} dt = \hat{x}(m\omega_0).$$

In other words, when $\alpha_m = j\omega_0 m$, it is possible to retrieve the Fourier coefficients $\hat{x}(0), \hat{x}(\omega_0), \ldots, \hat{x}(N\omega_0)$ of $x(t)$ from its samples. This partial knowledge of the spectrum of $x(t)$, which is incidentally not bandlimited, is sufficient to reconstruct the original signal exactly as shown by Theorem 5 and its extensions. This analysis also establishes an even tighter link between our sampling scheme and the one presented in [34] which is also based on the reconstruction of $x(t)$ from its Fourier coefficients. Our method, however, remains very attractive in many situations since it is local and can be used with a wider range of different kernels as shown in the next section.

## VI. Kernels With Rational Fourier Transform

Consider a classical continuous-time linear time-invariant system where the input $x(t)$ and output $y(t)$ are related by linear differential equations of the form

$$\frac{d^n y}{dt^n} + a_{n-1}\frac{d^{n-1} y}{dt^{n-1}} + \cdots + a_0 y(t) = b_m \frac{d^m x}{dt^m} + b_{m-1}\frac{d^{m-1} x}{dt^{m-1}} + \cdots + b_0 x(t).$$

The transfer function of this system is rational and is similar to the one in (6). This system represents the basic building block in classical system theory, and is sufficiently general to model most linear electrical, mechanical or electro-mechanical systems. Most important, many such systems can be used to sample signals with finite rate of innovation. The reason why we can use such kernels in our sampling formulation is that we can *convert* a kernel $\varphi(t)$ with rational Fourier transform as in (6) into a kernel that reproduces exponentials. This is achieved by filtering the samples $y_n = \langle x(t), \varphi(t-n)\rangle$[4] with an FIR filter with $z$ transform $H(z) = \prod_{m=0}^{N}(1 - e^{\alpha_m} z)$.

For example, assume that $\hat{\varphi}(\omega) = \frac{1}{j\omega - \alpha}$ and $y_n = \langle x(t), \varphi(t-n)\rangle$. Then $H(z) = (1 - e^{\alpha} z)$, and we have

$$\begin{aligned} z_n = h_n * y_n &= y_n - e^{\alpha} y_{n+1} \\ &\overset{(a)}{=} \langle x(t), \varphi(t-n) - e^{\alpha}\varphi(t-n-1)\rangle \\ &\overset{(b)}{=} \langle x(t), \beta_\alpha(t-n)\rangle \end{aligned}$$

where $(a)$ follows from the linearity of the inner product and $(b)$ from the fact that the Fourier transform of $\varphi(t-n) - e^{\alpha}\varphi(t-n-1)$ is $e^{-j\omega n}\hat{\beta}_\alpha(\omega)$. Therefore, by filtering the samples $y_n$ with the filter $H(z) = (1 - e^{\alpha} z)$, we obtain a new set of samples $z_n$ that are equivalent to those that would have been obtained by sampling the original signal $x(t)$ with the E-spline $\beta_\alpha(t)$.

Likewise, when the original kernel has $N+1$ poles at locations $\vec{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_N)$, by filtering the samples $y_n = \langle x(t), \varphi(t-n)\rangle$ with the filter $H(z) = \prod_{m=0}^{N}(1 - e^{\alpha_m} z)$ we have that $z_n = h_n * y_n = \langle x(t), \beta_{\vec{\alpha}}(t-n)\rangle$ and the new kernel is of compact support and reproduces the exponentials $\{e^{\alpha_0 t}, e^{\alpha_1 t}, \ldots, e^{\alpha_N t}\}$.

In the most general case, the kernel has a frequency response as in (6) and by filtering the samples with the digital filter $H(z) = \prod_{m=0}^{N}(1 - e^{\alpha_m} z)$, we obtain a new kernel with Fourier transform $\prod_{i=0}^{I}(j\omega - b_i)\hat{\beta}_{\vec{\alpha}}(\omega)$. Functions with such Fourier transform are sometimes called generalized E-spline [28] and clearly are still able to reproduce the exponentials $\{e^{\alpha_0 t}, e^{\alpha_1 t}, \ldots, e^{\alpha_N t}\}$. Moreover, notice that since we are assuming $I < N$, these new kernels have compact support.

Thus, the above analysis together with Theorem 5 and its extensions, allows us to say that acquisition devices with rational Fourier transform can be used to sample signal with FRI. The condition that need to be satisfied, however, is that the poles of this system are located at $\alpha_m = \alpha_0 + m\lambda$, $m = 0, 1, \ldots, N$, and the necessary number of such poles depends as usual on the local complexity of the signal. Clearly, these linear systems also need to be stable or at least meta-stable. For this reason, the real part of their poles has to be negative or zero.

As an example, we now show how to estimate a piecewise constant signal at the output of an $RC$ circuit.

*Example 1:* Consider the classical $RC$ circuit shown in Fig. 8 and call $H(\omega) = \alpha/(\alpha + j\omega)$ with $\alpha = 1/RC$ its transfer function. Assume that the input voltage is a step function $x(t) = Au(t - t_0)$. The output $y(t) = h(t) * x(t)$ is given by $y(t) = Au(t-t_0) - Ae^{-\alpha(t-t_0)}u(t-t_0)$. The output voltage is then uniformly sampled with sampling period $T = 1$ leading to the samples $y_n = Au(n-t_0) - Ae^{-\alpha(n-t_0)}u(n-t_0)$. Alternatively, we can say that $y_n = \langle x(t), \varphi(t-n)\rangle$ with $\varphi(t) = h(-t)$.[5] Our aim is to retrieve $x(t)$ from the samples $y_n$. Notice that at this stage neither the original signal $x(t)$ nor the sampling kernel $\varphi(t)$ have compact support. Yet, we know from the theories developed in the previous sections that $x(t)$ can be reconstructed from its samples. Two actions are needed in order to reconstruct

[4] We are assuming $T = 1$ for simplicity.

[5] Recall that the identity $\varphi(t) = h(-t)$ implies that $\hat{\varphi}(\omega) = H(-\omega)$.

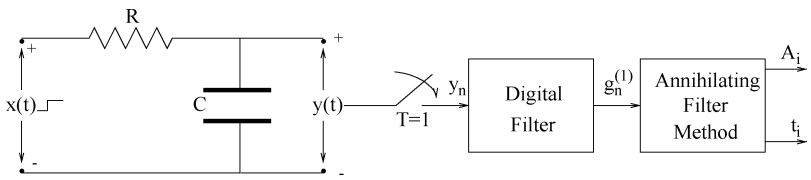


Fig. 8. Reconstruction of a piecewise constant signal $x(t)$ from its samples $y_n$. In this example, the acquisition device is a typical $RC$ circuit. The input voltage $x(t)$ is piecewise constant and the output $y(t)$ is uniformly sampled. From the samples $y_n$ and after a proper digital filtering, it is possible to retrieve the exact locations $t_i$ and amplitudes $A_i$ of the discontinuities of $x(t)$.

it. First, the sampling kernel needs to be converted into an exponential spline. Second, the derivative of the signal needs to be computed. As indicated before, this is achieved by filtering the samples $y_n$ with a proper digital filter. In this case, the digital filter has $z$ transform $H_1(z) = (z-1)(e^{\alpha}z-1)$.

The filtering of $y_n$ with $(e^{\alpha}z-1)$ yields

$$g_n = e^{\alpha}y_{n+1} - y_n = \langle x(t), e^{\alpha}\varphi(t-n-1) - \varphi(t-n)\rangle = \langle x(t), \alpha\beta_{\alpha}(t-n)\rangle .$$

The filtering of $g_n$ with $(z-1)$ gives [see also (15)]

$$g_n^{(1)} = g_{n+1} - g_n = \left\langle \frac{dx(t)}{dt}, \alpha\beta_{\alpha}(t-n) * \beta_0(t-n)\right\rangle .$$

Thus, the new samples $g_n^{(1)}$ are equivalent to those that one would obtain by sampling $dx(t)/dt$ with the new kernel $\varphi_{\alpha}(t) = \alpha\beta_{\alpha}(t) * \beta_0(t)$. The signal $dx(t)/dt$ is a Dirac centered at $t_0$ and with amplitude $A$. The new kernel $\varphi_{\alpha}(t) = \alpha\beta_{\alpha}(t) * \beta_0(t)$ is of compact support $L = 2$ and can reproduce a constant function and the exponential $e^{\alpha t}$. This means that the hypotheses of Theorem 5 are satisfied and that we can retrieve $A$ and $t_0$ from the samples $g_n^{(1)}$. To show how to retrieve these two parameters, recall that $y_n = Au(n-t_0) - Ae^{-\alpha(n-t_0)}u(n-t_0)$ and assume for simplicity that $t_0 \in ]0,1]$. Then

$$Y(z) = \frac{A}{z-1} - \frac{Ae^{\alpha t_0}}{e^{\alpha}z-1}$$

and $G^{(1)}(z) = H_1(z)Y(z) = (z-1)(e^{\alpha}z-1)Y(z) = A(e^{\alpha t_0}-1) + A(e^{\alpha}-e^{\alpha t_0})z$. For this reason, we have

$$\frac{1}{e^{\alpha}-1}\sum_n g_n^{(1)} = \frac{1}{e^{\alpha}-1}G^{(1)}(z)\Big|_{z=1} = A$$

and

$$\frac{1}{1-e^{-\alpha}}\sum_n e^{\alpha n}g_n^{(1)} = \frac{1}{1-e^{-\alpha}}G^{(1)}(z)\Big|_{z=e^{-\alpha}} = Ae^{\alpha t_0}.$$

Thus, we retrieve the amplitude $A$ from the first sum and the location $t_0$ from the second one.

Notice that with this $RC$ circuit we can sample any piecewise constant signal that has at most one discontinuity in an interval of length $4T$. To sample signals with higher local rate of innovation, we need an electrical circuit with more than one pole. For instance, to sample a piecewise constant signal that may contain two arbitrarily close discontinuities, we need a third-order $RC$ circuit with three poles at location $\alpha$, $2\alpha$, and $3\alpha$.

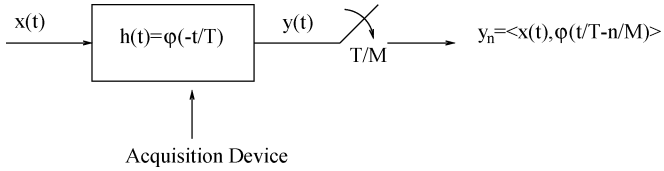


Fig. 9. Oversampled acquisition for noise reduction. In the oversampled scenario, the sampling period is reduced to $T/M$, where $M$ is integer and represents the oversampling factor.

## VII. Noisy Scenario

In many practical situations, the samples we have access to are corrupted by noise and the usual assumption is that the noise is additive, white and Gaussian. Thus, we measure $\hat{y}_n = y_n + e_n$ where $e_n$ is the additive noise.[6] We have seen that in our reconstruction schemes we first estimate the signal's moments from the samples and then the signal is reconstructed from its moments using techniques encountered in spectral estimation. This second problem, i.e., robust estimation of the signal from its noisy moments, has been extensively studied, in particular, by the array processing community and any of the techniques developed in that context can, in principle, be used in our scenario, as well. For an insightful review of those techniques, we refer to the recent paper [13] and to the book [22].

In this section, instead, we concentrate on the estimation of the moments from the noisy samples. In particular, we present a simple algorithm that reduces the estimation error by oversampling. We also show that the proposed estimator is asymptotically unbiased.

The new sampling setup is shown in Fig. 9. Here, the support of the kernel remains the same, but the sampling period is reduced by an integer factor $M$. The observed (noiseless) samples can be written as $y_n = \langle x(t), \varphi(t/T - n/M)\rangle$. From now on, we assume that $x(t)$ is a stream of $K$ Diracs, $x(t) = \sum_{k=0}^{K-1} a_k\delta(t-t_k)$, that the signal has finite length and call the finite set of samples generated by the signal $y_0, y_1, \ldots, y_{MJ-1}$. Moreover, we assume that the sampling kernel has compact support and can reproduce polynomials up to degree $N$. This means that, if we choose a function $\tilde{\varphi}(t)$ that is quasi-biorthonormal of order $M$ with $\varphi(t)$, we have

$$\sum_{n\in\mathbb{Z}} c_{m,Mn+i}\varphi(t/T - n - i/M) = t^m \qquad m = 0,1,\ldots N \text{ and } i = 0,1,\ldots M-1 \quad (19)$$

[6]In our setup, we are assuming that the input continuous-time signal $x(t)$ is noiseless and that noise is introduced by the acquisition device after sampling. In many practical situations, the original signal may already have noise. In this case, if the sampling kernels are orthogonal with respect to their shifts, the noise is still uncorrelated after sampling, but this is not the case otherwise. This latter situation (e.g., correlated noise) is not considered here.

where

$$c_{m,Mn+i} = \frac{1}{T}\int t^m \tilde{\varphi}(t/T - n - i/M)dt.$$

Note that $\tilde{\varphi}(t)$ can be chosen to have compact support.

Let us concentrate for the moment on the case $M = 2$, or the case where we are oversampling by a factor two. Since two consecutive samples are now obtained by shifting the kernel by $T/2$ rather than $T$, we can separate even and odd samples and treat them independently. In particular, the even and odd samples are given by

$$\begin{aligned} y_{2n} &= \langle x(t), \varphi(t/T - n)\rangle \\ y_{2n+1} &= \langle x(t), \varphi(t/T - n - 1/2)\rangle . \end{aligned} \tag{20}$$

For this reason, we can write $\tau_m$ [see (7) and (20)] in two different ways. First, using (19) with $M = 2$ and $i = 0$

$$\begin{aligned} \tau_m &= \sum_{n=0}^{J-1} c_{m,2n} y_{2n} = \left\langle x(t), \sum_n c_{m,2n}\varphi(t/T - n)\right\rangle \\ &= \int_{-\infty}^{\infty} \sum_{k=0}^{K-1} a_k \delta(t - t_k) t^m dt = \sum_{k=0}^{K-1} a_k t_k^m, \quad m = 0, 1, \ldots, N. \end{aligned}$$

Second, using (19) with $M = 2$ and $i = 1$

$$\begin{aligned} \tau_m &= \sum_{n=0}^{J-1} c_{m,2n+1} y_{2n+1} \\ &= \left\langle x(t), \sum_n c_{m,2n+1}\varphi(t/T - n - 1/2)\right\rangle \\ &= \int_{-\infty}^{\infty} \sum_{k=0}^{K-1} a_k \delta(t - t_k) t^m dt = \sum_{k=0}^{K-1} a_k t_k^m, \quad m = 0, 1, \ldots, N. \end{aligned}$$

In other words, in the oversampled case, the moments of $x(t)$ can be retrieved from either the even or the odd samples.

In the noisy scenario, we can use this fact to reduce the effect of noise. Call $\hat{y}_n = \langle x(t), \varphi(t/T - n/2)\rangle + e_n$ the noisy samples. We separate the even from the odd samples and estimate the moments of $x(t)$ from these two sets independently. We have

$$\hat{\tau}_m^{(i)} = \sum_{n=0}^{J-1} c_{m,2n+i}\hat{y}_{2n+i} = \tau_m + \sum_{n=0}^{J-1} c_{m,2n+i} e_{2n+i}, \quad i = 0, 1.$$

Our final estimated moment is then the average

$$\begin{aligned} \bar{\tau}_m &= \frac{1}{2}\left(\hat{\tau}_m^{(0)} + \hat{\tau}_m^{(1)}\right) \\ &= \tau_m + \frac{1}{2}\left(\sum_{n=0}^{J-1} c_{m,2n} e_{2n} + \sum_{n=0}^{J-1} c_{m,2n+1} e_{2n+1}\right). \end{aligned}$$

The extension to the case where $M > 2$ is now clear. The observed samples $\hat{y}_n = \langle x(t), \varphi(t/T - n/M)\rangle + e_n$ are divided into their polyphase components $\hat{y}_n^{(i)} = \hat{y}_{Mn+i}$, $i = 0, 1, \ldots, M - 1$. The moments of $x(t)$ are retrieved from each set of samples $\hat{y}_n^{(i)}$ independently and the final estimated moments are given by the average

$$\begin{aligned} \bar{\tau}_m &= \frac{1}{M}\sum_{i=0}^{M-1} \hat{\tau}_m^{(i)} \\ &= \tau_m + \frac{1}{M}\left(\sum_{n=0}^{J-1}\sum_{i=0}^{M-1} c_{m,Mn+i} e_{Mn+i}\right), \quad m = 0, 1, \ldots, N. \end{aligned}$$

Since the noise is assumed i.i.d Gaussian with variance $\sigma^2$, the mean-square estimation error is

$$E\left[(\bar{\tau}_m - \tau_m)^2\right] = \frac{1}{M^2}\sigma^2\left(\sum_{n=0}^{J-1}\sum_{i=0}^{M-1} c_{m,Mn+i}^2\right) \le \frac{1}{M} J A^2 \sigma^2 \tag{21}$$

where $A = \max |c_{m,Mn+i}|$ and, since $m$ and $n$ are bounded and $\tilde{\varphi}(t)$ is of compact support, $A$ is upper bounded independently of $M$. Therefore, (21) indicates that, on average, we reduce the effect of noise by a factor $\sqrt{M}$ and that our estimation of the moments is asymptotically consistent and unbiased. Moreover, since the reconstruction of $x(t)$ is obtained through operations that are continuous for small perturbations, the reconstruction of $x(t)$ from the moments is asymptotically unbiased, as well. For a precise proof of this last statement we refer to [22] and [23].

TABLE I
EFFECT OF THE OVERSAMPLING IN THE RECONSTRUCTION OF A STREAM OF DIRACS. THE SAMPLES HAVE BEEN CORRUPTED WITH ADDITIVE GAUSSIAN NOISE. HERE, $M$ INDICATES THE OVERSAMPLING FACTOR. THE ERROR IS MEASURED AS THE AVERAGE ABSOLUTE DIFFERENCE BETWEEN THE TRUE AND THE ESTIMATED DIRACS' LOCATIONS AND IS EXPRESSED AS A FUNCTION OF THE SAMPLING PERIOD $T$. THUS, AN AVERAGE ERROR $0.25T$ MEANS THAT THE RECONSTRUCTION HAS AN AVERAGE ACCURACY OF A QUARTER OF THE SAMPLING PERIOD

| | M=1 | M=2 | M=4 | M=8 |
|---|---|---|---|---|
| $\sigma = 10^{-1}$ | 0.25T | 0.22T | 0.19T | 0.17T |
| $\sigma = 10^{-2}$ | 0.15T | 0.10T | 0.07T | 0.041T |
| $\sigma = 5 \cdot 10^{-3}$ | 0.075T | 0.06T | 0.046T | 0.03T |

A quantitative analysis of the benefit of oversampling is presented in Table I. In this experiment, we assume that the input signal is made of $K = 2$ Diracs with fixed amplitude one and that the Diracs are uniformly distributed in $t \in [-16T, 16T]$. The observed samples are corrupted by additive Gaussian noise with variance $\sigma^2$. The sampling kernel is given by the convolution of the E-spline $\beta_\alpha(t)$ with $\alpha = -0.1$ and a B-spline of order $2K - 1$. This means that the kernel can reproduce $t^m e^{\alpha t}$ with $m = 0, 1, \ldots, 2K - 1$ and $\alpha = -0.1$. In this context, the E-spline is only used to increase the stability of the reconstruction algorithm since the function $t^m e^{\alpha t}$ is more stable than $t^m$. The table shows the average absolute error in estimating the Diracs' locations over 1000 experiments for different values of the noise. The reconstruction algorithm operates as follows: the noisy samples are hard thresholded with a threshold equal to $3\sigma$, then the moments are estimated using the above scheme. Finally, locations and amplitude of the Diracs are retrieved from the moments using the usual approach of Section III.

We can notice from the table that the reconstruction error is indeed reduced by oversampling. Also, the results are consistent with the theory. For instance, the two cases ($M = 1, \sigma =$

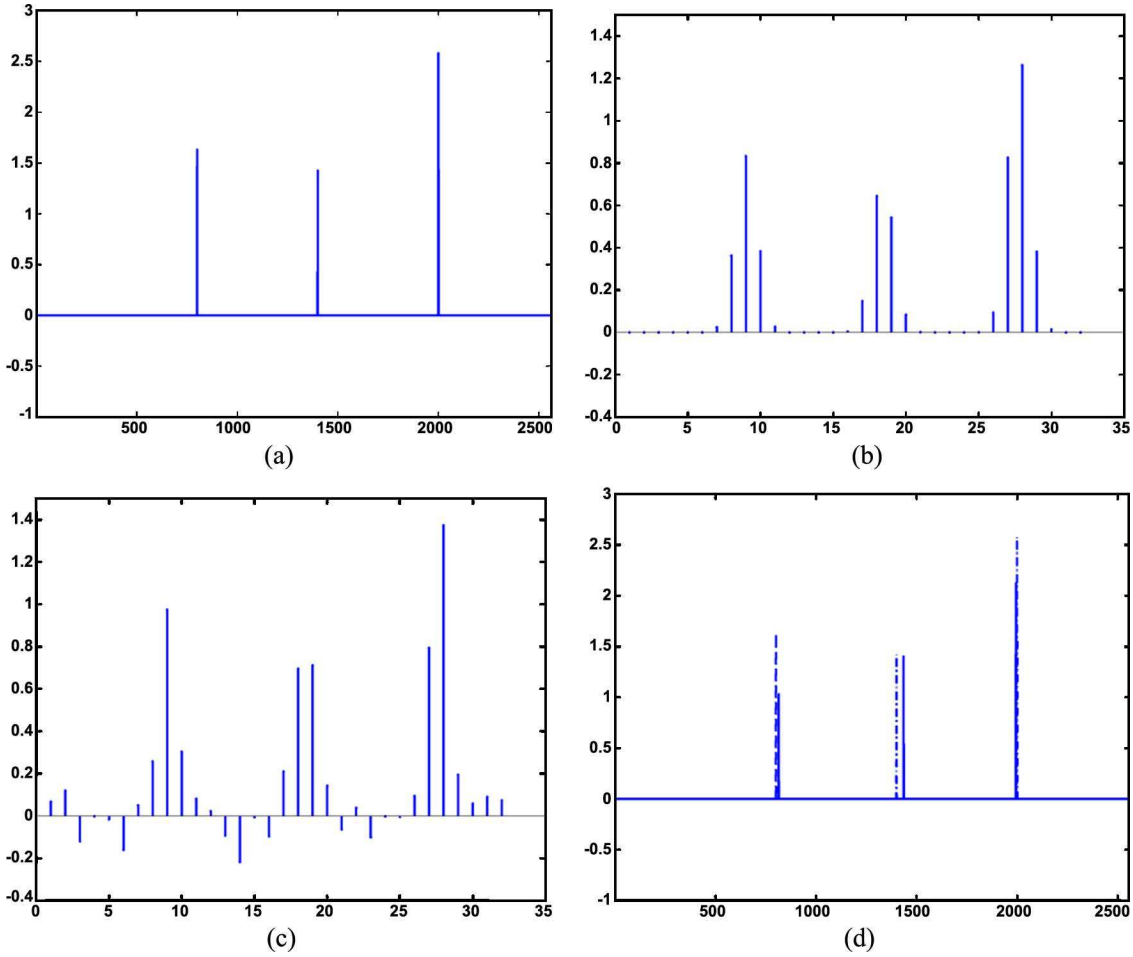


Fig. 10. Illustration of the reconstruction of a stream of $K = 3$ Diracs in the noisy case. Notice that in this simulation we are not using oversampling ($M = 1$). (a) Original signal. (b) Sample values before adding noise. (c) Noisy samples. In this case, the noise is additive white Gaussian with standard deviation $\sigma = 0.1$. (d) Reconstructed signal (continuous line) and original signal (dashed line).

TABLE II
LOCAL COMPLEXITY VERSUS RECONSTRUCTION FIDELITY. IN THIS SIMULATIONS WE ARE NOT USING OVERSAMPLING ($M = 1$)

| | K=1 | K=2 | K=4 | K=8 |
|---|---|---|---|---|
| $\sigma = 10^{-1}$ | 0.17T | 0.25T | 2.5T | 3.2T |
| $\sigma = 10^{-2}$ | 0.02T | 0.15T | T | 2.8T |
| $\sigma = 10^{-3}$ | $10^{-4}$T | 0.02T | 0.6T | 1.7T |

$5 \cdot 10^{-3}$) and ($M = 4, \sigma = 10^{-2}$) lead to roughly the same performance as anticipated by the theory. An example of the reconstruction is shown in Fig. 10. In this case, we do not use oversampling ($M = 1$), the input signals has $K = 3$ Diracs and the noise has standard deviation $\sigma = 0.1$.

It is also of interest to analyze the stability of the algorithm for different values of $K$. Table II shows the average error for the case $K = 1, 2, \ldots, 4$. As it can be noticed, the reconstruction process is very stable for small values of $K$. This seems to indicate that, in the case of noisy measurements and strong noise, our reconstruction scheme is reliable when the input signal has a low local rate of innovation (e.g., at most one Dirac in an interval of size $2LT$).

To conclude, we show in Fig. 11 a last simulation for the case of piecewise constant signals. In this example, the sampling kernel is the box function and the algorithm tries to estimate one discontinuity per time. The noise standard deviation is $\sigma = 10^{-2}$. The reconstruction shown is in fact quite faithful with an SNR = 27 dB.

## VIII. CONCLUSION

We have presented new schemes to sample signals with finite rate of innovation. We have shown that it is possible to sample and perfectly reconstruct many FRI signals using a wide range of sampling kernels. The classes of kernels that can be used include functions satisfying Strang–Fix conditions and, therefore, scaling functions for wavelet bases, E-splines, and functions with rational Fourier transform. This last class of kernels is of particular interest in engineering since many acquisition devices used in communications, control, and, also, A/D conversion, can be modeled in this way. Thus, these new sampling schemes may have an impact in these engineering areas in the future.

Another important feature of our sampling scheme is that the reconstruction algorithm is local. This makes this technique more resilient to noise. We have, in fact, shown that signals with small local rate of innovation can be well reconstructed

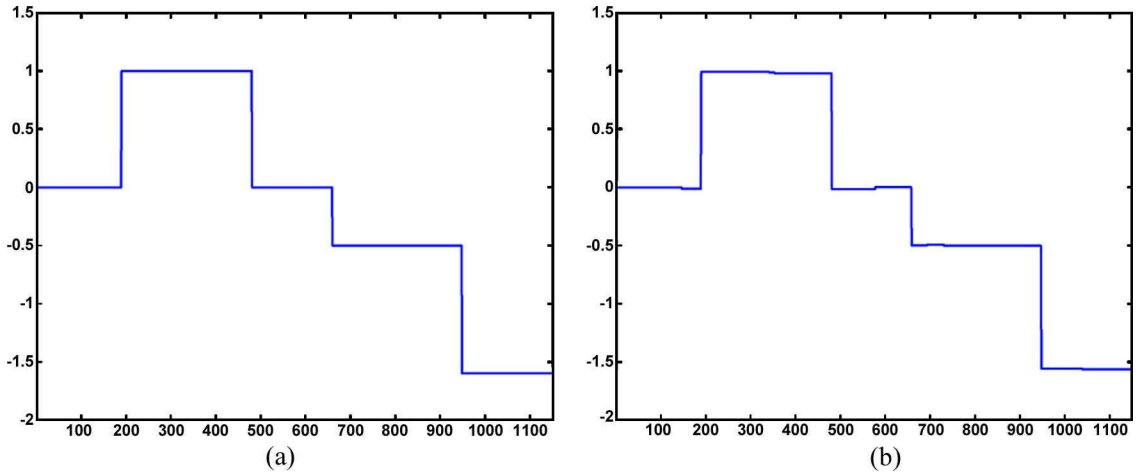


Fig. 11. Reconstruction of a piecewise constant signal. (a) The original signal. (b) Reconstructed signal (SNR = 27 dB).

also in the presence of noise. In this context, we have also presented a novel algorithm that reduces the effect of the noise by oversampling.

Extensions to the multidimensional case, as well as to the case of piecewise sinusoidal signals are presented in [2] and [21], respectively, and may have applications in image super-resolution and spread-spectrum communication.

## ACKNOWLEDGMENT

The authors would like to thank the two reviewers for their comments and M. Unser for stimulating discussions.

## REFERENCES

[1] N. I. Akhiezer and M. G. Krein, *Some Questions in the Theory of Moments*. Providence, RI: AMS, 1990, vol. 2.
[2] J. Berent and P. L. Dragotti, "Perfect reconstruction schemes for sampling piecewise sinusoisal signals," presented at the IEEE Int. Conf. Acoustic, Speech, Signal Processing, Toulouse, France, May 2006.
[3] R. E. Blahut, *Theory and Practice of Error Control Codes*. Reading, MA: Addison-Wesley, 1983.
[4] T. Blu, P. Thevenaz, and M. Unser, "MOMS maximal-order interpolation of minimal support," *IEEE Trans. Image Process.*, vol. 10, no. 7, pp. 1069–1080, Jul. 2001.
[5] ——, "Complete parameterization of piecewise-polynomial interpolation kernels," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1297–1309, Nov. 2003.
[6] T. Blu and M. Unser, "Approximation errors for quasiinterpolators and (multi-) wavelet expansions," *Appl. Comput. Harmon. Anal.*, vol. 6, no. 2, pp. 219–251, Mar. 1999.
[7] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principle: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
[8] E. W. Cheney and W. A. Light, "Quasiinterpolation with basis functions having noncompact support," *Constr. Approx.*, vol. 8, pp. 35–48, 1992.
[9] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.
[10] C. de Boor, R. A. Devore, and A. Ron, "Approximation from shift-invariant subspaces of $L_2(\mathbb{R})$," *Trans. Amer. Math. Soc.*, vol. 341, no. 2, pp. 787–806, Feb. 1994.
[11] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[12] D. L. Donoho, M. Vetterli, R. A. DeVore, and I. Daubechies, "Data compression and harmonic analysis," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2435–2476, Oct. 1998.
[13] M. Elad, P. Milanfar, and G. H. Golub, "Shapes from moments—An estimation theory perspective," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1814–1829, Jul. 2004.
[14] P. J. S. G. Ferreira and J. M. N. Vieira, "Locating and correcting errors in images," presented at the IEEE Int. Conf. Image Processing, Santa Barbara, CA, Oct. 1997.
[15] B. Gustafsson, C. He, P. Milanfar, and M. Putinar, "Reconstructing planar domains from their moments," *Inv. Probl.*, vol. 16, no. 54, pp. 1053–1070, Aug. 2000.
[16] Y. Lu and M. N. Do, "A geometrical approach to sampling signals with finite rate of innovation," presented at the IEEE Int. Conf. Acoustics, Speech, Signal Processing, Montréal, Canada, May 2004.
[17] S. Mallat, *A Wavelet Tour of Signal Processing*. New York: Academic, 1998.
[18] I. Maravic and M. Vetterli, "Exact sampling results for some classes of parametric nonbandlimited 2-D signals," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 175–189, Jan. 2004.
[19] P. Milanfar, G. C. Verghese, W. C. Karl, and A. S. Willsky, "Reconstructing polygons from moments with connections to array processing," *IEEE Trans. Signal Process.*, vol. 43, no. 2, pp. 432–443, Feb. 1995.
[20] A. Ron, "Factorization theorems for univariate splines on regular grids," *Israel J. Math.*, vol. 70, no. 1, pp. 48–68, 1990.
[21] P. Shukla and P. L. Dragotti, "Sampling schemes for 2-D signals with finite rate of innovation using kernels that reproduce polynomials," presented at the Proc. IEEE Int. Conf. Image Processing, Genova, Italy, Sep. 2005.
[22] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 2000.
[23] P. Stoica, T. Soderstrom, and F. Ti, "Asymptotic properties of the high-order Yule–Walker estimates of sinusoidal frequencies," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 11, pp. 1721–1734, Nov. 1989.
[24] G. Strang and G. Fix, "A Fourier analysis of the finite element variational method," in *Constructive Aspect of Functional Analysis*. Rome, Italy: Edizioni Cremonese, 1971, pp. 796–830.
[25] G. Strang and T. Nguyen, *Wavelets and Filter Banks*. Boston, MA: Wellesley-Cambridge, 1996.
[26] M. Unser, "Splines: A perfect fit for signal and image processing," *IEEE Signal Process. Mag.*, vol. 16, no. 6, pp. 22–38, Nov. 1999.
[27] ——, "Sampling—50 years after Shannon," *Proc. IEEE*, vol. 88, no. 4, pp. 569–587, Apr. 2000.
[28] ——, "Cardinal exponential splines: Part II—Think analogue, act digital," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1439–1449, Apr. 2005.
[29] M. Unser and A. Aldroubi, "A general sampling theory for nonideal acquisition devices," *IEEE Trans. Signal Process.*, vol. 42, no. 11, pp. 2915–2925, Nov. 1994.
[30] M. Unser and T. Blu, "Wavelet theory demystified," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 470–483, Feb. 2003.
[31] ——, "Cardinal exponential splines: Part I—Theory and filtering algorithms," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1425–1438, Apr. 2005.

[32] M. Vetterli, "Wavelets, approximation and compression," *IEEE Signal Process. Mag.*, vol. 18, no. 9, pp. 59–73, Sep. 2001.
[33] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*. Englewood Cliffs, NJ: Prentice-Hall, 1995.
[34] 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.
[35] J. M. N. Vieira and P. J. S. G. Ferreira, "Interpolation, spectrum analysis, error-control coding and fault-tolerant computing," presented at the IEEE Int. Conf. Acoustics, Speech, Signal Processing, Munich, Germany, Apr. 1997.

**Pier Luigi Dragotti** (M'02) received the Laurea degree (*summa cum laude*) in electrical engineering from the University Federico II, Naples, Italy, and the M.S. degree in communications systems and the Ph.D. degree from the Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, Switzerland, in 1997, 1998, and 2002, respectively.

He is currently a Lecturer in the Electrical and Electronic Engineering Department, Imperial College, London, U.K. He was a visiting student at Stanford University, Stanford, CA, in 1996, and, from July to October 2000, he was a Summer Researcher in the Mathematics of Communications Department at Bell Labs, Lucent Technologies, Murray Hill, NJ. Before joining Imperial College in November 2002, he was a Senior Researcher within the Swiss National Competence Center in Research on Mobile Information and Communication Systems. His research interests include wavelet theory, approximation and sampling theory, multirate signal processing, and distributed image/video processing and compression.

**Martin Vetterli** (F'95) received the Dipl. El.-Ing. degree from ETH Zürich (ETHZ), Switzerland, in 1981, the M.S. degree from Stanford University, Stanford, CA, in 1982, and the Doctorat ès Sciences degree from EPF Lausanne (EPFL), 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, 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 at Berkeley, where he was a Professor in 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 EPF Lausanne where he chaired the Communications Systems Division (1996/1997), and heads the Audiovisual Communications Laboratory. Since 2001, he has directed the National Competence Center in Research on Mobile Information and Communication Systems. He is also a Vice-President (International Affairs) at EPFL. He has held visiting positions at ETHZ (1990) and Stanford (1998). He is the co-author (with J. Kovačević) 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.

He is a member of SIAM and was the Area Editor for Speech, Image, Video, and Signal Processing of the IEEE TRANSACTIONS ON COMMUNICATIONS. He is also on the editorial boards of *Annals of Telecommunications, Applied and Computational Harmonic Analysis*, and *The Journal of Fourier Analysis and Application*. 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, 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 was a member of the Swiss Council on Science and Technology until December 2003.

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

He is with the Biomedical Imaging Group, Swiss Federal Institute of Technology, Lausanne, Switzerland, on leave from the France Télécom National Center for Telecommunications Studies (CNET), Issy-les-Moulineaux, France. His research interests include (multi)wavelets, multiresolution analysis, multirate filterbanks, approximation and sampling theory, psychoacoustics, optics, and wave propagation.

Dr. Blu is the recipient of the 2003 Best Paper Award (SP Theory and Methods) from the IEEE Signal Processing Society. He is currently an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING.