Economical Sampling of Parametric Signals

by

Julius Kusuma

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

August 2006

Author .....
Department of Electrical Engineering and Computer Science
August 28, 2006

Certified by .....
Vivek K Goyal
Assistant Professor
Thesis Supervisor

Accepted by .....
Arthur C. Smith
Chairman, Department Committee on Graduate Students

Economical Sampling of Parametric Signals

by

Julius Kusuma

Submitted to the Department of Electrical Engineering and Computer Science
on August 28, 2006, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis proposes architectures and algorithms for digital acquisition of parametric signals. It furthermore provides bounds for the performance of these systems in the presence of noise. Our simple acquisition circuitry and low sampling rate enable accurate parameter estimation to be achieved economically. In present practice, sampling and estimation are not integrated: the sampling device does not take advantage of the parametric model, and the estimation assumes that noise in the data is signal-independent additive white Gaussian noise.

We focus on estimating the timing information in signals that are linear combinations of scales and shifts of a known pulse. This signal model is well-known in a variety of disciplines such as ultra-wideband signaling, neurobiology, etc. The signal is completely determined by the amplitudes and shifts of the summands. The delays determine a subspace that contains the signals, so estimating the shifts is equivalent to subspace estimation. By contrast, conventional sampling theory yields a least-squares approximation to a signal from a fixed shift-invariant subspace of possible reconstructions.

Conventional acquisition takes samples at a rate higher than twice the signal bandwidth. Although this may be feasible, there is a trade-off between power, accuracy, and speed. Under the signal model of interest, when the pulses are very narrow, the number of parameters per unit time—the rate of innovation—is much lower than the Fourier bandwidth. There is thus potential for much lower sampling rate so long as nonlinear reconstruction algorithms are used.

We present a new sampling scheme that takes simultaneous samples at the outputs of multiple channels. This new scheme can be implemented with simple circuitry and has a successive approximation property that can be used to detect undermodeling. In many regimes our algorithms provide better timing accuracy and resolution than conventional systems. Our new analytical and algorithmic techniques are applied to previously proposed systems, and it is shown that all the systems considered have super-resolution properties.

Finally, we consider the same parameter estimation problem when the sampling instances are perturbed by signal-independent timing noise. We give an iterative algorithm that achieves accurate timing estimation by exploiting knowledge of the pulse shape.

Thesis Supervisor: Vivek K Goyal
Title: Assistant Professor

Acknowledgments

Throughout my academic training I have been blessed with the presence of many wonderful teachers, friends, and supporters. Without their help, I will either have lost my sanity or failed to finish my PhD! A comprehensive list of those to whom I owe their support and generosity will by itself comprise a larger volume than this thesis; hence I can only offer a small list and apologize to those who are not mentioned on this list.

I would like to begin by thanking my family for all their unconditional love, limitless support, and patience. I could not have asked for a better Father, Mother, nor Brother! I thank my advisor Vivek K Goyal for being my last advisor, and for being a great advisor. I am flattered to be his first PhD graduate and hope to do him proud. I would also like to thank Moe Z. Win, Sanjoy K. Mitter and Denny M. Freeman for serving on my thesis committee, and for their encouragements through the worst times at MIT.

My work has benefited from collaborations, discussions, and suggestions from many researchers. Among others, I am grateful for the continuing collaboration with Martin Vetterli, Andrea Ridolfi, and Kannan Ramchandran. I would not have been half the student I was without the advising of Sekhar Tatikonda, Vincent Chan, Muriel Medard, Bob Gallager, Greg Wornell, and Vahid Tarokh. My stay at LIDS was thanks to Vincent Chan and Muriel Medard, I would like to thank the support staff at RLE especially Eric J. Strattman. I also give Marilyn Pierce and the staff members at the EECS Graduate Office special thanks for their support and patience as I fail again and again to submit anything on time. In the last part of my graduate career Joel S. Schindall and Manish Bhardwaj have given generous advice and encouragement, leading well into my job search and the completion of my thesis.

During my stay in Boston I have made many friends, starting with my roomates who tolerated all my cooking and coffee roasting: Ariel, Abby, Justin, Chris Brown, Shauna, Alex Feldman, and Chris Fang-Yen. I probably will not be here today if not for the great care of my teammates Mark Abramson, Shane Mulrooney, Zach Hartwig, and Eric Fleming. I probably will not have maintained my sanity if not for the bike rides and bike geeking with Chris Dever, Lodrina Cherne, Eric Sakalowsky, and Julie Monagle. I owe Mario, Tricia and Ian Valenti (or is it The Valentii) special thanks for all the good times and support in the bad times. I will miss the drinks and musings about the virtues of Belgian versus Swiss chocolate with Tom Schouwenaars, Hadley Yates, Charles Swannack-Willenborg, Nicole Willenborg-Swannack, Lav Varshney, Guy Weichenberg, Raj Rao, Michael J. Neely, Constantine Caramanis, Raul Blazquez-Fernandez, and Hayden Marcollo.

I owe Trevor Meyerowitz a big thanks for the patient read and re-read of large portions of my thesis, and to Kush and Lav Varshney, Manish, Mitra, Mario, and Adam for reviewing and commenting on my thesis defense preparation. I am grateful for old friends Paul Twohey, Dan Tabion, Jay Freyensee, and the occasional but very wise advising of Eric A. Jacobsen. My stays in Switzerland have been very productive and enjoyable thanks to the hospitality of Nils Rinaldi and his family, Guillermo, Thibaut, Aslan, Thomas, Robin, Irena, Emre, Rüdiger, Cyril, and Jocelyn Plantefol. Je vous adresse mes plus vifs remerciements.

Finally, I thank Sarah Ann Slavoff for her wonderful company and support through the final stages of the writing and defense of this thesis.

My work at MIT was supported by the MIT Presidential Fellowship, the Alan T. Waterman Award, Grant No. CCR-0139398, and the Analog Devices Corporation.

Contents

Chapter 1

Introduction

This thesis revisits the problem of estimating signal parameters from samples. The conventional approach has two steps: first, the signal is sampled uniformly in time above its Nyquist rate, with anti-aliasing prefiltering if needed; then, a parameter estimation algorithm is applied on the discrete-time samples. In this approach the first step does not take advantage of the parametric model of the signal, and the second step usually assumes that noise in the data is signal-independent additive white Gaussian noise. Whether the conventional approach is effective and economical depends on the parameters to be estimated. If the goal of the parameter estimation is to find the best approximation to the acquired signal from a subspace of possible estimates, the conventional approach is ideal. However, many important parameter estimation problems are not of this simple form.

In this thesis we propose the integration of *signal acquisition* and *parameter estimation*. In particular, we are interested in keeping the acquisition circuitry simple and the sampling rate low so that accurate parameter estimation can be achieved economically. We present new architectures, several analyses of the effect of noise, algorithms for mitigating noise, and extensive simulations using realistic device models. These demonstrate that the modularity and linearity of the conventional approach come at the price of significant performance degradation. Our goal is to improve the economy of signal acquisition in terms of hardware complexity, sampling rate, and associated estimation algorithms.

The variety of parameter estimation problems seen in practice is unlimited, and it is difficult to make concrete statements without an explicit model. So here we consider signals that are a superposition of shifted versions of a known pulse $h(t)$,

$$x(t) = \sum_k c_k h(t - t_k). \tag{1.1}$$

These signals have parameters $\{c_k\}$ and $\{t_k\}$ which we refer to as the amplitudes and the offsets. Some practical motivations for considering this signal class are discussed in Section 1.1. From a theoretical and esthetic perspective, this signal class is attractive for several reasons. Principal among them is that the bandwidth of $h(t)$ does not fundamentally determine whether the parameters can be recovered from samples of $x(t)$. The generality of (1.1) is discussed further in Chapter 2.

This thesis focuses on the prototypical case when the pulse shape in (1.1) is the Dirac delta function, $h(t) = \delta(t)$. This case is of particular interest because its support in the Fourier domain is infinite: conventional sampling theory suggests that it is not possible to perfectly reconstruct this signal.

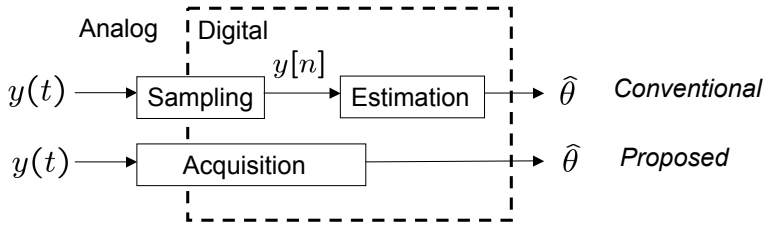


Figure 1-1. An illustration of the difference between the classical approach and the proposed approach. In the classical approach, the signal is first sampled in time at its Nyquist rate, and the obtained samples are then used for estimation. In the proposed approach, we design the data acquisition system with parameter estimation explicitly in mind.

There has been much recent work on sampling parametric signals. The *rate of innovation* of a parametric signal is the minimum number of parameters per unit time that can completely describe it. Methods and algorithms were proposed that can perfectly reconstruct certain parametric signals from samples taken at the signal rate of innovation rather than the Nyquist rate [32, 33, 60, 69, 72, 117]. This thesis extends this new sampling framework to consider signal parameter estimation in the presence of noise, explicitly taking advantage of parametric signal models.

The contributions of this thesis are threefold:

1. We propose a new signal acquisition device that is very simple to implement. It takes samples at the outputs of multiple channels simultaneously in time. The associated reconstruction procedures are a departure from the framework of reconstructing signals only in a *fixed* subspace and use nonlinear algorithms which in many cases are subspace estimation algorithms. We draw connections to the classical estimation problem of Prony, for which many algorithms of various complexities have been proposed. The proposed system has a successive approximation property that can be used to detect undermodeling.[1] We analyze the performance of the proposed system by deriving the Cramér-Rao bounds for timing estimation and the resolution of the system. We implement several reconstruction algorithms, and we show that in many cases they are superior to the classical methods and have a super-resolution property.

2. Using the analytical and algorithmic tools developed in the previous part, we revisit prior work on acquiring signals with finite rates of innovation using uniform sampling in time. We derive the Cramér-Rao bounds for timing estimation and resolution, and we show that it is possible to obtain better resolution than the conventional method using simple algorithms. We show that one of the schemes reviewed also has a successive approximation property that can be used to detect undermodeling. Further, this scheme allows for signal segmentation and reconstruction of different orders *after* the samples are taken, up to some constraints. This is particularly attractive for problems where the signal model order is not known exactly.

[1] Part of this work will be presented at the 2006 IEEE Intl. Conf. on Image Processing, Atlanta, GA [58].

3. In the last part of the thesis we consider the problem of estimating the parameters of a signal when the sampling instances are perturbed by signal-independent *timing noise*. The classical techniques consider timing noise to induce signal-independent additive white Gaussian noise on the sample values. We reject this simplification of the problem and give alternative methodologies. We focus on the problem of delay estimation when the pulse shape and amplitude of the signal are known, and we give an iterative algorithm that shows superior performance compared to the traditional method which relies on maximizing the cross-correlation. Using the signal parameters based on the single-chip ultra-wideband transceiver developed at the Massachusetts Institute of Technology (MIT) [11–13,119] we show that in some regimes we can cancel the effect of jitter completely.[2]

■ 1.1 Motivation

With the accelerating use of computers the need for digital signal processing has increased. In order to use an analog signal on a computer it must be digitized using an Analog-to-Digital Converter (ADC). In some cases the only information that we have about the signal of interest is its Fourier bandwidth. The Nyquist-Shannon sampling theorem states that it is possible to faithfully reproduce a signal when it is sampled at twice its bandwidth, at a rate called the Nyquist rate [68, 79]. Using this method, the reconstruction of the signal is done by sinc interpolation. If the ADC sampling rate is lower than this rate, an anti-aliasing lowpass filter with cutoff at half the available sampling rate can be used. In this case, the reconstructed signal is given by

$$x(t) = \sum_{n\in\mathbb{Z}} c_n \operatorname{sinc}(t - nT_s). \tag{1.2}$$

The bandwidth of this signal is $1/(2T_s)$ and its Nyquist rate is $1/T_s$. The signal (1.2) belongs to a fixed subspace that is spanned by $\{\operatorname{sinc}(t - nT_s)\}, n \in \mathbb{Z}$. Applying an anti-aliasing filter of bandwidth $1/T_s$ and sampling the filter output is equivalent to projecting the original signal into this subspace. Applying an anti-aliasing filter of smaller bandwidth and sampling at twice this lower bandwidth is equivalent to projecting the signal into a smaller subspace. This classical approach is illustrated in Figure 1-2.

■ 1.1.1 Why Parametric Signals

Many signals of interest consist of pulses of unknown amplitude and time delay such as (1.1), and we are interested in the information carried in these amplitudes and delays. This is one example of *parametric signals*, meaning signals that can be completely described by their parameters. Figure 1-3 shows one example of this class of signals. When the pulse shape is known, the signal can be described by its parameters: the amplitudes and delays of the pulses. For such a signal, the subspace that the signal of (1.1) belongs to is determined by its shifts. Hence, estimating its shifts $\{t_k\}$ is equivalent to subspace estimation. For the same reason, conventional methods that reconstruct the signal only within a fixed subspace

[2]Part of this work was presented at the 2006 Conf. on Information Sciences and Systems, Princeton, NJ [59].

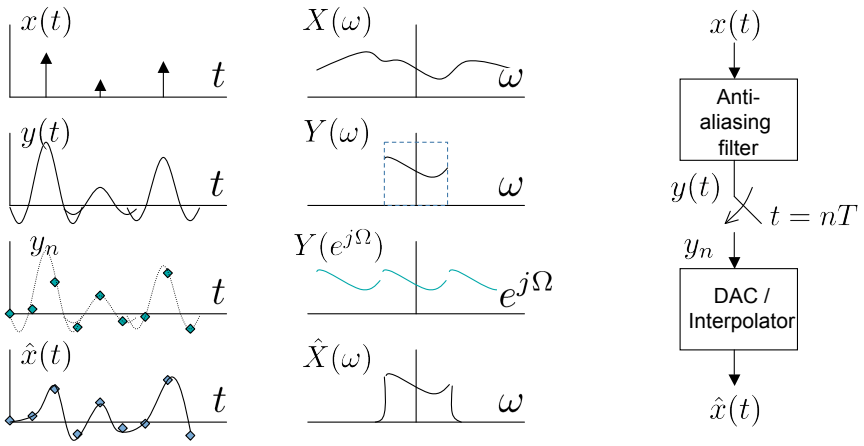


Figure 1-2. The classical framework of sampling signals, due to Nyquist, Shannon and Kotelnikov. The input signal $x(t)$ is first filtered using a anti-aliasing lowpass filter, and samples are taken at twice the cutoff frequency of this filter. The reconstruction is done by applying a reconstruction lowpass filter of the same bandwidth after translation to the continuous-time domain. We show a mismatched pair of anti-aliasing and reconstruction filters, resulting in further smoothing of the signal.

cannot obtain perfect reconstruction. In many cases the Fourier bandwidth of this signal is much larger than the number of parameters per unit time, called the signal *rate of innovation*. We call systems that generate this type of signal *bandwidth-expanding* systems.

Examples include ultra-wideband signals used for communication, ranging and identification applications [22, 38, 75, 76, 81, 121–123], and naturally-occurring signals in biological and physical phenomena [93, 96, 97, 120]. Modern systems such as the Global Positioning System (GPS), Code-Division Multiple Access (CDMA) systems, and neural circuits, are also bandwidth-expanding. That is, the transmitted signal has larger Fourier bandwidth than information rate or rate of innovation [60, 61, 69, 70].

Recent increase of interest in the integration of digital computers and neural systems means that conversion of signals from the analog domain to the digital domain and vice versa is becoming more important. Much recent work has been done in developing biologically-inspired hybrid systems that operate both in the analog and digital domains [40, 95–97]. To quote Sarpeshkar [96],

> "The brain is an excellent example of a hybrid of an analog computer and a digital computer. Neurons use 'spikes' or pulses of voltage to communicate with each other, and these spikes have an inherently hybrid nature: The presence of the spike is a digital event, but the time between spikes is a continuous analog variable."

There are several challenges in translating signals from the analog domain to the digital domain. A neural signal is by nature spiky [93, 120], and therefore its bandwidth is higher than its rate of innovation. In the Nyquist sampling framework, this necessitates a high-

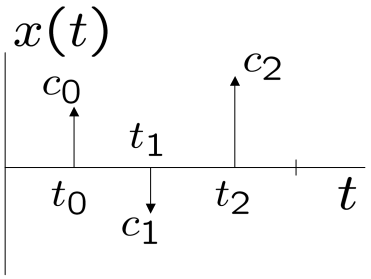


Figure 1-3. An example of a parametric signal. The signal $x(t)$ consists of K Dirac delta functions scaled and delayed. It is completely determined by its parameters $\{c_k, t_k\}_{k=0}^{K-1}$.

rate ADC, which in turn leads to high power consumption, complex circuitry, and stringent noise and circuit quality requirements.

Many applications admit a signal model that is particularly suitable for our proposed signal acquisition system. For example, it is known that due to the refractory period in the nervous system, inter-spike timing in a neural system is at least 1 ms, meaning that within any 1 ms period, there can be at most one spike [66, 93]. Recent interest in developing brain-machine interfaces means that economical methods for spike detection are desirable. Current methods either use a threshold detector for detecting the occurrence of the spikes or take samples at the Nyquist rate of the spikes and use digital signal processing for denoising [80]. The first approach requires sufficient analog processing to isolate the spike signal, and the second approach requires a high-rate ADC that samples at a rate higher than twice the bandwidth of the pulses.

In this thesis we are concerned with three challenges that are acute for acquisition of spiky signals: how to sample signals of very high bandwidth, how to deal with timing noise, and how to build a signal acquisition device from simple building blocks.

■ 1.1.2 Why Revisit Signal Acquisition

Although wideband ADCs are becoming more commonplace, there are many challenges in the design and implementation of such devices. Analog and digital electronics have advanced tremendously, but not at the same pace. For most applications it is preferable to shift functionality from analog to digital as much as possible. However, the fraction of chip area, power consumption, and manufacturing cost associated with the analog portion often grows over time, because of the need for higher-rate and more accurate ADCs to enable faster signaling and communication [12, 78, 119].

The recent thesis of Blazquez [11] proposed a single-chip UWB transceiver that sends Gaussian pulses with a bandwidth of 1 GHz. The manufactured chip is shown in Figure 1-4 with its subsystems highlighted. The signal acquisition system is a significant part of the proposed receiver, with the Flash ADC taking up nearly half of the baseband processor! The performance of an ADC is usually given in terms of its sampling rate, power consumption, dynamic range, and Effective Number Of Bits (ENOB). There is a tradeoff between these factors. We focus on the tradeoff between speed, accuracy in terms of ENOB, and power.

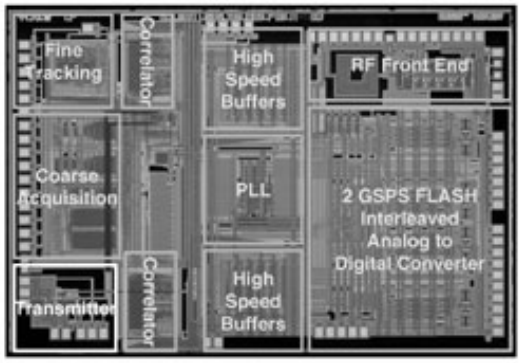


Figure 1-4. A UWB baseband processor on a single chip developed by the MIT Microsystem Technologies Lab.

For a Complementary-symmetry/Metal-Oxide Semiconductor (CMOS) system it is well-known that this tradeoff is given by [116]:

$$\frac{(\text{Speed [Hz]}) \times (\text{Accuracy [RMS]})^2}{(\text{Power [Watts]})} \approx \text{Constant}. \tag{1.3}$$

Although many architectures can obtain good tradeoff of these factors, there are several limiting factors such as timing noise. It is known that timing noise is a significant source of impairment in wideband ADCs, decreasing resolution by about 1 bit for every doubling of the sampling rate in the Mega- and Giga-sample per second regime [3, 67, 118].

It was recently shown that for some bandwidth-expanding communication systems it is possible to adjust the receiver sampling rate to be below the Nyquist rate even in the presence of noise [60,61,69,70,117]. As long as the sampling rate is above the rate of innovation, the resulting performance is commensurate to the performance of the conventional system when the transmission bandwidth is adjusted so that there is no aliasing. We will review this body of work in Section 2.3.2 and Section 2.3.4. Although Maravić and Vetterli [71] considered the performance of the sampling scheme of [117] in the presence of AWGN, their analysis assumes that the reconstruction is done by first estimating the second-order statistics of the observation and then using the matrix pencil method proposed by Hua and Sarkar [48, 49]. While the matrix pencil method is known to perform well for large block sizes, parameter estimation via second-order statistics is known to perform poorly for small block sizes. Further, the result of Vetterli and Maravic holds only for signals which contain only one component.

The effect of timing noise on the observation is signal-dependent. We consider the effect of timing noise on conventional uniform-in-time analog-to-digital conversion systems, with the goal of estimating the unknown parameters of the signal. As mentioned before, timing noise is a dominant source of noise in wideband ADCs. However, the standard approach is to consider the effect of timing noise to be signal-independent additive white Gaussian noise. The variance of this noise term is obtained by considering the highest frequency that can be sampled without aliasing, and computing the effect of timing noise on the zero-crossings of this sinusoidal signal. This is called approximation via the maximum signal "slew" rate

and gives

$$\text{AWGN [RMS]} = \text{Jitter [RMS]} \times \text{ADC Bandwidth [Hz]}.$$

We reject this simplification and show that there are large gains from explicitly considering the effect of timing noise on the samples of a parametric signal model. We focus on the canonical problem of signal delay estimation, for which timing accuracy is critical [38, 81]. We simulate the proposed approach using the signal models based on the ultra-wideband (UWB) testbed developed at the MIT Microsystem Technologies Laboratory (MTL), and show significant gains. In some regimes, we can completely cancel the effect of jitter. The characterization of timing noise is well-studied. One common model for timing noise is a stationary, zero-mean, filtered Gaussian white noise or Wiener process [23, 24, 41, 64]. Our algorithm can also be adapted to take advantage of this more accurate timing noise model.

In the presence of AWGN, the optimal estimator of pulse delays from samples obtained via Nyquist sampling finds the maxima at the output of a correlator. When there is only one pulse in the signal, this can be implemented using a gradient search algorithm. But when there are K pulses in the signal, the estimator is a K-dimensional maximization. It is known that the objective function has many local maxima, hence an accurate initial estimate is required to implement a gradient search [106]. In many cases a good initial estimate is not available, and the performance of the estimator suffers. Moreover, the resolution of the conventional system is limited.

We define the resolution of a signal acquisition system as the minimum spacing of two signal components that can still be resolved by the system. The resolution of a conventional system is determined solely by its sampling rate. In frequency estimation there is a class of algorithms that have a *super-resolution* property, that is at high SNR its resolution is higher than that of the conventional system. We will show that our proposed systems have super-resolution properties.

■ 1.2 Thesis Outline

We start by reviewing previous work in sampling and reconstruction of parametric signals in Chapter 2. Then we present the main results in three parts:

1. **Multichannel sampling of parametric signals:** In Chapter 3 we depart from the framework of linear subspace projection and use nonlinear algorithms that in many cases are subspace estimation algorithms. We reject the notion of Nyquist sampling for parametric signals and give an alternative method. The proposed system takes samples simultaneously in time at the output of different channels, and is particularly simple to implement using either integrators or B-spline kernels. We show that algorithms developed for line spectra estimation and Prony estimation can be used in our proposed system. We also derive the Cramér-Rao bound of our proposed system, and we show that our system has superior resolution compared to the classical method analytically and by simulation.

2. **Sampling signals with a finite rate of innovation:** In Chapter 4 we re-examine the sampling system proposed by Vetterli *et al.* that takes uniform samples in the time domain, at the rate of innovation of the signal. In the noisy case, the sampling rate can be adjusted anywhere above the rate of innovation, without requiring adaptation

of the signal pulse bandwidth. We show that below the Nyquist rate the performance of the proposed system is the same as that of the classical approach, if the bandwidth of the pulse shape in the latter case were adjusted according to the sampling rate available at the signal acquisition device. We use methods developed in the previous chapter to derive the resolution limits of the proposed system and show by numerical simulation that it has superior performance compared to the classical method.

3. **Signal parameter estimation in the presence of timing noise:** In Chapter 5 we explicitly consider timing noise instead of taking the standard approach of assuming timing noise noise to induce signal-independent additive white Gaussian noise. Taking advantage of a parametric signal model, we focus on the problem of delay estimation and show improved accuracy compared to the classical approach in terms of signal-to-noise ratio, jitter power, and sampling rate.

Chapter 2

Background

In this chapter we review the previous work in exponential fitting and sampling signals with a finite rate of innovation, which form the mathematical basis for the first part of the thesis. We also review wideband signal acquisition and examine the problem of timing noise mitigation.

■ 2.1 Powersums and Exponential Fitting

The nonlinear parameter estimation problems that we consider in this thesis are based on observations that have the form of *powersum series*. We start by giving the following definition.

Definition 2.1 (Powersum series). *A powersum series of K components is given by*

$$x_n = \sum_{k=0}^{K-1} c_k (u_k)^n, \quad n = 0, 1, \ldots, N-1. \tag{2.1}$$

We say that $\{c_k\}$ are the amplitudes and $\{u_k\}$ are the coefficients or poles.

The desired parameters are the K pairs $\{c_k, u_k\}$. A sequence of form (2.1) was first studied by Baron de Prony in 1795 as he attempted to find the decay rates of chemical processes [25]. In the original problem, the observations and parameters are real-valued. This is sometimes called "real exponential fitting" or "exponential analysis" in the natural sciences literature [4,20,36,50,124]. For the case when $K = 1$ and $N = 2$, we have a system of equations with 2 unknowns. Then we can quickly solve $\hat{c}_0 = x_0$, $\hat{u}_0 = x_1/\hat{c}_0$. This is the only instance we know of in which there is a closed-form solution, and one that gives $\hat{c}_k$ before $\hat{u}_k$. We will take advantage of the relative simplicity of the lower order problems.

De Prony showed that in the noiseless case it is possible to find $\{c_k, u_k\}$ exactly based on $N = 2K$ observations. In the signal processing literature, his method is called the *annihilating filter* method, which we review below. We review several other known methods in Appendix A. Instead of presenting Prony's method in terms of a constant-coefficient difference equation, we use signal processing terminology instead. The one-sided z-transform of (2.1) is given by

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

In this form the parameters $\{u_k\}$ appear as poles of $X(z)$. Now let h_n be a finite impulse response (FIR) filter of length K, with z-transform $H(z)$. If the central values in the convolution of h_n and x_n are zero,

$$h_n * x_n = 0, \quad n = K+1, \ldots, N-1, \tag{2.2}$$

then the K zeros of $H(z)$ must equal the K poles of $X(z)$. In other words, up to a scaling factor we must have that:

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

Suppose that the time-domain representation of this filter is given by $h_n = 1 + \sum_{n=1}^{K} h_n z^{-n}$. The solution to (2.2) can be obtained by writing the convolution in matrix form. The matrix equation that we need to solve is given by

$$\begin{bmatrix} x_{K-1} & x_{K-2} & \ldots & x_0 \\ x_K & x_{K-1} & \ldots & x_1 \\ \vdots & \vdots & \ddots & \vdots \\ x_{2K-2} & x_{2K-3} & \ldots & x_{K-1} \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_K \end{bmatrix} = - \begin{bmatrix} x_K \\ x_{K+1} \\ \vdots \\ x_{2K-1} \end{bmatrix} \tag{2.3}$$

In order to form the system of linear equations (2.3), we require $N \geq 2K$ observations. After finding $h_0, h_1, \ldots, h_{K-1}$ we obtain u_k by finding the roots of the corresponding filter $H(z)$. It is known that this method is not a consistent estimator of the parameters [51,104]. The number of required samples is summarized in the following lemma, which will be used to prove several sampling theorems later in this thesis.

Lemma 2.1 (Computing the parameters of a powersum series [25]). *Consider a powersum series consisting of K components*

$$x_n = \sum_{k=0}^{K-1} c_k (u_k)^n, \quad n = 0, 1, \ldots, N-1. \tag{2.4}$$

It is possible to obtain the parameters $\{c_k, u_k\}_{k=0}^{K-1}$ from $N \geq 2K$ observations x_n.

■ 2.2 Parametric Signals

A parametric signal is a signal that can be expressed in a small number of parameters. One well-studied example is a signal comprising of K complex exponentials, say

$$x(t) = \sum_{k=0}^{K-1} c_k \exp(j\omega_k n), \tag{2.5}$$

where $\{c_k\}_{k=0}^{K-1}$ are its complex amplitudes and $\{\omega_k\}_{k=0}^{K-1}$ are its frequencies. A signal of form (2.5) is a powersum series with coefficients $u_k = \exp(j\omega_k)$, which we introduced in Definition 2.1. Given that the signal contains K such terms it is possible to obtain perfect reconstruction using $N \geq 2K$ samples taken uniformly in time [106]. The minimum number

of parameters that can define a signal is called the rate of innovation of the signal. We make this precise following the definition given in [117].

Definition 2.2 (Signal rate of innovation). *Let $C_x(t_a, t_b)$ be a counting function that counts the number of degrees of freedom in a signal $x(t)$ over the interval $[t_a, t_b)$. Given a signal $x(t)$, its rate of innovation ρ is defined as*

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

Another example of parametric signals that is of particular interest in this thesis are signals consisting of scaled-and-delayed pulses. When the pulses are narrow, those signals have very high Fourier bandwidth, in many cases orders of magnitudes higher than their rates of innovation. We have seen that these types of signals occur in many biological systems and are used in ultra-wideband communication and signaling.

Definition 2.3 (Parametric Signals: Sum of Diracs). *Given a set of K amplitudes $\{c_k\}_{k=0}^{K-1}$ and time delays $\{t_k\}_{k=0}^{K-1}$, the signal of interest $x(t)$ is a superposition of Diracs*

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

The parameters $\{c_k, t_k\}_{k=0}^{K-1}$ are real-valued.

The problem of signal parameter estimation in continuous-time domain is well-studied [52, 111]. However, translation into the discrete-time domain is usually done by implicitly assuming sampling above the Nyquist rate. The problem with this approach is that while the number of desired signal parameters tends to be small, the Fourier bandwidth of the signal may be orders of magnitude higher. This is the main motivation of the work of sampling signals with a finite rate of innovation [117].

Both uniform and non-uniform sampling of analog signals have been studied extensively, including approximation and noise robustness properties [45, 115]. This theory builds upon the celebrated sampling theorem of Shannon, Nyquist and Kotelnikov, which applies to the class of *bandlimited signals* [53, 79, 100]. Many generalizations have been proposed: see for example [45, 83, 115]. All of this work relies on reconstructing the signal using linear projection on the span of a Riesz basis. This span is a fixed subspace. In contrast, in this thesis we propose methods that estimate a subspace via estimating the timing parameters of a signal.

■ 2.3 Sampling Signals with a Finite Rate of Innovation

The recent work of Vetterli *et al.* [117] provides examples of non-bandlimited signals that can be recovered perfectly from uniform samples. Though not bandlimited, the signals they can recover are specified by a finite number of degrees of freedom per unit time, a property termed *finite rate of innovation* (FRI).

Extensions to more general sampling kernels have also been developed and algorithms for their perfect reconstruction in the exact and noiseless case have been proposed [32, 58].

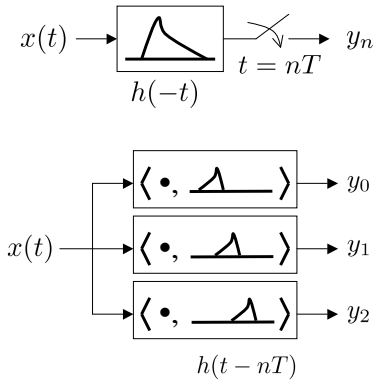


Figure 2-1. Filtering and uniform sampling in time written as a sequence of inner products with shifts of the filter function.

In general the reconstruction algorithms are not linear; most of them rely on solving the powersum series either by the Prony/annihilating filter method or the Matrix Pencil method. We present a few of these methods in Appendix A.

■ 2.3.1 Sampling and inner products

Define the usual $\mathcal{L}^2$ inner product of a signal $x(t)$ and a function $f(t)$ as

$$\langle x(t), f(t) \rangle = \int x(t) f(t) dt.$$

Filtering a signal $x(t)$ with a sampling kernel $h(-t)$ and taking uniform samples $y_n = (x * h)(nT)$ at the output of the sampling kernel can be seen as a sequence of inner products with shifts of the kernel $h(t - nT)$. Indeed,

$$\begin{aligned} y_n &= (x * h)(nT) = \int_{-\infty}^{\infty} x(\tau) h(t - \tau) d\tau \Big|_{t=nT} && (2.6) \\ &= \langle x(t), h(t - nT) \rangle. && (2.7) \end{aligned}$$

We show this in Figure 2-1. When the signal being considered consists of K weighted Diracs, inner products with a function $f(t)$ take a particularly simple form

$$\begin{aligned} \langle x(t), f(t) \rangle &= \left\langle \sum_{k=0}^{K-1} c_k \delta(t - t_k), f(t) \right\rangle \\ &= \sum_{k=0}^{K-1} c_k \langle \delta(t - t_k), f(t) \rangle \\ &= \sum_{k=0}^{K-1} c_k f(t_k). \end{aligned}$$

We will use these properties throughout this thesis. Because of the relationship (2.6) we use the term filter and kernel interchangeably.

■ 2.3.2 Periodic sums of Diracs

Consider signals that are K-sums of Diracs, such as the example shown in Figure 1-3. Such a signal can be written as

$$x(t) = \sum_{k=0}^{K-1} c_k \delta(t - t_k). \tag{2.8}$$

For the time being, we assume that the signal is **periodic** with period 1, hence its rate of innovation is $2K$. From Fourier analysis, a periodic signal can be expressed in terms of its Fourier series coefficients:

$$x(t) = \sum_{m=-\infty}^{+\infty} X[m] \exp(j2\pi mt). \tag{2.9}$$

For the signal model we are interested in, the Fourier series coefficients are given by:

$$X[m] = \sum_{k=0}^{K-1} c_k \exp\left(-j2\pi m t_k\right), \quad m \in \mathbb{Z}. \tag{2.10}$$

The Fourier domain representation given in (2.10) has infinite length, hence we say that this signal is not bandlimited. However, $x(t)$ has *finite* rate of innovation, $2K$.

In conventional linear sampling theory – assuming a non-parametric signal model – then given an ADC that takes samples at f_s Hz, we must first apply an anti-aliasing/lowpass filter with cutoff frequency $f_s/2$ Hz. Suppose that this cutoff corresponds to Fourier series index L. The linear reconstruction of the signal based on its discrete-time samples is then given by:

$$\hat{x}(t) = \sum_{m=-L}^{L} \sum_{k=0}^{K-1} c_k \exp\left(-j2\pi m t_k\right) \exp(j2\pi mt). \tag{2.11}$$

Compare the original signal given by (2.9) and its reconstruction given by (2.11). The reconstruction of the signal involves Fourier-domain truncation. The truncation boundary given by $[-L, +L]$ is determined by the lowpass filter to ensure that there is no aliasing. Clearly, because in general the original signal is non-zero in the entire Fourier domain, the reconstruction error is nonzero. The reconstruction error decays at a rate proportional to the sampling rate f_s. This reconstruction strategy is based on linear projection into a fixed subspace determined by the lowpass filter.

Vetterli, Marziliano and Blu [117] showed that it is possible to obtain *perfect reconstruction* of the signal $x(t)$ by determining its parameters $\{c_k, t_k\}_{k=0}^{K-1}$. They observed that (2.10) is a powersum series with $u_k = \exp(-j2\pi t_k)$, and therefore from Lemma 2.1 the parameters can be estimated from $2L + 1 \geq 2K$ Fourier series coefficients. The parametric reconstruction scheme is based on a nonlinear reconstruction algorithm, which in fact is based on subspace estimation.

Suppose we select the Fourier series coefficients around DC. Computing these Fourier

series coefficients can be done by first applying a lowpass filter corresponding to the cutoff band index $L \leq K$ in the Fourier domain, sampling at twice this cutoff frequency, and then computing the Fourier series from the given samples in the time domain. Note that using this algorithm there is no aliasing in the Nyquist sense either. The primary difference is that we compute the parameters of the signal based on the obtained samples using a nonlinear reconstruction method. Therefore in the noiseless case it is possible to obtain perfect reconstruction by sampling at the rate of innovation of the signal, which is well below the Nyquist rate, after applying a lowpass filter corresponding to this sampling rate [72,117]. Vetterli, Marziliano and Blu gave the following theorem.

Theorem 2.1 (Sampling periodic sums of Diracs [117]). *Let $x(t)$ be a periodic sum of K Diracs as given in (2.8), with period T. It is possible to obtain perfect reconstruction from $N \geq 2K$ uniform samples of $(h_B * x)(t)$, where $h_B(t) = B\,\mathrm{sinc}(Bt)$ is a sinc kernel with bandwidth B chosen such that it is greater than or equal the rate of innovation of the signal $2K/T$.*

Proof. Let the output of the sinc kernel be sampled at twice the bandwidth. Taking the discrete Fourier transform of this sequence we obtain

$$X[m] = \sum_{k=0}^{K-1} c_k \exp\left(-j2\pi m t_k\right), \quad m = -L, \ldots, +L.$$

By construction, the number of Fourier series coefficients obtained is greater than $2K$. This Fourier series is a powersum series, with poles $\{\exp(-j2\pi t_k)\}$ and amplitudes $\{c_k\}$. From Lemma 2.1 we know that it is possible to compute the parameters from $N \geq 2K$ observations of the powersum series. □

The estimation problem associated with this sampling theorem is closely related to the problem of line spectra estimation, where the poles are given by $u_k = \exp(j\omega_k)$. Many algorithms have been proposed that are suitable for this problem [106]. In many cases, their performance in the presence of white additive noise is known [48, 49, 54, 55, 57, 71, 105, 107]. In Section 2.4 we will review some of the results.

Vetterli, Marziliano and Blu also showed that it is possible to use sampling kernels other than an ideal lowpass filter for different classes of signals. The samples are obtained by filtering and sampling

$$y_n = \langle x(t), \varphi(t - nT) \rangle,$$

where φ is the sampling kernel. This scheme is shown in Figure 2-2. The original signal is given in the Fourier domain by $X[m]$ and has infinite support in that domain. The filtered and sampled signal in the Fourier domain is given by $Y[m]$, where we have that $Y[m] = X[m]$ for $m = -2, -1, 0, 1, 2$. It is periodic in the Fourier domain with period 5 due to the sampling. In Figure 2-3 we illustrate the difference between this proposed system and the conventional system based on Nyquist sampling. In the proposed system, we obtain perfect reconstruction by first estimating the parameters of the signal. These parameters allow for infinite reconstruction of the signal in the Fourier domain, despite having applied an anti-aliasing lowpass filter at the input.

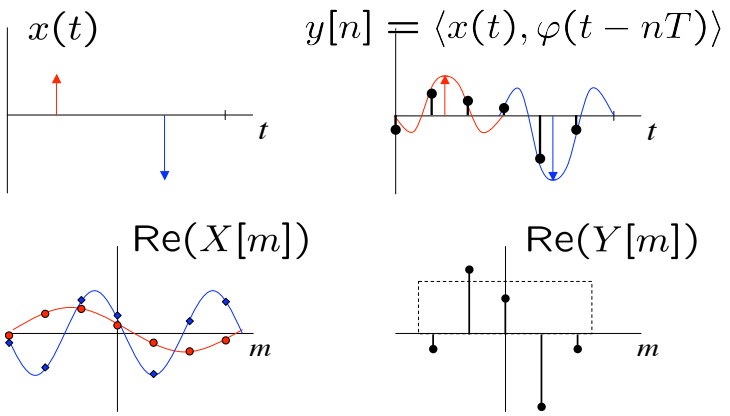


Figure 2-2. FRI sampling of a periodic signal consisting of two Diracs. The original signal $x(t)$ is filtered and sampled to obtain $y[n]$, and the corresponding Fourier series are given below the time-domain signals.

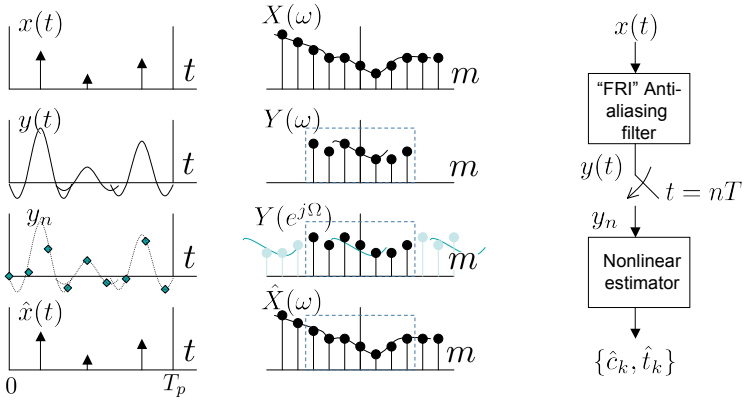


Figure 2-3. FRI sampling of a periodic signal consisting of K Diracs. In this example we illustrate the difference between the parametric approach and the non-parametric approach given in Section 1.1.

■ 2.3.3 Sum of a finite number of Diracs

Now consider the case when the signal model given in (2.8) is **aperiodic**. It was shown in [117] that it is possible to perfectly reconstruct the signal from $N \geq 2K$ measurements. Suppose that we take uniform samples in time at the output of a Gaussian filter $g(t) = \exp(-t^2/(2\rho^2))$. In the noiseless case, the samples at the output of a Gaussian filter are given by

$$\begin{aligned} y[n] &= \sum_{k=0}^{K-1} c_k \exp\left(-\frac{(nT - t_k)^2}{2\rho^2}\right) \\ &= \sum_{k=0}^{K-1} \left(c_k \exp\left(-\frac{t_k^2}{2\rho^2}\right)\right) \cdot \exp\left(\frac{2nTt_k}{2\rho^2}\right) \cdot \exp\left(-\frac{(nT)^2}{2\rho^2}\right) \\ &= \sum_{k=0}^{K-1} a_k (u_k)^2 \exp\left(-\frac{(nT)^2}{2\rho^2}\right), \end{aligned}$$

where $a_k = c_k \exp(-t_k^2/2\rho^2)$ and $u_k = \exp(Tt_k/\rho^2)$. Let $v[n] = y[n] \cdot \exp(+(nT)^2/2\rho^2)$. Then $v[n]$ is a powersum series with real values, which we have examined in the previous chapter. Just like in the periodic case, we have obtained a theorem.

Theorem 2.2 (Sampling aperiodic sums of Diracs [117]). *Let $x(t)$ be an aperiodic sum of K Diracs as given in (2.8). It is possible to obtain perfect reconstruction from $N \geq 2K$ uniform samples of $(h_G * x)(t)$, where $h_G(t) = \exp(-t^2/2\rho^2)$ is a Gaussian kernel with width ρ.*

There are disadvantages to this nonlinear reconstruction technique. Its performance when there is *undermodeling* is not yet known. Therefore the samples must strictly be taken above the rate of innovation. On the other hand, an anti-aliasing filter can be used in the linear sampling and reconstruction scheme given by Nyquist to prevent undermodeling. We do not yet have an analogous facility for the parametric signal models that we consider in this thesis.

It has been shown that this new sampling theory can be applied to *bandwidth-expanding communication systems*. It can also be used for channel acquisition and characterization in many applications such as GPS, UWB and CDMA systems. The model of (2.8) has been extended to include more general pulse shapes and different "smooth" pulse shapes arriving at different times [60,61,69–71]. It was shown that the performance of the proposed system in the case of delay estimation is the same as that of the classical approach for the same sampling rate assuming no aliasing in both cases, as shown in Figure 2-4. However, the classical approach requires that the bandwidth of the transmitter be adapted to ensure that the receiver sampling rate is above the corresponding Nyquist rate. The FRI receiver does not require adaptation at the transmitter side, as long as its sampling rate is above the transmitted signal's rate of innovation. We show later in this thesis that this system has *super-resolution property*: it can discriminate two pulses that are spaced more closely than the sampling rate.

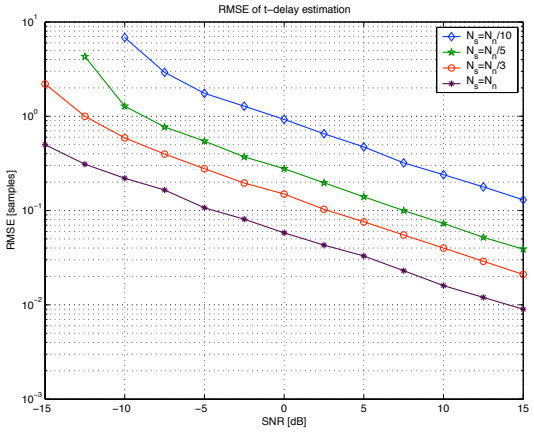


Figure 2-4. Delay estimation performance of an FRI receiver, sampling at rates at and below the Nyquist rate. In the figure, N_n is the Nyquist rate and N_s is the sampling rate. The different curves show the performance when the sampling rate is set to be $N_n/10$, $N_n/5$, $N_n/3$, and N_n in increasing order of performance. This figure is taken from [69].

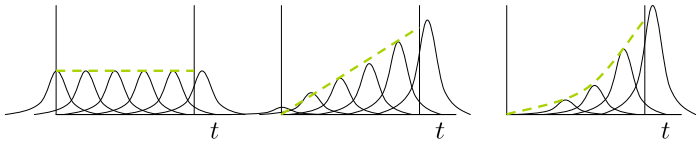


Figure 2-5. Production of t^0, t^1 and t^2 within an interval of interest, from uniformly-spaced B-splines.

■ 2.3.4 Parametric sampling using finite kernels

Recently the class of sampling filters has been extended [31–33] to include *local* reconstruction methods even for infinite-length signals from samples taken uniformly in time. The proposed method includes sampling filters which satisfy the Strang-Fix conditions, or in other words can reproduce polynomials and exponentials up to a certain order.

Definition 2.4 (Kernels that reproduce polynomials). *A kernel $\varphi(t)$ is said to be able to reproduce polynomials up to order R in interval $[0, I)$ if there exists $c_{m,n}$ such that*

$$\sum_n c_{m,n}\varphi(t-n) = t^m, \quad m = 0, 1, \ldots, R, \text{ and all} t \in [0, I).$$

We show an example of the production of monomials in Figure 2-5. Many kernels satisfy such conditions, such as the B-splines. It is known that B-splines of order R can reproduce polynomials up to that same order [114]. The coefficients $c_{m,n}$ are easily obtained for the case where we use B-splines. Since B-splines are interpolating functions, we simply set

$$c_{m,n} = n^m, \quad m = 0, 1, \ldots, R, \quad n \in [0, I)$$

Clearly, we require that $N \geq R$. Due to the compact support of the finite-order B-splines we only need to find the coefficients for the terms that overlap the interval of interest.

Now we can move on to the problem of reconstructing an *infinite* stream of Diracs. Dragotti *et al.* gave the following theorem.

Theorem 2.3 (Sampling infinite sums of Diracs [32]). *Consider an infinite stream of Diracs and let $\varphi(t)$ be a sampling kernel that can reproduce polynomials of maximum degree $N-1 \geq 2K-1$ and of compact support C. Then the signal is uniquely determined from N samples $y_n = \langle x(t), \varphi(t-nT) \rangle$ if and only if there are at most K Diracs in any interval of length KCT.*

To obtain the proof, it is sufficient to observe that $s[m] = \sum_n c_{m,n} y_n = \int x(t) t^m \, dt = \sum_{k=0}^{K-1} a_k t_k^m$, for $m = 0, 1, \ldots, N$. This is again a powersum series (2.1), and we can find the unknown terms from a sufficient number of observations according to Lemma 2.1. To finish the proof, in the case that $x(t)$ contains infinitely many Diracs, we can reconstruct it exactly if there is only at most K of them within a compact interval of length KCT.

Proof. Without loss of generality, assume $T = 1$ and that $x(t)$ has K Diracs. Let $s[m] = \sum_n d_{m,n} y_n$, for $m = 0, 1, \ldots, N$. This is a weighted sum of observed samples, with weights $d_{m,n}$ given by the polynomial reproduction conditions above. Then,

$$s[m] = \sum_n d_{m,n} y_n = \int x(t) t^m \, dt \tag{2.12}$$

$$= \sum_{k=0}^{K-1} c_k t_k^m, \quad m = 0, 1, \ldots, N. \tag{2.13}$$

This is again a powersum series (2.1), and we can find the unknown terms from a sufficient number of observations. To finish the proof, in the case that $x(t)$ contains infinitely many Diracs, we can reconstruct it exactly if there is only at most K of them within a compact interval of length KCT. □

The same approach can be applied to kernels which reproduce exponentials.

■ 2.4 Performance Limits of Signal Parameter Estimation

Throughout this thesis we make use of the Cramér-Rao bound (CRB) [21, 52, 89] to obtain lower bounds on the parameter estimation problems of interest. Let θ be a vector containing the parameters of interest, and let $\hat{\theta}$ be the estimate of this vector. Let the observations be given by $\mathbf{y}$, and the conditional distribution given by $p(\mathbf{y}; \theta)$. The Fisher information matrix is defined as,

$$\mathbf{J}(\theta) = \mathbb{E}\left[\left(\frac{\partial}{\partial \theta} \log p(\mathbf{y}; \theta) \right) \left(\frac{\partial}{\partial \theta} \log p(\mathbf{y}; \theta) \right)^T \right] = \mathbb{E}\left[\left(\frac{\partial}{\partial \theta} \log p(\mathbf{y}; \theta) \right)^2 \right]. \tag{2.14}$$

Cramér and Rao gave the following theorem.

Theorem 2.1 (Cramér-Rao estimation lower bound [21,89]). *The covariance matrix of an unbiased estimator $\hat{\theta}$ of a parameter θ can be bounded below by the inverse of the Fisher information matrix:*

$$\mathbb{E}[(\theta - \hat{\theta})(\theta - \hat{\theta})^T] \geq \mathbf{J}(\theta)^{-1}.$$

It is possible to consider biased estimators; however, simple forms of the CRB usually only exist for unbiased estimators. Moreover, an unbiased estimator does not always exist. Fortunately, in our case of interest it has been shown that the algorithms proposed are unbiased or asymptotically unbiased [20, 105, 106].

When the observations are subjected to additive white Gaussian noise, then the expression for the Fisher information matrix (2.14) can be simplified [52]. Let L be the log-likelihood function $L = \log p(y_n; \theta)$ and the parameterized noiseless signal model be given by $x[m; \theta]$. Then it can be shown that

$$\begin{aligned} \mathbf{J}(\theta) &= \left[\sum_n \left(\frac{\partial}{\partial \theta} L \right) \left(\frac{\partial}{\partial \theta} L \right)^T \right]. \\ &= \left[\sum_n \left(\frac{\partial}{\partial \theta} x[m;\theta] \right) \left(\frac{\partial}{\partial \theta} x[m;\theta] \right)^T \right]. \end{aligned}$$

In the general case it is difficult to compute the FIM and the CRB [52,111]. There has been some work on computing the FIM when the noise is colored, but most of the work in signal processing has been concentrated on noise which admits an auto-regressive model [39, 108].

Now we are ready to examine the problem of parameter estimation from a noisy powersum series as given in Definition 2.1. We observe

$$y_n = \sum_{k=0}^{K-1} c_k (u_k)^n + w_n, \quad n = 0, \ldots, N-1. \tag{2.15}$$

We are interested in estimating the coefficients $\{c_k\}$ and the roots $\{u_k\}$. We assume $c_k \in \mathbb{R}$ and either $u_k \in \mathbb{R}$ or $u_k \in \mathbb{C}$. The performance limits of estimation problems of form (2.15) have been studied before. We now review some of the results from [30, 103, 105, 106].

Let the signal be interest be of line spectrum form with additive noise

$$y_n = \sum_{k=0}^{K-1} \alpha_k \exp(j(\omega_k n + \varphi_k)) + w_n, \quad n = 0, \ldots, N-1. \tag{2.16}$$

The signal is given by its amplitudes α_k, phase terms φ_k, and frequencies ω_k. The poles of this signal all lie on the unit circle. When the noise w_n is additive white Gaussian noise (AWGN) with variance σ^2, for large numbers of samples N and $\{\omega_k\}$ spaced far from one another it can be shown that the Cramer-Rao bound on the estimates of ω_k is approximately [106, (4.3.9)]

$$\mathrm{var}(\hat{\omega}_k) \geq \frac{6\sigma^2}{N^3} \frac{1}{\alpha_k^2} \tag{2.17}$$

Several algorithms give performance close to this bound, such as the nonlinear least-squares

algorithm [108] and regularized maximum-likelihood algorithm [35]. However, these algorithms require very good initial conditions and perform poorly when the number of samples N is small. They are often simulated and implemented using initial values obtained from the Prony and Matrix Pencil methods.

Dilaveroğlu derived the Cramér-Rao bound for the estimation of two undamped complex exponentials in the presence of AWGN [30]. The expressions obtained were given in terms of the phase difference and frequency difference between the two complex exponential terms. He showed that the largest lower bound is obtained when the phase difference is zero. We will see that this case is of particular interest when we examine the performance of the FRI sampling scheme of Section 2.3.2. For large N, the bound he obtained is

$$\mathrm{var}(\hat{\omega}_k) \geq \frac{1}{4 \cdot \mathrm{SNR}_k \cdot N^3} \left(N \cdot \delta\omega\right)^{-4} + O\left(\left(N \cdot \delta\omega\right)^{-2}\right).$$

Several papers [54, 57, 105] considered a more general signal model given by

$$y_n = \sum_{k=0}^{K-1} \alpha_k \exp(j(\omega_k n + \varphi_k) + p_k n) + w_n, \quad n = 0, \ldots, N-1. \tag{2.18}$$

This signal has damping factors p_K, so the poles do not all lie on the unit circle. It is convenient to re-write the signal as,

$$y_n = \sum_{k=0}^{K-1} (\alpha_k \exp(j\varphi)) \exp((j\omega_k + p_k)n) + w_n, \quad n = 0, \ldots, N-1.$$

The poles of the signal have magnitude p_k and angle ω_k. It was shown in [105] that the Cramér-Rao bound of the parameters have dependence on some of the other parameters. Further, the Total Least-Squares (TLS) algorithm was shown to perform close to this bound, even for small sample sizes. It was shown empirically that the best frequency and phase estimation performance is obtained when the poles lie on the unit circle, *i.e.,* when the signal is undamped.

Line spectra estimation algorithms that do not consider the parametric representation directly are based on periodogram and the Fourier transform. The resolutions of these non-parametric algorithms are limited by the Fourier resolution limit, which is the size of the Fourier bins in the Fourier domain. It is known that algorithms such as MUSIC and ESPRIT (see Appendix A) have *super-resolution* properties, that is they are able to distinguish frequency terms that are spaced more closely than the Fourier resolution limit. Recent work by Dilaveroğlu [30] and Smith [103] give simplified expressions for the Cramér-Rao bound for the case when $K = 2$. In Chapter 3 we will adapt their methods to show similar bounds for our cases of interest and demonstrate that our proposed multi-channel acquisition system also has a super-resolution property. Then in Chapter 4 we will use the same technique to show that the previously proposed systems of Vetterli *et al.* [117] and Dragotti *et al.* [33] also have super-resolution properties.

Chapter 3

Multiscale Sampling of Parametric Signals

This chapter presents a novel analog-to-digital architecture for parametric signals that are sums of pulses with different amplitudes and delays, as we introduced in Chapter 1. Unlike previous methods, the proposed method takes simultaneous samples from multiple channels. The required hardware is very simple, and in the most basic cases the desired parameters can be estimated using very simple computations. The number of required samples is determined by the signal innovation within the interval of interest. In the noiseless case, we can obtain perfect reconstruction of the signal. We show that the proposed system has a successive refinement property that can be used to detect aliasing.[1] In the noisy case, we derive performance bounds and propose algorithms that are simple to implement and perform well. We make comparisons to conventional techniques that rely on the notion of Fourier bandwidth. When we examine the performance of the proposed scheme in the presence of additive noise, in many regimes it outperforms the classical method in terms of accuracy and resolution.

■ 3.1 Introduction

In many applications, the signals of interest are defined by a pulse shape, the number of components, and their amplitudes and delays. For example, consider the class of signals given by

$$x(t) = \sum_{k=0}^{K-1} c_k \delta(t - t_k), \quad c_k \in \mathbb{R}, t_k \in [0, T). \tag{3.1}$$

These signals are not bandlimited. However, any such signal is fully specified by K pairs of coefficients (c_k, t_k). Hence these signals have finite rate of innovation. We are interested in finding the parameters of the signal—its amplitudes and delays—given by $\theta = [c_0, \ldots, c_{K-1}, t_0, \ldots, t_{K-1}]^T$.

The conventional method for reconstructing signals of the form (3.1) is to first apply an anti-aliasing lowpass filter, sample the output at or above the Nyquist rate, and then use a correlator to find the locations and amplitudes of the signal components. The problem of signal location and amplitude estimation from uniformly-spaced samples in the presence of additive noise is a standard problem in estimation theory [52]. In the presence of AWGN,

[1] Part of this work will be presented at the 2006 IEEE Intl. Conf. on Image Processing, Atlanta, GA [58].

the optimal estimate of the signal is given by finding the K highest peaks at the output of a correlator to find the timing information, and then estimating the amplitudes based on the timing estimates. This amounts to a K-dimensional search. It is known that the objective function has many local maxima, and hence a gradient search algorithm requires a good initial estimate. The quality of this method is commensurate to the width of the sinc sampling filter used—the wider the bandwidth, the better the resolution in the presence of additive noise. Thus its performance depends on the available sampling rate. Further, the resolution of correlation-based methods is bounded below by the sampling rate. We will show that we can use *super-resolution techniques* to estimate the desired parameters, namely the time delays and amplitudes. We will show by simulation that the proposed system can outperform the resolution of classical methods in many regimes.

■ 3.2 Previous Work

Recall from Section 2.3.4 that Dragotti, Vetterli and Blu proposed reconstruction methods for infinite-length parametric signals using sampling filters that satisfy the Strang-Fix conditions, and uniform sampling in time. Following Definition 2.4, a filter that satisfies the Strang-Fix conditions can reproduce polynomials up to a certain order using superposition of its uniformly-spaced shifts. Some of these filters have finite support, enabling *local* reconstruction.

We have seen in Chapter 2 that previously studied sampling methods for reconstructing signals of form (3.1) given in [60,61,69,117] and [31–33] require the solution of a powersum series. All of these methods take uniform samples in time, and several algorithms are known: see Appendix A for an overview. In this chapter we present a new sampling method that is particularly simple to implement, has a simple reconstruction algorithm in its basic case, and samples *simultaneously in time* [58]. We compare its performance to that of the conventional method.

Multi-channel sampling for *bandlimited* signals was considered by Papoulis [84] and reviewed by Higgins [45]. However, they considered only filtered versions of the original signal, and drew connections to forming a Riesz basis from the filters. Their reconstruction methods are restricted to a *fixed* subspace. However, signals of form (3.1) do not form a subspace. Indeed, the subspace that the signal belongs to is determined by its pulse shifts $\{t_k\}$. Hence, perfect reconstruction using conventional methods is not possible. Throughout this paper we will focus on a specialized model of (2.1), in which any finite interval length T contains at most K Diracs.

The estimation of parameters in a *powersum series* has been studied primarily in the context of line spectra estimation [30,48,49,88,104–106]. In that problem, the poles are of form $\exp(j\omega_k)$—meaning that they are complex roots of unity and all lie on the unit circle—whereas in our case the desired parameters are real-valued. As a consequence, we cannot obtain simplifications that are possible in the line spectra case and resort to numerical evaluation.

■ 3.3 Proposed Method

Consider the following set of signals, which are obtained through successive integration as shown in Figure 3-2(b):

$$x_{\ell+1}(t) = \int_0^t x_\ell(\tau)\, d\tau, \quad x_0(t) = \int_0^t x(\tau)\, d\tau.$$

We are given samples of these signals at time T, denoted as $y_\ell = x_\ell(T)$. Then the samples can be written as:

$$y_\ell = x_\ell(T) = \frac{1}{\ell!} \sum_{k=0}^{K-1} c_k (T - t_k)^\ell, \quad \ell = 0, 1, \ldots, L-1. \tag{3.2}$$

We call the proposed approach the *integral sampling* system. It is possible to obtain the parameters $\{c_k, t_k\}$ from $L \geq 2K$ samples y_ℓ.

Theorem 3.1 (Multi-channel Sampling via Integrals). *Let the observed signal be $x(t) = \sum_{k=0}^{K-1} c_k \delta(t - t_k)$ where $t_k \in [0, T)$. Let $x_\ell(t)$ be the $(\ell+1)$-th integral of $x(t)$. We are given samples $y_\ell = x_\ell(T)$ for $\ell = 0, 1, \ldots, L-1$. Then it is possible to perfectly reconstruct $x(t)$ from $\{y_\ell\}_{\ell=0}^{L-1}$ as long as $L \geq 2K$. The algorithm can also be applied to non-contiguous integral powers of $x(t)$.*

Proof. Let $s_\ell = (\ell!) \cdot y_\ell$. Then we have that

$$s_\ell = \sum_{k=0}^{K-1} c_k (T - t_k)^\ell.$$

The sequence s_ℓ is a powersum series of the form (2.1); hence we can solve for $\{t_k\}$ and $\{c_k\}$ from $L \geq 2K$ samples y_ℓ. □

Let $h_\ell(t) = (T-t)^\ell 1_{[0,T)}$, where $1_{[t_a,t_b)}$ is the indicator function that is 1 within $[t_a, t_b)$ and zero elsewhere. The integrate-and-sample operation can be written as inner products between the input signal and functions given by $h_\ell(t)$, which we call the sampling kernel. It can be seen that

$$\begin{aligned} s_\ell &= \langle x(t), h_\ell(t) \rangle \\ &= \sum_{k=0}^{K-1} c_k h_\ell(t_k) \\ &= \sum_{k=0}^{K-1} c_k (T - t_k)^{\ell-1}. \end{aligned} \tag{3.3}$$

We illustrate this in Figure 3-1.

The above operation can be implemented by using a chain of integrators $x_\ell(t) = \int_0^t x_{\ell-1}(\tau)\, d\tau$ and $x_0(t) = \int_0^t x(\tau)\, d\tau$, or in parallel as shown in Figure 3-5. It is also easy to

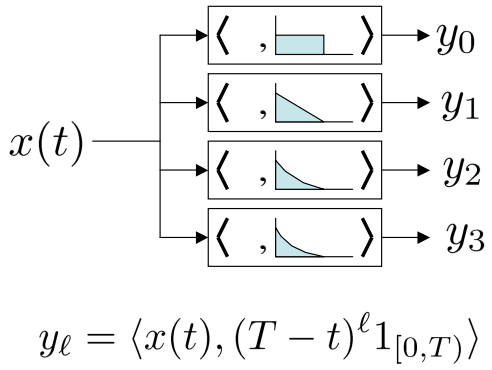


Figure 3-1. The proposed multichannel sampling scheme as a sequence of projections onto monomials of different orders.

extend this scheme to an infinite-length signal by segmenting, as we obtain measurement from local portions of the signal.

■ 3.3.1 Sampling using B-spline kernels

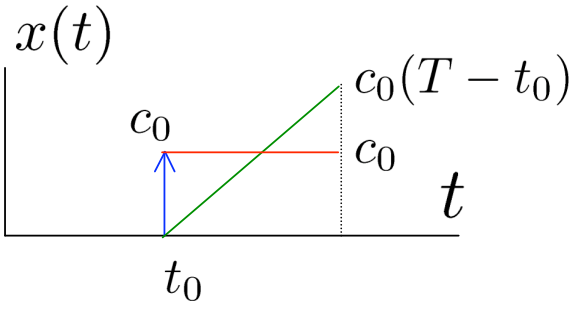
It is also possible to use B-spline kernels of various orders for sampling. Let $\beta_0(t)$ be the 0-th order box B-spline, defined as $\beta_0(t) = u(t) - u(t-1)$, where $u(t)$ is the usual step function. The higher-order B-splines can be generated recursively as $\beta_{k+1} = \beta_k * \beta_0$. Consider the following set of observations:

$$y_\ell = \langle \beta_\ell(t), x(t) \rangle, \quad \ell = 0, 1, \ldots, L-1.$$

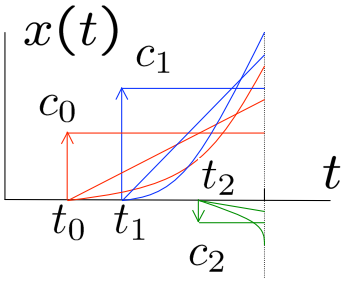
Since the first segment of a B-spline is a monomial of the order of the spline, sufficiency is achieved when $L \geq 2K$ just like in the previous case. It is important to note that the samples above are taken all at the same time, but across different splines scales. This is different from the spline interpolation scheme which takes samples within one spline scale but at different, uniformly spaced time instances. Similarly, the B-spline sampling scheme also has a successive refinement property.

Theorem 3.2 (Multi-channel Sampling via B-splines.). *Let the observed signal be* $x(t) = \sum_{k=0}^{K-1} c_k \delta(t - t_k)$ *where* $t_k \in [0, T)$. *Without loss of generality, let* $T = 1$. *Further let* $y_\ell = \langle \beta_\ell(t), x(t) \rangle, \ell = 0, 1, \ldots, L-1$, *the samples taken at time* $t = T$ *at the outputs of B-spline filters of different orders* ℓ. *Then it is possible to perfectly reconstruct* $x(t)$ *from* $\{y_\ell\}_{\ell=0}^{L-1}$ *as long as* $L \geq 2K$. *It is also possible to solve from* L *non-contiguous B-spline-filtered samples.*

The system proposed above can be implemented by a chain of β_0 filters, similar to the chain of integrators of the previous part. The sampling scheme proposed above can also be applied to signals with different numbers of Diracs at different segments, allowing for lower-order local reconstruction of the signal. However, the knots of the splines are at integer multiples of T, and hence the segments must all have the same length.



(a) Sampling of one Dirac within a time interval.



(b) Sampling of multiple Diracs within a time interval by successive integration and simultaneous sampling across the different integral levels.

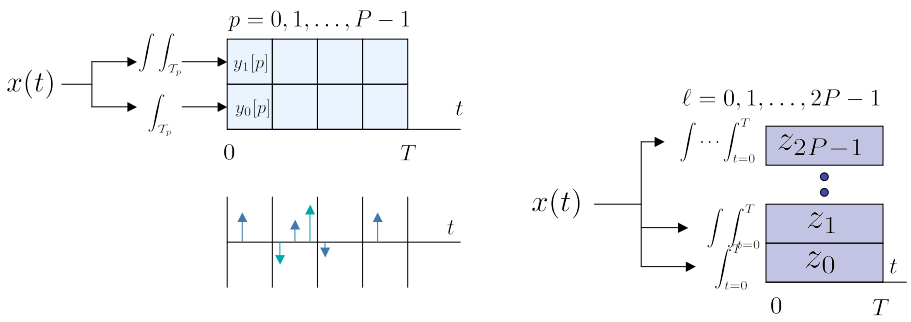
Figure 3-2. Sampling a single Dirac, and its generalized counterpart. The samples are taken simultaneously at $t = T$.

■ 3.4 Successive Refinement and Detecting Undermodeling

The sampling schemes proposed above have a *successive refinement* property. Suppose that we know that there is a finite but unknown number K^* of Diracs within an interval $[0, T)$. We can start by taking 2 samples and reconstructing 1 Dirac, and compare the results with that obtained by taking another 2 samples and reconstructing 2 Diracs. If the results are consistent, then we stop the sampling operation. Otherwise, having reconstructed L Diracs from $2L$ samples we take another 2 samples and compare the earlier reconstruction with that for $L + 1$ Diracs from the larger set of $2L + 2$ samples. We stop when the result is consistent with that of the previous reconstruction. In the presence of noise or model mismatch, a stopping criteria other than exact reconstruction can be defined.

We have seen that reconstructing 1 Dirac in a finite interval using integral-based sampling can be done using a simple computation. If we are able to partition a signal $x(t)$ into segments such that there is at most 1 Dirac in each segment, then we can use the simpler algorithm on each segment, taking 2 samples in each segment—basically a divide-and-conquer approach. Consider P such consecutive segments $\mathcal{S}_p = [T_p, T_{p+1})$, yielding say signal model A. A more general signal model is one which contains P Diracs anywhere within the union of those segments, say $[0, T)$, which we call signal model B. In both cases we have to take at least $2P$ samples, but these samples are not the same for both cases.

Signal model A is more restrictive than B, but gives a simpler reconstruction formula. In multichannel sampling it is possible to detect undermodeling, for example when we mistakenly assume signal model A. Consider the reconstruction algorithms for models A and B respectively, shown in Figure 3-3:



(a) System A (b) System B

Figure 3-3. Two systems with different reconstruction orders. System A is suited to a more restrictive signal model than System B.

1. **System A:** Obtain samples $y_0[p] = \int_{\mathcal{T}_p} x(t)dt$, $y_1[p] = \int_{\mathcal{T}_p}^{T_{p+1}} \int_{\mathcal{T}_p}^{t} x(t_1)dt_1dt$. From each p use $y_0[p]$ and $y_1[p]$ to compute estimates $\hat{c}_p, \hat{t}_p$.
2. **System B:** Obtain samples $z_\ell = \langle x(t), s_\ell(t/T)\rangle$ the ℓ-th integral of $x(t)$ sampled at $t = T$, for $\ell = 0, \ldots, 2P-1$. Use these samples to simultaneously obtain estimates $\{\hat{c}_p, \hat{t}_p\}_{p=0}^{P-1}$.

Let $\{c_p, t_p\}$ be the parameters of the Dirac in segment $\mathcal{S}_p$. When there is no undermodeling we can derive $\{z_\ell\}$ from $\{y_0[p], y_1[p]\}$ by first computing estimates $\{\hat{c}_p, \hat{t}_p\}$. From these estimates we can compute an estimate of what the samples of System B should be based on the samples of System A which we call $\{\hat{z}_\ell\}$. We give the following lemma.

Lemma 3.1 (Consistency for lower-order estimates). *Consider the two systems A and B defined above, giving samples $\{y_0[p], y_1[p]\}$ and $\{z_\ell\}$ respectively. We say that these samples are consistent when we can derive the latter from the former via signal parameter estimation. Let $\{\hat{c}_p, \hat{t}_p\}$ be the signal parameter estimates from samples of System A, and let $\{\hat{z}_\ell\}$ be the estimate of what the samples of System B should be, given the parameter estimates from A. Then for $\ell = 0, 1$ we always have consistency, whereas for $\ell \geq 2$ the samples of the two systems are consistent only if there is no undermodeling.*

Proof. Consider signal model A, and without loss of generality let every segment have 1 Dirac except for segment-a which has two Diracs. We denote this Dirac $\{c_u, t_u\}$. Then for $p \neq a$ the estimates obtained from A are correct, but for $p = a$ we have that $\hat{c}_a = c_a + c_u$, $\hat{t}_a = T_a - (c_a(T_{a+1} - t_a) + c_u(T_{a+1} - t_u))/(c_a + c_u)$. Since $\hat{z}_\ell = \sum_p \hat{c}_p(T - \hat{t}_p)^\ell$, we have that $z_0 = \hat{z}_0$ and $z_1 = \hat{z}_1$, but $z_\ell \neq \hat{z}_\ell$ for $\ell \geq 2$. □

The lemma can be stated more simply: for $\ell = 0, 1$ we always have consistency regardless of whether there is undermodeling. For $\ell > 1$ we have consistency only if there is no undermodeling. We summarize this in the table below.

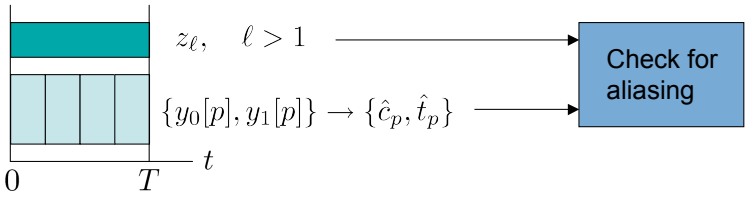


Figure 3-4. Checksum system for detecting aliasing in an FRI sampling system.

	$\ell = 0, 1$	$\ell > 1$
No undermodeling	Consistent	Consistent
Undermodeling	Consistent	Not consistent

This property can be used to detect aliasing, by considering a higher-order integral to act as a checksum.

Theorem 3.3 (Detecting aliasing in low-order FRI signals). *Let $\mathcal{S}_p$, $p = 0, \ldots, P-1$ be a P-segmentation of a signal $x(t)$ such that we expect at most 1 Dirac in each segment. From each segment, we compute $y_0[p] = \int_{\mathcal{S}_p} x(t)\,dt$ and $y_1[p] = \int_{\mathcal{T}_p}^{T_{p+1}} \int_{\mathcal{T}_p}^{t} x(t_1)dt_1dt$, from which we can compute the estimate of one Dirac in each $\mathcal{S}_p$. Then it is possible to detect aliasing in any of the P segments by taking one additional sample from a higher-order integral of the signal.*

The theorem follows immediately from Lemma 3.1. The method for detecting aliasing is analogous to computing a checksum based on z_ℓ, for $\ell > 1$, as shown in Figure 3-4. The lemma also suggests that we can "recycle" the samples already taken from system A for reconstruction based on samples taken from system B, which admits a less restrictive signal model. The lemma and theorem can be applied to any system that uses powersums for reconstruction, and we will see that they are an attractive feature for a previously proposed sampling scheme based on finite kernels that we reviewed in Section 2.3.4. We will discuss this in detail in Section 4.2. Finally, we note that these properties also apply to sampling with B-splines.

■ 3.5 Multiscale Signal Acquisition in the Presence of Noise

We now examine the performance of the proposed system in the presence of noise. Recall from Section 3.3 that the parameters of the signal are obtained from a powersum series with real-valued coefficients. Real exponential fitting is one of the most important, difficult and frequently appearing problems of applied data analysis [104]. Moreover, the original Prony method is not constrained to real-valued solutions. In order to do this, nonlinear least-squares methods have been proposed [86]. However, these methods have high complexity and require large sample sizes to perform well. We instead focus on algorithms that work well for small numbers of samples.

A method and analysis for solving (2.1) in its most general form (complex magnitudes and complex poles) was presented in [105]. The authors showed that for large sample sizes

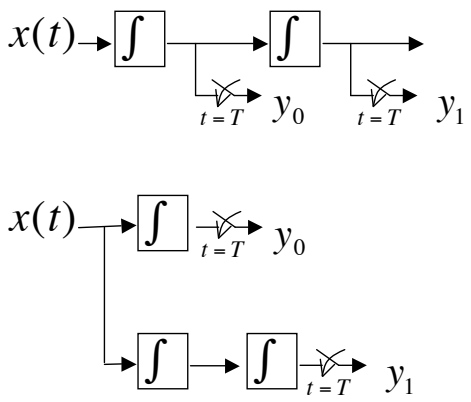


Figure 3-5. Illustration of the multichannel sampling scheme, implemented using integrators and samplers. The samples are taken simultaneously in time at different integral levels of the input signal.

the parameter estimates approach an unbiased normal distribution to a first order, and that the estimation of the radii of the poles is best when the poles are on the unit circle (see their Figure 2). Another algorithm of interest is Cornell's algorithm [20], which is based on partial sums. Cornell gave simple formulas for the cases of $K = 1$ and $K = 2$. We review the algorithms in Appendix A.

It is important to note that unlike the case of uniform sampling in time, when we take our samples across scales the nature of noise depends on the implementation. Two of the possible architectures are shown in Figure 3-5. When using cascaded filters, we may have noise propagation and the noise is no longer independent nor identically distributed across the samples anymore. Therefore we need to consider particular implementations of the system to obtain a noise model. Regardless of the statistics of the noise, we obtain

$$s_\ell = \sum_{k=0}^{K-1} c_k (T - t_k)^\ell + w_\ell, \quad \ell = 0, 1, \ldots, L-1, \tag{3.4}$$

where w_ℓ is the additive noise term.

Throughout this section we will use the model of noisy integrate-and-dump circuit given in Figure 3-6. In this model, there are two sources of noise in the integrator circuit: the input noise $n_I(t)$ and output noise $n_O(t)$. Both of these noise terms are white Gaussian, but clearly although the effect on the output signal will be Gaussian, it will not be white in general. As the prototype case, consider zero-mean white Gaussian sequence $x(t)$ with variance σ_x^2 filtered by $h(t)$ and its output $y(t)$ sampled at $t = T$, say $Z = y(T)$. Then clearly Z is Gaussian, $\mathbb{E}[Z] = 0$ and $\mathbb{E}[Z^2] = \sigma_x^2 \int h^2(t)dt$.

We consider two implementations: by concatenating integrators in series and taking simultaneous samples at the output of the different integrators, or by implementing the integrators in parallel. We show the different implementations in Figure 3-5.

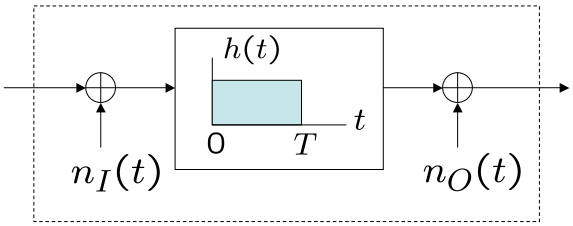


Figure 3-6. Model of a noisy integrator. There are two noise terms $n_I(t)$ and $n_O(t)$ at the input and output of the integrator respectively. Both these noise sources are white Gaussian.

■ 3.5.1 Conventional method: filtering and Nyquist sampling

For comparison, consider the conventional method of first applying an anti-aliasing filter and uniform sampler, and then estimating the signal pulse delay from these samples. The problem of delay estimation in the presence of AWGN from uniformly samples is a well-known estimation problem [52]. 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 given by:

$$\mathrm{var}(\hat{\tau}) \geq \frac{1}{\mathrm{SNR} \cdot F^2}, \tag{3.5}$$

where

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

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.

■ 3.5.2 Performance limits of proposed system

We derive the Cramér-Rao bound (CRB) for several basic cases of interest, in the presence of AWGN. We consider the powersum series of form $y_n = \sum_{k=0}^{K-1} c_k (u_k)^n + w_n$. In this chapter we are interested in the real-valued case, for $u_k = T - t_k$. For convenience, we define θ to be the vector of desired parameters

$$\theta = [c_0, \ldots, c_{K-1}, t_0, \ldots, t_{K-1}]^T.$$

Single component

We start with the case when $K = 1$. Suppose that we observe $s_n = c_0(u_0)^n + w_n$, where w_n is additive white Gaussian noise (AWGN). Then we show in Appendix B.1 that the Cramér-Rao Bound for the estimation of time delay is given in the following theorem.

Theorem 3.4 (Performance limit of single-component real powersum in AWGN). *Suppose that we are given $s_n = c_0(T - t_0)^n + w_n$, for $n = 0, 1, \ldots, N-1$, where w_n is*

additive white Gaussian noise with variance σ^2. Let $u_0 = T - t_0$ and $SNR = c_0^2/\sigma^2$. Then the Cramér-Rao bound for the estimation of t_0 is given by

$$\mathbb{E}[(\hat{t}_0 - t_0)^2] \geq \frac{u_0^2}{SNR \cdot N^3} \left(\frac{G_0(u_0^2)}{G_0(u_0^2)G_2(u_0^2) - G_1(u_0^2)G_1(u_0^2)} \right), \tag{3.6}$$

where $G_r(x)$ is given by

$$G_r(x) = \frac{1}{N^{r+1}} \sum_{n=0}^{N-1} n^r (x)^n. \tag{3.7}$$

The behavior of (3.6) is consistent with the behavior of frequency estimates in line spectra estimation which we reviewed in Section 2.1. However, a caveat is in order here. Suppose for now that the signal of interest consists of complex-valued poles. This case was studied by Steedly and Moses in [105]. 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. This is a different behavior from what we obtained in this section, where we observe that the larger the magnitude of the pole, the better its performance. This illustrates that translating the results from the complex-valued case to the real-valued case is not straightforward and can be misleading.

Two components

Now consider the case when $K = 2$. Then $s_n = c_0(u_0)^n + c_1(u_1)^n + w_n$, where w_n is additive white Gaussian noise. For convenience, let

$$\mathbf{A} = \begin{bmatrix} G_0(u_0^2) & G_0(u_0u_1) \\ G_0(u_0u_1) & G_0(u_1^2) \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} G_1(u_0^2) & G_1(u_0u_1) \\ G_1(u_0u_1) & G_1(u_1^2) \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} G_2(u_0^2) & G_2(u_0u_1) \\ G_2(u_0u_1) & G_2(u_1^2) \end{bmatrix},$$

where $G_r(x)$ is as given in (3.7). Then we show in Appendix B.1 that the Cramér-Rao bound can be found via the Schur complement:

Theorem 3.5 (Performance limit of two-component real powersum in AWGN). *Let there be two Diracs in $[0, T)$. Consider the estimation of t_k using the proposed method when the N observations are subjected to AWGN with RMS power σ. Let $u_k = T - t_k$ and $SNR_k = c_k^2/\sigma^2$. Then the CRB is given by,*

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

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

■ 3.5.3 Performance comparison

Many methods have been proposed for solving powersums of (2.1). We focus our attention on two methods: one proposed by Cornell [20], roughly speaking based on the rotational invariance property and one by Rahman and Yu [88,105] which uses the Total-Least Squares (TLS) approach applied to Prony's method.

Cornell's method is particularly simple to implement for the case of a single component: it uses N additions and 2 divisions. On the other hand, the TLS-based method in general

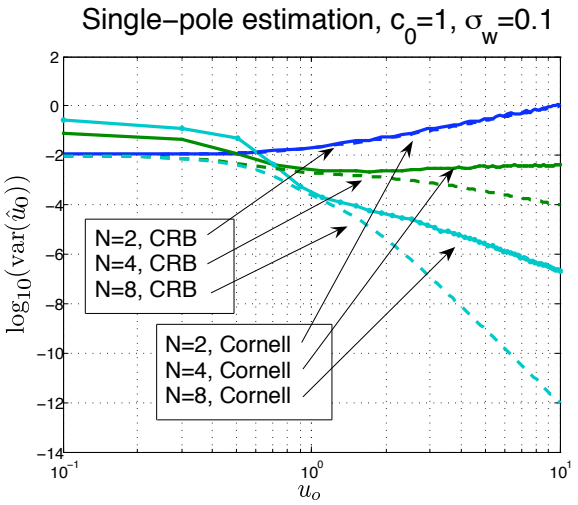


Figure 3-7. Comparison between Cramér-Rao bound and performance of Cornell algorithm in the presence of AWGN.

uses singular-value decomposition, but is known to perform well for situations in which the SNR is very low.

White powersum noise

Now we begin with the simplest model: each sample for both the conventional and the proposed systems are subjected to AWGN with variance σ^2. We show the performance comparison in Figure 3-7 for the Cornell method and in Figure 3-8 for the TLS method. On the x-axis of both figures we change the value of the root u_0 from 0.1 to 10. When the root is less than unity, both algorithms perform close to the Cramér-Rao bound. However, as the root is increased the TLS-based algorithm shows significantly better performance than the Cornell algorithm and is very close to the derived bound. Interestingly, for $N = 2$ the Cornell algorithm is better than the TLS-Prony algorithm, whereas for larger values of N the opposite is true.

Next we compare the performance of the system to the classical method which takes only few samples in time, after using a low-pass filter appropriate for the given sampling rate. We show the results in Figure 3-9. For the conventional system, we normalize the sampling period to be $1/N$, with a single Dirac randomly placed within $t \in [0, 1)$. The conventional system is implemented by using a gradient search at the output of a cross-correlator. For the case when $N = 2$ we note that the performance the same regardless of the value of the desired parameter $u_0 = (T - t_0)$. For larger values of N the performance improves for u_0 near unity and above unity. Roughly speaking, the proposed system is commensurate to the performance of the classical system at the cost of 10 dB of SNR requirement. However, the proposed system operates at $1/N$ the sampling rate of the conventional system. Moreover, implementing a full-search on the output of a cross-correlator can be expensive.

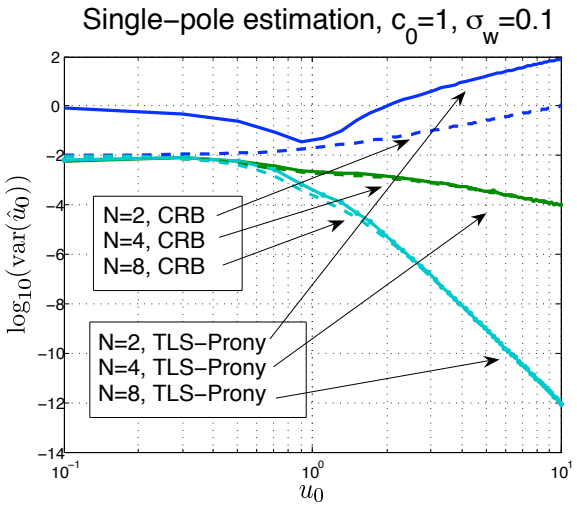


Figure 3-8. Comparison between Cramér-Rao bound and performance of TLS algorithm in the presence of AWGN.

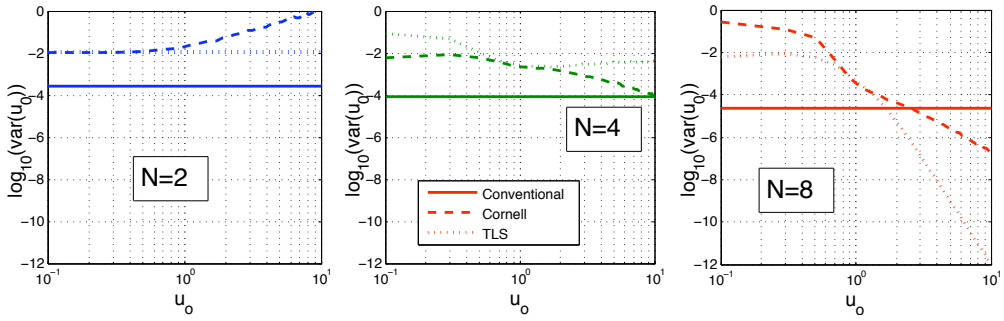


Figure 3-9. Comparison between the performance of the conventional sampling system, Cornell algorithm, and the TLS algorithm, for different numbers of samples within a segment, and reconstruction of only one Dirac in the presence of AWGN. In this case, $c_0 = 1$ and $\sigma_w = 0.1$.

White continuous-time input noise

The previous comparison uses a primitive noise model and it is difficult to make any quantitative statement in comparing the proposed system to the conventional system. We now use a different noise model that takes into account the system architectures and sampling methods. Let $w(t)$ be white Gaussian noise with spectral density $N_0/2$, and let the continuous-time signal be:

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

Classical system: Suppose we sample every T_s, and the sampling filter used is $\varphi(t) = \text{sinc}(t/T_s)$. Then the noise per sample is:

$$\begin{aligned}\sigma_c^2 &= \mathbb{E}\left[(\langle\varphi(t), w(t)\rangle)^2\right] \\ &= \frac{1}{2\pi}\int_{-\pi/T_s}^{+\pi/T_s} T_s^2\frac{N_0}{2}d\omega \\ &= \frac{N_0}{2}T_s.\end{aligned}$$

The observations are then given by:

$$y_c[n] = \text{sinc}(nT_s - t_0) + w_c[n], \tag{3.8}$$

where $w_c[n] \sim \mathcal{N}(0, \sigma_c^2)$. We take N samples within the signal segment $[0, T)$, so we set $T_s = T/N$ for fairness and obtain $\sigma_c^2 = N_0/2 \cdot T/N$.

Proposed system: In this case we compute

$$s_\ell = \langle x(t), h_\ell(t)\rangle = c_0(T - t_0)^\ell + w_\ell,$$

where $h_\ell(t) = t^\ell 1_{[0,T)}$ as given before. The noise output variance is then

$$\begin{aligned}\sigma_p^2[\ell] &= \int h_\ell^2(t)N_0/2dt \\ &= N_0/2\int_0^T t^{2\ell}dt \\ &= N_0/2\frac{1}{2\ell+1}T^{2\ell+1}.\end{aligned}$$

Since the noise source is in the original signal, the noise terms will be correlated. The covariance matrix is given by

$$\mathbb{E}[w[\ell]w[k]] = \frac{N_0}{2}\frac{1}{k+\ell+1}T^{k+\ell+1}.$$

In this case there is no difference in the noise covariance regardless of the implementation, because the only noise present in the system is in the continuous-time input signal. We simulate and show the results for the TLS-based algorithm in Figure 3-10 and the Cornell algorithm in Figure 3-11. The algorithms that we implement are exactly the same as that used in the previous part; they are not cognizant of the correlation in the system. In this

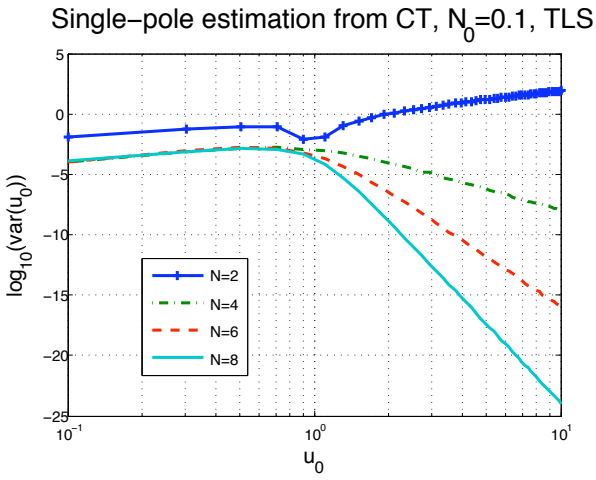


Figure 3-10. Performance of the TLS-based algorithm when the noise is in continuous-time with spectral density $N_0/2 = 0.1$ and $c_0 = 1$.

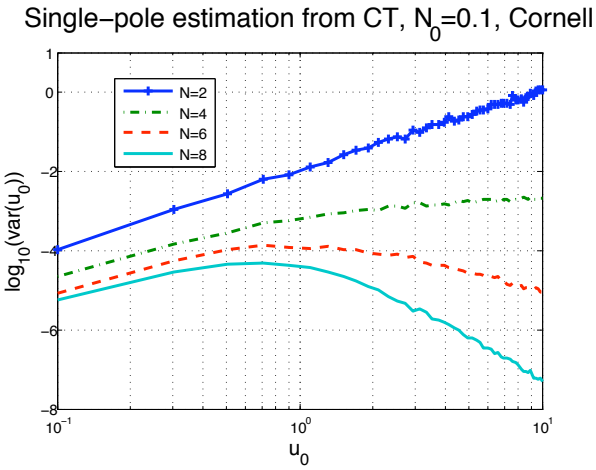


Figure 3-11. Performance of the Cornell algorithm when the noise is in continuous-time with spectral density $N_0/2 = 0.1$ and $c_0 = 1$.

case the performance difference between the proposed system and the classical system (see Figure 3-9) is very small, except for the case when $N = 2$. For small values of u_0 there is not much difference between the two proposed methods. The computation of the FIM with colored noise is not straightforward. Several methods have been proposed, but deal mainly with auto-regressive noise models. We refer the reader to [39] and the references therein.

Noisy components

Another noise model that can be used is one in which each component has a white continuous-time additive Gaussian noise at its *input*, similar to that shown in Figure 3-6. With this model in mind, we propose several possible implementations of the system. We show these in Figure 3-12. The output noise terms will be correlated, and we derive the covariance matrices in the following.

In our reconstruction algorithms, we do not compensate for the correlation in the noise terms of the samples. We examine the proposed signal acquisition system given in Chapter 3. The proposed system can be implemented in various ways. In this section we consider the serial and parallel implementations of the proposed system. We assume that the input noise terms are at the input interface of the integrator blocks as shown in Figure 3-6, and that they are zero-mean white Gaussian. Since the operations on the samples are linear, the resulting sample noise terms are zero-mean Gaussian and can be characterized completely by their covariance matrix.

Following the previous part, let $h_m(t) = t^m 1_{[0,T)}$ be the sampling kernels. Consider the prototype case in which a white Gaussian noise signal $x(t)$ with zero mean and variance σ_x^2 is filtered by kernels $h_m(t)$ and $h_n(t)$ and sampled at $t = T$ to give samples y_m and y_n respectively. Then clearly y_m, y_n are both zero-mean Gaussian, and

$$\mathbb{E}[y_m^2] = \sigma_x^2 \int h_m^2(t)dt, \quad \mathbb{E}[y_n^2] = \sigma_x^2 \int h_n^2(t)dt.$$

Moreover,

$$\mathbb{E}[y_m y_n] = \sigma_x^2 \int h_m(t)h_n(t)dt = \frac{\sigma_x^2}{m+n+1} T^{(m+n+1)}. \tag{3.9}$$

We are now ready to derive the covariance structures for the different implementations.

- Parallel implementation

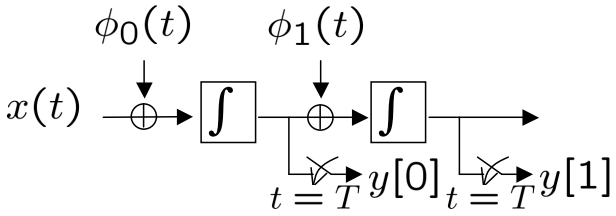
  We begin by considering the case when the integrate-and-dump components which yield the different integral levels of the input signal are implemented in parallel. The only noise in the system is introduced within each branch independently. In this case, we obtain *iid* noise terms at each sample output $w[\ell]$. The performance of algorithms for solving powersum equations in the presence of AWGN is well-studied [71, 106].

  We write the finite integration and sampling as convolution with $h(t) = 1_{[0,T)}$ and sampling at $t = T$. Since convolution is linear and time invariant, then given that the input noise is Gaussian, the distribution at the output in still Gaussian. The convolution will make the output signal correlated, but since we are only interested in the sampled output, we only need to compute its variance.

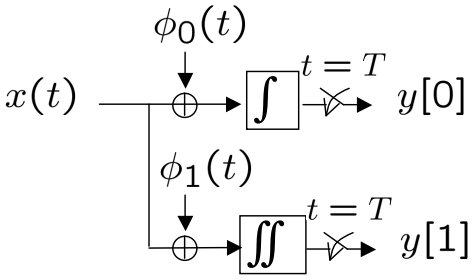
  Let w_ℓ be the output of the ℓ-th level integrator, and let $h_\ell(t) = t^\ell 1_{[0,T)}$. The sampled output can then be written as

  $$w_\ell = \big(w_{\ell,i}(t) * h_\ell(t)\big)\big|_{t=T} + w_{\ell,o}, \tag{3.10}$$

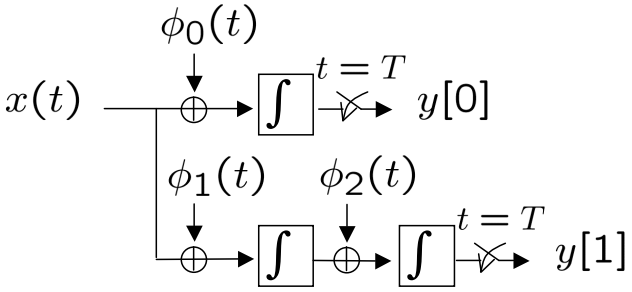
  where we separate the input and output noise at each channel as $w_{\ell,i}(t)$ and $w_{\ell,o}(t)$ respectively. Clearly, the contributions of these two terms are both Gaussian and zero



(a) Serial implementation



(b) Parallel implementation with noise in each branch



(c) Parallel implementation with noise in each first-order integrator

Figure 3-12. Three possible implementations of the proposed system. Each component has a white continuous-time additive Gaussian noise term, independent of the others.

mean. For simplicity, we consider only input noise terms. It is easy to derive that the variance of the first term is

$$\sigma_{\ell,i}^2 \int_0^T h_\ell^2(t)dt = \frac{\sigma_{\ell,i}^2}{2\ell+1} T^{2\ell+1}.$$

- Parallel System A

  Now consider the system shown in Figure 3-12 (b). For simplicity, we consider only *iid* input noise terms with variance σ^2, since the characterization of the output terms is very similar. Let the sample noise terms be $w[\ell]$. By virtue of the implementation, the noise terms are zero-mean Gaussian and uncorrelated. The variances are given by:

$$\mathbb{E}[w_\ell^2] = \sigma^2 \frac{1}{2\ell+1} T^{2\ell+1}.$$

- Parallel System B

  For the system shown in Figure 3-12 (c), we model each ℓ-th order integrator as a cascade of ℓ first-order integrator, each with its own input noise terms. We again model the noise terms as *iid* zero-mean Gaussian with variance σ^2. In this model the sampled noise terms are still zero-mean uncorrelated Gaussian, and the variances are given by:

$$\mathbb{E}[w_\ell^2] \quad = \quad \sigma^2 \sum_{k=0}^{\ell} \frac{T^{2(\ell-k)+1}}{2(\ell-k)+1}. \tag{3.11}$$

- Serial implementation

  Unlike in parallel implementation, when the system is implemented in series there may be noise propagation depending on where noise is introduced into the system. However, modeling the source and propagation of noise in circuitry is a difficult research area. We use a simplified model shown in Figure 3-12 (a).

  In our system, because the noise terms are zero-mean and jointly Gaussian, they can be characterized completely by their covariance matrices. The system is implemented by a concatenation of L integrate-dump circuit blocks. Let the input noise terms at block-ℓ be $\phi_\ell(t)$ with variances σ_ℓ^2 and the output noise terms be $\chi_\ell(t)$ with variances ρ_ℓ^2. All noise terms are zero-mean white Gaussian sequences.

  Let z_ℓ be the noise terms at the output of the ℓ-th sampler. We divide this into w_ℓ which is the noise due to the input noise terms $\phi_k(t)$, and v_ℓ which is due to the output noise terms $\chi_k(t)$. For convenience, let $h_k(t) = (T-t)^k 1_{[0,T)}$ represent the integrate-and-dump operation.

Without loss of generality let $m \leq n$. It can be shown that

$$\begin{aligned}
\mathbb{E}[w_m w_n] &= \sum_{k_1=0}^{m}\sum_{k_2=0}^{n}\int_{t_1}\int_{t_2}\mathbb{E}[\phi_{k_1}(t_1)\phi_{k_2}(t_2)]h_{m-k_1}(t_1)h_{n-k_2}(t_2)dt_1dt_2 \\
&= \sum_{k=0}^{m}\sigma_k^2\int h_{m-k}(t)h_{n-k}(t)dt \\
&= \sum_{k=0}^{m}\frac{\sigma_k^2}{m+n-2k+1}T^{m+n-2k+1}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\mathbb{E}[v_m v_n] &= \sum_{k_1=0}^{m}\sum_{k_2=0}^{n}\int_{t_1}\int_{t_2}\mathbb{E}[\phi_{k_1}(t_1)\phi_{k_2}(t_2)]h_{m-k_1-1}(t_1)h_{n-k_2-1}(t_2)dt_1dt_2 \\
&= \sum_{k=0}^{m}\sigma_k^2\int h_{m-k-1}(t)h_{n-k-1}(t)dt \\
&= \sum_{k=0}^{m}\frac{\sigma_k^2}{m+n-2k-1}T^{m+n-2k-1}.
\end{aligned}$$

We show the results in Figure 3-13 for the serial implementation. The performance of the system when it is implemented in parallel is shown in Figures 3-14 and 3-15. In the former, there is only one input noise term for each branch. In the latter, there is an input noise term for each first-order integrator. Clearly, the performance of the former is better than the performance of the latter. Interestingly, for $N = 2$ the Cornell algorithm is better than that of the TLS-Prony algorithm. For larger values of N we notice the opposite: as N is increased the TLS-Prony algorithm is superior to the Cornell algorithm. Note that the Cornell algorithm cannot be made aware of what the distribution of the noise is, it only requires that the noise is zero-mean. It is notable that the performance of both algorithms is not very sensitive to the correlation structures.

■ 3.5.4 Resolution limit

When there is more than one component in the signal but the spacing is large, then it is well-known that the Cramér-Rao bound and the performance are similar to that of the single-component case. We now examine the case when the two components are closely spaced. The line-spectra analog has been studied before, see for example [30, 103, 109]. In those cases, it is known that there is a strong dependence on phase offset between the two components, something which does not apply to our case.

One advantage of using super-resolution techniques is the ability to increase the resolution of the estimation method beyond the sampling rate [106]. Smith proposed that the minimum requirement to resolve two signals is that

$$\text{RMS of source separation} \leq \text{source separation}. \tag{3.12}$$

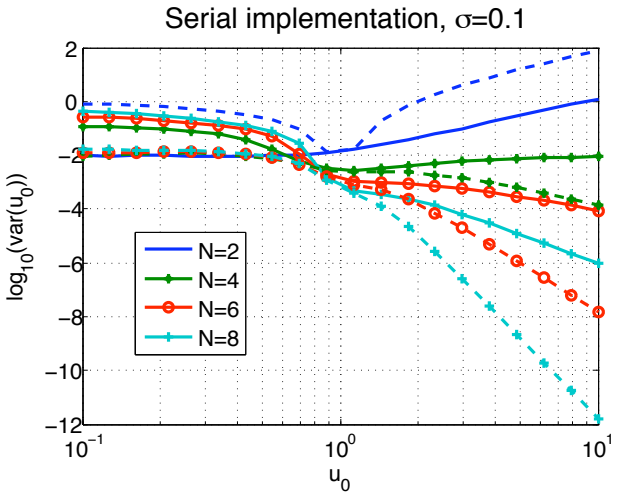


Figure 3-13. Performance of the TLS (dashed) and Cornell (solid) algorithms when each noise term has RMS power $\sigma = 0.1$. The system is implemented in series.

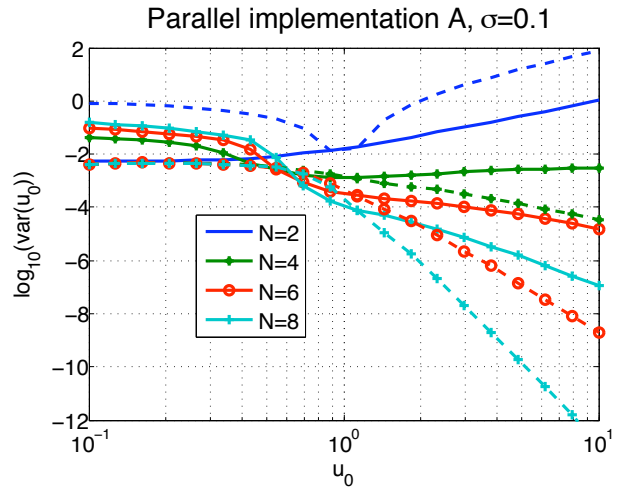


Figure 3-14. Performance of the TLS (dashed) and Cornell (solid) algorithms when each noise term has RMS power $\sigma = 0.1$. The system is implemented in parallel, and each branch has one noise term.

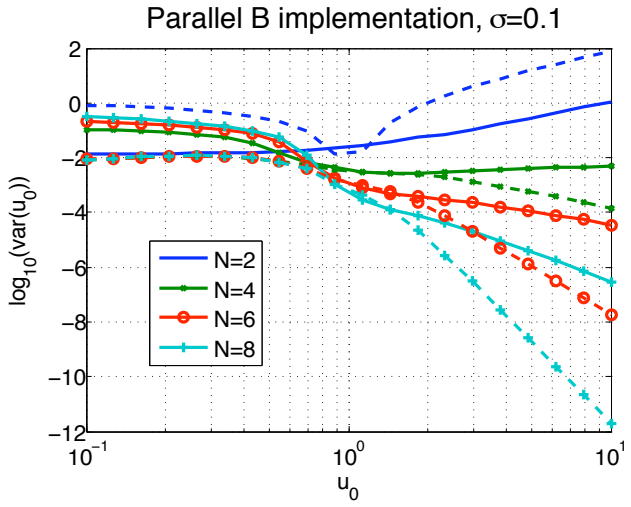


Figure 3-15. Performance of the TLS (dashed) and Cornell (solid) algorithms when each noise term has RMS power $\sigma = 0.1$. The system is implemented in parallel, and each first-order integrator has one noise term.

He then defines the *statistical resolution limit* as the source separation at which (3.12) 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 desired parameters be $\theta = (u_0, u_1)^T$. We are interested in how the estimate of θ depends on $\delta u = |u_0 - u_1|$. Since the roots that we are interested in are not complex roots of unity, we cannot obtain the simplification in [30], where CRB expressions that depend only on the separation between the two poles are obtained. The performance that we derive is shown in Figures 3-16, 3-17 and 3-18. The performance of our system depends on the actual location of the roots. In Figures 3-16 and 3-17 we compare the performance of the system with the classical method, shown as the line with downward slope. The dashed lines correspond to the actual spacing between the roots. It can be seen that in some cases, the performance of the proposed system exceeds the resolution limit of the classical system. In Figure 3-18 we show the dependence of the performance on the actual locations of the roots. The Cornell algorithm for $K = 2$ show performance that is superior to that of the TLS-based algorithm. We show this in Figure 3-19. When the spacing is very small, the Cornell algorithm starts to give complex-valued answers which is not valid. We call this a failure event. For the cases we show, the success rate is close to 1. There are other techniques that constrain the answer to lie on the real line, or the positive part of the real line [86]. However, these techniques are not straightforward to implement. From the simulation result we can see that in some regimes the performance of the proposed system exceeds the resolution limit of the conventional system, hence the proposed system has a super-resolution property.

We note certain trends in Figures 3-20 and 3-21. The performance of the system decays proportionately with decrease in spacing, whereas the resolution limit of the conventional method behaves the opposite. Moreover, the performance increases with SNR. From Figure 3-20 it can be seen that the relationship is $\mathrm{var}(u_0) \propto \mathrm{SNR}^{-1}$. We note that this result is

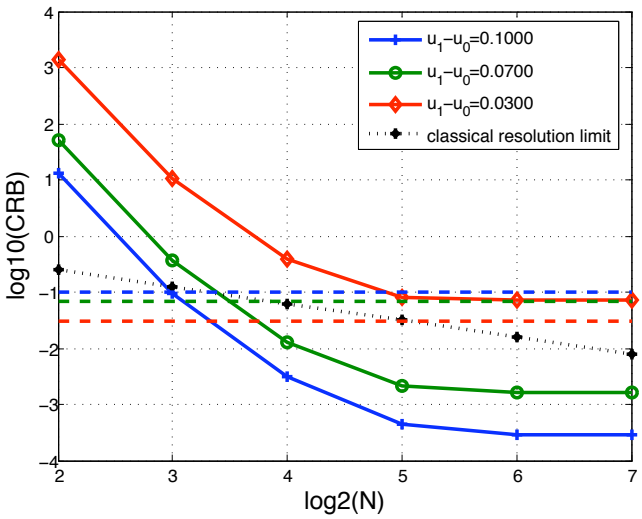


Figure 3-16. Comparison between Cramér-Rao bound of the proposed and classical systems in the presence of AWGN, in terms of number of samples. In this case the roots are near unity, and we consider $\delta u = 0.1, 0.07, 0.03$. The dashed lines show $(\delta u)^2$. The simulation is done at SNR=20 dB.

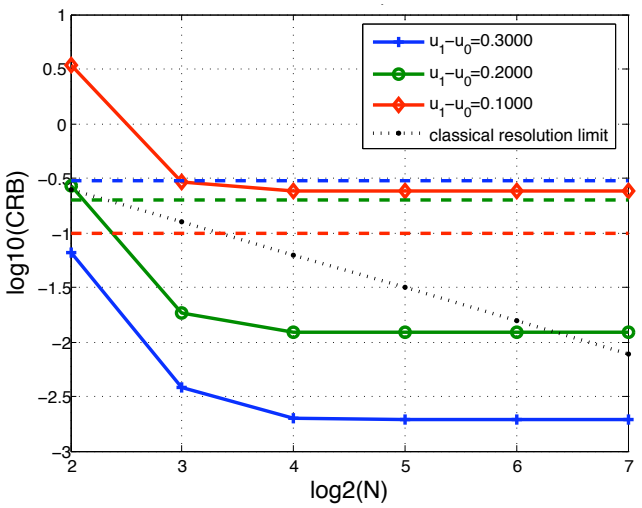


Figure 3-17. Comparison between Cramér-Rao bound of the proposed and classical systems in the presence of AWGN, in terms of number of samples. In this case the roots are near zero, $u_0 = 0.3$, and we consider $\delta u = u_1 - u_0 = 0.3, 0.2, 0.1$ unlike in the previous case. The dashed lines show $(\delta u)^2$. The bounds are evaluated at SNR=20 dB.

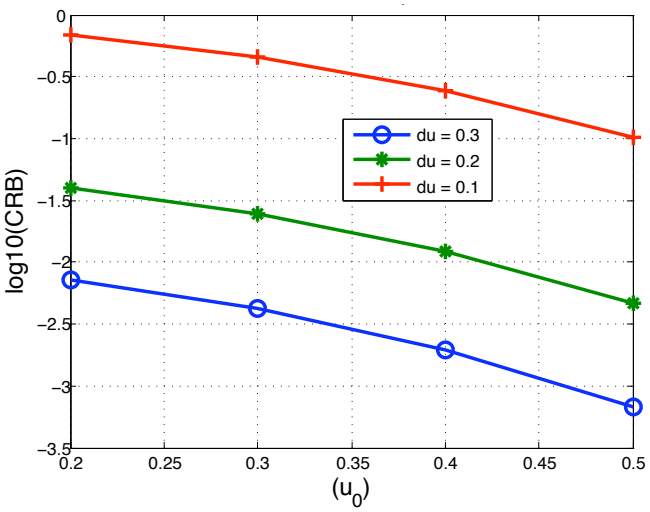


Figure 3-18. Comparison between Cramér-Rao bound of the proposed and classical systems in the presence of AWGN. The dependence of the performance on the actual locations of the roots is shown. The bounds are evaluated at SNR=20 dB.

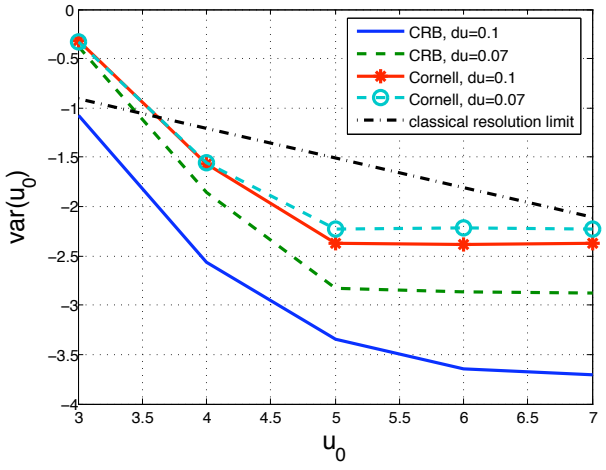


Figure 3-19. Comparison between Cramér-Rao bound of the proposed system and the Cornell algorithm in the presence of AWGN, in terms of number of samples. In this case the roots are near unity, $u_0 = 0.8$, and we consider $\delta u = u_1 - u_0 = 0.1, 0.07, 0.03$. The simulation is done at SNR=20 dB.

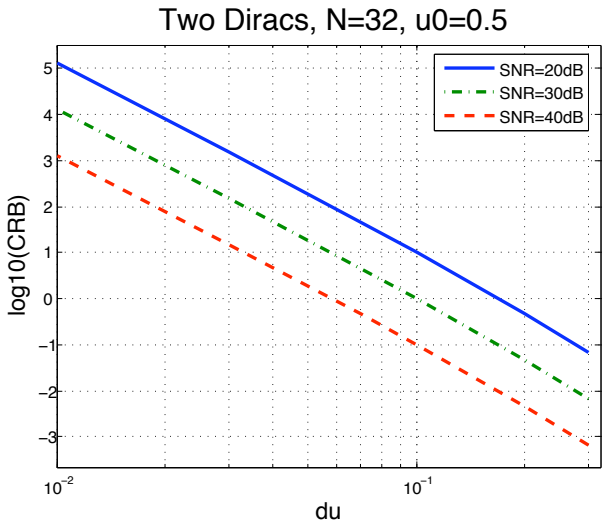


Figure 3-20. Comparison between Cramér-Rao bound of the proposed system in the presence of AWGN. We show the dependence on SNR.

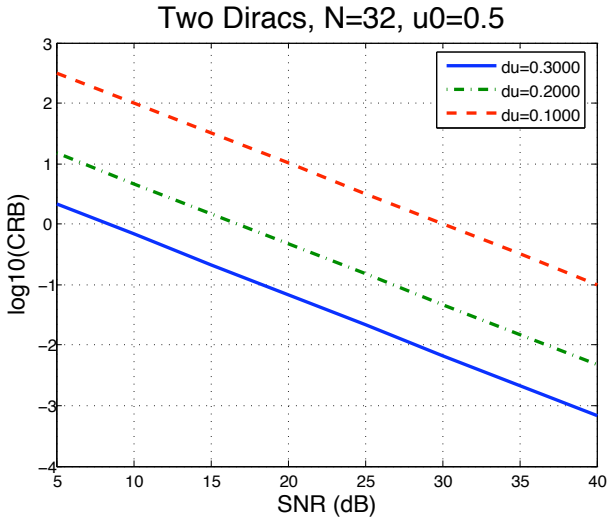


Figure 3-21. Comparison between Cramér-Rao bound of the proposed system in the presence of AWGN. We show the dependence on the performance on spacing between the roots.

different from that of Smith's examination of line spectra estimation, where he obtained $\text{var}(\omega_0) \propto \text{SNR}^{-1/4}$. Now we consider the effect of spacing on the performance in Figure 3-21. We observe that the relationship is $\text{var}(u_0) \propto (u_1 - u_0)^{-4}$. In contrast, Smith's study of line spectra estimation obtained $\text{var}(\omega_0) \propto (\omega_1 - \omega_0)^{-2}$. These observations further illustrate the difference between the line spectra estimation case and the real-valued case.

We observed that the TLS-Prony algorithm tends to give either a duplicate pair at very small spacings, or give one "good" estimate and one "poor" estimate. The performance of the "good" estimate tends to improve with N, but the performance of the "poor" estimate worsens with N. The performance of the "good" estimate tends to be better than the estimate of the Cornell algorithm, but the performance of the "poor" estimate is worse. On average, the performance is only slightly better than the Cornell algorithm.

■ 3.6 Conclusions

In this work we proposed a novel signal acquisition method suited to certain parametric signals. In the simplest case, a particularly simple reconstruction algorithm exists, suggesting a divide-and-conquer approach. We showed that it is possible to detect undermodeling, justifying the divide-and-conquer approach. We compared the performance of the proposed system to that of the conventional system based on the notion of Nyquist sampling at the Fourier bandwidth.

Much of the work relies on solving a powersum series. We considered two algorithms: the Cornell algorithm and the Total Least-Squares algorithm. Both are suited for small sample sizes, and we show that they are robust to the correlation structure in the noise. We derived a Cramér-Rao bound for the basic cases of interest. Although this problem has been studied in line specta estimation, translation of analytical results is not straightforward and can be misleading. We demonstrated that the proposed system has *super-resolution* properties, meaning that its performance in the high-SNR case exceeds that of the conventional system. Since the performance of the proposed system is dependent on the relative location of the Diracs within the integration and sampling period, the performance of the system can be improved by staggering several integration and sampling periods.

The number of integrate-and-dump circuitries required for the proposed method is on the order of the rate of the innovation of the signal, which is considerably lower than the Nyquist bandwidth of the signal.

Chapter 4

Acquisition of Parametric Signals via Uniform Sampling in Time

The previous chapter introduced a new sampling scheme with associated reconstruction algorithms and performance analysis for several cases of interest. The scheme proposed takes samples simultaneously at the output of multiple channels.

This chapter re-examines the proposed sampling schemes of Vetterli *et al.* and Dragotti *et al.* and uses the previously introduced algorithms and analysis techniques. In Chapter 2 we reviewed several of these schemes:

1. Sampling of periodic Diracs using the sinc kernel [117, Sec. 3] in Section 2.3.2.

2. Sampling of finite-length, aperiodic Diracs using the Gaussian kernel [117, Sec. 4B] in Section 2.3.3.

3. Sampling of infinite-length Diracs using finite-length kernels [32, 33] in Section 2.3.4.

These schemes all take uniform samples in time at the output of the sampling kernel, and find the signal parameters via the powersum series, given in Definition 2.1. In this chapter we focus on the first and last schemes. We will see that the first scheme requires the solution of the powersum series when the roots are complex roots of unity, and for the last scheme the roots are real-valued, similar to the previous chapter. Indeed, we will see later that these schemes are closely related. We use the analytical and algorithmic tools that we developed in the previous chapter, and prove new performance bounds for powersum series when the roots are complex roots of unity. In the more general case we give numerical evaluation of some bounds and demonstrate via Monte Carlo simulation that the sampling systems proposed in [33, 117] have *super-resolution* properties.

Further, we show that the scheme for sampling infinite-length Diracs using finite-length kernels given in [33] have a successive refinement property. Unlike the proposed scheme of the previous chapter, the successive refinement can be done *after* the samples are taken, subject to some constraints.

■ 4.1 Periodic Signals with Finite Rate of Innovation

In this section we examine the performance of the sampling scheme proposed by Vetterli *et al.* and examine its performance in the presence of additive noise. Recall that this scheme

was reviewed in Section 2.3.2. Following Definition 2.3, let the noiseless signal model be

$$x(t) = \sum_{\ell \in \mathbb{Z}} \sum_{k=0}^{K-1} c_k \delta(t - t_k - \ell T_p),$$

where $t_k \in [0, T_p)$. In the presence of additive noise, let the noisy signal model be

$$y(t) = \sum_{\ell \in \mathbb{Z}} \sum_{k=0}^{K-1} c_k \delta(t - t_k - \ell T_p) + w(t - \ell T_p). \tag{4.1}$$

Both $x(t)$ and $y(t)$ are periodic with period T_p. The rate of innovation of $x(t)$ is $2\pi/T_p$. The Fourier series of $y(t)$ is given by

$$\begin{aligned} Y[m] &= \sum_{k=0}^{K-1} c_k \exp\left(j\frac{2\pi}{T_p} t_k m\right) + W[m] && (4.2) \\ &= \sum_{k=0}^{K-1} c_k \left(\exp\left(j\frac{2\pi}{T_p} t_k\right)\right)^m + W[m]. && (4.3) \end{aligned}$$

Note that the above is a *noisy powersum series*, with poles $u_k = \exp(j\frac{2\pi}{T_p} t_k)$. We use techniques developed in [30] and [103] to derive the CRB for the *resolution* of the this sampling system. Recall Theorem 2.1. We have seen in Chapter 2.3.2 that in the noiseless case, the parameters $\{c_k, t_k\}_{k=0}^{K-1}$ can be estimated from $N \geq 2K$ observations of the sequence $Y[m]$. In turn, these observations $Y[m]$ can be obtained by taking samples of $y(t)$ the output of a lowpass filter and computing the Discrete Fourier Transform (DFT). The samples must be taken at a rate of $f_s = N/T_p$ and the bandwidth of the lowpass filter must be chosen to be $f_s/2$ so that there is no aliasing [117]. In the presence of noise, we can increase the sampling rate and the bandwidth of the lowpass filter and obtain more observations of $Y[m]$ [60, 61, 69, 71].

Vetterli, Marziliano and Blu proposed the following sampling scheme [117]:

1. Apply lowpass filter $\varphi(t)$ of bandwidth $N/2T_p$.
2. Obtain uniform samples $y[n] = \langle y(t), \varphi(t - nT) \rangle$ at sampling period $T = T_p/N$.
3. Compute N Fourier series coefficients $Y[m]$ from $y[n]$.
4. Solve for unknown parameters $\{c_k, t_k\}$ in powersum series.

We show the system in Figure 4-1 and illustrate the steps in Figure 4-2. In Figure 4-2 the original signal is given in the Fourier domain by $X[m]$ and is aperiodic in that domain. For the signal class of interest, the support of the signal $X[m]$ is infinite. The filtered and sampled signal in the Fourier domain is given by $Y[m]$, and in the noiseless case $Y[m] = X[m]$ for $m = -2, -1, 0, 1, 2$ because there is no aliasing in the Fourier domain. The sampled signal $Y[m]$ is periodic in the Fourier domain with period 5 due to the sampling. The estimation of unknown parameters in the powersum series can be done using one of many known methods. We give some of these in Appendix A.

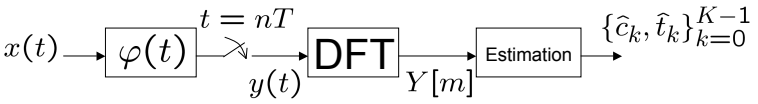


Figure 4-1. Illustration of FRI sampling system.

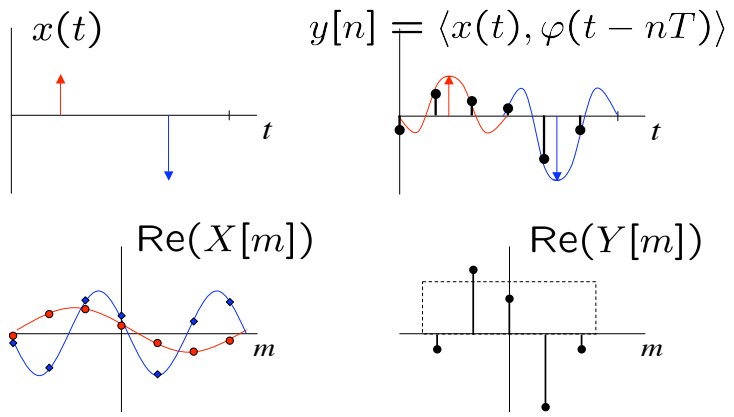


Figure 4-2. FRI sampling of a periodic signal consisting of two Diracs. The original signal $x(t)$ is filtered and sampled to obtain $y[n]$, and the corresponding Fourier series are given below the time-domain signals.

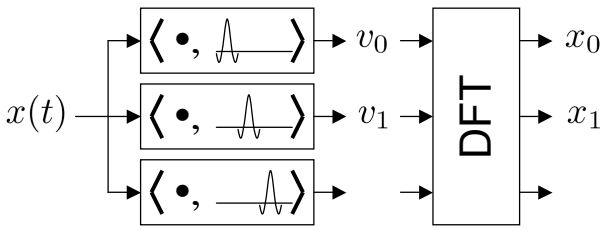


Figure 4-3. Implementation of the sampling scheme using the Discrete Fourier Transform.

Although Maravić and Vetterli [71] considered the performance of the FRI sampling scheme of [117] in the presence of AWGN, their analysis assumes that the reconstruction is done by first estimating the second-order statistics of the observation and then using the matrix pencil method proposed by Hua and Sarkar [48, 49]. While the matrix pencil method is known to perform well for large block sizes, parameter estimation via second-order statistics is known to perform poorly for small block sizes. Further, the result of Vetterli and Maravic holds only for signals which contain only one component. In this chapter we prove the performance limits and demonstrate that for simple cases of interest it is possible to obtain performance close to the limit by using the algorithms described in Appendix A. We show by numerical simulation that the proposed scheme of [117] also has a super-resolution property.

■ 4.1.1 Cramér-Rao bounds and performance evaluation

In this section we derive the Cramér-Rao bound for the timing estimation of a single pulse and the resolution bound for the discrimination of two pulses. The derivation closely follows the derivation of Dilaveroğlu [30].

Single-component estimation

First consider the case when the case when $K = 1$. The desired parameters are $\theta = (c_0, t_0)^T$. The noiseless signal is given by Let the noiseless signal be given by:

$$X[m;\theta] = c_0 \exp\left(j\frac{2\pi}{T_p}t_0 m\right), \tag{4.4}$$

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

$$\begin{aligned} y(t) &= c_0\delta(t-t_0) + w(t) \\ Y[m] &= c_0 \exp\left(j\frac{2\pi}{T_p}t_0 m\right) + W[m]. \end{aligned}$$

Suppose that we observe M samples of $Y[m]$, $m = 0, 1, \ldots, M-1$. For convenience, let $\delta t = |t_1 - t_0|$, and

$$\Gamma_r = \frac{1}{N^{r+1}}\sum_{m=0}^{N-1} m^r, \tag{4.5}$$

Further, let $\bar{s}[m] = \Re\{X[m;\theta]\}$ and $\tilde{s}[m] = \Im\{X[m;\theta]\}$. In the presence of AWGN, it is easy to see that the Fisher Information Matrix is given by

$$\mathbf{J} = \frac{2}{\sigma^2}\left[\sum_m \left(\bar{\mathbf{s}}[m]\bar{\mathbf{s}}^T[m] + \tilde{\mathbf{s}}[m]\tilde{\mathbf{s}}^T[m]\right)\right] = \frac{2}{\sigma^2}\left(\mathbf{K}\cdot\mathbf{L}\cdot\begin{bmatrix}\Gamma_0 & 0\\ 0 & \Gamma_2\end{bmatrix}\cdot\mathbf{L}\cdot\mathbf{K}\right), \tag{4.6}$$

where we have

$$\bar{\mathbf{s}}[m] = \begin{bmatrix}\cos\left(\frac{2\pi}{T_p}t_0 m\right)\\ -\alpha_0 m\sin\left(\frac{2\pi}{T_p}t_0 m\right)\end{bmatrix}, \quad \tilde{\mathbf{s}}[m] = \begin{bmatrix}\sin\left(\frac{2\pi}{T_p}t_0 m\right)\\ \alpha_0 m\cos\left(\frac{2\pi}{T_p}t_0 m\right)\end{bmatrix}, \tag{4.7}$$

and $\mathbf{K} = \text{diag}(\sqrt{N}, N\sqrt{N})$, $\mathbf{L} = \text{diag}(1, \alpha_0)$, and $\alpha_0 = c_0(2\pi/T_p)$. We obtain the following theorem, the proof of which is given in Appendix B.2.

Theorem 4.1 (Cramér-Rao bound for periodic single-Dirac estimation using sinc kernels). *Let $\alpha_0 = c_0(2\pi/T_p)$ and $SNR = \alpha_0^2/\sigma^2$. For convenience, let $\Gamma_r = \frac{1}{N^{r+1}}\sum_{m=0}^{N-1} m^r$. Suppose that we obtain N samples of the signal after filtering using an anti-aliasing filter with bandwidth π/N rad. For the case of $K = 1$, the CRB for time estimation is given by*

$$\mathbb{E}[(\hat{t}_0 - t_0)^2] \geq \frac{1}{2SNR}\frac{1}{N^3}\frac{1}{\Gamma(2)}. \tag{4.8}$$

The result obtained in the above theorem is noteworthy:

1. It has the same scaling law as what is known for single-component line spectra estimation [106].
2. Similarly to the line spectra estimation case, it is independent of the location of the pulse t_0.
3. In single-component line spectra estimation, it is known that in some regimes several algorithms achieve this lower bound, such as the TLS-Prony algorithm [105]. For small sample sizes and high SNR, we show that the Prony method and the rotational invariance algorithm perform close to the lower bound for our problem of interest.

We show the performance of the Prony method and the rotational-invariance algorithm in Figure 4-4. It can be seen that these two algorithms perform very close to the Cramér-Rao bound that we derived.

Resolution of two components

Now we examine the resolution performance of the proposed system by considering the case when $K = 2$. 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. Let $\theta = (c_0, c_1, t_0, t_1)^T$ be the vector of unknown parameters. For convenience, let the noiseless Fourier series be given by

$$X[m;\theta] = c_0 \exp(j(2\pi/T_p)t_0 m) + c_1 \exp(j(2\pi/T_p)t_1 m) \tag{4.9}$$

Further let $\bar{s}[m] = \Re\{X[m;\theta]\}$ and $\tilde{s}[m] = \Im\{X[m;\theta]\}$. Then the Fisher Information Matrix is given by

$$\mathbf{J} = \frac{2}{\sigma^2}\Big[\sum_m \left(\tilde{\mathbf{s}}[m]\bar{\mathbf{s}}^T[m] + \bar{\mathbf{s}}[m]\tilde{\mathbf{s}}^T[m]\right)\Big]. \tag{4.10}$$

For convenience, let $\alpha_0 = c_0 2\pi/T_p$ and $\alpha_1 = c_1 2\pi/T_p$. We define

$$\bar{\mathbf{s}}[m] = \begin{bmatrix} \cos\left(\frac{2\pi}{T_p}t_0 m\right) \\ \cos\left(\frac{2\pi}{T_p}t_1 m\right) \\ -\alpha_0 \cdot m \cdot \sin\left(\frac{2\pi}{T_p}t_0 m\right) \\ -\alpha_1 \cdot m \cdot \sin\left(\frac{2\pi}{T_p}t_1 m\right) \end{bmatrix}, \quad \tilde{\mathbf{s}}[m] = \begin{bmatrix} \sin\left(\frac{2\pi}{T_p}t_0 m\right) \\ \sin\left(\frac{2\pi}{T_p}t_1 m\right) \\ \alpha_0 \cdot m \cdot \cos\left(\frac{2\pi}{T_p}t_0 m\right) \\ \alpha_1 \cdot m \cdot \cos\left(\frac{2\pi}{T_p}t_1 m\right) \end{bmatrix} \tag{4.11}$$

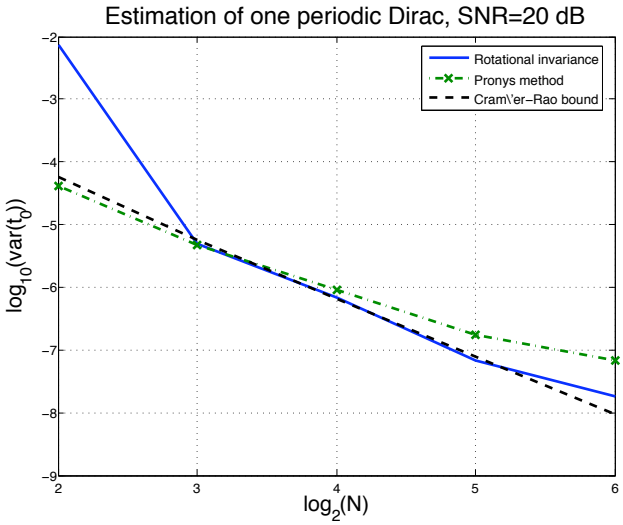


Figure 4-4. Timing estimation result for the case of one periodic Dirac using the FRI in-time sampling system. The SNR is 20 dB, and we compare the Prony method, the rotational invariance algorithm, and the Cramér-Rao Bound.

We obtain the following theorem, which is our main result on the resolution of our periodic signal signal model. The proof is given in Appendix B.2.

Theorem 4.2 (Cramér-Rao bound for periodic two-Dirac estimation using sinc kernels). *For the case of $K = 2$, let $SNR_k = \alpha_k^2/\sigma^2$. Suppose that we obtain N samples of the signal after filtering using an anti-aliasing filter with bandwidth π/N rad.For convenience, let $\delta t = |t_1 - t_0|$, and*

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

Further, let

$$\begin{aligned} \mathbf{M} &= \begin{bmatrix} \Gamma_2(\Gamma_0^2 - C_0^2) - S_1^2\Gamma_0 & C_2(\Gamma_0^2 - C_0^2) - S_1^2 C_0 \\ C_2(\Gamma_0^2 - C_0^2) - S_1^2 C_0 & \Gamma_2(\Gamma_0^2 - C_0^2) - S_1^2\Gamma_0 \end{bmatrix} \\ \mathbf{P} &= \frac{1}{\Gamma_0^2 - C_0^2}\mathbf{M}. \end{aligned}$$

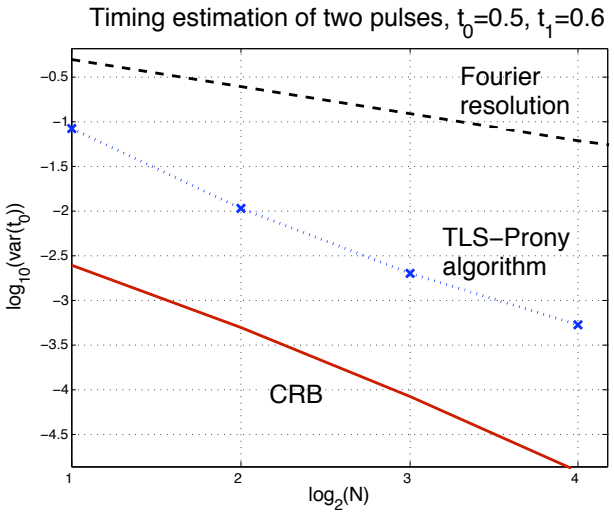


Figure 4-5. Timing estimation result for the case of two signal components using the FRI in-time sampling system. The SNR is 20 dB, and we compare the Fourier resolution, the Cramér-Rao Bound, and the performance of the TLS-Prony algorithm.

Then the CRB is given by,

$$\mathbb{E}[(\hat{t}_k - t_k)^2] \geq \mathbb{E}[(\hat{t}_k - t_k)^2] \geq \frac{1}{SNR_k}\frac{1}{N^3}[\mathbf{P}]_{k,k}. \tag{4.12}$$

We make several observations regarding Theorem 4.2:

1. The CRB depends only on the spacing $\delta t = |t_1 - t_0|$. All the terms Γ_r, S_r and C_r depend only on sample size N and spacing δt.

2. The CRB is not much affected by the spacing δt.

3. The decay rate in SNR and N is consistent with previously known results, and is consistent with Theorem 4.1.

4. The formula obtained in (4.12) is similar to that of Diraveroğlu [30, Theorem 2].

Given Theorem 4.2 we evaluate the performance of the TLS-Prony algorithm and Cornell's simple formula. In this set of simulations we set the period of the signal to be $T_p = 1$. We show the discrimination result in Figure 4-5.

■ 4.2 Infinite-length Signals with Finite Rate of Innovation

In this section we examine the sampling scheme proposed by Dragotti *et al.*, that uses kernels with compact support and is suitable for infinite-length signals with finite rate of innovation. Recall from Section 2.3.4 that this scheme allows for local reconstruction.

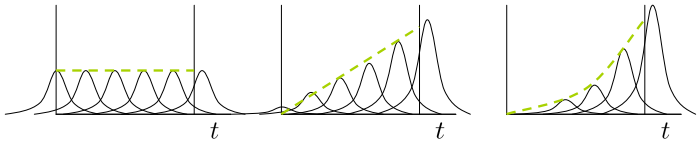


Figure 4-6. Production of t^0, t^1 and t^2 within an interval of interest, from uniformly-spaced B-splines.

Recall Definition 2.4, and let the kernel $\varphi(t)$ satisfy the Strang-Fix condition and able to reproduce polynomials up to order R. The samples are obtained: $x_n = \langle x(t), \varphi(t - nT) \rangle$. Without loss of generality, assume $T = 1$ and that $x(t)$ only has K Diracs. Let $s[m] = \sum_n c_{m,n} y_n, m = 0, 1, \ldots, N$. This is a weighted sum of observed samples, with weights $c_{m,n}$ given by the polynomial reproduction conditions above. Then,

$$\begin{aligned} s_m &= \sum_n c_{m,n} x_n = \int x(t) t^m \, dt \\ &= \sum_{k=0}^{K-1} a_k t_k^m, \quad m = 0, 1, \ldots, M-1. \end{aligned} \tag{4.13}$$

Clearly, we require that $R \geq M$. The signal model (2.1) is again a powersum series, as introduced in Definition 2.1. In the presence of additive noise, we write $y_n = x_n + w_n$. For the remainder of this section we assume that $\varphi(t)$ is a B-spline of a certain order.

■ 4.3 Successive Refinement

This scheme is closely related to the proposed system in Chapter 3. Given N samples of the output of an R-order B-spline, it is possible to obtain a powersum series of length $M = \min\{N, R\}$. Therefore it is possible to obtain perfect reconstruction of a powersum series with up to $M/2$ components.

Suppose that we know that there is a finite but unknown number K^* of Diracs within an interval $[0, T)$. We can start by taking 2 samples and reconstructing 1 Dirac, and compare the results with that obtained by taking another 2 samples and reconstructing 2 Diracs. If the results are consistent, then we stop the sampling operation. Otherwise, having reconstructed L Diracs from $2L$ samples we take another 2 samples and compare the earlier reconstruction with that for $L + 1$ Diracs from the larger set of $2L + 2$ samples. We stop when the result is consistent with that of the previous reconstruction. In the presence of noise or model mismatch, a stopping criteria other than exact reconstruction can be defined. This is possible so long as the number of Diracs K^* is less than $M/2$, half the maximal order of polynomial reconstruction possible. Unlike the proposed scheme of the previous chapter, the successive refinement can be done *after* the samples are taken, subject to some constraints.

■ 4.3.1 Single-component estimation

We focus first on a signal consisting of one Dirac, $c_0\delta(t-t_0)$, where $t_0 \in [0,T)$. Without loss of generality, let $T=1$. We examine the performance as we take N samples within this interval. Since $K=1$, we can simply use the first-order B-spline, $\beta^1(t)$. See [114] for an excellent introduction to B-splines. In the noiseless case, the samples are given by:

$$y_n = \langle x(t), \beta^1((t-n)/N)\rangle, \quad n=0,1,\ldots,N-1.$$

As the sampling rate and the number of samples is increased, we scale the sampling kernel appropriately.

Recall Section 2.3.4. To reconstruct t^0 and t^1 within the interval of interest $[0,1)$ from $\beta^1(t/N)$, the coefficients are simply given by the evaluation of the n-th order polynomial at

$$c_{m,n} = (n/N)^m$$

such that within $[0,1)$ we have

$$t^m = \sum_{n=0}^{N-1} c_{m,n}\beta^1(n/N).$$

Then we form the powersum series from the samples by

$$\begin{aligned} s_m &= \sum_n c_{m,n} y_n && (4.14)\\ &= \sum_{k=0}^{K-1} a_k (t_k)^m + v_m, \quad m=0,1,\ldots,M-1, && (4.15)\end{aligned}$$

where v_m is the noise term in the powersum series.

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 tri-diagonal form. The diagonal entries are given by:

$$\mathbb{E}[w_n w_m] = \begin{cases} \frac{N_0}{2}\frac{1}{N^2} 2\int_{t=0}^{1} t\cdot dt = \frac{N_0}{2}\frac{1}{N^2}, & m=n \\ \frac{N_0}{2}\frac{1}{N^2}\int_{t=0}^{1}(1-t)t\cdot dt = \frac{N_0}{2}\frac{1}{N^2}\frac{1}{6}, & |n-m|=1 \\ 0, & \text{else.}\end{cases}$$

The results are shown in Figure 4-7. 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. Hence regardless of the number of samples N taken in time, the powersum series that we obtain as an intermediate form has two terms: s_0 and s_1. Note that the noise is modeled as a continuous-time AWGN with infinite support in both the time and frequency domains. Recall that s_0 and s_1 are obtained from the projection of the noisy signal $y(t)$ onto kernels given by $f_0(t) = \sum_n d_{0,n}\beta^1(t-n/(N-1))$ and $f_1(t) = \sum_n d_{1,n}\beta^1(t-n/(N-1))$ respectively. By design, we have that $f_0(t) = t^0$ and

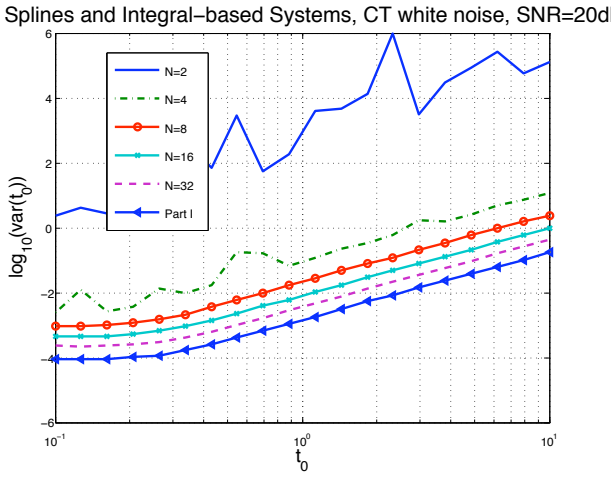


Figure 4-7. Performance results for estimation of one Dirac using a first-order B-spline. The system is implemented using the simple Cornell algorithm. The plot shows different number of samples N, but the reconstruction first forms a length-2 powersum series. The additive white Gaussian noise is added in the continuous-time domain, with spectral density $N_0 = 0.1$.

$f_1(t) = t^1$ for $t \in [0, T)$. But neither $f_0(t)$ nor $f_1(t)$ is zero outside that interval of interest. By assumption, the signal exists only in $t \in [0, T)$, and hence the kernels capture noise outside of this interval. As the number of time-domain samples N is increased, the splines used for sampling become narrower, and the footprint of $f_0(t)$ and $f_1(t)$ becomes closer to $[0, T)$. Clearly, the performance then approaches that of the multichannel sampling system given in Chapter 3. In other words, the multichannel sampling scheme given in Chapter 3 is more efficient in the acquisition of the desired signal as the system takes data only from the desired portion of the signal.

■ 4.3.2 Two-component estimation

For the discrimination of two Diracs, we show the results in Figure 4-8. 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 super-resolution property.

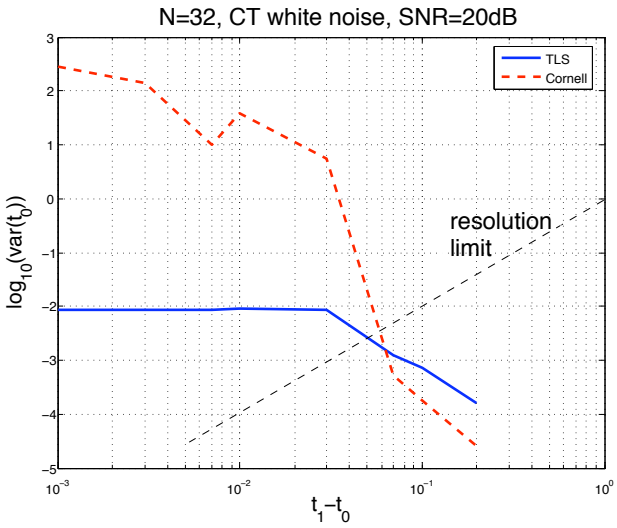


Figure 4-8. Performance results for estimation of two Diracs using a first-order B-spline. The reconstruction first forms a length-4 powersum series. The additive white Gaussian noise is added in the continuous-time domain, with spectral density $N_0 = 0.1$.

■ 4.4 Performance Comparison

It is possible to make performance comparisons between the proposed sampling schemes that we have considered so far, even if the signal models do not match exactly. We use as an example an aperiodic signal with one Dirac located within $t \in [0, T)$ subjected to continuous-time white noise of infinite support both in time and frequency domains. The multichannel sampling scheme given in Chapter 3 and the sampling scheme given in Section 2.3.4 and analyzed in Section 4.2 are suitable for this signal model.

The sampling scheme given in Section 2.3.2 and analyzed in Section 4.1 is suitable for periodic sums of Diracs, but we can use a periodic approximation from a finite number of samples and compute the Discrete Fourier Transform of these samples.

We show the comparison in Figure 4-9. The conventional method is implemented using a full gradient search of the peak of a correlator. However, suppose now that $K > 1$. In this general case, the conventional method requires a simultaneous K-dimensional search for peaks at the output of a correlator. It is known that this method requires very accurate initial conditions, and that the objective function contains many local maxima. Regardless, we have demonstrated that the proposed systems have super-resolution properties and in many regimes outperform the conventional method.

■ 4.5 Conclusions

We examined the proposed method of Vetterli *et al.* for reconstructing signals of finite rate of innovation by filtering and uniformly sampling the signal in time. We proved lower bounds on the timing estimation performance of the proposed system for the case when

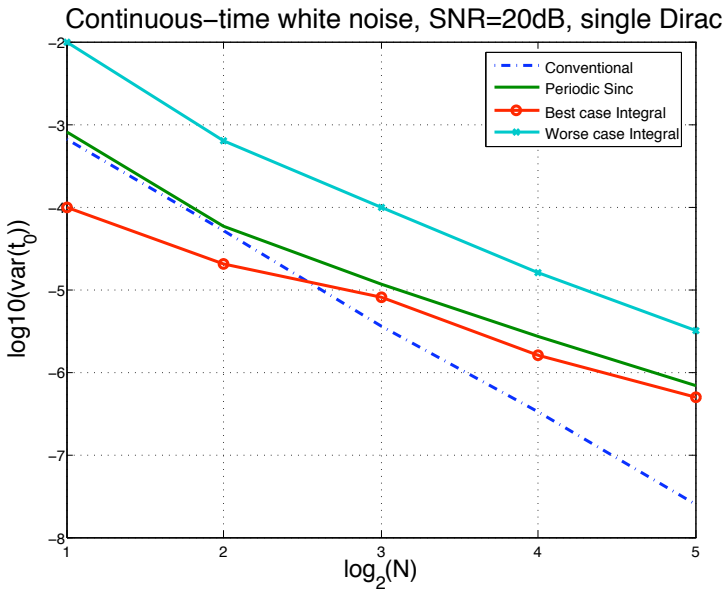


Figure 4-9. Performance comparison for single-Dirac estimation, located within $[0, T)$, for the same numbers of samples.

there is only one component in the signal, and the resolution of the proposed system when there are two components in the signal.

In the both cases we obtained a bound that is independent of the parameter of interest and derive significant simplifications that give strong insight into the problem of interest. Further, we showed that well-known algorithms that are very simple to implement give performance close to the obtained bounds. We showed that the proposed system has a *super-resolution* property, that is its resolution exceeds that of the classical system.

We then applied the analytical and algorithmic tools of the previous chapter to the method of Dragotti *et al.* for reconstructing infinite-length signals with finite rate of innovation, using finite kernels such that local reconstruction is possible. We used the same algorithms from the previous chapter to show that in many regimes this system also has superior performance and resolution when compared to the conventional system.

Finally, we showed that the scheme of Dragotti *et al.* for sampling infinite-length Diracs using finite-length kernels given in [33] have a successive refinement property similar to that of the multichannel scheme of Chapter 3. However, unlike the proposed scheme of the previous chapter, the successive refinement can be done *after* the samples are taken, subject to some constraints.

Chapter 5

Signal Parameter Estimation in the Presence of Timing Noise

In this chapter we consider the problem of estimating parameters of a signal when the sampling instances are perturbed by signal-independent timing noise. The classical techniques consider timing noise to induce a signal-independent additive noise term on the sample values, and in practice this additive noise is modeled as white Gaussian if considered at all. We present tractable alternative methodologies and simulation results that indicate significant potential for improvement over linear processing. A specific problem studied in depth is delay estimation when the pulse shape and amplitude of the signal are known. We give an iterative algorithm that shows superior performance compared to the traditional method which relies on maximizing the cross-correlation.[1]

■ 5.1 Introduction

The standard abstraction for analog-to-digital conversion (ADC) starts with the sampling of a continuous-time signal at precisely known, even-spaced times [115]. In any real application, the actual sampling instances are not the same as the ideal, desired sampling instances. We call the difference *timing noise*. Making timing noise small in practice comes at a large cost in terms of power, manufacturing cost (for tight tolerances), and device size. Thus algorithmic mitigation of timing noise can significantly improve system design.

In many current and emerging applications, timing noise has become a significant impairment. It is a dominant source of noise in wideband ADC, decreasing resolution by about 1 bit for every doubling of the sampling rate in the Mega- and Giga-sample per second range [3, 67, 118]. In many communication applications this timing noise is even more pronounced as the receiver circuitry has to synchronize to the transmitter circuitry, and both have their own timing noise. Applications that rely on accurate timing, such as the Global Positioning System (GPS) have strict timing noise specifications [27]. A more abstract (and unwieldy) replacement for "timing noise" could be "function domain noise." This would emphasize that similar problems arise when acquired signals are not time series. For example, a crawler that traverses the length of an oil pipe looking for signs of wear and tear is propelled by wheels, which may slip or may not have constant velocity. Air- or water-borne observation devices may have sensing elements in which inter-sensor distance

[1] Part of this work was presented at the 2006 Conf. on Information Sciences and Systems, Princeton, NJ [59].

and inter-sample timing is not constant.

ADCs are triggered by clock signals which in turn are often derived from lower-frequency clock signals. In most digital integrated systems a single clock generator is used to trigger many different components within the same system. The problem of *clock distribution* has been widely studied, in which propagation error resulting from additive noise, capacitance, and delay is considered and routing methodologies are optimized to minimize the effect of clock propagation across the integrated circuit [29,37,46]. Clock signals are not perfect, and characterization of their noise has been studied extensively [42,62]. In electronic system design, timing noise is often modeled as additional input-independent additive white Gaussian noise (AWGN) [1,14,15,34,118]. The effect of timing noise is specified in terms of increased noise floor in the system, given in terms of Effective Number of Bits (ENOB). However, when viewed as additive, the effect of timing noise in not signal independent and furthermore neither white nor Gaussian. This is the key insight of this work. We consider signals with a known parametric model, but this insight can also be applied in non-parametric acquisition.

We do not consider quantization here, though the observation-domain noise could be redefined to include the effects of quantization, if quantization noise was modeled as signal independent. In some ADC architectures, timing noise would also affect value accuracy; see [47] for an overview of the various implementations. We focus our attention on the unquantized case because the three fundamental limitations in ADC are thermal noise, timing noise, and comparator ambiguity, with the last of these prohibitive only at extremely high sampling rates [118].

In this work we focus on the classical problem of signal delay estimation, which is severely affected by timing noise. It is known that in the presence of only AWGN, the optimal delay estimator is the one that maximizes the output of the cross-correlator. Given the standard simplification of the effect of timing noise given above, this same algorithm is used even when it is known that the ADC is subjected to timing noise. We refer to this method as the "standard" method. In contrast, we propose new algorithms that take into account timing noise and its effect on the observations. Throughout this chapter we use the term Additive White Gaussian Noise (AWGN) to refer to *observation-domain* AWGN.

The main contributions of this work are threefold:

1. We reject the simplification that the effect of timing noise on the observed samples is well-modeled by signal-independent AWGN.

2. Instead of directly computing the maximum-likelihood (ML) estimate of the desired signal parameter, we use the dependence between the desired signal and the effect of timing noise to estimate the timing noise explicitly. This simplifies the computation.

3. We propose algorithms that are simple to implement, and demonstrate the efficacy of our approach.

■ 5.2 Previous Work

The problem of sampling in the presence of timing noise is not new. In this section we review previous work divided into three parts: signal reconstruction, spectral estimation,

and characterization of timing noise. The term "jitter" is often used to describe timing noise. More broadly speaking, jitter is also used to denote the deviation of a signal's phase due to noise or other impairments.

■ 5.2.1 Signal reconstruction in the presence of timing noise

The case of bandlimited signals, including deterministic line spectral terms, was considered by Balakrishnan [6] and Brown [18]. These works focused on computing spectra, the optimal LTI interpolation filters, and the resulting expected $\mathcal{L}_2$ reconstruction errors.

Butzer [19] also studied the effect of timing noise on bandlimited signals; see the newer book of Higgins for a summary of the work [45]. Butzer considered the $\mathcal{L}_2$ error between the original signal, say $f(t)$, and its sinc interpolation after sampling with timing noise $\hat{f}(t) = f(nT + z_n) * \mathrm{sinc}(t/T)$. He further assumed that the signal decays sufficiently quickly in time, and that the timing noise term is bounded. The square reconstruction error can then be bounded by a constant that is proportional to the signal decay rate and the supremum of its first derivative. This result is intuitive: the steeper the slopes of the signal, the larger the projection of timing noise from the timing domain to the amplitude domain.

■ 5.2.2 Randomized sampling and spectral estimation

The effect of timing noise on the spectra of the signal has been studied. See [16, 17, 91] and the references therein. Roughly speaking, the spectrum of the input signal is filtered by the spectrum of the timing noise sequence in the Fourier domain. This is the motivation for randomized sampling, which was proposed as a method for alias-free spectral estimation. See [10] for a summary of the results. Instead of lowpass filtering at the input of the ADC, they proposed randomizing the sampling instances instead. Much work has been done in this area, starting with the spectral characterization of random processes sampled by stationary point processes and impulse processes [8, 9, 65]. In this body of work the interest is to find a consistent estimator of the input signal spectrum. The same results has also been used to study the spectra of UWB signals with random coding in the timing domain [92].

■ 5.2.3 Characterizations of timing noise

Noise in an ADC comes from various sources, such as from harmonic distortion, nonlinearities, additive and timing noise, quantization error, and filtering imperfections. There are many architectures that implement an ADC, and they are usually characterized by their stated resolution (in terms of Effective Number of Bits – ENOB), signal-to-noise ratio (SNR), dynamic range, and power dissipation [118]. Often the effect of timing noise on ENOB is approximated only by considering the signal bandwidth and timing noise RMS value [1, 14, 15]. Clearly, there is a tradeoff between speed, power, and accuracy in ADC design. For a CMOS system it is well-known that this tradeoff is given by [116]:

$$\frac{(\text{Speed [Hz]}) \times (\text{Accuracy [RMS]})^2}{(\text{Power [Watts]})} \approx \text{Constant}. \tag{5.1}$$

Sampling in an ADC is triggered by detecting zero-crossings of a clock signal from an

oscillator. In the past, sampling ADCs were built up from a separate Sample-and-Hold Aperture (SHA) and ADC, with *aperture jitter* a separately specified parameter. It is generally assumed that the effect of oscillator phase noise and aperture uncertainty are independent of each other, and that the overall timing noise variance is simply the sum of the variances of the two [5]. In contrast, today's ADCs contain an integral SHA circuitry and often only the overall timing noise RMS is specified.

Circuit designers are aware that the effect of timing noise is signal-dependent. In many papers [1,14,15], it is suggested that the effect of jitter on a sinusoidal input signal amplitude A at frequency f is approximated by its "slew", or derivative, at the zero-crossing of the signal. For a sinusoidal signal with frequency f, this gives the expression

$$\text{SNR} = 20 \log_{10} \left(\frac{1}{2\pi f \cdot \sigma_z} \right),$$

where σ_z is the standard deviation of the jitter. In order to generalize this approximation to bandlimited input signals it is often assumed that the input signal has a flat power spectra within the bandwidth of interest, and the overall error is averaged over this bandwidth [102]. Da Dalt *et al.* considered a more general signal model, that is a wide-sense stationary input signal with a given autocorrelation r_x [23, 24]. They showed that by linearizing the effect of jitter around the ideal sampling instances, they obtain

$$\text{SNR} = 20 \log_{10} \left(\frac{r_x(0)}{-r_x''(0) \cdot \sigma_z} \right).$$

In their work, the error that they are interested in is still the square error of the non-parametric reconstruction of the signal.

Much work has been done in characterizing *phase noise* of an oscillator, starting with the most well-cited pioneering work of Leeson [64]. Lee and Hajimiri improved upon Leeson's model by considering linear time-invariant and linear time-varying models [42, 62, 63]. These models can also be extended to general pulse shapes other than a sinusoid. One common model for timing noise is stationary, zero-mean, filtered white Gaussian noise or a Wiener process [23, 24, 41, 64]. This filter is usually modeled as a lowpass filter whose cut-off frequency is called the *loop bandwidth* [2, 41]. Many other models have been proposed, for example based on phase trajectories of the circuit elements, in an attempt to simplify the characterization of phase noise based on only a few parameters [43]. It is possible to design clock circuitry with very narrow loop bandwidth even for wideband applications, but at the cost of potentially significant complexity [73, 90]. A feedforward system has also been proposed which can perform better within the loop bandwidth than the typical PLL because of the absence of division in the feedforward system, but this system has only been demonstrated to work for low oscillator frequencies and at the cost of significant hardware [112, 113].

One notable previous work is that of Narasimhamurti and Awad [77]. The authors considered the problem of estimating the phase of a sinusoid sampled using a noisy clock with accumulated jitter when the frequency and amplitude are known. They proposed a state-space approach via writing the relationship between successive sample phase using trigonometric identities. The problem of measuring phase noise itself is well-known and has

been studied in the context of oscilloscope measurement [2]. There are several methods to separate the different sources of jitter in a circuit [101].

■ 5.3 Problem Statement

Consider the problem of signal delay estimation when the pulse shape is perfectly known to the receiver. That is, the input signal is

$$x(t) = g(t - \tau), \tag{5.2}$$

where τ is the unknown delay and $g(t)$ is the known pulse shape. Our model for acquisition with sampling period T, timing noise sequence z_n, and additive noise sequence w_n is

$$y_n = x(nT + z_n) + w_n.$$

We also use the notation

$$s_n = nT + z_n \tag{5.3}$$

for the noisy sampling instances and refer to z_n as in the *timing domain* and w_n as in the *observation domain*. The problem is to estimate τ from $\{y_n\}$.

We are concerned with z_n and w_n that are independent of the signal $x(t)$. However, clearly the effect of z_n on the observed samples y_n depends on the signal of interest $x(t)$. We are especially interested in situations where z_n is a prominent source of impairment.

■ 5.4 Example Toy Problems and Main Idea

To gain some insight on the problem, first consider the case with no additive noise in the observation domain. The observed samples are given by

$$y_n = x(s_n), \tag{5.4}$$

and an example of the effect of timing noise is shown in Figure 5-1. The original signal $x(t)$ has only one degree of freedom, namely its delay. However, each new sample y_n will be subjected to timing noise and contain one additional unknown parameter z_n. As a consequence, taking N samples means having N constraints on $N + 1$ unknown parameters. In fact, one can write explicitly that if $(\hat{\tau}, \hat{z}_0, \hat{z}_1, \ldots, \hat{z}_{N-1})$ is consistent with the observations, then so is $(\hat{\tau} + \alpha, \hat{z}_0 - \alpha, \hat{z}_1 - \alpha, \ldots, \hat{z}_{N-1} - \alpha)$ for any $\alpha \in \mathbb{R}$. In one can find this line and has a probabilistic model for $\{z_n\}$, then a reasonable estimate can be chosen.

Ambiguity on a single line is the best case scenario, as $g(\cdot)$ will not generally be a one-to-one function. However, it is worth noting that as compared to the nonparametric case, (5.2) has given us a relatively easy problem. For example, if (5.2) is replaced by

$$x(t) = \sum_n x_n \operatorname{sinc}(t - nT),$$

then there are roughly two degrees of freedom per observation. In this work, we are exploiting both having a parametric signal model and a willingness to depart from linear, time-invariant (LTI) processing of samples. The reader is referred to [117] for formal results

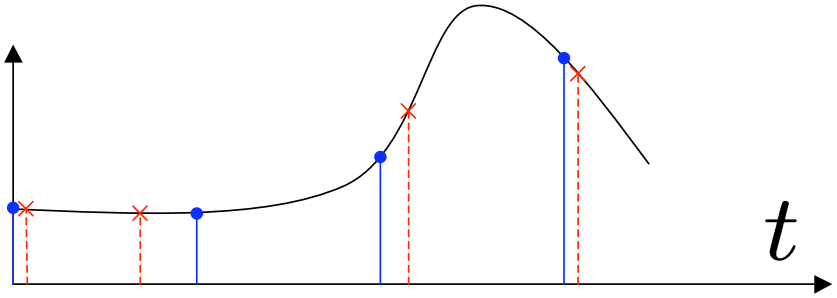


Figure 5-1. An illustration of the effect of timing noise on an input signal. The ideal sampling instances are solid stems marked by 'O' and ones subjected to timing noise are dashed stems marked by 'X'.

on sampling parametric signals.

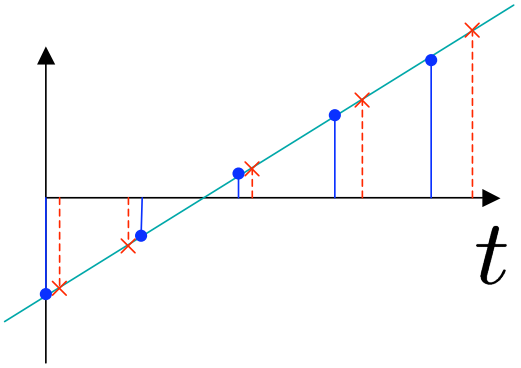
We are now ready to graduate to a more interesting problem. Let $x(t) = t - \tau$, where τ is an unknown delay term. We observe a set of samples $y_n = x(nT + z_n) = nT + z_n - \tau$. Let the timing noise z_n be an *i.i.d.* white Gaussian sequence. Given a set of samples $\{y_0, \ldots, y_{N-1}\}$ we obtain an estimate for $z_n = y_n - nT + \tau$. Since z_n appears as an additive, zero-mean, Gaussian noise term, the Maximum-Likelihood (ML) estimate for τ is simply given by:

$$\hat{\tau}_{ML} = \frac{1}{N} \sum_{n=0}^{N-1} (y_n - nT).$$

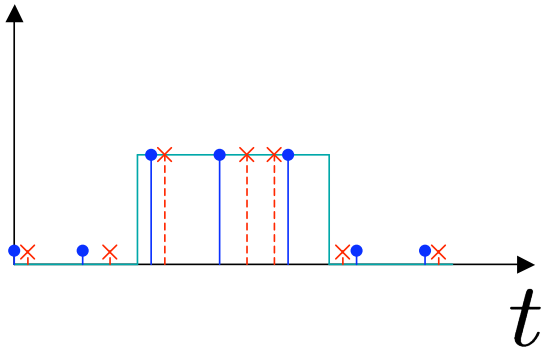
Once the estimate $\hat{\tau}$ has been obtained, we can then find the estimate of the timing noise sequence $\hat{z}_n = y_n - nT + \hat{\tau}_{ML}$. In Figure 5-2 we show two examples. The first example is equivalent to the problem that we just considered. In the first example, the signal is such that knowledge of its parameter can be used to determine the timing noise sequence. In the second example, knowledge of its parameter is not very useful for determining the timing noise sequence.

In general the pulse shape of interest is not an invertible map. Let us first assume that timing noise z_n is small compared to the sampling period T, and that the signal is sufficiently smooth. Then given τ we can compute the sampling instances s_n, and the timing noise z_n, from the obtained samples y_n. Given a model for the timing noise, we can then estimate it. Conversely, we will show that if we have an estimate of the timing noise, we can use it explicitly in our delay estimation routine.

First we consider the case when the only source of noise in the system is timing noise. We will then extend to the case when there is also AWGN.



(a) An example in which the location is uniquely determined by the sampled value.



(b) An example in which the location is not uniquely determined by the sampled value.

Figure 5-2. Toy examples of how knowledge of the signal can be used to estimate the timing noise. The ideal sampling instances are solid stems marked by 'O' and the ones subjected to timing noise are dashed stems marked by 'X'.

■ 5.5 Delay Estimation in the Presence of Timing Noise Only

We are now ready to tackle the first problem which is when the only noise in the system is white timing noise [59]. The observed samples are

$$y_n = g(nT + z_n - \tau), \quad n = 0, 1, \ldots, N-1, \tag{5.5}$$

and we wish to estimate τ. An estimate of delay and timing noise of the signal $\{\tau; z_0, \ldots, z_{N-1}\}$ is consistent with observations $\{y_n\}$ for a given pulse shape $g(t)$ when $y_n = g(nT + z_n - \tau)$ for $n = 0, 1, \ldots, N-1$. We denote the set of consistent values of $\{y_n, z_n; \tau\}$ by $\mathcal{C}$.

We propose to find the estimate $\hat{\tau}$ that minimizes the Euclidean norm of the timing noise while giving a *consistent* solution. When g is invertible, let $a_n = g^{-1}(y_n)$. In this case the term $a_n = nT + z_n - \tau$ is completely determined by y_n. Given any estimate of $\mathbf{a}$, the optimal choice of τ is the mean of $\{a_n - nT\}$. Therefore we need to only optimize over τ, or over $\mathbf{z}$. From here it is straightforward to obtain $\hat{\tau}$ that minimizes $\|\mathbf{z}\|^2$:

$$\hat{\tau}_{ML} = \frac{1}{N} \sum_n (a_n - nT). \tag{5.6}$$

In general the pulse shapes of interest are not invertible maps. For example, the Gaussian pulse is not an invertible map. In this case, for any given estimate of the delay τ there may be more than one timing noise value at each sample point that produces a consistent reconstruction, hence direct evaluation of the ML estimate is difficult. We represent all the consistent sequences via the *ambiguity set*.

Definition 5.1. *Consider the problem of estimating the delay of a signal consisting of a known pulse shape $g(t)$. Given a sequence of samples y_n we define the ambiguity set for each n as follows:*

$$\mathcal{A}_n^{(g)} = \{t \mid g(t) = y_n\}. \tag{5.7}$$

The use of this set is to allow treatment of pulse shapes $g(\cdot)$ which are not invertible maps. When the reconstruction algorithm makes a mistake in choosing the correct member of the ambiguity set, we refer to this error as an ambiguity error*.*

If we restrict ourselves only to values of $\mathbf{z}$ and τ which are consistent, then the values of $\mathbf{z}$ are completely determined by τ assuming small timing noise. Hence given the ML estimate of one we can derive the ML estimate of the other term. This observation means that the solution to the above can also be obtained by optimizing over $\mathbf{z}$. We optimize the "complete" log-likelihood:

$$\hat{\tau} = \arg\max_{\tau, \mathbf{z}} \ln Pr(\mathbf{y} \mid \mathbf{z}; \tau) Pr(\mathbf{z}; \tau) \tag{5.8}$$

Since the jitter z_n is assumed to be Gaussian, we wish to find the consistent estimate that minimizes $\|\mathbf{z}\|^2$:

$$\min_{\tau, \mathbf{z}} \|\mathbf{z}\|^2 \quad \text{s.}t. \quad y_n = g(nT + z_n - \tau), \, \forall n. \tag{5.9}$$

The above optimization can be computed in the timing domain, but considering all ambigu-

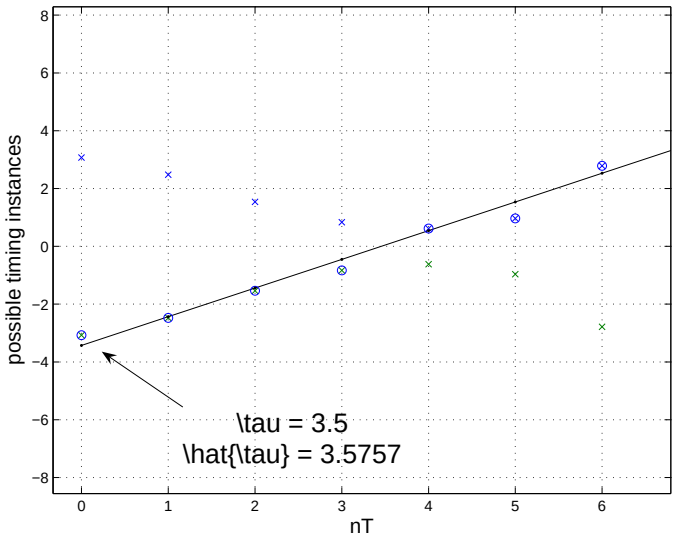


Figure 5-3. Sampling at noisy instances, with only white timing noise. In this set, $N = 16$, Gaussian pulse with width 2, $T = 2$, $\tau = 3.4$. The X's are the ambiguity set members, and the O's are those that are estimated to be the true sampling points. The line is the regression, and τ is given by the intersection of the line with the y-axis.

ous sets of a_n for each n which yield the same y_n. This corresponds to finding the minimization of $\|\mathbf{z}\|^2$ over all possible ambiguous sets given by the cardinal product $\mathbf{a} \in \vec{\mathcal{A}} = \bigotimes_n \mathcal{A}_n^{(g)}$.

In summary, a brute-force method for solving the optimization problem of (5.9) is given by:

1. For each n, compute $\mathcal{A}_n^{(g)}$ from y_n, and obtain $\vec{\mathcal{A}} = \bigotimes_n \mathcal{A}_n^{(g)}$.

2. For each $\mathbf{a} \in \vec{\mathcal{A}}$, compute

$$\hat{\tau}(\mathbf{a}) = \frac{1}{N}\sum_n (a_n - nT), \quad z_n(\mathbf{a}) = a_n - (nT - \hat{\tau}).$$

3. Choose $\hat{\tau}(\mathbf{a})$ which minimizes $\|\mathbf{z}\|_2^2$.

In most cases, the ambiguity sets $\mathcal{A}_n$ will consist only of discrete points, and a simplified solution is possible. Recall that any admissible a_n is of form $a_n = nT + z_n - \tau$. When z_n is small, this is simply a line of slope T in the variable n, and its intercept is the delay term $-\tau$. A simple heuristic algorithm for resolving the ambiguity is to iterate between *line regression*—fixing the slope to always be nT—and *nearest-neighbor association*. The standard line regression is optimal in the mean-square sense, and therefore optimal for white Gaussian timing noise z_n.

The proposed system is compared to the pulse fitting procedure which seeks to minimize the $\mathcal{L}_2$ norm by maximizing the cross-correlation between the estimate and the observation.

This latter system is a gradient-search algorithm, and we allow multiple initial conditions and pick the best result. For concreteness, we emulate the ultra-wideband (UWB) system of the MIT Microsystems Technology Laboratories (MTL) [11, 13, 119], in which a 2 ns Gaussian pulse located near the center of a 100 ns frame is subjected to a Gaussian jitter of standard deviation of 75 ps or 200 ps. There are 200 samples taken within each frame. The results are given in the table below.

	$\sigma_z = 0.2$-ns no AWGN	$\sigma_z = 0.075$-ns no AWGN	$\sigma_z = 0.075$-ns 10-dB AWGN
var($\hat{\tau}$) Proposed	0.0109	0.0031	0.9407
var($\hat{\tau}$) Standard	0.0625	0.0557	0.1064

The above algorithm is in fact an iterative algorithm that alternates between finding the ML estimate of τ and z_n. It can be further improved by considering all possible z_n given a previous estimate of τ. From (5.8) note that we can write:

$$\hat{\tau}^{(k+1)} = \arg\max_{\tau,\mathbf{z}} \ln Pr(\mathbf{y} \mid \mathbf{z}; \hat{\tau}^{(k)}) Pr(\mathbf{z}; \tau^{(k)}),$$
$$\ln Pr(\mathbf{z}; \tau^{(k)}) = \ln Pr(\mathbf{a} - \mathbf{n}T + \hat{\tau}^{(k)}).$$

This is the basis of the Expectation-Maximization (EM) algorithm [28], where instead of finding the ML estimate of the hidden variable, the expected likelihood is computed instead. We found that given the signal parameter used in the previous part, this modification gives a modest gain in performance.

While the new algorithm consistently outperforms the conventional method, it is not robust to observation-domain noise. A small amount of observation-domain noise may have an adverse effect on the performance depending on the pulse shape. We will address this in the next section.

Now we consider the delay estimation problem in which the observations are given by

$$y_n = x(nT + \Delta n) = g(nT + \Delta n - \tau).$$

In this case, the timing noise consists of a pure drift term determined by Δ. Drift occurs when the clock frequencies between two references are not exactly the same. Suppose that the drift term Δ is relatively small; in most electronic applications the drift can be guaranteed to be below 100 ppm.

From the set of measurements $\{y_n\}$, we can again find the ambiguity sets $\mathcal{A}_n$ as defined above. The problem now becomes a clustering and line regression problem. We wish to find a set of points from the ambiguity set such that a line regression of these points will yield a line with slope close to T, because we know that the drift term is small relative to T. We implement a simple iterative routine which repeats the following two steps, except that the slope is no longer T but is instead $(1 + \Delta)T$:

1. From each ambiguity set indexed in n, find the point closest to the current estimate of the line slope and offset.

2. From the set of nearest points, update the estimate of the slope and offset.

In the presence of drift and white Gaussian jitter, the above algorithm can be used with no

modification. After all, line regression is a least-squares fit, which is optimal for additive Gaussian noise. We obtain results similar to that shown in Figure 5-3.

■ 5.6 Delay Estimation in the Presence of Timing Noise and AWGN

In the previous part we have seen that one problem with the consistency-based algorithm in the presence of AWGN is that a small amount of additive noise may lead to a very different answer to the ambiguity set $\mathcal{A}_n$, and hence give a very poor estimate $\hat{\tau}$.

■ 5.6.1 Extending consistency to include AWGN

We can extend the model to include the additive noise term w_n in our definition of consistency.

Definition 5.2 (Consistency with timing noise and AWGN). *Let parameter vector θ parameterize the signal $x_\theta(t)$, and let $\{y_n\}_{n=0}^{N-1}$ be samples taken at times $\{nT + z_n\}_{n=0}^{N-1}$. Estimates of the unknown parameter vector θ, timing noise vector $\{z_n\}_{n=0}^{N-1}$ and AWGN $\{w_n\}_{n=0}^{N-1}$ are said to be* consistent with the observations *when $y_n = x_\theta(nT + z_n) + w_n$ for $n = 0, 1, \ldots, N-1$. Let $\mathcal{C}$ denote the set of consistent estimates.*

Directly expressing $Pr(\mathbf{y}; \theta)$ is difficult in general. We again follow the Expectation-Maximization algorithm (EM) approach by working with the complete data set:

$$\arg\max_{\theta,\mathbf{z}} \ln Pr(\mathbf{y}, \mathbf{z}; \theta) = \arg\max_{\theta,\mathbf{z}} \ln Pr(\mathbf{w}|\mathbf{z}; \theta) Pr(\mathbf{z}; \theta). \tag{5.10}$$

The computation of (5.10) can be done either in a nested manner or iteratively between the two variables τ and $\mathbf{z}$. For the problem of delay estimation, we are concerned only with the delay τ as the unknown parameter. The model is then given by

$$y_n = g(nT + z_n - \tau) + w_n, \quad n = 0, 1, \ldots, N-1. \tag{5.11}$$

Similarly to the case where only timing noise is present, we can obtain an ML estimate of τ from $\{z_n\}$ and vice-versa. An estimate of delay, timing noise and additive noise of the signal $\{\tau; z_0, \ldots, z_{N-1}; w_0, \ldots, w_{N-1}\}$ is consistent with observations $\{y_n\}$ for a given pulse shape $g(t)$ if and only if $y_n = g(nT + z_n - \tau) + w_n$ for $n = 0, 1, \ldots, N-1$.

Conditioned on τ, because of the whiteness, the estimates of $\{w_n, z_n\}$ can be computed separately for each n. Moreover, for each n a consistent w_n is completely determined by fixing τ and z_n. Indeed, using the model we can set:

$$w_n = y_n - g(nT + z_n - \tau), \quad n = 0, 1, \ldots, N-1. \tag{5.12}$$

Now consider computing the estimates of the hidden variables $\{w_n, z_n\}$ and the unknown parameter τ iteratively. The two estimation problems are given by:

- **Estimating hidden variables:**

$$(\hat{\mathbf{w}}_{k+1}, \hat{\mathbf{z}}_{k+1}) = \arg\max_{\mathbf{w},\mathbf{z}} Pr(\mathbf{w}, \mathbf{z} \mid \mathbf{y}, \hat{\tau}_k).$$

- **Estimating unknown parameter:**

$$\hat{\tau}_{k+1} = \arg\max_{\tau} Pr(\mathbf{y} \mid \hat{\mathbf{w}}_k, \hat{\mathbf{z}}_k; \tau).$$

In the estimation of the hidden variables we can use the insights of the previous part, namely that the estimation can be distributed in n, and that w_n is uniquely determined by (τ, z_n) following (5.12). We obtain

$$\begin{aligned} &\arg\max_{(w_n,z_n)} Pr(w_n, z_n|y_n, \hat{\tau}_k) \\ &= \arg\max_{(w_n,z_n)} \ln Pr(z_n|y_n;\tau) Pr(w_n|z_n, y_n;\tau) \\ &= \arg\min_{z_n} \frac{\|z_n\|^2}{\sigma_z^2} + \frac{\|y_n - g(nT + z_n - \tau)\|^2}{\sigma_w^2}. \end{aligned} \tag{5.13}$$

Estimation of the unknown delay is simpler: given $\mathbf{z}$ the only uncertainty in $\mathbf{y}$ is the AWGN term. Hence,

$$\arg\max_{\tau} Pr(\mathbf{y} \mid \hat{\mathbf{w}}_k, \hat{\mathbf{z}}_k; \tau) = \arg\min_{\tau} \|\mathbf{y} - g(\mathbf{n}T + \mathbf{z} - \tau)\|^2, \tag{5.14}$$

where $\mathbf{n}$ denotes the vector $[0, 1, \ldots, N-1]^T$ and g denotes the appropriate N-fold product of the pulse shape. The optimization in (5.14) has an intuitive interpretation: given our knowledge of the timing noise we can construct a template pulse shape which we can then use for maximizing the cross-correlation with the sampled signal.

We apply the optimization procedures of (5.13) and (5.14) iteratively. For both steps we use a steepest-descent approach. The performance is strictly better than that of using only correlation, because it also allows for compensation in the timing domain. The performance evaluation of this method is promising: in all cases, the performance of the proposed technique is superior to that of the standard, correlation-based method.

■ 5.6.2 Algorithm based on linearization of effect of timing noise

To be able to incorporate correlation properties of the noise sequences $\{w_n\}$ and $\{z_n\}$, we now develop an algorithm based on linearization of $x(\cdot)$ around the nominal sampling instances. We lineare the effect of z_n around nT. Following our previous notation, let $s_n = nT + z_n$ denote the sampling instances. We are interested in the pulse delay τ, which is contained in the argument of the pulse shape $g(\cdot)$. For convenience, let

$$a_n = s_n - \tau = nT + z_n - \tau. \tag{5.15}$$

The noisy samples we obtain are given by $y_n = g(a_n) + w_n$. In the amplitude domain the noise term w_n is additive and is independent of the signal. However, the effect of z_n is signal-dependent. We are interested in estimating $\tau = a_n - nT - z_n$, given the observation y_n.

Suppose for a moment that we already know τ, and wish to characterize z_n and w_n. Around each sampling instance, we use a first-order Taylor expansion to approximate the

pulse shape.

- **In the amplitude domain:** We obtain

$$y_n \approx g(nT-\tau) + z_n g'(nT-\tau) + w_n. \tag{5.16}$$

  For convenience, let $u_n^{(\tau)} = z_n g'(nT-\tau)+w_n$. From statistical characterizations of the processes z_n and w_n, we can derive a statistical characterization of $u_n^{(\tau)}$ conditional on τ.

- **In the timing domain:** We obtain

$$a_n \approx nT + z_n - \tau + \frac{1}{g'(nT-\tau)} w_n. \tag{5.17}$$

  Let $v_n^{(\tau)} = z_n + w_n/g'(nT-\tau)$. We can similarly derive a statistical characterization of $v_n^{(\tau)}$.

In all cases of interest, both w_n and z_n are zero-mean stationary Gaussian processes [26,64]. Therefore, by linearization of the pulse shape around nT, both $u_n^{(\tau)}$ and $v_n^{(\tau)}$ are also zero-mean Gaussian processes. However, they are not stationary because their variances depend on the local slope of the pulse shape at each nT. The amplitude-domain additive noise w_n arises from Brownian motion in the circuitry, and hence is modeled as white noise. On the other hand, timing-domain additive noise z_n arises from aperture and clock phase jitter as we discussed in Section 5.2.3, and more restrictive models are often used.

The characterization of $u_n^{(\tau)}$ and $v_n^{(\tau)}$ requires knowledge of τ, which is the unknown, desired parameter. Therefore we use an iterative approach which refines our estimate of τ. The refinement of the estimate τ can be done in both domains. In the timing domain, given the characterization of the timing noise z_n, we can attempt to estimate it as we did in the previous subsection.

The linear approximation given above is valid when the jitter is small and the pulse is smooth. We can linearize around $z_n = 0$ to obtain

$$g(nT + z_n - \tau) \approx g(nT-\tau) + z_n g'(nT-\tau). \tag{5.18}$$

Then our signal model can be simplified to

$$y_n = g(nT-\tau) + u_n^{(\tau)}, \quad n = 0, \ldots, N-1, \tag{5.19}$$

where $u_n^{(\tau)} \sim N(0, (\sigma_z g'(nT-\tau))^2 + \sigma_w^2)$. This approximation is good for high SNR and signal-to-jitter ratio (SJR). Then for fixed τ, $u_n^{(\tau)}$ is white and Gaussian, but not identically distributed. Its variance depends on $g'(nT-\tau)$. For convenience, we write the variance as $(\sigma_n^{(\tau)})^2 = (\sigma_z g'(nT-\tau))^2 + \sigma_w^2$.

The ML estimation problem for the model of (5.19) is given by:

$$\hat{\tau}_{ML} = \arg\min_{\tau} \sum_{n=0}^{N-1} \left| \frac{y_n - g(nT - \hat{\tau} + \hat{z}_n)}{\sigma_n^{(\tau)}} \right|^2. \tag{5.20}$$

The problem of (5.20) is still hard to evaluate because of the dependence of $\sigma_n^{(\tau)}$ on both n and hypothesis τ via $g'[nT + z[n] - \tau]$. However, given a previous estimate $\hat{\tau}_k$, we can instead use $\sigma_n^{(\hat{\tau}_k)}$. We obtain the following iterative algorithm:

- Step 0: Start with initial estimate $\hat{\tau}_0$.
- Step 1: Given previous estimate $\hat{\tau}_k$, compute $\sigma_n^{(\hat{\tau}_k)}$.
- Step 2: Perform the following optimization:

$$\hat{\tau}_{k+1} = \arg\min_{\tau} \sum_{n=0}^{N-1} \left| \frac{y_n - g(nT - \hat{\tau}_k + \hat{z}_n^{(k)})}{\sigma_n^{(\tau)}} \right|^2 . \tag{5.21}$$

- Step 3: Given τ_{k+1}, find new estimate $\hat{z}_n^{(k+1)}$ given the characterization of $z[n]$.
- Step 4: Repeat as necessary, $k = k + 1$.

The optimization in step 2 is basically a cross-correlation which takes jitter into account, both as a penalty term and as a compensation term. This can also be implemented as a gradient search.

Given a spectral characterization of the timing noise z_n, the noise term v_n is a non-stationary, white, zero-mean Gaussian process. The optimal estimate of z_n in this case is still given by the Wiener filter [44]. The approach above can easily be modified to use different models of the timing noise z_n, such as when z_n is a deterministic periodic sequence with known or unknown period.

■ 5.7 Towards the Optimal Pulse Shape for Delay Estimation in the Presence of Timing Noise

Recall the basic example in which the pulse shape is given by $g(t) = at$ with $a \neq 0$ a fixed constant. Clearly, this function is invertible. It can easily be derived that the conditional distribution of y_n given τ can easily be derived to be normal with mean $a(nT - \tau)$ and variance $\sigma_w^2 + a^2\sigma_z^2$. Hence the ML estimate and Cramér-Rao bound are given by:

$$\hat{\tau}_{ML} = \frac{1}{N} \sum_{n=0}^{N-1} \left(\frac{y_n}{a} - nT \right), \quad \text{var}(\hat{\tau}) \geq \frac{\sigma_z^2 + \frac{\sigma_w^2}{a^2}}{N}. \tag{5.22}$$

In contrast to the case where only white timing noise is present, in this case the slope a determines the relative effect of the timing noise versus the additive noise: the steeper the slope, the more the additive noise is attenuated. This result should be compared with the non-parametric analytical result of Butzer [19] which we discussed in Section 5.2.1. Butzer's analysis of non-parametric signals suggests that steeper slopes are worse because it amplifies the effect of timing noise on the observations in the amplitude domain, where the error is measured.

In order to gain an insight into the best design of pulse shapes, consider a pulse consisting of two piecewise-linear regimes, denoted A_1 and A_2, as shown in Figure 5-4. The slopes at

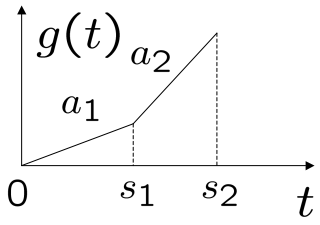


Figure 5-4. A simple pulse shape with two different slopes.

the different regimes are a_1 and a_2 respectively. Without loss of generality, let $a_2 \geq a_1$. Let n_1 and n_1 be the number of samples that fall within each regime.

The ML estimator can be derived from the log-likelihood function and is a linear combination of the samples $\{y_n\}_{n=0}^{N-1}$. The Cramér-Rao bound is given by

$$\mathrm{var}(\hat{\tau}) \geq \left(\frac{n_1}{\sigma_z^2 + \frac{\sigma_w^2}{a_1^2}} + \frac{n_2}{\sigma_z^2 + \frac{\sigma_w^2}{a_2^2}} \right)^{-1}. \tag{5.23}$$

The result of (5.23) can be used to obtain an insight into the performance of delay estimation in the presence of timing noise, for different pulse shapes. For fairness, we will fix the same N, T and signal energy. Let the signal energy of a pulse $g(t)$ be denoted by $\mathcal{E} = \int |g(t)|^2 \, dt$. Then it can be shown that for all values of N, T, $\mathcal{E}$, σ_z, and σ_w, the minimum value of the right hand side of (5.23) is achieved by $s_1 = 0$; it is strictly decreasing in s_1 by the assumption that $a_2 \geq a_1$, and for any fixed s_1 it is strictly decreasing in a_2 and a_1. This means that the triangle pulse is optimal in the sense that it gives the lowest Cramér-Rao bound for the estimate of pulse delay, which we know can be achieved with the estimator derived from the log-likelihood function.

■ 5.8 Simulation Results

We now verify the efficacy of our proposed algorithms via simulation. In several simulation sets, we emulate the signal and system parameters used in the MIT Microsystems Technology Laboratories (MTL) single-chip UWB transceiver system [11,13,119]. The system uses a Gaussian pulse of width 2 ns with 400 samples taken at sampling period 0.25 ns.

■ 5.8.1 Delay RMS vs. AWGN RMS

We begin by comparing the performance of the proposed method against the conventional method for various values of σ_w and σ_z. The signal parameters are based on the MTL UWB system. In this set of simulations the horizontal asymptotes correspond to the mean jitter value given the width of the pulse: in the timing domain, the mean of the jitter and the delay are not distinguishable from one another. We show this in Figure 5-5.

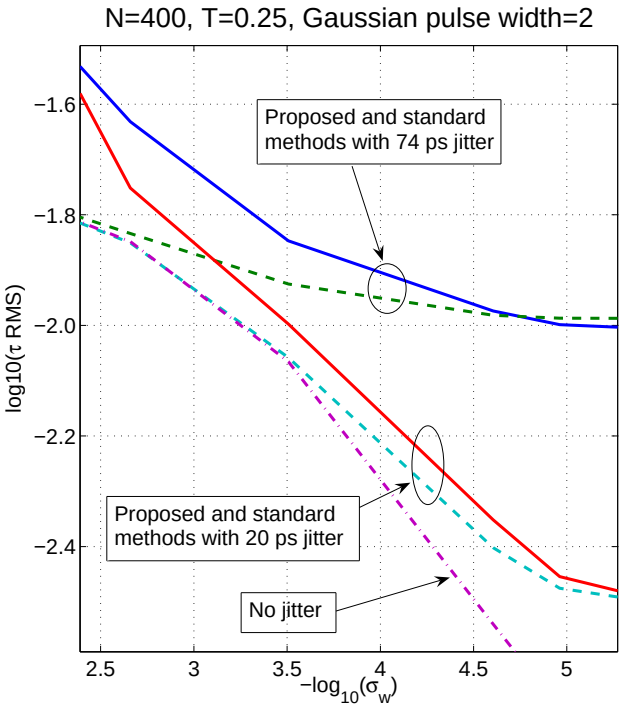


Figure 5-5. The performance of the proposed algorithm based on the MIT MTL single-chip UWB transceiver. The algorithm uses gradient search for the delay τ, and LMMSE for the jitter term. The system is simulated for different jitter RMS values. The system uses a Gaussian pulse of width 2 ns with 400 samples taken at every 0.25 ns.

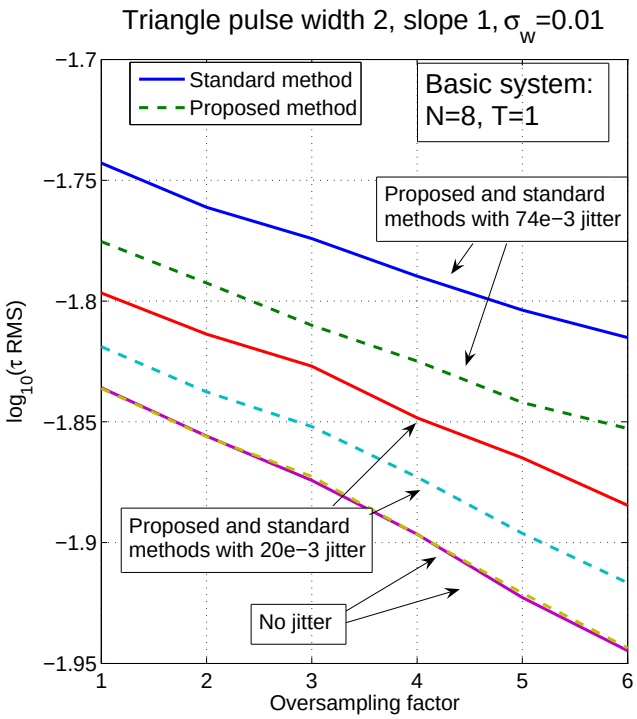


Figure 5-6. Simulation result of scaling of performance as a function of oversampling rate. The pulse shape is triangular.

■ 5.8.2 Delay RMS vs. Oversampling

Now we fix σ_w and σ_z, and show a family of curves as a function of oversampling. The "baseline" system uses $N = 8$ and $T = 1$. The oversampled signal increases N while keeping $N \times T$ fixed. The results are shown in Figure 5-6.

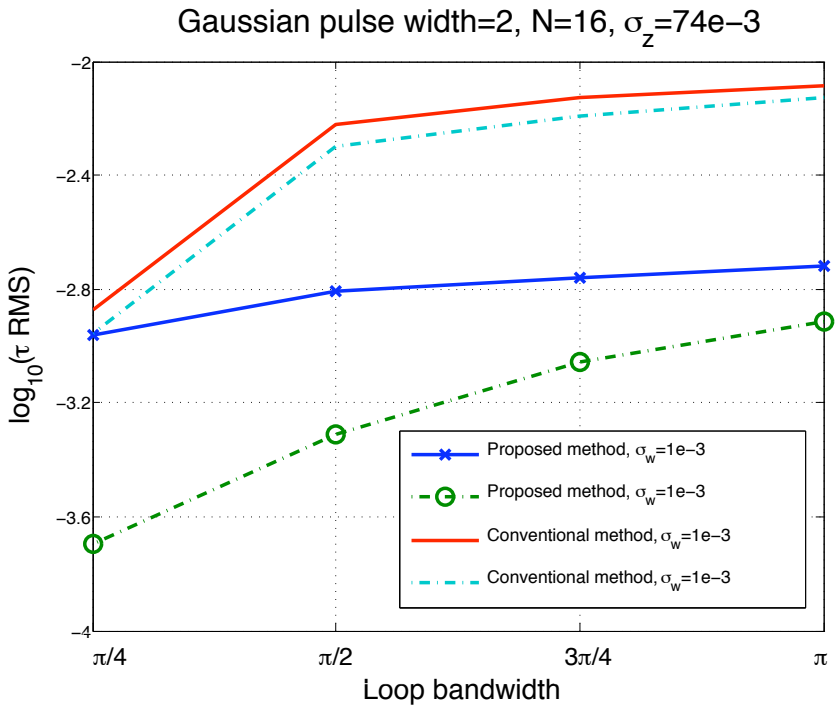


Figure 5-7. RMS of delay estimation. In this case z_n is bandlimited to digital frequencies shown on the X-axis, $N = 16$, $T = 0.5$, $\sigma_z = 74e-3$, and $\sigma_w = 1e-3, 1e-4$.

■ 5.8.3 Delay RMS vs. Loop Bandwidth

When we know that z_n has a certain loop bandwidth, we can take advantage of this in the estimation of z_n given $\hat{\tau}$. Timing noise with lower loop bandwidth is easier to estimate and cancel, hence we can obtain better performance. In comparing timing noise with different loop bandwidths, we make sure that the RMS of the timing noise remains the same. We show this in Figure 5-7. In this batch of simulations we set the mean of the jitter sequence to zero so that the mean jitter does not appear as a noise floor. We show the performance of the system when $N = 16$, $T = 0.5$, $\sigma_z = 74e-3$, and $\sigma_w = 1e-3$. The signal pulse shape is Gaussian with amplitude 2 and width 2. However, loop bandwidth does not appear to speed up the converge of the algorithm as seen in Figure 5-8 where we show the average performance at different iteration steps. The performance gains can be significant, as shown in Figures 5-9 and 5-10.

Using a similar system setup, we examine the effectiveness of jitter cancellation in Figure 5-9. We also used the signal parameters used in the MIT MTL UWB testbed and show the performance in Figure 5-10. The horizontal asymptote refers to the performance when no jitter is present in the system. In this batch of simulation we also make sure that the mean of the jitter sequence is zero. These systems are implemented using gradient search for the delay-estimation part.

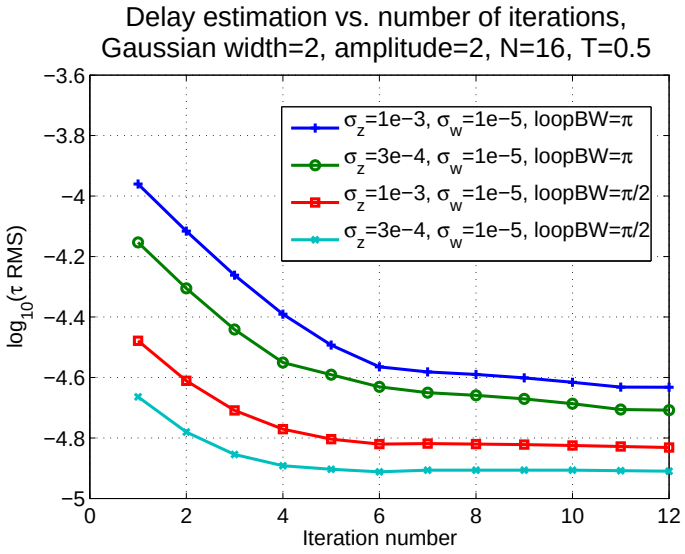


Figure 5-8. RMS of delay estimation versus number of steps. This set of simulations use $N = 16$, $T = 0.5$, for different σ_z, σ_w and loop bandwidth. This version uses gradient search in the observation domain, with jitter estimation updated every 6 gradient descent steps.

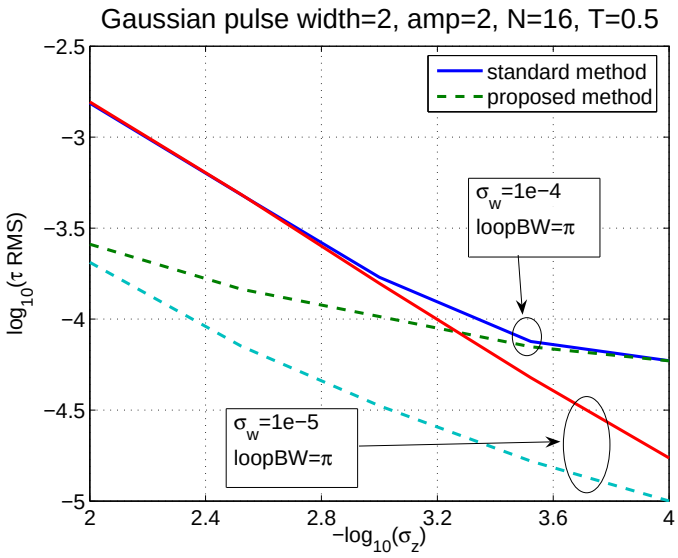


Figure 5-9. Performance of delay estimation versus jitter. In this case $N = 16$, $T = 0.5$, for different values of σ_w and σ_z. We show the effect of σ_w on the performance.

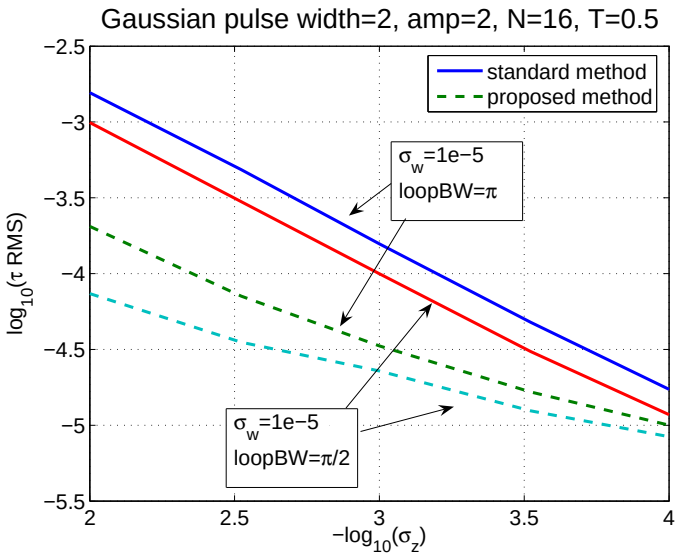


Figure 5-10. Performance of delay estimation versus jitter. In this case $N = 16$, $T = 0.5$, for different values of σ_w and σ_z. We show the effect of loop bandwidth on the performance.

■ 5.9 Conclusion

We have considered the problem of estimating parametric signals in the presence of timing noise, focusing on the classical problem of delay estimation. We proposed new algorithms which take into account the nature of timing noise, and showed that when timing noise is a prominent source of impairment, our new approach is superior to that of the old method that uses only cross-correlation. Unlike the classical approach which considers only AWGN and use LTI processing, our algorithm explicitly takes timing noise into account and uses an iterative approach to simplify the computation. The proposed approach can be easily modified to use more accurate models of timing noise.

Chapter 6

Closing remarks

This thesis revisited the problem of estimating signal parameters. It proposed the integration of sampling and estimation, and proposed several economical architectures with low sampling rate, requiring simple hardware and associated algorithms. We gave several analytical lower bounds on estimation error variance and proposed algorithms that give good performance for the cases of interest. This thesis also considered timing noise, which is a prominent source of impairment in wideband ADCs. We proposed algorithms that take timing noise into account, and showed that our algorithms are superior to that of the old method that uses only cross-correlation.

Most of the parameter estimation was done from a powersum series: the timing parameters are encoded into the roots in the series. This problem is called "Prony analysis," "exponential fitting" or "exponential analysis" in other fields of studies. Although this problem has been studied in line spectra estimation, translation of the result is not straightforward and can be misleading. We demonstrated that the proposed systems have *super-resolution* properties, meaning that their performances in the high-SNR case exceed that of the conventional system. Since the performance of the proposed system is dependent on the relative locations of the Diracs within the integration and sampling period, in many cases adaptation of the integration intervals can improve the performance of the system.

Finally, we showed that the analysis method that was developed can also be used to give new performance results on prior work of Vetterli *et al.* [117], in which a parametric signal is filtered and sampled uniformly in time. Since the desired parameters appear as complex roots of unity and thus lie on the unit circle, we were able to obtain simplification from trigonometric identities that gave insight into the behavior of the system.

■ 6.1 Algorithms

The problem of exponential fitting is not new and is well-studied in various fields ranging from natural sciences [4, 36, 50, 99, 110, 124], to medicine [74], statistics [82], mathematics, computer science [56], and engineering. Many algorithms are known, but as the work is widely distributed across many different fields, there is still no comprehensive study of which algorithms are suitable for which scenarios.

The 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 [48, 49, 71, 106], and on the case when we have a large number of samples with multiple snapshots. Since the parameter of greatest interest is in the angles

of the coefficients of the powersum series, 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 algorithms proposed to estimate parameters in an experiment for which they know the correct answer [7, 20, 50]. Moreover, the number of observations tends to be small. This is more suitable for our problems of interest, since we focus on the economy of small-scale problems and small numbers of observations. However, where appropriate we still use Monte-Carlo simulation and give the resulting mean-square error.

For an overview of the work in signal processing, we refer the reader to Stoica's textbook [106] and the newer work of Maravic and Vetterli [71]. For an overview of the work in natural sciences, we refer the reader to the recent overview by Istratov and Vyvenko [50]. It is an interesting exercise to see the overlap and non-overlap between the different fields.

■ 6.2 Analytical Challenges

We have seen that constraining the coefficients of the powersum series to lie on the real line instead of on the unit circle significantly adds to the difficulty of the analysis: no longer can we use trigonometric identities to decompose the problem into the form that we prefer. Translating the results from the case where the coefficients lie on the unit circle to when the coefficients lie on the real line is not straightforward and in some cases can be misleading. For example, one may be tempted to translate the analysis of Steedly, Ying and Moses [105], which considered complex-valued poles, to real-valued poles as follows: set the angle of a complex-valued coefficient to zero and consider the estimate of the magnitude to represent the coefficient itself on the real line. Their results suggest that the best performance is obtained when the magnitude is unity, or when the poles lie on the unit circle on the complex plane. However, we have seen in Section 3.5.2 and Theorem 3.4 that this approach is incorrect.

It is possible to obtain asymptotic results on the performance of the estimates as $\delta\omega \to \infty$ for line spectra estimation, such as in the papers by Dilaveroğlu [30] and Smith [103], although neither authors proposed an algorithm. We are not able to obtain analogous results for our cases of interest. Regardless, lower bounds on performance are most interesting when an algorithm that performs close to it is also presented, which we did in several cases.

■ 6.3 Future Work

Much of the contribution of this thesis is on rethinking the design of ADC for parametric signal acquisition. As such, more realistic device noise models should be studied. Moreover, a study of the actual implementation cost is in order to justify the economical efficacy of our proposed approach and also those that were given by other researchers.

■ 6.3.1 Generalizing sampling

Throughout this thesis we viewed sampling as inner products with a set of functions. These functions are chosen for two reasons:

1. That the operations are easy to implement.

2. That the signal parameters can easily be obtained from the mapping induced by the functions.

There is an infinite multitude of possible approaches for these two goals—we have only considered very specialized cases for which there are well-known simple devices such as integrators, and fast computational tools such as the Fast Fourier Transform. Broadly speaking, there are three avenues for further work:

1. Consider more general methods for obtaining samples of a signal, including methods that are not inner products.

2. Consider other systems of equations from which the parameters of a signal can be obtained, other than the powersum series.

3. Consider other devices for obtaining the desired inner products.

The last avenue is related to the connection between the multi-channel sampling scheme of Chapter 3 and the Dragotti-Vetterli-Blu sampling scheme of Section 4.2. Essentially, within the interval of interest the Dragotti scheme uses Strang-Fix theory to obtain the inner product that the multi-channel scheme obtained via integration. The primary difference is that the front-end of the Dragotti scheme is implemented as linear shift-invariant sampling followed by uniform sampling in time, whereas in the multi-channel scheme it is implemented using integrate-and-dump circuitry.

There are advantage and disadvantages to both schemes: in the multi-channel scheme the integrate-and-dump operates exactly on the interval of interest, but if we were to segmentize a signal, the integration intervals must be decided in advance.

By contrast, in the Dragotti scheme the intervals of interest can be decided after the samples are taken. However, by virtue of the sampling kernels being interpolating functions the footprints of the inner products are necessarily larger than the interval of interest. When there is continuous-time white noise, this means higher noise power. When there is unwanted signal within this overflow, then the estimate obtained will be incorrect.

Even if the sampling schemes are limited to inner products, it is easy to generate new sampling kernels by taking linear combinations of the proposed sampling kernels, as long as it can be recombined into the desired form.

■ 6.3.2 Successive refinement and detecting undermodeling

In Chapter 3 and Section 4.2 we considered two sampling schemes that are closely related. We call these the multichannel and Dragotti schemes respectively. In the latter, given N samples of the output of an R-order B-spline, it is possible to obtain a powersum series of length $M = \min\{N, R\}$. Therefore it is possible to obtain perfect reconstruction of a powersum series with up to $M/2$ components.

An attractive feature of the Dragotti scheme is its ability to segmentize a signal arbitrarily, and compute reconstructions of different orders at different segments. To start, suppose that we take N samples of a signal filtered using a B-spline kernel of order R within a segment $[0, T)$. Suppose that we segmentize the signal into segments $p = 0, 1, \ldots, P-1$ with N_p samples within each segment. Then within each segment we can compute a powersum series of length $M_p = \min\{N_p, R\}$, which in turn can be used to obtain reconstruction of a

powersum series with up to $M_p/2$ components within each segment. The segments can also overlap across time and scale.

However, care must be given to the fact that the reconstruction of polynomials have support outside its interval of interest. This means that we must account for this when segmenting a signal into several intervals of interest, within each we wish to reproduce polynomials of the desired orders. Hence, even if the segments overlap across time and scale, we must take into account the overlaps in the inner products.

■ 6.3.3 Reviewing timing and signaling

In Chapter 5, the last part of the thesis, we proposed an iterative algorithm that estimates both the signal parameter and the timing noise sequence. The viewpoint that we introduced is very general: timing-domain noise and observation-domain noise are two processes with different structures, and the map from one domain to the other is determined by the signal, which in turn is determined by its parameters.

We considered very simple cases when the timing noise sequence is a stationary stochastic process. Further extensions are possible. For example, we can consider timing noise to be a deterministic sequence or a non-stationary sequence admitting a special structure. When the timing noise sequence has memory, the estimate can be used to predict the timing noise term in other portions of the signal. Furthermore, the proposed algorithm can be used as a method for decoding covert messages encoded in the timing noise at the signal transmitter.

The preliminary work on designing transmission pulse shapes with timing noise and observation-domain additive noise that we explored in Section 5.7 can be extended to include other design considerations such as spectral efficiency and dynamic range.

Appendix A

Powersum Series

The nonlinear parameter estimation problems that we consider in this thesis are based on observations of form

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

The desired parameters are the K pairs $\{c_k, u_k\}$. A sequence of form (A.1) is called a *powersum series*, first studied by Baron de Prony in 1795 as he attempted to find the decay rate of chemical processes [25]. In the original problem, the observations and parameters are real-valued. This is sometimes called "real exponential fitting" in the natural sciences literature. De Prony showed that in the noiseless case it is possible to find $\{c_k, u_k\}$ exactly based on $N = 2K+1$ observations. In the signal processing literature, his method is called the *annihilating filter* method, which we review below. It is known that this method is not a consistent estimator of the parameters [51, 104].

In the noisy case we observe

$$y_n = \sum_{k=0}^{K-1} c_k (u_k)^n + w_n, \quad n = 0, \ldots, N-1. \tag{A.2}$$

We are interested in recovering the coefficients $\{c_k\}$ and the roots $\{u_k\}$. We assume that $c_k \in \mathbb{R}$ and u_k either $u_k \in \mathbb{R}$ or $\{u_k \in \mathbb{C} \mid \|u_k\| = 1\}$.

It is possible to rewrite (A.2) in matrix form:

$$\mathbf{y} = \mathbf{A}\mathbf{c} + \text{noise}, \tag{A.3}$$

where $[\mathbf{y}]_n = y_n$, $[\mathbf{c}]_k = c_k$, and $[\mathbf{A}]_{n,k} = (u_k)^n$. Throughout this document, we index matrices starting with 0 instead of 1 for consistency. Clearly, the matrix $\mathbf{A}$ is a Vandermonde matrix. Hence our problem of finding $\{u_k\}$ is equivalent to finding the elements of a Vandermonde matrix.

Several methods of solution have been proposed, which we cover below. Most of these methods first find $\{u_k\}$, and then use the estimate to find $\{c_k\}$. We are interested in two classes of algorithms: one based on Prony's method and one based on the Matrix Pencil method, also known as the Rotational Invariance Property [48, 49, 94]. These two classes of algorithms are closely related: see [35] for a quick overview of their similarities and differences. For low noise levels the Matrix Pencil-based methods are slightly superior to the Prony-based methods, and for high noise levels it is the other way around.

Several algorithms give performance close to the Cramér-Rao bounds in the presence of AWGN, such as the Nonlinear Least-Squares algorithm [108] and a regularized Maximum-Likelihood algorithm [35]. However, while their performance can exceed that of the Prony and Matrix Pencil-based methods, these algorithms requires 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.1 Prony's Method: Annihilating Filter Solution

Instead of presenting Prony's method in terms of a constant-coefficient difference equation, we use signal processing terminologies instead. The one-sided Z-transform of (A.1) is given by

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

In this form the parameters $\{u_k\}$ appear as poles of $X(z)$. Now let h_n be a Finite Impulse Response (FIR) filter of length K, with Z-transform $H(z)$. If the central values in the convolution of h_n and x_n are zero,

$$h_n * x_n = 0, \quad n = K+1, \ldots, N-1, \tag{A.4}$$

then the K zeros of $H(z)$ must equal the K poles of $X(z)$. In other words, up to a scaling factor we must have that:

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

The solution to (A.4) can be obtained by writing the convolution in matrix form. The rank of the Toeplitz matrix will be K, hence by writing a $K \times (K+1)$ matrix the solution is the nullspace of the matrix. If we have more measurements, we can write a $K \times (K+L)$ Toeplitz matrix instead, and obtain a nullspace of rank L. Each basis vector that spans this nullspace must satisfy (A.4). Hence we can improve the performance of the system by averaging the basis vectors.

When we have a large number of samples, we can estimate the covariance matrix of the samples. In the presence of white noise, the covariance matrix will have the same property, that is its rank is K. Hence we can use the same procedure to obtain the estimates of $\{u_k\}$. This is the main idea behind Pisarenko's algorithm and the MUSIC (MUltiple SIgnal Classification) algorithm [87, 98].

■ A.2 Matrix Pencil, Rotational Invariance Technique

Another method for finding the parameters of a powersum series is based on self-similarity, which we will make precise in the following. Let $\mathbf{X}$ be a $P \times Q$ Hankel matrix containing

the observations y_n, and let $P, Q > K$:

$$\mathbf{X} = \begin{bmatrix} y_0 & y_1 & \cdots & y_{Q-1} \\ y_1 & y_2 & \cdots & y_Q \\ \vdots & & & \vdots \\ y_{P-1} & y_P & \cdots & y_{P+Q-2} \end{bmatrix}.$$

One possible decomposition of $\mathbf{X}$ is $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^H$, where

$$[\mathbf{U}]_{i,j} = (u_j)^i, \mathbf{S} = \text{diag}(c_0, \ldots, c_{K-1}, 0, \ldots, 0), [\mathbf{V}]_{i,j} = (u_j)^i.$$

However, the factorization above is not unique. For example, $\mathbf{X} = (\mathbf{U}\mathbf{W}_1)\mathbf{W}_1^{-1}\mathbf{S}\mathbf{W}_2(\mathbf{W}_2^{-1}\mathbf{V}^H)$ is also a valid factorization for any unitary $\mathbf{W}_1, \mathbf{W}_2$. In other words, what we really want is a decomposition with a structural constraint, namely that $\mathbf{U}, \mathbf{V}^H$ are Vandermonde. Fortunately there is an algebraic structure that we can take advantage of.

Let $\mathbf{\Phi} = \text{diag}(u_k)$ be a $K \times K$ diagonal matrix containing the signal poles. Now let $\mathbf{b}$ be a row vector of all 1's. Because $\mathbf{U}, \mathbf{V}$ are Vandermonde, we can write:

$$\mathbf{U} = \begin{bmatrix} \mathbf{b} \\ \mathbf{b} \cdot \mathbf{\Phi} \\ \mathbf{b} \cdot \mathbf{\Phi}^2 \\ \vdots \\ \mathbf{b} \cdot \mathbf{\Phi}^{M-1} \end{bmatrix}, \mathbf{V} = \begin{bmatrix} \mathbf{b} \\ \mathbf{b} \cdot (\Phi^H) \\ \mathbf{b} \cdot (\Phi^H)^2 \\ \vdots \\ \mathbf{b} \cdot (\Phi^H)^{M-1} \end{bmatrix}.$$

Now we introduce truncations of these matrices. Let $\overline{(\cdot)}$ and $\underline{(\cdot)}$ denote the operation of omitting the first and last row of $(\cdot)$ respectively. Then it can be shown that

$$\overline{\mathbf{U}} = \underline{\mathbf{U}} \cdot \mathbf{\Phi}, \quad \overline{\mathbf{V}} = \underline{\mathbf{V}} \cdot \mathbf{\Phi}. \tag{A.5}$$

This is called the *rotational invariance property*. This property was used in line spectra estimation for finding frequencies which appear as $u_k = \exp(j\omega_k)$. We now show that this property also applies to $\mathbf{U}\mathbf{W}_1$ and $\mathbf{V}\mathbf{W}_2$. To see this, we write:

$$\mathbf{U}\mathbf{W}_1 = \begin{bmatrix} \mathbf{b}\mathbf{W}_1 \\ \mathbf{b}\mathbf{W}_1 \cdot \mathbf{W}_1^{-1}\mathbf{\Phi} \\ \mathbf{b}\mathbf{W}_1 \cdot \mathbf{W}_1^{-1}\mathbf{\Phi}^2\mathbf{W}_1 \\ \vdots \\ \mathbf{b}\mathbf{W}_1 \cdot \mathbf{W}_1^{-1}\mathbf{\Phi}^{M-1}\mathbf{W}_1 \end{bmatrix}, \mathbf{V}\mathbf{W}_2 = \begin{bmatrix} \mathbf{b}\mathbf{W}_2 \\ \mathbf{b}\mathbf{W}_2 \cdot (\mathbf{\Phi}^H)\mathbf{W}_2 \\ \mathbf{b}\mathbf{W}_2 \cdot (\mathbf{\Phi}^H)^2\mathbf{W}_2 \\ \vdots \\ \mathbf{b}\mathbf{W}_2 \cdot (\mathbf{\Phi}^H)^{M-1}\mathbf{W}_2 \end{bmatrix}.$$

Because $\mathbf{W}_1$ is unitary, $(\mathbf{W}_1^{-1}\mathbf{\Phi}\mathbf{W}_1)^k = \mathbf{W}_1^{-1}\mathbf{\Phi}^k\mathbf{W}_1$. Therefore,

$$\overline{\mathbf{U}\mathbf{W}_1} = \underline{\mathbf{U}\mathbf{W}_1} \cdot \mathbf{W}_1^{-1}\mathbf{\Phi}\mathbf{W}_1, \quad \overline{\mathbf{V}\mathbf{W}_2} = \underline{\mathbf{V}\mathbf{W}_2} \cdot \mathbf{W}_2^{-1}\mathbf{\Phi}\mathbf{W}_2. \tag{A.6}$$

Recall that we are interested in $\mathbf{P}hi$. Because the rotational invariance property applies to A.6 then it applies to any decomposition of the Hankel data matrix $\mathbf{X}$. For convenience, we pick the singular-value decomposition (SVD) of $\mathbf{X}$, and use the rotational invariance

property of the left- and right-singular matrices to obtain $\mathbf{\Phi}$.

In summary, the algorithm is as follows:

1. Form $\mathbf{X}$ the Hankel matrix from observations y_n.
2. SVD $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^H$, form principal left and right singular vectors $\mathbf{U}_s, \mathbf{V}_s$ corresponding to the K largest singular values of $\mathbf{X}$.
3. Form $\mathbf{Z} = \underline{\mathbf{U}_s}^\dagger \cdot \overline{\mathbf{U}_s}$.
4. Find $u_k = \text{eig}(\mathbf{Z})$.

Similarly, when the number of samples is large, we can estimate the covariance of the signal and use it in place of the data. This is the main idea behind ESPRIT (Estimation of Signal Parameters via Rotational Invariance Techniques) [94].

■ A.3 TLS-Prony Algorithm

We briefly review the algorithm proposed by Rahman and Yu, and analyzed by Steedly and Moses [88, 105].

Suppose that we are given observations $y[n], n = 0, \ldots, N-1$. Pick an integer $L > K$, recommended to be around $N/3$.

1. Form the Hankel matrix $\mathbf{Y}$ of size $(N - L \times L)$ from observations.
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 $(\cdot)^\dagger$ denote the pseudo-inverse.
4. Find the L roots of polynomial representation $\hat{B}(z)$, obtaining estimates $\hat{u}_\ell$ for $\ell = 0, 1, \ldots, L-1$.
5. Do least-squares fitting to find amplitudes $\hat{c}_\ell$ 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.

■ A.4 Cornell's Algorithm

Cornell [20] proposed a procedure for finding the coefficients of a powersum series from uniformly spaced observations based on segmenting and computing partial sums of the set of observations. He gave simple formulas for when $K = 1$ and $K = 2$, and Petersson and Holmström gave formulas for the case when $K = 3, 4$ [85]. In this thesis we review only the simple formulas for $K = 1$ and $K = 2$. It is known that the algorithm performs poorly for

large K. Suppose that we are given observations $y[n], n = 0, \ldots, N-1$. For convenience, let $N = 2^{2q}$ for some integer q.

For $K = 1$, the steps are given by:

1. Segment $y[n]$ and take sums:

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

2. Compute $W = S_2/S_1$.

3. Set estimate $\hat{u_0} = (W)^{N/2}$.

For $K = 2$, the steps are given by:

1. Segment $y[n]$ and take sums:

$$S_1 = \sum_{n=0}^{N/4-1} y[n], \quad S_2 = \sum_{N/4}^{N/2-1} y[n].$$

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

2. Compute

$$\begin{aligned} L_1 &= (S_1S_4 - S_2S_3)/(S_1S_3 - S_2^2), \\ L_2 &= (S_2S_4 - S_3^2)/(S_1S_3 - S_2^2). \end{aligned}$$

3. Find the roots of $x^2 - L_1x + L_2 = 0$, say W_0, W_1.

4. Set estimates $\hat{u}_0 = (W_0)^{N/4}, \hat{u}_1 = (W_1)^{N/4}$.

Cornell's algorithm requires only very mild conditions on the noise, that is that the noise is zero-mean. It does not even require that the noise is additive, but it does have to satisfy:

$$\mathbb{E}[y[n]] = \sum_{k=0}^{K-1} c_k(u_k)^n.$$

He showed that under this mild condition the algorithm is a consistent estimator. Cornell's algorithm has been extended by Agha [4] 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 has also been modified to allow for non-uniform spacing of samples by Foss [36].

Appendix B

Noise Analysis of Powersum Series

In this appendix we derive the performance limits of the estimation of the parameters of a powersum series. Recall from Section 2.4 that the Cramér-Rao Bound (CRB) is a powerful tool for finding the lower bound of estimation error for any unbiased estimator under some mild conditions. Further, recall that in the case of AWGN the CRB has a particularly simple form. We start with the noise analysis for the case when the roots are real-valued. This is the case of interest for Chapter 3, where the timing information appear as $u_k = (T - t_k)$. We continue by analyzing the case when the roots are complex roots of unity, and thus lie on the unit circle. This is useful for Chapter 4, where the timing information appears as $u_k = \exp(j2\pi t_k/T_p)$.

■ B.1 Real-valued Roots

We are interested in estimating parameters $\{u_k\}$ and $\{c_k\}$ from observations of the *noisy powersum*

$$y[n] = \sum_{k=1}^{K-1} c_k (u_k)^n + w[n], \quad n = 0, 1, \ldots, N-1. \tag{B.1}$$

■ B.1.1 Cramér-Rao Bounds

Consider the observation model (B.1) with all quantities real-valued. Suppose that additive noise $\mathbf{w}$ is zero-mean Gaussian with covariance matrix $\mathbf{\Sigma}$. Then the likelihood is:

$$L = \frac{1}{|\Sigma|^{1/2}} \exp\left(-(\mathbf{y} - \mathbf{A}\mathbf{c})^H \mathbf{\Sigma}^{-1} (\mathbf{y} - \mathbf{A}\mathbf{c})\right), \tag{B.2}$$

where $\mathbf{A}$ is a Vandermonde matrix as in Appendix A. We are interested in finding the Cramér-Rao bound for the estimating $\{u_k\}$.

To derive the Fisher information matrix (FIM) we start with the log-likelihood:

$$-\ln L = \mathbf{y}^H \mathbf{\Sigma}^{-1} \mathbf{y} + \mathbf{c}^H \mathbf{A}^H \mathbf{\Sigma}^{-1} \mathbf{A}\mathbf{c} - \mathbf{c}^H \mathbf{A} \mathbf{\Sigma}^{-1} \mathbf{y} - \mathbf{y}^H \mathbf{\Sigma}^{-1} \mathbf{c}.$$

For convenience, define

- $\mathbf{F}_{\mathbf{c}} = \partial_{\mathbf{c}}^T \partial_{\mathbf{c}} \ln L$.
- $\mathbf{F}_{\mathbf{u},\mathbf{c}} = \partial_{\mathbf{u}}^T \partial_{\mathbf{c}} \ln L$
- $\mathbf{F}_{\mathbf{u}} = \partial_{\mathbf{u}}^T \partial_{\mathbf{u}} \ln L$.

Then we obtain that the FIM is given by,

$$\mathbf{F}_{\mathbf{c},\mathbf{u}} = \mathbb{E}\begin{bmatrix} \mathbf{F}_{\mathbf{c}} & \mathbf{F}_{\mathbf{u},\mathbf{c}}^T \\ \mathbf{F}_{\mathbf{u},\mathbf{c}} & \mathbf{F}_{\mathbf{u}} \end{bmatrix}. \tag{B.3}$$

We make the following definition:

- $\boldsymbol{\Xi} = \mathbf{A}^H \boldsymbol{\Sigma}^{-1} \mathbf{A}$,
- following [103] let $\boldsymbol{\Xi}_{[\mathbf{u}]}$ denote the substitution of the first derivative w.r.t. u_k when the k-th index of $\mathbf{u}$ is required,
- and similarly let let $\boldsymbol{\Xi}_{[\mathbf{u},\mathbf{u}]}$ denote the substitution of the first derivative w.r.t. u_k and u_ℓ when the k, ℓ-th index of $\mathbf{u}$ is required.

Then we can derive:

$$-\begin{pmatrix} \partial_{\mathbf{c}}^T \\ \partial_{\mathbf{u}}^T \end{pmatrix} \ln L = \begin{pmatrix} (2\mathbf{c}^H \boldsymbol{\Xi} + \mathbf{y}^H \boldsymbol{\Sigma}^{-1} \mathbf{A}) \\ (\mathbf{c}^H \boldsymbol{\Xi}_{\mathbf{u}} \mathbf{c} - \mathbf{y}^H \boldsymbol{\Sigma}^{-1} \mathbf{A}_{\mathbf{u}} \mathbf{c} - \mathbf{c}^H \mathbf{A}_{\mathbf{u}}^H \boldsymbol{\Sigma}^{-1} \mathbf{y}) \end{pmatrix} \tag{B.4}$$

$$-\partial_{\mathbf{c}}^T \partial_{\mathbf{c}} \ln L = 2\boldsymbol{\Xi} \tag{B.5}$$

$$-\partial_{\mathbf{c}} \partial_{\mathbf{u}}^T \ln L = 2\boldsymbol{\Xi}_{\mathbf{u}} \mathbf{c} - 2\mathbf{A}_{[\mathbf{u}]}^H \boldsymbol{\Sigma}^{-1} \mathbf{y} \tag{B.6}$$

$$-\partial_{\mathbf{u}}^T \partial_{\mathbf{u}} \ln L = \mathbf{c}^H \boldsymbol{\Xi}_{[\mathbf{u},\mathbf{u}]} \mathbf{c} - \mathbf{c}^H \mathbf{A}_{[\mathbf{u},\mathbf{u}]} \boldsymbol{\Sigma}^{-1} \mathbf{y} - \mathbf{y}^H \boldsymbol{\Sigma}^{-1} \mathbf{A}_{[\mathbf{u},\mathbf{u}]} \mathbf{c}. \tag{B.7}$$

Now we are ready to find the Fisher information matrix by taking the expectation of the above. To evaluate the above expression, note that the noise is assumed to be zero mean, therefore we simply need to substitute $\mathbb{E}[\mathbf{y}] = \mathbf{A}\mathbf{c}$.

To finally obtain the CRB, note that the FIM given in (B.3) is of block matrix form. Therefore we can use the Schur complement of the FIM to obtain a lower bound on the estimate of $\mathbf{u}$:

$$\mathbb{E}[(\hat{\mathbf{u}} - \mathbf{u})(\hat{\mathbf{u}} - \mathbf{u})^T] \geq (\mathbf{F}_{\mathbf{u}} - \mathbf{F}_{\mathbf{u},\mathbf{c}}^T \mathbf{F}_{\mathbf{c}}^{-1} \mathbf{F}_{\mathbf{u},\mathbf{c}})^{-1}. \tag{B.8}$$

■ B.1.2 White noise case

When the noise is white we obtain a simpler expression for the FIM (for example see [52]). For convenience, let θ be the vector of parameters and $x[n;\theta]$ be the noiseless signal given by θ. In this case,

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

The partial derivative is particularly simple:

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

Then,

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

Let $\mathbf{p}[n] = [(u_0)^n, \ldots, (u_{K-1})^n, c_0 \cdot n \cdot (u_0)^{n-1}, \ldots, c_{K-1} \cdot n \cdot (u_{K-1})^{n-1}]^T$. Therefore,

$$\mathbf{J} = \frac{1}{\sigma^2}\left[\sum_n \mathbf{p}^T[n]\mathbf{p}[n]\right]. \tag{B.9}$$

To examine the resolution limit, following Dilaveroğlu [30] we wish to derive the performance for $K = 2$ in terms of $\delta_u = u_1 - u_2$. However, Diraveroğlu 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, ω_2 that depends on $\delta\omega = \omega_1 - \omega_2$, but independent of absolute terms ω_1, ω_2. By contrast, we do not obtain such convenient decompositions and have to rely on numerical evaluation.

■ B.1.3 Single-pole case

We now use the result to consider the simplest case when there is only one pole in the signal. First consider the case when $w[n]$ is white Gaussian with variance σ^2. We can write the FIM as:

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

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:

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

The behavior of (3.6) is consistent with the behavior of frequency estimates in line spectra estimation which we reviewed in Section 2.1. However, a caveat is in order here. Suppose for now that the signal of interest consists of complex-valued poles. This case was studied by Steedly and Moses in [105]. 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. Clearly, this is a different behavior from what we obtained in this section. This illustrates that translating the results from the complex-valued case to the real-valued case is not straightforward and can be misleading.

A caveat is in order here. Translation of the currently known results to our specific case is not straightforward and can be misleading. When the poles are complex-valued, the lack of knowledge of the exact pole angle lead 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

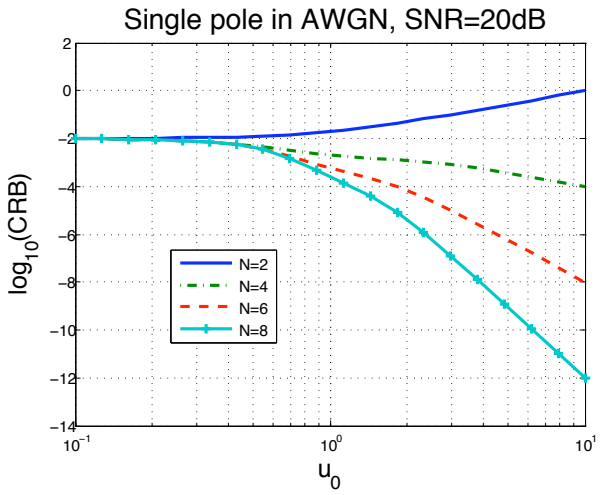


Figure B-1. The Cramér-Rao bound for the estimation of a single pole with unit amplitude and AWGN with RMS 0.1 per sample.

magnitude of the pole, as shown in Figure 2 of [105]. Hence in the complex case, the error in pole magnitude estimate is best for unit magnitude, and becomes worse as the true pole magnitude increases. In our case, the poles have positive real values. The error in the pole value (magnitude) estimate decreases as the pole value (magnitude) is increased, as there is no phase ambiguity. We show this in Figure B-1.

■ B.1.4 Two-pole case

We now consider the case when $K = 2$. Then $y[n] = c_0(u_0)^n + c_1(u_1)^n + w[n]$. For convenience, we define the following:

$$\mathbf{R} = \begin{bmatrix} \sqrt{N} & & & \\ & \sqrt{N} & & \\ & & N\sqrt{N} & \\ & & & N\sqrt{N} \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & (c_0/u_0) & \\ & & & (c_1/u_1) \end{bmatrix}, \tag{B.10}$$

and

$$G_r(x) = \frac{1}{N^{r+1}} \sum_{n=0}^{N-1} n^r (x)^n. \tag{B.11}$$

Then the FIM can be written as,

$$\mathbf{J} = \frac{1}{\sigma^2} \cdot \mathbf{R} \cdot \mathbf{S} \cdot \left[\begin{array}{cc|cc} G_0(u_0^2) & G_0(u_0u_1) & G_1(u_0^2) & G_1(u_0u_1) \\ G_0(u_0u_1) & G_0(u_1^2) & G_1(u_0u_1) & G_1(u_1^2) \\ \hline G_1(u_0^2) & G_1(u_0u_1) & G_2(u_0^2) & G_2(u_0u_1) \\ G_1(u_0u_1) & G_1(u_1^2) & G_2(u_0u_1) & G_2(u_1^2) \end{array}\right] \cdot \mathbf{S} \cdot \mathbf{R}. \tag{B.12}$$

For convenience, let

$$\mathbf{A} = \begin{bmatrix} G_0(u_0^2) & G_0(u_0u_1) \\ G_0(u_0u_1) & G_0(u_1^2) \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} G_1(u_0^2) & G_1(u_0u_1) \\ G_1(u_0u_1) & G_1(u_1^2) \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} G_2(u_0^2) & G_2(u_0u_1) \\ G_2(u_0u_1) & G_2(u_1^2) \end{bmatrix},$$

Let $\mathrm{SNR}_k = c_k^2/\sigma^2$. Then the CRB is given by,

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

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

■ B.2 Complex Roots on the Unit Circle

In this section we examine the case when the roots of the powersum series are complex roots of unity. Unlike in the previous case where we were forced to make a numerical evaluation, in this section we are able to obtain useful factorizations and gain further insight into the problem at hand. We assume that the observation is subjected to AWGN for this section.

■ B.2.1 Single-component case

We prove Theorem 4.1. It can be shown that the FIM is given by:

$$\mathbf{J} = \frac{2}{\sigma^2}\left(\begin{bmatrix}\sqrt{N} & \\ & N\sqrt{N}\end{bmatrix}\begin{bmatrix}1 & \\ & \alpha_0\end{bmatrix}\begin{bmatrix}\Gamma(0) & 0 \\ 0 & \Gamma(2)\end{bmatrix}\begin{bmatrix}1 & \\ & \alpha_0\end{bmatrix}\begin{bmatrix}\sqrt{N} & \\ & N\sqrt{N}\end{bmatrix}\right). \tag{B.13}$$

The inverse is given by:

$$\mathbf{J}^{-1} = \frac{\sigma^2}{2}\begin{bmatrix}\frac{1}{N}\frac{1}{\Gamma_0} & 0 \\ 0 & \frac{1}{N^3}\frac{1}{\alpha_0^2\Gamma(2)}\end{bmatrix}. \tag{B.14}$$

We have then obtained Theorem 4.1.

■ B.2.2 Resolution of FRI method

Now we consider Theorem 4.2. For convenience we define

$$\alpha_0 = \left(c_0\frac{2\pi}{T_p}\right), \quad \alpha_1 = \left(c_1\frac{2\pi}{T_p}\right), \tag{B.15}$$

and

$$\mathbf{K} = \begin{bmatrix}\sqrt{N} & & & \\ & \sqrt{N} & & \\ & & N\sqrt{N} & \\ & & & N\sqrt{N}\end{bmatrix}, \quad \mathbf{L} = \begin{bmatrix}1 & & & \\ & 1 & & \\ & & \alpha_0 & \\ & & & \alpha_1\end{bmatrix} \tag{B.16}$$

We segmentize the FIM as follows:

$$\mathbf{J} = \frac{2}{\sigma^2}\left(\mathbf{K}\cdot\mathbf{L}\cdot\begin{bmatrix}\mathbf{A} & \mathbf{b} \\ \mathbf{b}^T & \mathbf{c}\end{bmatrix}\cdot\mathbf{L}\cdot\mathbf{K}\right). \tag{B.17}$$

After some algebra, we obtain

$$\mathbf{A} = \begin{bmatrix} \Gamma(0) & C(0) \\ C(0) & \Gamma(0) \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 0 & -S(1) \\ S(1) & 0 \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} \Gamma(2) & C(2) \\ C(2) & \Gamma(2) \end{bmatrix}$$

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

$$\mathbf{J}^{-1} = \frac{2}{\sigma^2} \left(\mathbf{K}^{-1} \cdot \mathbf{L}^{-1} \cdot \begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{b}^T & \mathbf{c} \end{bmatrix}^{-1} \cdot \mathbf{L}^{-1} \cdot \mathbf{K}^{-1} \right). \tag{B.18}$$

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 $\mathbf{J}$:

$$\begin{aligned} \mathbf{P} &= \mathbf{c} - \mathbf{b}^T \mathbf{A}^{-1} \mathbf{b}. \\ &= \begin{bmatrix} \Gamma_2 & C_2 \\ C_2 & \Gamma_2 \end{bmatrix} - \frac{1}{\Gamma_0^2 - C_0^2} \begin{bmatrix} S_1^2 \Gamma_0 & S_1^2 C_0 \\ S_1^2 C_0 & S_1^2 \Gamma_0 \end{bmatrix}. \\ \mathbf{M} &= \begin{bmatrix} \Gamma_2(\Gamma_0^2 - C_0^2) - S_1^2 \Gamma_0 & C_2(\Gamma_0^2 - C_0^2) - S_1^2 C_0 \\ C_2(\Gamma_0^2 - C_0^2) - S_1^2 C_0 & \Gamma_2(\Gamma_0^2 - C_0^2) - S_1^2 \Gamma_0 \end{bmatrix} \\ \mathbf{P} &= \frac{1}{\Gamma_0^2 - C_0^2} \mathbf{M}. \\ \mathbf{P}^{-1} &= (\Gamma_0^2 - C_0^2) \mathbf{M}^{-1}. \end{aligned}$$

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

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

Let $X_0 = \Gamma_0^2 - C_0^2$. Then,

$$[\mathbf{P}]_{k,k} = \frac{\Gamma_2 X_0 - S_1^2 \Gamma_0 X_0}{\Gamma_2^2 X_0^2 - 2 S_1^2 \Gamma_0 X_0 + S_1^4 \Gamma_0^2 - C_2^2 X_0^2 + 2 S_1^2 C_0 C_2 X_0 - S_1^4 C_0^2}.$$

Appendix C

List of Acronyms

ADC	Analog-to-Digital Converter
AWGN	Additive White Gaussian Noise
CDMA	Code-Division Multiple Access
CMOS	Complementary-symmetry/Metal-Oxide Semiconductor
CRB	Cramér-Rao Bound
DFT	Discrete Fourier Transform
EM	Expectation-Maximization algorithm
ENOB	Effective Number Of Bits
FFT	Fast Fourier Transform
FIM	Fisher Information Matrix
FRI	Finite Rate of Innovation
GPS	Global Positioning System
IID	Independent and Identically Distributed
MIT	Massachusetts Institute of Technology
ML	Maximum Likelihood
RMS	Root Mean Square
SHA	Sample-and-Hold Apperture
SNR	Signal-to-Noise Ratio
UWB	Ultra-wideband

Bibliography

[1] The ABCs of ADCs: Understanding how ADC errors affect system performance. Technical Report 748, Maxim IC, 2001.

[2] Loop bandwidth and clock data recovery (CDR) in oscilloscope measurements. Technical Report 1304-6, Agilent Technologies, 2002.

[3] Analog-to-Digital Converter design guide. Technical report, Microchip Technology, Inc., 2004.

[4] M. Agha. A direct method for fitting linear combinations of exponentials. *Biometrics*, 27(2):399–413, June 1971.

[5] B. Baker. When undersampling, clock jitter does matter. EDN Magazine, August 2005.

[6] A. V. Balakrishnan. On the problem of time jitter in sampling. *IRE Trans. Inf. Theory*, 3:226–236, April 1962.

[7] L. G. Beavais and S. J. Lippard. Reactions of the peroxo intermediate of soluble methane monooxygenase hydroxylase with ethers. *J. Amer. Chem. Soc.*, 127(20):7371–7378, 2005.

[8] F. J. Beutler and O. A. Z. Leneman. Random sampling of random processes: Stationary point processes. *Information and Control*, 9:325–346, 1966.

[9] F. J. Beutler and O. A. Z. Leneman. The spectral analysis of impulse processes. *Information and Control*, 12:236–258, 1968.

[10] I. Bilinskis and A. Mikelsons. *Randomized Signal Processing*. Prentice Hall, United Kingdom, 1992.

[11] R. Blazquez. *Ultra-wideband Baseband*. PhD thesis, Massachusetts Institute of Technology, 2006.

[12] R. Blazquez, F. S. Lee, P. Newaskar, J. Powell, D. Wentzloff, and A. Chandrakasan. Digital architecture for an ultra-wideband radio receiver. In *IEEE Vehicular Technology Conference*, 2003.

[13] R. Blazquez, P. P. Newaskar, F. S. Lee, and A. P. Chandrakasan. A baseband processor for impulse ultra-wideband communications. *IEEE Journal of Solid-State Circuits*, 40(9):1821 – 1828, September 2005.

[14] B. Brannon. Aperture uncertainty and ADC system performance. Technical Report AN-501, Analog Devices, 2000.

[15] B. Brannon. Sampled systems and the effects of clock phase noise and jitter. Technical Report AN-756, Analog Devices, 2004.

[16] P. Brémaud, L. Massoulié, and A. Ridolfi. Power spectra of random spike fields and related processes. *Journal of Applied Probability*, 2005.

[17] P. Brémaud and A. Ridolfi. Power spectral measure and reconstruction error of randomly sampled signals. In *Proc. of IEEE Information Theory Workshop 2002*, pages 159–162, 2002.

[18] W. M. Brown. Sampling with random jitter. *Journal of SIAM*, 11(2):460–473, June 1963.

[19] P. L. Butzer. A survey of the Whittaker-Shannon sampling theorem and some of its extensions. *J. Math. Res. Expos.*, 3:183–212, 1983.

[20] R. G. Cornell. A method for fitting linear combinations of exponentials. *Biometrica*, 18(1):104–113, March 1962.

[21] H. Cramér. *Mathematical Methods of Statistics*. Princeton University Press, Princeton, NJ, 1946.

[22] R. Jean-Marc Cramer, Robert A. Scholtz, and Moe Z. Win. Evaluation of an ultra wide-band propagation channel. *IEEE Trans. on Antennas and Propagation*, 50(5):561 – 569, May 2002.

[23] N. Da Dalt. Effect of jitter on asynchronous sampling with finite number of samples. *IEEE Trans. Circuits and Systems II*, 51(12):660–664, December 2004.

[24] N. Da Dalt, M. Harteneck, C. Sandner, and A. Wiesbauer. On the jitter requirements of the sampling clock for analog-to-digital converters. *IEEE Trans. Circuits and Systems I*, 49(9):1354–1360, September 2002.

[25] Gaspard Riche de Prony. Essai éxperimental et analytique: sur les lois de la dilatabilit de fluides élastique et sur celles de la force expansive de la vapeur de l'alkool, à différentes températures. *Journal de l'École Polytechnique*, 1(22):24–76, 1795.

[26] A. Demir, A. Mehrotra, and J. Roychowdhury. Phase noise in oscillators: A unifying theory and numerical methods for characterization. *IEEE Trans. on Circuits and Systems I*, 47(5):655–674, May 2000.

[27] A. Dempster. Aperture jitter effects in software radio GNSS receivers. *Journal of GPS*, 3(1-2):45–48, 2004.

[28] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc., Ser. B*, 39(1):1–38, 1977.

[29] A. Deutsch. Electrical characteristics of interconnections for high-performance systems. *Proc. IEEE*, 86(2):315–355, February 1998.

[30] E. Dilaveroğlu. Nonmatrix Cramér-Rao bound expressions for high-resolution frequency estimators. *IEEE Transactions on Signal Processing*, 46(2):463–475, February 1998.

[31] P. L. Dragotti and M. Vetterli. Wavelet and footprint sampling of signals with a finite rate of innovation. In *IEEE Conference on Acoustics, Speech and Signal Processing*, volume 2, pages 941–944, Montreal, Canada, May 2004.

[32] P. L. Dragotti, M. Vetterli, and T. Blu. Exact sampling results for signals with finite rate of innovation using Strang-Fix conditions and local kernels. In *Proc. IEEE Conf. on Acoustics, Speech, and Signal Proc.*, Philadelphia, PA, March 2005.

[33] P. L. Dragotti, M. Vetterli, and T. Blu. Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets Strang-Fix. *IEEE Transactions on Signal Processing*, 2006. Submitted for publication.

[34] B. Drakhlis. Calculate oscillator jitter by using phase-noise analysis. *Microwaves & RF*, pages 82–90, 157, January 2001.

[35] M. Elad, P. Milanfar, and G. H. Golub. Shape from moments – an estimation theory perspective. *IEEE Transactions on Signal Processing*, 52(7):1814–1829, July 2004.

[36] S. D. Foss. A method of exponential curve fitting by numerical integration. *Biometrics*, 26(4):815–821, December 1970.

[37] E. G. Friedman. Clock distribution networks in synchronous digital integrated circuits. *Proc. of the IEEE*, 89(5):665–692, May 2001.

[38] S. Gezici, Z. Tian, G. B. Giannakis, H. Kobayashi, A. F. Molisch, H. V. Poor, and Z. Sahinoglu. Localization via ultra-wideband radios: A look at positioning aspects for future sensor networks. *IEEE Signal Processing Magazine*, 22(4):70–84, July 2005.

[39] M. Ghogho and A. Swami. Fast computation of the exact FIM for deterministic signals in colored noise. *IEEE Transactions on Signal Processing*, 47(1):52–61, January 1999.

[40] R. Hahnloser, R. Sarpeshkar, M. Mahowald, R. Douglas, and S. Seung. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature*, 405:947–951, June 2000.

[41] A. Hajimiri and T. Lee. *The Design of Low Noise Oscillators*. Kluwer, Norwell, MA, 1999.

[42] A. Hajimiri and T. H. Lee. A general theory of phase noise in electrical oscillators. *IEEE Journal of Solid-State Circuits*, 33(2):179–194, February 1998.

[43] D. Ham, W. Andress, and D. S. Ricketts. Phase noise in oscillators. In *Intl. Workshop on SiP/SoC Integration of MEMS and Passive Components with RF-ICs*, March 2004. Invited paper.

[44] M. Hayes. *Digital Signal Processing and Modeling*. Wiley, New York, NY, 1996.

[45] J. R. Higgins. *Sampling Theory in Fourier and Signal Analysis*. Oxford Science Publications, Oxford, UK, 1996.

[46] M. Horowitz, C. K. K. Yang, and S. Sidiropoulos. High-speed electrical signaling – overview and limitations. *IEEE Micro*, 18(1):12–24, 1998.

[47] P. Horowitz and W. Hill. *The Art of Electronics*. Cambridge University Press, 2 edition, 1989.

[48] Y. Hua and T. K. Sarkar. Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Transactions on Signal Processing*, 38(5):814–824, May 1990.

[49] Y. Hua and T. K. Sarkar. On SVD for estimating generalized eigenvalues of singular matrix pencil in noise. *IEEE Transactions on Signal Processing*, 39(4):892–900, 1991.

[50] A. A. Istratov and O. F. Vyenko. Exponential analysis in physical phenomena. *Review of Scientific Instruments*, 70(2):1233–1257, February 1999.

[51] M. Kahn, M. Mackisack, M. S. Osborne, and G. K. Smyth. On the consistency of Prony's method and related algorithms. *Journal of Computational and Graphical Statistics*, 1:329–349, 1992.

[52] S. M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall, New Jersey, 1993.

[53] V. A. Kotelnikov. On the carrying capacity of the ether and wire in telecommunications. In *Material for the First All-Union Conference on Questions of Communication*, Izd. Red. Upr. Svyazi RKKA, Moscow, 1933.

[54] R. Kumaresan and D. W. Tufts. Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-30:833–840, December 1982.

[55] R. Kumaresan and D. W. Tufts. Estimating the angles of arrival of multiple plane waves. *IEEE Trans. Aerospace and Elec. Syst.*, AES-19(1):134–139, January 1983.

[56] D. Kundu and N. Kannan. Constrained maximum likelihood estimators for superimposed exponential signals. *Comm. Statist. Simulation Comput.*, 26(2):733–764, 1997.

[57] S. Y. Kung, K. S. Arun, and D. V. B. Rao. State-space and singular-value decomposition-based approximation methods for the harmonic retrieval problem. *Journal of the Optical Society of America*, 73(12):1799 – 1811, December 1983.

[58] J. Kusuma and V. K Goyal. Multichannel sampling for parametric signals with a successive approximation property. In *IEEE Conf. on Image Processing*, August 2006. Accepted for publication.

[59] J. Kusuma and V. K Goyal. Signal parameter estimation in the presence of timing noise. In *Conference on Information Sciences and Systems*, March 2006. Accepted for publication.

[60] J. Kusuma, I. Maravić, and M. Vetterli. Sampling with finite rate of innovation: Applications to GPS and UWB. In *Proc. IEEE Conference on Communications*, pages 3540–3544, July 2003.

[61] J. Kusuma, A. Ridolfi, and M. Vetterli. Sampling of communication systems with bandwidth expansion. In *Proc. of the IEEE Conference on Communication*, pages 1601 – 1605, April 2002.

[62] T. Lee and A. Hajimiri. Oscillator phase noise: A tutorial. *IEEE Journal of Solid-State Circuits*, 35(3):326–336, March 2000.

[63] T. H. Lee. Oscillator phase noise: A tutorial. In *Proc. IEEE Custom Integrated Circuits*, pages 373–380, San Diego, CA, May 1999.

[64] D. B. Leeson. A simple model of feedback oscillator noise spectrum. *Proceedings of the IEEE*, 54(2):329–330, February 1966.

[65] O. A. Z. Leneman. Random sampling of random processes: Impulse processes. *Information and Control*, 9:347–363, 1966.

[66] M. S. Lewicki. A review of methods for spike sorting: the detection and classification of neural action potentials. *Network: Computation in Neural Systems*, 9(4):53–78, 1998.

[67] M. Löhning and G. Fettweis. The effects of aperture jitter and clock jitter in wideband ADCs. In *Proc. International Workshop on ADC Modelling and Testing*, 2003.

[68] H. D. Lüke. The origins of the sampling theorem. *IEEE Communications Magazine*, pages 106–108, April 1999.

[69] I. Maravić, J. Kusuma, and M. Vetterli. Low-sampling rate channel estimation and timing for UWB systems. *Journal of Communication Networks*, 5(4):319–327, December 2003.

[70] I. Maravić and M. Vetterli. Digital DS-CDMA receivers working below the chip rate: Theory and design. *IEEE Transactions on Communications*, 2002. In review.

[71] I. Maravić and M. Vetterli. Sampling and reconstruction of signals with finite rate of innovation in the presence of noise. *IEEE Transactions on Signal Processing*, 53(8):2788–2805, July 2005.

[72] P. Marziliano. *Sampling innovations*. PhD thesis, Ecole Polytechnique Fédérale de Lausanne (EPFL), April 2001.

[73] S. E. Meninger and M. H. Perrott. A 1-MHz bandwdith 3.6-GHz 0.18-μm CMOS fractional-N synthesizer utilizing a hybrid PFD/DAC structure for reduced broadband phase noise. *IEEE Journal of Solid-State Circuits*, 41(4):966–980, April 2006.

[74] B. B. Mittal, V. Sathiaseelan, A. W. Rademaker, M. C. Pierce, P. M. Johnson, and W. N. Brand. Feasibility studies of an ingestible telemetric temperature sensor for deep hyperthermia applications. *Int. J. Radiation Oncology Biology and Physics*, 21(5):1353–1361, 1991.

[75] A. F. Molisch. *Wireless Communications*. Wiley-IEEE Press, 2005.

[76] A. F. Molisch and I. Oppermann, editors. *UWB Communication Systems: A Comprehensive Overview*. Hindawi Publishing Corporation, 2006.

[77] N. Narasimhamurti and S. Awad. Estimating the parameters of a sinusoid sampled with a clock with accumulated jitter. In *Proc. IEEE Instrumentation and Measurement Tech. Conf.*, pages 1132–1135, May 1997.

[78] P.P. Newaskar, R. Blazquez, and A.R. Chandrakasan. A/D precision requirements for an ultra-wideband radio receiver. In *IEEE Workshop on Signal Processing Systems (SIPS)*, pages 270 – 275, October 2002.

[79] H. Nyquist. Certains topics in telegraph transmission theory. *Trans. AIEE*, 47:617–644, April 1928.

[80] I. Obeid and P. D. Wolf. Evaluation of spike-detection algorithm for a brain-machine interface application. *IEEE Trans. Biomedical Engineering*, 51(6):905–911, June 2004.

[81] I. Oppermann, M. Hämäläinen, and J. Iinatt, editors. *UWB: Theory and Applications*. John Wiley & Sons, 2004.

[82] M. R. Osborne and G. K. Smyth. A modified prony algorithm for exponential function fitting. *SIAM Journal on Scientific Computing*, 16(1):119–138, 1995.

[83] A. Papoulis. Generalized sampling expansion. *IEEE Trans. Circuits and Systems*, CAS-24(11):652–654, November 1977.

[84] A. Papoulis. *Signal Analysis*. McGraw-Hill, New York, 1977.

[85] J. Petersson and K. Holmström. Methods for parameter estimation in exponential sums. Technical Report IMa-TOM-1997-5, Mälardalen University, Västerâs, Sweden, August 1997.

[86] J. Petersson and K. Holmström. A review of the parameter estimation problem of fitting positive exponential sums to empirical data. Technical Report IMa-TOM-1997-08, Mälardalen University, Västerâs, Sweden, August 1998.

[87] V. F. Pisarenko. The retrieval of harmonics from a covariance function. *Geophysics J. Roy. Astron. Soc.* , 33:347–366, 1973.

[88] Md. A. Rahman and K. B. Yu. Total Least Squares apparoch for frequency estimation using linear prediction. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-35(10):1440–1454, October 1987.

[89] C. R. Rao. Information and accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, 37:81–91, 1945.

[90] B. Razavi, T. Aytur, C. Lam, F.-R. Yang, K.-Y. Li, R.-H. Yan, H.-C. Kang, C.-C. Hsu, and C.-C. Lee. A UWB CMOS transceiver. *IEEE Journal of Solid-State Circuits*, 40(12):2555–2562, December 2005.

[91] A. Ridolfi. *Power Spectra of Random Spikes and Related Complex Signals*. PhD thesis, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 2005.

[92] A. Ridolfi and M. Z. Win. Ultrawide bandwidth signals as shot-noise: a unifying approach. *IEEE Journal on Selected Areas in Communications*, 2005.

[93] F. Rieke, D. Warland, R. de Ruyter van Steveninck, and W. Bialek. *Spikes: Exploring the Neural Code*. MIT Press, 1999.

[94] R. Roy and T. Kailath. ESPRIT - Estimation of Signal Parameters via Rotational Invariance Techniques. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-37:984–995, July 1989.

[95] R. Sarpeshkar. Analog versus digital: Extrapolating from electronics to neurobiology. *Neural Computation*, 10:1601–1638, 1998.

[96] R. Sarpeshkar. Brain power: Borrowing from biology makes for low-power computing. *IEEE Spectrum*, 43(5):24–29, May 2006. Invited paper.

[97] R. Sarpeshkar and M. O'Halloran. Scalable hybrid computation with spikes. *Neural Computation*, 14(9):2003–2038, September 2002.

[98] R. O. Schmidt. *A Signal Subspace Approach to Multiple Emitter Location and Spectral Estimation*. PhD thesis, Stanford University, Stanford, CA, 1981.

[99] S. Schreurs and J. P. Franois. The combined use of the improved methods of Gardner and Prony for the qualitative and quantitative analysis of multicomponent decay curves. *Chemometrics and Intelligent Laboratory Systems*, (43):107–121, 1998.

[100] C. E. Shannon. Communication in the presence of noise. *Proc. Institute of Radio Engineers*, 37(1):10–21, January 1949.

[101] M. Shinagawa, Y. Akazawa, and T. Wakimoto. Jitter analysis of high-speed sampling systems. *IEEE Journal of Solid-State Circuits*, 25(1):220–224, February 1990.

[102] P. Smith. Little known characteristics of phase noise. Technical Report AN-741, Analog Devices, 2004.

[103] S. T. Smith. Statistical resolution limits and the complexified Cramér-Rao bound. *IEEE Transactions on Signal Processing*, 53(5):1597–1609, May 2005.

[104] G. Smyth. The Prony method. `http://www.statsci.org/other/prony.html`, 2005.

[105] W. M. Steedly, C. J. Ying, and R. L. Moses. Statistical analysis of TLS-based Prony techniques. *Automatica*, 30(1):115–129, January 1994. Special Issue on Statistical Signal Processing and Control.

[106] P. Stoica and R. Moses. *Spectral Analysis of Signals*. Prentice-Hall, Upper Saddle River, NJ, 2005.

[107] P. Stoica and A. Nehorai. MUSIC, maximum likelihood and Cramér-Rao bound. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(5):446–453, May 1989.

[108] P. Stoica and A. Nehorai. Statistical analysis of two nonlinear least-squares estimators of sine-wave parameters in the colored-noise case. *Circuits, Systems, and Signal Processing*, 8(1):3–15, 1989.

[109] D. N. Swingler. Frequency estimation for closely spaced sinusoids: Simple approximations to the Cramér-Rao lower bound. *IEEE Transactions on Signal Processing*, 41(1):489–495, January 1993.

[110] B. Z. Taibin. A modern approach to Prony's method (in connection with two hundred anniversary of this method). *Vestnik St. -Petersburg University. Ser. 4: Physics, chemistry*, 3(18):102–109, 1996.

[111] H. L. Van Trees. *Detection, Estimation, and Modulation Theory: Part I*. John Wiley and Sons, New York, 1968.

[112] M. J. Underhill. The adiabatic anti-jitter circuit. *IEEE Trans. Ultrasonics, Ferroelectrics and Frequency Control*, 48(3):666–674, May 2001.

[113] M. J. Underhill. The noise and suppression transfer functions of the anti-jitter circuit. In *Proc. IEEE Intl. Frequency Control Sympposium*, pages 490–498, May 2003.

[114] M. Unser. Splines: A perfect fit for signal and image processing. *IEEE Signal Proc. Magazine*, pages 22–38, November 1999.

[115] M. Unser. Sampling-50 years after Shannon. *Proceedings of the IEEE*, 88-4:569–587, April 2000.

[116] K. Uyttenhove and M. S. J. Steyaert. Speed-power-accuracy tradeoff in high-speed CMOS ADCs. *IEEE Trans. Circuits and Systems II: Analog and Digital Signal Proc.*, 49(4):280–286, April 2002.

[117] M. Vetterli, P. Marziliano, and T. Blu. Sampling signals with finite rate of innovation. *IEEE Transactions on Signal Processing*, 50(6):1417–1428, June 2002.

[118] R. H. Walden. Analog-to-digital converter survey and analysis. *IEEE Journal on Selected Areas in Communications*, 17(4):539–550, April 1999.

[119] D. D. Wentzloff, R. Blazquez, F. S. Lee, B. P. Ginsburg, J. Powell, and A. Chandrakasan. System design considerations for ultra-wideband communication. *IEEE Communications Magazine*, 43(8):114 – 121, August 2005.

[120] H. R. Wilson. *Spikes, Decisions, and Actions: Dynamical Foundations of Neuroscience*. Oxford University Press, New York, NY, 1999.

[121] M. Z. Win and R. A. Scholtz. Impulse radio: How it works. *IEEE Communication Letters*, 2(2):36–38, February 1998.

[122] M. Z. Win and R. A. Scholtz. Characterization of ultra-wide bandwidth wireless indoor communications channel: A communication theoretic view. *IEEE Journal on Selected Areas in Communications*, 20(9):1613–1627, December 2002. **Invited Paper**.

[123] M. Z. Win, R. A. Scholtz, and M. A. Barnes. Ultra-wide bandwidth signal propagation for indoor wireless communications. In *Proc. IEEE ICC*, pages 56–60, April 1997.

[124] B. H. Worsley and L. C. Lax. Selection of a numerical technique for analyzing experimental data of the decay-type with special reference to the use of tracers in biological systems. *Biochim. Biophysis. Acta*, 59:1–24, 1962.