

On the Accuracy and Resolution of Powersum-Based Sampling Methods

Julius Kusuma, *Member, IEEE*, and Vivek K Goyal, *Senior Member, IEEE*

Abstract—Recently, several sampling methods suitable for signals that are sums of Diracs have been proposed. Though they are implemented through different acquisition architectures, these methods all rely on estimating the parameters of a powersum series. We derive Cramér–Rao lower bounds (CRBs) for estimation of the powersum poles, which translate to the Dirac positions. We then demonstrate the efficacy of simple algorithms due to Prony and Cornell for low-order powersums and low oversampling relative to the rate of innovation. The simulated performance illustrates the possibility of superresolution reconstruction and robustness to correlation in the powersum sample noise.

Index Terms—Analog-to-digital conversion, Cramér–Rao bound (CRB), estimation, parametric modeling, Prony’s method.

I. INTRODUCTION

DIGITAL processing of continuous-time signals relies first and foremost on accurate data acquisition. In the classical paradigm, acquisition involves filtering a continuous-time signal and then measuring uniformly spaced samples; the samples are construed to specify a unique signal in a particular subspace of continuous-time signals. Importantly, the combination of Hilbert-space geometry and the representative signals forming a subspace makes the influence of noise, as measured by \mathcal{L}^2 error, easy to analyze [1].

The focus of this paper is on signal acquisition for certain classes of signals that do not form subspaces. Through recently developed architectures and algorithms, these signals can be acquired from a small number of samples, but the greater geometric complexity of these signal sets makes the performance when samples are subject to noise more difficult to analyze. We provide a unification of the techniques of [2]–[4], showing that they each yield a *powersum series fitting* problem. We analyze the performance limits for powersum series fitting and the performance of several algorithms. Our analysis method is adapted for real-valued and complex-roots-of-unity cases, corresponding to these different sampling schemes. In particular, this enables comparison between architectures and highlights the importance of modeling sources of noise.

Manuscript received December 05, 2007; revised September 03, 2008. First published October 31, 2008; current version published January 06, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Pierre Vanderghyest. This work was supported in part by the NEC Corporation Fund for Research in Computers and Communications and by the Texas Instruments Leadership University Consortium Program. This work was presented in part at the IEEE International Conference on Image Processing, Atlanta, GA, October 2006.

J. Kusuma is with the Schlumberger Technology Corporation, Cambridge, MA 02139 USA (e-mail: kusuma@alum.mit.edu).

V. K Goyal is with the Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: vgoyal@mit.edu).

Digital Object Identifier 10.1109/TSP.2008.2007102

As a specific instance, we are interested in acquiring real-valued signals from the set

$$\mathcal{M}_K = \left\{ x : x(t) = \sum_{k=0}^{K-1} a_k \delta(t - t_k) \right\} \quad (1)$$

where the number of components K is known.¹ Sets of this type have been used to model many naturally occurring signals [5] and in ranging and wideband communication systems [6]. A signal in \mathcal{M}_K is uniquely determined by K pairs of parameters $\{(a_k, t_k)\}_{k=0}^{K-1}$, so it can be specified in various ways by $2K$ real numbers. One could hope that $N \geq 2K$ samples of $y(t) = h(t) * x(t)$ would suffice as such a representation. Indeed, it is shown in [2]—constructively through an algorithm that recovers $\{(a_k, t_k)\}$ —that certain sampling kernels $h(t)$ do enable unique specification of $x(t)$ through samples of $y(t)$.

It is useful to separate the (approximate) acquisition of a signal from \mathcal{M}_K into two interrelated phases: *measurement* and *estimation*. In the measurement phase, analog hardware takes $x(t)$ as an input and creates certain quantized samples. As described further in Section III, several *architectures* for measurement have been proposed. These each yield a powersum series fitting problem. In the estimation phase, some *algorithm* is applied to the samples to solve the fitting problem.

An important open question is: How robustly can a signal in \mathcal{M}_K be estimated when the measurement process is subject to noise? Because of the form of \mathcal{M}_K , when the t_k s are fixed the estimation of the a_k s is a standard linear problem. The most interesting issue is thus the accuracy of estimating the t_k s. We address this question by explicitly exhibiting the Cramér–Rao bound (CRB) for the powersum series fitting problem and by comparing this bound numerically to the performance obtained with two practical algorithms.

We limit our attention to the $K = 1$ and $K = 2$ cases and hint at the infeasibility of an explicit approach for larger values of K . Note also that we are interested in the performance when the number of samples N is at or near the minimum possible ($2K$). An adequate understanding of the performance when the number of samples is large can be obtained by interpretation of results for spectral analysis [7]. As a final caveat, note that we consider Gaussian additive noise models. These are appropriate for cases in which thermal noise—rather than quantization noise, aperture uncertainty, and comparator ambiguity—is the dominant analog-to-digital conversion (ADC) impairment; this is the case for high-resolution ADC [8].

The remainder of the paper is organized as follows. We first introduce powersum series and solution for their parameters in

¹The use of a Dirac delta simplifies the discussion. It can be replaced by a known pulse $g(t)$ and then absorbed into the sampling kernel $h(t)$, yielding an effective sampling kernel $g(t) * h(t)$.

the noiseless case in Section II. Then Section III shows how powersum series arise from several sampling architectures for signals of the class (1) and other signal classes. In particular, this puts architectures from [2]–[4] into a common framework. In Section IV we turn to algorithms for fitting powersum series which have noise. We focus on algorithms that are simple, work well for small numbers of samples, and do not require initialization. CRBs for powersum series estimation problems are developed in Section V, where we give results for real-valued and complex-roots-of-unity powersums. These are applied in Section VI, which compares various architectures and algorithms, using two models for the sources of noise in the measurement architectures: continuous-time white noise and powersum white noise.

The estimation error analysis presented here appeared first in [9], and the architecture of Section III-C appeared first in [4].

II. POWERSUM SERIES

The nonlinear parameter estimation problems that we consider in this paper are all reduced to estimation problems involving a *powersum series*. We first introduce the powersum series, before we review the estimation problems that are relevant in Section III.

Definition 1 (Powersum Series): Samples $\{x_n\}_{n=0}^{N-1}$ are said to be generated by a *powersum series* of order K with *amplitudes* $\{c_k\}_{k=0}^{K-1}$ and *poles* $\{u_k\}_{k=0}^{K-1}$ when

$$x_n = \sum_{k=0}^{K-1} c_k (u_k)^n, \quad n = 0, 1, \dots, N-1. \quad (2)$$

Sequences of form (2) were first studied by G. C. M. R. de Prony in 1795 as he attempted to find the decay rates of chemical processes [10]. In de Prony’s original problem, the observations and parameters are real-valued. This is sometimes called “real exponential fitting” or “exponential analysis” in the natural sciences literature [11], [12].

de Prony’s method is based on an idea that is quite intuitive to readers of this Transactions. Suppose x_n is of the powersum form (2) for $n \geq 0$ and zero for $n < 0$. Then the z -transform of this infinite sequence is given by

$$X(z) = \sum_{k=0}^{K-1} c_k \frac{1}{1 - u_k z^{-1}}.$$

Since $X(z)$ has K poles, there is a monic *annihilating filter* b_n supported on $\{0, 1, \dots, K\}$ such that $d_n = x_n * b_n$ is zero outside of $\{0, 1, \dots, K-1\}$.² This fact can be written in matrix form as

$$\begin{bmatrix} x_K & x_{K-1} & x_{K-2} & \cdots & x_0 \\ x_{K+1} & x_K & x_{K+1} & \cdots & x_1 \\ \vdots & & & & \vdots \\ x_{2K-1} & x_{2K-2} & x_{2K-3} & \cdots & x_{K-1} \end{bmatrix} \begin{bmatrix} 1 \\ b_1 \\ b_2 \\ \vdots \\ b_K \end{bmatrix} = \begin{bmatrix} d_K \\ d_{K+1} \\ \vdots \\ d_{2K-1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (3)$$

²Monic means that $b_0 = 1$; some arbitrary normalization is needed because constant multiples of b will have the same property of annihilating x outside of $\{0, 1, \dots, K-1\}$.

where we have written K equations to have enough to solve for the K unknowns $\{b_k\}_{k=1}^K$. Looking at the matrix in (3), we see that $N = 2K$ samples of x_n are generally sufficient for recovery of the b_k s. (If the solution is not unique, the data are fit by a lower-order model.) Now factoring $1 + \sum_{k=1}^K b_k z^{-k}$ yields the u_k s because the annihilating filter satisfies

$$1 + \sum_{k=1}^K b_k z^{-k} = \prod_{k=0}^{K-1} (1 - u_k z^{-1}).$$

With the u_k s fixed, (2) describes a linear relationship between $(c_0, c_1, \dots, c_{K-1})$ and $(x_0, x_1, \dots, x_{K-1})$; thus the c_k s are easily determined.

We return to the fitting of powersum series—there in the presence of noise—in Section IV. That will be after we exhibit several sampling architectures that generate powersum series.

III. SIGNAL MODELS AND ARCHITECTURES YIELDING POWERSUM SERIES

As aforementioned, we are interested in signal estimation problems involving powersum series. The form of powersum series that arises depends on the signal model and the acquisition architecture. In this section, we consider the following three scenarios in their order of publication:

- 1) a signal that is a periodic sum of Diracs, acquired using a sinc sampling kernel and uniform sampling in time [2];
- 2) a sum of Diracs signal with a known local rate of innovation, acquired using a sampling kernel that satisfies a Strang-Fix condition and uniform sampling in time [3];
- 3) a sum of Diracs signal with a known local rate of innovation, acquired using integrators and simultaneous sampling in multiple channels [4], [9].

This is not an exhaustive review of the literature; in particular other scenarios are presented in [2]. The coverage is selected to include powersums with both real and complex poles and to facilitate a comparison between 2) and 3) in Section VI.

A. Periodic Sum of Diracs Acquired With Sinc Kernel

Consider a signal $x(t)$ that is a 1-periodic extension of $\sum_{k=0}^{K-1} a_k \delta(t - t_k)$, where $\{t_k\}_{k=0}^{K-1} \subset [0, 1)$. Because of the periodicity, $x(t)$ can be represented using Fourier series coefficients as

$$x(t) = \sum_{m=-\infty}^{\infty} X_m \exp(j2\pi mt). \quad (4)$$

For the given signal model, the Fourier series coefficients are given by

$$X_m = \sum_{k=0}^{K-1} a_k \exp(-j2\pi mt_k), \quad m \in \mathbb{Z}. \quad (5)$$

The Fourier domain representation given in (5) has infinite length, hence we say that this signal is not bandlimited. However, since (5) is a powersum series, it is possible to estimate the coefficients from $N \geq 2K$ samples of X_m . In this case, the poles of the powersum series are complex roots of unity.

Vetterli *et al.* [2] showed that Fourier series coefficient X_m can be obtained by linear processing of uniform samples of the

output of a $\text{sinc}(2Kt)$ sampling filter with input $x(t)$. Specifically, X_m for $m \in \{-M, -M+1, \dots, M\}$ are obtained from $N \geq 2M+1$ samples.

B. Aperiodic Sum of Diracs Acquired With Compactly Supported Kernel

The technique described in Section III-A is an idealized abstraction because it involves unrealizable filters and infinite periodic extension. Dragotti *et al.* [3] introduced more practical schemes that use compactly supported sampling kernels and causal sampling kernels with rational transfer functions. We concentrate here on sampling kernels $\varphi(t)$ that satisfy the Strang-Fix conditions for polynomial reproduction up to degree M for shifts by $1/N$.

Consider $x(t)$ as in (1) where $\{t_k\}_{k=0}^{K-1} \subset [0, 1)$. This is representative of having finite local rate of innovation normalized to $2K$. Let $r_{m,n}$ be coefficients for polynomial reproduction, such that

$$\sum_n r_{m,n} \varphi\left(\frac{t-n}{N}\right) = d_m(t)$$

where $d_m(t) = t^m$, for $t \in [0, 1)$ and $m = 0, 1, \dots, M$. Taking N uniform samples of $x(t) * \varphi(-t)$ in $[0, 1)$ yields

$$\begin{aligned} x_n &= \left\langle x(t), \varphi\left(\frac{t-n}{N}\right) \right\rangle \\ &= \int \left(\sum_{k=0}^{K-1} a_k \delta(t-t_k) \right) \varphi\left(\frac{t-n}{N}\right) dt \\ &= \sum_{k=0}^{K-1} a_k \varphi\left(\frac{t_k-n}{N}\right) \end{aligned}$$

for $n = 0, 1, \dots, N-1$. Then we can compute

$$\begin{aligned} y_m &= \sum_n r_{m,n} x_n \\ &= \sum_n r_{m,n} \sum_{k=0}^{K-1} a_k \varphi\left(\frac{t_k-n}{N}\right) \\ &= \sum_{k=0}^{K-1} a_k \sum_n r_{m,n} \varphi\left(\frac{t_k-n}{N}\right) \\ &= \sum_{k=0}^{K-1} a_k t_k^m \end{aligned}$$

which is a powersum series. Hence, $x(t)$ can be perfectly reconstructed from $N \geq 2K$ samples $\{x_n\}_{n=0}^{N-1}$.

C. Aperiodic Sum of Diracs Acquired With Parallel Sampling

In [4], [9] we proposed a sampling architecture that is implemented by parallel integrators. Consider the same signal model as in Section III-B. Let $x_k(t)$ be the k th integral of $x(t)$; i.e., $x_0(t) = x(t)$ and $x_{k+1}(t) = \int_0^t x_k(\tau) d\tau$. Samples are taken simultaneously at the outputs of multiple channels: $y_n = x_{n+1}(1)$, $n = 0, 1, \dots, N-1$.

To see how a powersum series is obtained, note that

$$\begin{aligned} x_1(t) &= \int_0^t x(\tau) d\tau = \int_0^t \sum_{k=0}^{K-1} a_k \delta(\tau-t_k) d\tau \\ &= \sum_{k=0}^{K-1} a_k H(t-t_k) \\ x_2(t) &= \int_0^t x_1(\tau) d\tau = \int_0^t \sum_{k=0}^{K-1} a_k H(\tau-t_k) d\tau \\ &= \sum_{k=0}^{K-1} a_k (t-t_k) H(t-t_k) \end{aligned}$$

and

$$\begin{aligned} x_3(t) &= \int_0^t x_2(\tau) d\tau \\ &= \int_0^t \sum_{k=0}^{K-1} a_k (\tau-t_k) H(\tau-t_k) d\tau \\ &= \sum_{k=0}^{K-1} \frac{1}{2} a_k (t-t_k)^2 H(t-t_k) \end{aligned}$$

etc., where $H(t)$ is the Heaviside (unit step) function. Thus $y_n = x_{n+1}(1) = (1/n!) \sum_{k=0}^{K-1} a_k (1-t_k)^n$, $n = 0, 1, \dots, N-1$.

IV. ALGORITHMS FOR POWERSUM ESTIMATION

Work on exponential fitting in signal processing has been concentrated in the areas of angle-of-arrival estimation and direction finding, often using multiple antennas or sensors. This body of work is focused on estimating the signal parameters by first estimating the signal covariance structure [7], [13]–[15], and on the case where we have large numbers of samples with multiple snapshots. Since the parameters of greatest interest are the angles of the coefficients of the powersum series—corresponding to frequencies of the series components—it is often assumed that the coefficients lie on the unit circle. Most of the publications in this area demonstrate the efficacy of their algorithms by Monte Carlo simulation and give the resulting mean-square error.

On the other hand, the papers on exponential fitting in the natural sciences often give proof of concept by using the proposed algorithms to estimate parameters in an experiment for which the correct answer is known [11], [12]. Moreover, the number of observations tends to be small. This is matched to our interest here, since we focus on signals with low local rate of innovation and sampling rates near the rate of innovation. However, where appropriate we still use Monte Carlo simulation and give the resulting mean-square error.

Throughout, we are interested in estimating the parameters of a powersum series in the presence of additive noise

$$y_n = \sum_{k=0}^{K-1} c_k (u_k)^n + w_n, \quad n = 0, 1, \dots, N-1. \quad (6)$$

We consider two classes of algorithms: algorithms based on Prony's method and those based on the Matrix Pencil method,

also known as the Rotational Invariance Property [13], [14], [16]. These two classes of algorithms are closely related. (See [17] for a quick overview of their similarities and differences.) Several algorithms give performance close to the CRBs in the presence of additive white Gaussian noise (AWGN), such as the nonlinear least-squares algorithm [18] and a regularized maximum-likelihood algorithm [17]. However, while their performances can exceed those of the Prony- and Matrix Pencil-based methods, these algorithms require very good initial conditions and perform poorly when the number of samples N is small. These algorithms are often simulated and implemented using initial values obtained from the Prony and Matrix Pencil methods. A review of algorithms such as the annihilating filter, ESPRIT and MUSIC is given in [15]. We focus on the two methods less known within the signal processing community.

A. TLS-Prony

We briefly review the algorithm proposed by Rahman and Yu [19] and analyzed by Steedly and Moses [20] called total least squares-Prony (TLS-Prony). Suppose that we are given observations y_n , $n = 0, 1, \dots, N-1$. Pick an integer $L \geq K$, recommended to be around $N/3$.

- 1) Form the Hankel matrix \mathbf{Y} of size $(N-L) \times L$ from observations, where $[\mathbf{Y}]_{i,j} = y_{j-i}$.
- 2) Compute the SVD of \mathbf{Y} and reconstruct using only the K largest singular values. Call this reconstruction $\hat{\mathbf{Y}}$, and the first column $\hat{\mathbf{y}}$.
- 3) Compute the least-squares estimate $\hat{\mathbf{b}} = (\hat{\mathbf{Y}})^\dagger \hat{\mathbf{y}}$, where $(\cdot)^\dagger$ denotes the pseudo-inverse.
- 4) Find the L roots of polynomial representation $\hat{B}(z)$, obtaining estimates \hat{u}_ℓ for $\ell = 0, 1, \dots, L-1$.
- 5) Do least-squares fitting to find amplitudes \hat{c}_ℓ for each of the L estimates.
- 6) For each of the L estimates, compute energy

$$E_\ell = \sum_n |\hat{c}_\ell (\hat{u}_\ell)^n|^2.$$

- 7) Pick K estimates with the largest energies.

B. Cornell's Algorithm

Cornell [11] proposed a procedure for finding the coefficients of a powersum series from uniformly spaced observations based on segmenting the observations and computing partial sums. He gave simple formulas for the $K = 1$ and $K = 2$ cases. Petersson and Holmström [21] gave formulas for the $K = 3$ and $K = 4$ cases. These are dramatically more complicated, and to quote the authors, even for $K = 3$ they found the formula "troublesome," both due to the complexity of the algebraic expressions and their poor performance and instability in the presence of noise. Thus, in this paper we review and utilize only the simple formulas for $K = 1$ and $K = 2$.

Suppose that we are given observations y_n , $n = 0, 1, \dots, N-1$. For convenience, let $N = 4q$ for some integer q .

For $K = 1$, the steps are given by the following.

- 1) Compute partial sums of y_n as follows:

$$S_1 = \sum_{n=0}^{N/2-1} y_n, \quad S_2 = \sum_{N/2}^{N-1} y_n.$$

- 2) Compute $W = S_2/S_1$.
- 3) Set estimate $\hat{u}_0 = W^{N/2}$.

For $K = 2$, the steps are given by the following.

- 1) Compute partial sums of y_n as follows:

$$S_1 = \sum_{n=0}^{N/4-1} y_n, \quad S_2 = \sum_{N/4}^{N/2-1} y_n$$

$$S_3 = \sum_{n=N/2}^{3N/4-1} y_n, \quad S_4 = \sum_{3N/4}^{N-1} y_n.$$

- 2) Compute

$$L_1 = \frac{(S_1 S_4 - S_2 S_3)}{(S_1 S_3 - S_2^2)}$$

$$L_2 = \frac{(S_2 S_4 - S_3^2)}{(S_1 S_3 - S_2^2)}.$$

- 3) Find the roots W_0 and W_1 of $x^2 - L_1 x + L_2 = 0$.
- 4) Set estimates $\hat{u}_0 = W_0^{N/4}$ and $\hat{u}_1 = W_1^{N/4}$.

Cornell showed that under the mild condition $\mathbf{E}[w_n] = 0$, this algorithm is a consistent estimator. Cornell's algorithm has been extended by Agha [22] in order to avoid having to take powers of real numbers, although Agha's modified algorithm gives similar performance for small sample sizes. Cornell's algorithm has also been modified to allow for nonuniform spacing of samples by Foss [23].

V. CRAMÉR-RAO LOWER BOUNDS

In the derivation of CRBs we focus on the cases $K = 1$ and $K = 2$. We treat the complex-roots-of-unity and real-valued cases separately. We write the noisy powersum series as (6). In this paper, we focus on the case where the additive noise w_n is i.i.d. zero-mean Gaussian with variance σ^2 , although we have results for additive Gaussian noise with arbitrary covariance given in [9].

The derivations of the CRBs are done via the Fisher information matrix (FIM), which is derived from the log-likelihood of the vector of parameters of interest [26]. Proofs of the theorems in this section appear in the Appendix.

A. $K = 1$, Complex Roots of Unity

This case is applicable to the sampling scheme in Section III-A. Let the signal of (6) be periodic with period T_p . The desired parameters are $\boldsymbol{\theta} = [c_0, t_0]^T$. The noiseless signal is given by

$$X_m = c_0 \exp\left(j \frac{2\pi}{T_p} t_0 m\right). \quad (7)$$

In the presence of additive noise, the signal is given by

$$\begin{aligned} y(t) &= c_0 \delta(t - t_0) + w(t) \\ Y_m &= c_0 \exp\left(j \frac{2\pi}{T_p} t_0 m\right) + W_m. \end{aligned} \quad (8)$$

Suppose that we observe N samples of Y_m , $m = 0, 1, \dots, N - 1$. Define

$$\Gamma_r = \frac{1}{N^{r+1}} \sum_{m=0}^{N-1} m^r. \quad (9)$$

Further, let $\bar{x}[m] = \Re\{X_m\}$ and $\tilde{x}[m] = \Im\{X_m\}$. In the presence of AWGN, the FIM is given by [26]

$$\begin{aligned} \mathbf{J} &= \frac{2}{\sigma^2} \left[\sum_m (\bar{\mathbf{x}}[m] \bar{\mathbf{s}}^H[m] + \tilde{\mathbf{x}}[m] \tilde{\mathbf{s}}^H[m]) \right] \\ &= \frac{2}{\sigma^2} \left(\mathbf{K} \cdot \mathbf{L} \cdot \begin{bmatrix} \Gamma_0 & 0 \\ 0 & \Gamma_2 \end{bmatrix} \cdot \mathbf{L} \cdot \mathbf{K} \right) \end{aligned} \quad (10)$$

where we have

$$\begin{aligned} \bar{\mathbf{s}}[m] &= \begin{bmatrix} \cos\left(\frac{2\pi}{T_p} t_0 m\right) \\ -\alpha_0 m \sin\left(\frac{2\pi}{T_p} t_0 m\right) \end{bmatrix} \\ \tilde{\mathbf{s}}[m] &= \begin{bmatrix} \sin\left(\frac{2\pi}{T_p} t_0 m\right) \\ \alpha_0 m \cos\left(\frac{2\pi}{T_p} t_0 m\right) \end{bmatrix} \end{aligned}$$

$\mathbf{K} = \text{diag}(\sqrt{N}, N\sqrt{N})$, $\mathbf{L} = \text{diag}(1, \alpha_0)$, and $\alpha_0 = c_0(2\pi/T_p)$. From this computation, we obtain the following theorem:

Theorem 1: Consider the noisy powersum (8) where W_n is zero-mean white Gaussian noise with variance σ^2 . Let $\alpha_0 = c_0(2\pi/T_p)$ and $\text{SNR} = \alpha_0^2/\sigma^2$. Suppose that we obtain N samples of the signal after filtering using an antialiasing filter with bandwidth π/N rad. The CRB for time estimation is given by

$$\mathbf{E}[(\hat{t}_0 - t_0)^2] \geq \frac{1}{2\text{SNR}} \frac{1}{N^3} \frac{1}{\Gamma_2}. \quad (11)$$

The bound of Theorem 1 scales as $1/N^2$, consistent with the scaling law of single-component line spectrum estimation [15]. Further, in both cases the bounds do not depend on the location of the pulse t_0 (or the angle of the pole in line spectrum estimation). In single-component line spectrum estimation, it is known that in some regimes several algorithms achieve this lower bound, such as the TLS-Prony algorithm [20]. For small sample sizes and high SNR, we will show in Section VI-A that the Prony method and the Cornell algorithm perform close to the lower bound for our problem of interest.

B. $K = 2$, Complex Roots of Unity

We now turn the $K = 2$ case with poles on the unit circle, following the technique of Dilaveroglu [24]. Let the noiseless Fourier series be given by

$$X_m = c_0 \exp\left(j \left(\frac{2\pi}{T_p}\right) t_0 m\right) + c_1 \exp\left(j \left(\frac{2\pi}{T_p}\right) t_1 m\right). \quad (12)$$

Further let $\bar{x}[m] = \Re\{X_m\}$ and $\tilde{x}[m] = \Im\{X_m\}$. The FIM is given by

$$\mathbf{J} = \frac{2}{\sigma^2} \left[\sum_m (\tilde{\mathbf{x}}[m] \tilde{\mathbf{x}}^T[m] + \bar{\mathbf{x}}[m] \bar{\mathbf{x}}^T[m]) \right]. \quad (13)$$

For convenience, let $\alpha_0 = c_0(2\pi/T_p)$, $\alpha_1 = c_1(2\pi/T_p)$

$$\bar{\mathbf{x}}[m] = \begin{bmatrix} \cos\left(\frac{2\pi}{T_p} t_0 m\right) \\ \cos\left(\frac{2\pi}{T_p} t_1 m\right) \\ -\alpha_0 \sin\left(\frac{2\pi}{T_p} t_0 m\right) \\ -\alpha_1 \sin\left(\frac{2\pi}{T_p} t_1 m\right) \end{bmatrix}$$

and

$$\tilde{\mathbf{x}}[m] = \begin{bmatrix} \sin\left(\frac{2\pi}{T_p} t_0 m\right) \\ \sin\left(\frac{2\pi}{T_p} t_1 m\right) \\ \alpha_0 \cos\left(\frac{2\pi}{T_p} t_0 m\right) \\ \alpha_1 \cos\left(\frac{2\pi}{T_p} t_1 m\right) \end{bmatrix}.$$

We obtain the following theorem:

Theorem 2: Let $\text{SNR}_k = \alpha_k^2/\sigma^2$, $k = 1, 2$. Suppose that we obtain N samples of the signal after filtering using an antialiasing filter with bandwidth π/N rad. For convenience, let $\delta t = |t_1 - t_0|$

$$\begin{aligned} \Gamma_r &= \frac{1}{N^{r+1}} \sum_{m=0}^{N-1} m^r \\ C_r &= \frac{1}{N^{r+1}} \sum_{m=0}^{N-1} m^r \cos\left(\frac{2\pi}{T_p} \delta t \cdot m\right) \quad \text{and} \\ S_r &= \frac{1}{N^{r+1}} \sum_{m=0}^{N-1} m^r \sin\left(\frac{2\pi}{T_p} \delta t \cdot m\right). \end{aligned}$$

Further, let

$$\mathbf{M} = \begin{bmatrix} \Gamma_2(\Gamma_0^2 - C_0^2) - S_1^2 \Gamma_0 & C_2(\Gamma_0^2 - C_0^2) - S_1^2 C_0 \\ C_2(\Gamma_0^2 - C_0^2) - S_1^2 C_0 & \Gamma_2(\Gamma_0^2 - C_0^2) - S_1^2 \Gamma_0 \end{bmatrix}$$

and

$$\mathbf{P} = \frac{1}{\Gamma_0^2 - C_0^2} \mathbf{M}.$$

Then the CRB is given by

$$\mathbf{E}[(\hat{t}_k - t_k)^2] \geq \frac{1}{\text{SNR}_k} \frac{1}{N^3} [\mathbf{P}]_{k,k}. \quad (14)$$

The bound of Theorem 2 scales as $1/N^2$, consistent with previously known results in line spectrum estimation and with Theorem 1. Further, the formula obtained in (14) is similar to that of Dilaveroglu [24, Theorem 2].

C. $K = 1$, Real-Valued Case

This case is applicable to the sampling schemes of Section III-B (where $u_k = t_k$) and Section III-C (where $u_k = 1 - t_k$). Consider the simplest case where w_n is white Gaussian with variance σ^2 . Define the finite summation

$$G_r(x) = \frac{1}{N^{r+1}} \sum_{n=0}^{N-1} n^r x^n. \quad (15)$$

We can write the FIM as

$$\mathbf{J} = \frac{2}{\sigma^2} \left(\begin{bmatrix} \sqrt{N} & 0 \\ 0 & N\sqrt{N} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \left(\frac{c_0}{u_0}\right) \end{bmatrix} \begin{bmatrix} G_0(u_0^2) & G_1(u_0^2) \\ G_1(u_0^2) & G_2(u_0^2) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \left(\frac{c_0}{u_0}\right) \end{bmatrix} \begin{bmatrix} \sqrt{N} & 0 \\ 0 & N\sqrt{N} \end{bmatrix} \right).$$

We are interested in finding the CRB for u_0 , which is the last entry of the inverse of the FIM. This can be obtained by using direct matrix inversion.

Theorem 3: Let a noisy real-valued powersum be given by $y_n = c_0 u_0^n + w_n$, for $n = 0, 1, \dots, N-1$, where w_n is zero-mean white Gaussian noise with variance σ^2 . Then the CRB for estimation of u_0 from $\{y_n\}_{n=0}^{N-1}$ is given by

$$\mathbf{E}[(\hat{u}_0 - u_0)^2] \geq \frac{1}{\text{SNR}} \frac{u_0^2}{N^3} \left(\frac{G_0(u_0^2)}{G_0(u_0^2)G_2(u_0^2) - G_1(u_0^2)G_1(u_0^2)} \right) \quad (16)$$

where $\text{SNR} = c_0^2/u_0^2$.

The bound of (16) scales as $1/N^2$, consistent with the behavior of frequency estimates in line spectrum estimation [15].

Suppose for now that the signal of interest consists of complex-valued poles, not necessarily complex roots of unity. This case was studied by Steedly and Moses in [20]. The magnitude of the poles in that case corresponds to the damping coefficient of the signal. They showed that the CRB for the estimation of this parameter is minimized around the unit circle. By contrast, (16) is not minimized by $u_0 = 1$. We will revisit this comparison later in Section VI-A.

When the poles are complex-valued, the lack of knowledge of the exact pole angle leads to large errors in the estimate of the pole magnitude: a small error in the phase estimate of the pole will be amplified by the magnitude of the pole, as shown in Fig. 2 of [20]. Hence, in the complex case, the variance in a pole magnitude estimate is best near unit magnitude and becomes worse as the true pole magnitude increases. In our case, the poles have positive real values. The variance in the pole estimate decreases as the pole magnitude is increased, as there is no phase ambiguity.

D. $K = 2$, Real-Valued Case

Finally, we examine the case with $K = 2$ and real-valued poles. Let $\boldsymbol{\theta} = [c_0, c_1, t_0, t_1]^T$ be the vector of unknown parameters. We wish to derive the CRB for the estimation of t_0 and t_1 in terms of $\delta t = t_1 - t_0$ when the observations are subjected to AWGN.

Consider two-term noisy powersum $y_n = c_0(u_0)^n + c_1(u_1)^n + w_n$, where w_n is white Gaussian noise. For convenience, let

$$\mathbf{A} = \begin{bmatrix} G_0(u_0^2) & G_0(u_0 u_1) \\ G_0(u_0 u_1) & G_0(u_1^2) \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} G_1(u_0^2) & G_1(u_0 u_1) \\ G_1(u_0 u_1) & G_1(u_1^2) \end{bmatrix}$$

$$\mathbf{C} = \begin{bmatrix} G_2(u_0^2) & G_2(u_0 u_1) \\ G_2(u_0 u_1) & G_2(u_1^2) \end{bmatrix}$$

where $G_r(x)$ is as given in (15). Then the CRB can be found via the Schur complement, yielding the following theorem.

Theorem 4: Consider the estimation of $\{u_0, u_1\}$ from N observations of a two-term noisy powersum subject to AWGN with variance σ^2 . Let $\text{SNR}_k = c_k^2/\sigma^2$, $k = 1, 2$. Then the CRB is given by

$$\mathbf{E}[(\hat{t}_k - t_k)^2] \geq \frac{u_0^2}{\text{SNR}_k \cdot N^3} [(\mathbf{C} - \mathbf{B}^T \mathbf{A}^{-1} \mathbf{B})^{-1}]_{k,k}$$

where \mathbf{A} , \mathbf{B} , \mathbf{C} are as given above.

VI. PERFORMANCE EVALUATION

In this section, we first compare the performance of the schemes of Section IV against the CRBs of Section V, suitable for the cases where the powersum series is subjected to AWGN.

Then we consider the case of signal parameter estimation in the presence of continuous-time AWGN, where the different sampling schemes yield different noise structures in the powersum. We compare the powersum-based sampling schemes with the conventional method of applying an antialiasing filter, taking uniform samples, and estimating the signal pulse delay by finding the maxima of the cross correlation.

A. Powersum AWGN

It is known from the line spectrum estimation literature that both the Prony method and rotational invariance algorithms work well in the presence of AWGN when powersum poles are complex roots of unity, and that both algorithms have a superresolution property. Further, the performances of the algorithms are independent of the actual values of the powersum poles. We show this in Fig. 1 for the estimation of one Dirac, and in Fig. 2 for the case of two Diracs. We compare the two algorithms with the derived CRB from Theorem 1 and Theorem 2. In this set of simulations, we set the period of the signal to be $T_p = 1$. The results of the Cornell algorithm are not shown as they are similar to the results of the TLS-Prony algorithm.

The real-valued case is less known. From Fig. 3 we see that the performance depends on the actual value of the powersum poles. The TLS-Prony algorithm outperforms the Cornell algorithm, except when the poles are small and the number of samples is very small, e.g., $N = 2$.

Suppose for now that the signal of interest consists of complex-valued poles, not necessarily complex roots of unity. The results shown in Fig. 3 are very different from those of Steedly and Moses in [20], where the poles are complex-valued but are not necessarily roots of unity. The magnitude of the poles in that case corresponds to the damping coefficient of the signal. They

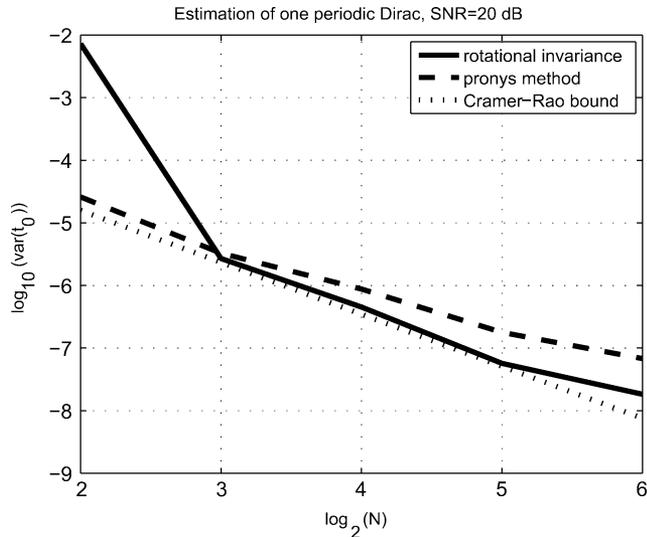


Fig. 1. Comparison between CRB and performance of the Cornell and TLS algorithms in estimating a complex roots of unity powersum with one component in the presence of powersum AWGN. The SNR is 20 dB.

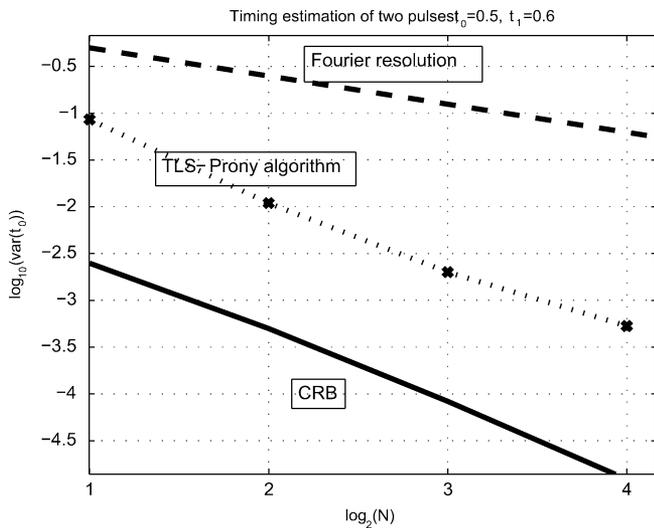


Fig. 2. Timing estimation result for the case of two signal components using the FRI in-time sampling system. The SNR is 20 dB, and we compare the Fourier resolution, the CRB, and the performance of the TLS-Prony algorithm.

showed that the CRB for the estimation of this parameter is minimized when the poles are on the unit circle. By contrast, neither the bound (16) nor the MSE performance of the algorithms in Fig. 3 is minimized by $u_0 = 1$. This illustrates that translating the results from the complex-valued case to the real-valued case is not straightforward and can be misleading.

Now we examine the superresolution property of the proposed multichannel sampling method in Fig. 4. Smith proposed in [25] that the minimum requirement to resolve two signals is that

$$\text{RMS of source separation} \leq \text{source separation.} \quad (17)$$

The *statistical resolution limit* is then defined as the source separation at which (17) is achieved with equality. Consider a signal with two components: $y_n = c_0(u_0)^n + c_1(u_1)^n + w_n$. Let the

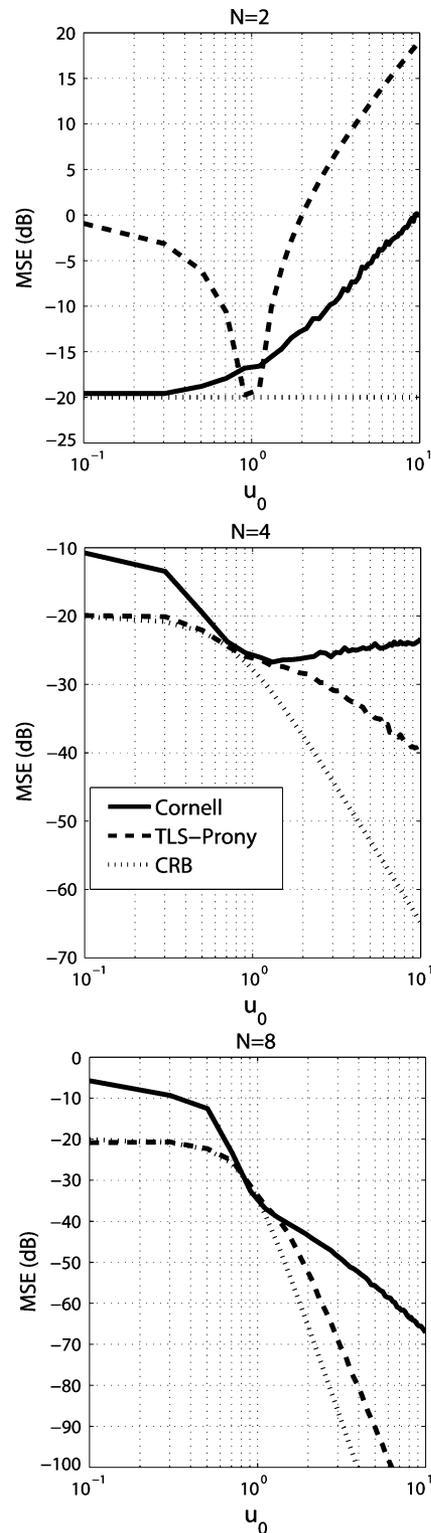


Fig. 3. Comparison between CRB and performance of the Cornell and TLS algorithms in estimating a real-valued powersum with one component in the presence of powersum AWGN. The SNR is 20 dB.

desired parameters be $\theta = (u_0, u_1)^T$. We are interested in how the estimate of θ depends on $\delta u = |u_0 - u_1|$.

It can be seen that in some cases, the performance of the proposed system exceeds the resolution limit of the classical system. The performance depends on the actual locations of the

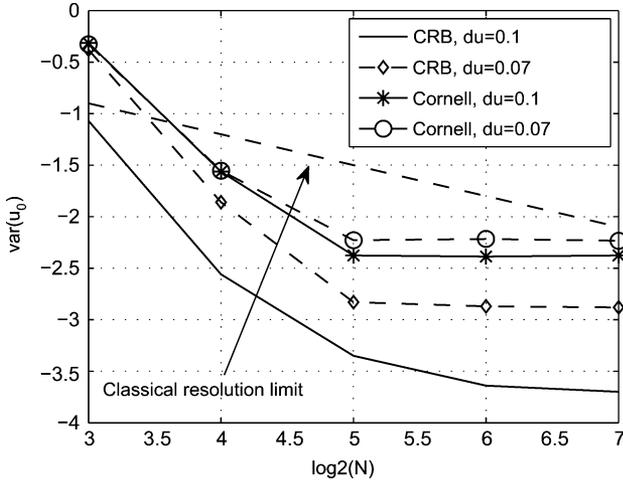


Fig. 4. Comparison between CRB of the proposed system and the Cornell algorithm in the presence of AWGN, in terms of number of samples. In this case the poles are near unity, $u_0 = 0.8$, and we consider $\delta u = u_1 - u_0 = 0.1, 0.07$. The simulation is done at SNR = 20 dB. As the performance of the TLS algorithm is very poor, we omit this from presentation.

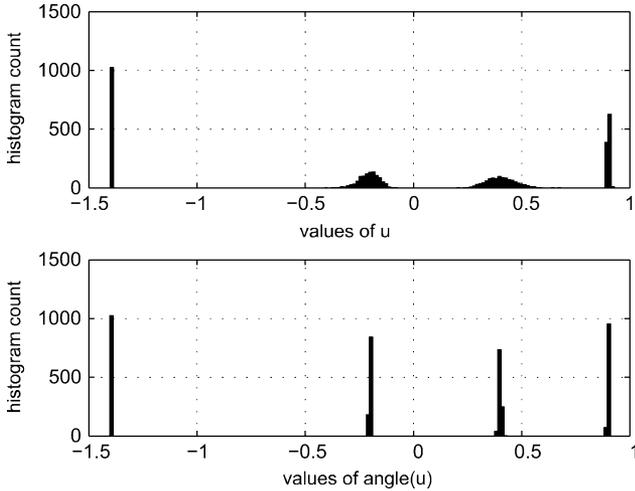


Fig. 5. Simulation results for $K = 4$ for both real-valued and complex-roots of polynomial cases. The estimation is done using the TLS-Prony algorithm. The poles all have amplitude 1. The values of the real-valued poles, and the angles of the complex poles in radians, are $-0.2, 0.9, -1.4, 0.4$. The number of samples is 16 and the SNR is at 40 dB. The histogram bin widths are 0.02, and we show the results of 1024 trials.

poles. The Cornell algorithm for $K = 2$ shows performance that is far superior to that of the TLS-based algorithm. We show the mean-square error result from the Cornell algorithm in Fig. 4 and omit the results from the TLS-based algorithm.

We also demonstrate that the algorithms considered perform well for larger values of K . We show the case where $K = 4$ and SNR is at 40 dB in Fig. 5. As expected, the performance of the TLS-Prony algorithm is worse for the poles that are closer together, and in the real-valued case for poles with smaller magnitudes.

B. Continuous-Time AWGN

The problem of delay estimation in the presence of AWGN from uniformly spaced samples is a well-known estimation

problem [26]. Let the energy of the signal be \mathcal{E}_s . Given N samples of a signal with bandwidth B , it is known that the optimal estimate is the one that maximizes the cross-correlation, and its performance is bounded by

$$\text{var}(\hat{\tau}) \geq \frac{1}{\text{SNR} \cdot F^2} \tag{18}$$

where

$$F^2 = \frac{\int s'(t)^2 dt}{\int s^2(t) dt}, \quad \text{SNR} = \frac{\mathcal{E}_s}{\left(\frac{N_0}{2}\right)}.$$

The sampling rate is $\Delta = 1/(2B)$ and $\sigma^2 = N_0 B$. In this case we must choose $s(t)$ the lowpass sampling filter to be commensurate to our desired sampling rate. When the original pulse is a Dirac, it is well-known that the resulting mean square-error decays as the square of the sampling rate. Finally, by brute-force search of the cross-correlation peak, it is known that the bound of (18) is achievable.

1) *Vetterli-Marziliano-Blu*: For this case, in Section III-A we have derived that the operation of lowpass filter—sample—Discrete Fourier Transform is equivalent to projection of the input signal into an orthonormal basis. Hence, white continuous-time AWGN becomes AWGN powersum noise, which we considered in Fig. 1 for the estimation of one Dirac and in Fig. 2 for the case of two Diracs.

2) *Dragotti-Vetterli-Blu*: In Sections III-B and III-C we saw that the sampling scheme of Dragotti *et al.* is equivalent to the multichannel sampling scheme except for the sampling kernels used. The span of the union of kernels of the former is larger than that of the kernels of the latter. More importantly, the extraneous span falls outside the interval where the desired signal is located [9].

The performance of the Dragotti scheme in the presence of powersum AWGN is identical to that of the multichannel scheme, which we consider in Section VI-B-3).

Now consider the noise characterization when the noise in the system arises from continuous-time AWGN. In this case, w_n will be correlated. When the sampling kernel is a first-order B-spline, the covariance matrix has a tridiagonal form. The diagonal entries are given by

$$\mathbf{E}[w_n w_m] = \frac{N_0}{2} \frac{1}{N^2} 2 \int_{t=0}^1 t \cdot dt = \frac{N_0}{2} \frac{1}{N^2}$$

and the off-diagonal entries are given by

$$\mathbf{E}[w_n w_m] = \frac{N_0}{2} \frac{1}{N^2} \int_{t=0}^1 (1-t)t \cdot dt = \frac{N_0}{2} \frac{1}{N^2} \frac{1}{6}.$$

Simulation results are shown in Fig. 6. In this simulation, we show the effect of different numbers of samples N . The kernel used is a simple first-order B-spline, which can reconstruct t^0 and t^1 within the interval of interest. For the scheme of Dragotti, we used a B-spline of order 1 as the sampling kernel. In this comparison we plot the estimate of $(1 - \hat{t}_0)$ from the multichannel scheme versus the estimates of \hat{t}_0 of the Dragotti scheme for consistency. Clearly, the performance of the Dragotti scheme is strictly worse than that of the multichannel scheme, due to the difference in the footprints of the sampling kernels. In the Dragotti scheme, the width of the B-spline is

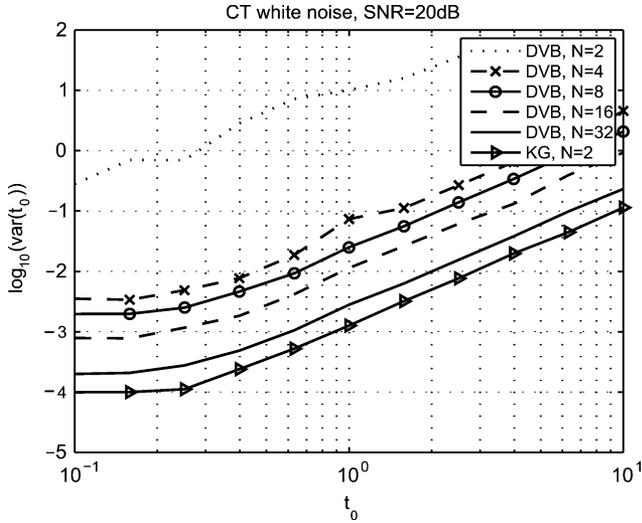


Fig. 6. Performance results for estimation of one Dirac using a first-order B-spline. The system is implemented using the simple Cornell algorithm. The plot shows different number of samples N , but the reconstruction first forms a length-2 powersum series. The AWGN is added in the continuous-time domain, with spectral density $N_0 = 0.1$. DVB refers to the Dragotti-Vetterli-Blu architecture and KG refers to the architecture proposed by the authors.

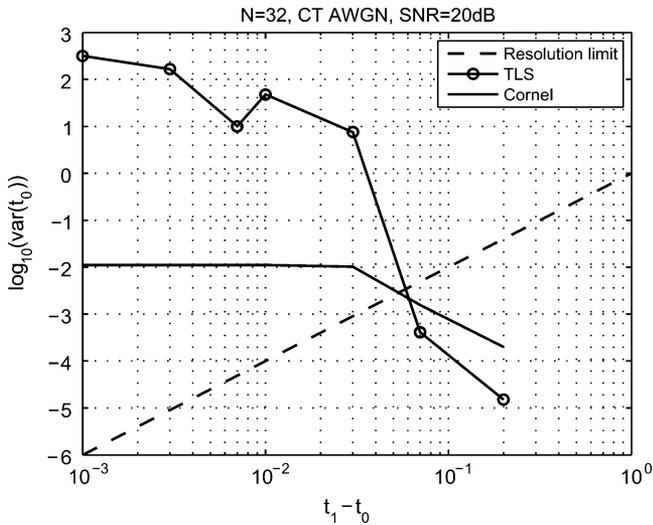


Fig. 7. Performance results for estimation of two Diracs using a first-order B-spline. The reconstruction first forms a length-4 powersum series. The AWGN is added in the continuous-time domain, with spectral density $N_0 = 0.1$.

scaled inversely to the number of samples to be taken. Hence, as the number of samples and the sampling rate N grows, the extraneous support of the kernels become smaller and approach that of the multichannel sampling scheme.

For the discrimination of two Diracs, we show the results in Fig. 7. From the figure we can see that in some regime the RMS error of the estimate is below the spacing of the two Diracs, and hence the system under consideration has a superresolution property.

3) *Kusuma-Goyal*: We finally come to the case where the noise is induced in the continuous-time domain. We focus on the case of continuous-time AWGN. Due to the structure of the multichannel sampling scheme, the sample domain noise will

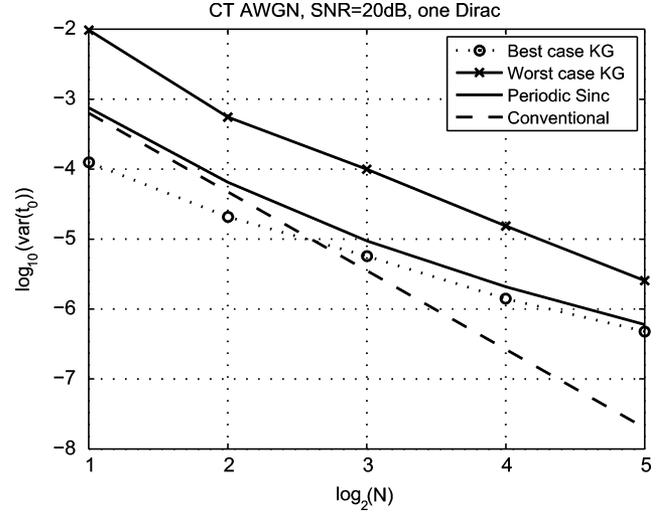


Fig. 8. Performance comparison for single-Dirac estimation, located within $[0, T)$, for the same numbers of samples. KG refers to the architecture proposed by the authors.

be correlated additive Gaussian noise. We derive the covariance structure in the following.

Let $w(t)$ be white Gaussian noise with spectral density $N_0/2$, and let the continuous-time signal be:

$$y(t) = c_0\delta(t - t_0) + w(t).$$

Following Section III-C, let the sampled signal be:

$$y_\ell = c_0(1 - t_0)^\ell + w_\ell \quad (19)$$

where w_ℓ is the additive noise term. The covariance of the noise term w_ℓ can be written as

$$\begin{aligned} \mathbf{E}[w_\ell w_k] &= \int_0^T (t^\ell t^k) \frac{N_0}{2} dt \\ &= \frac{N_0}{2} \frac{1}{k + \ell + 1} T^{k+\ell+1}. \end{aligned} \quad (20)$$

C. Comparison of the Sampling Schemes Using Continuous-Time AWGN

Although the scheme of Section III-A is suitable for a periodic signal, it is possible to apply this scheme to an aperiodic signal by applying a lowpass filter, taking samples uniformly within the time interval of interest, and computing the Discrete Fourier Transform instead of the Fourier series coefficients. By this method, we can compare the three architectures together as applied to an aperiodic signal. In the previous, we have compared the multichannel scheme and the Dragotti scheme and showed that the former is strictly better than the other in the presence of continuous-time AWGN. Further, when white sampling noise is present, the Dragotti scheme suffers from noise amplification.

Using the same continuous-time AWGN model we compare the periodic sinc scheme, the multichannel scheme, and the conventional scheme based on cross-correlation in Fig. 8. Since the

performance of the multichannel scheme is dependent on the actual location of the pulse, we show the mean-square error of the best-case and worst-case parameters when the pulse is located in $t \in [0, 1)$. The conventional scheme gives the best result for the estimation of a single pulse, but it requires either a brute-force search or a gradient search to find the peak of the correlation.

VII. CLOSING REMARKS

We examined several sampling architectures that are based on estimating the parameters of a powersum series. We introduced less-known algorithms that are suitable for small sample sets, do not require initialization, and give superresolution properties. We derived CRBs for when the powersum series is subjected to additive white Gaussian noise. For cases where the number of components is $K = 1$ or $K = 2$, we showed that the proposed algorithms work well even when the noise in the powersum is correlated.

In the real-valued case, although the TLS-Prony method gives superior performance for estimating single or well-spaced Diracs when $N > 2$, the Cornell algorithm is better for separating two closely spaced Diracs. This is also true for continuous-time white noise [9]. Further, the bounds and performances depend on the true values of the parameters.

In the complex-valued case, TLS-Prony and Cornell algorithms give nearly identical performance. For separating two closely spaced Diracs, both algorithms again give very similar results. Unlike the real-valued cases, the performance of the system does not depend on the true values of the parameters.

We used a continuous-time white Gaussian noise model to compare the three measurement architectures considered. For the estimation of a single Dirac, the conventional scheme of using a lowpass filter and using a correlation gives the best mean-square error performance. However, estimating multiple Diracs requires a multi-dimensional peak finding algorithm. By contrast, the proposed parametric schemes can give simultaneous solutions. Further, the parametric sampling schemes have a superresolution property.

Finally, we showed that the performance of the multi-channel scheme is strictly better than that of the scheme based on Strang-Fix kernels. It also compares favorably with the standard method and the scheme for periodic Diracs via periodic approximation.

We have also considered several hardware-centric noise models that depend on the topology of the system in [9]. We demonstrated that the systems and algorithms proposed work well even in the presence of correlation in the noise term of the powersum series. Some of the CRBs derived in this paper can also be extended to white Gaussian noise with arbitrary covariance, which is suitable for these hardware-centric models.

APPENDIX

In this appendix we derive the performance limits of Section V. We are interested in estimating parameters $\{u_k\}$ and $\{c_k\}$ from observations of the *noisy powersum*

$$y_n = \sum_{k=1}^{K-1} c_k (u_k)^n + w_n, \quad n = 0, 1, \dots, N-1. \quad (21)$$

The additive noise w_n has covariance Σ . In vector notation, let $\mathbf{c} = [c_0, c_1, \dots, c_{K-1}]^T$ and $\mathbf{u} = [u_0, u_1, \dots, u_{K-1}]^T$. The rows of the Vandermonde matrix \mathbf{U} are defined to contain scalar powers of \mathbf{u}^T . Then we can write (21) in vector notation as:

$$\mathbf{y} = \mathbf{U}\mathbf{c} + \mathbf{w}.$$

Suppose that additive noise \mathbf{w} is zero-mean Gaussian with covariance matrix Σ . Then the likelihood is:

$$L = \frac{1}{|\Sigma|^{1/2}} \exp\left(-(\mathbf{y} - \mathbf{U}\mathbf{c})^H \Sigma^{-1} (\mathbf{y} - \mathbf{U}\mathbf{c})\right) \quad (22)$$

where \mathbf{U} is a Vandermonde matrix containing powers of the poles $\{u_k\}$. While it is possible to derive a CRB for additive Gaussian noise with arbitrary covariance matrix Σ (see [9]), in this paper we focus on the white noise case.

When the noise is white we obtain a simpler expression for the FIM (for example see [26]). For convenience, let $\boldsymbol{\theta}$ be the vector of parameters and let $x[n; \boldsymbol{\theta}]$ be the noiseless signal given by $\boldsymbol{\theta}$. In this case,

$$-\ln L = \text{constant} + \frac{1}{2\sigma^2} \sum_n (y_n - x[n; \boldsymbol{\theta}])^2.$$

The partial derivative is particularly simple:

$$-\frac{\partial}{\partial \boldsymbol{\theta}} \ln L = \frac{2}{2\sigma^2} \sum_n (y_n - x[n; \boldsymbol{\theta}]) \frac{\partial}{\partial \boldsymbol{\theta}} x[n; \boldsymbol{\theta}].$$

Then,

$$\begin{aligned} \mathbf{J} &= \frac{1}{\sigma^2} \left[\left(\frac{\partial}{\partial \boldsymbol{\theta}} \ln L \right)^H \left(\frac{\partial}{\partial \boldsymbol{\theta}} \ln L \right) \right] \\ &= \frac{1}{\sigma^2} \left[\sum_n \left(\frac{\partial}{\partial \boldsymbol{\theta}} x[n; \boldsymbol{\theta}] \right)^H \left(\frac{\partial}{\partial \boldsymbol{\theta}} x[n; \boldsymbol{\theta}] \right) \right]. \end{aligned}$$

Let $\mathbf{p}[n] = [(u_0)^n, \dots, (u_{K-1})^n, c_0 \cdot n \cdot (u_0)^{n-1}, \dots, c_{K-1} \cdot n \cdot (u_{K-1})^{n-1}]^T$. Then we can write the FIM compactly as

$$\mathbf{J} = \frac{1}{\sigma^2} \left[\sum_n \mathbf{p}^T[n] \mathbf{p}[n] \right]. \quad (23)$$

Complex Poles on the Unit Circle

In this section, we examine the case when the poles of the powersum series are complex roots of unity. This is suitable for the sampling scheme of Vetterli, Marziliano, and Blu, which we reviewed in Section III-A.

Single-Component Case: We prove Theorem 1. Recall the FIM from (10). The inverse of the FIM is given by

$$\mathbf{J}^{-1} = \frac{\sigma^2}{2} \begin{bmatrix} \frac{1}{N} \frac{1}{\Gamma_0} & 0 \\ 0 & \frac{1}{N^3} \frac{1}{\alpha_0^2 \Gamma_2} \end{bmatrix}. \quad (24)$$

Since the desired estimation bound is for u_0 , we have obtained Theorem 1 from the bottom right corner of \mathbf{J}^{-1} .

Resolution of FRI Method: Now we consider Theorem 2. For convenience define

$$\begin{aligned} \mathbf{K} &= \text{diag}(\sqrt{N}, \sqrt{N}, N\sqrt{N}, N\sqrt{N}) \\ \mathbf{L} &= \text{diag}(1, 1, \alpha_0, \alpha_1). \end{aligned} \quad (25)$$

We segment the FIM as follows:

$$\mathbf{J} = \frac{2}{\sigma^2} \left(\mathbf{K} \cdot \mathbf{L} \cdot \begin{bmatrix} \mathbf{E} & \mathbf{W} \\ \mathbf{W}^T & \mathbf{Q} \end{bmatrix} \cdot \mathbf{L} \cdot \mathbf{K} \right). \quad (26)$$

After some algebra, we obtain

$$\mathbf{E} = \begin{bmatrix} \Gamma_0 & C_0 \\ C_0 & \Gamma_0 \end{bmatrix}, \quad \mathbf{W} = \begin{bmatrix} 0 & -S_1 \\ S_1 & 0 \end{bmatrix}$$

$$\mathbf{Q} = \begin{bmatrix} \Gamma_2 & C_2 \\ C_2 & \Gamma_2 \end{bmatrix}.$$

The CRB is found by computing the inverse of the FIM:

$$\mathbf{J}^{-1} = \frac{2}{\sigma^2} \left(\mathbf{K}^{-1} \cdot \mathbf{L}^{-1} \cdot \begin{bmatrix} \mathbf{E} & \mathbf{W} \\ \mathbf{W}^T & \mathbf{Q} \end{bmatrix}^{-1} \cdot \mathbf{L}^{-1} \cdot \mathbf{K}^{-1} \right). \quad (27)$$

We are interested in the bound on the estimates of t_0 and t_1 , which we obtain via the inverse of the Schur complement of \mathbf{c} in \mathbf{J} :

$$\mathbf{P} = \mathbf{Q} - \mathbf{W}^T \mathbf{E}^{-1} \mathbf{W}.$$

$$= \begin{bmatrix} \Gamma_2 & C_2 \\ C_2 & \Gamma_2 \end{bmatrix} - \frac{1}{\Gamma_0^2 - C_0^2} \begin{bmatrix} S_1^2 \Gamma_0 & S_1^2 C_0 \\ S_1^2 C_0 & S_1^2 \Gamma_0 \end{bmatrix}.$$

$$\mathbf{M} = \begin{bmatrix} \Gamma_2(\Gamma_0^2 - C_0^2) - S_1^2 \Gamma_0 & C_2(\Gamma_0^2 - C_0^2) - S_1^2 C_0 \\ C_2(\Gamma_0^2 - C_0^2) - S_1^2 C_0 & \Gamma_2(\Gamma_0^2 - C_0^2) - S_1^2 \Gamma_0 \end{bmatrix}.$$

$$\mathbf{P} = \frac{1}{\Gamma_0^2 - C_0^2} \mathbf{M}.$$

$$\mathbf{P}^{-1} = (\Gamma_0^2 - C_0^2) \mathbf{M}^{-1}.$$

Finally, we define $\text{SNR}_k = 2\alpha_k^2/\sigma^2$ and obtain

$$\mathbb{E}[(\hat{t}_k - t_k)^2] \geq \frac{1}{\text{SNR}_k} \frac{1}{N^3} [\mathbf{P}]_{k,k}$$

proving Theorem 2.

Real-Valued Poles

We now examine the case when the poles are real-valued. This is suitable for the sampling schemes of the Dragotti, Vetterli, and Blu—reviewed in Section III-B—and that of the authors, reviewed in Section III-C.

To examine the resolution limit, following Dilaveroglu [24] we wish to derive the performance for $K = 2$ in terms of $\delta u = u_1 - u_2$. However, Dilaveroglu only considered the undamped line spectra case, which contains complex exponentials on the unit circle, say $\exp(j\omega_1)$ and $\exp(j\omega_2)$. He used trigonometric identities to decompose the parameters into the desired form. In his case, the final result gives a CRB for ω_1 and ω_2 that depends on $\delta\omega = \omega_1 - \omega_2$, but independent of absolute terms ω_1 and ω_2 . By contrast, we do not obtain such convenient decompositions and have to rely on numerical evaluation.

Single-Pole Case: We now use the result to consider the simplest case when there is only one pole in the signal, as in Theorem 3. First consider the case when w_n is white Gaussian with variance σ^2 .

We can write the FIM as:

$$\mathbf{J} = \frac{2}{\sigma^2} \left(\begin{bmatrix} \sqrt{N} & 0 \\ 0 & N\sqrt{N} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \frac{c_0}{u_0} \end{bmatrix} \begin{bmatrix} G_0(u_0^2) & G_1(u_0^2) \\ G_1(u_0^2) & G_2(u_0^2) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \frac{c_0}{u_0} \end{bmatrix} \begin{bmatrix} \sqrt{N} & 0 \\ 0 & N\sqrt{N} \end{bmatrix} \right).$$

We are interested in finding the CRB for $u_0 = (T - t_0)$, which is the last entry of the inverse of the FIM \mathbf{J}^{-1} . This can be obtained by using direct matrix inversion:

$$\mathbb{E}[(\hat{u}_0 - u_0)^2] \geq \frac{\sigma^2}{N^3} \left(\frac{u_0}{c_0} \right)^2 \left(\frac{G_0(u_0^2)}{G_0(u_0^2)G_2(u_0^2) - G_1(u_0^2)G_1(u_0^2)} \right).$$

Two-Pole Case: We now consider the case when $K = 2$ as in Theorem 4. For convenience, we define the following:

$$\mathbf{R} = \text{diag}(\sqrt{N}, \sqrt{N}, N\sqrt{N}, N\sqrt{N})$$

$$\mathbf{S} = \text{diag} \left(1, 1, \left(\frac{c_0}{u_0} \right), \left(\frac{c_1}{u_1} \right) \right)$$

and

$$G_r(x) = \frac{1}{N^{r+1}} \sum_{n=0}^{N-1} n^r(x)^n. \quad (28)$$

Recall the definitions from Section V-D. Then the FIM can be written as

$$\mathbf{J} = \frac{1}{\sigma^2} \cdot \mathbf{R} \cdot \mathbf{S} \cdot \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{C} \end{bmatrix} \cdot \mathbf{S} \cdot \mathbf{R}. \quad (29)$$

By defining $\text{SNR}_k = c_k^2/\sigma^2$, then the CRB is given by

$$\mathbb{E}[(\hat{t}_k - t_k)^2] \geq \frac{u_0^2}{\text{SNR}_k \cdot N^3} [(\mathbf{C} - \mathbf{B}^T \mathbf{A}^{-1} \mathbf{B})^{-1}]_{k,k}.$$

Unfortunately, no further simplification has been found in finding the inverse of the FIM, and we obtain the CRB by numerical evaluation instead.

REFERENCES

- [1] M. Unser, "Sampling—50 years after Shannon," *Proc. IEEE*, vol. 88, pp. 569–587, Apr. 2000.
- [2] 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.
- [3] 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.
- [4] J. Kusuma and V. K. Goyal, "Multichannel sampling for parametric signals with a successive approximation property," in *Proc. IEEE Conf. Image Process.*, Atlanta, GA, Oct. 2006, pp. 1265–1268.
- [5] F. Rieke, D. Warland, R. de Ruyter van Stevenink, and W. Bialek, *Spikes: Exploring the Neural Code*. Cambridge, MA: MIT Press, 1999.
- [6] A. Ridolfi and M. Z. Win, "Ultrawide bandwidth signals as shot-noise: A unifying approach," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 4, pp. 899–905, Apr. 2006.

- [7] I. Maravić 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, Jul. 2005.
- [8] R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 4, pp. 539–550, Apr. 1999.
- [9] J. Kusuma, "Economical sampling of parametric signals," Ph.D. dissertation, Mass. Inst. Technol., Cambridge, 2006.
- [10] G. R. de Prony, "Essai Éxperimental et analytique: Sur les lois de la dilatabilité de fluides Élastique et sur celles de la force expansive de la vapeur de l'alkool, à différentes températures," (in French) *J. l'École Polytech.*, vol. 1, no. 22, pp. 24–76.
- [11] R. G. Cornell, "A method for fitting linear combinations of exponentials," *Biometrika*, vol. 18, no. 1, pp. 104–113, Mar. 1962.
- [12] A. A. Istratov and O. F. Vyenko, "Exponential analysis in physical phenomena," *Rev. Scientif. Instrum.*, vol. 70, no. 2, pp. 1233–1257, Feb. 1999.
- [13] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Trans. Signal Process.*, vol. 38, no. 5, pp. 814–824, May 1990.
- [14] Y. Hua and T. K. Sarkar, "On SVD for estimating generalized eigenvalues of singular matrix pencil in noise," *IEEE Trans. Signal Process.*, vol. 39, no. 4, pp. 892–900, 1991.
- [15] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Upper Saddle River, NJ: Prentice-Hall, 2005.
- [16] R. Roy and T. Kailath, "ESPRIT—Estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-37, pp. 984–995, Jul. 1989.
- [17] M. Elad, P. Milanfar, and G. H. Golub, "Shape from moments—An estimation theory perspective," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1814–1829, Jul. 2004.
- [18] P. Stoica and A. Nehorai, "Statistical analysis of two nonlinear least-squares estimators of sine-wave parameters in the colored-noise case," *Circuits Syst., Signal Process.*, vol. 8, no. 1, pp. 3–15, 1989.
- [19] M. A. Rahman and K. B. Yu, "Total least squares approach for frequency estimation using linear prediction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, no. 10, pp. 1440–1454, Oct. 1987.
- [20] W. M. Steedly, C. J. Ying, and R. L. Moses, "Statistical analysis of TLS-based Prony techniques," *Automatica*, vol. 30, no. 1, pp. 115–129, Jan. 1994.
- [21] J. Petersson and K. Holmström, *Methods for Parameter Estimation in Exponential Sums*. Mälardalen Univ. Västerås, Sweden, 1997, Tech. Rep. IMA-TOM-1997-5.
- [22] M. Agha, "A direct method for fitting linear combinations of exponentials," *Biometrics*, vol. 27, no. 2, pp. 399–413, Jun. 1971.
- [23] S. D. Foss, "A method of exponential curve fitting by numerical integration," *Biometrics*, vol. 26, no. 4, pp. 815–821, Dec. 1970.
- [24] E. Dilaveroğlu, "Nonmatrix Cramér–Rao bound expressions for high-resolution frequency estimators," *IEEE Trans. Signal Process.*, vol. 46, no. 2, pp. 463–475, Feb. 1998.

- [25] S. T. Smith, "Statistical resolution limits and the complexified Cramér–Rao bound," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1597–1609, May 2005.
- [26] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.



Julius Kusuma (S'97–M'06) received the B.S.E.E. degree with highest distinction from Purdue University, West Lafayette, where he won the Rappaport Scholarship award; and the M.S.E.E. degree from the University of California, Berkeley, where he received the Demetri Angelakos Memorial Award. He received the Ph.D. degree from the Massachusetts Institute of Technology, Cambridge, in 2006, where he was an MIT Presidential Fellow in 2001.

He was a Visiting Scientist with the École Polytechnique Fédérale de Lausanne, Switzerland, in 2001 and 2004. He joined Schlumberger Technology Corporation in 2006, where he is currently a Senior Research Scientist.



Vivek K Goyal (S'92–M'98–SM'03) received the B.S. degree in mathematics and the B.S.E. degree in electrical engineering (both with highest distinction) from the University of Iowa, Iowa City, in 1993. He received the M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley, in 1995 and 1998, respectively.

He was a Research Assistant with the Laboratoire de Communications Audiovisuelles at École Polytechnique Fédérale de Lausanne, Switzerland, in 1996; a Member of Technical Staff with the Mathematics of Communications Research Department, Bell Laboratories, Lucent Technologies, during 1998–2001; and a Senior Research Engineer for Digital Fountain, Inc., during 2001–2003. He is currently the Esther and Harold E. Edgerton Associate Professor of Electrical Engineering and a member of the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge. His research interests include source coding theory, sampling, quantization, and information gathering and dispersal in networks.

Dr. Goyal is a member of Phi Beta Kappa, Tau Beta Pi, Sigma Xi, Eta Kappa Nu, and SIAM. In 1998, he received the Eliahu Jury Award of the University of California, Berkeley, awarded to a graduate student or recent alumnus for outstanding achievement in systems, communications, control, or signal processing. He was also awarded the 2002 IEEE Signal Processing Society Magazine Award and an NSF CAREER Award. He serves on the IEEE Signal Processing Society's Image and Multiple Dimensional Signal Processing Technical Committee and as permanent Co-Chair of the SPIE Wavelets conference series.