Multichannel Sampling with Unknown Gains and Offsets: A Fast Reconstruction Algorithm

Yue M. Lu*† and Martin Vetterli†‡
*School of Engineering and Applied Sciences
Harvard University, Cambridge, MA 02138, USA
†School of Computer and Communication Sciences
Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland
‡Department of Electrical Engineering and Computer Sciences
University of California, Berkeley, CA 94720, USA

***Abstract*—We study a multichannel sampling scheme, where different channels observe scaled and shifted versions of a common bandlimited signal. The channel gains and offsets are unknown *a priori*, and each channel samples at sub-Nyquist rates. This setup appears in many practical signal processing applications, including time-interleaved ADC with timing skews, unsynchronized distributed sampling in sensor networks, and superresolution imaging. In this paper, we propose a new algorithm to efficiently estimate the unknown channel gains and offsets. Key to our algorithm is a novel linearization technique, which converts a system of trigonometric polynomial equations of the unknown parameters to an overparameterized linear system. The computation steps of the proposed algorithm boil down to forming a fixed data matrix from the discrete-time Fourier transforms of the observed channel samples and computing the singular value decomposition of that matrix. Numerical simulations verify the effectiveness, efficiency, and robustness of the proposed algorithm in the presence of noise. In the high SNR regime (40 dB and above), the proposed algorithm significantly outperforms a previous method in the literature in terms of estimation accuracy, at the same time being three orders of magnitudes faster.**

I. Introduction

Consider a multichannel sampling scheme shown in Figure 1, where each channel takes uniform samples of a scaled and shifted version of a common signal $x(t)$. We assume that $x(t)$ is bandlimited, with its Fourier transform supported on $[-\sigma, \sigma]$. Each channel samples at sub-Nyquist rates, *i.e.*, $1/T < \sigma/\pi$.

When the channel gains $\{\alpha_k\}_{k=1}^K$ and offsets $\{\tau_k\}_{k=1}^K$ are known, the problem of reconstructing the input $x(t)$ from its samples $\{y_k[n]\}_{k=1}^K$ is linear. In fact, this task becomes a special case of the classical Papoulis generalized sampling scheme [1], whose extensions and variations have been extensively studied in the literature (see, *e.g.*, [2]–[6]). In this paper, we consider the more challenging case where—in addition to $x(t)$—the gains and offsets are also part of the unknown.

A. Motivations

The multichannel sampling setup described above appears in many practical signal processing applications, some of which we highlight below.

Example 1 (Time-Interleaved ADCs): Designing a single analog-to-digital converter (ADC) with very high sampling rate can be expensive in terms of hardware costs and power consumption. An attractive alternative is to use a parallel array of lower-rate ADCs, working in a time-interleaved fashion [7]. In an ideal time-interleaved ADC, the channel gains (*i.e.*, $\{\alpha_k\}_{k=1}^K$ in Figure 1) are uniform, and the offsets are

$$\tau_k = \frac{k-1}{K} T, \quad (1)$$

where K is the total number of ADCs, and $1/T$ is the sampling rate at each channel. The samples $\{y_k[n]\}$ from different channels can then be directly multiplexed into a single stream, emulating the effect of a virtual single ADC with a higher sampling rate K/T. In practice, however, mismatches among the ADCs lead to nonuniform channel gains as well as timing skews, *i.e.*,

$$\tau_k = \frac{k-1}{K} T + \delta_k,$$

for some unknown $\{\delta_k\}$. Consequently, in order to maintain the performance of the ADC, it is necessary to first estimate the unknown gains and offsets, and then to apply digital compensation to the samples (see, *e.g.*, [8], [9]).

Example 2 (Distributed Sampling): The model in Figure 1 can also describe a distributed sampling scenario [10], where we use K sensors to observe a common bandlimited source signal $x(t)$. In many applications (*e.g.*, sound recording in

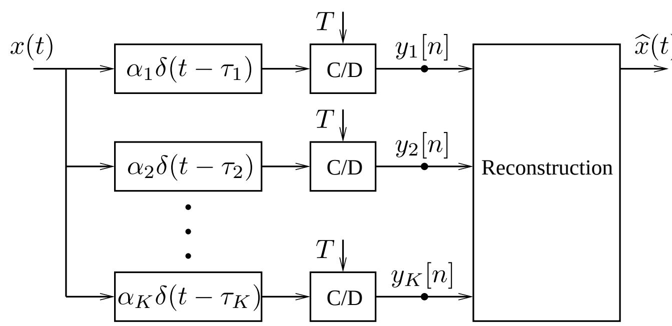

Fig. 1. A multichannel sampling scheme, where each channel observes and uniformly samples a scaled and shifted version of a bandlimited signal $x(t)$. The channel gains $\{\alpha_k\}$ and offsets $\{\tau_k\}$ are not known *a priori*. Our goal is to reconstruct $x(t)$ from its samples $y_k[n] = \alpha_k x(nT - \tau_k)$, taken at sub-Nyquist rates.

non-reverberant rooms, underwater acoustics, etc.), the unknown channel from the source to each sensor can be well-approximated by a one-tap filter $\alpha_k \delta(t - \tau_k)$. The coefficient α_k and delay τ_k are determined by the relative distance between the source and the sensor, as well as the physical properties of the medium. The signals observed at different sensors are highly correlated—after all, they are just filtered versions of the same signal, albeit with unknown filter parameters. Therefore, intuitively, each sensor should be able to sample at a sub-Nyquist rate, but still allowing for perfect reconstruction at a central decoder.

Example 3 (Superresolution Imaging): A 2-D extension of the sampling setup in Figure 1 serves as a fundamental model in super-resolution imaging, where one wants to reconstruct a higher-resolution image from a set of lower-resolution images that are slightly shifted with respect to each other. In this case, the filter at each channel can be written as $\alpha_k \delta(t_1 - \tau_k, t_2 - \xi_k)$, where the coefficient α_k models the exposure differences between images and τ_k, ξ_k the relative shifts of the kth image along the horizontal and vertical axes, respectively.

B. Contributions and Paper Outline

The main contribution of this paper is a new noniterative algorithm that can efficiently estimate the unknown channel gains and offsets. Our algorithm relies on a subspace-based rank condition derived in an earlier work of Vandewalle *et al.* [11]. However, unlike in [11] where the unknown system parameters are estimated by exhaustively testing the rank condition, our algorithm exploits the rank condition much more efficiently, converting a nonlinear minimization problem into a linear system of equations via overparameterization.

The rest of this paper is organized as follows. In Section II, we briefly review the multichannel sampling setup and derive a matrix-vector model in the Fourier domain, linking the observed channel samples to the unknown input signal and system parameters. Based on this forward model, we show in Section III the minimum sampling rate each channel should use to ensure unique signal recoveries. The focus of this paper is on Section IV, where we first have a streamlined derivation of the rank condition of [11] and then present a new estimation algorithm based on a linearization technique. In particular, we convert a system of trigonometric polynomial equations of the unknown parameters derived from the rank condition to an overparameterized linear system. The computation steps of the proposed algorithm then boil down to forming a fixed data matrix from the discrete-time Fourier transforms of the observed channel samples and computing the singular value decomposition of that matrix. Numerical simulations in Section V verify the effectiveness, efficiency, and robustness of the proposed algorithm in the presence of noise. We conclude in Section VI.

Notations: We summarize below the main notations used in this paper. $X(\omega)$ denotes the continuous-time Fourier transform of $x(t)$, defined as

$$X(\omega) \stackrel{\text{def}}{=} \int_{\mathbb{R}} x(t)\, e^{-j\omega t} dt.$$

In this work, we assume that $X(\omega)$ is bandlimited to a fixed interval $[-\sigma, \sigma]$ for some $\sigma > 0$. The corresponding Nyquist rate is σ/π. We use $\lfloor \alpha \rfloor$ to denote the largest integer less than or equal to a real number α; similarly, $\lceil \alpha \rceil$ is the smallest integer greater than or equal to α.

II. Preliminaries

Suppose that an input signal $x(t)$ is sampled by K channels, as in Figure 1. We denote by

$$y_k[n] \stackrel{\text{def}}{=} \alpha_k\, x(nT - \tau_k)$$

the samples taken at the kth channel. Applying the standard sampling formula in the frequency domain, we calculate the discrete-time Fourier transform of $y_k[n]$ as

$$\begin{aligned} Y_k(\omega) &\stackrel{\text{def}}{=} \sum_{n \in \mathbb{Z}} y_k[n]\, e^{-jnT\omega} \\ &= \alpha_k \sum_{n \in \mathbb{Z}} x(nT - \tau_k)\, e^{-jnT\omega} \\ &= \frac{\alpha_k}{T} \sum_{m \in \mathbb{Z}} X(\omega + mc)\, e^{-j(\omega + mc)\,\tau_k}, \end{aligned} \tag{2}$$

where $c \stackrel{\text{def}}{=} 2\pi/T$ is a constant that will appear in many of our later derivations.

The expression in (2) implies that $Y_k(\omega)$ is a periodic function with period c, since $Y_k(\omega) = Y_k(\omega + c)$ for all ω. Consequently, we only need to focus on $Y_k(\omega)$ within a single period of length c. In what follows, we shall always assume that the frequency variable ω falls within the interval

$$\omega \in \mathcal{P} \stackrel{\text{def}}{=} [-\sigma, -\sigma + c). \tag{3}$$

Since $X(\omega)$ has a finite support, the summation in (2) involves only a finite number of nonzero terms. First, consider a special case when $M \stackrel{\text{def}}{=} 2\sigma/c$ is an integer. This implies that the sampling rate at each channel is exactly $1/M$th of the Nyquist rate of the input signal $x(t)$. Under this setup, we can verify that (2) reduces to a finite sum of M terms

$$Y_k(\omega) = \frac{\alpha_k}{T} \sum_{m=0}^{M-1} X(\omega + mc)\, e^{-j(\omega + mc)\,\tau_k} \tag{4}$$

for ω satisfying (3).

In general, when $2\sigma/c$ is not an integer, the situation is slightly more complicated. We demonstrate this through an example in Figure 2. It is clear from the visualization that we need to distinguish between two sub-intervals

$$\mathcal{P}_1 \stackrel{\text{def}}{=} [-\sigma, -\sigma + r) \quad \text{and} \quad \mathcal{P}_2 \stackrel{\text{def}}{=} [-\sigma + r, -\sigma + c), \tag{5}$$

where $r \stackrel{\text{def}}{=} 2\sigma - \lfloor 2\sigma/c \rfloor c$ is the remainder of the "floored division" of 2σ by c. As shown in the figure, when $\omega \in \mathcal{P}_1$, up to $\lceil 2\sigma/c \rceil$ spectral segments will be "aliased" on top of each other in forming (2). When $\omega \in \mathcal{P}_2$, the number of overlapping spectral segments becomes $\lfloor 2\sigma/c \rfloor$. Defining

$$M(\omega) \stackrel{\text{def}}{=} \begin{cases} \lceil 2\sigma/c \rceil & \text{for } \omega \in \mathcal{P}_1 \\ \lfloor 2\sigma/c \rfloor & \text{for } \omega \in \mathcal{P}_2 \end{cases},$$

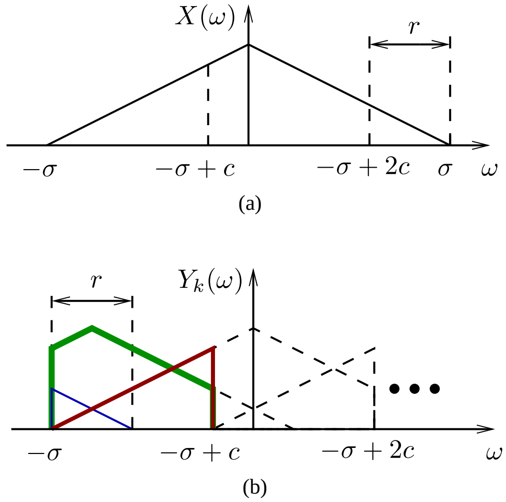

Fig. 2. Illustration of frequency aliasing (overlapping) due to downsampling. (a) The continuous-time Fourier transform $X(\omega)$ of the input signal. (b) The discrete-time Fourier transform $Y_k(\omega)$ of the sampled sequence according to formula (2). For simplicity, we assume that $\tau_k = 0$. Intuitively, $Y_k(\omega)$ consists of several overlapping spectral segments of $X(\omega)$. More specifically, for $\omega \in \mathcal{P}_1 = [-\sigma, -\sigma + r)$, there are three (equal to $\lceil 2\sigma/c \rceil$) spectral segments "folded" on top of each other. For $\omega \in \mathcal{P}_2 = [-\sigma + r, -\sigma + c)$, the number of overlapping spectral segments is two (equal to $\lfloor 2\sigma/c \rfloor$).

we can combine the above two cases, and simplify (2) as

$$Y_k(\omega) = \frac{\alpha_k}{T} \sum_{m=0}^{M(\omega)-1} X(\omega + mc)\, e^{-j(\omega+mc)\,\tau_k}. \tag{6}$$

Note that when $2\sigma/c$ is an integer, the sub-interval $\mathcal{P}_1$ is empty and (6) reduces to (4).

Let $\boldsymbol{Y}(\omega) \stackrel{\text{def}}{=} [Y_1(\omega), Y_2(\omega), \ldots, Y_K(\omega)]^T$ be a vector of the Fourier transforms of the K channels, and let

$$\boldsymbol{X}(\omega) \stackrel{\text{def}}{=} [X(\omega), X_2(\omega + c), \ldots, X(\omega + (M(\omega) - 1)c)]^T$$

be an $M(\omega)$-dimensional vector formed from $X(\omega)$. The equality (6) can be written in a compact matrix-vector form

$$\boldsymbol{Y}(\omega) = \boldsymbol{\Lambda}_{\boldsymbol{\tau}}(\omega)\, \boldsymbol{\Lambda}_{\boldsymbol{\alpha}}\, \boldsymbol{V}_{\boldsymbol{\tau}}\, \boldsymbol{X}(\omega), \tag{7}$$

where $\boldsymbol{\Lambda}_{\boldsymbol{\tau}}(\omega) \stackrel{\text{def}}{=} \operatorname{diag}\left\{e^{-j\omega\tau_k}\right\}$ and $\boldsymbol{\Lambda}_{\boldsymbol{\alpha}} \stackrel{\text{def}}{=} \operatorname{diag}\left\{\alpha_k/T\right\}$ are two diagonal matrices,

$$\boldsymbol{V}_{\boldsymbol{\tau}} \stackrel{\text{def}}{=} \begin{bmatrix} 1 & e^{-jc\tau_1} & \ldots & e^{-jc(M(\omega)-1)\tau_1} \\ 1 & e^{-jc\tau_2} & \ldots & e^{-jc(M(\omega)-1)\tau_2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-jc\tau_K} & \ldots & e^{-jc(M(\omega)-1)\tau_K} \end{bmatrix}$$

is a $K \times M(\omega)$ Vandermonde matrix, and $\boldsymbol{\alpha} \stackrel{\text{def}}{=} [\alpha_1, \ldots, \alpha_K]^T$ and $\boldsymbol{\tau} \stackrel{\text{def}}{=} [\tau_1, \ldots, \tau_K]^T$ denote the unknown channel gains and offsets, respectively.

Remark 1: In (7), we could have combined $\boldsymbol{\Lambda}_{\boldsymbol{\tau}}(\omega)$ and $\boldsymbol{\Lambda}_{\boldsymbol{\alpha}}$ into a single diagonal matrix. We choose to separate the two, because the former is a function of ω whereas the latter is a constant matrix. This property will be exploited in the derivation of the proposed reconstruction algorithm in Section IV.

III. Minimum Sampling Rates for Unique Reconstructions

In this section, we study the following question: For the input $x(t)$ to be uniquely determined by its samples $\{y_k[n]\}$, how many channels do we need and how often should each channel take samples? In particular, we are interested in deriving minimum sampling rates which still allow for unique reconstructions.

We start by considering the simple case where the channel gains $\boldsymbol{\alpha}$ and offsets $\boldsymbol{\tau}$ are known.

Proposition 1: Supposed that the channel gains $\boldsymbol{\alpha}$ and offsets $\boldsymbol{\tau}$ are known. The input signal $x(t)$ is uniquely specified by its samples *if and only if* the following two conditions holds:

1) The sampling rate at each channel is lower bounded by

$$\frac{1}{T} \geq \frac{\sigma}{K\pi}; \tag{8}$$

2) Among the K offsets $\{\tau_k\}$, there exist at least $\lceil 2\sigma/c \rceil$ of them such that their pairwise differences satisfy

$$(\tau_k - \tau_\ell)/T \notin \mathbb{Z}. \tag{9}$$

Proof: We note that the unicity condition is equivalent to the invertibility of the matrix-vector equation (7) linking the unknown $\boldsymbol{X}(\omega)$ to the observations $\boldsymbol{Y}(\omega)$. This, in turn, boils down to checking whether the matrix $\boldsymbol{\Lambda}_{\boldsymbol{\tau}}(\omega)\boldsymbol{\Lambda}_{\boldsymbol{\alpha}}\boldsymbol{V}_{\boldsymbol{\tau}}$ in (7) has full column rank for (almost) all ω satisfying (3).

Sufficiency: The bound (8) on sampling rates implies that $K \geq \lceil 2\sigma/c \rceil$. Consequently, the matrix $\boldsymbol{\Lambda}_{\boldsymbol{\tau}}(\omega)\boldsymbol{\Lambda}_{\boldsymbol{\alpha}}\boldsymbol{V}_{\boldsymbol{\tau}}$ always has at least as many rows as its columns. Furthermore, for any τ_k, τ_ℓ satisfying (9),

$$e^{-jc\tau_k}/e^{-jc\tau_\ell} = e^{-j2\pi(\tau_k - \tau_\ell)/T} \neq 1,$$

i.e., $e^{-jc\tau_k} \neq e^{-jc\tau_\ell}$. Since we have at least $\lceil 2\sigma/c \rceil$ offset values satisfying this pairwise condition, we can conclude that the Vandermond matrix $\boldsymbol{V}_{\boldsymbol{\tau}}$ has full column rank. It follows that $\boldsymbol{\Lambda}_{\boldsymbol{\tau}}(\omega)\boldsymbol{\Lambda}_{\boldsymbol{\alpha}}\boldsymbol{V}_{\boldsymbol{\tau}}$ has full column rank as well, since the diagonal matrices $\boldsymbol{\Lambda}_{\boldsymbol{\tau}}(\omega)$ and $\boldsymbol{\Lambda}_{\boldsymbol{\alpha}}$ are always invertible.

Necessity: If (8) does not hold, then $K < \lceil 2\sigma/c \rceil$. In this case, and for $\omega \in \mathcal{P}_1$, the matrix $\boldsymbol{\Lambda}_{\boldsymbol{\tau}}(\omega)\boldsymbol{\Lambda}_{\boldsymbol{\alpha}}\boldsymbol{V}_{\boldsymbol{\tau}}$ has more columns than rows, preventing it from having full column rank. Similarly, if we cannot find $\lceil 2\sigma/c \rceil$ offset values satisfying the pairwise condition (9), the Vandermonde matrix $\boldsymbol{V}_{\boldsymbol{\tau}}$ will not contain $\lceil 2\sigma/c \rceil$ linearly independent rows, and thus it does not have full column rank. ■

Remark 2: The requirement in (9) is intuitive. If we have two channels such that their offsets satisfy $\tau_k - \tau_\ell = NT$, for some integer N, then

$$\begin{aligned} y_k[n] &= \alpha_k\, x(nT - \tau_k) \\ &= (\alpha_k/\alpha_\ell)\, \alpha_\ell\, x(nT - NT - \tau_\ell) \\ &= (\alpha_k/\alpha_\ell)\, y_\ell[n - N]. \end{aligned}$$

In this case, the samples taken at the kth channel are merely scaled and shifted versions of those taken at the ℓth channel, and thus do not carry any additional information. The requirement in (9) is almost always satisfied in practice. In fact, if we are to draw the offset values $\{\tau_k\}$ from some continuous probability distributions (*e.g.*, uniform distributions), then (9) holds with probability one.

Next, we consider the more challenging case where the channel gains and offsets are unknown, which is the focus of this work. To be clear, it is impossible to completely determine the unknown signal $x(t)$ and parameters $\{\alpha_k\}$ and $\{\tau_k\}$ from the channel samples. In fact, for any $\alpha \neq 0$ and $\tau \neq 0$, the following signal/parameter combinations

$$\{x(t), \alpha_k, \tau_k, 1 \leq k \leq K\}$$

and

$$\left\{\alpha x(t-\tau), \frac{\alpha_k}{\alpha}, \tau_k + \tau, 1 \leq k \leq K\right\}$$

will generate the same channel outputs. To avoid this inherent ambiguity, we fix the gain and offset of the first channel to $\alpha_1 = 1$ and $\tau_1 = 0$ in our following discussions.

Let $\text{BL}(\sigma)$ represent the space of bandlimited functions with frequency support $[-\sigma, \sigma]$. Denote by

$$\mathcal{X} \stackrel{\text{def}}{=} \text{BL}(\sigma) \times \left\{\boldsymbol{\alpha} \in \mathbb{R}^K : \alpha_1 = 1\right\} \times \left\{\boldsymbol{\tau} \in \mathbb{R}^K : \tau_1 = 0\right\}$$

the combined set of all unknowns in our problem. Let

$$\mathcal{B} \stackrel{\text{def}}{=} \Big\{(x(t), \boldsymbol{\alpha}, \boldsymbol{\tau}) \in \mathcal{X} : \exists\, (\widetilde{x}(t), \widetilde{\boldsymbol{\alpha}}, \widetilde{\boldsymbol{\tau}}) \in \mathcal{X} \text{ s.t. } x(t) \neq \widetilde{x}(t) \\ \text{but } \boldsymbol{\Lambda}_{\boldsymbol{\tau}}(\omega)\, \boldsymbol{\Lambda}_{\boldsymbol{\alpha}}\, \boldsymbol{V}_{\boldsymbol{\tau}}\, \boldsymbol{X}(\omega) = \boldsymbol{\Lambda}_{\widetilde{\boldsymbol{\tau}}}(\omega)\, \boldsymbol{\Lambda}_{\widetilde{\boldsymbol{\alpha}}}\, \boldsymbol{V}_{\widetilde{\boldsymbol{\tau}}}\, \widetilde{\boldsymbol{X}}(\omega)\Big\}. \tag{10}$$

It consists of those instances of the input signal $x(t)$, channel gains $\boldsymbol{\alpha}$ and offsets $\boldsymbol{\tau}$ that cannot be uniquely determined by the channel samples. In other words, we cannot distinguish between the channel outputs $\boldsymbol{Y}(\omega)$ generated by $\{x(t), \boldsymbol{\alpha}, \boldsymbol{\tau}\}$ and those by $\{\widetilde{x}(t), \widetilde{\boldsymbol{\alpha}}, \widetilde{\boldsymbol{\tau}}\}$, for some $\widetilde{x}(t) \neq x(t)$.

Proposition 2: The "unrecoverable set" $\mathcal{B}$ in (10) is a set of measure zero *if and only if*

$$\frac{1}{T} > \frac{\sigma}{K\pi}. \tag{11}$$

We leave the proof of this result to [12].

Remark 3: Comparing the above bound to the one in (8), we see that each channel needs to be (slightly) oversampled, when the channel gains and offsets are unknown. Intuitively, we need this redundancy (by oversampling) to compensate for the parameter uncertainties.

Remark 4: Strictly speaking, we do not have the usual notion of "measure zero" in $\text{BL}(\sigma)$, as there is no analog of Lesbegue measure on infinite-dimensional Hilbert spaces. A more rigorous statement of Proposition 2 should thus be based on the concept of *prevalence* [13], which extends Lesbegue "almost everywhere" to infinite-dimensional spaces. More details about this technicality can be found in [12].

IV. An Efficient Reconstruction Algorithm

A. Subspace Condition for Parameter Estimation

In principle, given the forward model (7) of the multichannel sampling process, we can estimate the input signal and unknown parameters as follows:

$$(\widehat{x}(t), \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\tau}}) \\ = \underset{(x(t),\boldsymbol{\alpha},\boldsymbol{\tau}) \in \mathcal{X}}{\arg\min} \int_{\omega \in \mathcal{P}} \|\boldsymbol{Y}(\omega) - \boldsymbol{\Lambda}_{\boldsymbol{\tau}}(\omega)\, \boldsymbol{\Lambda}_{\boldsymbol{\alpha}}\, \boldsymbol{V}_{\boldsymbol{\tau}}\, \boldsymbol{X}(\omega)\|^2 d\omega. \tag{12}$$

Solutions to this optimization problem are maximum likelihood estimators of the unknown signal/parameters under white Gaussian noise. We observe that (12) is a separable nonlinear least-squares problem [14]: If we knew the parameters $\boldsymbol{\alpha}, \boldsymbol{\tau}$, the corresponding optimal estimate of $x(t)$—given via its equivalent frequency domain representation $\boldsymbol{X}(\omega)$—is

$$\boldsymbol{X}(\omega) = (\boldsymbol{\Lambda}_{\boldsymbol{\tau}}(\omega)\, \boldsymbol{\Lambda}_{\boldsymbol{\alpha}}\, \boldsymbol{V}_{\boldsymbol{\tau}})^{\dagger} \boldsymbol{Y}(\omega),$$

where $(\cdot)^{\dagger}$ denotes the Moore–Penrose generalized inverse of a matrix. On substituting this $\boldsymbol{X}(\omega)$ into (12), the original minimization problem becomes

$$(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\tau}}) = \underset{\boldsymbol{\alpha},\boldsymbol{\tau}}{\arg\min} \int_{\omega \in \mathcal{P}} \|(\boldsymbol{I} - (\boldsymbol{\Lambda}_{\boldsymbol{\tau}}(\omega)\, \boldsymbol{\Lambda}_{\boldsymbol{\alpha}}\, \boldsymbol{V}_{\boldsymbol{\tau}}) \\ \times (\boldsymbol{\Lambda}_{\boldsymbol{\tau}}(\omega)\, \boldsymbol{\Lambda}_{\boldsymbol{\alpha}}\, \boldsymbol{V}_{\boldsymbol{\tau}})^{\dagger}) \boldsymbol{Y}(\omega)\|^2\, d\omega, \tag{13}$$

where the unknown signal $\boldsymbol{X}(\omega)$ has been eliminated.

In practice, it is still challenging to solve the simplified problem in (13). Local descent algorithms (such as the variable projection method [14]) are not useful in this particular problem, because the structure of the parametric matrix $\boldsymbol{\Lambda}_{\boldsymbol{\tau}}(\omega)\, \boldsymbol{\Lambda}_{\boldsymbol{\alpha}}\, \boldsymbol{V}_{\boldsymbol{\tau}}$ lead to many local minima in the cost functional.

In what follows, we further simplify the parameter estimation task by separating $\boldsymbol{\alpha}$ and $\boldsymbol{\tau}$. Our derivation relies on a subspace-based rank condition originally proposed in [11].

Let $\boldsymbol{\Lambda}^*_{\boldsymbol{\tau}}(\omega) \stackrel{\text{def}}{=} \text{diag}\left\{e^{j\omega\tau_k}\right\}$ denote the complex conjugate of $\boldsymbol{\Lambda}_{\boldsymbol{\tau}}(\omega)$. Multiplying both sides of (7) by $\boldsymbol{\Lambda}^*_{\boldsymbol{\tau}}(\omega)$, we get

$$\boldsymbol{\Lambda}^*_{\boldsymbol{\tau}}(\omega) \boldsymbol{Y}(\omega) = \boldsymbol{\Lambda}_{\boldsymbol{\alpha}}\, \boldsymbol{V}_{\boldsymbol{\tau}}\, \boldsymbol{X}(\omega). \tag{14}$$

When $\omega \in \mathcal{P}_2$ as defined in (5), $\boldsymbol{\Lambda}_{\boldsymbol{\alpha}} \boldsymbol{V}_{\boldsymbol{\tau}}$ is a constant $K \times \lfloor 2\sigma/c \rfloor$ matrix. The equality (14) implies that

$$\boldsymbol{\Lambda}^*_{\boldsymbol{\tau}}(\omega) \boldsymbol{Y}(\omega) \in \mathcal{R}\left(\boldsymbol{\Lambda}_{\boldsymbol{\alpha}} \boldsymbol{V}_{\boldsymbol{\tau}}\right), \tag{15}$$

where $\mathcal{R}\left(\boldsymbol{\Lambda}_{\boldsymbol{\alpha}} \boldsymbol{V}_{\boldsymbol{\tau}}\right)$ denotes the range space of $\boldsymbol{\Lambda}_{\boldsymbol{\alpha}} \boldsymbol{V}_{\boldsymbol{\tau}}$.

More generally, let $\{\omega_1, \omega_2, \ldots, \omega_N\} \subset \mathcal{P}_2$ be a set of N frequency points, and construct a $K \times N$ data matrix

$$\boldsymbol{D}_{\boldsymbol{\tau}} \stackrel{\text{def}}{=} \left[\boldsymbol{\Lambda}^*_{\boldsymbol{\tau}}(\omega_1) \boldsymbol{Y}(\omega_1), \boldsymbol{\Lambda}^*_{\boldsymbol{\tau}}(\omega_2) \boldsymbol{Y}(\omega_2), \ldots, \boldsymbol{\Lambda}^*_{\boldsymbol{\tau}}(\omega_N) \boldsymbol{Y}(\omega_N)\right]. \tag{16}$$

It follows from (15) that

$$\mathcal{R}\left(\boldsymbol{D}_{\boldsymbol{\tau}}\right) \subseteq \mathcal{R}\left(\boldsymbol{\Lambda}_{\boldsymbol{\alpha}} \boldsymbol{V}_{\boldsymbol{\tau}}\right). \tag{17}$$

Proposition 3 (Rank Test [11]): Suppose that the sampling rate at each channel satisfies (11) and that $(\tau_k - \tau_\ell)/T \notin \mathbb{Z}$ for all $1 \leq k < \ell \leq K$. Then

$$\mathrm{rank}\ (\boldsymbol{D}_{\boldsymbol{\tau}}) \leq \lfloor 2\sigma/c \rfloor, \tag{18}$$

with equality holding when the "signal matrix"

$$\boldsymbol{S} \stackrel{\text{def}}{=} [\boldsymbol{X}(\omega_1), \boldsymbol{X}(\omega_2), \ldots, \boldsymbol{X}(\omega_N)] \tag{19}$$

is of full row rank.

Proof: It follows from (17) that

$$\mathrm{rank}\ (\boldsymbol{D}_{\boldsymbol{\tau}}) \leq \mathrm{rank}\ (\boldsymbol{\Lambda}_{\boldsymbol{\alpha}} \boldsymbol{V}_{\boldsymbol{\tau}}). \tag{20}$$

Furthermore, when the matrix in (19) is of full row rank, $\boldsymbol{D}_{\boldsymbol{\tau}}$ spans the full range space of $\boldsymbol{\Lambda}_{\boldsymbol{\alpha}} \boldsymbol{V}_{\boldsymbol{\tau}}$, in which case (20) becomes an equality. In what follows, we just need to show that $\mathrm{rank}\ (\boldsymbol{\Lambda}_{\boldsymbol{\alpha}} \boldsymbol{V}_{\boldsymbol{\tau}}) = \lfloor 2\sigma/c \rfloor$.

The condition (11) on sampling rate implies that $K > \sigma T/\pi = 2\sigma/c$. This makes $\boldsymbol{\Lambda}_{\boldsymbol{\alpha}} \boldsymbol{V}_{\boldsymbol{\tau}}$ a "tall" matrix, having more rows than columns. Meanwhile, the condition on offset values guarantees that $\boldsymbol{\Lambda}_{\boldsymbol{\alpha}} \boldsymbol{V}_{\boldsymbol{\tau}}$ has full column rank, thanks to the Vandermonde structure of $\boldsymbol{V}_{\boldsymbol{\tau}}$ (see the proof of Proposition 1 for a similar argument). ■

In practice, we can choose the number of frequency points N to be much greater than K, and hence the data matrix $\boldsymbol{D}_{\boldsymbol{\tau}}$, of size $K \times N$, is a very "wide" matrix. Proposition 3 implies that $\boldsymbol{D}_{\boldsymbol{\tau}}$ is rank deficient. This suggest that we can estimate the unknown offsets $\boldsymbol{\tau}$ by solving the following minimization problem [11]

$$\widehat{\boldsymbol{\tau}} = \arg \min_{\boldsymbol{\tau}} \sigma_K (\boldsymbol{D}_{\boldsymbol{\tau}}), \tag{21}$$

where $\sigma_K(\cdot)$ represents the Kth singular value of a matrix.

Once we obtain $\widehat{\boldsymbol{\tau}}$ from (21), we can subsequently estimate the channel gains $\boldsymbol{\alpha}$ as follows. Without loss of generality, we assume that $K = \lfloor 2\sigma/c \rfloor + 1$. (If $K > \lfloor 2\sigma/c \rfloor + 1$, we then choose to work with the first $\lfloor 2\sigma/c \rfloor + 1$ channels.) Under this setting, both $\boldsymbol{D}_{\boldsymbol{\tau}}$ and $\boldsymbol{V}_\tau$ are of co-dimension one. Consequently, we can find two vectors $\boldsymbol{n}_1, \boldsymbol{n}_2$ satisfying

$$\boldsymbol{n}_1^T \boldsymbol{D}_{\boldsymbol{\tau}} = \boldsymbol{0} \quad \text{and} \quad \boldsymbol{n}_2^T \boldsymbol{V}_{\boldsymbol{\tau}} = \boldsymbol{0},$$

which are unique up to scalar multiplications.

Recall that $\boldsymbol{D}_{\boldsymbol{\tau}} = \boldsymbol{\Lambda}_{\boldsymbol{\alpha}} \boldsymbol{V}_{\boldsymbol{\tau}} \boldsymbol{S}$, where $\boldsymbol{S}$ is the signal matrix defined in (19). It follows that

$$\boldsymbol{0} = \boldsymbol{n}_1^T \boldsymbol{D}_{\boldsymbol{\tau}} = \boldsymbol{n}_1^T \boldsymbol{\Lambda}_{\boldsymbol{\alpha}} \boldsymbol{V}_{\boldsymbol{\tau}} \boldsymbol{S}.$$

When $\boldsymbol{S}$ is of full row rank, the above equality implies that $(\boldsymbol{n}_1^T \boldsymbol{\Lambda}_{\boldsymbol{\alpha}}) \boldsymbol{V}_{\boldsymbol{\tau}} = \boldsymbol{0}$. It follows from the unicity of $\boldsymbol{n}_2$ that

$$\boldsymbol{n}_1^T \boldsymbol{\Lambda}_{\boldsymbol{\alpha}} = s\, \boldsymbol{n}_2,$$

for some unknown scalar s. The channel gains are obtained as $\alpha_k = sT n_{2,k}/n_{1,k}$, with the uncertainty about s resolved by the assumption that $\alpha_1 = 1$. We can also show [12] that $n_{1,k} \neq 0$ and thus the division operations are always feasible.

It is clear from the above discussions that estimating the channel offsets $\boldsymbol{\tau}$ is the key step in the reconstruction algorithm. Once we obtain estimates of $\boldsymbol{\tau}$, the other unknowns $\boldsymbol{\alpha}$ and $\boldsymbol{X}(\omega)$ can be obtained by standard linear inversions. In [11], the channel offsets $\boldsymbol{\tau}$ are estimated by solving (21) in an exhaustive fashion, with very high computational cost. In the following section, we propose a novel noniterative way to exploit the rank condition (18), which allows us to estimate the unknown offsets in a single step.

B. A Noniterative Algorithm for Estimating the Channel Offsets

For simplicity of exposition, we focus on a specific case where we have $K = 3$ channels in the system. The sampling rate $1/T$ at each channel is chosen so that $\lfloor 2\sigma/c \rfloor = 2$. Under these settings, we can write the matrix $\boldsymbol{D}_{\boldsymbol{\tau}}$ defined in (16) as

$$\boldsymbol{D}_{\boldsymbol{\tau}} = \begin{bmatrix} Y_1(\omega_1) & Y_1(\omega_2) & \ldots & Y_1(\omega_N) \\ Y_2(\omega_1) e^{j\omega_1 \tau_2} & Y_2(\omega_2) e^{j\omega_2 \tau_2} & \ldots & Y_2(\omega_N) e^{j\omega_N \tau_2} \\ Y_3(\omega_1) e^{j\omega_1 \tau_3} & Y_2(\omega_2) e^{j\omega_2 \tau_3} & \ldots & Y_3(\omega_N) e^{j\omega_2 \tau_3} \end{bmatrix}.$$

Since we set $\tau_1 = 0$, we have omitted the terms $\left\{e^{j\omega_n \tau_1}\right\}$ in the above expression.

To further simplify $\boldsymbol{D}_{\boldsymbol{\tau}}$, we set the frequency points $\{\omega_n\}$ to be

$$\omega_n = \triangle\omega\, n, \tag{22}$$

for some $\triangle\omega > 0$. Substituting (22) into the above formula for $\boldsymbol{D}_{\boldsymbol{\tau}}$, and writing

$$Y_{k,n} \stackrel{\text{def}}{=} Y_k(\omega_n), \quad u \stackrel{\text{def}}{=} e^{j\triangle\omega\tau_1}, \quad \text{and } v \stackrel{\text{def}}{=} e^{j\triangle\omega\tau_2},$$

we can rewrite $\boldsymbol{D}_{\boldsymbol{\tau}}$ as

$$\boldsymbol{D}_{\boldsymbol{\tau}} = \begin{bmatrix} Y_{1,1} & Y_{1,2} & \ldots & Y_{1,N} \\ Y_{2,1}\, u & Y_{2,2}\, u^2 & \ldots & Y_{2,N}\, u^N \\ Y_{3,1}\, v & Y_{3,2}\, v^2 & \ldots & Y_{3,N}\, v^N \end{bmatrix}. \tag{23}$$

The task of estimating the unknown channel offsets $\{\tau_2, \tau_3\}$ becomes that of estimating the two parameters u and v.

From Proposition 3, the above $3 \times N$ matrix is of rank $\lfloor 2\sigma/c \rfloor = 2$. It follows that the null space of $\boldsymbol{D}_{\boldsymbol{\tau}}^T$ is nontrivial, and thus there must exist some a and b such that

$$\begin{bmatrix} 1 & -a & -b \end{bmatrix} \boldsymbol{D}_{\boldsymbol{\tau}} = \boldsymbol{0}, \tag{24}$$

Combining (23) and (24) leads to the following set of equations

$$Y_{1,n} = Y_{2,n}\, a\, u^n + Y_{3,n}\, b\, v^n, \quad \text{for } 1 \leq n \leq N, \tag{25}$$

and our goal is to estimate u and v from the observations $\{Y_{1,n}, Y_{2,n}, Y_{3,n}\}_{n=1}^N$.

The above setup is reminiscent of the classical harmonic retrieval problem [15]. The challenging aspect is that, in our problem, the unknown exponential sequences au^n and bv^n are "modulated" by two sequences $Y_{2,n}$ and $Y_{3,n}$, respectively. This complication makes the classical algorithms for harmonic retrieval (such as the annihilation filter method [15]) inapplicable.

In what follows, we propose a new algorithm for estimating u and v, based on the idea of overparameterization.

Proposition 4: If the matrix $\boldsymbol{D_\tau}$ in (23) is rank deficient, then

$$(Y_{1,n+2}Y_{2,n+1}Y_{3,n})\, u - (Y_{1,n+2}Y_{2,n}Y_{3,n+1})\, v \\ - (Y_{1,n+1}Y_{2,n+2}Y_{3,n})\, u^2 + (Y_{1,n}Y_{2,n+2}Y_{3,n+1})\, u^2 v \\ + (Y_{1,n+1}Y_{2,n}Y_{3,n+2})\, v^2 - (Y_{1,n}Y_{2,n+1}Y_{3,n+2})\, uv^2 = 0, \tag{26}$$

for $1 \leq n \leq N-2$.

Proof: It follows from the rank deficiency of $\boldsymbol{D_\tau}$ that any three consecutive columns of $\boldsymbol{D_\tau}$ must be linearly dependent. In other words, we have

$$\det\left(\begin{bmatrix} Y_{1,n} & Y_{1,n+1} & Y_{1,n+2} \\ Y_{2,n}\, u^n & Y_{2,n+1}\, u^{n+1} & Y_{2,n+2}\, u^{n+2} \\ Y_{3,n}\, v^n & Y_{3,n+1}\, v^{n+1} & Y_{3,n+2}\, v^{n+2} \end{bmatrix}\right) = 0,$$

for $1 \leq n \leq N-2$. Since $|u| = |v| = 1$, we can multiply the above matrix on the left by $\mathrm{diag}\{1, u^{-n}, v^{-n}\}$ without changing its determinant. It follows that

$$\det\left(\begin{bmatrix} Y_{1,n} & Y_{1,n+1} & Y_{1,n+2} \\ Y_{2,n} & Y_{2,n+1}\, u & Y_{2,n+2}\, u^2 \\ Y_{3,n} & Y_{3,n+1}\, v & Y_{3,n+2}\, v^2 \end{bmatrix}\right) = 0. \tag{27}$$

Expressing (27) by using the Leibniz formula [16] for matrix determinants, we are done. ■

Note that (26) in Proposition 4 is a linear equality involving six unknowns u, v, u^2, u^2v, v^2 and uv^2. By varying the frequency index n from 1 to $N-2$, we can construct from (26) the following matrix equation

$$\boldsymbol{A}\begin{bmatrix} u & v & u^2 & u^2v & v^2 & uv^2 \end{bmatrix}^T = \boldsymbol{0}, \tag{28}$$

where $\boldsymbol{A}$ is a matrix of size $(N-2) \times 6$ whose rows are formed by the constant coefficients in (26).

When $N \geq 7$, we can show [12] that $\mathrm{rank}(\boldsymbol{A}) = 5$, except for certain degenerate cases. This ensures that the null space of $\boldsymbol{A}$ has dimension one, and thus the unknown parameter vector $\begin{bmatrix} u & v & u^2 & u^2v & v^2 & uv^2 \end{bmatrix}^T$ can be uniquely determined (up to a scalar) by computing the singular value decomposition (SVD) of $\boldsymbol{A}$ and picking its right singular vector living in the null space.

Remark 5: To be sure, the six unknown parameters here (u, v, u^2, u^2v, v^2 and uv^2) are redundant. From a "parsimonious" perspective, the expression in (28) should be interpreted as a system of trigonometric polynomial equations of variables u and v. However, numerically solving systems of polynomial equations is a challenging task. Existing techniques (such as the Gröbner basis method [17]) are highly sensitive to numerical precisions and noise in the data. In the proposed approach, we treat the six parameters as if they were independent. This overparameterized perspective, albeit slightly redundant, leads to a linear problem formulation, for which many mature and robust numerical procedures (*e.g.*, SVD) are available.

Remark 6: The result of Proposition 4 can be extended to arbitrary K channels [12]. In general, the rank condition in Proposition 3 leads to a system of linear equations with $K!$ unknowns. As long as $N \geq K! + 1$, we can solve these unknowns by computing the SVD of a constant matrix of size $(N-2) \times K!$. It appears daunting that the computational complexity is combinatorial with respect to K. However, for relatively small values of K (*e.g.*, $K \leq 6$)—which is often the case in many practical applications—the problem dimensions ($6! = 720$) is still well-within the capabilities of standard numerical algorithms and computers.

V. Numerical Results

In this section, we verify the performance of the proposed algorithm through several numerical simulations. In all our experiments, we assume that $K = 3$ and $1/T = \sigma/(2\pi)$. This corresponds to a sampling setup where we have three channels, each sampling at one-half of the Nyquist rate. The input signal $x(t)$ is randomly generated, with its Nyquist samples drawn from an i.i.d. Gaussian distribution. We also add white Gaussian noise to the channel samples $\{y_k[n]\}$, to emulate a wide range of signal-to-noise ratios (SNR). At each SNR value, we randomly choose the two unknown offsets τ_2, τ_3 from a uniform distribution on $[0, T]^2$. We run the proposed estimation algorithm over 5000 such realizations.

We show in Figure 3 the mean absolute error in offset estimation as a function of SNR values (ranging from 5 dB to 70 dB). An error of 1.0 corresponds to an entire sampling period T. For comparison, we also show the results obtained by the rank-based method proposed in [11], which directly solves the minimization problem (21) through an exhaustive search on a coarse grid followed by local refinements. We observe from the figure that the proposed algorithm achieves a lower estimation error than the rank-based method, and tends to be more consistent (*i.e.*, with lower standard deviations).

Figure 4 plots the success rate as a function of the SNR. We define successful estimation as those whose absolute errors are smaller than 0.001. In medium to high SNR regimes (above 40 dB), the proposed algorithm possesses superior success rates than the rank-based method. At lower SNR values, the latter becomes better, mainly due to its exhaustive search strategy. Nonetheless, the performance of the proposed method remains close.

The proposed algorithm is noniterative. Its computation steps boil down to forming a fixed data matrix (as in (28)) from the frequency values of the channel samples and computing the SVD of that matrix. Consequently, its computational complexity is much lower than that of the rank-based method in [11], which contains an exhaustive search step. We have implemented both the proposed method and the rank-based method in MATLAB. In our experiments, the former is about 1600-times faster than the latter in terms of CPU time. At lower SNR regimes (*e.g.*, 35 dB or lower), we find it beneficial to add a local refinement step (see [12]) to the proposed method. Even with the additional computational cost incurred by this local refinement stage, the proposed method is still about 30 to 40 times faster than the rank-based method, all the while delivering similar or superior estimation performances.

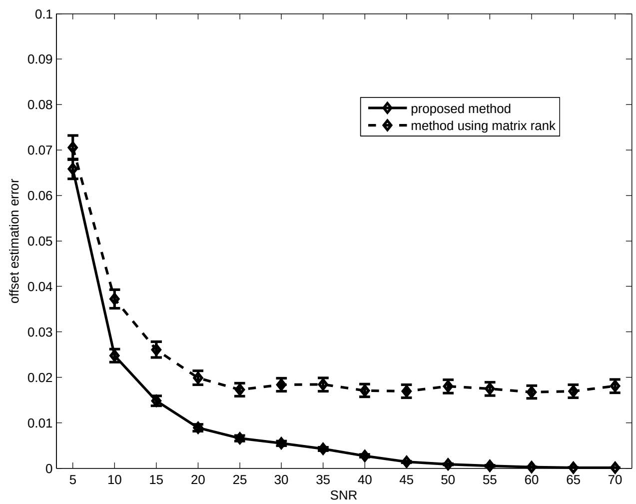

Fig. 3. Mean absolute error in offset estimation as a function of the SNR. An error of 1.0 corresponds to an entire sampling period T. Results shown in the figure are averaged over 5000 simulations with randomly selected offsets.

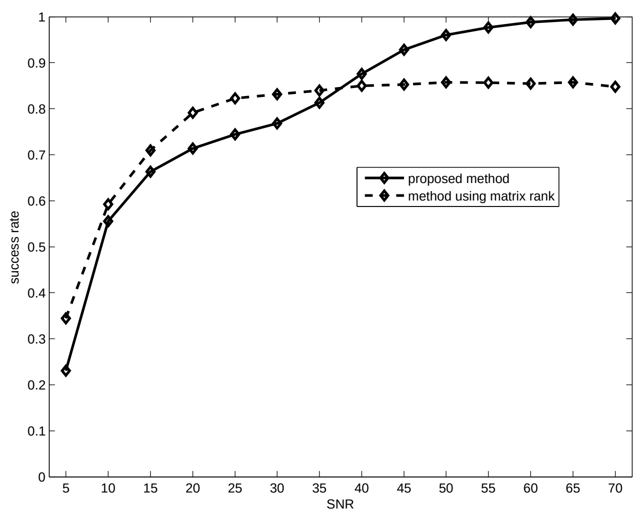

Fig. 4. Success rate as a function of the SNR. An estimation is deemed successful if the absolute error in offset estimation is smaller than 0.001.

VI. Conclusion

This paper presents a detailed treatment of a multichannel sampling problem, where the channel gains and offsets are unknown. We derived the minimum sampling rates for unique signal recovery and, as our main contribution, proposed a novel algorithm that can efficiently estimate the unknown gains and offsets. Our developments centered around ways to efficiently exploit a subspace-based rank condition. By using a linearization technique, we convert the original nonlinear problem to a system of linear equations, whose solutions lead to the unknown system parameters. The proposed algorithm has low computational complexity, and can be solved by computing the SVD of a fixed data matrix. Numerical results confirm the effectiveness, efficiency, and robustness of the proposed algorithm in the presence of noise.

References

[1] A. Papoulis, "Generalized sampling expansion," *IEEE Trans. Circuits Syst.*, vol. CAS-24, no. 11, pp. 652–654, Nov. 1977.

[2] J. L. Brown, Jr., "Multi-Channel sampling of low-pass signals," *IEEE Trans. Circuits Syst.*, vol. CAS-28, no. 2, pp. 101–106, Feb. 1981.

[3] J. L. Brown, Jr. and S. D. Cabrera, "On well-posedness of the Papoulis generalized sampling expansion," *IEEE Trans. Circuits Syst.*, vol. 38, no. 5, pp. 554–556, May 1991.

[4] M. Unser and J. Zerubia, "A generalized sampling theory without bandlimiting constraints," *IEEE Trans. Circuits Syst. II*, vol. 45, pp. 959–969, May 1998.

[5] R. S. Prendergast, B. C. Levy, and P. J. Hurst, "Reconstruction of band-limited periodic nonuniformly sampled signals through multirate filter banks," *IEEE Trans. Circuits Syst. I*, vol. 51, no. 8, pp. 1612–1622, Aug. 2004.

[6] P. Sommen and K. Janse, "On the relationship between uniform and recurrent nonuniform discrete-time sampling schemes," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5147–5156, Oct. 2008.

[7] W. C. Black and D. A. Hodges, "Time interleaved converter arrays," *IEEE J. Solid-State Circuits*, vol. 15, no. 6, pp. 1022–1029, Dec. 1980.

[8] S. Huang and B. C. Levy, "Blind calibration of timing offsets for four-channel time-interleaved ADCs," *IEEE Trans. Circuits Syst. I*, vol. 54, no. 4, pp. 863–876, Apr. 2007.

[9] J. A. McNeill, C. David, M. Coln, and R. Croughwell, "Split ADC calibration for all-digital correction of time-interleaved ADC errors," *IEEE Trans. Circuits Syst. II*, vol. 56, no. 5, pp. 344–348, May 2009.

[10] A. Hormati, O. Roy, Y. M. Lu, and M. Vetterli, "Distributed sampling of correlated signals linked by sparse filtering: Theory and applications," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1095–1109, Mar. 2010.

[11] P. Vandewalle, L. Sbaiz, J. Vandewalle, and M. Vetterli, "Super-Resolution from unregistered and totally aliased signals using subspace methods," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3687–3703, Jul. 2007.

[12] Y. M. Lu and M. Vetterli, "Multichannel sub-Nyquist sampling with unknown gains and offsets," *IEEE Trans. Signal Process.*, to be submitted.

[13] B. R. Hunt, T. Sauer, and J. A. Yorke, "Prevalence: A translational-invariant "almost every" on infinite-dimensional spaces," *Bull. Amer. Math. Soc. (N.S.)*, vol. 27, no. 2, pp. 217–238, Oct. 1992.

[14] G. H. Golub and V. Pereyra, "The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate," *SIAM J. Numer. Anal.*, vol. 10, no. 2, pp. 413–432, Apr. 1973.

[15] P. Stoica and R. L. Moses, *Introduction to spectral analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1997.

[16] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, UK: Cambridge University Press, 1985.

[17] D. Cox, J. Little, and D. O'Shea, *Ideals, Varieties, and Algorithms*, 3rd ed. Springer-Verlag, 1992.