Sampling Continuous-Time Sparse Signals: A Frequency-Domain Perspective

Benjamín Béjar Haro and Martin Vetterli, *Fellow, IEEE*

***Abstract*—We address the problem of sampling and reconstruction of sparse signals with finite rate of innovation. We derive general conditions under which perfect reconstruction is possible for sampling kernels satisfying Strang-Fix conditions. Previous results on the subject consider two particular cases; when the kernel is able to reproduce (complex) exponentials, or when it has the polynomial reproduction property. In this paper, we extend such analysis to the case where both properties could be found in the sampling kernel and show that the former two situations can be regarded as special cases. As a result of our analysis, we provide general conditions under which perfect recovery in the noiseless case is possible. In practice, a given sampling kernel might not satisfy Strang-Fix conditions. When dealing with arbitrary sampling kernels, we propose a unified view for sampling and reconstruction in the frequency domain. Our formulation generalizes previous approaches and provides new insights for devising optimal reconstruction schemes. We also propose a novel algorithm for denoising treating the problem as a particular instance of structured low-rank approximation. Finally, we provide some numerical experiments and a comparison between different state-of-the-art methods showing the improved estimation performance of the proposed approach.**

***Index Terms*—Sampling methods, sparsity, spectral estimation.**

I. Introduction

SAMPLING and reconstruction of analog signals is at the heart of signal processing and communications. The work of Claude E. Shannon [1] paved the way for the success of digital communications by formalizing the sampling theorem for bandlimited signals. Later, a geometrical viewpoint of sampling and reconstruction allowed the generalization of the sampling theorem to more general shift-invariant subspaces [2]. Going back and forth from the continuous (analog) to the discrete (digital) domain requires that the signals being sampled can be predicted provided the sampling rate is above some (finite) critical value (i.e. signals with *Finite Rate of Innovation* (FRI) [3]). This is true for signals living in shift-invariant subspaces such as bandlimited signals, but it is also true for a broader class of signals that admit a parametric representation. A sampling theorem for a class of sparse FRI signals was proposed in [3]. More concretely, the works in [3], [4] focused on sampling and reconstruction of signals such as streams of (differentiated) Diracs, piecewise polynomials or non-uniform splines. Later extensions of the theory also included piecewise sinusoidal signals [5]. Perfect reconstruction from lowpass filtered observations of such signals can be achieved for some classes of sampling kernels such as ideal lowpass filters and gaussian kernels [3], or kernels that satisfy Strang-Fix conditions [4], [6]. Kernels such as B-Splines [7], E-Splines [8], Sum of Sincs (SoS) [9], or E-MOMS [10] fall within this category. Robust approaches to handle noise have also been developed in [11]–[13] as well as extensions to higher dimensions [14]–[16].

We start by considering the situation where the sampling kernel satisfies Strang-Fix conditions. Under this setup, two situations have been separately explored in the literature; the case where the kernel is able to reproduce (complex) exponentials, and the case where it has the polynomial reproduction property [3], [4]. However, these situations only cover two subsets of a more general setup, namely the case where both properties can be found in the sampling kernel. We analyze such situations and show that they generalize the previous two cases. As a result of our analysis, we first provide general conditions for perfect recovery in the noiseless case.

Since perfect recovery puts restrictions in the choice of the sampling kernel (e.g. Strang-Fix conditions) it is of interest to consider the general case of arbitrary sampling kernels. An extension of the FRI theory to work with arbitrary kernels was proposed in [10] in what it has been called *approximate FRI*. The approximate FRI framework has been developed exploiting the property of arbitrary kernels to (approximately) reproduce (complex) exponentials. The method applies not only to those kernels for which perfect reconstruction is not possible, but also to kernels such as B-Splines for which perfect reconstruction in the noiseless case is possible (e.g. using time-domain moments of the signal) but that suffer from ill-conditioning when noise is present. The authors in [10] start their derivation considering the reproduction of general exponentials to end up concluding that the best strategy is to consider purely imaginary exponents that are as evenly spaced on the unit circle as possible. In other words, it is best to consider frequency-domain information as in the ideal case of perfect reproduction of complex exponentials [3], [4].

Such an observation motivates us to propose a common sampling and reconstruction framework that suits both the ideal

Manuscript received July 11, 2017; revised December 10, 2017; accepted December 19, 2017. Date of publication January 11, 2018; date of current version February 1, 2018. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ian Clarkson. *(Corresponding author: Benjamín Béjar Haro.)*

The authors are with the I&C Audiovisual Communications Laboratory (LCAV), École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland (e-mail: benjamin.bejarharo@epfl.ch; martin.vetterli@epfl.ch).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org. This includes MATLAB source code for reproducing figures and results in the paper. The material is 57.8 KB in size.

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.2018.2791973

and approximate cases. The key idea is precisely to consider the frequency-domain information about the signal as the quantity of interest. We treat both the locations and amplitudes of the stream of Diracs as random quantities in our analysis and show that this point of view simplifies the derivations and provides a more intuitive explanation of the entire reconstruction procedure. Furthermore, it also shows that the approximate FRI framework of [10] is suboptimal in the minimum Mean Squared Error (MSE) sense. This is so since the latter one focuses on the problem of approximating deterministic signals (e.g. complex exponentials) while not taking into account the effect of noise. In order to make our approach more robust to noise, we also address the problem of denoising in the context of sparse FRI signals. As already noted in [13], the denoising problem can be seen as a particular instance of the Structured Low-Rank Approximation (SLRA) problem. In [13] an iterative method is proposed for denoising sparse FRI signals assuming that the sampling kernel is an ideal lowpass filter. However, the method is restricted to smooth (differentiable) loss functions and does not consider the more general setup of arbitrary sampling kernels. Here we provide a more general formulation of the problem and a solution that is more flexible in the choice of the loss function. In particular, we consider in detail the use of weighted ℓ_2 norms. In order to solve for the structured low-rank approximation problem we propose to use the Alternating Direction Method of Multipliers (ADMM) optimization framework [17] that leads to iterative procedures with closed-form (or easy to compute) updates. In the last part of the paper we provide numerical experiments and compare our estimation pipeline with current state-of-the-art methods showing its improved estimation performance.

In short, our main contributions are:

- A generalization of the conditions for perfect reconstruction in the noiseless case when the sampling kernel satisfies Strang-Fix conditions.
- A frequency-domain estimation framework for arbitrary sampling kernels.
- A novel denoising algorithm using the ADMM optimization framework that outperforms the state-of-the-art.

II. On Spectral Estimation

The problem of sampling and reconstruction of sparse FRI signals is closely related to that of spectral line estimation [3], [18]. In fact, under some conditions on the sampling kernel, these two problems are equivalent [3], [4]. This equivalence comes as no surprise since spikes are Fourier duals of complex exponentials. Consider then the model problem where we observe a sequence of the form:

$$X_k = \sum_{i=0}^{K-1} a_i\, u_i^k, \qquad k \in \mathcal{K}, \tag{1}$$

where $\mathcal{K} \subseteq \mathbb{Z}$ is a sequence of consecutive elements, and where $a_i, u_i \in \mathbb{C}$ are some complex numbers with $a_i \neq 0$ and $u_i \neq u_j$ for $i \neq j$. Note that for $u_i = e^{j\omega_i}$ the problem is that of identifying a set of complex exponentials and there is a very rich literature about the topic [18]. High-resolution line-spectral estimation methods such as ESPRIT [19], [20] that exploit the rotation invariance property of Vandermonde matrices, or methods based on the eigen-decomposition of the data covariance matrix such as Pisarenko's method [21] and MUSIC [22]–[24] as well as quadratic approximations to the maximum likelihood estimation problem (IQML) [25], [26] are among the most popular methods for this task. We shall briefly review here Prony's method (*a.k.a. annihilating filter method*) as basic algorithm [27], [28].

A. Annihilating Filter Method

The algebraic structure of the sequence X_k allows for a decoupled estimation of the unknown variables u_i and a_i, provided enough measurements are available. In order to see this, consider a filter with roots at the u_i values:

$$H(z) = \sum_{k=0}^{K} h_k\, z^{-k} = \prod_{i=0}^{K-1} \left(1 - u_i z^{-1}\right). \tag{2}$$

The above filter $H(z)$ annihilates the sequence X_k since:

$$X_k * h_k = 0. \tag{3}$$

Let $\boldsymbol{x} = [X_0, \ldots, X_L]^T$ with $L \geq 2K-1$, and rewrite (3) as

$$\underbrace{\begin{bmatrix} X_K & X_{K-1} & \cdots & X_0 \\ X_{K+1} & X_K & \cdots & X_1 \\ \vdots & \vdots & \vdots & \vdots \\ X_L & X_{L-1} & \cdots & X_{L-K} \end{bmatrix}}_{\mathrm{T}_K(\boldsymbol{x}) = \boldsymbol{X} = [-\boldsymbol{b}\ \tilde{\boldsymbol{B}}]} \boldsymbol{h} = \boldsymbol{0}, \tag{4}$$

where $\boldsymbol{h} = [h_0, h_1, \ldots, h_K]^T$ and $\mathrm{T}_K(\cdot)$ is a linear operator that maps a sequence to a Toeplitz matrix. Equation (4) implies that $\boldsymbol{X}$ is rank deficient. In fact, it has rank K provided all u_i values are distinct. In such case, $\boldsymbol{X}$ has a non-trivial one-dimensional nullspace and the annihilating filter $\boldsymbol{h}$ can be uniquely determined up to a scale factor. Since scaling does not affect the position of the roots, we can simply fix $h_0 = 1$ which results in the following system of equations:

$$\tilde{\boldsymbol{B}}\, \boldsymbol{h}_{1:K} = \boldsymbol{b}, \tag{5}$$

where $\boldsymbol{h}_{1:K} = [h_1, \ldots, h_K]^T$. Note that, for $L \geq 2K-1$ the above system of equations can be solved exactly in the noiseless case, and in the least-squares sense in the presence of noise. Alternatively, we could use a normalization of $\|\boldsymbol{h}\| = 1$. In that case, the solution to (4) can be obtained from the Singular Value Decomposition (SVD) of $\boldsymbol{X}$, and it would correspond to the left-singular vector associated to the smallest singular value of $\boldsymbol{X}$. The two solutions are equivalent provided $h_0 \neq 0$. In the noisy case, the SVD-based solution corresponds to a Total Least Squares approach to solve (5).

Once we have the u_i values, the retrieval of the amplitudes boils down to solving the following linear system of equations:

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ u_0 & u_1 & \cdots & u_{K-1} \\ \vdots & \vdots & & \vdots \\ u_0^L & u_1^L & \cdots & u_{K-1}^L \end{bmatrix} \boldsymbol{a} = \boldsymbol{V}_L(\boldsymbol{u})\, \boldsymbol{a} = \boldsymbol{x}, \tag{6}$$

where $\boldsymbol{a} = [a_0, \ldots, a_{K-1}]^T$ is the vector of amplitudes and $\boldsymbol{u} = [u_0, u_1, \ldots, u_{K-1}]^T$. Since the u_i values are known and distinct, the Vandermonde structure of $\boldsymbol{V}_L(\boldsymbol{u})$ ensures that the system of equations in (6) has a unique solution provided $L \geq K-1$. In the presence of noise, (6) can be solved in the least-squares sense.

III. Sampling Sparse Signals

In this section we describe the class of sparse FRI signals of interest and the considered sampling framework (see Fig. 1). In particular, we will consider both finite and infinite (periodic) streams of Dirac impulses as an abstraction of some real-world phenomenon. We will be analyzing both cases separately showing the relationship between them. We will start first by highlighting the relationship between our sampling problem and the model problem in (1).

A. Sparse Signals and Fourier Series

Let $x(t)$ be a finite stream of K weighted Diracs:

$$x(t) = \sum_{i=0}^{K-1} a_i\, \delta(t - t_i), \tag{7}$$

where $t_i \in [0, 1)$ and $a_i \in \mathbb{C}$. Sparse signals of the form of (7) are common in different domains such as biology, communications, or astronomy, to name a few. Note that a complete characterization of the signal $x(t)$ requires the knowledge of $2K$ parameters: those corresponding to the amplitudes a_i and locations t_i of the spikes. For the case of a periodic stream of Diracs we will assume that $x(t)$ corresponds to its fundamental period. In any case, regardless whether the signal is periodic or not we can always represent it by its *Fourier Series* (FS) expansion on some interval $[0, \tau)$, $\tau \geq 1$. The (scaled) FS coefficients of $x(t)$ are thus given by

$$X_k(\tau) = \left\langle x(t), e^{j2\pi kt/\tau} \right\rangle = \sum_{i=0}^{K-1} a_i\, e^{-j2\pi k t_i/\tau},\ k \in \mathbb{Z}. \tag{8}$$

Note that if we let $u_i = e^{-j2\pi t_i/\tau}$, then (8) has the form of our model problem (1) and therefore, a subsequence of $2K$ FS coefficients suffices to fully characterize $x(t)$ (e.g. by using the annihilating filter method [3]).

In all practical situations, we won't be able to directly observe the signal of interest. Instead, the acquisition process is usually modeled as filtering followed by sampling. In that case, the desired signal goes through some (typically lowpass) filter and then, the resulting output is recorded at discrete-time locations as illustrated in Fig. 1. Our goal is then to retrieve the parameters of the signal from the available set of observations.

B. Periodic Stream of Diracs

Consider the input signal $z(t) = \sum_{m\in\mathbb{Z}} x(t - m\tau)$ to be the periodization of $x(t)$ and, without loss of generality, let the period be $\tau = 1$. Let $f(t)$ denote the output of the lowpass filtering operation (see Fig. 1) with the scaled filter $\varphi^*(-t/T)$:

$$f(t) = \sum_{m\in\mathbb{Z}} \sum_{i=0}^{K-1} a_i\, \varphi^*\Big(\frac{m - (t - t_i)}{T}\Big) \tag{9}$$

$$= T \sum_{k\in\mathbb{Z}} X_k\, \hat{\varphi}^*\big(2\pi kT\big) e^{j2\pi kt}, \tag{10}$$

where $\hat{\varphi}(\omega)$ is the Fourier transform of $\varphi(t)$, and where we have identified X_k as the FS coefficients of the periodic stream of Diracs $z(t)$ whose fundamental period corresponds to $x(t)$. The second equality in (10) follows from the application of the Poisson[1] sum formula together with Fourier transform properties for shift and scaling. Finally, the samples are thus

$$y_n = f(nT) = T \sum_{k\in\mathbb{Z}} X_k\, \hat{\varphi}^*\big(2\pi kT\big)\, e^{j2\pi knT},\ n \in \mathbb{Z}. \tag{11}$$

Consider now the situation where we observe the signal over one period with a sampling rate of N samples per period (i.e. $T = 1/N$). In such case, the sequence of samples is

$$y_n = \frac{1}{N} \sum_{k\in\mathbb{Z}} X_k\, \hat{\varphi}^*\left(\frac{2\pi k}{N}\right) e^{j2\pi kn/N},\ n = 0, \ldots, N-1. \tag{12}$$

Since we are interested in retrieving the FS coefficients of the signal it is natural to work in the frequency domain. From the samples y_n we can compute the Discrete Fourier Transform (DFT) at $\omega_k = \frac{2\pi k}{N}$, $k = 0, \ldots, N-1$, to obtain

$$Y_k = \sum_{n=0}^{N-1} y_n\, e^{-j2\pi nk/N} \tag{13}$$

$$= \sum_{\ell\in\mathbb{Z}} X_\ell\, \hat{\varphi}^*\left(\frac{2\pi\ell}{N}\right) \frac{1}{N} \underbrace{\sum_{n=0}^{N-1} e^{j2\pi\ell n/N} e^{-j2\pi kn/N}}_{N\delta(k-\ell_{\bmod N})} \tag{14}$$

$$= X_k\, \hat{\varphi}^*\left(\frac{2\pi k}{N}\right) + \underbrace{\sum_{\substack{\ell\in\mathbb{Z}\\ \ell\neq 0}} X_{k+mN}\, \hat{\varphi}^*\left(\frac{2\pi k}{N} + 2\pi\ell\right)}_{\text{aliasing}}. \tag{15}$$

Aliasing and the choice of the sampling kernel: From (15) we can identify the contribution of two different terms: one corresponding to the desired FS coefficient X_k, and an aliasing term. The presence or not of aliasing will depend on the chosen sampling kernel as we illustrate in the example below.

Example—Ideal lowpass filter: Assume the sampling kernel is an ideal lowpass filter with cut-off frequency $1/2$ that is,

$$\varphi(t) = \frac{\sin(\pi t)}{\pi t} \overset{\mathcal{F}}{\longleftrightarrow} \hat{\varphi}(\omega) = \begin{cases} 1 & |\omega| \leq \pi \\ 0 & \text{otherwise} \end{cases}. \tag{16}$$

In such a case, it is easy to verify from (15) that the DFT of the measured samples will be free of aliasing, i.e. $Y_k = X_k$.

[1] Poisson sum formula: $\sum_{n\in\mathbb{Z}} f(n) = \sum_{k\in\mathbb{Z}} \hat{f}(2\pi k)$.

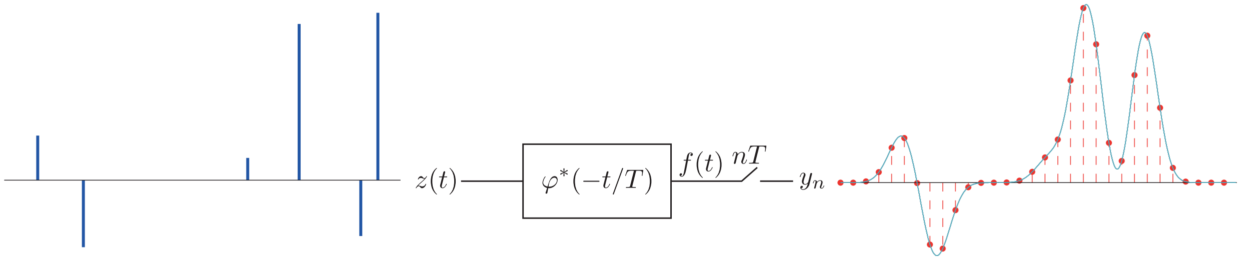


Fig. 1. Sampling scheme. The input sparse signal $z(t)$ passes through the filter with impulse response $\varphi^*(-t/T)$ and it is then sampled at intervals T seconds apart to produce the output sequence y_n.

Therefore, we have access to the FS coefficients X_k and hence, it is possible to retrieve the locations and amplitudes of the spikes using Prony's method provided $N \geq 2K$.

C. Finite Stream of Diracs and Kernels With Finite Support

Consider now the case where $z(t) = x(t)$ is a finite stream of Diracs with the additional assumption that the kernel $\varphi(t)$ has finite support. The output signal $f(t)$ is thus given by

$$f(t) = x(t) * \varphi^*(-t/T) = \sum_{i=0}^{K-1} a_i\, \delta(t - t_i) * \varphi^*(-t/T). \tag{17}$$

The discrete-time signal y_n is then:

$$y_n = f(nT) = \sum_{i=0}^{K-1} a_i\, \varphi^*(t_i/T - n), \qquad n \in \mathbb{Z}. \tag{18}$$

If we now compute the Discrete-Time Fourier Transform (DTFT) of the sequence y_n at frequency ω we get

$$Y(e^{j\omega}) = \sum_{n\in\mathbb{Z}} \sum_{i=0}^{K-1} a_i\, \varphi^*(t_i/T - n)\, e^{-j\omega n} \tag{19}$$

$$= \sum_{\ell\in\mathbb{Z}} \sum_{i=0}^{K-1} a_i\, e^{-j(\omega+2\pi\ell)t_i/T}\, \hat{\varphi}^*(\omega + 2\pi\ell), \tag{20}$$

where (20) follows from the Poisson summation formula.

A word on the equivalence of periodic and non-periodic cases: Since both the kernel and the input signal have finite support the sampled signal y_n is also supported on a finite interval. Without loss of generality, assume that the samples y_n are zero for values of $n < 0$ and $n \geq M_0$, where M_0 is some positive integer. In order to highlight the connection between the periodic and non-periodic cases consider frequencies of the form $\omega_k = 2\pi k/M$, where $M \geq M_0$. Particularizing (20) for the case of $T = 1/N$ we end up with

$$Y(e^{j\omega_k}) = \sum_{\ell\in\mathbb{Z}} X_{k+\ell M} \left(\frac{M}{N}\right) \hat{\varphi}^* \left(\frac{2\pi k}{M} + 2\pi\ell\right). \tag{21}$$

Note that (21) is analogous to the expression in (15) for the periodic case. The only difference is that now the signal has an *equivalent* period of $\tau = M/N$ rather than 1. Therefore, we can argue that for frequencies lying on a uniform grid (i.e. take the M-point DFT of y_n, $M \geq M_0$) sampling a periodic stream of Diracs or sampling a finite stream of Diracs with a sampling kernel with finite support are two equivalent situations.

IV. Sampling With Strang-Fix Kernels

Perfect reconstruction of the original stream of Diracs is possible for particular choices of the sampling kernel [3], [4], [9], [10]. In a general setting, sampling and reconstruction of streams of Diracs fits the framework of sampling in a union of subspaces [29]. Particular choices of kernels that allow for perfect recovery are kernels such as E-Splines, SoS, or E-MOMS, that satisfy Strang-Fix conditions [6]. We would like to highlight the relationship between those kernels and the frequency-domain information about the signal of interest.

A. Strang-Fix Kernels

Consider kernels of finite support that reproduce (complex) exponential polynomials of degree P (see Fig. 2), that is

$$\sum_{n\in\mathbb{Z}} c_n(\omega_k, P)\varphi(t - n) = t^P\, e^{j\omega_k t}, \tag{22}$$

for some frequency ω_k and appropriate choice of the sequence of coefficients $c_n(\omega_k, P) \in \mathbb{C}$. Equation (22) holds if $\varphi(t)$ satisfies the so-called Strang-Fix conditions [6], [30]:

$$\hat{\varphi}(\omega_k) \neq 0 \qquad \text{and} \qquad \frac{d^p\hat{\varphi}}{d\omega^p}(\omega_k + 2\pi\ell) = 0, \tag{23}$$

for $p = 0, \ldots, P$ and $\ell \in \mathbb{Z} - \{0\}$.

Definition 1: ($(\mathcal{W}, \mathcal{P})$ Strang-Fix conditions) Let $\mathcal{W} = \{\omega_0, \ldots, \omega_L\}$ denote a set of frequencies with $\omega_i \neq \omega_j + 2\pi\ell$ for all $i, j = 0, \ldots, L$ and $\ell \in \mathbb{Z} - \{0\}$. And let also denote $\mathcal{P} = \{P_0, \ldots, P_L\}$ as the corresponding set of orders. We say that a kernel satisfies the $(\mathcal{W}, \mathcal{P})$ Strang-Fix conditions if (23) holds for all (ω_k, P_k) pairs, $k = 0, \ldots, L$.

For the case where the order of polynomial reproduction is the same for all frequencies (i.e. $P_k = P$) we will simply say that the kernel satisfies the $(\mathcal{W}, P)$ Strang-Fix conditions.

B. Sampling Sparse Signals With Strang-Fix Kernels

When it comes to sampling sparse signals two particular cases of interest that have appeared in the FRI literature namely, the $(\mathcal{W}, 0)$ (complex exponential reproduction) and the $(0, P)$ (polynomial reproduction) cases. In this section we will start by reviewing the former two situations but will also consider the

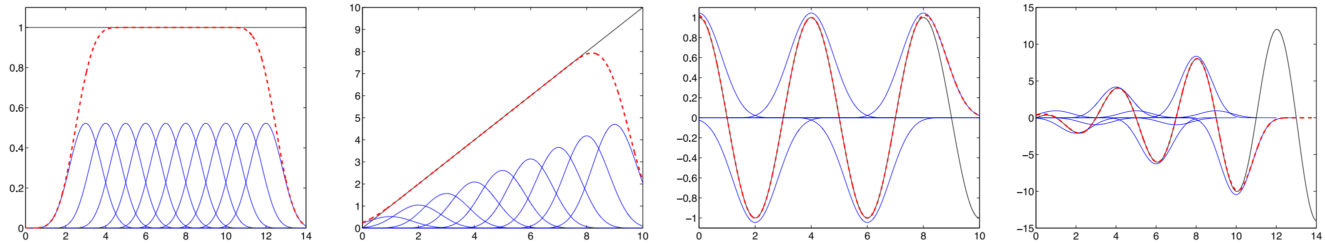


Fig. 2. Exponential polynomial reproduction property of a kernel that satisfies the $\left(\mathcal{W} = \{-\pi/2, 0, \pi/2\}, P = 1\right)$ Strang-Fix conditions. From left to right, reproduction of: a constant, t, $\cos(\pi t/2)$, $t\cos(\pi t/2)$.

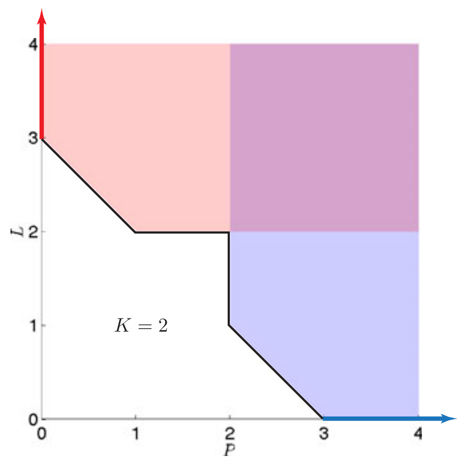


Fig. 3. Regions in the L-P plane that allow for perfect recovery for the case of $K = 2$ Diracs. The blue region corresponds to the case where Theorem 1 applies. Likewise, the red region corresponds to Theorem 2. The red and blue arrows on the L and P axes represent respectively, the $(\mathcal{W}, 0)$ complex exponential reproduction and the $(0, P)$ polynomial reproduction cases.

more general $(\mathcal{W}, \mathcal{P})$ case (see Fig. 3). We will provide conditions under which perfect reconstruction of a stream of Diracs from the set of samples is possible. In our analysis we will consider the non-periodic case but the results presented naturally extend to the periodic case, too.

Exponential Reproduction: A kernel that reproduces complex exponentials satisfies the $(\mathcal{W}, 0)$ Strang-Fix conditions:

$$\hat{\varphi}(\omega_k) \neq 0 \qquad \text{and} \qquad \hat{\varphi}(\omega_k + 2\pi\ell) = 0, \tag{24}$$

for all $\omega_k \in \mathcal{W}$ and for $\ell \in \mathbb{Z} - \{0\}$. An equivalent interpretation of the Strang-Fix conditions in (24) is that the DTFT of the sampled kernel is *free of aliasing* at the set of frequencies $\mathcal{W}$. Now, consider the case where the set of frequencies $\mathcal{W}$ follow a uniform progression, e.g. $\mathcal{W} = \{\frac{2\pi}{M}k\}_{k=0}^{L}$. If we look back to (21) and because (24) holds, we immediately realize that we have access to the continuous-time frequency information of the original signal since $Y_k = X_k\, \hat{\varphi}^*(\omega_k)$. The effect of the sampling kernel can then be compensated to give back the FS coefficients of $x(t)$. Having access to a set of $2K$ consecutive FS coefficients (i.e. $L = 2K - 1$) allows the estimation of the parameters of the signal using Prony's method. This is precisely the idea behind E-Splines, SoS, and E-MOMS kernels [3], [4], [9], [10].

Polynomial Reproduction: A kernel that reproduces polynomials of degree P satisfies the $(\{0\}, P)$ Strang-Fix conditions:

$$\hat{\varphi}(0) \neq 0 \qquad \text{and} \qquad \frac{d^p\hat{\varphi}}{d\omega^p}(2\pi\ell) = 0, \tag{25}$$

for $p = 0, \ldots, P$ and for $\ell \in \mathbb{Z} - \{0\}$. This property allows the computation of the temporal moments of the continuous signal from the samples [4]. For the case of streams of Diracs, the sequence of temporal moments has the form of (1) and hence, perfect reconstruction of the signal's parameters is possible using Prony's method provided at least $2K$ moments (i.e. $P = 2K - 1$) are available [4]. In the frequency domain we can think of this situation as having access to P *alias-free* measurements of the signal and its derivatives up to order P at $\omega = 0$.

Exponential Polynomial Reproduction: We have just described two situations where it is possible to retrieve the stream of Diracs: one where we have multiple frequencies but no polynomial reproduction property, and another one where we have a single frequency (e.g. $\omega = 0$) but can reproduce polynomials of multiple orders. It becomes intuitive that a combination of these two situations could allow the definition of more general conditions under which perfect signal reconstruction is possible. To see this, let us continue our discussion from the polynomial reproduction case. Note that, in principle, we could have used any other frequency different from $\omega = 0$ and the same principle would apply to the modulated signal. Assume then that the sampling kernel $\varphi(t)$ satisfies the (ω_0, P) Strang-Fix conditions for some (not necessarily zero) frequency ω_0. Since the kernel reproduces exponential polynomials we can compute the *modulated moments* of the signal $x(t)$ from the samples y_n as follows:

$$\tau_p(\omega_0) = \sum_{n\in\mathbb{Z}} c_n^*(\omega_0, P)\, \underbrace{\langle x(t), \varphi(t/T - n)\rangle}_{y_n} \tag{26}$$

$$= \left\langle x(t), \sum_{n\in\mathbb{Z}} c_n(\omega_0, P)\, \varphi(t/T - n)\right\rangle \tag{27}$$

$$= \int_{-\infty}^{\infty} x(t) \left(\frac{t}{T}\right)^p e^{-j\omega_0 t/T}\, dt \tag{28}$$

$$= \sum_{i=0}^{K-1} a_i \left(\frac{t_i}{T}\right)^p e^{-j\omega_0 t_i/T}. \tag{29}$$

We can now use the continuous moments to generalize the sampling results that have appeared in the literature. We consider two situations – one where estimation is done using the polynomial reproduction property and another where estimation exploits the reproduction of complex exponentials.

Assume that $\varphi(t)$ satisfies the $(\mathcal{W},\mathcal{P})$ Strang-Fix conditions and consider the sequences of moments $\tau_p(\omega_k)$, $p=0,\ldots,P_k$ for $k=0,\ldots,L$. The key observation is to realize that for every ω_k in the set $\mathcal{W}$, the same filter annihilates the sequence of moments $\tau_p(\omega_k)$, $p=0,\ldots,P_k$ for all k provided that $P_k \geq K$. This can be easily seen by rewriting $\tau_p(\omega_k)$ as

$$\tau_p(\omega_k)=\sum_{i=0}^{K-1} a_i\left(\frac{t_i}{T}\right)^p e^{-j\omega_k t_i/T}=\sum_{i=0}^{K-1}\tilde{a}_{ik}\left(\frac{t_i}{T}\right)^p, \quad (30)$$

which is of the same form as our model problem (1). In order to retrieve the signal parameters we just need to make sure that we can form a system of equations that uniquely specifies the annihilating filter up to a scale factor.

Theorem 1: (Temporal moments) Consider a non-periodic stream of K Diracs $x(t)$ as in (7) and the sampling setup of Fig. 1. Assume that the sampling kernel $\varphi(t)$ satisfies the $(\mathcal{W},\mathcal{P})$ Strang-Fix conditions with

$$\sum_{k=0}^{L} P_k + L(1-K)\geq 2K-1 \qquad \text{and} \qquad P_k\geq K, \quad (31)$$

then y_n provides a complete characterization of $x(t)$ if

$$\boldsymbol{B}=\begin{bmatrix}\boldsymbol{V}_{n_0}(\boldsymbol{t})\mathrm{diag}\,(\boldsymbol{\beta}_0)\\ \vdots\\ \boldsymbol{V}_{n_L}(\boldsymbol{t})\mathrm{diag}\,(\boldsymbol{\beta}_L)\end{bmatrix}, \quad (32)$$

has full column rank, where $\boldsymbol{t}=\frac{1}{T}[t_0,\ldots,t_{K-1}]^T$ denotes the vector of normalized locations, $n_k=P_k-K$, and $\boldsymbol{\beta}_k^T=[e^{-j\omega_k t_0/T},\ldots,e^{-j\omega_k t_{K-1}/T}]$.

Proof: Consider the sequence (vector) of moments $\boldsymbol{\tau}(\omega_k)=[\tau_0(\omega_k),\tau_1(\omega_k),\ldots,\tau_{P_k}(\omega_k)]^T$ for some frequency $\omega_k\in\mathcal{W}$. Form the Toeplitz matrix $\mathrm{T}_K(\boldsymbol{\tau}(\omega_k))$ and express it as:

$$\mathrm{T}_K(\boldsymbol{\tau}(\omega_k))=\boldsymbol{V}_{n_k}(\boldsymbol{t})\mathrm{diag}\,(\boldsymbol{\beta}_k)\,\boldsymbol{A}, \quad (33)$$

where $n_k=P_k-K$, $\boldsymbol{\beta}_k^T=[e^{-j\omega_k t_0/T},\ldots,e^{-j\omega_k t_{K-1}/T}]$ and where the matrix $\boldsymbol{A}$ is given by

$$\boldsymbol{A}=\mathrm{diag}\,(\boldsymbol{a})\begin{bmatrix}(t_0/T)^K & (t_0/T)^{K-1} & \cdots & 1\\ (t_1/T)^K & (t_1/T)^{K-1} & \cdots & 1\\ \vdots & \vdots & & \vdots\\ (t_{K-1}/T)^K & (t_{K-1}/T)^{K-1} & \cdots & 1\end{bmatrix}. \quad (34)$$

It is clear that a filter $\boldsymbol{h}$ whose roots correspond to the normalized locations of the spikes (i.e. t_i/T) annihilates the sequence of moments $\{\tau_p(\omega_k)\}_{p=0}^{P_k}$ provided that $P_k\geq K$ and therefore $\boldsymbol{A}\boldsymbol{h}=\boldsymbol{0}$. Since this is true for all frequencies in the set $\mathcal{W}$ we can form the following system of equations:

$$\begin{bmatrix}\mathrm{T}_K(\boldsymbol{\tau}(\omega_0))\\ \vdots\\ \mathrm{T}_K(\boldsymbol{\tau}(\omega_L))\end{bmatrix}\boldsymbol{h}=\begin{bmatrix}\boldsymbol{V}_{n_0}(\boldsymbol{t})\mathrm{diag}\,(\boldsymbol{\beta}_0)\\ \vdots\\ \boldsymbol{V}_{n_L}(\boldsymbol{t})\mathrm{diag}\,(\boldsymbol{\beta}_L)\end{bmatrix}\boldsymbol{A}\boldsymbol{h}=\boldsymbol{B}\boldsymbol{A}\boldsymbol{h}=\boldsymbol{0}, \quad (35)$$

where $\boldsymbol{B}$ is defined as in (32). The filter $\boldsymbol{h}$ in (35) is uniquely determined (up to scale) provided the product $\boldsymbol{BA}$ has full column rank. Matrix $\boldsymbol{A}$ is a $K\times(K+1)$ matrix and it is of full rank since it is the product of a full rank diagonal matrix (since $a_i\neq 0$) and a $K\times(K+1)$ Vandermonde matrix also of full row rank. From Sylvester's rank inequality we have

$$\mathrm{rank}\,(\boldsymbol{BA})\geq\mathrm{rank}\,(\boldsymbol{B})+\mathrm{rank}\,(\boldsymbol{A})-K=\mathrm{rank}\,(\boldsymbol{B}), \quad (36)$$

where the last equality follows from the fact that $\mathrm{rank}\,(\boldsymbol{A})=K$. On the other hand, it is clear that $\mathrm{rank}\,(\boldsymbol{BA})\leq K$. Therefore, if $\boldsymbol{B}$ has full column rank, then $\mathrm{rank}\,(\boldsymbol{BA})=K$ and the vector $\boldsymbol{h}$ is uniquely specified. But in order for $\boldsymbol{B}$ to have full column rank it must have at least the same number of rows than columns (i.e. K). From its definition, $\boldsymbol{B}$ has dimensions $\sum_k(n_k+1)\times K$ and hence it is required that

$$\sum_{k=0}^{L}(n_k+1)=\sum_{k=0}^{L}(P_k-K+1)\geq K. \quad (37)$$

Rearranging terms we require $\sum_k P_k+L(1-K)\geq 2K-1$ which is true by assumption and the proof is complete. ■

From Theorem 1 we see that perfect recovery in the general case depends on the parameters to be estimated. Nevertheless, there exist specific configurations for which perfect reconstruction is always possible regardless of the parameters of the signal. A simple way to ensure perfect recovery is to have $P_k\geq 2K-1$ for some k. In such case, perfect reconstruction is always possible [4]. We illustrate this result as a corollary of Theorem 1.

Corollary 1: (Polynomial reproduction) Consider a non-periodic stream of K Diracs $x(t)$ as in (7) and the sampling setup of Fig. 1. Assume that $\varphi(t)$ satisfies the (ω_0,P) Strang-Fix conditions. Then for the case of $P\geq 2K-1$ perfect recovery is always possible.

Proof: Note that in that case, matrix $\boldsymbol{B}$ would be given by

$$\boldsymbol{B}=\boldsymbol{V}_{P-K}(\boldsymbol{t})\mathrm{diag}\,(\boldsymbol{\beta}_0), \quad (38)$$

which is of rank K (full rank) and, therefore perfect recovery is possible. The full rank property of matrix $\boldsymbol{B}$ follows from the fact that for $P\geq 2K-1$, matrix $\boldsymbol{V}_{P-K}$ is a tall Vandermonde matrix which is full rank and of rank K since the t_is are distinct. The diagonal matrix is also full rank since the elements of $\boldsymbol{\beta}_0$ are nonzero. Then, from Sylvester's rank inequality it can be shown that $\boldsymbol{B}$ is of full rank K. ■

Moreover, if we restrict the set of frequencies to lie on a uniform grid then perfect recovery is also guaranteed:

Corollary 2: Under the same assumptions as in Theorem 1 and if, in addition, the set of frequencies is of the form $\mathcal{W}=\{\alpha+\frac{2\pi k}{M}\}_{k=0}^{L}$, for some $\alpha\in\mathbb{R}$ and with $L\geq K-1$, then perfect recovery is always possible.

Proof: From (32) it is clear that

$$\tilde{\boldsymbol{B}}=\boldsymbol{V}_L(\boldsymbol{u})\mathrm{diag}\left([e^{-j\alpha t_0/T},\ldots,e^{-j\alpha t_{K-1}/T}]\right), \quad (39)$$

where $\boldsymbol{u}^T=[e^{-j\frac{2\pi}{M}t_0/T},\ldots,e^{-j\frac{2\pi}{M}t_{K-1}/T}]$, is a subset of the rows of $\boldsymbol{B}$ in (32). Matrix $\tilde{\boldsymbol{B}}$ (hence $\boldsymbol{B}$) has full column rank if $\boldsymbol{V}_L(\boldsymbol{u})$ has also full column rank which, due to the Vandermonde structure of $\boldsymbol{V}_L(\boldsymbol{u})$, is true for $L\geq K-1$. ■

Theorem 1 establishes conditions under which perfect reconstruction of a stream of Diracs is possible if we have access to the modulated (temporal) moments of the original signal at an arbitrary set of frequencies not necessarily lying on a grid. It also defines a region in the $L-P$ plane (blue region in Fig. 3) where it applies. On the other hand, if we restrict the set of frequencies to be on a uniform grid, then it becomes clear that a similar condition for perfect reconstruction using frequency-domain information is possible for the case of $L \geq K$ by just switching the roles between complex exponential moments (frequency-domain information) and temporal moments. The following result formalizes these ideas and the corresponding region in the $L-P$ plane is illustrated in red in Fig. 3.

Theorem 2: (Complex moments) Consider a non-periodic stream of K Diracs $x(t)$ as in (7) and the sampling setup of Fig. 1. Without loss of generality assume that $t_i \in [0,1)$, $i = 0,\ldots,K-1$. Assume that the sampling kernel $\varphi(t)$ satisfies the $(\mathcal{W},P)$ Strang-Fix conditions where the set of frequencies lie on a uniform progression of the form $\mathcal{W} = \{\alpha + \frac{2\pi k}{M}\}_{k=0}^{L}$ for some $\alpha \in \mathbb{R}$ and $\frac{L}{MT} < 1$. Assume also that

$$(P+1)(L-K+1) \geq K \qquad \text{and} \qquad L \geq K. \tag{40}$$

Then y_n provides a complete characterization of $x(t)$ if

$$\hat{\boldsymbol{B}} = \begin{bmatrix} \boldsymbol{V}_{L-K}(\boldsymbol{u}) \\ \boldsymbol{V}_{L-K}(\boldsymbol{u})\mathrm{diag}\,(\boldsymbol{t}) \\ \vdots \\ \boldsymbol{V}_{L-K}(\boldsymbol{u})\mathrm{diag}\,(\boldsymbol{t}^P) \end{bmatrix}, \tag{41}$$

has full column rank, where $\boldsymbol{u}^T = [u_0, u_1, \ldots, u_{K-1}]$ with $u_i = e^{-j\frac{2\pi}{M}t_i/T}$.

Proof: Without loss of generality, let us assume that $\alpha = 0$. Note that the effect of any α can be removed by considering the modulated sequence $y_n e^{-\alpha n}$. Consider now the sequence of moments (indexed by frequency) $\boldsymbol{\tau}_p = [\tau_p(\omega_0), \tau_p(\omega_1), \ldots, \tau_p(\omega_L)]^T$ for some $p \leq P$ and form the Toeplitz matrix $\mathrm{T}_K(\boldsymbol{\tau}_p)$.

Then it holds that:

$$\mathrm{T}_K(\boldsymbol{\tau}_p) = \boldsymbol{V}_{L-K}(\boldsymbol{u})\mathrm{diag}\,(\boldsymbol{t}^p)\,\hat{\boldsymbol{A}}, \tag{42}$$

where $\boldsymbol{u}^T = [u_0, u_1, \ldots, u_{K-1}]$ with $u_i = e^{-j\frac{2\pi}{M}t_i/T}$, and the matrix $\hat{\boldsymbol{A}}$ is given by

$$\hat{\boldsymbol{A}} = \mathrm{diag}\,(\boldsymbol{a}) \begin{bmatrix} u_0^K & u_0^{K-1} & \cdots & 1 \\ u_1^K & u_1^{K-1} & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ u_{K-1}^K & u_{K-1}^{K-1} & \cdots & 1 \end{bmatrix}. \tag{43}$$

Since $L \geq K$ and due to the fact that the frequencies lie on a uniform progression, there exist a filter with roots at the u_is that annihilates the sequence $\boldsymbol{\tau}_p$ (i.e. $\hat{\boldsymbol{A}}\boldsymbol{h} = 0$) and thus

$$\mathrm{T}_K\left(\{\tau_p(\omega_k)\}_{k=0}^{L}\right)\boldsymbol{h} = 0, \qquad p = 0,\ldots,P. \tag{44}$$

Therefore, the annihilation equation can be written as:

$$\begin{bmatrix} \mathrm{T}_K(\boldsymbol{\tau}_0) \\ \mathrm{T}_K(\boldsymbol{\tau}_1) \\ \vdots \\ \mathrm{T}_K(\boldsymbol{\tau}_P) \end{bmatrix} \boldsymbol{h} = \begin{bmatrix} \boldsymbol{V}_{L-K}(\boldsymbol{u}) \\ \boldsymbol{V}_{L-K}(\boldsymbol{u})\mathrm{diag}\,(\boldsymbol{t}) \\ \vdots \\ \boldsymbol{V}_{L-K}(\boldsymbol{u})\mathrm{diag}\,(\boldsymbol{t}^P) \end{bmatrix} \hat{\boldsymbol{A}}\boldsymbol{h} = \boldsymbol{0}, \tag{45}$$

which will have a unique solution (up to scale) provided $\hat{\boldsymbol{B}}$ has full column rank. Matrix $\hat{\boldsymbol{B}}$ will have full column rank provided it has more rows than columns:

$$(P+1)(L-K+1) \geq K, \tag{46}$$

which is true by assumption. From the roots of the filter we can retrieve the locations of the spikes provided $(L/M/T) < 1$. This last condition is required since otherwise there will be an ambiguity in the location of the spikes due to the periodicity of the complex exponentials. The amplitudes can be computed using the retrieved locations solving the system in (42).

For the case of complex moments we can also define situations for which perfect reconstruction is possible regardless of the signal's parameters:

Corollary 3: (Complex exponential reproduction) Under the same assumptions as in Theorem 2 and for the case of $L \geq 2K-1$, perfect recovery is always possible.

Proof: It follows from the fact that for $L \geq 2K-1$, $\boldsymbol{V}_{L-K}(\boldsymbol{u})$ is a Vandermonde matrix of full column rank and thus, the conditions of Theorem 2 are met. ■

Corollary 4: Under the same assumptions as in Theorem 2 and for $P \geq K-1$, perfect recovery is always possible.

Proof: Note that $\hat{\boldsymbol{B}}$ has $\boldsymbol{V}_P(\boldsymbol{t})$ as submatrix and since $P \geq K-1$ it follows that K linearly-independent (row) vectors can be found, which makes the rank of $\hat{\boldsymbol{B}}$ equal to K. ■

If the conditions of Theorem 1 or Theorem 2 (or both) are met, a general procedure for estimating the parameters of a finite stream of Diracs is summarized in Algorithm 1.

C. From Discrete to Continuous Moments

The reconstruction method for sparse FRI signals in Algorithm 1 requires the computation of the modulated moments of the signal. This can be achieved if we have access to the sequences $c_n(\omega_k, p)$ for exponential polynomial reproduction. Next we describe a procedure that allows the computation of the modulated moments in a recursive way by using frequency domain information about the kernel and its derivatives at a given frequency. Define the *discrete modulated moments* as

$$m_p(\omega_0) = \sum_{n\in\mathbb{Z}} n^p\, e^{-j\omega_0 n}\, y_n,\; p = 0,\ldots,P. \tag{47}$$

Making use of the following Fourier identities:

$$\sum_{n\in\mathbb{Z}} n^p\, y_n \stackrel{\mathcal{F}}{\longleftrightarrow} j^p \left.\frac{d^p Y}{d\omega^p}(e^{j\omega})\right|_{\omega=0} \tag{48}$$

$$y_n\, e^{-j\omega_0 n} \stackrel{\mathcal{F}}{\longleftrightarrow} Y(e^{j(\omega+\omega_0)}), \tag{49}$$

Algorithm 1: (Reconstruction for Strang-Fix kernels).

1: Compute the modulated moments $\tau_p(\omega_k)$ for all $\omega_k \in \mathcal{W}$ and $p = 0, \ldots, P_k$

2: Form the (block) Toeplitz matrix

$$\boldsymbol{T} = \begin{bmatrix} \mathrm{T}_K(\boldsymbol{\tau}(\omega_0)) \\ \mathrm{T}_K(\boldsymbol{\tau}(\omega_1)) \\ \vdots \\ \mathrm{T}_K(\boldsymbol{\tau}(\omega_L)) \end{bmatrix}, \quad \text{or} \quad \boldsymbol{T} = \begin{bmatrix} \mathrm{T}_K(\boldsymbol{\tau}_0) \\ \mathrm{T}_K(\boldsymbol{\tau}_1) \\ \vdots \\ \mathrm{T}_K(\boldsymbol{\tau}_P) \end{bmatrix}$$

3: Find the annihilating filter using the SVD

$$\boldsymbol{h} = \arg\min_{\|\boldsymbol{x}\|=1} \|\boldsymbol{T}\boldsymbol{x}\|$$

4: Find the roots of the filter $H(z) = \sum_{k=0}^{K} h_k z^{-k}$ and retrieve the locations of the spikes

5: Find the amplitudes solving a linear system based on

$$\tau_p(\omega_k) = \sum_{i=0}^{K-1} a_i \left(\frac{t_i}{T}\right)^p e^{-j\omega_k t_i / T}$$

we immediately realize that

$$m_p(\omega_0) = \sum_{n\in\mathbb{Z}} n^p\, e^{-j\omega_0 n}\, y_n = j^p\, \frac{d^p Y}{d\omega^p}(e^{j\omega_0}). \tag{50}$$

Consider again the expression for the DTFT of y_n given in (20). By successively differentiating (20) and, under the assumptions made on the sampling kernel, it follows that

$$\frac{d^p Y}{d\omega^p}(e^{j\omega_0}) = \sum_{r=0}^{p} \binom{p}{r} \frac{d^r \hat{\varphi}^*}{d\omega^r}(\omega_0)(-j)^{(p-r)} \tau_{p-r}(\omega_0). \tag{51}$$

By combining (50) and (51) we end up with

$$m_p(\omega_0) = \sum_{r=0}^{p} \binom{p}{r} j^r\, \frac{d^r \hat{\varphi}^*}{d\omega^r}(\omega_0)\, \tau_{p-r}(\omega_0), \tag{52}$$

which gives us an explicit relationship between the continuous and discrete moments of the signal. Let us rewrite (52) in matrix-vector form as

$$\begin{bmatrix} m_0(\omega_0) \\ m_1(\omega_0) \\ \vdots \\ m_P(\omega_0) \end{bmatrix} = \underbrace{\begin{bmatrix} L_{00}(\omega_0) & 0 & \cdots & 0 \\ L_{10}(\omega_0) & L_{11}(\omega_0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ L_{P0}(\omega_0) & \cdots & & L_{PP}(\omega_0) \end{bmatrix}}_{\boldsymbol{L}} \boldsymbol{\tau}(\omega_0), \tag{53}$$

where the nonzero entries of matrix $\boldsymbol{L}$ are given by

$$L_{pr}(\omega_0) = \binom{p}{p-r} j^{p-r}\, \frac{d^{p-r} \hat{\varphi}^*}{d\omega^{p-r}}(\omega_0). \tag{54}$$

Note that since matrix $\boldsymbol{L}$ is lower-triangular, the system of equations in (53) can be solved efficiently using forward substitution. Note also, that the above system is invertible since, by assumption (Strang-Fix conditions), the diagonal elements $L_{rr}(\omega_0) = \hat{\varphi}^*(\omega_0)$ are different from zero. Based on these derivations we can now state the following result:

Lemma 1: (Recursive moments) Let $\varphi(t)$ be a finite-support kernel that satisfies the $(\mathcal{W}, P)$ Strang-Fix conditions for some set of frequencies $\mathcal{W} = \{\omega_0, \ldots, \omega_{L-1}\}$ and some order $P \in \mathbb{Z}_+$. Let $x(t)$ be as in (7) and consider the sampling setup of Fig. 1. Then, the modulated moments of $x(t)$ can be computed in a recursive way from the discrete moments as

$$\tau_p(\omega_k) = \hat{\varphi}^*(\omega_k)^{-1}\Big(m_p(\omega_k) - \sum_{r=0}^{p-1} L_{pr}(\omega_k)\, \tau_r\Big), \tag{55}$$

where $L_{pr}(\omega_k) = \binom{p}{p-r} j^{p-r} \frac{d^{p-r}\hat{\varphi}^*}{d\omega^{p-r}}(\omega_k)$.

Proof: The result follows directly from (53) and the fact that $L_{rr}(\omega_k) = \hat{\varphi}^*(\omega_k)$ for all $r = 0, \ldots, P$.

We have derived an expression that allows us to compute the continuous moments of the signal from the samples. The interesting property of (55) is that it allows us to compute the moments exactly without the need to compute or approximate the dual basis of $\varphi(t)$. All we need is to evaluate the pth derivative of the sampling kernel at the set of reproducing frequencies $\mathcal{W}$. This is particularly convenient in situations where we either have an explicit expression for the derivatives or an efficient way for their evaluation. Note also that from (55) we can derive the reproducing sequences:

Lemma 2: (Reproducing sequence) Let $\varphi(t)$ be a finite-support kernel that satisfies the (ω_0, P) Strang-Fix conditions. Then the sequence $c_n(\omega_0, p)$ for exponential polynomial reproduction can be computed as

$$c_n(\omega_0, p) = e^{j\omega_0 n} \sum_{r=0}^{p} C_{pr}\, n^r, \qquad p = 0, \ldots, P, \tag{56}$$

where C_{pr} is the (p, r)th element of matrix $(\boldsymbol{L}^*)^{-1}$ with $(0, 0)$ denoting the first (top-left) element.

Proof: The result follows directly from inverting the system in (53) and the definition of $m_p(\omega_0)$ in (50). ∎

Example - B-Spline: Consider, as an illustration, that $\varphi(t)$ is a B-Spline [7] of order $P = 2$, whose frequency response is $\hat{\varphi}(\omega) = \left[\frac{\sin(\omega/2)}{\omega/2}\right]^3$. Its derivatives at $\omega = 0$ are given by

$$\hat{\varphi}(0) = 1, \qquad \frac{d\hat{\varphi}}{d\omega}(0) = 0, \qquad \frac{d^2\hat{\varphi}}{d\omega^2}(0) = 1/4.$$

Computing $\boldsymbol{L}^{-1}$ it can be easily verified that

$$c_n(0,0) = 1,\ c_n(0,1) = n,\ c_n(0,2) = n^2 - 1/4.$$

Example - E-Spline: Polynomial exponentials can be reproduced by cardinal E-Splines [8]. Following the notation in [8] let $\beta_{\boldsymbol{\alpha}}(t)$ denote a centered E-Spline of order K and parameter vector $\boldsymbol{\alpha} = [\alpha_0, \ldots, \alpha_{K-1}]$ where

$$\beta_{\boldsymbol{\alpha}}(t) = \beta_{\alpha_0}(t) * \cdots * \beta_{\alpha_{K-1}}(t),\ \beta_\alpha(t) = \begin{cases} e^{\alpha t} & -\frac{1}{2} \le t < \frac{1}{2} \\ 0 & \text{otherwise} \end{cases} \tag{57}$$

Consider an E-Spline with $\boldsymbol{\alpha} = [-j\frac{\pi}{2}, -j\frac{\pi}{2}, 0, 0, j\frac{\pi}{2}, j\frac{\pi}{2}]$. That is, purely imaginary frequencies (complex exponentials) with double multiplicity. Since frequencies appear in complex conjugate pairs the resulting kernel is a real function (see Fig. 2). The double multiplicity of the frequencies gives the kernel the reproduction property of polynomial (degree one) exponentials at the set of frequencies in $\boldsymbol{\alpha}$, in other words it satisfies the $(\mathcal{W} = \{-\pi/2, 0, \pi/2\}, P = 1)$ Strang-Fix conditions. For these parameters, the Fourier transform of $\beta_{\boldsymbol{\alpha}}(t)$ is given by

$$\hat{\beta}_{\boldsymbol{\alpha}}(\omega) = \left[\operatorname{sinc}\left(\frac{\omega - \pi/2}{2}\right) \operatorname{sinc}\left(\frac{\omega}{2}\right) \operatorname{sinc}\left(\frac{\omega + \pi/2}{2}\right)\right]^2, \tag{58}$$

where $\operatorname{sinc}(\omega) = \sin(\omega)/\omega$. For the considered sampling kernel, it can be verified that

$$\boldsymbol{L}(0) = \frac{64}{\pi^4}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \ \boldsymbol{L}(\frac{\pi}{2}) = \boldsymbol{L}^*\left(-\frac{\pi}{2}\right) = \frac{32}{\pi^4}\begin{bmatrix} 1 & 0 \\ j\frac{\pi-6}{\pi} & 1 \end{bmatrix}.$$

V. Arbitrary Kernels and Noise

We have been discussing situations where it is possible to retrieve the original signal from its samples provided some assumptions on the sampling kernel are met, and for the case of uncorrupted (noiseless) measurements. But in practice a given sampling kernel might not satisfy the desired properties. Additionally, as it is inherent to any real-world measurement system, there will be some noise during the acquisition process. We address how to handle such situations.

Assume an additive noise model for the observations $\tilde{y}_n$ as

$$\tilde{y}_n = y_n + v_n, \qquad n = 0, \ldots, M_0 - 1, \tag{59}$$

where v_n is independent and identically distributed (i.i.d.) white noise with zero mean and variance $\mathrm{E}\left[|v_n|^2\right] = \sigma_v^2$. Note that (59) holds for both the periodic ($M_0 = N$) and non-periodic cases. From a practical perspective, it is also convenient to do estimation using complex moments (i.e. Fourier-based information) since they yield better conditioned systems [10]. Therefore, we will restrict our attention to frequencies lying on a uniform grid when dealing with arbitrary sampling kernels. With all these considerations, let $\tilde{Y}_k$ be the M-point DFT of $\tilde{y}_n$, where $M = N$ in the periodic case and $M \geq M_0$ for the non-periodic case. Making use of (21), we can now write the DFT of $\tilde{y}_n$ as

$$\tilde{Y}_k = X_k(\tau)\hat{\varphi}^*\left(\frac{2\pi k}{M}\right) + \underbrace{\sum_{\substack{\ell \in \mathbb{Z} \\ \ell \neq 0}} X_{k+\ell M}(\tau)\hat{\varphi}^*\left(\frac{2\pi k}{M} + 2\pi\ell\right)}_{\text{aliasing}} + V_k, \tag{60}$$

where $\tau = \frac{M}{N}$ and V_k is the M-point DFT of v_n. We will omit the dependence of X_k w.r.t. τ to simplify notation.

A. Estimating FS Coefficients

Recall that $2K$ consecutive FS coefficients X_k provide a sufficient characterization of the stream of Diracs $x(t)$. Therefore, our focus would be on retrieving those FS coefficients from the set of noisy filtered observations $\tilde{y}_n$. We already know that the ability to recover such coefficients would be intimately related to the choice of the sampling kernel. In a general setup, a reasonable strategy would be to estimate those coefficients as accurately as possible. This is precisely the approach that we will follow here which turns out to have a lot of parallelism to the approximate FRI framework [10]. The proposed estimation framework can be seen as a general procedure for FRI signal estimation in the frequency domain.

From (60) we can identify three different components of $\tilde{Y}_k$ – one corresponding to the desired FS coefficient X_k, a second term corresponding to aliasing, and the noise term V_k. A simple way to estimate the FS coefficients is to completely disregard the effect of aliasing and noise, and just compensate for the attenuation introduced by the frequency response of the sampling kernel. In analogy to communication systems we call this approach *zero-forcing* estimation and it corresponds to

$$\tilde{X}_k^{ZF} = \frac{1}{\hat{\varphi}^*(\omega_k)}\tilde{Y}_k, \ \omega_k = \frac{2\pi k}{M}. \tag{61}$$

Note that (61) might result in a significant noise amplification depending on the kernel's frequency response and noise level.

A more robust estimate can be obtained by minimizing the mean squared error. In order to derive such an estimate we treat both the locations and amplitudes of the spikes as random variables. More concretely, we assume that a_i are zero-mean i.i.d. random variables independent from the locations t_i. The locations of the spikes are i.i.d. random variables uniformly distributed over the unit interval $t_i \sim \mathcal{U}\big([0,1)\big)$. Both amplitudes and locations are also independent from the noise v_n.

Under these assumptions the FS coefficients X_k are zero-mean uncorrelated random variables since:

$$\mathrm{E}\left[X_k\right] = \sum_{i=0}^{K-1} \underbrace{\mathrm{E}\left[a_i\right]}_{0} \mathrm{E}\left[e^{-j2\pi k t_i}\right] = 0, \ \mathrm{E}\left[X_k X_\ell^*\right] = \sigma_a^2\,\delta_{k-\ell}, \tag{62}$$

where we have defined $\sigma_a^2 = \sum_{i=0}^{K-1}\mathrm{E}\left[|a_i|^2\right]$.

Let $\tilde{X}_k^{MSE} = C_k\tilde{Y}_k$ be the linear minimum MSE estimate. The orthogonality principle states that the error should be orthogonal to the signal subspace, that is

$$\mathrm{E}\left[(X_k - C_k\tilde{Y}_k)\tilde{Y}_k^*\right] = 0, \tag{63}$$

which means that the optimal coefficient is

$$C_k = \mathrm{E}\left[X_k\tilde{Y}_k^*\right] / \,\mathrm{E}\left[\tilde{Y}_k\tilde{Y}_k^*\right]. \tag{64}$$

Since X_k and V_k are zero-mean independent random variables, and $\mathrm{E}\left[X_k X_\ell^*\right] = \sigma_a^2\,\delta_{k-\ell}$, then it follows that

$$\mathrm{E}\left[X_k\tilde{Y}_k^*\right] = \sigma_a^2\,\hat{\varphi}(\omega_k). \tag{65}$$

Similarly, for $\mathrm{E}\left[\tilde{Y}_k\tilde{Y}_k^*\right]$ it follows that

$$\mathrm{E}\left[\tilde{Y}_k\tilde{Y}_k^*\right] = \sigma_a^2\,\hat{a}_\varphi(e^{j\omega_k}) + \hat{\sigma}_v^2, \tag{66}$$

where $\hat{\sigma}_v^2 = M_0\sigma_v^2$ and $\hat{a}_\varphi(e^{j\omega_k}) = \sum_{\ell\in\mathbb{Z}}|\hat{\varphi}(\omega_k + 2\pi\ell)|^2$ is the DFT of the auto-correlation sequence $a_n = \langle\varphi(t-n), \varphi(t)\rangle$. Using (65) and (66), we can write the LMMSE estimate of the FS coefficient X_k as

$$\tilde{X}_k^{MSE} = \frac{\sigma_a^2\,\hat{\varphi}(\omega_k)}{\sigma_a^2\,\hat{a}_\varphi(e^{j\omega_k}) + \hat{\sigma}_v^2}\tilde{Y}_k. \tag{67}$$

Not surprisingly, (67) is nothing but a Wiener filtering operation in the frequency domain. Note also that, as expected, (67) simplifies to (61) if we do not consider aliasing and noise. In the noiseless case ($\sigma_v^2 = 0$) the coefficients in (67) correspond to the least-squares coefficients in the approximate FRI framework of [10]. Another interpretation of this result is that, in the noiseless case, finding the best coefficients that reproduce complex exponentials is equivalent to minimizing the estimation error of the FS coefficients since we are effectively trying to compensate for the distortion introduced by the sampling kernel. The procedure can also be seen (also in the noiseless case) as using the optimal digital filter for approximating an ideal lowpass filter [31].

B. Denoising—Exploiting Low-Rank Structure

Once we have an estimate of the FS coefficients, we can use Prony's method to retrieve the locations of the spikes. However, it is possible to improve our estimates of the FS coefficients by exploiting the additional structure present in the estimation problem. As already noted, the *annihilation equation* (4) reveals the low-rank property of matrix $\boldsymbol{X}$. However, this situation only holds in the noiseless case. In the presence of noise, the low-rank property of $\boldsymbol{X}$ is very unlikely to hold. Nevertheless, we can take advantage of such structure in order to improve the estimation of the FS coefficients and hence, of the annihilating polynomial. Then it seems natural to enforce this low-rank property during the estimation process. In its basic form, the problem is equivalent to a structured total least squares problem and it appears in different applications in signal processing and control [32], [33]. More generally, the problem can be seen as producing a structured low-rank approximation of the noisy input data matrix. Assume that we have access to a set of $L+1$ noisy FS coefficients given by

$$\tilde{X}_k = X_k + E_k, \qquad k = 0, \ldots, L, \tag{68}$$

where E_k is some error. Let $\tilde{\boldsymbol{x}} = [\tilde{X}_0, \ldots, \tilde{X}_L]^T$ denote the vector of estimates and form the general convolution matrix $\boldsymbol{X}_0 = \mathrm{T}_B\left(\tilde{\boldsymbol{x}}\right)$, where $K \leq B \leq L - K + 1$ and $L \geq 2K - 1$. Let denote by $\mathcal{R}_K$ the set of matrices of rank at most K, and by $\mathcal{T}$ the set of Toeplitz matrices of appropriate dimensions (i.e. $(L - B + 1) \times (B + 1)$). The problem can be then formulated as the following structured low-rank approximation problem:

$$\hat{\boldsymbol{X}} = \arg\min_{\boldsymbol{X}} \frac{1}{2}\|\boldsymbol{X} - \boldsymbol{X}_0\|_F^2 + \mathbb{I}(\boldsymbol{X} \in \mathcal{R}_K) + \mathbb{I}(\boldsymbol{X} \in \mathcal{T}) \tag{69}$$

where $\mathbb{I}(\cdot)$ is the indicator function defined as

$$\mathbb{I}(x) = \begin{cases} 0 & x \text{ is true} \\ +\infty & \text{otherwise} \end{cases} \tag{70}$$

Solving (69) is, in general, NP-hard except for a few special cases. The problem is then usually solved locally [34], [35]. An alternative strategy would be to simplify the problem and replace the rank constraint by its convex surrogate using the nuclear norm. The work in [36] shows that it is possible to recover minimum rank solutions subject to linear constraints by minimizing the nuclear norm. However, for the problem at hand, the Toeplitz structure of the matrix and the fact that two spikes might be arbitrarily close, might result in the rows of the matrix being very coherent thus, violating the assumptions for perfect recovery in [36]. The authors in [13] evaluated the convex relaxation approach but it resulted in a poor performance compared to current methods for spike retrieval. A popular method in the FRI literature is to use Cadzow's iterative denoising [37]. Given the noisy Toeplitz matrix $\boldsymbol{X}_0$, the method iteratively alternates two projection steps. The first step is a projection onto the set of matrices with rank at most K, which can be efficiently computed using the SVD of the input matrix [38]. The second step projects back to the convex set of Toeplitz matrices and it corresponds to averaging over the diagonals of the input matrix. The iterative process continues until some convergence criterion is met. The procedure can be interpreted as a particular instance of alternating projections and it is closely related to singular spectrum analysis [39].

A more recent method has been proposed in [13] that is closely related to the Douglas-Rachford iteration method. The method works well in practice but it is restricted to differentiable loss functions such as the ℓ_2-norm (or Frobenius norm). We take a similar approach here but instead of treating the problem as a matrix approximation problem we keep the focus on the observed FS sequence. The mapping to a Toeplitz matrix is then accounted for by the appropriate linear operator. We propose the following general formulation for the denoising problem:

$$\begin{array}{ll} \underset{\boldsymbol{X},\,\boldsymbol{e}}{\text{minimize}} & \ell(\boldsymbol{e}) + \mathbb{I}(\boldsymbol{X} \in \mathcal{R}_K) \\ \text{subject to} & \boldsymbol{X} = \mathrm{T}_B(\tilde{\boldsymbol{x}} - \boldsymbol{e}) \end{array}, \tag{71}$$

where $\ell(\cdot)$ is some convex loss function. The vector $\boldsymbol{e}$ in (71) represents the error between the true and the estimated FS coefficients. Note that the proposed formulation is very general and that it can be tailored to different setups (e.g. correlated noise, sparsity) by properly choosing the loss function. Additionally, one could also use some of the ideas coming from robust statistics [40] and robust subspace learning (see [41] and references therein) in order to handle the presence of outliers and/or missing data.

In order to solve (71) we propose to use the ADMM optimization framework (see [17] for an overview). By doing so, we end up with iterative schemes with easy-to-compute updates that are of the same complexity as in Cadzow denoising (complexity dominated by SVD). In general, there are no convergence guarantees for Cadzow's iterative denoising nor there are for the approach in [13]. Recent work [42] shows that ADMM is guaranteed to converge even for non-convex and non-smooth problems provided some conditions are met. The particular problem at hand satisfies such conditions when the loss function is smooth (differentiable) and hence, the iterates given by ADMM are guaranteed to converge to a stationary point of the augmented Lagrangian of (71) [42]. In this contribution we restrict our attention to the case where the loss function is a weighted ℓ_2 norm given by $\ell(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x}$, for some positive semi-definite matrix $\boldsymbol{Q}$. The denoising procedure is summarized in Algorithm 2. The matrix $\boldsymbol{\Gamma}$ in the algorithm is a diagonal matrix whose diagonal entries are the number of elements on each diagonal of the Toeplitz matrix. For the ordering

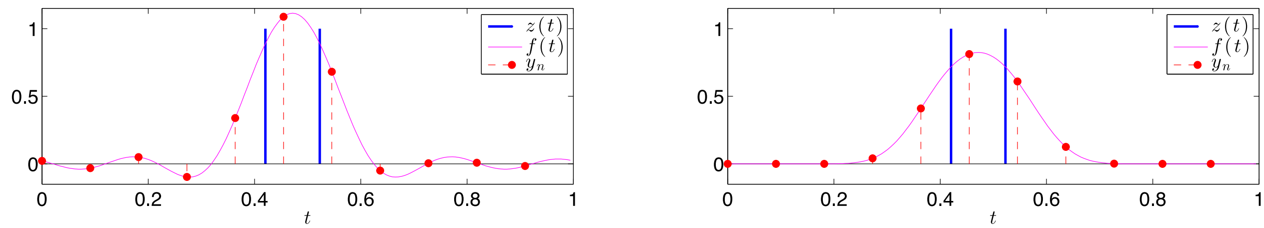


Fig. 4. Signals in the sampling setup of Fig. 1 for a stream of $K = 2$ Diracs located at $\boldsymbol{t} = [0.42, 0.52]$. The left plot correspond to the use of an ideal lowpass filter as the sampling kernel, while the plot on the right corresponds to filtering with B-Spline of order 5. The sampling rate is $N = 1/T = 11$.

we use the convention that the first diagonal corresponds to the top-right one. The operator $\mathrm{T}_B^*(\cdot)$ denotes the adjoint of $\mathrm{T}_B(\cdot)$ and maps a matrix to a sequence by adding up the elements along the diagonals. It is then easy to realize that $\boldsymbol{\Gamma}^{-1}\,\mathrm{T}_B^*(\boldsymbol{X})$ results in a sequence that is the average over the diagonals of the input matrix $\boldsymbol{X}$. The operator $\mathrm{Proj}_{\mathcal{R}_K}(\cdot)$ denotes the projection onto the set of matrices of rank at most K (e.g. using SVD). The scalar $\rho > 0$ is a parameter of the algorithm.

Algorithm 2: (Denoising - $\ell(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x}$).

1: $\boldsymbol{e}^{(0)} = \boldsymbol{0}, \boldsymbol{X}^{(0)} = \boldsymbol{X}_0, \boldsymbol{U}^{(0)} = \boldsymbol{0}$
2: **while** !Exit condition **do**
3: $\boldsymbol{e}^{(k+1)} = \left(\frac{1}{\rho}\boldsymbol{\Gamma}^{-1}\boldsymbol{Q}+\boldsymbol{I}\right)^{-1}\left(\tilde{\boldsymbol{x}}-\boldsymbol{\Gamma}^{-1}\,\mathrm{T}_B^*\left(\boldsymbol{X}^{(k)}+\boldsymbol{U}^{(k)}\right)\right)$
4: $\boldsymbol{X}^{(k+1)} = \mathrm{Proj}_{\mathcal{R}_K}\left(\boldsymbol{X}^{(k)}+\mathrm{T}_B\left(\boldsymbol{e}^{(k+1)}\right)-\boldsymbol{X}_0+\boldsymbol{U}^{(k)}\right)$
5: $\boldsymbol{U}^{(k+1)} = \boldsymbol{U}^{(k)} + \boldsymbol{X}^{(k+1)} + \mathrm{T}_B\left(\boldsymbol{e}^{(k+1)}\right) - \boldsymbol{X}_0$
6: **end while**

Algorithm 3: (Frequency-domain FRI signal estimation).

1: Compute M-point DFT of the noisy input sequence
2: Estimate FS coefficients
3: Denoising using Algorithm 2
4: Annihilating filter for location and amplitude estimation

We have just described a general pipeline for the estimation of sparse FRI signals using frequency-domain information. We outline the overall procedure in Algorithm 3.

TABLE I
ESTIMATION METHODS EVALUATED AND PARAMETERS

Method	No. Iter.	Parameters
Prony [27]	1	–
ESPRIT [19]	1	–
IQML [25]	50	–
Cadzow [37]	50	–
Condat-Hirabayashi [13]	50	$\mu = 0.1,\ \gamma = 0.51\mu$
Proposed (ADMM)	50	$\rho = 0.5$

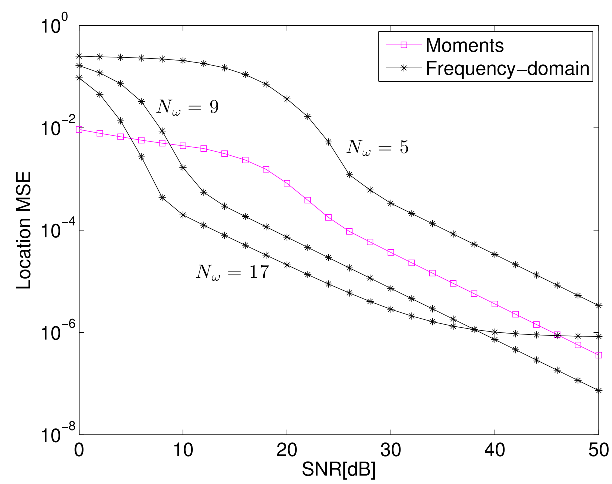


Fig. 5. MSE in location estimation for a stream of $K = 2$ Diracs when the sampling kernel satisfies the generalized Strang-Fix conditions. The plot compares moment-based reconstruction to the frequency-domain approach.

VI. NUMERICAL EXPERIMENTS

In this section we conduct numerical examples to evaluate the proposed estimation framework. For the simulations we use additive zero-mean white Gaussian noise to corrupt our measurements and define the Signal to Noise Ratio (SNR) as $SNR = 10\log_{10}\left(\|\boldsymbol{y}\|^2/N\sigma_v^2\right)$, where $\boldsymbol{y}$ is the vector of noiseless signal samples in the time domain and σ_v^2 is the noise variance. Since the accuracy in the estimation of the amplitudes depends on the accuracy of the location estimates (recall it is a linear problem given the locations) we evaluate only the location estimation error for the estimation methods listed in Table I. For those methods that are iterative we fix the number of iterations to 50. For the method in [13] we have used the implementation (with default parameters) provided by the authors. The code to reproduce the results in this section can be found at https://infoscience.epfl.ch and in the supplementary material.

Reconstruction With Generalized Strang-Fix Kernels: In this first experiment we consider the estimation of a non-periodic stream of $K = 2$ Diracs located at $\boldsymbol{t} = [0.12, 0.32]$ when the sampling kernel satisfies the generalized Strang-Fix conditions. For that purpose we use the estimation procedure outlined in Algorithm 1. The sampling kernel considered is the E-Spline with parameters $\boldsymbol{\alpha} = [-j\frac{\pi}{2}, -j\frac{\pi}{2}, 0, 0, j\frac{\pi}{2}, j\frac{\pi}{2}]$ used in previous examples (see Fig. 2). Under these assumptions, we fall within the conditions of Theorem 2 and thus, perfect recovery is possible in the noiseless case. Note that the kernel is not able to reproduce $2K$ moments in either of the two axis

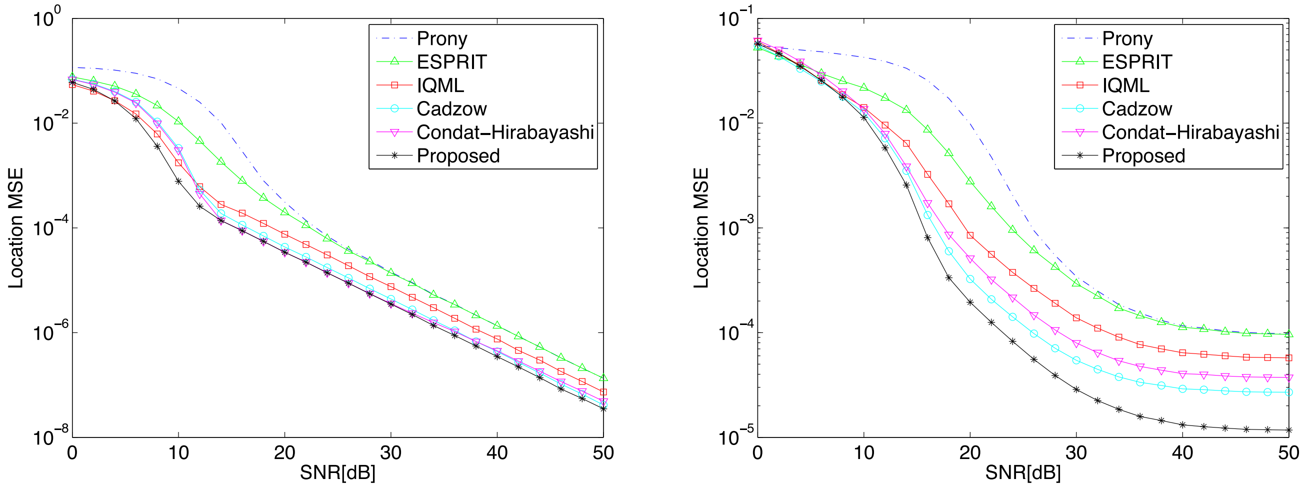


Fig. 6. MSE in location estimation for a periodic stream of $K = 2$ spikes with unit amplitude. The results correspond to an average over 10000 noise realizations for each SNR value. The sampling kernel is an ideal lowpass filter (left) and a B-Spline of order 5 (right). The sampling rate is set to $N = 11$.

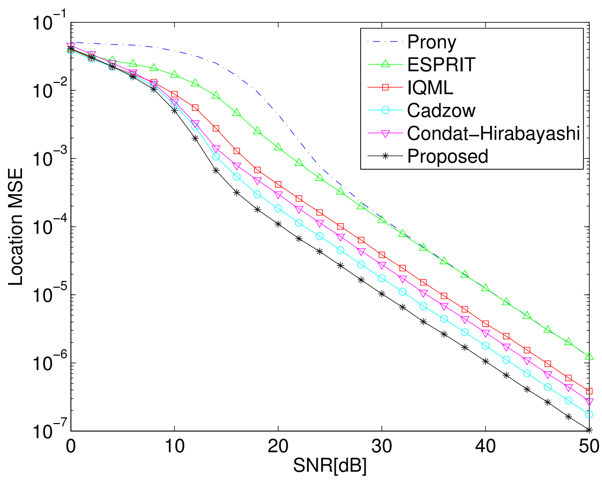


Fig. 7. MSE in location estimation for a periodic stream of $K = 2$ spikes with unit amplitude. The results correspond to an average over 10000 realizations for each SNR value. The sampling kernel is a B-Spline of order 5 and the sampling rate is set to $N = 22$. For estimation 11 frequencies were used.

in the $L - P$ plane separately, which prevents the use of previous results on FRI signal reconstruction. We run a simulation for different SNR values and provide a comparison with the frequency-domain estimation approach described in Algorithm 3. The sampling rate is set to $N = 11$ and the number of realizations per SNR is 10000. The estimation results in terms of MSE are given in Fig. 5. It can be appreciated that exploiting the reproduction properties of the kernel gives more accurate estimates than the general frequency-domain approach for the same model complexity (i.e. use $N_\omega = 2K + 1 = 5$ frequencies for estimation). However, the frequency-domain method can perform better by increasing the number of frequencies used for estimation. At some point, the frequency-domain method saturates due to signal attenuation in the frequency domain and aliasing. This result suggests that a proper combination of the two methods could lead to an overall improved reconstruction. Choosing an optimal subset of moments and frequencies is going to be kernel-dependent and it is left as an open question for further research.

Effect of the Sampling Kernel: Let us now compare the estimation performance of the proposed frequency-domain estimation framework when the sampling kernel deviates from the ideal scenario. More concretely, we consider the estimation of a periodic stream of $K = 2$ Diracs with unit amplitude (see Fig. 4) when the sampling kernel $\varphi(t)$ is an ideal lowpass filter (alias-free, no attenuation in frequency) or a B-Spline of order 5. We run a simulation for different noise levels and compute the average localization error for different state-of-the-art methods (see Table I). For the estimation of the FS coefficients we use the zero-forcing estimate as per (61). For the denoising part we use the weighted ℓ_2-loss function with $\boldsymbol{Q} = \text{diag}\left(|\hat{\varphi}(\omega_0)|^2, \ldots, |\hat{\varphi}(\omega_{N-1})|^2\right)$. The results are displayed in Fig. 6. We can appreciate that the proposed scheme achieves the best performance at all SNR values (except for the very low-ones) and for both sampling kernels. We can also observe a saturation state in Fig. 6 (right) due to the presence of aliasing. Since the error floor is due to aliasing introduced by the sampling kernel, an obvious strategy to reduce its effect is to increase the sampling rate. We repeat the previous experiment but now we use twice as many samples per unit time (i.e. $N = 22$). For the estimation of the locations of the spikes we use the same number ($N_\omega = 11$) of frequencies as in the previous case. In Fig. 7 we present the results. We can observe now that the error floor is not present at the considered range of SNR values and that all methods exhibit a linear error decay in the high SNR regime.

Random Spike Generation: Instead of considering a fix separation as in the previous examples, let us now consider the case where the spikes are randomly located. For that purpose, we use again a periodic stream of two spikes sampled using a B-Spline sampling kernel of order 5 and a sampling rate of $N = 22$. The locations of the spikes are now independent random variables uniformly distributed over the interval $[0, 1)$. The amplitudes are also independent random variables that take the values ± 1 with equal probability. As in the previous

TABLE II
MEDIAN PSNR FOR A STREAM OF $K = 2$ SPIKES GENERATED AT RANDOM, SAMPLED WITH A B-SPLINE KERNEL AND USING MMSE ESTIMATES

Method - SNR [dB]	0	5	10	15	20	25	30	35	40	45	50
Prony [27]	20.70	27.67	31.63	34.09	37.26	41.90	46.98	52.04	57.08	62.09	67.09
ESPRIT [19]	22.18	29.21	33.17	35.13	37.85	42.16	47.07	52.08	57.10	62.10	67.09
IQML [25]	17.77	22.04	25.57	29.94	34.82	40.04	45.26	50.34	55.34	60.34	65.34
Cadzow [37]	**24.86**	**31.01**	35.57	39.80	44.32	49.39	54.53	59.59	64.61	69.62	74.61
Condat-Hira. [13]	24.52	29.69	32.97	36.44	41.03	46.38	51.68	56.82	61.88	66.88	71.87
Proposed (ADMM)	24.42	30.71	**36.08**	**41.43**	**46.77**	**52.07**	**57.20**	**62.25**	**67.27**	**72.28**	**77.28**

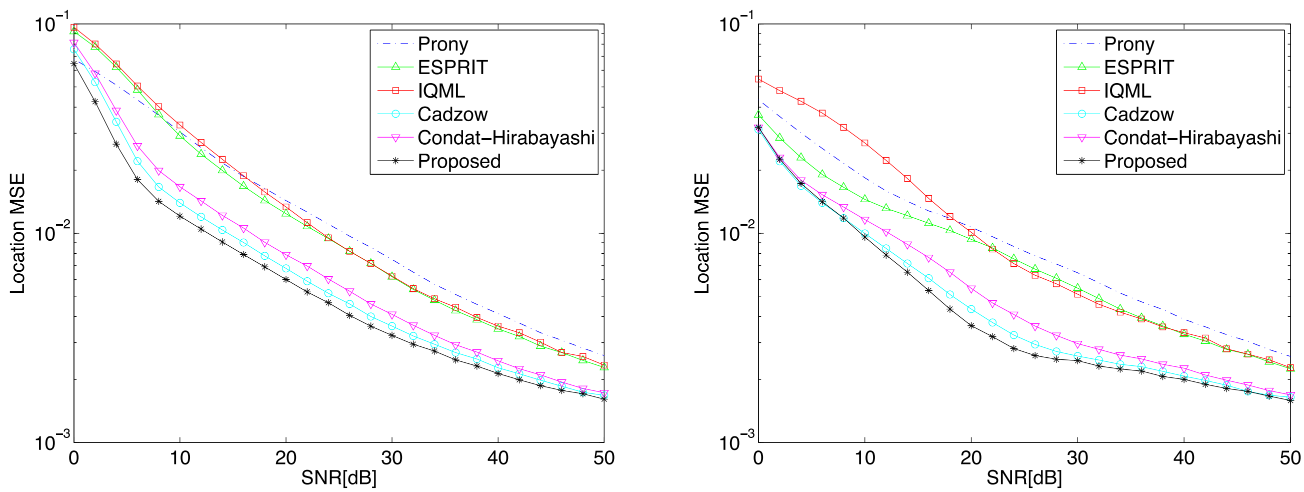


Fig. 8. MSE in location estimation for a periodic stream of $K = 2$ spikes generated at random from a uniform distribution in $[0, 1)$. The results correspond to an average over 10^5 independent realizations for each SNR value. The sampling kernel is a B-Spline of order 5 and the sampling rate is set to $N = 22$. The number of frequencies used for estimation is set to 11. The left plot uses ZF estimates for the FS coefficients while the right plot uses MMSE estimates.

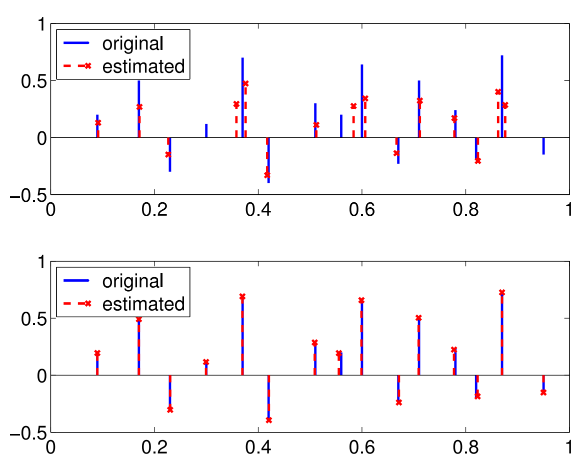


Fig. 9. Reconstruction of $K = 15$ spikes using a B-Spline sampling kernel of order 5 and the MMSE FS estimates. The top plot corresponds to the use of 63 frequencies for estimation while the bottom one uses only $2K + 1 = 31$. The sampling rate is $N = 76$ and the SNR is 30 dB.

experiments we evaluate the error in location estimation for different SNR values. In this case, the localization error is an average of 10^5 independent realizations for each SNR value. We have considered two different scenarios – one where the initial estimate of the FS coefficients is obtained using the ZF method as per (61) and another where the initial estimates are obtained using the MMSE method as per (67). In both cases we use a weighted ℓ_2-norm where $\boldsymbol{Q}$ is chosen, as before, to be a diagonal matrix whose main diagonal is proportional to the squared magnitude of the kernel's frequency response. The results are depicted in Fig. 8. We observe that the proposed estimation scheme provides the best performance at almost all SNR values. It can also be appreciated that the MMSE estimates consistently outperform the ZF estimates for all tested methods, particularly at low SNR values. As an additional metric, we also provide Peak Signal to Noise Ratio (PSNR) values for the case of MMSE estimates. The PSNR is defined in dB as $PSNR = 10\log_{10}(K \max^2(\boldsymbol{t})/\|\boldsymbol{t} - \hat{\boldsymbol{t}}\|^2)$, where $\boldsymbol{t}$ and $\hat{\boldsymbol{t}}$ are the vectors of true and estimated locations, respectively. We report in Table II the median PSNR values over the 10^5 realizations for each SNR value. It can be seen that the proposed method achieves the highest values except at very low SNRs.

A Larger Example: Finally, let us look at a larger example where we try to estimate a periodic stream of $K = 15$ spikes distributed over the unit interval (see Fig. 9). We use the B-spline sampling kernel of order 5 and a sampling rate of $N = 76$ samples. In this case we fix the SNR value to 30 dB and obtain the initial FS estimates using the MMSE estimate. We test two different situations where we compare the effect of the number of frequencies used for estimation. The results are illustrated in Fig. 9.

We can observe the estimation results for the case where we use 63 frequencies for estimation (top plot) and the case where we only use 31 (bottom plot). It can be appreciated that better results are obtained using fewer measurements. The reason for that behavior can be explained based on the frequency response of the sampling kernel. Since the sampling kernel exhibits a lowpass behavior with a fast decay in the frequency domain, using more (higher) frequencies results in an effective decrease of the signal to noise ratio due to the attenuation introduced by the sampling kernel. If instead we use an ideal lowpass filter, better results are obtained as the number of frequencies increases.

VII. Conclusion

We considered the problem of sampling and reconstruction of sparse signals with finite rate of innovation. When the kernel satisfies Strang-Fix conditions, previous approaches have focused on the reconstruction from temporal or complex (frequency) moments of the signal. We have generalized these results providing conditions over the entire time-frequency plane of moments for which perfect signal reconstruction is possible. In the more general case of arbitrary sampling kernels, we have proposed a unified frequency-domain approach to signal estimation and a novel iterative denoising scheme that improves over the state-of-the-art at almost all SNR regimes in the tested scenarios.

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 623–656, 1948.

[2] M. Unser, "Sampling—50 years after Shannon," *Proc. IEEE*, vol. 88, no. 4, pp. 569–587, Apr. 2000.

[3] 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.

[4] P. L. Dragotti, M. Vetterli, and T. Blu, "Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets Strang-Fix," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1741–1757, May 2007.

[5] J. Berent, P. L. Dragotti, and T. Blu, "Sampling piecewise sinusoidal signals with finite rate of innovation methods," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 613–625, Feb. 2010.

[6] G. Strang and G. Fix, "A Fourier analysis of the finite element variational method," in *Constructive Aspects of Functional Analysis*. Rome, Italy: Edizioni Cremonese, 1971, pp. 793–840.

[7] M. Unser, A. Aldroubi, and M. Eden, "B-spline signal processing. I. Theory," *IEEE Trans. Signal Process.*, vol. 41, no. 2, pp. 821–833, Feb. 1993.

[8] M. Unser and T. Blu, "Cardinal exponential splines: Part I—Theory and filtering algorithms," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1425–1438, Apr. 2005.

[9] R. Tur, Y. C. Eldar, and Z. Friedman, "Innovation rate sampling of pulse streams with application to ultrasound imaging," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1827–1842, Apr. 2011.

[10] J. A. Urigüen, T. Blu, and P. L. Dragotti, "FRI sampling with arbitrary kernels," *IEEE Trans. Signal Process.*, vol. 61, no. 21, pp. 5310–5323, Nov. 2013.

[11] I. Maravic and M. Vetterli, "Sampling and reconstruction of signals with finite rate of innovation in the presence of noise," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2788–2805, Aug. 2005.

[12] T. Blu, P. L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, "Sparse sampling of signal innovations," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 31–40, Mar. 2008.

[13] L. Condat and A. Hirabayashi, "Cadzow denoising upgraded: A new projection method for the recovery of dirac pulses from noisy linear measurements," *Sampling Theory Signal Image Process.*, vol. 14, no. 1, pp. 17–47, 2015.

[14] 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.

[15] C. Chen, P. Marziliano, and A. C. Kot, "2D finite rate of innovation reconstruction method for step edge and polygon signals in the presence of noise," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2851–2859, Jun. 2012.

[16] H. Pan, T. Blu, and P. L. Dragotti, "Sampling curves with finite rate of innovation," *IEEE Trans. Signal Process.*, vol. 62, no. 2, pp. 458–471, Jan. 2014.

[17] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[18] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*. Upper Saddle River, NJ, USA: Prentice-Hall, 1997.

[19] A. Paulraj, R. Roy, and T. Kailath, "Estimation of signal parameters via rotational invariance techniques- esprit, in *Proc. 19th Asilomar Conf. Circuits, Syst. Comput.*, Nov. 1985, pp. 83–89.

[20] A. Paulraj, R. Roy, and T. Kailath, "A subspace rotation approach to signal parameter estimation," *Proc. IEEE*, vol. 74, no. 7, pp. 1044–1046, Jul. 1986.

[21] V. F. Pisarenko, "The retrieval of harmonics from a covariance function," *Geophys. J.*, vol. 33, pp. 347–366, Sep. 1973.

[22] N. L. Owsley, "An adaptive search and track array (ASTA)," NUSL Tech. Memo. no. 2242166-69, Jul. 1969.

[23] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," in *Proc. RADC Spectrum Estimation Workshop*, Rome, NY, USA, 1979, pp. 243–258.

[24] G. Bienvenu, "Influence of the spatial coherence of the background noise on high resolution passive methods," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 1979, vol. 4, pp. 306–309.

[25] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 5, pp. 1081–1089, Oct. 1986.

[26] R. Kumaresan, L. Scharf, and A. Shaw, "An algorithm for pole-zero modeling and spectral analysis," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 3, pp. 637–640, Jun. 1986.

[27] R. Prony, "Essai experimental et analytique: Sur les lois de la dilatabilité des fluides élastiques et sur celles de la force expansive de la vapeur de l'alkool à différentes températures," *J. de l'École Polytechnique, Paris*, vol. 1, no. 22, pp. 24–76, 1795.

[28] D. W. Tufts and R. Kumaresan, "Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood," *Proc. IEEE*, vol. 70, no. 9, pp. 975–989, Sep. 1982.

[29] Y. M. Lu and M. N. Do, "A theory for sampling signals from a union of subspaces," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2334–2345, Jun. 2008.

[30] I. Khalidov, T. Blu, and M. Unser, "Generalized L-Spline wavelet bases," in *Proc. SPIE Conf. Math. Imaging: Wavelet XI*, vol. 5914, San Diego CA, USA, Jul. 31–Aug. 3, 2005, pp. 59140F-1–59140F-8.

[31] T. Blu and M. Unser, "Quantitative Fourier analysis of approximation techniques: Part I—Interpolators and projectors," *IEEE Trans. Signal Process.*, vol. 47, no. 10, pp. 2783–2795, Oct. 1999.

[32] Bart De Moor, "Structured total least squares and L2 approximation problems," *Linear Algebra Appl.*, vol. 188, pp. 163–205, 1993.

[33] P. Lemmerling, L. Vanhamme, S. Van Huffel, and B. De Moor, "IQML-like algorithms for solving structured total least squares problems: A unified view," *Signal Process.*, vol. 81, no 9, pp. 1935–1945, 2001.

[34] I. Markovsky, "Structured low-rank approximation and its applications," *Automatica*, vol. 44, no. 4, pp. 891–909, 2008.

[35] I. Markovsky and K. Usevich, "Software for weighted structured low-rank approximation," *J. Comput. Appl. Math.*, vol. 256, pp. 278–292, 2014.

[36] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm Minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, 2010.

[37] J. A. Cadzow, "Signal enhancement—A composite property mapping algorithm," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 1, pp. 49–62, Jan. 1988.

[38] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, 1936.

[39] J. W. Gillard, "Cadzow's basic algorithm, alternating projections and singular spectrum analysis," *Statist. Interface*, vol. 3, no. 3, pp. 335–343, 2010.

[40] P. J. Huber and E. M. Rochetti, *Robust Statistics*, vol. 2. New York, NY, USA: Wiley, 2009.

[41] Danijel Skočaj, Aleš Leonardis, and Horst Bischof, “Weighted and robust learning of subspace representations,” *Pattern Recog.*, vol. 40, no. 5, pp. 1556–1569, May 2007.

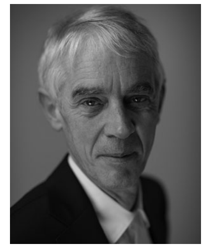
[42] Y. Wang, W. Yin, and J. Zeng, “Global convergence of ADMM in nonconvex nonsmooth optimization,” UCLA CAM Report 15-62, 2015.

Benjamín Béjar Haro received the Electrical Engineering degree from both the Universitat Politècnica de Catalunya (UPC), Barcelona, Spain, and the Technische Universität Darmstadt (TUD), Darmstadt, Germany, in 2006 under the framework of the Double Degree Exchange program, and the Ph.D. degree in electrical engineering from the Universidad Politécnica de Madrid (UPM), Madrid, Spain, in 2012. From September 2011 to September 2013, he was a member of the Vision, Dynamics and Learning Lab, Johns Hopkins University (JHU), Baltimore, MD, USA, as part of the MSE program in biomedical engineering.

He has held research appointments as a visiting Ph.D. student at the Università degli Studi di Udine, Udine, Italy, in 2009 and, at the Hong Kong University of Science and Technology (HKUST), Hong Kong, China, in 2011.

In 2012, he was awarded with the 2012 Best Paper Award in Medical Robotics and Computer Assisted Intervention Systems from the Medical Image Computing and Computer Assisted Intervention Society (MICCAI).

He served as a Postdoctoral Researcher and a Lecturer in the Audiovisual Communications Laboratory (LCAV) at EPFL from October 2013 to October 2017. Currently, he is an Associate Research Scientist in the Center for Imaging Science, The Johns Hopkins University. His research interests include digital signal processing, sampling and reconstruction of sparse signals, image processing, convex optimization, machine learning, and inverse problems in biomedical imaging.



Martin Vetterli (S’86–M’86–SM’90–F’95) received the Dipl. El.-Ing. degree from ETH Zurich (ETHZ), Switzerland, in 1981, the M.S. degree from Stanford University, Stanford, CA, USA, 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 in New York, where he was last an Associate Professor of electrical engineering and the codirector 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 has held an Adjunct Professor position until June 2010. Since 1995, he has been a Professor of Communication Systems at EPF Lausanne, Switzerland, where he chaired the Communications Systems Division (1996/97), and heads the Audiovisual Communications Laboratory. From 2001 to 2004, he directed the National Competence Center in Research on mobile information and communication systems. He also was a Vice-President at EPFL from October 2004 to February 2011 in charge, among others, of international affairs and computing services. He has held visiting positions at ETHZ (1990) and Stanford (1998). From March 2011 to 2012, he was the Dean of the School of Computer and Communication Sciences of EPFL. From 2013 to 2016, he led the Swiss National Science Foundation and since 2017 he is President of EPFL. He is a fellow of IEEE, a fellow of ACM, a fellow of EURASIP, and a member of SIAM and NAE. He is on the editorial boards of *Applied and Computational Harmonic Analysis*, the *Journal of Fourier Analysis and Application* and the IEEE JOURNAL ON SELECTED TOPICS IN SIGNAL PROCESSING and has been elected Foreign Member of the NAE in 2015. He received the Best Paper Award of EURASIP in 1984, the Research Prize of the Brown Bovery Corporation (Switzerland) in 1986, the IEEE Signal Processing Society’s Senior Paper Awards in 1991, in 1996 and in 2006 (for papers with D. LeGall, K. Ramchandran, and Marziliano and Blu, respectively). He won the Swiss National Latsis Prize in 1996, the SPIE Presidential award in 1999, the IEEE Signal Processing Technical Achievement Award in 2001, the IEEE Signal Processing Society Award in 2010 for fundamental contributions to signal processing theory, technology and education, and the IEEE Jack S. Kilby Signal Processing Medal in 2017. He is an ISI highly cited researcher in engineering. He was a member of the Swiss Council on Science and Technology from 2000 to 2003. He was a plenary speaker at various conferences (e.g. IEEE ICIP, ICASSP, ISIT) and is the co-author of three books with J. Kovacevic, “Wavelets and Subband Coding,” 1995, with P. Prandoni “Signal Processing for Communications,” 2008 and with J. Kovacevic and V.K. Goyal, “Foundations of Signal Processing,” 2015. He has published about 170 journal papers on a variety of topics in signal/image processing and communications and holds about 50 patents and patent applications. His research interests include sampling, wavelets, multirate signal processing, computational complexity, signal processing for communications, digital image/video processing, joint source/channel coding, signal processing for sensor networks and inverse problems like acoustic tomography.