Multichannel Sampling of Finite Rate of Innovation Signals

Hojjat Akhondi Asl

April 2011

A Thesis submitted in fulfilment of requirements for the degree
of Doctor of Philosophy of Imperial College London

Communications and Signal Processing Group
Department of Electrical and Electronic Engineering
Imperial College London

Abstract

Recently there has been a surge of interest in sampling theory in signal processing community. New efficient sampling techniques have been developed that allow sampling and perfectly reconstructing some classes of non-bandlimited signals at sub-Nyquist rates. Depending on the setup used and reconstruction method involved, these schemes go under different names such as compressed sensing (CS), compressive sampling or sampling signals with finite rate of innovation (FRI).

In this thesis we focus on the theory of sampling non-bandlimited signals with parametric structure or specifically signals with finite rate of innovation. Most of the theory on sampling FRI signals is based on a single acquisition device with one-dimensional (1-D) signals. In this thesis, we extend these results to the case of 2-D signals and multichannel acquisition systems. The essential issue in multichannel systems is that while each channel receives the input signal, it may introduce different unknown delays, gains or affine transformations which need to be estimated from the samples together with the signal itself. We pose both the calibration of the channels and the signal reconstruction stage as a parametric estimation problem and demonstrate that a simultaneous exact synchronization of the channels and reconstruction of the FRI signal is possible. Furthermore, because in practice perfect noise-free channels do not exist, we consider the case of noisy measurements and show that by considering Cramér-Rao bounds as well as numerical simulations, the multichannel systems are more resilient to noise than the single-channel ones.

Finally, we consider the problem of system identification based on the multi-

channel and finite rate of innovation sampling techniques. First, by employing our multichannel sampling setup, we propose a novel algorithm for system identification problem with known input signal, that is for the case when both the input signal and the samples are known. Then we consider the problem of blind system identification and propose a novel algorithm for simultaneously estimating the input FRI signal and also the unknown system using an iterative algorithm.

Declaration of Originality

I declare that the whole content of this thesis is the outcome and product of my own research work under the helpful guidance and support of my supervisor, Dr Pier Luigi Dragotti. Any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline. The material of this thesis has not been submitted for any degree at any other academic or professional institution.

Acknowledgment

First and foremost, I would like to deeply thank my thesis supervisor Dr. Pier Luigi Dragotti for his unmatched support and quality supervision throughout my research and writing-up. I have been so grateful and blessed to have a PhD supervisor whose motivations and encouragements in our research discussions have been second-to-none. I have known Pier Luigi since the day I started Electrical Engineering at Imperial College London. Then, he was our Communications-I lecturer and I greatly enjoyed his lectures with his energetic smiles and wisdom. For me, it has been an honour to undertake my Master's and PhD thesis under his supervision and I am thankful for his wonderful advises in our regular weekly meetings which has ensured my work come to a satisfactory and timely conclusion.

Despite the usual stresses, these research years have been quite enjoyable, thanks to my friends and colleagues in the CSP group. In particular, I want to thank Pradeep and Dushyant (Dush) for the good moments we had together during our lunch times and coffee breaks.

Last but not least and probably most important of all, I would like to thank my mum and dad for their self-less support and encouragement and I am especially thankful to my dear wife Fatemeh for her constant support, patience and understanding.

Contents

List of Figures

Abbreviations

FRI :	Finite Rate of Innovation
CS:	Compressed Sensing
ADC:	Analogue-To-Digital Converter
CRB:	Cramér-Rao Bound
C/D:	Continuous-To-Digital
1-D:	One Dimensional
2-D:	Two Dimensional
E-splines:	Exponential Splines
SVD:	Singular-Value-Decomposition
EVD:	Eigen-Value-Decomposition
SNR:	Signal-To-Noise Ratio
LMS:	Least Mean Square
FIR:	Finite Impulse Response
ACMP:	Algebraically Coupled Matrix Pencils
MEMP:	Matrix Enhancement and Matrix Pencil
ECG:	Electrocardiogram
RC Circuit:	Circuit composed of resistors and capacitors

Mathematical Notations

$*$:	Convolution operator
$\langle , \rangle$:	Inner product operator
$\mathbf{B}_{M,N}$:	Matrix $\mathbf{B}$ of size $M \times N$
$[.]^T$:	Matrix transpose operator
$[.]^H$:	Matrix Hermitian operator
$[.]^{-1}$:	Matrix inversion operator
$[.]^\dagger$:	Matrix pseudo-inverse operator
$\lceil . \rceil$:	Ceiling operator
diag(.) :	Diagonal operator
E(.) :	Expectation operator
var(.) :	Variance operator
$\|\|.\|\|$:	Euclidean norm operator
$g(x)$:	Continuous-time 1-D input signal
$\bar{g}(x)$:	Estimated version of the signal $g(x)$
$g(x, y)$:	Continuous domain 2-D input signal
$\hat{g}(j\omega)$:	Fourier transform of 1-D input signal $g(x)$
$\hat{g}(j\omega_x, j\omega_y)$:	Fourier transform of 2-D input signal $g(x, y)$
$h(x)$:	Impulse response of 1-D acquisition device
$h(x, y)$:	Impulse response of 2-D acquisition device
$\phi(x)$:	1-D sampling kernel

$\phi(x,y)$:	2-D sampling kernel
$\hat{\phi}(j\omega)$:	Fourier transform of 1-D sampling kernel $\phi(x)$
$\hat{\phi}(j\omega_x, j\omega_y)$:	Fourier transform of 2-D sampling kernel $\phi(x,y)$
T :	Sampling interval
(T_x, T_y) :	Sampling interval along the Cartesian axes
s_k :	Samples of input signal $g(x)$
$\hat{s}_k$:	Samples s_k corrupted with noise
σ^2 :	Noise variance
ϵ_k :	White additive Gaussian noise at the sample index k
$s_{j,k}$:	Samples of input signal $g(x,y)$
N :	Total number of samples s_k
(N_x, N_y) :	Total number of samples $s_{j,k}$ along the Cartesian axes
$c_{m,k}$:	Coefficients of 1-D sampling kernels
$c_{j,k}^{m,n}$:	Coefficients of 2-D sampling kernels
τ_m :	Exponential or polynomial (geometric) moments in 1-D
$\tau_{m,n}$:	Exponential or polynomial moments in 2-D
ρ :	Rate of innovation
K :	The degrees of freedom of the input signal
L :	Signal interval in 1-D
(L_x, L_y) :	Signal interval in 2-D along the Cartesian axes
$\tilde{\phi}(x)$:	Dual function of $\phi(x)$
$R_g(x,\theta)$:	Radon transform of input signal $g(x,y)$
$\psi(x)$:	The function of the unknown system within the system identification setup
s_k^{SIG} :	Samples of input signal $g(x)$ within the system identification setup
s_k^{SYS} :	Samples of input signal convolved with the unknown system
τ_m^{SIG} :	Exponential moments of the input signal

τ_m^{SYS} : Exponential moments of the convolved signal $g(x) * \psi(x)$

Chapter 1

Introduction

1.1 Background

Real world signals such as for example communication signals, audio signals and video signals are all analog signals; signals that are continuous in time and amplitude. In order to process such signals, we need to convert them in discrete form. The process in which a continuous-time signal $g(x)$ is represented by a discrete set of values or samples $g[k]$, where $k \in \mathbb{Z}$, is called "Sampling" in signal processing. The process of sampling plays a fundamental role in modern signal processing and communications where it is employed in majority of signal processing related applications such as computers, mobile phones, MP3 players, satellite receivers, radar communications, complex telescopes and many more.

A typical sampling setup often used in practice is described by a pre-filtering module together with a sampling device. The sampling device observes a filtered version of the input analog signal $g(x)$, where the filter $h(x)$ is an antialiasing filter or a sampling kernel. Depending on the application, the samples are taken at predefined intervals where if the samples are taken at equal time instances, that is, $g[k] = g(kT)$ for every T seconds, then the signal is uniformly sampled. When

the samples are not taken at equal time instances but at arbitrary points, then we call this form of sampling, non-uniform [51]. Considering that the sampling devices takes uniform samples from the observed signal, the samples $g[k]$ will be given by:

$$\begin{aligned} g[k] &= g_s(kT) \\ &= g(x) * h(x)|_{x=kT}, \end{aligned}$$

where $g_s(x)$ is the filtered signal observed by the sampling device and $*$ denotes the convolution operator. Given the samples, the fundamental questions of interest for such a process are, 1) Under what conditions signal $g(x)$ is perfectly and uniquely recovered from the set of samples $g[k]$ and 2) What are the methods of reconstruction?

The classical yet powerful answer to these key questions was given by Whittaker [85], Kotel'nikov [39, 40], and Shannon [67], in a well-known sampling theorem[1], which states that any bandlimited continuous-time signal $g(x)$ can be perfectly reconstructed from its samples, if the sampling rate is chosen to be equal or greater than twice the maximum non-zero frequency of the signal[2]. If this condition is met, then the reconstruction of the bandlimited signal from its samples can be obtained with a sinc interpolation function.

This extremely fruitful result however, has two major drawbacks. First, real world signals are never exactly bandlimited and second, an ideal sinc interpolation function is physically not implementable. Thus, different approximations need to be taken into account for such a sampling scheme to be used in practice [74, 37, 76]. These limitations have led researchers to re-examine some of the core ideas of the classical sampling theory and take it further into more advanced sampling techniques, including extensions to a larger class of signals that are not necessarily

[1]From here on, we will be referring to the theorem as Shannon's sampling theory.
[2]This is also referred to as Nyquist rate [56].

bandlimited.

Recently in [25, 17, 84], it has been shown that it is possible to sample and perfectly reconstruct some classes of non-bandlimited signals. In these schemes, the prior that the signal is sparse in a basis or in a parametric space is taken into account and perfect reconstruction is achieved based on a set of suitable measurements. Depending on the setup used and reconstruction method involved, these sampling methods go under different names such as compressed sensing (CS), compressive sampling [25, 17, 10] or sampling signals with finite rate of innovation (FRI) [84, 52, 27].

Signals with finite rate of innovation, first introduced by Vetterli et al. in [84], are parametric non-bandlimited signals that posses a finite number of parameters per unit of time or, as mentioned in [84], they posses a finite number of degrees of freedom per unit of time. Some examples of FRI signals include streams of Diracs, piecewise-polynomial [84, 27] and piecewise-sinusoidal signals [11]. The reconstruction of these FRI signals is based on the annihilating filter method (also known as Prony's method [70]). More recently, the extensions to the sampling of multi-dimensional FRI signals have been considered in [48, 47] and [69] where sampling schemes for 2-D FRI signals, such as set of 2-D Diracs and bi-level polygons have been presented. In-depth discussions of sampling 2-D FRI signals can be found in [46] and [68].

The core idea behind the sampling theory of FRI signals can be thought of an estimation problem where by recovering the degrees of freedom of the signal from its samples, the original signal can be perfectly recovered. The sampling theory introduced in [84], assumes a prior knowledge of the acquisition device employed, i.e. its sampling kernel. The sampling kernels considered in [84], are infinite-support kernels, however, it has been shown in [27] that many 1-D FRI signals with local finite rate of innovation can be sampled and perfectly reconstructed using a wide

range of sampling kernels that have finite support. Such kernels have the property of reproducing polynomials or exponentials and deliver practical implementation of the same sampling and retrieval techniques used in [84].

1.2 Motivation and Problem Statement

The process of sampling has allowed us to manipulate, store and transmit vast amount of data with increasing convenience. However, in data-intensive and/or power-limited applications such as sensor networks, the information contained is normally far less than the data observed, therefore, efficient sampling techniques is vital and necessary in such applications. Sparse sampling theories [25, 17, 84] are considered to be in the category of efficient sampling techniques as they allow sub-Nyquist sampling rates while achieving perfect retrieval of the observed signal. However, most of the papers on sparse sampling and in particular FRI sampling focus on a single-channel acquisition model and this normally requires expensive acquisition devices working at high sampling rates. On the contrary, multichannel sampling allows a reduction in the complexity of the acquisition device while keeping higher rates of conversion. For example, modern and fast Analogue-to-Digital Converters (ADC) use a parallel array of lower-rate ADCs working in a time-interleaved fashion [12]. Sensor networks and acquisition of images in a multi-camera system are other examples of multichannel systems where highly efficient sampling schemes are of interest. Therefore, given the practical importance of multichannel acquisition devices, it is natural to investigate extensions of sparse sampling theories to the multichannel scenario.

In this thesis we present a possible extension of the theory of sampling signals with finite rate of innovation to the case of multichannel acquisition systems. The critical issue in our proposed multichannel sampling setup, shown in Figure

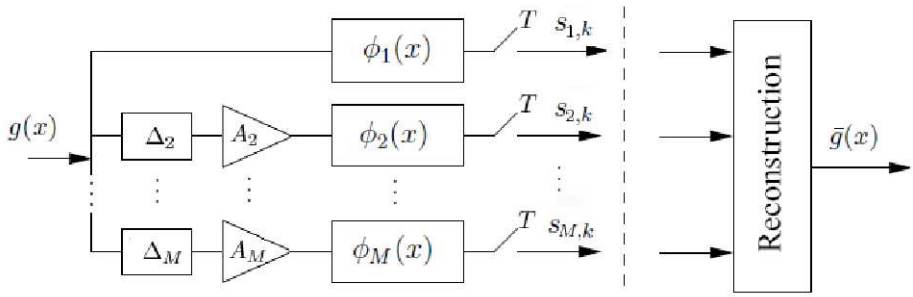


Figure 1.1: The proposed multichannel sampling setup.

1.1, is the precise synchronization of the various channels, since different devices introduce different drifts and gains in 1-D and geometric transformations such as affine transformation in 2-D, within each channel. For the signal reconstruction, these parameters need to be estimated in advance, which we refer to as the channel synchronization stage.

In this thesis we consider the multichannel sampling of FRI signals and extend the results in [27] to this new scenario. Furthermore, we also consider multichannel sampling of FRI signals under the presence of noise and assume that the samples are corrupted by white Gaussian noise. By evaluating the Cramér-Rao bounds (CRB) and taking numerical simulations into account, we asses the resilience to noise of multichannel sampling systems compared to single-channel ones.

Multichannel sampling was first proposed by Papoulis in 1977 in the context of bandlimited signals [58] and extended by Unser et al. [80, 81] for signals lying in shift-invariant subspaces. A further extension related to union of shift-invariant subspaces has been recently considered in [28]. The multichannel sampling of FRI signals has been considered in [43], [57] and [66], but in all cases the channel synchronization problem nor the channel noise was addressed.

Also in this thesis, we consider the problem of system identification based on multichannel and finite rate of innovation sampling techniques. First, by em-

ploying our multichannel sampling setup, we propose a novel algorithm for system identification problem with known input signal, that is for the case when both the input signal and the samples are known. Then we consider the problem of blind system identification and propose a novel algorithm for simultaneously estimating the unknown system as well as the input FRI signal using an iterative algorithm.

1.3 Organization of the Thesis

The organization of this thesis is as follows:

In Chapter 2, we provide a background on FRI sampling theory where we present and discuss the different elements of the sampling setup used for sampling 1-D FRI signals, including the sampling kernels and the reconstruction methods involved. In this chapter, we also consider the case of noisy measurements and discuss the common methods and tools involved for retrieving FRI signals from noisy samples.

In Chapter 3, we present a possible extension of sampling 1-D FRI signals to the case of multichannel sampling. In this chapter, by considering both the synchronization stage and the signal reconstruction stage as a parametric estimation problem, we propose our novel algorithm for multichannel sampling of FRI signals and demonstrate that it is possible to estimate simultaneously the channel parameters (i.e., delays and gains) and the signal itself from the measured samples. Then we consider the noisy scenario and assume that the noise samples are corrupted by additive white Gaussian noise. By evaluating the Cramér-Rao bounds and taking numerical simulations into account, we asses the resilience to noise of multichannel sampling systems compared to single-channel ones.

In Chapter 4, we propose our novel algorithms for system identification based on the theories of FRI sampling. The novelty of this chapter is divided into two

sections; first, by employing the multichannel sampling setup presented in Chapter 3, we propose a novel algorithm for system identification problem with known input signal. Then we consider the problem of blind system identification where by blind we mean that only the output samples are given and the input signal is not known. We will propose a novel algorithm for simultaneously estimating the input FRI signal and also the unknown system using an iterative algorithm.

In Chapter 5, we consider the problem of multichannel sampling of multi-dimensional FRI signals. In this chapter, we first introduce the multidimensional sampling framework for FRI signals which include the definition of 2-D FRI signals, sampling setup used and the properties of the sampling kernels involved. We then introduce our novel algorithms for sampling and perfectly reconstructing set of 2-D Diracs and bi-level polygons using exponential splines. Finally, we will extend the multichannel sampling setup presented in Chapter 3, to the case of multichannel sampling of 2-D FRI signals and present our novel algorithm for signal and channel estimation under simple 2-D translations and also affine transformations.

We finally conclude in Chapter 6, and present some ideas and remarks for future works.

1.4 Original Contribution

The main contribution of this thesis is the extension of finite rate of innovation sampling to the case of multichannel sampling along with channel synchronization, both in 1-D and 2-D. Furthermore, as the channel synchronization stage could be thought of a system identification problem, we also consider the problem of system identification based on FRI sampling techniques and propose a novel algorithm for identifying unknown systems by employing a FRI sampling setup.

To the best of our knowledge, Chapters 3, 4 and 5 contain original research

work where these contributions have led to the following publications:

- H. Akhondi Asl and P.L. Dragotti. **Simultaneous Estimation of Sparse Signals and Systems at sub-Nyquist Rates**. To appear in the 19th European Signal Processing Conference (EUSIPCO11), Barcelona, Spain, 2011.

- H. Akhondi Asl and P.L. Dragotti. **Multichannel Sampling of Multidimensional Parametric Signals**. To appear in the special issue of Sampling Theory in Signal and Image Processing Journal, 2011.

- H. Akhondi Asl and P.L. Dragotti and L. Baboulaz. **Multichannel Sampling of Signals with Finite Rate of Innovation**. IEEE Signal Processing Letters, vol.17, no.8, pp.762-765, August 2010.

- H. Akhondi Asl and P.L. Dragotti. **Multichannel Sampling of Translated, Rotated and Scaled Bi-level Polygons using Exponential Splines**. 8th International Conference on Sampling Theory and Applications (SAMPTA09), Marseille, France, May 2009.

- H. Akhondi Asl and P.L. Dragotti. **Single and Multichannel Sampling of Bi-level Polygons using Exponential Splines**. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP09), Taipei, Taiwan, April 2009.

- H. Akhondi Asl and P.L. Dragotti. **A Sampling Theorem For Bi-level Polygons using E-Splines**, 8th International Conference on Mathematics in Signal Processing (IMA08), Royal Agriculture College, Cirencester, UK, December 2008.

Chapter 2

Finite Rate of Innovation Sampling Theory

2.1 Introduction

In 2002, Vetterli et al. [84] introduced the notion of signals with finite rate of innovation. In their work, they showed that it is possible to sample and perfectly reconstruct some classes of signals that are neither bandlimited nor belong to a fixed subspace. Signals that can be reconstructed using this framework are called signals with finite rate of innovation (FRI) as they can be completely defined by a finite number of parameters per unit of time.

Since its introduction in 2002, finite rate of innovation sampling is finding more and more applications, for example, for compression of ECG signals [31], resolution enhancement [27, 52], distributed compression [30, 18], synchronization and channel estimation for ultra-wideband signals [46, 50, 42], ADC converters [38] and image super-resolution [7, 8, 9].

In this chapter we provide a background on finite rate of innovation sampling theory which is the core foundation of this thesis. The organization of this chapter

is as follows: In Section 2.2 we start by presenting a mathematical definition for 1-D FRI signals. Then, the different elements of the sampling setup used for sampling 1-D FRI signals, including the sampling kernels and the reconstruction methods involved will be presented and fully discussed. In Section 2.4, we will consider the case of noisy scenario and present and discuss the common methods and tools involved for retrieving FRI signals from noisy samples. We will finally conclude with a summary at the end of this chapter.

2.2 Finite Rate of Innovation Sampling Framework

2.2.1 Signals with Finite Rate of Innovation

Let us consider a 1-D signal of the form [27]:

$$g(x) = \sum_{r=0}^{N} \sum_{j \in \mathbb{Z}} \gamma_{j,r} \ \phi_r(x - x_j). \tag{2.1}$$

The degrees of freedom of the signal $g(x)$ are the shifts x_j and the coefficients $\gamma_{j,r}$, assuming that the set of functions $\phi_r(x)$ are known. If we introduce a counting function $C_g(x_a, x_b)$ which counts the number of free parameters of $g(x)$ over the interval $L = [x_a, x_b]$, then the rate of innovation ρ of the signal $g(x)$ is defined as:

$$\rho = \lim_{L \to \infty} \frac{1}{L} \, C_g(-L/2, L/2). \tag{2.2}$$

If ρ is finite, then the signal is said to have a finite rate of innovation. It is important to note that all shift-invariant signals, including bandlimited signals can be defined with the above definition. The rate of innovation of real-valued bandlimited signals is: $\rho = 2 \times f_{max}$ where f_{max} is the maximum frequency of the bandlimited signal.

2.2.2 Sampling Setup

Figure 2.1 shows the typical sampling setup employed for 1-D FRI signals where $g(x)$ represents the input 1-D FRI signal, $h(x)$ the impulse response of the acquisition device, $\phi(x)$ a re-scaled and time-reversed version of $h(x)$ (also known as the sampling kernel), $g_s(x)$ the sampled version of the input signal, s_k the samples and T the sampling interval. The box C/D (continuous-to-discrete) reads out the sample values s_k from $g_s(x)$. From the setup shown in Figure 2.1, the following expression

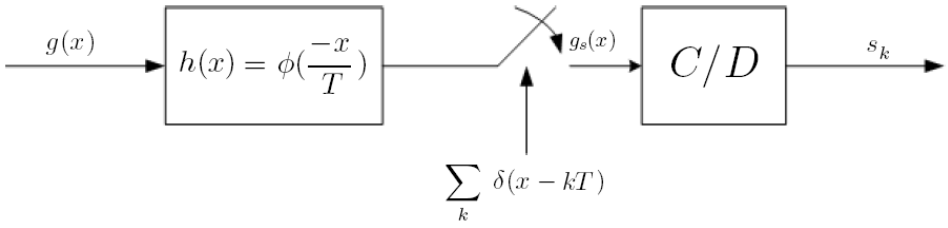


Figure 2.1: A typical sampling setup for 1-D FRI signals. Here, $g(x)$ is the continuous-time input signal, $h(x)$ the impulse response of the acquisition device, $\phi(x)$ the sampling kernel and T the sampling period. The measured samples are $s_k = \langle g(x), \phi(x/T - k)\rangle$.

for the samples s_k can be deduced:

$$\begin{aligned} s_k &= g(x) * h(x)|_{x=kT} \\ &= \int_{-\infty}^{\infty} g(x)\ \phi(\frac{x}{T} - k)\ dx \\ &= \langle g(x), \phi(\frac{x}{T} - k)\rangle. \end{aligned} \tag{2.3}$$

Having achieved the samples from the setup described above, we want to see under what conditions perfect reconstruction of the signal $g(x)$ can be obtained from the samples s_k. This includes the type of sampling kernels that can be employed and also the reconstruction techniques that are required. The answers to these questions are discussed in detail in the subsequent subsections.

2.2.3 Sampling Kernels

Sampling kernels are characterized by the physical properties of the acquisition device which are normally specified and cannot be modified. Unlike the classical sampling schemes, FRI sampling schemes provide a larger choice of kernels that allow perfect reconstruction of the input signal. The sampling kernels considered in [84] are the sinc and the Gaussian kernels. Such kernels have an infinite support and are therefore not physically realizable. Moreover, the use of such kernels make the reconstruction algorithm unstable. Dragotti et al. [27] showed that FRI signals with local finite rate of innovation can be sampled and reconstructed using a wide range of sampling kernels that have finite support. Such kernels have the property of reproducing polynomials or exponentials and deliver practical implementation of the same sampling and retrieval techniques used in [84] for 1-D FRI signals.

In this thesis, we will focus on polynomial and exponential reproducing kernels and in particular exponential splines (E-splines) [79, 75], splines that can reproduce real or complex exponentials. E-splines are compact support splines that are practically implementable (using RC circuits for example [27]) and our simulation results show that they tend to be more stable than other kernels.

Polynomial Reproducing Kernels

Any kernel $\phi(x)$ that together with its shifted versions can reproduce polynomials of maximum degree M is called a polynomial reproducing kernel. That is any kernel satisfying the following property:

$$\sum_{k\in\mathbb{Z}} c_{m,k}\phi(x-k) = x^m, \tag{2.4}$$

for a proper choice of coefficients $c_{m,k}$ with $m = 0, 1, \ldots, M$. Here, the subscript k represents the shifts index and the superscript m represents the polynomial degree.

The choice of M depends on the local rate of innovation of the signal $g(x)$ and will be discussed later on. Furthermore, the coefficients $c_{m,k}$ can be calculated as follows:

$$c_{m,k} = \int_{-\infty}^{\infty} x^m \tilde{\phi}(x-k)\, dx, \tag{2.5}$$

where $\tilde{\phi}(x)$ is chosen to form with $\phi(x)$ a quasi-biorthonormal set [15]. This includes the particular case where $\tilde{\phi}(x)$ is the dual of $\phi(x)$, that is, $\langle \tilde{\phi}(x-j), \phi(x-k) \rangle = \delta_{j,k}$. We should mention that, given the specified kernel and the required polynomial degree, the coefficients can also be calculated numerically.

Polynomial reproducing kernels include any function satisfying the so-called Strang-Fix conditions [71] which states that, the kernel $\phi(x)$ satisfies Equation (2.4) if and only if its Fourier transform $\hat{\phi}(j\omega)$ satisfies:

$$\begin{cases} \hat{\phi}(0) \neq 0 & \text{and,} \\ \hat{\phi}^{(m)}(2k\pi) = 0 & \text{for } k \neq 0 \text{ and } m = 0, 1, \ldots, M, \end{cases} \tag{2.6}$$

where the superscript (m) stands for the m-th order derivative of $\hat{\phi}(j\omega)$. B-splines [77, 78, 73] are one of the most well-known examples of kernels satisfying the Strang-Fix conditions. A function $\beta(x)$ with Fourier transform:

$$\hat{\beta}^M(j\omega) = \prod_{m=0}^{M} \frac{1 - e^{-j\omega}}{j\omega}, \tag{2.7}$$

is called a B-spline of order $M+1$. The resulting spline has compact support and can reproduce any polynomial in the subspace spanned by $\{1, x, x^2, \ldots, x^M\}$. In the time-domain, the expression of a B-spline of order 1 is given as follows:

$$\beta^0(x) = \begin{cases} 1 & 0 \leq x < 1 \\ 0 & \text{otherwise,} \end{cases} \tag{2.8}$$

and the higher order B-spline functions are obtained by successive convolutions of the first order B-spline, defined above. B-splines are biorthogonal functions and their dual basis is defined in [65]. Figures 2.2(a)(b)(c) and (d) show B-splines of order 1, 2, 3 and 4 respectively.

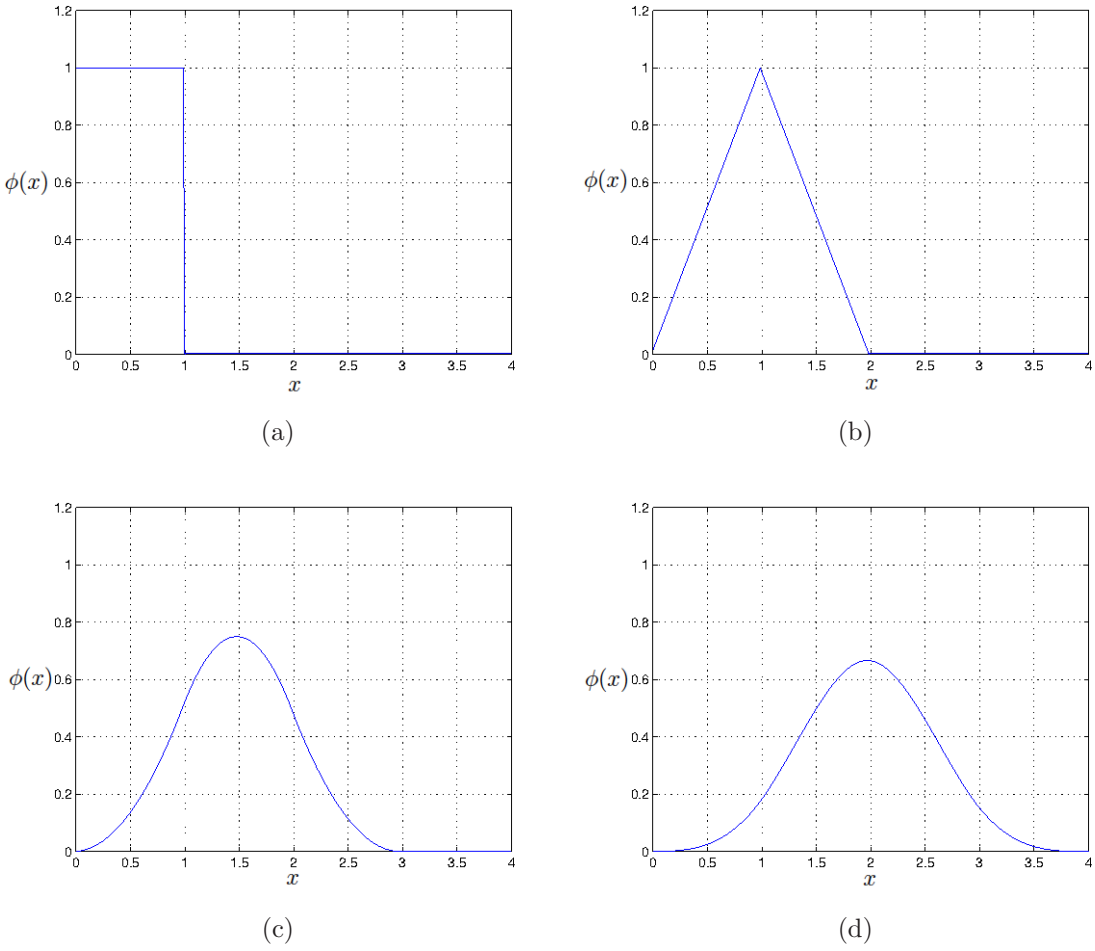


Figure 2.2: B-splines of orders $1, 2, 3$ and 4. (a) B-spline of order 1 (b) B-spline of order 2 (c) B-spline of order 3 (d) B-spline of order 4.

Strang-Fix conditions are also used extensively in wavelet theory and Daubechies scaling functions satisfy such conditions [20, 83]. More precisely, a wavelet with $M+1$ vanishing moments is generated by a scaling function that can reproduce polynomials of degree M. Daubechies scaling functions are orthogonal functions, and their dual basis is: $\tilde{\phi}(x) = \phi(x)$.

Exponential Reproducing Kernels

Any kernel $\phi(x)$ that together with its shifted versions can reproduce real or complex exponentials of the form $e^{\alpha_m x}$ with $m = 0, 1, \ldots, M$ is called an exponential reproducing kernel. That is any kernel satisfying the following property:

$$\sum_{k\in\mathbb{Z}} c_{m,k}\phi(x-k) = e^{\alpha_m x}, \qquad \text{with } \alpha_m \in \mathbb{C}, \tag{2.9}$$

for a proper choice of coefficients $c_{m,k} \in \mathbb{C}$. The coefficients $c_{m,k}$ in the above equation are given by the following expression:

$$c_{m,k} = \int_{-\infty}^{\infty} e^{\alpha_m x}\tilde{\phi}(x-k)dx, \tag{2.10}$$

where, as in the polynomial reproducing kernel's case, $\tilde{\phi}(x)$ is chosen to form with $\phi(x)$ a quasi-biorthonormal set. The choice of the exponents in Equation (2.9) is restricted to $\alpha_m = \alpha_0 + m\lambda$ with $\alpha_0, \lambda \in \mathbb{C}$ and $m = 0, 1, ..., M$. This is done to allow the use of the annihilating filter method at the reconstruction stage. This fact will be more evident when the reconstruction methods are described later on.

The theory of exponential reproducing kernels is quite recent and is based on the notion of E-splines [79]. A function $\beta_{\vec{\alpha}}(x)$ with Fourier transform:

$$\hat{\beta}_{\vec{\alpha}}(j\omega) = \prod_{m=0}^{M} \frac{1 - e^{\alpha_m - j\omega}}{j\omega - \alpha_m}, \tag{2.11}$$

is called an E-spline of order $M+1$ where $\vec{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_M)$. The produced spline has compact support and can reproduce any exponential in the subspace spanned by $(e^{\alpha_0 x}, e^{\alpha_1 x}, \ldots, e^{\alpha_M x})$. Moreover, the values of α_0 and λ can be chosen arbitrarily, but too small or too large values could lead to unstable results for the reproduction of exponentials. In the time-domain, the expression of an E-spline of order one is

given by:

$$\beta_{\alpha_0}(x) = \begin{cases} e^{\alpha_0 x} & 0 \leq x < 1 \\ 0 & \text{otherwise,} \end{cases} \tag{2.12}$$

where the higher order E-splines are obtained by successive convolutions of lower order ones with their specific α_m parameters. Moreover, since the exponential reproduction formula is preserved through convolution [79], any composite function of the form $\phi(x) * \beta_{\vec{\alpha}}(x)$ is also able to reproduce exponentials. Figure 2.3 shows E-splines of order 1, 2, 3 and 4 with the following α_m values: $\alpha_{0:3} = [-0.2 + 0.3j, -0.1 + 0.1j, 0.5, 0.2 - 0.1j]$. In the figure, the blue and the red lines show the real and imaginary parts of the E-splines respectively.

2.3 Reconstruction Algorithms

Having gone through the sampling stage, we will now discuss the reconstruction process of 1-D FRI signals. For most 1-D FRI signals, the problem of reconstructing the signal $g(x)$ is reduced to the problem of reconstructing a set of 1-D Diracs. For example, in [84] it is shown that the problem of reconstructing non-uniform splines can be reduced to the problem of reconstructing stream of 1-D Diracs. Also, in the case of piecewise polynomials, the reconstruction procedure is reduced to reconstructing a sum of derivatives of Diracs. For this reason we concentrate on stream of Diracs only.

Let us assume our input signal $g(x)$ consists of K Diracs, with amplitudes a_k located at distinct instants $x_k \in [0, L[$, that is:

$$g(x) = \sum_{k=1}^{K} a_k \delta(x - x_k). \tag{2.13}$$

The signal is sampled using the sampling setup shown in Figure 2.1 with a sampling

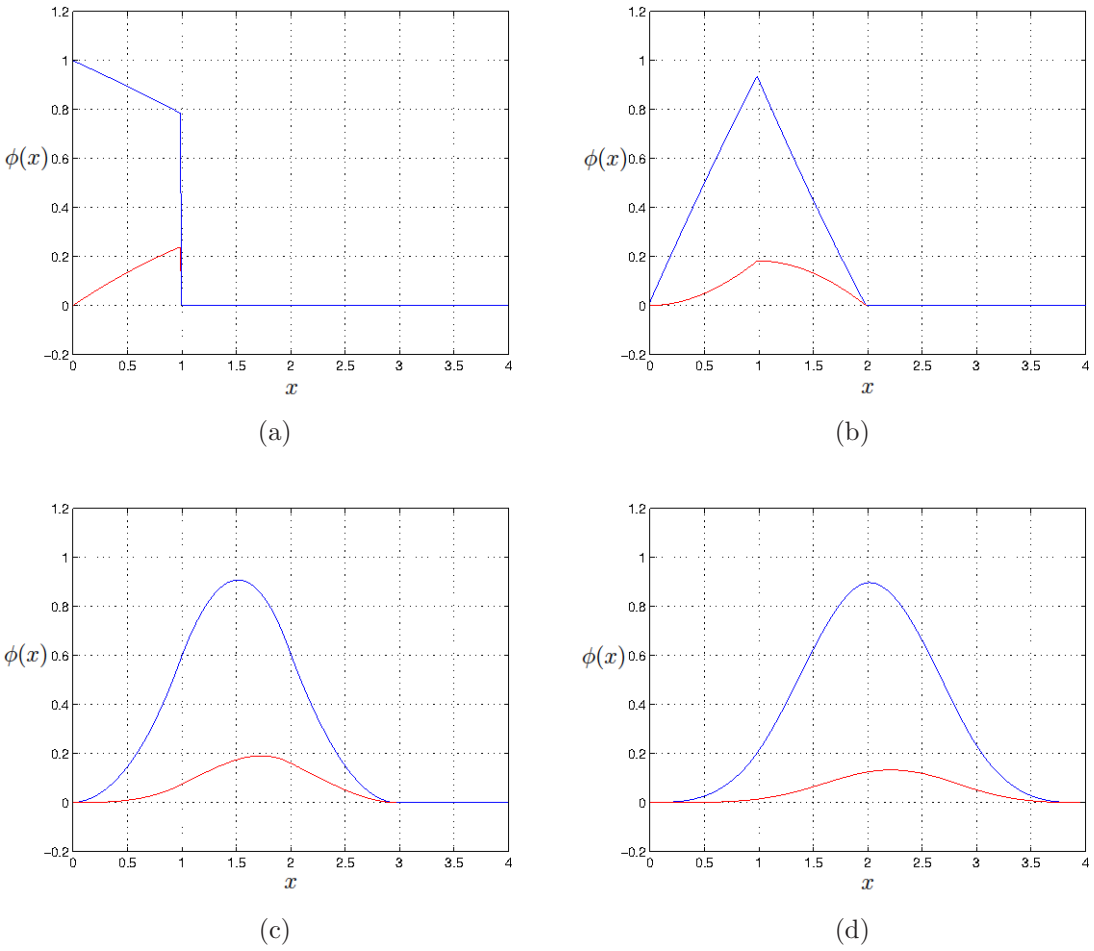


Figure 2.3: E-splines of orders $1, 2, 3$ and 4. (a) E-spline of order 1 with $\alpha_0 = -0.2+0.3j$ (b) E-spline of order 2 with $\alpha_{0:1} = [-0.2+0.3j, -0.1+0.1j]$ (c) E-spline of order 3 with $\alpha_{0:2} = [-0.2+0.3j, -0.1+0.1j, 0.5]$ (d) E-spline of order 4 with $\alpha_{0:3} = [-0.2+0.3j, -0.1+0.1j, 0.5, 0.2-0.1j]$. The blue and red lines show the real and imaginary parts of the E-splines respectively.

kernel $\phi(x)$. Furthermore, we assume the sampling period is $T = L/N$ where N is the number of samples. Consequently, the samples s_k are given by: $s_k = \langle g(x), \phi(x/T - k)\rangle$. In [84] and [27], it is shown that such a stream of Diracs can be perfectly reconstructed using sinc, Gaussian, polynomial and exponential reproducing kernels. As we mainly focus on polynomial and exponential reproducing kernels in this thesis,

let us consider the following weighted sum of the samples:

$$\tau_m = \sum_k c_{m,k} s_k. \tag{2.14}$$

Substituting Equation (5.4) into the above equation, yields (for simplicity we have assumed $T = 1$):

$$\tau_m = \langle g(x), \sum_k c_{m,k} \phi(x - k) \rangle, \tag{2.15}$$

where we have used the linearity of the inner product to move the sum operator inside the inner product. The second term in the inner product can be replaced by one of the equations defined in (2.4) or (2.9) depending on the sampling kernel used. If polynomial reproducing kernels are employed, then from Equation (2.15), the polynomial moments of the signal are obtained:

$$\tau_m = \int_{-\infty}^{\infty} g(x)\ x^m\ dx. \tag{2.16}$$

Given that our input signal is a set of K 1-D Diracs, measurements τ_m will lead to a power-sum series form, that is:

$$\tau_m = \int_{-\infty}^{\infty} g(x)\ x^m\ dx \tag{2.17}$$

$$= \int_{-\infty}^{\infty} \sum_{k=1}^{K} a_k\ \delta(x - x_k)\ x^m\ dx \tag{2.18}$$

$$= \sum_{k=1}^{K} a_k\ x_k^m, \qquad m = 0, 1, \ldots, M. \tag{2.19}$$

Likewise, if exponential reproducing kernels are employed as the sampling kernel,

then the exponential moments of the signal are obtained, that is:

$$\tau_m = \int_{-\infty}^{\infty} g(x)\ e^{\alpha_m x}\ dx \tag{2.20}$$

$$= \int_{-\infty}^{\infty} \sum_{k=1}^{K} a_k\ \delta(x - x_k)\ e^{\alpha_m x}\ dx \tag{2.21}$$

$$= \sum_{k=1}^{K} a_k\ e^{\alpha_m x_k} \tag{2.22}$$

$$= \sum_{k=1}^{K} \hat{a_k}\ u_k^m, \qquad m = 0, 1, \ldots, M, \tag{2.23}$$

where $\hat{a_k} = a_k e^{\alpha_0 x_k}$ and $u_k = e^{\lambda x_k}$. In the case of purely imaginary E-splines, that is with $\alpha_m = jm\lambda$, the Fourier transform of the signal $g(x)$ at α_m are obtained from the exponential moments, that is:

$$\tau_m = \hat{g}(\alpha_m),$$

where $\hat{g}(j\omega)$ represents the Fourier transform of the signal $g(x)$.

For both cases explained above, the given expressions for polynomial and exponential reproducing kernels leads to a power-sum series structure in the form:

$$\tau_m = \sum_{k=1}^{K} a_k\ u_k^m, \quad m = 0, 1, \ldots, M. \tag{2.24}$$

In 1795 Prony showed that the unknown parameters a_k and u_k can be exactly recovered, provided that the number of measurements τ_m is at least $2K$. Prony's method, also referred as the annihilating filter method in [84, 27], is a widely used technique in spectral estimation [70, 55] and error-correction coding [13]. In the following subsection, the annihilating filter method will be discussed and explained.

Annihilating Filter Method

The field of spectral estimation, which is related to estimating frequency contents of a signal, has a vast range of applications in signal processing. Although there are many spectral estimation methods available, most of them suffer from resolution inaccuracy and large amount of computational burden. Subspace spectral estimation methods are generally more efficient than the classical methods [70, 41, 60]. The well-known methods such as ESPRIT [61] (Estimation of Signal Parameters via Rotational Invariance Techniques) and MUSIC [64] (MULtiple Signal Classification) are examples of subspace spectral estimation methods for one dimensional signals where matrix decomposition techniques are used to estimate unknown parameters such as amplitude, phase and frequency (see [60] for a detailed tutorial). In this section we will explain the annihilating filter method, which is the most popular estimation method in the FRI sampling community.

Let us define a filter h_m with $m = 0, 1, \ldots, K$, such that the locations u_k are the roots of the filter. The z-transform of such a filter is:

$$H(z) = \sum_{m=0}^{K} h_m z^{-m} = \prod_{k=1}^{K} (1 - u_k z^{-1}). \tag{2.25}$$

The observed signal τ_m convolved with the filter defined above, results in:

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

The under-braced term in the set of equations above equals to zero, as $H(u_k) = 0$,

thus:

$$h_m * \tau_m = 0. \tag{2.26}$$

The filter $H(z)$ is called the annihilating filter as it annihilates the observed signal τ_m. The zeros of such a filter uniquely define the distinct locations u_k. To retrieve the locations, the convolution equation is written in the following matrix form:

$$\boldsymbol{\tau} \cdot H = \begin{bmatrix} \tau_K & \tau_{K-1} & \cdots & \tau_0 \\ \tau_{K+1} & \tau_K & \cdots & \tau_1 \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{2K} & \tau_{2K-1} & \cdots & \tau_K \\ \vdots & \vdots & \ddots & \vdots \\ \tau_M & \tau_{M-1} & \cdots & \tau_{M-K} \end{bmatrix} \times \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_K \end{bmatrix} = 0, \tag{2.27}$$

where $\boldsymbol{\tau}$ is a Toeplitz matrix of size $(M-K+1) \times (K+1)$, H is a column vector of length $K+1$ and $M \geq 2K-1$ as at least $2K$ consecutive values of τ_m are required in order to solve the matrix equation shown above. Notice that the above expression indicates that the matrix is rank deficient. By assuming $h_0 = 1$, it can be written as a system of Yule-Walker equations:

$$\begin{bmatrix} \tau_{K-1} & \tau_{K-2} & \cdots & \tau_0 \\ \tau_K & \tau_{K-1} & \cdots & \tau_1 \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{M-1} & \tau_{M-2} & \cdots & \tau_{M-K} \end{bmatrix} \times \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_K \end{bmatrix} = - \begin{bmatrix} \tau_K \\ \tau_{K+1} \\ \vdots \\ \tau_M \end{bmatrix}, \tag{2.28}$$

where by taking the inverse of the left-hand-side matrix we can solve for the coefficients h_m. Given the filter coefficients, the locations of the Diracs are found by taking the roots of the filter. The system of equations above gives a unique solution for u_k since the filter coefficients h_m are unique for a given signal. After finding the locations u_k, we are able to find the weights a_k from the power-sum series expression

given in Equation (2.24). By expanding the equation and writing it in the matrix form, we obtain:

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ u_1 & u_2 & \cdots & u_K \\ u_1^2 & u_2^2 & \cdots & u_K^2 \\ \vdots & \vdots & \ddots & \vdots \\ u_1^{K-1} & u_2^{K-1} & \cdots & u_K^{K-1} \end{bmatrix} \times \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_K \end{bmatrix} = \begin{bmatrix} \tau_0 \\ \tau_1 \\ \vdots \\ \tau_{K-1} \end{bmatrix}. \tag{2.29}$$

The above system of equations is also known as a Vandermonde system and leads to a unique solution for the amplitudes a_k since the u_k are distinct. The modified versions of the annihilating filter method can also be used for other 1-D FRI signals such as differentiated Diracs [27], piecewise-polynomial signals [27] and piecewise-sinusoidal signals [11].

Having discussed the annihilating filter method, we conclude that perfect reconstruction of a stream of K Diracs is possible with any kernel able to reproduce polynomials or exponentials. If $g(x)$ has more than K Diracs or possibly an infinite number of Diracs, we cannot use the above discussed method directly. However, since the kernels considered have compact support, the above scheme can be applied sequentially. More precisely, it was shown in [27] that if there are no more than K Diracs in an interval of size $L = 2KPT$, where P denotes the support of the kernel used, then we are guaranteed that two groups of K consecutive Diracs are sufficiently distant and that they are separated by some zero samples. By locating these zeros, one can separate the two groups and apply the above reconstruction method on each group independently. If $g(x)$ has more than K Diracs in an interval of size $2KPT$ then the only way to sample it is by increasing the sampling rate. In Chapter 3, we show that this can be avoided by using a multichannel acquisition system. We will now discuss the reconstruction of FRI signals in the presence of noise.

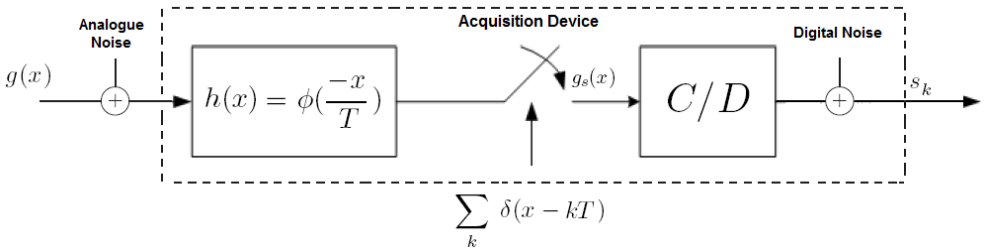


Figure 2.4: FRI sampling setup with possible sources of noise in the entire sampling process.

2.4 FRI Sampling in the Presence of Noise

The annihilating filter method, discussed in the previous section is ideal for the noiseless case. However, when noise is present the described method can become unstable. In this section our aim is to introduce more robust methods that yield more accurate reconstruction in the presence of noise [49, 14, 84]. The sources of noise could be both in the analog domain, for example in transmission, or in the digital domain. Some examples of the sources of digital noise which are also known as the sampling noise are the quantization process and the noise introduced by the acquisition devices. Figure 2.4 shows the block diagram of the possible sources of the noise.

In this thesis, we only consider the sampling noise, where noisy samples measured at the output are:

$$\hat{s}_k = s_k + \epsilon_k, \tag{2.30}$$

where ϵ_k is assumed to be additive white Gaussian (AWGN), independent of the sample s_k. Given the noisy samples, our goal is to recover the innovation parameters of the input signal $g(x)$. To achieve robustness to noise, however, there is no option but to increase the sampling rate and obtain more samples. We will now discuss some of the popular methods used for recovering FRI signals from noisy samples.

2.4.1 Total Least-Squares Method

Given the noisy samples, the annihilating filter equation $\boldsymbol{\tau} \cdot H = 0$ will not be satisfied exactly. However, as discussed in [14, 84, 59], a total least-squares approach can be applied to reduce the effect of noise, by minimizing the Euclidean norm $||\tau \cdot H||^2$ under the constraint that $||H||^2 = 1$. This can be done by evaluating the singular-value-decomposition (SVD) [55, 23] of the Toeplitz matrix $\boldsymbol{\tau} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$, and choosing the annihilating filter H to be the column vector of matrix $\mathbf{V}$ that corresponds to the smallest singular value. As the matrix $\boldsymbol{\Sigma}$ is a diagonal matrix, containing the singular value elements in a decreasing order, the last column vector of the matrix $\mathbf{V}$ will correspond to the annihilating filter [14].

2.4.2 Matrix Pencil Method

Another popular subspace spectral estimation method is the matrix pencil method [34, 35, 62] which makes use of Hankel matrices, singular-value-decomposition and Eigen-value-decomposition (EVD) [55]. Matrix pencil method, tends to perform well under noisy conditions and in our simulations it performs slightly better the total least-squares method.

The estimation algorithm is as follows: As before, let us assume that we have access to the measurements $\tau_m = \sum_{k=1}^{K} a_k \, u_k^m$, with unknown locations u_k and amplitudes a_k. Then we arrange the measurements τ_m into a Hankel matrix $\mathbf{H}$ of dimension $M_1 \times M_2$ as follows:

$$\mathbf{H}_{M_1 \times M_2} = \begin{bmatrix} \tau_0 & \tau_1 & \dots & \tau_{M_1-1} \\ \tau_1 & \tau_2 & \dots & \tau_{M_1} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{M_2-1} & \tau_{M_2-2} & \dots & \tau_{M_1+M_2-2} \end{bmatrix}, \tag{2.31}$$

where $M_1 \geq K+1$, $M_2 \geq K$. As for the annihilating filter method, the number of measurements of τ_m should be at least $2K$ in order to fully recover the unknown parameters. The matrix $\mathbf{H}$, with the arrangement shown, is a product of three matrices: $\mathbf{S}$, $\mathbf{A}$ and $\mathbf{T}$ where $\mathbf{S}$ and $\mathbf{T}$ are Vandermonde matrices of the locations u_k and $\mathbf{A}$ is a diagonal matrix containing the amplitudes a_k. More precisely, the Hankel matrix $\mathbf{H}$ can be written as:

$$\mathbf{H} = \mathbf{SAT}^T, \tag{2.32}$$

with the following decomposition:

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ u_1 & u_2 & \dots & u_K \\ \vdots & \vdots & \ddots & \vdots \\ u_1^{M_1-1} & u_2^{M_1-1} & \dots & u_K^{M_1-1} \end{bmatrix} \begin{bmatrix} a_1 & 0 & \dots & 0 \\ 0 & a_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_K \end{bmatrix} \begin{bmatrix} 1 & 1 & \dots & 1 \\ u_1 & u_2 & \dots & u_K \\ \vdots & \vdots & \ddots & \vdots \\ u_1^{M_2-1} & u_2^{M_2-1} & \dots & u_K^{M_2-1} \end{bmatrix}^T. \tag{2.33}$$

The singular-value-decomposition decomposes a matrix into a product of three matrices: $\mathbf{U}$, $\mathbf{\Sigma}$ and $\mathbf{V}$ where $\mathbf{U}^H\mathbf{U} = \mathbf{I}$, $\mathbf{V}^H\mathbf{V} = \mathbf{I}$ and $\mathbf{\Sigma}$ is a diagonal matrix containing the singular values. Here, the superscript H stands for the Hermitian transpose. If we take the SVD of the described Hankel matrix $\mathbf{H}$, we obtain:

$$\mathbf{H}_{M_1 \times M_2} = \mathbf{U}_{M_1 \times M_1} \mathbf{\Sigma}_{M_1 \times M_2} \mathbf{V}^H_{M_2 \times M_2}. \tag{2.34}$$

In order to obtain the signal subspace, only the product of the first K columns of the matrices $\mathbf{U}$ and $\mathbf{V}$, and also the $K \times K$ upper left matrix of $\mathbf{\Sigma}$ are taken into

account, resulting in a truncated version of the original matrix $\mathbf{H}$, that is:

$$\begin{aligned}\mathbf{H}_K &= \mathbf{U}_K \mathbf{\Sigma}_K \mathbf{V}_K^H \\ &= [U_1\ U_2\ \dots\ U_K] \begin{bmatrix} \Sigma_1 & 0 & \dots & 0 \\ 0 & \Sigma_2 & \vdots & 0 \\ \vdots & \dots & \ddots & 0 \\ 0 & \dots & 0 & \Sigma_K \end{bmatrix} [V_1\ V_2\ \dots\ V_K]^H.\end{aligned}$$

Since the matrices $\mathbf{S}$ and $\mathbf{U}_K$ span the same column space, the following relationship holds true:

$$\mathbf{U}_K = \mathbf{SQ}, \tag{2.35}$$

where $\mathbf{Q}$ is a non-singular matrix of dimension $K \times K$. We mentioned that the matrices $\mathbf{S}$ and $\mathbf{T}$ have a Vandermonde structure. Vandermonde matrices satisfy the "shift-invariant" subspace property which states that if $\overline{\mathbf{S}}$ and $\underline{\mathbf{S}}$ denote the matrix $\mathbf{S}$ after omission of the first and the last row respectively, then the following relationship is valid:

$$\overline{\mathbf{S}} = \underline{\mathbf{S}}\mathbf{\Phi}, \tag{2.36}$$

where $\mathbf{\Phi} = diag\{u_1, u_2, \dots, u_K\}$. Knowing that $\mathbf{U}_K = \mathbf{SQ}$, we clearly have (true for any matrix multiplication):

$$\underline{\mathbf{U}_K} = \underline{\mathbf{S}}\mathbf{Q} \tag{2.37}$$

$$\overline{\mathbf{U}_K} = \overline{\mathbf{S}}\mathbf{Q} \tag{2.38}$$

$$= \underline{\mathbf{S}}\mathbf{\Phi}\mathbf{Q}. \tag{2.39}$$

Now let us consider the matrix pencil $(\overline{\mathbf{U}_K}, \underline{\mathbf{U}_K})$ as follows:

$$\overline{\mathbf{U}_K} - \lambda \underline{\mathbf{U}_K} = \underline{\mathbf{S}}(\mathbf{\Phi} - \lambda \mathbf{I})\mathbf{Q}, \tag{2.40}$$

where λ is called the rank reducing number. We can solve for u_k by finding the Eigen-values of the matrix pencil. The problem of finding the Eigen-values of a matrix pencil is called the "Generalized Eigen-value problem". Therefore, to obtain the locations u_k we construct the following matrix equation:

$$\underline{\mathbf{U}_K}^{-1}.\overline{\mathbf{U}_K} = \mathbf{Q}^{-1}\mathbf{\Phi}\mathbf{Q}, \tag{2.41}$$

where by taking the Eigen-value-decomposition, we obtain the matrix $\mathbf{\Phi}$ which is a diagonal matrix containing all the locations u_k:

$$eig(\underline{\mathbf{U}_K}^{-1}\overline{\mathbf{U}_K}) = eig(\mathbf{Q}^{-1}\mathbf{\Phi}\mathbf{Q}) = \mathbf{\Phi}. \tag{2.42}$$

Moreover, since we have found the exact values of the locations u_k, we can now construct the matrices $\mathbf{S}$ and $\mathbf{T}$ to obtain the amplitudes a_k, using the following equation:

$$\mathbf{A} = (\mathbf{S}^{\dagger})\mathbf{H}(\mathbf{T}^{T})^{\dagger}, \tag{2.43}$$

where the dagger † stands for pseudo-inverse of the matrix.

2.4.3 Cadzow's Algorithm

The total least-squares method and the matrix pencil method are reliable only for moderate values of noise. In [14], an iterative denoising algorithm, known as Cadzow's algorithm [16, 72] is introduced where, when applied before one of the methods discussed above, yields more robust and reliable results.

In the noiseless case, the Toeplitz matrix $\boldsymbol{\tau}$, which is constructed from the measurements τ_m, has rank K, equal to the number of Diracs in the input signal. This is due to the fact that the annihilating filter H in $\boldsymbol{\tau} \cdot H = 0$, has $K + 1$ coefficients. When the signal is corrupted by noise, this rank deficiency property

is lost. However, by taking the SVD, we can assume that the K largest singular values in $\boldsymbol{\tau} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H$ correspond to the actual signal and the rest correspond to the noise. To restore the rank deficiency property of the matrix, we set the singular values of the noise to zero. This will in-turn alter the Toeplitz structure of the original matrix, but by taking the average of the diagonals of the matrix, Toeplitz structure can be restored back. By iterating this procedure, the matrix $\boldsymbol{\tau}$ converges to a well-approximated Toeplitz matrix of correct rank. The resulting matrix, is a properly denoised version of the original matrix $\boldsymbol{\tau}$ and the matrix pencil method or the total-least squares method can be applied to the denoised matrix. Given a noise contaminated discrete-time measurements of a signal, Cadzow's algorithm exploits the signal attributes (i.e. rank deficiency property and Toeplitz structure) from its matrix representation.

The Cadzow's algorithm combined with the matrix pencil method will be utilized in the next chapter, when we consider the multichannel sampling of FRI signal in the presence of noise.

2.5 Summary

In this chapter we presented a background on the 1-D sampling framework of FRI signals. We described the different elements of the sampling setup and showed how a set of 1-D Diracs can be sampled and perfectly reconstructed using both polynomial and exponential reproducing kernels. Then we considered the case of noisy measurements and presented some denoising techniques for the reconstruction of the innovation parameters from the noisy samples.

Chapter 3

Multichannel Sampling of Finite Rate of Innovation Signals

3.1 Introduction

Multichannel sampling was first proposed by Papoulis in the context of bandlimited signals [58] in 1977. In his work, Papoulis introduced a powerful extension of Shannon's sampling theory, showing that a bandlimited signal $g(x)$ could be reconstructed exactly from the samples of M linear shift-invariant systems, sampled at $1/M$th of the Nyquist rate.

In this chapter we present a possible extension of the theory of sampling signals with finite rate of innovation to the case of multichannel acquisition systems. The critical issue in our proposed multichannel sampling setup is the precise synchronization of the various channels, since different devices introduce different drifts and different gains within each channel. This could be due, for example, to imperfections of electronic circuits. For the signal reconstruction, these parameters need to be estimated in advance, which we refer to as the channel synchronization stage. In this chapter we consider the multichannel sampling of FRI signals and extend the

results in [27] to this new scenario. The material of this chapter has been in part published at [5].

The organization of this chapter is as follows: In the next section, we present an extension of 1-D FRI sampling framework, discussed in the previous chapter, to the case of multichannel sampling. In Section 3.3, by considering both the synchronization stage and the signal reconstruction stage as a parametric estimation problem, we propose our novel algorithm for multichannel sampling of FRI signals and demonstrate that it is possible to simultaneously estimate the channel parameters (i.e., delays and gains) and the signal itself from the measured samples. This is achieved by operating at a sampling rate proportional to $1/TM$, where M is the number of channels involved. In Section 3.5 we consider the noisy scenario and assume that the samples are corrupted by additive white Gaussian noise. By evaluating the Cramér-Rao bounds and taking numerical simulations into account, we asses the resilience to noise of multichannel sampling systems compared to single-channel ones. We finally summarize this chapter in Section 3.6.

3.2 Multichannel Sampling Framework

3.2.1 Sampling Setup

Figure 3.1 shows our proposed multichannel setup for 1-D FRI signals. As shown in the figure, the bank of acquisition devices or sampling kernels with functions $\phi_1(x), \phi_2(x), \ldots, \phi_M(x)$, receive drifted and scaled versions of the input FRI signal $g(x)$ denoted with Δ_i and A_i for $i = 2, 3, \ldots, M$. Given the setup model, the problem can be divided into two stages, the channel synchronization stage and the signal reconstruction stage. The samples $s_{i,k}$ measured at the output of the multichannel sampling setup are utilized jointly for the channel synchronization and the signal re-

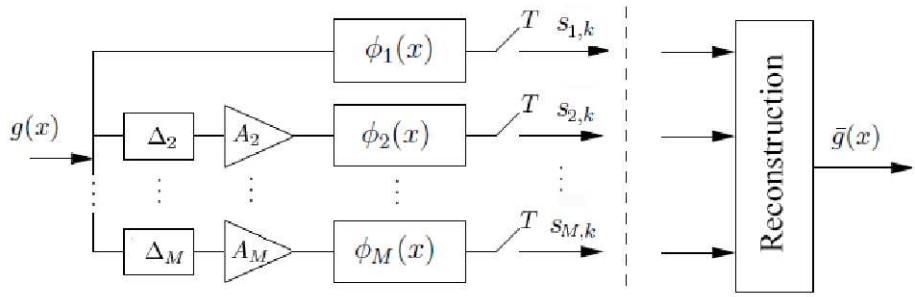


Figure 3.1: The proposed multichannel sampling setup for 1-D FRI signals. Here, the continuous-time signal $g(x)$ is received by multiple channels with multiple acquisition devices. The samples $s_{i,k}$ from each channel are utilized jointly for the reconstruction process.

construction process. Our goal is to see under what conditions the multiple channels can be synchronized, that is under what conditions the estimation of the unknown parameters Δ_i and A_i is possible. Moreover, given the synchronized channels, we also want to see under what conditions perfect reconstruction of the input FRI signal $g(x)$ can be achieved. In the following subsections all these questions are addressed and a novel algorithm will be presented.

3.2.2 Sampling Kernels

For the multichannel setup shown in Figure 3.1, we will assume that the acquisition devices are all E-spline sampling kernels. The reason for the use of such compact support kernel is that, unlike polynomial reproducing kernels, dynamic multichannel sampling is possible with E-splines. By dynamic sampling, we mean that the overall number of samples required from the multichannel setup, can be arbitrarily distributed within different channels, as long as the conditions necessary for channel synchronization are met. This is due to the fact that, arbitrary exponentials can be reproduced with E-splines and this rises from the arbitrary choice of the parameters α_0 and λ in $\alpha_m = \alpha_0 + m\lambda$. Moreover, the exponents of the E-spline can be chosen so

that a number of exponents are in common between the different sampling kernels and this will allow the synchronization of the channels which will be explained in detail in the next section.

3.3 Multichannel Sampling of Finite Rate of Innovation Signals

3.3.1 Channel Synchronization

Without loss of generality, let us assume that the input signal $g(x)$ is a stream of K Diracs at distinct instants x_k with amplitudes a_k. Let us also assume for simplicity that $M = 2$, that is, our multichannel system is restricted to two channels only. We specify each channel sampling kernel to be of the form:

$$\hat{\phi}_1(j\omega) = \prod_{m=0}^{P} \left(\frac{1 - e^{\alpha_m - j\omega}}{j\omega - \alpha_m} \right) \tag{3.1}$$

$$\hat{\phi}_2(j\omega) = \prod_{m=P-1}^{Q} \left(\frac{1 - e^{\alpha_m - j\omega}}{j\omega - \alpha_m} \right), \tag{3.2}$$

where Q depends on the structure of the input signal. For simplicity we will assume that $Q = 2P - 1$, so that both sampling kernels will be of equal order $P + 1$. As can be seen from the equations given above, both kernels can reproduce the exponentials $e^{\alpha_{P-1}x}$ and $e^{\alpha_P x}$. Given our pre-specified sampling kernels, the input signal is observed by each of the two channels and then sampled at a sampling interval T. The samples s_k of the i-th channel, where $i = 1, 2$ are thus given by:

$$s_{i,k} = \langle A_i g(x - \Delta_i), \phi_i(x/T - k) \rangle, \tag{3.3}$$

where $A_1 = 1$ and $\Delta_1 = 0$ within the first channel (or reference channel). Given the samples, we calculate the exponential moments from the two channels, with the proper coefficients $c_{m,k}^{(i)}$, $i = 1, 2$, as follows:

$$\tau_m^{(1)} = \sum_{k=0}^{N} c_{m,k}^{(1)} s_{1,k} = \int_{-\infty}^{\infty} g(x) e^{\alpha_m x} dx, \tag{3.4}$$

where $m = 0, 1, \ldots, P$ and,

$$\tau_m^{(2)} = \sum_{k=0}^{N} c_{m,k}^{(2)} s_{2,k} = \int_{-\infty}^{\infty} A_2 g(x - \Delta_2) e^{\alpha_m x} dx = A_2 \tau_m^{(1)} e^{\alpha_m \Delta_2}, \tag{3.5}$$

where $m = P - 1, P, \ldots, 2P - 1$. Here, the parameter N represents the number of samples measured. As shown in the equations above, in our setup we have two common parameters between the exponents of the two channels, that is at $m = P-1$ and $m = P$. Given these two common parameters, we can build the following system of equations:

$$\tau_P^{(2)} = A_2 \tau_P^{(1)} e^{\alpha_P \Delta_2} \tag{3.6}$$

$$\tau_{P-1}^{(2)} = A_2 \tau_{P-1}^{(1)} e^{\alpha_{P-1} \Delta_2}, \tag{3.7}$$

Now, from the above equations, the delay and the gain parameters are estimated as follows:

$$\Delta_2 = \frac{1}{\alpha_{P-1} - \alpha_P} ln \left(\frac{\tau_P^{(1)} \tau_{P-1}^{(2)}}{\tau_{P-1}^{(1)} \tau_P^{(2)}} \right) \tag{3.8}$$

and

$$A_2 = \frac{\tau_P^{(2)}}{\tau_P^{(1)}} e^{-\alpha_P \Delta_2}. \tag{3.9}$$

Therefore with only two common parameters, the delay and gain parameters can be estimated. We should point out that by having more than two common parameters, more accurate estimations can be achieved for the unknown parameters, however, one has to bear in mind that having more common parameters require higher spline

orders for each channel and thus more samples will be needed.

From the set of results obtained above we can see that independently of the input signal $g(x)$, it is possible to synchronize the two channels exactly from the samples $s_{i,k}$. However, a few considerations need to be addressed here. In the above analysis we have implicitly assumed that $\tau_m^{(i)} \neq 0$ for $m = P-1, P$ and $i = 1, 2$ and this is not always true. Thus, for a guaranteed synchronization of the channels, some constraints need to be imposed on the signal. In our context we are interested in streams of Diracs and in this case the simple assumption that, given the K locations of the Diracs, the amplitudes are drawn from a non-singular distribution over $\mathbb{R}^K$ guarantees that the event $\tau_m^{(i)} = 0$ has probability zero. This is clearly a fairly mild hypothesis and similar types of conditions can be imposed on any other FRI signal.

3.3.2 Signal Reconstruction

Let us now return to the original problem of reconstructing the stream of Diracs $g(x)$. Given the exact gain and delay of channel two, we can now estimate the moments $\tau_m^{(1)}$, with $m = P+1, P+2, ..., 2P-1$ from $\tau_m^{(2)}$ as follows:

$$\tau_m^{(1)} = \frac{\tau_m^{(2)}}{A_2} e^{-\alpha_m \Delta_2}, \quad m = P+1, P+2, ..., 2P-1. \tag{3.10}$$

It follows that if $2P-1 \geq 2K-1$ or, more simply, $P \geq K$ then a perfect recovery of $g(x)$ is possible from the moments $\tau_m^{(1)}$, $m = 0, 1, ...2P-1$ by using the annihilating filter method discussed in Chapter 2. The advantage of the new setup is that we now require splines of lower order (i.e., $P \geq K$ rather than $P \geq 2K-1$) and this leads to shorter kernels.

For example, as mentioned in Chapter 2, the sampling of an infinite stream of Diracs in the single-channel case requires that there are no more than K Diracs

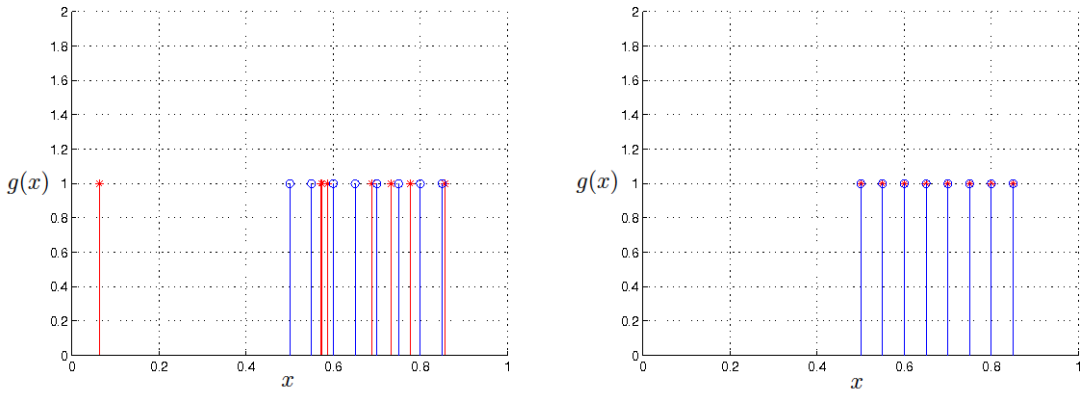
in an interval of size[1] of $2K(P+1)T \leq 4K^2T$, where we have used the fact that in the single-channel setup $P \geq 2K-1$. In the case of the two-channel acquisition system, the interval is reduced to $(2K^2+2K)T$ since $P \geq K$. This indicates that in the new setup we can either sample signals with a higher concentration of Diracs or alternatively for the same signal we can almost halve the sampling rate.

As an example, let us assume that our input signal $g(x)$ consists of 8 Diracs located within $x_k \in [0, 1s[$. We sample this signal at $T = 0.05s$ with a single-channel and also a two-channel sampling system where the unknown gain and delay parameters are $A_2 = 2$ and $\Delta_2 = 0.1s$ respectively. The order of the E-spline sampling kernel in the single-channel case is set to its least possible value of $2K = 16$, and in the two-channel case is set to $K+1 = 9$ for both channels. Figure 3.2 shows the simulations results for both scenarios. With $N = 20$ samples, the retrieval of the locations of the input signal is not possible with the single-channel case (Figure 3.2(a)). However, with the multichannel sampling system employed, we can perfectly retrieve all the 8 Diracs with shorter sampling kernels for each channel (Figure 3.2(b)).

3.3.3 Generalization

Thus far, we have restricted the E-spline sampling kernel to be of equal order $P+1 = K+1$ for each channel. It is also possible to employ sampling kernels of different orders, as long as the following criteria are met: Denoting the E-spline order of channel 1 and 2 with $Q1$ and $Q2$ respectively, we require the order of each kernel to be $Q1 \geq 2$ and $Q2 \geq 2$ and the overall E-spline order to be $Q1 + Q2 \geq 2K+2$. To make things clearer, Figure 3.3 shows the sampling region for the multichannel case. As can be seen from the figure, as long as the equality $Q1 + Q2 \geq 2K+2$ holds, then we can synchronize the two channels and perfectly reconstruct the input

[1]For E-splines of order $P+1$, the support is also $P+1$.



(a) (b)

Figure 3.2: Multichannel sampling with $M = 2$ versus single-channel sampling. (a) Single-channel sampling of 8 Diracs with $N = 20$ samples taken. (b) Multichannel sampling of the same signal with $N = 20$ samples taken for each channel, $M = 2$, $A_2 = 2$ and $\Delta_2 = 0.1s$. In both figures, Diracs with circles on top are the true locations and Diracs with asterisks on top are the reconstructed Diracs.

signal consisting of K Diracs.

Having gone through the two-channel case, the extension to the case of M channels is now straightforward. By designing each sampling kernel so that pairs of channels have two moments in common, it is possible to synchronize the channels using Equations (3.8) and (3.9) and then reconstruct $g(x)$. For M channels and K Diracs the requirement is now that $MP - (M - 1) \geq 2K - 1$ which implies $P \geq \lceil (2K - 2)/M \rceil + 1$. This indicates that by using an M-channel system we can either sample the same FRI signals with a reduced sampling rate proportional to $\sim 1/TM$ or sample signals with a much higher density of Diracs.

Having presented our algorithm for multichannel sampling of FRI signals, we now assess the resilience to noise of the proposed system.

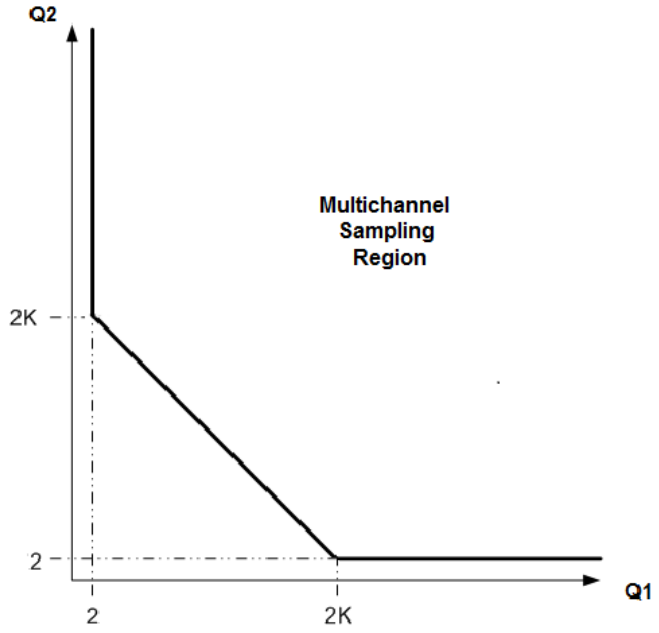


Figure 3.3: Multichannel sampling region with $M = 2$. Here, $Q1$ and $Q2$ represent the E-spline sampling kernel order for each channel.

3.4 Noisy Scenario

We have seen already that, in the noiseless case, a multichannel acquisition system achieves perfect reconstruction of FRI signals with a sampling rate proportional to $1/TM$. Thus, perfect reconstruction is achieved at a sampling rate lower than the single channel case. We now show both theoretically and numerically that, if we do not reduce the sampling rate and leave it fixed at $1/T$, a multichannel system is more resilient to noise than a single-channel one. For numerical evaluation, we will apply the denoising algorithms discussed in Chapter 2, particularly the matrix pencil method and the Cadzow's algorithm. For theoretical evaluation, we will use a common and useful tool known as the Cramér-Rao bound which gives the lowest achievable bound at a given noise level, for any unbiased estimator.

3.5 Multichannel Sampling of FRI Signals in the Presence of Noise

Consider the multichannel sampling setup of Figure 3.1 and let us now assume that, due to the noise, the samples we measure are given by:

$$\hat{s}_{i,k} = \langle A_i g(x - \Delta_i), \phi_i(x/T - k) \rangle + \epsilon_{i,k}, \tag{3.11}$$

where $k = 0, 1, ...N-1$, $i = 1, 2, ..., M$, $\epsilon_{i,k}$ is i.i.d. additive Gaussian noise with zero mean and variance σ^2 and $T = L/N$, with $L = 1$ seconds. We also assume that the gains introduced by the channels are all equal and known a-priori, that is, $A_i = 1$, $i = 1, 2, ..., M$, therefore only the delays need to be estimated. Moreover, we assume that the input signal $g(x)$ has $K = 3$ Diracs with known fixed amplitudes. We have considered these assumptions only for the sake of simplicity, however, in practice the gain parameters introduced and the amplitudes of the Diracs are not necessarily known a-priori.

Our aim is to compare the performance of the single-channel setup (i.e., $M = 1$) against the two-channel and three-channel systems. Since for all cases we have a standard parametric estimation problem, we use Cramér-Rao bounds (CRB) to compare the minimum bounds of the different setups and probe whether the multichannel systems are in theory more resilient to noise. The Cramér-Rao bound is given by the inverse of the Fisher information (see Appendix A for full derivation[2]):

$$CRB(\mathbf{\Theta}) = \sigma^2 \left(\sum_{k=0}^{N-1} \nabla f(\mathbf{\Theta}, k) \cdot \nabla f(\mathbf{\Theta}, k)^T \right)^{-1}, \tag{3.12}$$

[2]Regarding the single-channel case, we should point out that, Blu et al. [14] have computed the Cramér-Rao bounds for sampling a single Dirac with a sinc kernel and Homann et al. [26] have computed the Cramér-Rao bounds for sampling a single Dirac with polynomial and exponential reproducing kernels.

where $\mathbf{\Theta}$ is the vector of all unknown parameters to be estimated, σ^2 is the variance of the noise, N is the total number of the samples, k is the sample index, ∇ represents the divergence operator and $f(\mathbf{\Theta}, k)$ is a function that depends on the $\mathbf{\Theta}$ vector and the sample index k and is obtained from the samples s_k.

In the single-channel scenario, the reconstruction of $g(x)$ involves the estimation of three unknown parameters (i.e., the three locations of the Diracs) and therefore its corresponding $\mathbf{\Theta}$ vector will consist of three parameters $\mathbf{\Theta} = (x_1, x_2, x_3)$. The vector $\nabla f(\mathbf{\Theta}, k)$ for the single-channel case will therefore be:

$$\nabla f(\mathbf{\Theta}, k)_{M=1} = \left(\frac{\partial \hat{s}_k}{\partial x_1}, \frac{\partial \hat{s}_k}{\partial x_2}, \frac{\partial \hat{s}_k}{\partial x_3} \right)^T. \tag{3.13}$$

In the multichannel setup however, one has to retrieve the unknown delays for the synchronization process as well as the unknown locations of the Diracs. Therefore, there are four parameters to be estimated when $M = 2$, that is $\mathbf{\Theta} = (x_1, x_2, x_3, \Delta_2)$, and five parameters when $M = 3$, that is $\mathbf{\Theta} = (x_1, x_2, x_3, \Delta_2, \Delta_3)$. Their corresponding divergence vectors are as follows:

$$\nabla f(\mathbf{\Theta}, k)_{M=2} = \left(\frac{\partial \hat{s}_k}{\partial x_1}, \frac{\partial \hat{s}_k}{\partial x_2}, \frac{\partial \hat{s}_k}{\partial x_3}, \frac{\partial \hat{s}_k}{\partial \Delta_2} \right)^T \tag{3.14}$$

$$\nabla f(\mathbf{\Theta}, k)_{M=3} = \left(\frac{\partial \hat{s}_k}{\partial x_1}, \frac{\partial \hat{s}_k}{\partial x_2}, \frac{\partial \hat{s}_k}{\partial x_3}, \frac{\partial \hat{s}_k}{\partial \Delta_2}, \frac{\partial \hat{s}_k}{\partial \Delta_3} \right)^T. \tag{3.15}$$

To construct the Fisher information matrix, we evaluate the vectors shown above by taking partial derivatives of the samples, and then calculate the matrix multiplication $\nabla f(\mathbf{\Theta}, k) \cdot \nabla f(\mathbf{\Theta}, k)^T$, to obtain 3×3, 4×4 and 5×5 Fisher information matrices for each of the cases $M = 1, 2$ and 3. It should be pointed out that the total number of samples for $M = 2$ and $M = 3$ is $2N-1$ and $3N-1$ respectively. As the large number of unknown parameters leads to a fairly large Fisher information matrices, it is simpler to evaluate the CRB numerically for all cases. Therefore, after constructing the Fisher information matrices, the values of the sampling kernels and

their corresponding partial derivatives, at different sampling indices, are evaluated numerically.

3.5.1 Simulation Results

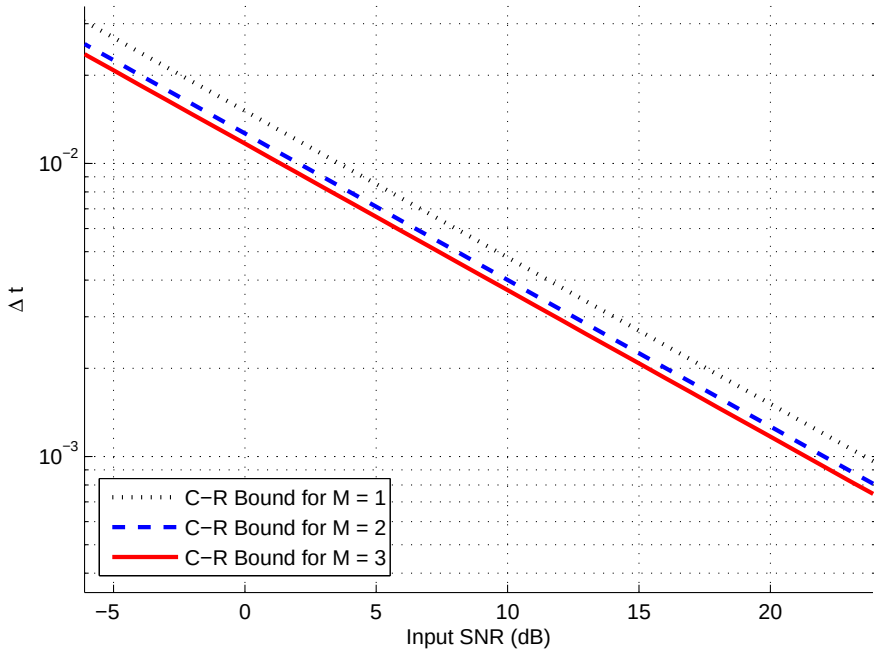


Figure 3.4: CRB for single-channel and multichannel sampling systems. The input SNR is calculated as $10log_{10}\frac{||s_k||^2}{\sigma^2}$ where σ^2 is the noise variance and Δt is the uncertainty on the estimated locations. Dirac locations are set at 0.5, 0.6 and 0.7 for all cases. The delays Δ_2 and Δ_3 are fixed at $\frac{T}{2}$ and T respectively.

The sampling kernel used in our analysis is a complex-valued E-spline function with equally spaced, purely imaginary exponents (for derivation of the CRB for complex-valued functions, see Appendix A). The order of E-splines for each channel is fixed and set to $P = 9$. The CRB for the estimation of the Diracs for the three cases discussed previously, assuming a fixed number of samples $N = 20$, are shown in Figure 3.4. Interestingly, the results reveal that the CRB improves with the number of channels. More precisely, the CRB improvement when going from single-channel to two-channels is approximately 0.86dB, while the improvement when going from

single-channel to three-channel system is approximately 1.1dB. It is interesting to see that, despite the fact that the unknown delays need to be estimated in order to synchronize the channels, there is still a noticeable gain by using multichannel sampling setup when compared to the single-channel sampling case.

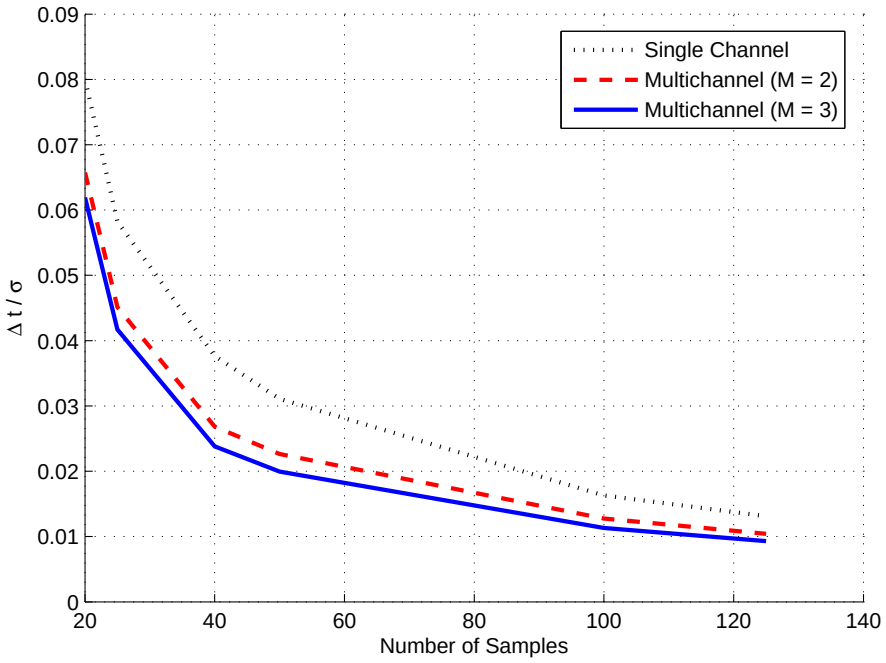


Figure 3.5: Theoretical uncertainties on the estimated locations with varying sampling rates. The input SNR is calculated as $10log_{10}\frac{||s_k||^2}{\sigma^2}$ where σ^2 is the noise variance and Δt is the uncertainty on the estimated locations. Dirac locations are set at 0.5, 0.6 and 0.7 for all cases. The delays Δ_2 and Δ_3 are fixed at $\frac{T}{2}$ and T respectively.

Furthermore, in Figure 3.5 we show the CRB of each sampling system at varying sampling rates. We can see that at a given target uncertainty for the estimation of the locations, there is a reduction in the number of samples needed when going from single-channel to multichannel sampling systems. For example, at the reconstruction quality of $\frac{\Delta t}{\sigma} = 0.04$, the number of samples could be reduced from 38 samples to 27 samples when going from the single-channel to the three-channel setup.

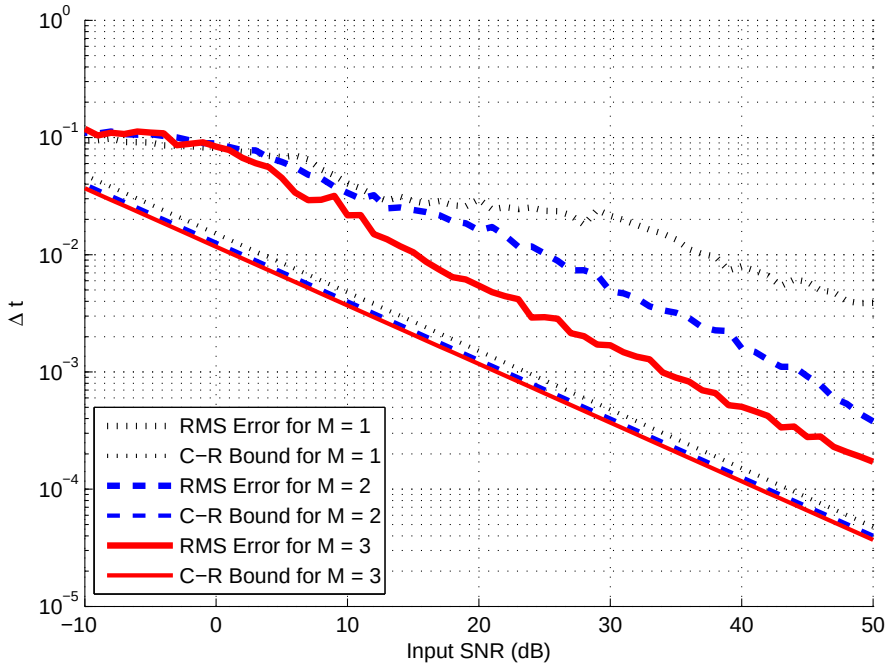


Figure 3.6: Numerical results with single and multichannel sampling. The input SNR is calculated as $10log_{10}\frac{||s_k||^2}{\sigma^2}$ where σ^2 is the noise variance and Δt is the uncertainty on the estimated locations. Dirac locations are set at 0.5, 0.6 and 0.7 for all cases. The delays Δ_2 and Δ_3 are fixed at $\frac{T}{2}$ and T respectively.

To analyse the performance of the reconstruction algorithm, Figure 3.6 presents some actual numerical results on the uncertainty of the estimated locations which are also compared against the theoretical bounds from Figure 3.4. The delays are estimated using Equation (3.8), while the locations of the Diracs are obtained using the matrix pencil method and also the Cadzow's algorithm to further denoise the exponential moments τ_m. While none of the algorithms achieve the CRB, our results show that the gain in performance with multichannel sampling over single-channel sampling can be significant. For instance, at input SNR= 15dB, the gain in performance from single-channel to three-channels is approximately 4.4dB. The additional noise robustness in our numerical simulations when compared to the CRB is mostly due to the denoising algorithms that we use. This will be further discussed in our 'Future Research' section in Chapter 6.

3.6 Summary

In this chapter we have extended the theory of sampling FRI signals to the case of multichannel acquisition systems. We illustrated that by synchronizing the different channels of the proposed multichannel sampling setup, one can estimate the unknown delays and gains introduced within the channels, regardless of the input FRI signal. In the case of noisy measurements, we showed that by evaluating the CRB, the multichannel system can achieve a target performance with a number of samples which is smaller than the number of samples needed in the single-channel setup. The improvement in the resilience to noise of the multi-channel architecture was also confirmed with numerical simulations.

Chapter 4

System Identification based on the Theories of Finite Rate of Innovation Sampling

4.1 Introduction

In many practical applications the impulse response of an unknown system is required to be estimated. This problem which is usually referred to as "System Identification Problem" in literature has a vast number of applications, such as line echo cancellation, channel equalization and model mismatch estimation. Figure 4.1 shows a system identification problem setup where a signal $g(x)$ is fed to an unknown system with impulse response $\psi(x)$. The output signal from the unknown system then goes through a sampling process with a sampling rate $1/T$ which outputs the samples s_k.

The aim of system identification is to completely determine the function $\psi(x)$ from the samples s_k. Unlike standard techniques for system identification, such as least mean squares (LMS) algorithm [63], which require the sampling rate to be at

or above the Nyquist rate, sparse sampling techniques could be employed to identify the unknown system at sub-Nyquist sampling rates.

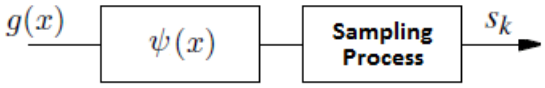


Figure 4.1: A system identification problem setup. Here, $g(x)$ represents the continuous-time input signal, $\psi(x)$ represents the unknown system and s_k represent the output samples

System identification using the sparse sampling techniques has already been considered (or partly considered) in [53, 19, 33]. In [53] McCormick et al. consider a similar problem to ours and present a novel sub-Nyquist sampling algorithm for system identification in the frequency domain by exploiting the sparsity feature of the unknown system. In this method, the impulse response of the unknown system is adaptively estimated using a frequency-domain LMS filter [53], however, the input signal is considered to be known and bandlimited. In this chapter we will propose a novel algorithm for simultaneous estimation of sparse signals along with system identification using the theories of sparse sampling.

The novelty of our work is divided into two sections; first, by employing the multichannel sampling setup presented in the previous chapter, we propose a novel algorithm for the system identification problem with known input signal, that is for the case when both the input signal and the output samples are known. Then we consider the problem of blind system identification where by blind we mean that only the output samples are given and the input signal is not known. We will propose a novel algorithm for simultaneously estimating the input FRI signal and also the unknown system using an iterative algorithm. We will show that, based on our numerical simulations, the solution to the blind system identification problem is normally convergent.

The organization of the chapter is as follows: In Section 4.2, we propose our

novel algorithm for system identification with known input signal. We will present few examples on system identification problems and highlight their corresponding applications. In Section 4.3, we present our iterative algorithm for blind system identification and show that simultaneous estimation of both the input FRI signal and the unknown system $\psi(x)$ is possible. We summarize the chapter in Section 4.4.

4.2 System Identification with Known Input Signal

In the previous chapter, we presented a novel algorithm for multichannel sampling of FRI signals. We assumed that in the multichannel sampling setup, each channel introduces an unknown delay and gain parameter on the incoming input signal. It can be shown that the channel parameters estimation stage or rather the channel synchronization stage, could be thought of a system identification problem, where the unknown system to be identified is $\psi(x) = A_i\delta(x - \Delta_i)$ for $i = 2, 3, \ldots, M$. In this section we will show how a similar sampling scheme could be employed to estimate the unknown function $\psi(x)$ from the samples, given the input signal.

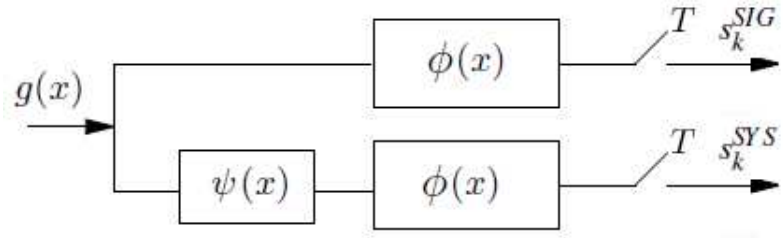


Figure 4.2: System identification setup with known input signal $g(x)$. Here, the function $\phi(x)$, identical in both channels, represents the E-spline sampling kernel and the output samples s_k^{SIG} and s_k^{SYS} represent the signal and the system samples respectively.

Figure 4.2 shows the sampling setup proposed, where $g(x)$ represents the input signal, $\psi(x)$ represents the stationary unknown system to be identified, $\phi(x)$

represents the sampling kernel which is considered to be an E-spline and s_k represent the samples. As can be seen, a two-channel sampling system with identical sampling kernels has been employed to sample the input signal with and without the unknown system $\psi(x)$. In the first channel, the input signal is directly sampled with our pre-specified E-spline sampling kernel $\phi(x)$. We recall here that the Fourier transform of $\phi(x)$ is given by:

$$\hat{\phi}(j\omega) = \prod_{m=0}^{P} \left(\frac{1 - e^{\alpha_m - j\omega}}{j\omega - \alpha_m} \right), \tag{4.1}$$

where P depends only on the structure of the unknown function $\psi(x)$. This will be more evident later on. The samples s_k at the output of the first channel are therefore:

$$s_k^{SIG} = \langle g(x), \phi(x-k) \rangle. \tag{4.2}$$

Given the samples, we then calculate the exponential moments of the input signal, as was shown in Equation (2.20). For the sake of clarity and without loss of generality, we will assume that the α_m parameters are purely imaginary $\alpha_m = jm\lambda$ and therefore:

$$\tau_m^{SIG} = \hat{g}(\alpha_m), \tag{4.3}$$

where $\hat{g}(\alpha_m)$ represents the Fourier transform of the signal $g(x)$ at α_m with $m = 0, 1, \ldots, P$. In the second channel, the same input signal is fed through the unknown function $\psi(x)$ and then sampled with the same sampling kernel. Therefore, its corresponding samples s_k are:

$$s_k^{SYS} = \langle g(x) * \psi(x), \phi(x-k) \rangle. \tag{4.4}$$

Given the samples of the second channel, we calculate its corresponding exponential moments. This will lead to:

$$\tau_m^{SYS} = \hat{g}(\alpha_m) \cdot \hat{\psi}(\alpha_m). \tag{4.5}$$

Here, the function $\hat{\psi}(\alpha_m)$ represents the Fourier transform of the function $\psi(x)$ at α_m. Moreover, the above equation is deduced from that fact that convolution in time domain corresponds to multiplication in Fourier domain. From the set of results obtained above, we derive that:

$$\frac{\tau_m^{SYS}}{\tau_m^{SIG}} = \frac{\hat{g}(\alpha_m) \cdot \hat{\psi}(\alpha_m)}{\hat{g}(\alpha_m)} = \hat{\psi}(\alpha_m), \tag{4.6}$$

where we have assumed that $\hat{g}(\alpha_m) \neq 0$. Therefore, by dividing the exponential moments obtained from the two-channels, we have shown that the Fourier transform of the unknown stationary function $\psi(x)$ can be obtained, regardless of the structure of the input signal. Now, given $\hat{\psi}(\alpha_m)$ with $m = 0, 1, .., P$ and $\alpha_m = jm\lambda$, we will have an inverse problem to solve for the unknown parameters of the function $\psi(x)$. Once the unknown parameters are estimated, the function $\psi(x)$ will be completely determined and the system will therefore be fully identified. In the following section, we show cases where we can solve the above inverse problem (i.e. we identify the system) and highlight the applications in which the proposed system model is of interest.

4.2.1 Identification of a System with K Diracs

Consider the unknown function $\psi(x)$ to be a stream of K Diracs with unknown locations and amplitudes. Applications of such a system could be acoustic room impulse response estimation or line echo cancellation. We already know that the Fourier transform of such a function has a power-sum series form:

$$\hat{\beta}(j\omega) = \sum_{k=1}^{K} a_k \; e^{j\omega x_k}, \tag{4.7}$$

where a_k and x_k correspond to the unknown amplitudes and locations respectively. From the setup shown in Figure 4.2, as previously described, the exponential mo-

ments of the output samples with and without the unknown function, that is τ_m^{SYS} and τ_m^{SIG} are obtained. The Fourier transform of the function $\psi(x)$ at $\alpha_m = jm\lambda$ can now be easily calculated as follows:

$$\frac{\tau_m^{SYS}}{\tau_m^{SIG}} = \hat{\psi}(\alpha_m), \tag{4.8}$$

where,

$$\begin{aligned} \hat{\psi}(\alpha_m) &= \hat{\beta}(\alpha_m) && (4.9)\\ &= \sum_{k=1}^{K} a_k e^{\alpha_m x_k} && (4.10)\\ &= \sum_{k=1}^{K} a_k u_k^m, \quad m = 0, 1, \ldots, P, && (4.11) \end{aligned}$$

where $u_k = e^{j\lambda x_k}$. As $\hat{\psi}(\alpha_m)$ has a power-sum series form, we apply the annihilating filter method to the measurements $\frac{\tau_m^{SYS}}{\tau_m^{SIG}} = \hat{\psi}(\alpha_m)$ to retrieve the unknown parameters a_k and x_k. For such a system, in order to recover the K Diracs, the E-spline sampling kernel is required to be of order $P \geq 2K$. This setup is similar to the one discussed in [33].

As an example, let us assume that our signal $g(x)$ is a piecewise-polynomial signal and the unknown function $\psi(x)$ is a stream of 3 Diracs. Our goal is to estimate the system, given the input signal and its corresponding s_k^{SIG} and s_k^{SYS} samples. Figure 4.3 shows the simulation results for this example. We can see that the unknown system is perfectly estimated, regardless of the structure of the input signal.

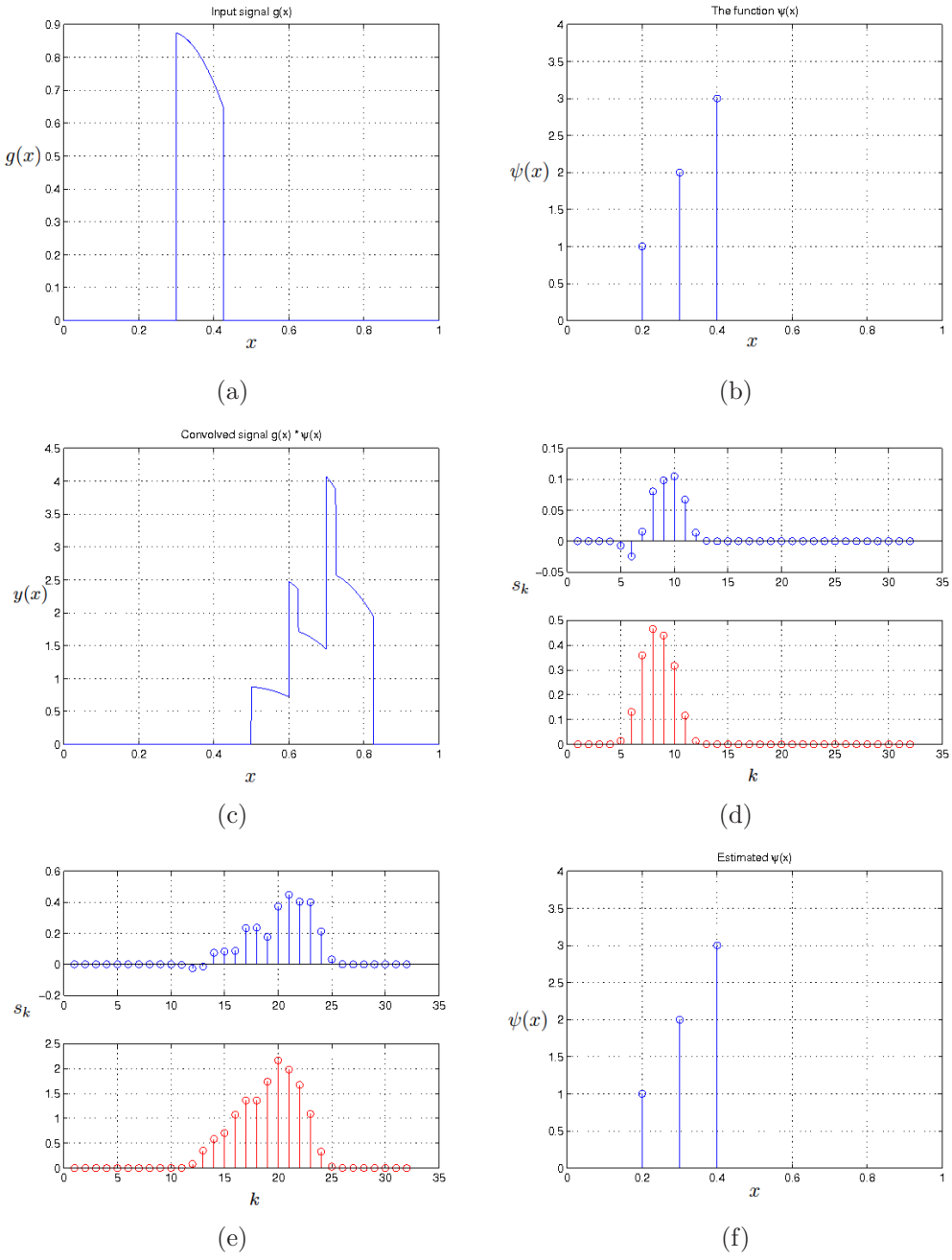


Figure 4.3: System identification of an unknown system $\psi(x)$ consisting of 3 Diracs. (a) The input piecewise-polynomial signal. (b) The true parameters of the function $\psi(x)$. (c) Input signal convolved with the function $\psi(x)$. (d) The real (blue) and imaginary (red) samples of the signal. (e) The real (blue) and imaginary (red) samples of the system. (f) Estimated parameters of the function $\psi(x)$.

4.2.2 B-Splines

Let us consider $\psi(x)$ to be a B-spline $\beta^K(x)$ of unknown order $K+1$. An application of such a system could be the camera lens calibration [6]. This is because the point

spread function of a camera lens is very often assumed to be a Gaussian pulse and B-splines of order $K \geq 2$ are increasingly similar to Gaussian functions. We already know that the Fourier transform function of a B-spline of order $K+1$ is given by:

$$\hat{\beta}^K(j\omega) = \prod_{k=0}^{K} \frac{1-e^{-j\omega}}{j\omega} = \left(\frac{1-e^{-j\omega}}{j\omega}\right)^{K+1}. \tag{4.12}$$

Assuming the unknown function in our setup shown in Figure 4.2 is a B-spline of unknown order, that is $\psi(x) = \beta^K(x)$, then by calculating the exponential moments of the output samples with and without the unknown filter, we can obtain the Fourier transform of $\psi(x)$ at $\alpha_m = jm\lambda$ as follows:

$$\frac{\tau_m^{SYS}}{\tau_m^{SIG}} = \hat{\psi}(\alpha_m), \tag{4.13}$$

where,

$$\hat{\psi}(\alpha_m) = \hat{\beta}^K(\alpha_m) \tag{4.14}$$

$$= \left(\frac{1-e^{-\alpha_m}}{\alpha_m}\right)^{K+1}. \tag{4.15}$$

By taking logarithms on both sides of the equation, the unknown order $K+1$ is calculated as follows:

$$\frac{\log\left(\hat{\psi}(\alpha_m)\right)}{\log\left(\frac{1-e^{-\alpha_m}}{\alpha_m}\right)} = K+1. \tag{4.16}$$

In order to estimate the unknown order of the B-spline, the E-spline sampling kernel is required to be of order $P \geq 1$.

As an example, let us assume that our signal $g(x)$ consists of 2 Diracs and the unknown function $\psi(x)$ to be estimated is a B-spline of order 4. Our goal is to estimate the order of the B-spline, given the input signal and its corresponding s_k^{SIG} and s_k^{SYS} samples. Figure 4.4 shows the simulation results for this example. As can be seen from the figure, the system with B-spline of order 4 has been perfectly

identified.

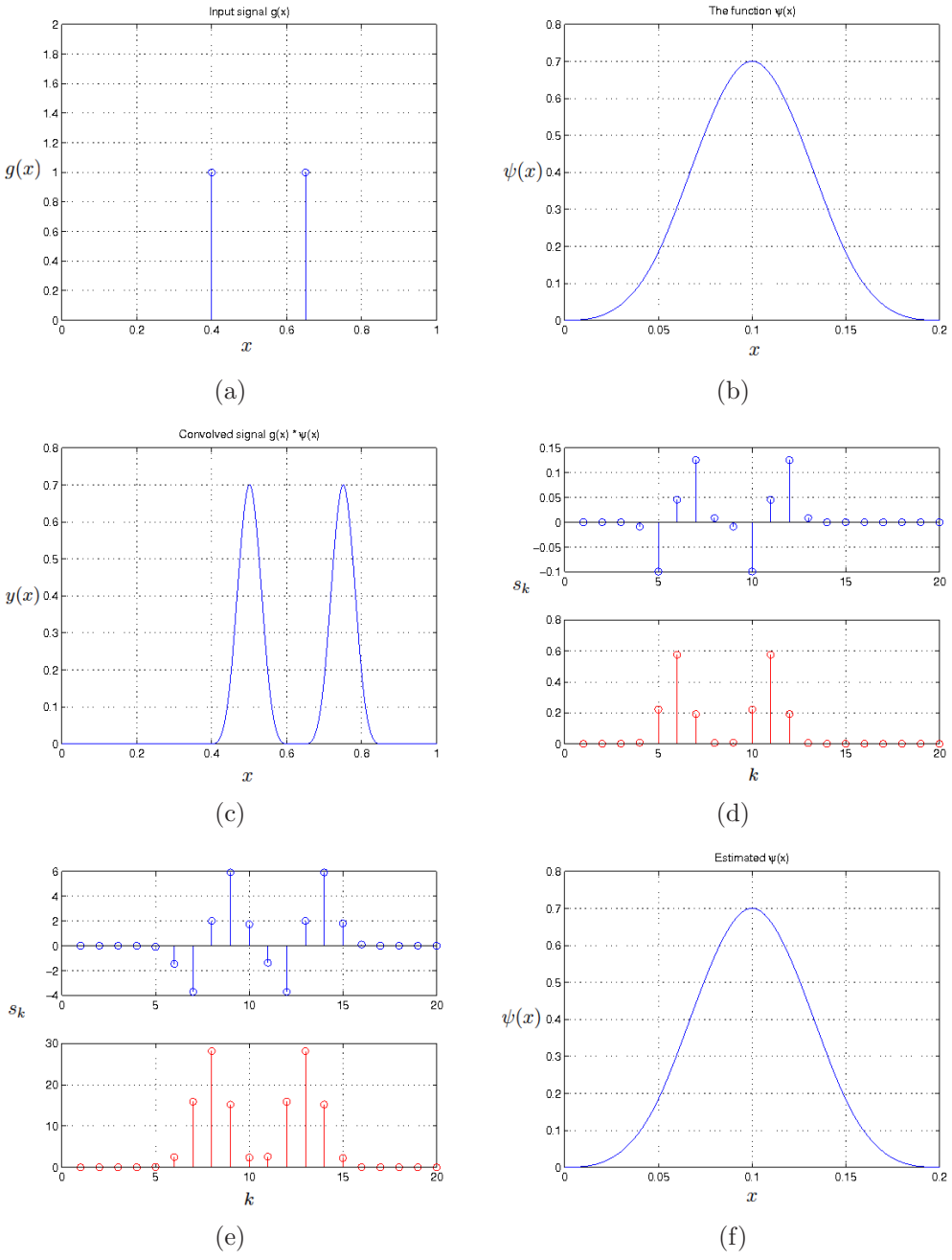


Figure 4.4: System identification of an unknown system $\psi(x)$ with a B-spline of order 4. (a) The input signal with 2 Diracs. (b) The true values of the function $\psi(x)$. (c) Input signal convolved with the function $\psi(x)$. (d) The real (blue) and imaginary (red) samples of the signal. (e) The real (blue) and imaginary (red) samples of the system. (f) The estimated function $\psi(x)$.

4.2.3 E-Splines

Let us now consider $\psi(x)$ to be an E-spline $\beta_{\vec{\gamma}}(x)$ of known order $K+1$ with unknown exponent parameters $\vec{\gamma}$. An application of such a system could be the estimation of the electronic components of a finite order electronic circuit, which will be fully described in the next subsection. As stated previously, the Fourier transform of an E-spline of order $K+1$ is:

$$\hat{\beta}_{\vec{\gamma}}(j\omega) = \prod_{k=0}^{K} \frac{1 - e^{\gamma_k - j\omega}}{j\omega - \gamma_k}. \tag{4.17}$$

From the setup shown in Figure 4.2, the exponential moments τ_m^{SYS} and τ_m^{SIG} are calculated from the samples s_k^{SYS} and s_k^{SIG} respectively. Given the exponential moments, the Fourier transform of the function $\psi(x)$ at $\alpha_m = jm\lambda$ can be obtained as follows:

$$\frac{\tau_m^{SYS}}{\tau_m^{SIG}} = \hat{\psi}(\alpha_m), \tag{4.18}$$

where,

$$\hat{\psi}(\alpha_m) = \hat{\beta}_{\vec{\gamma}}(\alpha_m) \tag{4.19}$$

$$= \prod_{k=0}^{K} \frac{1 - e^{\gamma_k - \alpha_m}}{\alpha_m - \gamma_k}. \tag{4.20}$$

Calculation of the unknown parameters of the E-splines (as the unknown function $\psi(x)$) is more involved. We first need to simplify both the numerator and denominator of the E-spline function. Simplifying the numerator gives:

$$\prod_{k=0}^{K} \left(1 - e^{\gamma_k - \alpha_m}\right) = \prod_{k=0}^{K} \left(1 - a_k u^m\right), \tag{4.21}$$

where $a_k = e^{\gamma_k}$ and $u^m = e^{\alpha_m}$. This can be further simplified as follows:

$$\prod_{k=0}^{K}(1 - a_k u^m) = (1 - a_1 u^m)(1 - a_2 u^m)...(1 - a_K u^m) \tag{4.22}$$

$$= \sum_{k=0}^{K} q_k u^{km} \tag{4.23}$$

$$= \sum_{k=0}^{K} q_k t_k^m, \tag{4.24}$$

where $t_k = u^k$. Simplifying the denominator gives:

$$\prod_{k=0}^{K}(\alpha_m - \gamma_k) = Q(m) = \sum_{k=0}^{K} r_k m^k, \tag{4.25}$$

where $Q(m)$ is a polynomial of degree $K + 1$. Rearranging Equation (4.20) using Equation (4.24) and (4.25), leads to:

$$\sum_{k=0}^{K} r_k m^k \cdot \hat{\psi}(\alpha_m) = \sum_{k=0}^{K} q_k t_k^m. \tag{4.26}$$

The above equation can be considered as a linear system, consisting of $2K$ unknowns with the unknown parameters being $\mathbf{\Theta} = (r_0, r_1, \ldots, r_{K-1}, q_1, \ldots, q_K)$. Here, $q_0 = 1$, $t_0 = 1$ and t_k are a known set of parameters. As we have a linear system, by constructing the following matrix equation and taking its inverse we are able to calculate the unknown parameters:

$$\begin{bmatrix} \psi(\alpha_0) & 0 & \ldots & 0 & 1 & \ldots & 1 \\ \psi(\alpha_1) & \psi(\alpha_1) & \ldots & \psi(\alpha_1) & t_2 & \ldots & t_K \\ \psi(\alpha_2) & 2\psi(\alpha_2) & \ldots & 2^{K-1}\psi(\alpha_2) & t_2^2 & \ldots & t_K^2 \\ \vdots & \ldots & \ldots & \ldots & \ldots & \ldots & \vdots \\ \psi(\alpha_{2K-1}) & (2K-1)\psi(\alpha_{2K-1}) & \ldots & (2K-1)^{K-1}\psi(\alpha_{2K-1}) & t_2^{2K-1} & \ldots & t_K^{2K-1} \end{bmatrix}$$

$$\times \begin{bmatrix} r_0 \\ r_1 \\ \vdots \\ r_{K-1} \\ q_1 \\ \vdots \\ q_K \end{bmatrix} = \begin{bmatrix} 1 \\ 1-\psi(\alpha_1)r_K \\ 1-2^K\psi(\alpha_2)r_K \\ \vdots \\ 1-(2K-1)^K\psi(\alpha_{2K-1})r_K \end{bmatrix}. \tag{4.27}$$

Once the parameters are estimated, by taking the roots of the polynomial q_k for $k = 0, 1, \ldots, K$ and then taking the logarithm of the roots we obtain the unknown γ_k parameters, since $roots(q_k) = a_k = e^{\gamma_k}$. The E-spline sampling kernel is required to be of order $P \geq 2K$ in order to estimate the parameters of the E-spline function $\psi(x)$.

As an example, let us assume that our input signal $g(x)$ consists of 2 Diracs and the unknown function $\psi(x)$ to be estimated is an E-spline of order 2 with $\gamma_0 = -1$ and $\gamma_1 = 2$. Our goal is to estimate the system, given the input signal and its corresponding s_k^{SIG} and s_k^{SYS} samples. Figure 4.5 shows the simulation results for this example. As can be seen from the figure, the function $\psi(x)$ is fully identified (The small error is due to numerical imprecision).

4.2.4 Linear Time-Invariant Circuits

Finite order linear time-invariant electronic circuits can be modelled as modified E-splines [27]. In general any p-th order electronic circuit has a transfer function ($s = j\omega$):

$$\hat{\beta}(s) = \frac{b_q s^q + b_{q-1}s^{q-1} + \ldots + b_1 q + b_0}{a_p s^p + a_{p-1}s^{p-1} + \ldots + a_1 p + a_0} = \frac{\sum_{q=0}^{Q} b_q s^q}{\sum_{p=0}^{P} a_p s^p}. \tag{4.28}$$

The equation above has a very similar structure to the E-spline case and one may suggest a similar simplifying procedure with constructing a system of linear equations with $Q + P + 1$ unknowns, as discussed previously. However, the transient response of a finite order electronic circuit is infinite and this would mean that the function $\psi(x)$ will not be time-limited. Assuming that the electronic circuit has a

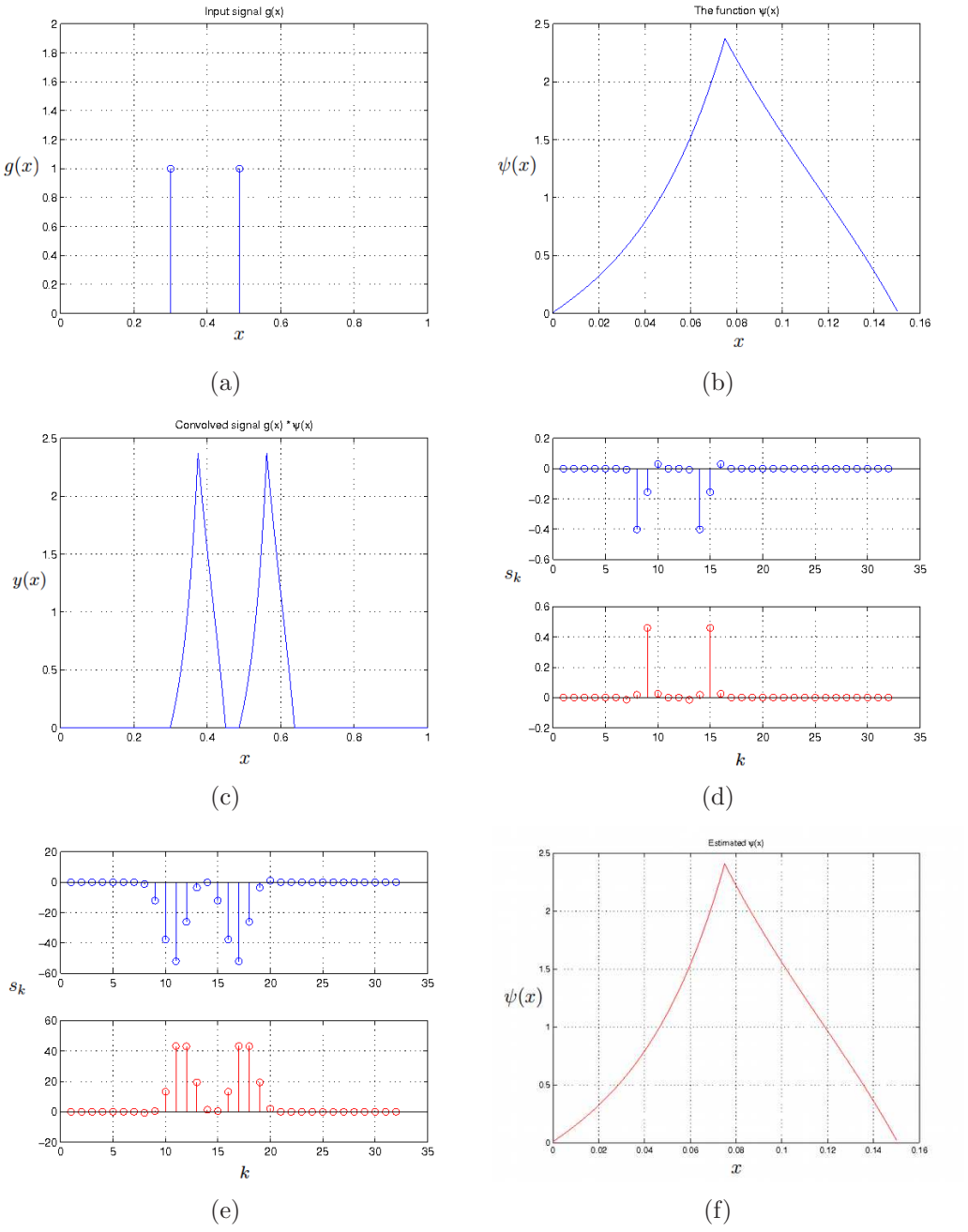


Figure 4.5: System identification of an unknown system $\psi(x)$ with an E-spline of order 2 and parameters $\gamma_0 = -1$ and $\gamma_1 = 2$. (a) The input signal with 2 Diracs. (b) The true values of the function $\psi(x)$. (c) Input signal convolved with the function $\psi(x)$. (d) The real (blue) and imaginary (red) samples of the signal. (e) The real (blue) and imaginary (red) samples of the system. (f) The estimated function $\psi(x)$ with MSE of 3×10^{-6} due to numerical imprecision.

fast decaying transient response, we can approximate the above function to have a finite duration and thus, as in the E-spline case, a linear system of matrix equations

with $Q + P + 1$ unknown parameters can be constructed.

As a simple example, consider the setup shown in Figure 4.6 where the unknown function $\psi(x)$ is a first order RC circuit (low-pass filter) with $R = 100k\Omega$ and $C = 20\mu F$. We already know that the transfer function of such a circuit is:

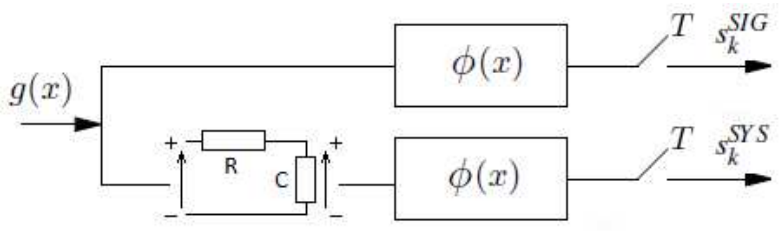


Figure 4.6: First order RC circuit as the unknown function $\psi(x)$. Here, the output samples s_k^{SIG} and s_k^{SYS} represent the input signal and the system samples respectively.

$$\hat{\beta}(j\omega) = \frac{\gamma}{\gamma + j\omega}, \tag{4.29}$$

where $\gamma = 1/RC$. Our goal is to estimate the parameter $\gamma = 1/RC = \frac{1}{100k \cdot 20\mu} = 0.5$ and therefore identify the transfer function. Like before, we obtain the exponential moments at $\alpha_m = jm\lambda$, which in turn will lead to the the Fourier transform of the function $\psi(x)$ at α_m:

$$\frac{\tau_m^{SYS}}{\tau_m^{SIG}} = \hat{\psi}(\alpha_m), \tag{4.30}$$

where,

$$\hat{\psi}(\alpha_m) = \hat{\beta}(\alpha_m) \tag{4.31}$$

$$= \frac{\gamma}{\gamma + \alpha_m}. \tag{4.32}$$

The above function has one unknown parameter only and therefore the product RC can be estimated as follows:

$$\gamma = \frac{\alpha_m \cdot \hat{\psi}(\alpha_m)}{1 - \hat{\psi}(\alpha_m)}. \tag{4.33}$$

Figure 4.7 shows the simulation results for this example. As can be seen from the figure, the transient response is perfectly estimated.

We have shown in this section that by having access to the input signal and by exploiting the sparsity feature of the unknown system, system identification problem can be solved at low sampling rates by employing a multichannel sampling setup. Sampling at lower rates, which generally results in the reduction of computational complexity, is indeed the main advantage of our proposed algorithm when compared to the classical methods. The amount of complexity reduction and other features such as accuracy and noise robustness of our proposed method is an area of future research and currently is out of the scope of our work.

4.3 Blind System Identification

In the previous section, we looked at the case where both the input signal and the output samples were known. For the case of blind system identification however, that is when both the input signal and the system are unknown, the above solution cannot be used directly and the problem is more involved. However, a recursive version of the previously discussed method, as shown in Figure 4.8(a) and 4.8(b), can be utilized to estimate both the input FRI signal and the unknown function $\psi(x)$. Let us assume that the input sparse signal is a stream of Diracs with unknown locations. As shown in Figure 4.8(a), the unknown input signal is fed to the unknown system $\psi(x)$ and then is sampled with our pre-specified E-spline sampling kernel with $\alpha_m = jm\lambda$. Therefore, its corresponding exponential moments are:

$$\tau_m^0 = \hat{\psi}(\alpha_m) \cdot \hat{g}(\alpha_m), \tag{4.34}$$

where both $\hat{\psi}(\alpha_m)$ and $\hat{g}(\alpha_m)$ are unknown. As our input signal is a stream of Diracs with unknown locations, we directly apply the annihilating filter method to the moments τ_m^0 and obtain an estimate of the input signal, denoted as $\bar{g}(x)$ (Figure 4.8(b)). Once an estimate of the input signal is obtained, we recursively feed the estimated signal $\bar{g}(x)$ back to our pre-specified sampling kernel and obtain

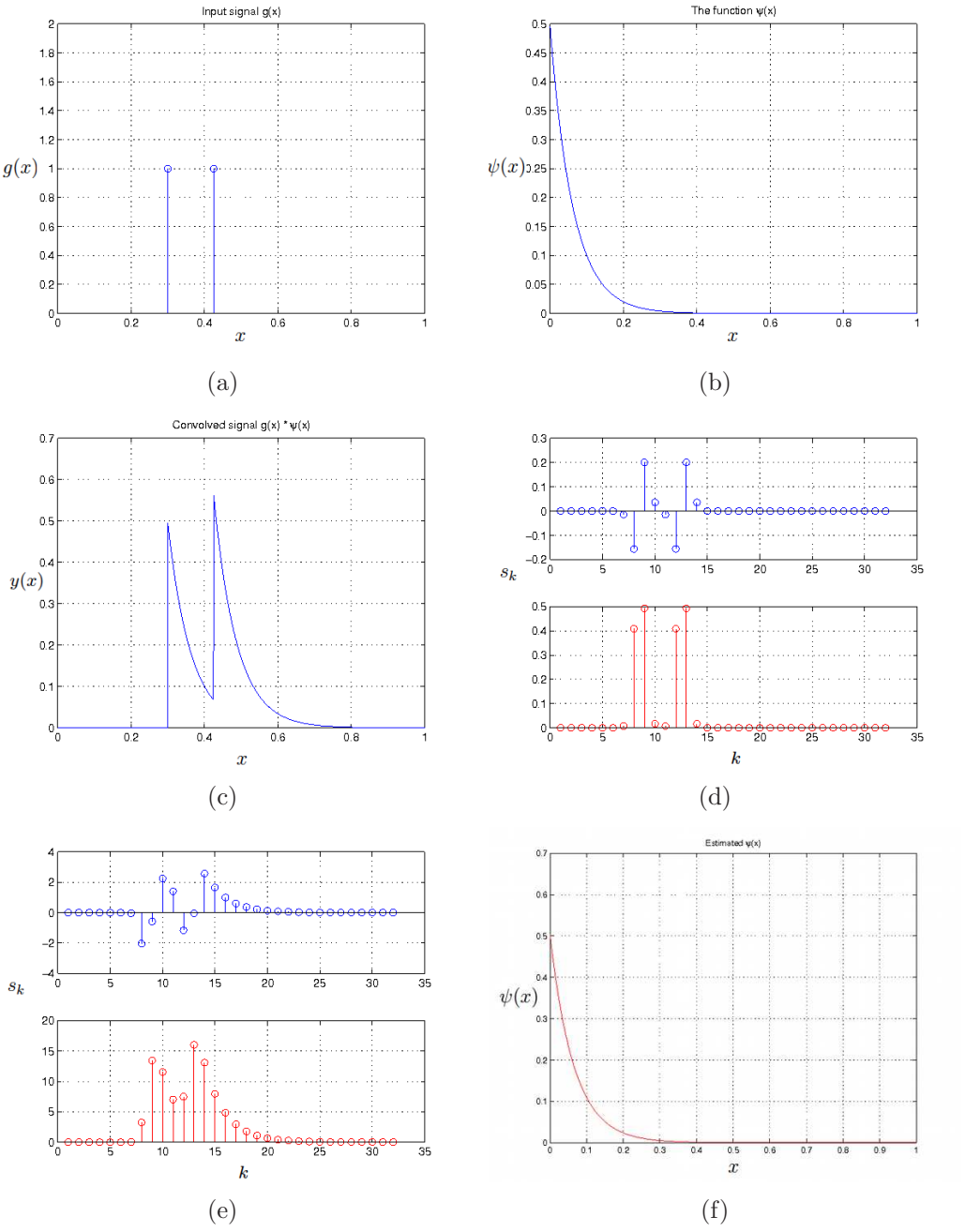


Figure 4.7: System identification of an unknown system $\psi(x)$ with a first order RC circuit where $\gamma = 1/RC = 0.5$. (a) The input signal with 2 Diracs. (b) The true values of the function $\psi(x)$. (c) Input signal convolved with the function $\psi(x)$. (d) The real (blue) and imaginary (red) samples of the signal. (e) The real (blue) and imaginary (red) samples of the system. (f) The estimated function $\psi(x)$ with MSE of 1.3×10^{-6} due to numerical imprecision.

its corresponding exponential moments at each recursion:

$$\tau_m^{upd} = \hat{\tilde{g}}(\alpha_m). \tag{4.35}$$

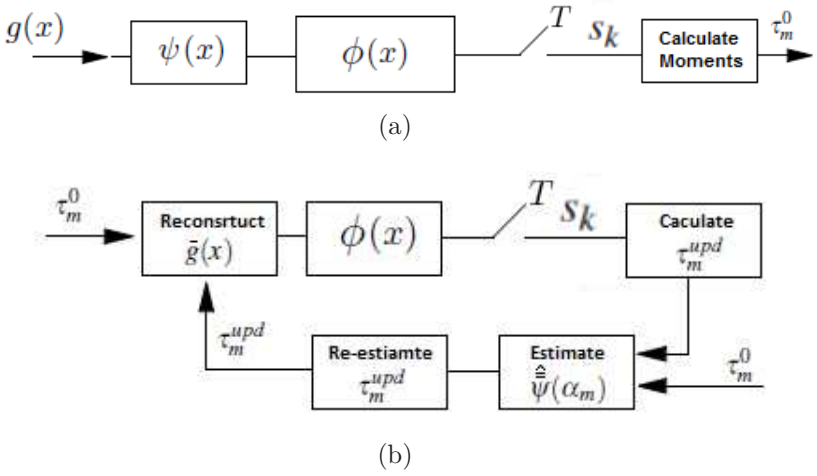


Figure 4.8: The setup proposed for recursive estimation. Here, $g(x)$ is the input FRI signal, $\psi(x)$ is the unknown system to be identified, $\phi(x)$ is the E-spline sampling kernel, τ_m^0 represent the initial measurements, τ_m^{upd} represent the updated measurements, $\hat{\hat{\psi}}(\alpha_m)$ represents the re-estimated Fourier transform of the unknown system and $\bar{g}(x)$ represents the estimated input FRI signal.

Here the superscript "upd" stands for "updated" and $\hat{\bar{g}}(\alpha_m)$ is an estimate of the Fourier transform of the input signal at α_m. Now, we divide the updated exponential moments τ_m^{upd} by the initial measurements τ_m^0 to obtain an estimate for the unknown system $\hat{\psi}(\alpha_m)$ as follows:

$$\frac{\tau_m^0}{\tau_m^{upd}} = \frac{\hat{g}(\alpha_m) \cdot \hat{\psi}(\alpha_m)}{\hat{\bar{g}}(\alpha_m)} = \hat{\hat{\psi}}(\alpha_m). \tag{4.36}$$

From $\hat{\hat{\psi}}(\alpha_m)$, as was shown in Section 4.2, the unknown parameters of the unknown system can be estimated. Once the parameters are estimated, from the model of the unknown system we re-estimate the Fourier transform of the function $\psi(x)$, denoted by $\hat{\hat{\psi}}(\alpha_m)$, and from that we re-estimate the measurements τ_m^{upd} as follows:

$$\tau_m^{upd} = \frac{\tau_m^0}{\hat{\hat{\psi}}(\alpha_m)}. \tag{4.37}$$

We apply the annihilating filter method on the re-estimated τ_m^{upd} and obtain an improved estimate of the unknown input signal $g(x)$. Our empirical results show that by applying the above method recursively, the estimations converge to the

actual input signal $g(x)$ and the unknown function $\psi(x)$.

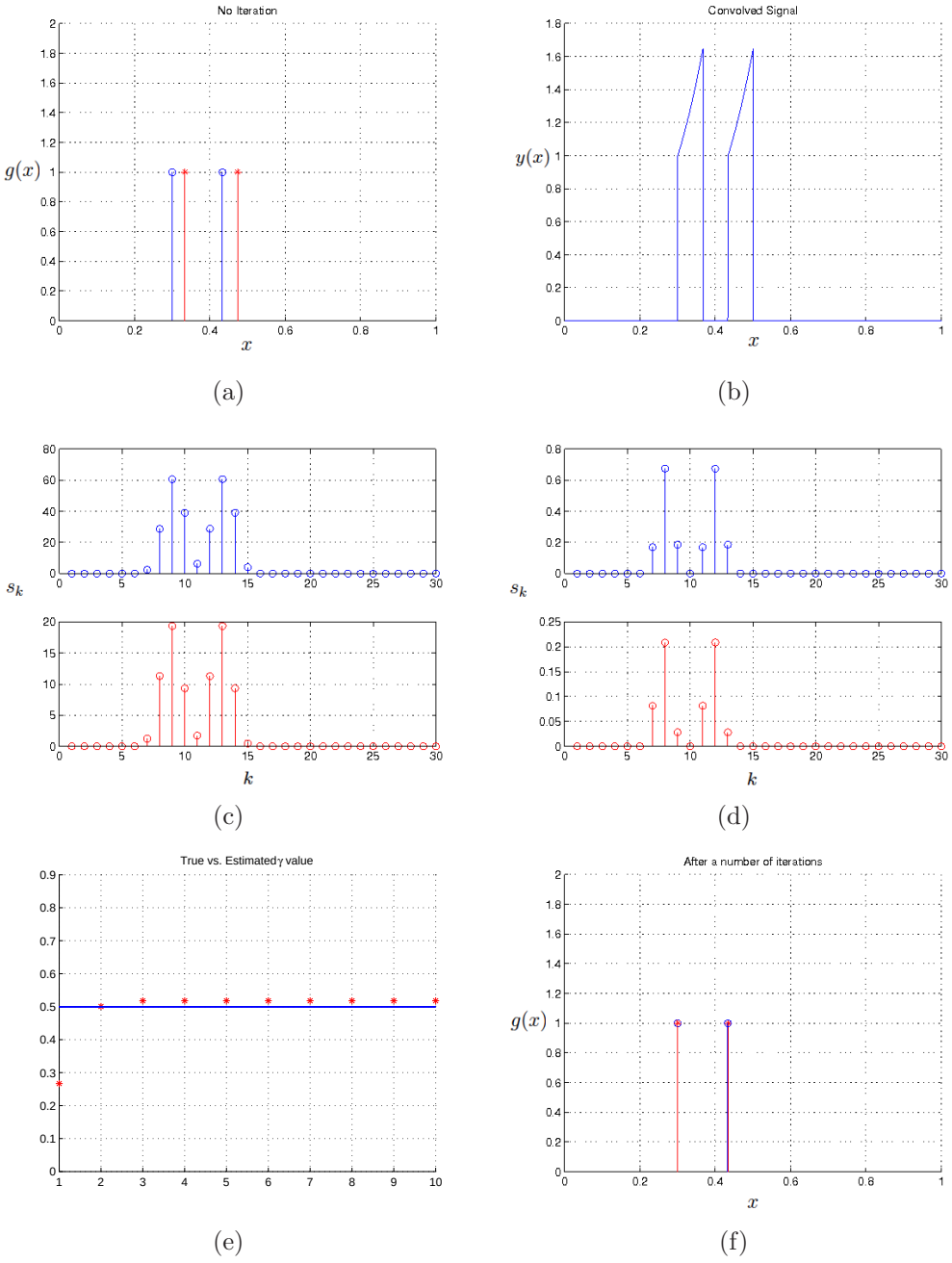


Figure 4.9: Simultaneous estimation of an input sparse signal with a first order E-spline as the unknown function $\psi(x)$. (a) The input signal with 2 Diracs (circle) along with the immediate estimate of the signal (asterisk) with no iterations. (b) Input signal convolved with the function $\psi(x)$. (c) The real (blue) and imaginary (red) samples with no iteration. (d) The real (blue) and imaginary (red) samples after 10 iterations. (e) True vs. estimated values of the parameter γ after 10 iterations. (f) True vs. estimated version of the input signal after 10 iterations.

As an example let us assume that our input signal consists of 2 Diracs with unknown location. Let us also assume that the unknown system to be identified is a first order E-spline with $\gamma = 2$. Our goal is to simultaneously estimate the input signal and also the unknown γ parameter. Figure 4.9 shows the results for the above example after 10 iterations. It can be seen that both the input signal and also the unknown system are estimated to a very good degree when compared to their true values.

As another example, let us consider the unknown system to be identified to be a first order RC circuit with $\gamma = 1/RC = 2$ where $R = 100k\Omega$ and $C = 5\mu F$. Figure 4.9 shows the results for the above example after 20 iterations. It can be seen that both the input signal and also the unknown system are estimated to a very good degree when compared to their true values.

4.4 Summary

In this chapter we proposed our novel algorithms for system identification problem, based on the finite rate of innovation sampling theories. The novelty of this chapter was divided into two section, where in the first section, we showed that by having access to the input signal, system identification problem could be solved at low sampling rates by employing the multichannel sampling setup and exponential moments. Then, in the second section, we considered the problem of blind system identification. We showed that a recursive method could be utilized to estimate both the input FRI signal and also the unknown system when we only have access to the output samples.

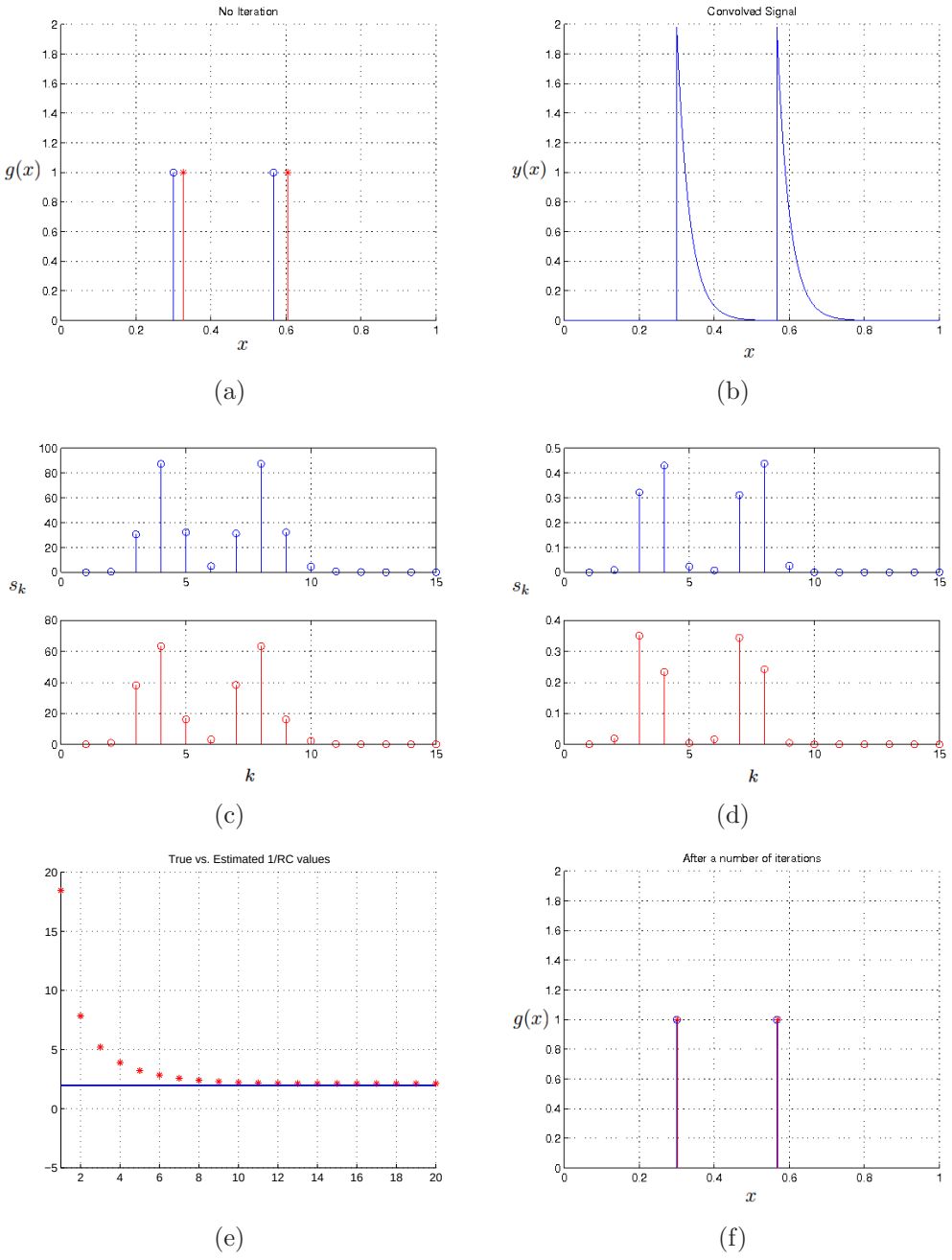


Figure 4.10: Simultaneous estimation of an input sparse signal with a first order RC circuit as the unknown function $\psi(x)$. (a) The input signal with 2 Diracs (circle) along with the immediate estimate of the signal (star) with no iterations. (b) Input signal convolved with the function $\psi(x)$. (c) The real (blue) and imaginary (red) samples with no iteration. (d) The real (blue) and imaginary (red) samples after 20 iterations. (e) True vs. estimated values of the parameter γ after 20 iterations. (f) True vs. estimated version of the input signal after 20 iterations.

Chapter 5

Multichannel Sampling of Multidimensional FRI Signals

5.1 Introduction

In Chapter 3, we presented a possible extension of sampling one-dimensional FRI signals to the case of multichannel acquisition systems. In the multichannel sampling setup presented, we assumed that the input signal observed by the system and also the channel parameters introduced within each channel (i.e. delay and gain parameters), have a 1-D structure. In this chapter, we aim to extend the framework presented in Chapter 3 to the multidimensional case, where both the input signal and also the channel parameters have a two-dimensional structure. The motivation of this extension comes from the fact that in applications such as acquisition of images in a multi-camera system, the estimation of the 2-D input signal as well as the 2-D channel parameters is required. Therefore, given the practical importance of such setups, it is essential to analyse the case of multidimensional FRI signals. The material of this chapter has been in part published in [4, 2, 3, 1].

The contribution of this chapter is two-fold; first, we present a novel approach for sampling and perfectly reconstructing 2-D signals with parametric structure or namely 2-D signals with FRI. The considered signals are bi-level polygons and set of

2-D Diracs and it will be shown that with the use of ACMP method (Algebraically Coupled Matrix Pencils), projection-slice theorem, Radon projections and exponential splines as sampling kernels, such signals can be perfectly reconstructed from their samples. Then, we consider the multichannel sampling of such 2-D signals and demonstrate that a simultaneous channel synchronization and signal reconstruction is possible. For the channel synchronization stage, we assume that the channel parameters are unknown geometric transformations such as translation, scaling, shearing and rotation.

The organization of this chapter is as follows: In the next section, we will introduce the multidimensional sampling framework for FRI signals which will include the definition of 2-D FRI signals, sampling setup used and the properties of the sampling kernels involved. In Section 5.3 we will introduce our novel algorithms for sampling and perfectly reconstructing set of 2-D Diracs by employing the ACMP method and bi-level polygons by utilizing Radon transformation and its relationship with the Fourier transform. In Section 5.4, we will extend the multichannel sampling setup presented in Chapter 3, to the case of multichannel sampling of 2-D FRI signals and present our novel algorithm for signal and channel estimation under simple 2-D translations and also affine transformations. The summary of the chapter will be given in Section 5.5.

5.2 Multidimensional Sampling Framework

5.2.1 2-D Signals with Finite Rate of Innovation

The definition of 2-D FRI signals is very similar to the 1-D case described in Chapter 2. Let us consider a 2-D signal with the following form:

$$g(x,y) = \sum_{r=0}^{N} \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} \gamma_{j,k,r}\ \phi_r(x - x_j, y - y_k). \tag{5.1}$$

The degrees of freedom of the signal $g(x, y)$ are the shifts x_j, y_k and the coefficients $\gamma_{j,k,r}$, assuming that the set of functions $\phi_r(x, y)$ are known. If we introduce a counting function which counts the number of free parameters of the signal $g(x, y)$ over the window of size (L_x, L_y), then the rate of innovation of the signal $g(x, y)$ is:

$$\rho = \lim_{L_x, L_y \to \infty} \frac{1}{L_x L_y} C_g \left[(-\tfrac{L_x}{2}, \tfrac{L_x}{2}), (-\tfrac{L_y}{2}, \tfrac{L_y}{2}) \right]. \tag{5.2}$$

A set of 2-D Diracs, bi-level polygons and classes of algebraic curves (ellipses, cardioids and lemniscates) are all examples of 2-D signals with finite rate of innovation. Also, as in the 1-D case, all two dimensional bandlimited signals could be defined with the above definition.

5.2.2 Sampling Setup

Figure 5.1 shows a typical sampling setup used for sampling 2-D FRI signals. In the figure, $g(x, y)$ represents the input FRI signal, $h(x, y)$ the impulse response of the acquisition device, $\phi(x, y)$ the sampling kernel, $g_s(x, y)$ the sampled version of the input signal, $s_{j,k}$ the samples and T_x, T_y are the sampling intervals along the Cartesian dimensions respectively. For the sampling setup shown in Figure 5.1, the

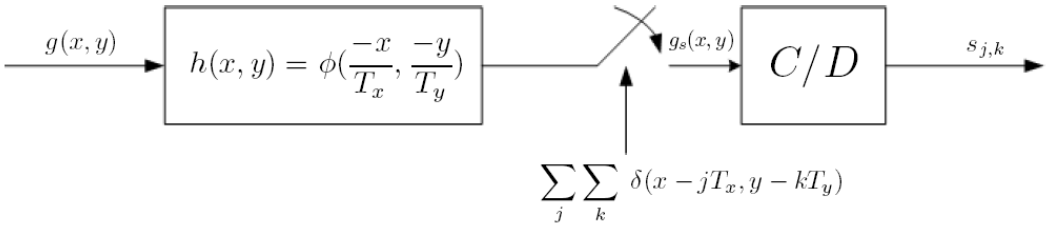


Figure 5.1: A typical sampling setup for 2-D FRI signals. Here, $g(x, y)$ represents the input FRI signal, $\phi(x, y)$ the sampling kernel, $g_s(x, y)$ the sampled version of the input signal, $s_{j,k}$ the samples and T_x, T_y are the sampling intervals along the Cartesian dimensions respectively.

samples $s_{j,k}$ are given by:

$$s_{j,k} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)\ \phi(\frac{x}{T_x} - j, \frac{y}{T_y} - k)\ dx\ dy \tag{5.3}$$

$$= \langle g(x,y), \phi(\frac{x}{T_x} - j, \frac{y}{T_y} - k)\rangle, \tag{5.4}$$

where the kernel $\phi(x,y)$ is the time-reversed version of the impulse response of the acquisition device. Given the samples from the setup described above, we want to see under what conditions, perfect reconstruction of the signal $g(x,y)$ can be obtained from the samples $s_{j,k}$. This includes the type of sampling kernels that can be employed and the methods of reconstruction.

5.2.3 Multidimensional Sampling Kernels

The sampling kernels that we consider in the setup shown in Figure 5.1 are given by the tensor product of two 1-D functions, that is: $\phi(x,y) = \phi(x)\phi(y)$. For the case of multidimensional polynomial reproducing kernels, if two functions $\phi(x)$ and $\phi(y)$ can reproduce the polynomials x^m and y^n along both the Cartesian axes, then the resulting 2-D kernel can reproduce polynomials along both dimensions, specifically:

$$\sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} c_{j,k}^{m,n} \phi(x-j, y-k) = x^m y^n, \tag{5.5}$$

for a proper choice of coefficients $c_{j,k}^{m,n}$ with $m = 0, 1, \ldots, M$ and $n = 0, 1, \ldots, N$. Two-dimensional orthogonal Daubechies scaling functions [20, 83] and two-dimensional biorthogonal B-splines [77, 78, 73], both satisfy the above property. A function $\beta^{P,Q}(x,y)$ with Fourier transform:

$$\hat{\beta}^{P,Q}(j\omega_x, j\omega_y) = \prod_{m=0}^{P} \prod_{n=0}^{Q} \left(\frac{1 - e^{-j\omega_x}}{j\omega_x}\right) \left(\frac{1 - e^{-j\omega_y}}{j\omega_y}\right), \tag{5.6}$$

is called a B-spline of order $(P+1) \times (Q+1)$. The resulting spline has compact support and can reproduce any polynomial in the subspace spanned by $\{1, x, \ldots, x^P y^Q\}$.

For the case of multidimensional exponential reproducing kernels, if two functions $\phi(x)$ and $\phi(y)$ can reproduce real or complex exponentials along both the Cartesian axes, then the resulting 2-D kernel can reproduce exponentials along both dimensions, specifically:

$$\sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} c_{j,k}^{m,n}\phi(x-j,y-k) = e^{\alpha_m x}e^{\beta_n y}, \tag{5.7}$$

for a proper choice of coefficients $c_{j,k}^{m,n}$. The choice of the exponents is restricted to $\alpha_m = \alpha_0 + m\lambda_1$ and $\beta_n = \beta_0 + n\lambda_2$ which is done to allow the use of specific reconstruction techniques, and are described later on. E-splines [79] satisfy the property defined in Equation (5.7) where a function $\beta_{\vec{\alpha},\vec{\beta}}(x,y)$ with Fourier transform:

$$\hat{\beta}_{\vec{\alpha},\vec{\beta}}(j\omega_x, j\omega_y) = \prod_{m=0}^{P}\prod_{n=0}^{Q}\left(\frac{1-e^{\alpha_m - j\omega_x}}{j\omega_x - \alpha_m}\right)\left(\frac{1-e^{\beta_n - j\omega_y}}{j\omega_y - \beta_n}\right), \tag{5.8}$$

is called an E-spline of order $(P+1)\times(Q+1)$ where $\vec{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_P)$ and $\vec{\beta} = (\beta_0, \beta_1, \ldots, \beta_Q)$ can be real or complex. The resulting spline has compact support and can reproduce any exponential in the subspace spanned by $\{e^{\alpha_0 x}e^{\beta_0 y}, e^{\alpha_1 x}e^{\beta_0 y}, \ldots, e^{\alpha_P x}e^{\beta_Q y}\}$.

Interestingly, E-splines can be regarded as a generalized version of B-splines. The reason is that, if we set the α_m and β_n parameters to zero, that is $\vec{\alpha} = \vec{0}$ and $\vec{\beta} = \vec{0}$, then we obtain a 2-D B-spline of order $(P+1)\times(Q+1)$. Moreover, E-splines can reproduce a combination of polynomials and exponentials. To illustrate this, let us assume that the set of parameters α_m and β_n are $\vec{\alpha} = (0,0,0,\alpha_3,\alpha_4,\ldots,\alpha_P)$ and $\vec{\beta} = (0,0,0,\beta_3,\beta_4,\ldots,\beta_Q)$ respectively. Then the resulting spline can reproduce

both polynomials and exponentials:

$$\sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} c_{j,k}^{m,n}\phi(x-j,y-k) = \begin{cases} x^m\ y^n & \text{for } m=0,1,2, \text{ and } n=0,1,2 \\ x^m\ e^{\beta_n y} & \text{for } m=0,1,2, \text{ and } n=3,4,\ldots,Q \\ e^{\alpha_m x}\ y^n & \text{for } m=3,4,\ldots,P, \text{ and } n=0,1,2 \\ e^{\alpha_m x}\ e^{\beta_n y} & \text{for } m=3,4,\ldots,P, \text{ and } n=3,4,\ldots,Q. \end{cases} \tag{5.9}$$

The use of this feature will be explained later on.

5.2.4 Geometric and Exponential Moments

As in the 1-D case, the geometric or exponential moments can be estimated to retrieve the degrees of freedom of $g(x,y)$. In order to obtain the geometric or exponential moments from the samples, let us consider $\tau_{m,n}$ to be:

$$\tau_{m,n} = \sum_j \sum_k\ c_{j,k}^{m,n}\ s_{j,k}, \tag{5.10}$$

where $c_{j,k}^{m,n}$ are the suitable coefficients used in Equation (5.5) or (5.7). By expanding the samples $s_{j,k}$ we have (we have assumed that $T_x = T_y = 1$):

$$\tau_{m,n} = \sum_j \sum_k\ c_{j,k}^{m,n}\ \langle g(x,y), \phi(x-j,y-k)\rangle \tag{5.11}$$

$$= \langle g(x,y), \sum_j \sum_k\ c_{j,k}^{m,n}\ \phi(x-j,y-k)\rangle. \tag{5.12}$$

Assuming that a polynomial reproducing kernel is used as the sampling kernel $\phi(x,y)$, then by substituting Equation (5.5) into the above equation, the 2-D geo-

metric moments of the signal are obtained:

$$\tau_{m,n} = \langle g(x,y), \sum_j \sum_k \; c_{j,k}^{m,n} \; \phi(x-j, y-k) \rangle \tag{5.13}$$

$$= \langle g(x,y), x^m \; y^n \rangle \tag{5.14}$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) \; x^m \; y^n \; dx \; dy. \tag{5.15}$$

Likewise, if an exponential reproducing kernel is used as the sampling kernel, then by substituting Equation (5.7), the exponential moments of the signal are obtained:

$$\tau_{m,n} = \langle g(x,y), \sum_j \sum_k \; c_{j,k}^{m,n} \; \phi(x-j, y-k) \rangle \tag{5.16}$$

$$= \langle g(x,y), e^{\alpha_m x} e^{\beta_n y} \rangle \tag{5.17}$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) \; e^{\alpha_m x} e^{\beta_n y} \; dx \; dy. \tag{5.18}$$

In the case of purely imaginary E-splines with $\alpha_m = jm\lambda_1$ and $\beta_n = jn\lambda_2$, the Fourier transform of the signal $g(x,y)$ at (α_m, β_n) are obtained from the exponential moments, that is:

$$\tau_{m,n} = \hat{g}(\alpha_m, \beta_n). \tag{5.19}$$

Here, $\hat{g}(j\omega_x, j\omega_y)$ represents the Fourier transform of the signal $g(x,y)$.

5.3 Reconstruction Techniques

The problem of perfectly reconstructing 2-D FRI signals from their samples is more involved and does not allow direct extension of the 1-D results. Recently, extensions to the multidimensional case were considered by Maravic et al. [48] and Shukla et al. [69]. Maravic et al. considered 2-D FRI signals, such as 2-D set of Diracs and bi-level polygons and used the sinc and Gaussian sampling kernels to sample and perfectly reconstruct such signals. Shukla et al. proposed algorithms, from the theory of complex moments [21, 22, 54], for sampling the same 2-D signals but with the use

of B-splines as the sampling kernel. As we previously mentioned, the sinc and the Gaussian sampling kernels have infinite support and are not physically realizable. Moreover, such kernels make the reconstruction algorithm unstable. In the following sections we present our novel algorithms for sampling and perfectly reconstructing 2-D set of Diracs and bi-level polygons using E-spline sampling kernels.

5.3.1 A Sampling Theorem for 2-D Diracs

Let us assume that a set of 2-D Diracs is sampled with the sampling setup shown in Figure 5.1. Assuming that there are K Diracs in the signal, such a signal can be written as:

$$g(x,y) = \sum_{k=1}^{K} a_k \, \delta(x - x_k, y - y_k), \tag{5.20}$$

where a_k are the amplitudes and (x_k, y_k) are the coordinates of the Diracs respectively. Since each Dirac has an amplitude and two coordinates, the signal has $3K$ degrees of freedom. From the samples obtained, we first calculate the exponential moments as follows:

$$\begin{aligned}
\tau_{m,n} &= \sum_j \sum_k c_{j,k}^{m,n} s_{j,k} && (5.21)\\
&= \sum_j \sum_k c_{j,k}^{m,n} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) \, \phi(x-j, y-k) \, dx \, dy && (5.22)\\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) \, e^{\alpha_m x} e^{\beta_n y} \, dx \, dy && (5.23)\\
&= \sum_{k=1}^{K} a_k \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(x - x_k, y - y_k) \, e^{\alpha_m x} e^{\beta_n y} \, dx \, dy && (5.24)\\
&= \sum_{k=1}^{K} a_k \, e^{\alpha_m x_k} e^{\beta_n y_k}. && (5.25)
\end{aligned}$$

We can see that the derived exponential moments $\tau_{m,n}$ has a 2-D power-sum series form. Our aim is to estimate the parameters (a_k, x_k, y_k) from $\tau_{m,n}$. At first sight, one might suggest an extension of the annihilating filter method for 1-D signals described in Chapter 2 to this scenario. However, this extension fails because the

relation $\tau_{m,n} * h_{m,n} = 0$, where $h_{m,n}$ is the annihilating filter, is not unique for the distinct locations (x_k, y_k) and thus there exists no unique solution to $\tau_{m,n} * h_{m,n} = 0$.

Another way to tackle this problem is by setting the indices m and n to zero one at a time and applying the 1-D annihilating filter method on both sets to find the values of x_k and y_k separately. There are two problems with this approach: first, the estimated locations have to be paired and this is a combinatorial problem and second, in the case of 2-D Diracs with a common coordinate, the annihilating filter method is unable to find the multiple poles, because of having non-unique filter coefficients.

This is indeed a spectral estimation problem and among the earliest techniques that addressed this problem was the matrix enhancement and matrix pencil (MEMP) algorithm by Hua [36]. For the case of common coordinates problem, Hua solves the rank deficiency problem by introducing an enhanced matrix of the original data matrix. In this way, a partitioned and stacked Hankel matrix of the original data matrix is constructed in a way such that the full-rank property of the original matrix is restored. For the pairing problem, an unattractive combinatorial approach is suggested, trying all the possible combinations to find the correct pairing. The combinatorial approach is computationally expensive and not efficient. The ACMP (Algebraically Coupled Matrix Pencils) method by F. Vanpoucke et al. [82] however, introduces a new technique to find the correct pairings by simultaneously solving two algebraically related generalized eigenvalue equations. As the matrix enhancement approach in [36] is not compatible with the algebraic pairing technique, an alternative rank restoration technique is introduced. We will now briefly outline the ACMP algorithm, but more detailed discussions can be found in [82].

Outline of the ACMP Method

Given $\alpha_m = \alpha_0 + m\lambda_1$ and $\beta_n = \beta_0 + n\lambda_2$, the obtained measurements $\tau_{m,n}$, which consists of a sum of K exponentials (complex or real) with unknown coordinate

pairs x_k and y_k, and amplitudes a_k, can be rewritten as:

$$\tau_{m,n} = \sum_{k=1}^{K} a_k \, e^{\alpha_m x_k} \, e^{\beta_n y_k} = \sum_{k=1}^{K} \widehat{a_k} \varphi_k^m \, \psi_k^n, \tag{5.26}$$

where $\widehat{a_k} = a_k e^{\alpha_0 x_k} e^{\beta_0 y_k}$, $\varphi_k = e^{\lambda_1 x_k}$ and $\psi_k = e^{\lambda_2 y_k}$. Let the $K(K+1) \times K(K+1)$ enhanced matrix $\mathbf{J}$ be defined as:

$$\mathbf{J} = \begin{bmatrix} \mathbf{H}^{(1,1)} & \mathbf{H}^{(2,1)} & \dots & \mathbf{H}^{(K,1)} \\ \mathbf{H}^{(1,2)} & \mathbf{H}^{(2,2)} & \dots & \mathbf{H}^{(K,2)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}^{(1,K)} & \mathbf{H}^{(2,K)} & \dots & \mathbf{H}^{(K,K)} \end{bmatrix}, \tag{5.27}$$

where each block matrix $\mathbf{H}^{(l,k)}$ of size $K \times K$ is given by:

$$\mathbf{H}^{(l,k)} = \tau_{l:K+l,k:K+k}. \tag{5.28}$$

For the construction of matrix $\mathbf{J}$, at least $2K \times 2K$ data points are required. If this condition is met, this new enhanced matrix can be decomposed as follows:

$$\mathbf{J} = \mathbf{X}' \mathbf{A} \mathbf{Y}'^{T}, \tag{5.29}$$

where

$$\mathbf{X}' = \begin{bmatrix} \mathbf{X}_{K+1}^T & \mathbf{\Psi}\mathbf{X}_{K+1}^T & \mathbf{\Psi}^2\mathbf{X}_{K+1}^T & \dots & \mathbf{\Psi}^{K-1}\mathbf{X}_{K+1}^T \end{bmatrix}^T \tag{5.30}$$

and,

$$\mathbf{Y}' = \begin{bmatrix} \mathbf{Y}_{K+1}^T & \mathbf{\Phi}\mathbf{Y}_{K+1}^T & \mathbf{\Phi}^2\mathbf{Y}_{K+1}^T & \dots & \mathbf{\Phi}^{K-1}\mathbf{Y}_{K+1}^T \end{bmatrix}^T. \tag{5.31}$$

Here, the matrices $\mathbf{X}'$ and $\mathbf{Y}'$ are both of sizes $K(K+1) \times K$. Furthermore, $\mathbf{\Phi}$, $\mathbf{\Psi}$ and $\mathbf{A}$ are all $K \times K$ diagonal matrices with parameters $\{\varphi_1, \varphi_2, \dots, \varphi_K\}$, $\{\psi_1, \psi_2, \dots, \psi_K\}$ and $\{\widehat{a_1}, \widehat{a_2}, \dots, \widehat{a_K}\}$ along the diagonals respectively. Finally, the

matrices $\mathbf{X}_{K+1}$ and $\mathbf{Y}_{K+1}$ are given by:

$$\mathbf{X}_{K+1} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \varphi_1 & \varphi_2 & \cdots & \varphi_K \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_1^K & \varphi_2^K & \dots & \varphi_K^K \end{bmatrix} \quad \mathbf{Y}_{K+1} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \psi_1 & \psi_2 & \dots & \psi_K \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1^K & \psi_2^K & \dots & \psi_K^K \end{bmatrix}. \tag{5.32}$$

The matrices $\mathbf{X}'$ and $\mathbf{Y}'$ have a Vandermonde structure and when the matrix $\mathbf{J}$ is at least of size $K(K+1) \times K(K+1)$, they are both full-rank . This matrix enhancement technique restores the full-rank property of the original matrix for the case of common coordinates problem [82]. Now, the four sub-matrices $\mathbf{J}_{tl}$, $\mathbf{J}_{tr}$, $\mathbf{J}_{bl}$ and $\mathbf{J}_{br}$ of matrix $\mathbf{J}$ are constructed, which correspond to the omission of the first and last rows and columns on each block of the matrix $\mathbf{J}$:

$$\mathbf{J}_{tl} = \underline{\mathbf{J}|} = \underline{\mathbf{X}'}\mathbf{A}\underline{\mathbf{Y}'}^T \tag{5.33}$$

$$\mathbf{J}_{tr} = \underline{|\mathbf{J}} = \underline{\mathbf{X}'}\mathbf{A}\overline{\mathbf{Y}'}^T \tag{5.34}$$

$$\mathbf{J}_{bl} = \overline{\mathbf{J}|} = \overline{\mathbf{X}'}\mathbf{A}\underline{\mathbf{Y}'}^T \tag{5.35}$$

$$\mathbf{J}_{br} = \overline{|\mathbf{J}} = \overline{\mathbf{X}'}\mathbf{A}\overline{\mathbf{Y}'}^T. \tag{5.36}$$

Here, over-line indicates the omission of the first row, under-line indicates the omission of the last row, left-line indicates the omission of the first column and right-line indicates the omission of the last column of the matrices. Because of the Vandermonde structure of the matrices $\mathbf{X}'$ and $\mathbf{Y}'$, it follows that:

$$\mathbf{J}_{tr} = \underline{\mathbf{X}'}\mathbf{A}\mathbf{\Psi}\underline{\mathbf{Y}'}^T \tag{5.37}$$

$$\mathbf{J}_{bl} = \underline{\mathbf{X}'}\mathbf{\Phi}\mathbf{A}\underline{\mathbf{Y}'}^T \tag{5.38}$$

$$\mathbf{J}_{br} = \underline{\mathbf{X}'}\mathbf{\Phi}\mathbf{A}\mathbf{\Psi}\underline{\mathbf{Y}'}^T. \tag{5.39}$$

From the matrices described above we can obtain two matrix pencils $\mathbf{J}_{tr} - \mu\mathbf{J}_{tl}$ and $\mathbf{J}_{bl} - \lambda\mathbf{J}_{tl}$. The ACMP method then operates as follows: First the SVD of $\mathbf{J}_{tl}$ is

computed:

$$\mathbf{J}_{tl} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H, \tag{5.40}$$

where H is Hermitian operator. By multiplying $\mathbf{U}^H$ to the left hand side and multiplying $\mathbf{V}$ to the right hand side of the two matrix pencils defined above we obtain:

$$\mathbf{U}^H(\mathbf{J}_{tr} - \mu\mathbf{J}_{tl})\mathbf{V} = \mathbf{U}^H\underline{\mathbf{X}'}\mathbf{A}\mathbf{\Psi}\underline{\mathbf{Y}'}^T\mathbf{V} - \mu\mathbf{U}^H\underline{\mathbf{X}'}\mathbf{A}\underline{\mathbf{Y}'}^T\mathbf{V} \tag{5.41}$$

$$= \mathbf{F}\mathbf{\Psi}\mathbf{G} - \mu\mathbf{F}\mathbf{G} \tag{5.42}$$

$$= \mathbf{C}_{tr} - \mu\mathbf{C}_{tl}, \tag{5.43}$$

and,

$$\mathbf{U}^H(\mathbf{J}_{bl} - \lambda\mathbf{J}_{tl})\mathbf{V} = \mathbf{U}^H\underline{\mathbf{X}'}\mathbf{A}\mathbf{\Phi}\underline{\mathbf{Y}'}^T\mathbf{V} - \lambda\mathbf{U}^H\underline{\mathbf{X}'}\mathbf{A}\underline{\mathbf{Y}'}^T\mathbf{V} \tag{5.44}$$

$$= \mathbf{F}\mathbf{\Phi}\mathbf{G} - \lambda\mathbf{F}\mathbf{G} \tag{5.45}$$

$$= \mathbf{C}_{bl} - \lambda\mathbf{C}_{tl}, \tag{5.46}$$

where $\mathbf{F} = \mathbf{U}^H\underline{\mathbf{X}'}\mathbf{A}$, $\mathbf{G} = \underline{\mathbf{Y}'}^T\mathbf{V}$, $\mathbf{C}_{tr} = \mathbf{U}^H\mathbf{J}_{tr}\mathbf{V} = \mathbf{F}\mathbf{\Psi}\mathbf{G}$, $\mathbf{C}_{bl} = \mathbf{U}^H\mathbf{J}_{bl}\mathbf{V} = \mathbf{F}\mathbf{\Phi}\mathbf{G}$ and $\mathbf{C}_{tl} = \mathbf{U}^H\mathbf{J}_{tl}\mathbf{V} = \mathbf{F}\mathbf{G} = \mathbf{\Sigma}$. By applying Eigen-value-decomposition (EVD) on the new matrix pencils $\mathbf{C}_{tr} - \mu\mathbf{C}_{tl}$ and $\mathbf{C}_{bl} - \lambda\mathbf{C}_{tl}$, each of the poles φ_k and ψ_k are obtained:

$$eig(\mathbf{C}_{tl}^{-1}\mathbf{C}_{tr}) = eig(\mathbf{G}^{-1}\mathbf{F}^{-1}\mathbf{F}\mathbf{\Psi}\mathbf{G}) = eig(\mathbf{G}^{-1}\mathbf{\Psi}\mathbf{G}) = \mathbf{\Psi} \tag{5.47}$$

$$eig(\mathbf{C}_{tl}^{-1}\mathbf{C}_{bl}) = eig(\mathbf{G}^{-1}\mathbf{F}^{-1}\mathbf{F}\mathbf{\Phi}\mathbf{G}) = eig(\mathbf{G}^{-1}\mathbf{\Phi}\mathbf{G}) = \mathbf{\Phi}. \tag{5.48}$$

The identical transformation $\mathbf{G}$ on both equations guarantees that we have the correct pairing for the estimated φ_k and ψ_k values. The steps below present a pseudo code of the algorithm:

1. Construct the enhanced matrix $\mathbf{J}$ from the moments.

2. Construct the sub-matrices $\mathbf{J}_{tl}$, $\mathbf{J}_{tr}$ and $\mathbf{J}_{bl}$ from the matrix $\mathbf{J}$.

3. Compute the singular value decomposition (SVD) of the sub-matrix $\mathbf{J}_{tl}$, that is $\mathbf{J}_{tl} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$.

4. Generate the matrices $\mathbf{C}_{tl}$, $\mathbf{C}_{bl}$ and $\mathbf{C}_{tr}$ with the following equations: $\mathbf{C}_{tl} = \mathbf{\Sigma}$, $\mathbf{C}_{tr} = \mathbf{U}^H\mathbf{J}_{tr}\mathbf{V}$, $\mathbf{C}_{bl} = \mathbf{U}^H\mathbf{J}_{bl}\mathbf{V}$.

5. Apply the EVD to $\mathbf{C}_{tl}^{-1}\mathbf{C}_{tr}$ and $\mathbf{C}_{tl}^{-1}\mathbf{C}_{bl}$. This leads to the unknown poles since: $eig(\mathbf{C}_{tl}^{-1}\mathbf{C}_{tr}) = \mathbf{\Psi}$ and $eig(\mathbf{C}_{tl}^{-1}\mathbf{C}_{bl}) = \mathbf{\Phi}$.

As the exact values of the poles φ_k and ψ_k are found using the above method, the matrices $\mathbf{X}'$ and $\mathbf{Y}'$ can now be constructed to obtain the parameters $\widehat{a_k}$, using the following equation:

$$\mathbf{A} = (\mathbf{X}'^{\dagger})\mathbf{J}(\mathbf{Y}'^{T})^{\dagger}, \tag{5.49}$$

where † stands for pseudo-inverse. From the estimated parameters $\widehat{a_k}$ and the poles φ_k and ψ_k we can easily find the amplitudes a_k and coordinates x_k and y_k as follows:

$$x_k = \frac{ln(\varphi_k)}{\lambda_1}, \qquad y_k = \frac{ln(\psi_k)}{\lambda_2}, \qquad a_k = \frac{\widehat{a_k}}{e^{\alpha_0 x_k} e^{\beta_0 y_k}}. \tag{5.50}$$

As mentioned above, for a set of K 2-D Diracs, at least $2K \times 2K$ data points are required for the construction of the enhanced matrix. This means that the 2-D E-spline order $(P+1) \times (Q+1)$ needs to be at least $2K \times 2K$ in order to reproduce the $2K$ exponential moments along both x and y axes. We can now summarize the above discussion with the following proposition:

Proposition I - A set of K 2-D Diracs is uniquely determined from the samples $s_{j,k} = \langle g(x,y), \phi(\frac{x}{T_x} - j, \frac{y}{T_y} - k)\rangle$ provided that the sampling kernel $\phi(x,y)$, can reproduce exponentials with an order $2K$ along both the Cartesian axes.

As an example, let us consider the input signal to be a set of 4 2-D Diracs in a frame size of 256×256, shown in Figure 5.2(a). The signal is passed through the E-spline sampling kernel of order $2K \times 2K = 8 \times 8$ shown in Figure 5.2(b) and then sampled at the intervals $T_x = T_y = 16$, as shown in Figure 5.2(c). Given the samples, the exponential moments are calculated and the ACMP method is applied

to retrieve the Diracs, which are shown in Figure 5.2(d).

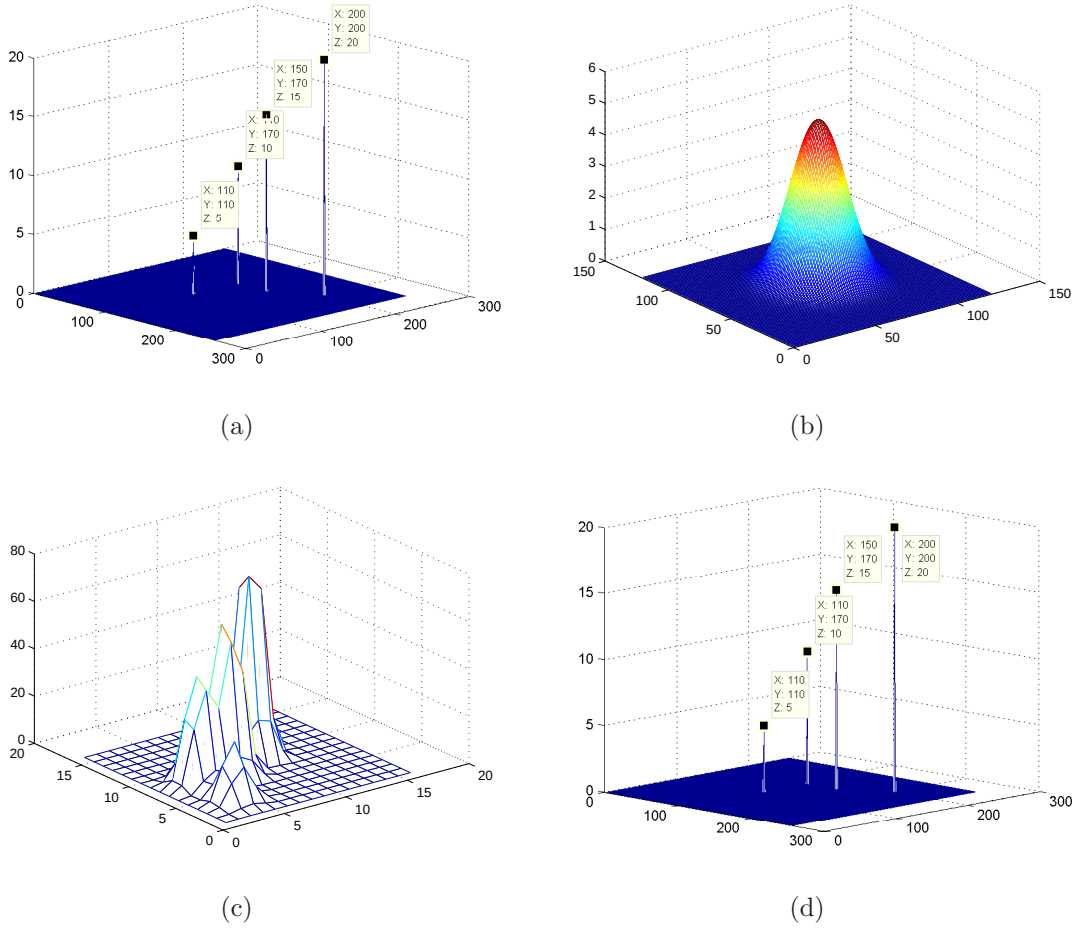


Figure 5.2: Sampling and reconstructing a set of 2-D Diracs using E-splines (a) The original input signal consisting of 4 Diracs in a frame size of 256×256. (b) The 2-D E-spline sampling kernel of order 8 along both dimensions. (c) The 16×16 samples of the input signal. (d) The reconstructed signal with the use of ACMP method.

Having presented a sampling theorem for set of 2-D Diracs, we now move on to our next considered signal, bi-level polygons.

5.3.2 A Sampling Theorem for Bi-level Polygons

Consider a non-intersecting, convex and bi-level K-sided polygon with vertices at points (x_k, y_k), $k = 1, 2, \ldots, K$. The described polygon can be uniquely specified by its K vertices and therefore has degrees of freedom equal to $2K$. Lee and Mittra [44]

derived a general formula for the Fourier transform of any K-sided bi-level polygon, where they showed that the Fourier transform is directly related to the location of the polygon's vertices (x_k, y_k):

$$\hat{g}(j\omega_x, j\omega_y) = \sum_{k=1}^{K} e^{j(\omega_x x_k + \omega_y y_k)} \frac{p_{k-1} - p_k}{(\omega_x + p_{k-1}\omega_y)(\omega_x + p_k\omega_y)}. \tag{5.51}$$

Here, p_k represent the gradients of the polygonal lines. The reader can refer to [44] for the derivation of this result. As $\tau_{m,n}$ are the exponential moments of the input signal and given $\alpha_m = jm\lambda$ and $\beta_n = jn\lambda$ for the case of purely imaginary E-splines (we have assumed that $\alpha_0 = \beta_0 = 0$ and $\lambda_1 = \lambda_2 = \lambda$), we can deduce the following:

$$\begin{aligned}
\tau_{m,n} &= \hat{g}(\alpha_m, \beta_n) && (5.52)\\
&= \sum_{k=1}^{K} e^{(\alpha_m x_k + \beta_n y_k)} \frac{p_{k-1} - p_k}{(m\lambda + p_{k-1}n\lambda)(m\lambda + p_k n\lambda)} && (5.53)\\
&= \sum_{k=1}^{K} a_{k,m,n}\, e^{\alpha_m x_k}\, e^{\beta_n y_k} && (5.54)\\
&= \sum_{k=1}^{K} a_{k,m,n} \varphi_k^m\, \psi_k^n, && (5.55)
\end{aligned}$$

where $a_{k,m,n} = \frac{p_{k-1}-p_k}{(m\lambda+p_{k-1}n\lambda)(m\lambda+p_k n\lambda)}$, $\varphi_k = e^{j\lambda x_k}$ and $\psi_k = e^{j\lambda y_k}$. The above equation closely follows the data model shown in the ACMP method, however, since the result for the Fourier transform has a frequency-varying amplitude, the ACMP method cannot be applied to find the locations (x_k, y_k). Having said that, by setting m and n indices to zero separately, we will end up with two equations in power-sum series form, which means that annihilating filter method can be applied to retrieve the parameters x_k and y_k separately. We already know that such a method has the problem of finding the correct pairings between the x_k and y_k coordinates, however, with the use of Radon transform [24] and the projection-slice theorem [32, 54] we can retrieve the locations of the vertices of bi-level polygons from their moments.

Projection-slice theorem states that the Fourier transform function $\hat{g}(j\omega_x, j\omega_y)$ evaluated along a line passing through the origin at an angle θ, is iden-

tical to the one dimensional Fourier transform of the Radon projection $R_g(x,\theta)$. In mathematical form:

$$\hat{g}(j\omega\cos(\theta), j\omega\sin(\theta)) = \hat{R}_g(j\omega,\theta). \tag{5.56}$$

In our setup, $\theta = \tan^{-1}(\frac{n}{m})$ with m and n being the indices of the moments and $\omega = \sqrt{(m^2+n^2)}$. With the help of this mapping, we can transform the Fourier coefficients of bi-level polygons, obtained from purely imaginary E-spline sampling kernel, to the Radon domain as follows:

$$\begin{aligned}
\hat{R}_g(j\omega,\theta) &= \tau_{\omega cos(\theta),\, \omega sin(\theta)} \\
&= \sum_{k=1}^{K} \frac{p_{k-1}-p_k}{(\omega cos(\theta)+p_{k-1}\omega sin(\theta))(\omega cos(\theta)+p_k\omega sin(\theta))}\; e^{\alpha_{\omega cos(\theta)}x_k+\beta_{\omega sin(\theta)}y_k} \\
&= \omega^{-2} \sum_{k=1}^{K} a_{k,\theta}\; e^{\alpha_{\omega cos(\theta)}x_k+\beta_{\omega sin(\theta)}},
\end{aligned}$$

where $a_{k,\theta} = \frac{p_{k-1}-p_k}{(cos(\theta)+p_{k-1}sin(\theta))(cos(\theta)+p_k sin(\theta))}$. Let us introduce $S(\omega,\theta) = \tau_{\omega cos(\theta),\, \omega sin(\theta)} \times \omega^2$. The above equation can now be rewritten as:

$$S(\omega,\theta) = \sum_{k=1}^{K} a_{k,\theta}\; e^{\alpha_{\omega cos(\theta)}x_k+\beta_{\omega sin(\theta)}y_k}. \tag{5.57}$$

At $\omega = 0$, $S(\omega,\theta) = 0$ so the minimum required spline order can be decreased by 1 as the first data sample is always zero. Since the angle θ is fixed for a given projection, then the mapped equation at different projections has a power-sum series form:

$$S_\theta(\omega) = \sum_{k=1}^{K} \widehat{a_k} e^{j\omega\lambda z_k} = \sum_{k=1}^{K} \widehat{a_k} u_k^{\omega}, \quad \omega \neq 0, \tag{5.58}$$

where $\widehat{a_k} = a_{k,\theta}$, $z_k = x_k cos(\theta) + y_k sin(\theta)$ and $u_k = e^{j\lambda z_k}$. By using the annihilating filter method we can retrieve the parameters z_k for each projection. For example, by setting $m = 0$ and $n = 0$ we have the projections at $\theta = 0$ and $\theta = 90$ degrees respectively, likewise if $m = n$ then we have the projection at 45 degrees. Further angles can be obtained by choosing different patterns such as $m = 2n$ or $n = 2m$.

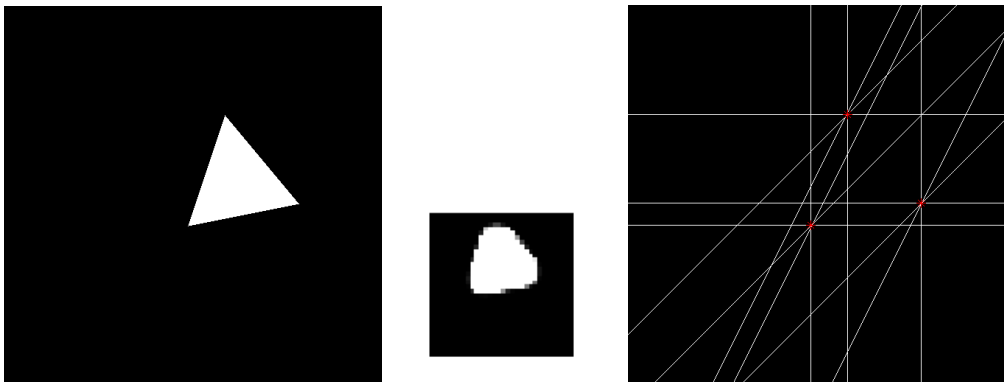
Each projection, as explained above, will result in a power-sum series form and annihilating filter method can be used to retrieve the parameters z_k which correspond to the sums of the vertices of the polygon in different directions. By back-projecting the parameters z_k according to their projection angle θ we are able to retrieve some information about the polygon's vertices. As any K-sided convex and bi-level polygon is completely specified by the location of its K vertices, it is known [47] that $K+1$ projections will entirely specify the vertices of the bi-level polygon, that is, points that have $K+1$ line intersections from the back-projections correspond to the K vertices.

To reconstruct a set of K 1-D Diracs from its samples, we need at least $2K$ data points, which means a minimum spline order of $2K$ is required. For bi-level polygons however, as the first data sample is always zero, a minimum number of $2K-1$ exponential moments at each projection angle is required. Thus, the minimum spline order required for a perfect reconstruction of a given K-sided bi-level polygon is $p.(2K-1)-(p-1)$ where p is the number needed in order to produce at least $K+1$ projections. For example for $K=3$ or 4 the correct value of is $p=2$ and for $K=5,6,7,8$ the value is $p=3$. The value of p can be found by inspection but it can be shown that p is $\mathcal{O}(K)$, thus, the order of the spline is $\mathcal{O}(K^2)$ along both dimensions. We can now summarize the above discussion with the following proposition:

Proposition II - A K-sided bi-level polygon is perfectly reconstructed from the samples $s_{j,k} = \langle g(x,y), \phi(\frac{x}{T_x} - j, \frac{y}{T_y} - k)\rangle$ provided that the sampling kernel $\phi(x,y)$, can reproduce exponentials with an order $p(2K-1)-(p-1)$ along both dimensions, where p is the number required in order to produce at least $K+1$ projections.

As an example, let us consider our input signal to be a bi-level triangle in a frame size of 256×256, shown in Figure 5.3(a). The signal is passed through the E-spline sampling kernel of order 9×9 and then sampled at intervals $T_x = T_y = 8$, as shown in Figure 5.3(b). Given the samples, the exponential moments are

calculated and $3+1=4$ projections are taken at the angles $0,\ 90,\ 45$ and $tan^{-1}(2)$. From the projections, all the z_n parameters are calculated and then normalized with $\sqrt{(m^2+n^2)}$ and finally back-projected. Figure 5.3(c) shows the reconstructed signal where the crosses are the actual vertices of the original signal. It can be seen that the locations of the vertices of the triangle have been perfectly recovered.



(a) (b) (c)

Figure 5.3: Sampling and reconstructing a bi-level polygon using E-splines (a) The original input bi-level triangle (b) The 32×32 samples of the input signal [not to scale] (c) The reconstructed vertices of the polygon where the crosses are the actual vertices of the original signal.

The 2-D order of the spline required is $\mathcal{O}(K^2) \times \mathcal{O}(K^2)$ and this suggests that as the complexity of the signal increases, a higher sampling rate will be required. In the next section we show how this can be avoided by using a multichannel acquisition system.

5.4 Multichannel Sampling Framework

In this section we investigate the problem of multichannel sampling of 2-D FRI signals. A model of the multichannel system observing multidimensional signals is shown in Figure 5.4 where the sampling kernels $\phi_1(x), \phi_2(x), \ldots, \phi_M(x)$ receive different geometrically transformed versions of the original signal $g(x, y)$. Here, the unknown transformation parameters are denoted by $\mathcal{T}_2, \ldots, \mathcal{T}_M$.

For the sake of clarity, we will first assume that the transformations or rather

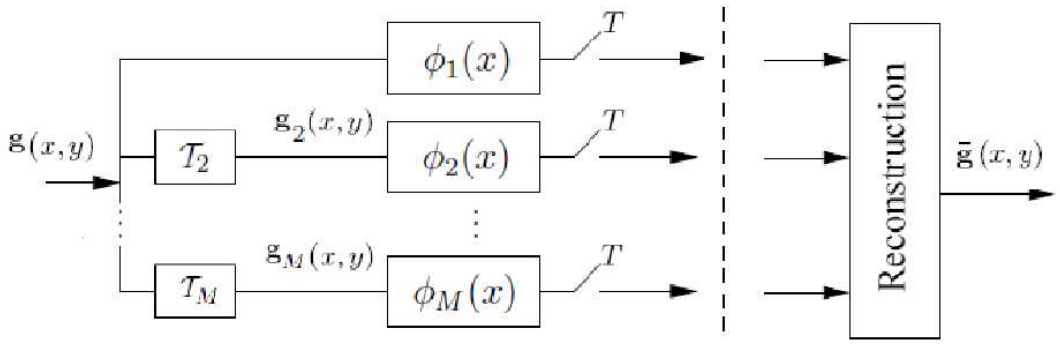


Figure 5.4: The proposed multichannel sampling setup for 2-D FRI signals. The sampling kernels $\phi_1(x), \phi_2(x), \ldots, \phi_M(x)$ receive different geometrically transformed versions of the original signal $g(x,y)$. Here, the unknown transformation parameters are denoted by $\mathcal{T}_2, \ldots, \mathcal{T}_M$.

the channel parameters $\mathcal{T}_2, \ldots, \mathcal{T}_M$ are 2-D translations $(\Delta_{xi}, \Delta_{yi})$, $i = 2, 3, \ldots, M$ and present an extension of the algorithm presented in Chapter 3 to this scenario. Then we consider the channel parameters to be affine transformations which include scaling, rotation, shearing and translations and present a novel algorithm for simultaneous estimation of the channel parameters as well as the input signal.

5.4.1 Channel Synchronization and Signal Reconstruction under 2-D Translations

In this section, we consider the geometric transformations in each channel to be 2-D translations $(\Delta x_i, \Delta y_i)$, only. For simplicity, we assume that the signal $g(x,y)$ is observed by a multichannel sampling system shown in Figure 5.4 with 2 channels and the sampling kernel in each channel is an E-spline of equal order $P \times Q$:

$$\hat{\phi}_1(j\omega_x, j\omega_y) = \prod_{m=0}^{P} \prod_{n=0}^{2Q-1} \left(\frac{1 - e^{\alpha_m - j\omega_x}}{j\omega_x - \alpha_m}\right) \left(\frac{1 - e^{\beta_n - j\omega_y}}{j\omega_y - \beta_n}\right) \tag{5.59}$$

$$\hat{\phi}_2(j\omega_x, j\omega_y) = \prod_{m=P}^{2P-1} \prod_{n=0}^{2Q-1} \left(\frac{1 - e^{\alpha_m - j\omega_x}}{j\omega_x - \alpha_m}\right) \left(\frac{1 - e^{\beta_n - j\omega_y}}{j\omega_y - \beta_n}\right). \tag{5.60}$$

As shown above, we have set one parameter to be common between the exponents of E-spline in one dimensions, specifically, both kernels can reproduce the exponentials $e^{\alpha_P x}e^{\beta_n y}$ for $n = 0, 1, \ldots, 2Q-1$. The samples of the i-th channel, with $i = 1, 2$, are given by:

$$s^i_{j,k} = \langle g(x - \Delta x_i, y - \Delta y_i), \phi(x - j, y - k)\rangle, \tag{5.61}$$

where $(\Delta x_1, \Delta y_1) = (0, 0)$. Our goal is to have a reconstruction technique that can perfectly retrieve the unknown translation parameters as well as the input signal. Give the samples, we can obtain the exponential moments $\tau^i_{m,n}$ for both channels:

$$\tau^1_{m,n} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y)\, e^{\alpha_m x}e^{\beta_n y}dx\, dy, \tag{5.62}$$

where $m = 0, 1, \ldots, P$ and $n = 0, 1, \ldots, 2Q-1$ and,

$$\tau^2_{m,n} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x - \Delta x_2, y - \Delta y_2)\, e^{\alpha_m x}e^{\beta_n y}dx\, dy = \tau^1_{m,n}e^{\alpha_m \Delta x_2}e^{\beta_n \Delta y_2}, \tag{5.63}$$

where $m = P, P+1, \ldots, 2P-1$ and $n = 0, 1, \ldots, 2Q-1$. We can see that there is a direct relationship between the exponential moments of the two signals. Now, by taking logarithms on both sides of Equation (5.63), we obtain a system of simple linear equations which we can solve for the translation parameters Δx_2 and Δy_2, that is:

$$\begin{pmatrix}\Delta x_2\\ \Delta y_2\end{pmatrix} = \begin{pmatrix}\alpha_P & \beta_0\\ \alpha_P & \beta_1\end{pmatrix}^{-1}\begin{pmatrix}\ln(\frac{\tau^2_{P,0}}{\tau^1_{P,0}})\\ \ln(\frac{\tau^2_{P,1}}{\tau^1_{P,1}})\end{pmatrix}. \tag{5.64}$$

Given the exact translation parameters, we can now estimate the exponential moments $\tau^1_{m,n}$, with $m = P, P+1, ..., 2P-1$ from $\tau^2_{m,n}$ as follows:

$$\tau^1_{m,n} = \tau^2_{m,n}e^{-\alpha_m \Delta x_2}e^{-\beta_n \Delta y_2}, \tag{5.65}$$

where $m = P, P+1, ..., 2P-1$ and $n = 0, 1, \ldots, 2Q-1$.

The result above, reveals that independently of $g(x, y)$ it is possible to estimate the channel parameters and synchronize the channels exactly from the obtained

measurements. Therefore, once the channels are synchronized and the entire exponential moments of the input signal are estimated, the corresponding reconstruction algorithm, as described in Section 5.3.1 and 5.3.2, is applied to perfectly recover the input signal. We now move on to the case where the transformation parameters are affine transformations.

5.4.2 Channel Synchronization and Signal Reconstruction under Affine Transformation

The above discussed method can be applied to any 2-D FRI signal as long as the transformation parameters are restricted to 2-D translations. This is because introducing more complicated geometric transformations such as rotation, would result in a non-linear relationship between the exponential moments of the different signals and therefore our introduced method cannot be applied. However, we will show that by employing a modified E-spline sampling kernel, the channel parameters and the input signal can still be perfectly estimated.

Let us assume that we have a two-channel sampling system and the input signal is a K-sided bi-level polygon (this method would also work for 2-D Diracs) with affine transformation introduced in the second channel. Affine transformation for 2-D signals is defined by a 6 degree parameter transformation and is defined as:

$$[x' \; y']^T = \mathbf{A} \cdot [x \; y]^T + [\Delta_x \; \Delta_y]^T, \tag{5.66}$$

where $[x \; y]$ represent the coordinates of the original signal, $[x' \; y']$ represent the affine transformed coordinates, $[\Delta_x \; \Delta_y]$ represents a 2-D translation and $\mathbf{A}$ is a 2×2 matrix. Matrix $\mathbf{A}$ is composed of a rotation parameter θ, 2-D scaling (X_{scale}, Y_{scale} and 2-D shearing (X_{shear}, Y_{shear}), and is given as follows:

$$\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \cdot \begin{bmatrix} X_{scale} & X_{shear} \\ Y_{shear} & Y_{scale} \end{bmatrix}. \tag{5.67}$$

The unknown parameters of the affine transformation, can be estimated with the use of geometric moments up to order 4 which can be obtained by a polynomial reproducing kernel [29, 45, 86]. As we mentioned earlier, E-splines are a generalized version of B-splines, thus, a combination of polynomials and exponentials from E-splines can be reproduced. The Fourier transform of the corresponding modified E-spline sampling kernel will then be as follows:

$$\hat{\beta}_{\vec{\alpha},\vec{\beta}}(\omega_x,\omega_y) = \left(\frac{1-e^{-j\omega_x}}{j\omega_x}\right)^3 \left(\frac{1-e^{-j\omega_y}}{j\omega_y}\right)^3 \prod_{m=0}^{P}\prod_{n=0}^{Q}\left(\frac{1-e^{\alpha_m-j\omega_x}}{j\omega_x-\alpha_m}\right)\left(\frac{1-e^{\beta_n-j\omega_y}}{j\omega_y-\beta_n}\right) \quad (5.68)$$

The produced spline is of order $(P+1+4)\times(Q+1+4)$, has compact support and can reproduce polynomials up to order 4 along both dimensions and can reproduce exponentials up to order $(P+1)\times(Q+1)$ (see Equation (5.9)). Therefore, the affine parameters can be estimated from the modified sampling kernel. Given the channel parameters, the next move is to recover the input signal.

As mentioned earlier, there is a non-linear relationship between the exponential moments of the different signals. However, as affine transformation is an invertible geometric transform, the following equation will also hold true:

$$[x\ y]^T = \mathbf{A}^{-1}\ ([x'\ y']^T - [\Delta_x\ \Delta_y]^T). \quad (5.69)$$

Given the equation above, we propose our algorithm for signal estimation as follows: First, we take two projections from the samples of the reference signal (i.e. first channel) and then calculate its corresponding back-projections. This will lead to K^2 solutions. Then, at the same angles, we take two projections from the samples of affine transformed signal and calculate its corresponding back-projections. This will also lead to K^2 solutions. We estimate the affine parameters from the geometric moments of the two signals and then apply the inverted affine parameters, as was shown in Equation (5.69), on each of the K^2 solutions of the affine transformed signal. We then map those points to the K^2 solutions of the reference signal. The mapped points that intersect with the K^2 solutions of the reference signal, will be

exactly K points and correspond to the vertices of the input bi-level polygon signal, as long as $A_{12} \neq 0$ and $A_{21} \neq 0$.

To show that there are exactly K intersected points, let us consider $g(x, y)$ to have 2 Diracs at the positions (x_1, y_1) and (x_2, y_2) and let $g^{'}(x, y)$ be an affine transformed version of $g(x, y)$. With 2 projections taken from the samples of each signal at the angles 0 and 90 degrees, we will end up with $K^2 = 4$ solutions for the signal $g(x, y)$ at the positions $(x_1, y_1), (x_1, y_2), (x_2, y_1), (x_2, y_2)$ and $K^2 = 4$ solutions for the transformed signal $g^{'}(x, y)$ at the positions $(x_1^{'}, y_1^{'}), (x_1^{'}, y_2^{'}), (x_2^{'}, y_1^{'}), (x_2^{'}, y_2^{'})$. Given that the signal $g^{'}(x, y)$ is an affine transformed version of the signal $g(x, y)$, we have:

$$\begin{cases} x_1^{'} = A_{11}x_1 + A_{12}y_1 + \Delta_x \\ y_1^{'} = A_{21}x_1 + A_{22}y_1 + \Delta_y \end{cases} \quad \text{and} \quad \begin{cases} x_2^{'} = A_{11}x_2 + A_{12}y_2 + \Delta_x \\ y_2^{'} = A_{21}x_2 + A_{22}y_2 + \Delta_y. \end{cases} \tag{5.70}$$

From the set of equations given above, we know that at least $K = 2$ points will match to the correct points. Now, by considering the above equations, and applying the affine transformation parameters on the rest of the solutions, that is at (x_1, y_2) and (x_2, y_1), we have:

$$\begin{cases} x_1^{'} \neq x_2^{'} \neq A_{11}x_1 + A_{12}y_2 + \Delta_x \\ y_1^{'} \neq y_2^{'} \neq A_{21}x_1 + A_{22}y_2 + \Delta_y \end{cases} \quad \text{and} \quad \begin{cases} x_2^{'} \neq x_1^{'} \neq A_{11}x_2 + A_{12}y_1 + \Delta_x \\ y_2^{'} \neq y_1^{'} \neq A_{21}x_2 + A_{22}y_1 + \Delta_y, \end{cases} \tag{5.71}$$

where $A_{12} \neq 0$ and $A_{21} \neq 0$. Therefore, exactly $K = 2$ points will match, which are the correct solutions. This will be true for any K as long as $A_{12}, A_{21} \neq 0$.

For the sake of clarity, the steps given below shows the procedures required to sample bi-level polygons in our proposed multichannel framework with $M = 2$, with affine transformation introduced on the second channel:

1. Choose the sampling kernel on both channels to be of equal order $(2K - 1 + 4) \times (2K - 1 + 4)$ where the number 4 corresponds to the spline order which is required for estimating the affine transformation parameters and the order

$(2K-1) \times (2K-1)$ corresponds to the minimum spline order required to produce at least 2 projections for each signal (normally at 0 and 90 degree).

2. Estimate the affine transformation parameters between the two-channels from their corresponding geometric moments.

3. Take two projections from the samples of both channels at the same angles and compute their corresponding back-projections. This will lead to K^2 solutions for each of the signals.

4. Invert the estimated affine parameters and apply them (see Equation (5.69)) on each of the K^2 solutions obtained from the second channel. Map those points to the reference signal.

5. The mapped points that intersect with the K^2 solutions of the reference signal correspond to the K vertices of the input bi-level polygon signal.

As an example, let us consider the input signal to be a 4-sided bi-level polygon shown in Figure 5.5(a), which is observed by a two-channel sampling system. The second channel observes an affine transformed version of the input signal with the following parameters: $(\Delta_x, \Delta_y) = (90, -90)$[1], $\theta = 50, (X_{scale}, Y_{scale}) = (1.05, 1.05)$ and $(X_{shear}, Y_{shear}) = (0.5, 0.2)$, shown in Figure 5.5(b). Both signals are sampled with an E-spline sampling kernel of order $2K - 1 + 4 = 11$, shown in Figures 5.5(c)(d)(e). Two projections are taken at the angles 0 and 90 degrees for both channels and this will lead to $K^2 = 16$ solutions, as shown in Figure 5.5(f)(g). The affine parameters, calculated from the geometric moments, are inverted and applied to each of the $K^2 = 16$ solutions of the affine transformed signal and then mapped to the solutions of the reference signal, as shown in Figure 5.5(h). As can be seen, only $K = 4$ points match with the $K^2 = 16$ solutions, which correspond to the true vertices of the input bi-level polygon.

[1]In pixels.

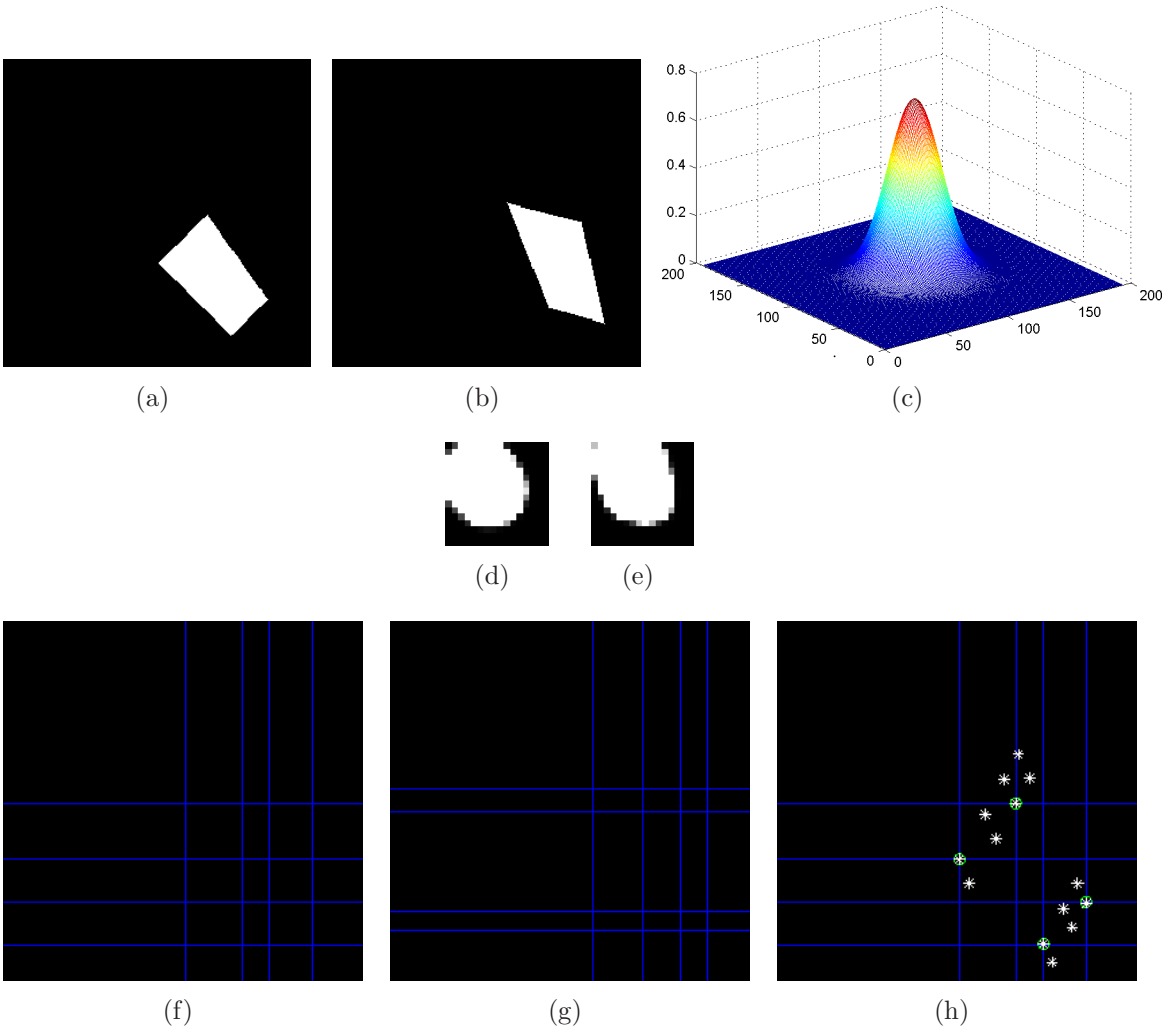


Figure 5.5: Multichannel sampling of bi-level polygons under affine transformation. (a) The reference signal in a frame data size of 256×256. (b) The affine transformed signal with $(\Delta_x, \Delta_y) = (90, -90), \theta = 50, (X_{scale}, Y_{scale}) = (1.05, 1.05)$ and $(X_{shear}, Y_{shear}) = (0.5, 0.2)$. (c) 2-D modified E-spline sampling kernel of order $2K - 1 + 4 = 11$. (d) & (e) The 16×16 samples of both signals. (f) & (g) K^2 solutions of both signals. (h) The reconstructed vertices of the reference signal. The green circles are the true vertices and the white asterisks are the mapped solutions.

5.5 Summary

In this chapter we showed that with the use of ACMP method, projection-slice theorem and Radon projections, multidimensional parametric signals such as set of 2-D Diracs and bi-level polygons can be sampled and perfectly reconstructed using E-splines as the sampling kernel. For the case of multichannel sampling scenario,

assuming that the geometric transformations are restricted to 2-D translations only, we showed that the different channels can be synchronized regardless of the input FRI signal. For the case of affine transformations however, we illustrated that, with the use of modified E-splines and also taking affine transforms' invertibility into account, we can simultaneously estimate the channel parameters and the input signal.

Chapter 6

Conclusion

In this thesis we presented a possible extension of the theory of sampling non-bandlimited 1-D and 2-D signals with finite rate of innovation to the case of multi-channel acquisition systems. What distinguishes our considered multichannel setup to other setups is that, we assume each channel receives a disrupted version of the observed input signal with some unknown parameters which could be delays and gains in the 1-D case or affine transformations in the 2-D case. We posed both the channel parameters estimation stage and the signal reconstruction stage as a parametric estimation problem and demonstrated that a simultaneous exact synchronization of the channels and reconstruction of the input 1-D and 2-D FRI signal is possible. Furthermore, for the case of noisy measurements, we showed in this thesis that by considering the Cramér-Rao bounds as well as the numerical simulations, the multichannel systems are more resilient to noise than the single-channel ones.

In the following section, we briefly summarize the contents of each chapter of this thesis.

6.1 Thesis Summary

In Chapter 2, we presented a background on the 1-D sampling framework of FRI signals. We described the different elements of the sampling setup used for sampling

FRI signals and showed how a set of 1-D Diracs can be sampled and perfectly reconstructed using both polynomial and exponential reproducing kernels. Then we considered the case of noisy scenario and presented some denoising techniques such as matrix pencil method and Cadzow's algorithm for the reconstruction of the innovation parameters from the noisy samples.

In Chapter 3, we presented a possible extension of the theory of sampling FRI signals to the case of multichannel acquisition systems. We showed that by synchronizing the different channels of the proposed multichannel sampling setup, one can estimate the unknown delays and gains introduced within the channels, regardless of the input FRI signal. For the case of noisy measurements, we showed by evaluating the CRB that, despite the fact that the unknown delays are required to be estimated for synchronization, the number of samples required can be reduced by the multichannel setup. Moreover, our numerical results with the multichannel sampling setup revealed that the performance of the reconstruction algorithm improves over the single-channel ones.

In Chapter 4, we proposed our novel algorithms for system identification problem, based on the finite rate of innovation sampling theory. The novelty of this chapter was divided into two section, where in the first section, we showed that by having access to the input signal, system identification problem could be solved at low sampling rates by employing the multichannel sampling setup and exponential moments. Then, in the second section, we considered the problem of blind system identification where we showed that a recursive method could be utilized to estimate both the input FRI signal and also the unknown system when we only have access to the output samples.

In Chapter 5, we first presented and discussed the sampling theory of multidimensional FRI signals. We showed that with the use of ACMP method, projection-slice theorem and Radon projections, multidimensional parametric signals such as set of 2-D Diracs and bi-level polygons can be sampled and perfectly reconstructed using exponential splines. For the case of multichannel sampling scenario, we first

assumed that the geometric transformations are restricted to 2-D translations only and showed that the different channels can be synchronized regardless of the input FRI signal. Then we considered the case of affine transformations as the channel parameters and illustrated that, with the use of modified E-splines we can simultaneously estimate the channel parameters as well as the input signal.

6.2 Future Research

There are many interesting while challenging open problems to look at in our research work and some of them are given below:

Numerical Simulations versus Cramér-Rao Bounds

In Chapter 3, we theoretically showed that the described multichannel sampling system is more resilient to noise than the single-channel ones. This was also confirmed with our numerical simulations. However, none of the denoising methods employed achieve the CRB and the gap between the CRB and the performance of the numerical simulations is quite significant. Although sampling FRI signals under the presence of noise has been extensively considered in [14], but we think that it will be interesting to see if more robust and reliable algorithms for sampling complex FRI signals under noisy conditions, such as adaptive hard-thresholding methods, can be introduced.

Convergence of the Proposed Blind System Identification Method

In Chapter 4, we mentioned that our empirical results show that by applying the proposed method for blind system identification recursively, the estimations converge to the actual input signal $g(x)$ and the unknown function $\psi(x)$. This was revealed to us by taking a number of simulations with different parameters. However, a more

subtle way to approach the convergence problem would be to provide a mathematical evidence that the proposed method is convergent.

Proposed Sampling Theorem for Bi-level Polygons

In Chapter 5, we presented a novel algorithm for sampling and perfectly reconstructing bi-level polygons using E-splines, where we showed that by using the projection-slice theorem, we can recover the vertices of bi-level polygons. Since for perfect reconstruction $K + 1$ projections are needed for a K-sided bi-level polygon, the order of E-spline used is $\mathcal{O}(N^2)$ which is quite redundant when compared to the degrees of freedom of the signal $\mathcal{O}(2K)$. It would be interesting to find more efficient ways of sampling and perfectly reconstructing bi-level polygons using E-splines.

The Potentials of E-splines

We have already mentioned that, E-splines are compact support splines that are practically implementable and tend to be more stable than other kernels. Moreover, from purely imaginary E-spline sampling kernels, the Fourier transform coefficients of an input signal can be obtained. Therefore, they have a great potential to be used as sampling kernels in various applications. For example, the model of neural signals are very similar to the model of E-splines convolved with Diracs and therefore E-splines can have vast number of applications in neuroscience. We have already considered E-splines for the problem of system identification in this thesis, however, we think that more and more applications could benefit from these stable and practically implementable sampling kernels.

Appendix A

Cramér-Rao Bound Derivation

Let us assume that we have an unbiased estimator $\hat{\mathbf{\Theta}}$ that is required to estimate a vector of K unknown parameters from noisy samples $\hat{s}_k$. We denote the vector of parameters by $\mathbf{\Theta}$ as follows:

$$\mathbf{\Theta} = (\Theta_1, \Theta_2, \ldots, \Theta_K)^T.$$

The measured noisy samples $\hat{s}_k$ depend on the unknown parameters $\mathbf{\Theta}$ and the added noise ϵ_k, that is:

$$\hat{s}_k = f(\mathbf{\Theta}, k) + \epsilon_k,$$

where we have assumed ϵ_k to be additive white Gaussian noise, independent of the samples s_k and the function $f(\mathbf{\Theta}, k)$ is a function that depends on the vector $\mathbf{\Theta}$ and the sample index k. Since we have an unbiased estimator, we have:

$$E(\hat{\mathbf{\Theta}}) = \mathbf{\Theta}.$$

The Cramér-Rao bound for any unbiased estimator $\hat{\mathbf{\Theta}}$ of the parameters $\mathbf{\Theta}$ gives the lowest achievable bound at a given noise level, and is given by the following expression:

$$var(\hat{\mathbf{\Theta}}) \geq \mathbf{I}^{-1}(\mathbf{\Theta}),$$

where $\mathbf{I}(\mathbf{\Theta})$ is the Fisher information matrix defined as:

$$\mathbf{I}(\mathbf{\Theta}) = E(\nabla l(\mathbf{\Theta}) \cdot \nabla l(\mathbf{\Theta})^T).$$

Here, $l(\mathbf{\Theta})$ is the log-likelihood function of $\mathbf{\Theta}$. The log-likelihood function of a set of parameters $\mathbf{\Theta}$ given the samples $\hat{s}_k$ is defined by the probability of the samples given the parameters $\mathbf{\Theta}$. Therefore, the log-likelihood function is given as follows:

$$l(\mathbf{\Theta}) = ln\ P(\hat{s}_0, \hat{s}_1, \ldots, \hat{s}_{N-1}|\mathbf{\Theta}),$$

where P is the probability distribution of the samples conditioned on the parameters $\mathbf{\Theta}$. Since ϵ_k is assumed to be white Gaussian with zero mean and variance σ^2, the probability distribution of ϵ_k is therefore given by [14, 26]:

$$p_\epsilon(\epsilon_k) = \frac{1}{\sqrt{2\pi\sigma^2}}\ exp\left(-\frac{\epsilon_k^2}{2\sigma^2}\right).$$

Given the probability distribution of the noise, we can deduce that:

$$\begin{aligned} p_{\hat{s}}(\hat{s}_k|\mathbf{\Theta}) &= p_\epsilon(g^{-1}(\hat{s}_k))\left|\frac{\partial g^{-1}(\hat{s}_k)}{\partial \hat{s}_k}\right| \\ &= \frac{1}{\sqrt{2\pi\sigma^2}}\ exp\left(-\frac{(\hat{s}_k - f(\mathbf{\Theta}, k))^2}{2\sigma^2}\right) \\ &= p_\epsilon(\hat{s}_k - f(\mathbf{\Theta}, k)), \end{aligned}$$

where $g(\epsilon_k) = \hat{s}_k = f(\mathbf{\Theta}, k) + \epsilon_k$. Since the noise samples are independent, the following can be deduced for the log-likelihood function:

$$\begin{aligned} l(\mathbf{\Theta}) &= ln\ P(\hat{s}_0, \hat{s}_1, \ldots, \hat{s}_{N-1}|\mathbf{\Theta}) \\ &= ln \prod_{k=0}^{N-1} p_{\hat{s}}(\hat{s}_k|\mathbf{\Theta}) \\ &= \sum_{k=0}^{N-1} ln\ p_{\hat{s}}(\hat{s}_k|\mathbf{\Theta}) \\ &= \sum_{k=0}^{N-1} ln\ p_{\epsilon}(\hat{s}_k - f(\mathbf{\Theta}, k)). \end{aligned}$$

To calculate the Fisher information matrix, we need to apply the divergence operator on the log-likelihood function. Therefore we compute the partial derivatives of $l(\mathbf{\Theta})$ with respect to the parameters $\mathbf{\Theta}$ as follows:

$$\begin{aligned} \frac{\partial l(\mathbf{\Theta})}{\partial \Theta_i} &= \frac{\partial l(\mathbf{\Theta})}{\partial \hat{s}_k - f(\mathbf{\Theta}, k)} \frac{\partial \hat{s}_k - f(\mathbf{\Theta}, k)}{\partial f(\mathbf{\Theta}, k)} \frac{\partial f(\mathbf{\Theta}, k)}{\partial \Theta_i} \\ &= -\sum_{k=0}^{N-1} \frac{\hat{p}_{\epsilon}}{p_{\epsilon}} (\hat{s}_k - f(\mathbf{\Theta}, k)) \frac{\partial f(\mathbf{\Theta}, k)}{\partial \Theta_i} \\ &= \sum_{k=0}^{N-1} \frac{\hat{s}_k - f(\mathbf{\Theta}, k)}{\sigma^2} \frac{\partial f(\mathbf{\Theta}, k)}{\partial \Theta_i} \\ &= \frac{1}{\sigma^2} \sum_{k=0}^{N-1} \epsilon_k \frac{\partial f(\mathbf{\Theta}, k)}{\partial \Theta_i}. \end{aligned}$$

Thus,

$$\nabla l(\mathbf{\Theta}) = \frac{1}{\sigma^2} \sum_{k=0}^{N-1} \epsilon_k \nabla f(\mathbf{\Theta}, k).$$

Since the noise samples are independents and uncorrelated, the Fisher information matrix can now be calculated as follows:

$$\begin{aligned}
\mathbf{I}(\mathbf{\Theta}) &= E(\nabla l(\mathbf{\Theta}) \cdot \nabla l(\mathbf{\Theta})^T) \\
&= E\left(\frac{1}{\sigma^4}\sum_{k=0}^{N-1}\sum_{j=0}^{N-1}\epsilon_k\epsilon_j \nabla f(\mathbf{\Theta},k)\nabla f(\mathbf{\Theta},k)^T\right) \\
&= \frac{1}{\sigma^4}\sum_{k=0}^{N-1}\sum_{j=0}^{N-1}\delta_{k,j}\nabla f(\mathbf{\Theta},k)\nabla f(\mathbf{\Theta},k)^T \\
&= \frac{1}{\sigma^2}\sum_{k=0}^{N-1}\nabla f(\mathbf{\Theta},k)\nabla f(\mathbf{\Theta},k)^T.
\end{aligned}$$

Having obtained the Fisher information matrix, the Cramér-Rao bound, as mentioned previously, is given by the inverse of the Fisher information matrix:

$$CRB(\mathbf{\Theta}) = \sigma^2\left(\sum_{k=0}^{N-1}\nabla f(\mathbf{\Theta},k)\cdot\nabla f(\mathbf{\Theta},k)^T\right)^{-1}.$$

The CRB equation derived above only holds true for real-valued functions. For the case of complex-valued E-spline sampling kernels, the samples are complex valued and the noisy samples are assumed to be corrupted by complex AWGN, defined as $\epsilon_k = u_k + jv_k$ where u and v are both real and independent random variables with zero mean and $\frac{\sigma^2}{2}$ variance. The joint probability distribution of u and v is given by:

$$\begin{aligned}
p_{u,v}(u_k, v_k) &= \frac{1}{\sqrt{2\pi\frac{\sigma^2}{2}}}\, exp\left(-\frac{u_k^2}{2\frac{\sigma^2}{2}}\right)\cdot\frac{1}{\sqrt{2\pi\frac{\sigma^2}{2}}}\, exp\left(-\frac{v_k^2}{2\frac{\sigma^2}{2}}\right) \\
&= \frac{1}{\pi\sigma^2}\, exp\left(-\frac{u_k^2+v_k^2}{\sigma^2}\right) \\
&= \frac{1}{\pi\sigma^2}\, exp\left(-\frac{|\tilde{\epsilon}_n|^2}{\sigma^2}\right) \\
&= p_{\tilde{\epsilon}}(\tilde{\epsilon}_n),
\end{aligned}$$

where $\tilde{\epsilon}_n = \sqrt{u_k^2 + v_k^2}$. Given the probability distribution of the noise, the log-

likelihood function $l(\mathbf{\Theta})$ will be:

$$l(\mathbf{\Theta}) = \sum_{k=0}^{N-1} ln\ p_{\tilde{\epsilon}}(\hat{s}_k - f(\mathbf{\Theta}, k)).$$

To calculate the Fisher information matrix, we need to apply the divergence operator on the log-likelihood function like before. Applying a similar procedure will yield:

$$\mathbf{I}(\mathbf{\Theta}) = \frac{2}{\sigma^2}\ Re\left(\sum_{k=0}^{N-1} \nabla f(\mathbf{\Theta}, k) \cdot \nabla f^*(\mathbf{\Theta}, k)\right),$$

where * stands for conjugate. The Cramér-Rao bound will therefore be as follows:

$$CRB(\mathbf{\Theta}) = \frac{\sigma^2}{2} Re\left(\sum_{k=0}^{N-1} \nabla f(\mathbf{\Theta}, k) \cdot \nabla f^*(\mathbf{\Theta}, k)\right)^{-1}.$$

Bibliography

[1] H. Akhondi Asl and P.L. Dragotti. A sampling theorem for bilevel polygons using E-splines. In *The 8th IMA International Conference on Mathematics in Signal Processing*, Royal Agriculture College, Cirencester, UK, December 2008.

[2] H. Akhondi Asl and P.L. Dragotti. Multichannel sampling of translated, rotated and scaled bilevel polygons using exponential splines. In *The 8th international conference on Sampling Theory and Applications*, Marseille, France, May 2009.

[3] H. Akhondi Asl and P.L. Dragotti. Single and multichannel sampling of bilevel polygons using exponential splines. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 3349–3352, Taipei, Taiwan, May 2009.

[4] H. Akhondi Asl and P.L. Dragotti. Multichannel sampling of multidimensional parametric signals. *To Appear in the Special Issue of Sampling Theory in Signal and Image Processing Journal*, December 2011.

[5] H. Akhondi Asl, P.L. Dragotti, and L. Baboulaz. Multichannel sampling of signals with finite rate of innovation. *IEEE Signal Processing Letter*, 17(8):762–765, August 2010.

[6] L. Baboulaz. *Feature Extraction for Image Super-resolution using Finite Rate of Innovation Principles*. PhD thesis, Imperial College London, UK, 2008.

[7] L. Baboulaz and P.L. Dragotti. Distributed acquisition and image super-resolution based on continuous moments from samples. In *IEEE International Conference on Image Processing (ICIP06)*, Atlanta (GA), October 2006.

[8] L. Baboulaz and P.L. Dragotti. Local feature extraction for image super-resolution. In *IEEE International Conference on Image Processing (ICIP07)*, September 2007.

[9] L. Baboulaz and P.L. Dragotti. Exact feature extraction using finite rate of innovation principles with an application to image super-resolution. *IEEE Transactions on Image Processing*, 18(2):281–298, February 2009.

[10] R. G. Baraniuk, E. Candès, R. Nowak, and M. Vetterli. Compressive sampling. *IEEE Signal Processing Magazine*, 25(2):12–13, March 2008.

[11] J. Berent, P.L. Dragotti, and T. Blu. Sampling piecewise sinusoidal signals with finite rate of innovation methods. *IEEE Transactions on Signal Processing*, 58(2):613–625, February 2010.

[12] W.C. Black and D.A. Hodges. Time interleaved converter arrays. *IEEE Journal of Solid-State Circuits*, 15(6):1022–1029, 1980.

[13] R.E. Blahut. *Theory and practice of error control codes*. Addison-Wesley, 1983.

[14] T. Blu, P.L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot. Sparse sampling of signal innovations: Theory, algorithms and performance bounds. *IEEE Signal Processing Magazine*, 25(2):31–40, March 2008.

[15] T. Blu and M. Unser. Approximation errors for quasi-interpolators and (multi-) wavelet expansions. *Applied and Computational Harmonic Analysis*, 6 (2):219–251, March 1999.

[16] J.A. Cadzow. Signal enhancement- A composite property mapping algorithm. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 36(1):49–62, January 1988.

[17] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principle: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, February 2006.

[18] V. Chaisinthop and P.L Dragotti. Distributed video coding based on sampling of signals with finite rate of innovation. In *Proc. of SPIE Conference on Wavelet Applications in Signal and Image Processing, Wavelets XII, San Diego, USA*, August 2007.

[19] Y. Chen, Y. Gu, and A.O. Hero-III. Sparse lms for system identification. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3125–3128, 2009.

[20] I. Daubechies. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, Philadelphia, USA, 1992.

[21] P.J. Davis. Triangle formulas in the complex plane. *Mathematics of Computation*, 18(88):569–577, October 1964.

[22] P.J. Davis. Plane regions determined by complex moments. *Journal of Approximation Theory*, 19(2):148–153, 1977.

[23] B. De Moor. The singular value decomposition and long and short spaces of noisy matrices. *IEEE Transactions on Signal Processing*, 41(9):2826–2838, September 1993.

[24] S.R. Deans. *The Radon transform and some of its applications*. Wiley New York, 1983.

[25] D.L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, April 2006.

[26] P.L. Dragotti and F. Homann. Sampling signals with finite rate of innovation in the presence of noise. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 2941–2944, April 2009.

[27] 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*, 55(5):1741–1757, May 2007.

[28] Y. Eldar. Compressed sensing of analog signals in shift-invariant spaces. *IEEE Transactions on Signal Processing*, 57(8):2986–2997, August 2009.

[29] J. Flusser and T. Suk. A moment-based approach to registration of images with affine geometric distortion. *IEEE Transactions on Geoscience and Remote Sensing*, 32(2):382–387, March 1994.

[30] N. Gehrig and P.L. Dragotti. Distributed sampling and compression of scenes with finite rate of innovation in camera sensor networks. In *Data Compression Conference (DCC'06)*, March 2006.

[31] Y. Hao, P. Marziliano, M. Vetterli, and T. Blu. Compression of ECG as a signal with finite rate of innovation. In *The 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 7564–7567. IEEE, January 2005.

[32] G.T. Herman. *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography*. Academic Press, New York, 1980.

[33] A. Hormati, O. Roy, Lu. Y.M., and M. Vetterli. Distributed sampling of correlated signals linked by sparse filtering: Theory and applications. *IEEE Transactions on Signal Processing*, 58(3):1095–1109, March 2010.

[34] Y. Hua and T.K. Sarkar. Matrix pencil method and its performance. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 2476–2479, April 1988.

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

[36] Y. Huy. Estimating two-dimensional frequencies by matrix enhancement and matrix pencil. *IEEE Transactions on Signal Processing*, 40(9):2267–2280, Sep. 1992.

[37] A.J. Jerri. The shannon sampling theorem- its various extensions and applications: A tutorial review. *Proceedings of the IEEE*, 65(11):1565–1596, November 1977.

[38] I. Jovanovic and B. Beferull-Lozano. Oversampled A/D conversion and error-rate dependence of nonbandlimited signals with finite rate of innovation. *IEEE Transactions on Signal Processing*, 54(6):2140–2154, June 2006.

[39] V.A. Kotel'nikov. On the transmission capacity of the ether and cables in electrical communications. In *Proceedings of the first All-Union Conference on the technological reconstruction of the communications sector and the development of low-current engineering. Moscow*, 1933.

[40] V.A. Kotelnikov. On the transmission capacity of the ether and wire in electrocommunications. *Modern Sampling Theory: Mathematics and Applications. Benedetto, J.J. and Ferreira, P. Birkhauser, Boston, MA*, 2000.

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

[42] J. Kusuma. *Economical sampling of parametric signals*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA 02139, USA, 2006.

[43] J. Kusuma and V.K. Goyal. On the accuracy and resolution of powersum-based sampling methods. *IEEE Transactions on Signal Processing*, 57(1):182–193, January 2009.

[44] S.W. Lee and R. Mittra. Fourier transform of a polygonal shape function and its application in electromagnetics. *IEEE Transactions on Antennas and Propagation*, 31(1):99–103, January 1983.

[45] H. Li, B. S. Manjunath, and S. K. Mitra. A contour-based approach to multisensor image registration. *IEEE Transactions on Image Processing*, 4(2):320–334, March 1995.

[46] I. Maravic. *Sampling methods for parametric non-bandlimited signals : extensions and applications*. PhD thesis, Audio-Visual Communcation Laboratory, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, 2004.

[47] I. Maravic and M. Vetterli. A sampling theorem for the radon transform of finite complexity objects. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pages 1197–1200, Orlando, Florida (USA), May 2002.

[48] I. Maravic and M. Vetterli. Exact sampling results for some classes of parametric nonbandlimited 2-D signals. *IEEE Transactions on Signal Processing*, 52(1):175–189, January 2004.

[49] I. Maravic 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, August 2005.

[50] I. Maravic, M. Vetterli, and K. Ramchandran. Channel estimation and synchronization with sub-Nyquist sampling and application to ultra-wideband systems. In *Proceedings of the International Symposium on Circuits and Systems (ISCAS'04)*, May 2004.

[51] F.A. Marvasti. *Nonuniform sampling: theory and practice*. Springer, NY, US, 2001.

[52] P. Marziliano. *Sampling innovations*. PhD thesis, Audio-Visual Communcation Laboratory, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, 2001.

[53] M. McCormick and M. Lu, Y. M. Vetterli. Learning sparse systems at subnyquist rates: A frequency-domain approach. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Dallas, Texas, USA, March 2010.

[54] P. Milanfar, G.C. Verghese, W.C. Karl, and A.S. Willsky. Reconstructing polygons from moments with connections to array processing. *IEEE Transactions on Signal Processing*, 43(2):432–443, February 1995.

[55] T.K. Moon and W.C. Stirling. *Mathematical methods and algorithms for signal processing*, volume 204. Prentice Hall, New York, 2000.

[56] H. Nyquist. Certain topics in telegraph transmission theory. *Transactions of the American Institute of Electrical Engineers*, 47(2):617–644, 1928.

[57] H. Olkkonen and J.T. Olkkonen. Measurement and reconstruction of impulse train by parallel exponential filter. *IEEE Signal Processing Letter*, 15:241–244, February 2008.

[58] A. Papoulis. Generalized sampling expansions. *IEEE Transactions on Circuits Syst.*, 24:652–654, November 1977.

[59] V.F. Pisarenko. The retrieval of harmonics from a covariance function. *Geophysical Journal of the Royal Astronomical Society*, 33(3):347–366, September 1973.

[60] B.D. Rao and K.S. Arun. Model based processing of signals: a state space approach. *IEEE Transactions on Signal Processing*, 80(2):283–309, February 1992.

[61] R. Roy and T. Kailath. ESPRIT-estimation of signal parameters via rotational invariance techniques. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37(7):984 –995, July 1989.

[62] T.K. Sarkar and O. Pereira. Using the matrix pencil method to estimate the parameters of a sum of complex exponentials. *Antennas and Propagation Magazine, IEEE*, 37(1):48–55, February 1995.

[63] A.H. Sayed. *Adaptive filters*. Wiley-IEEE Press, 2008.

[64] R. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, 34(3):276 – 280, March 1986.

[65] I. J. Schoenberg. Contribution to the problem of approximation of equidistant data by analytic functions. *Quarterly of Applied Mathematics*, 4(2):45–99 and 112–141, 1946.

[66] C.S. Seelamantula and M. Unser. A generalized sampling method for finite-rate-of-innovation-signal reconstruction. *IEEE Signal Processing Letter*, 15:813–816, November 2008.

[67] C.E. Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, January 1949.

[68] P. Shukla. *Sampling schemes for multidimensional nonbandlimited signals*. PhD thesis, Imperial College London, UK, 2007.

[69] P. Shukla and P.L. Dragotti. Sampling schemes for multidimensional signals with finite rate of innovation. *IEEE Transactions on Signal Processing*, 55(7):3670–3686, July 2007.

[70] P. Stoica and R. Moses. *Spectral Analysis of Signals*. Prentice Hall, April 2005.

[71] G. Strang and G. Fix. A Fourier analysis of the finite element variational method. In *Constructive Aspect of Functional Analysis*, pages 796–830, Rome (Italy), 1971.

[72] D.W. Tufts and R. Kumaresan. Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood. *Proceedings of the IEEE*, 70(9):975–989, September 1982.

[73] M. Unser. Splines: A perfect fit for signal and image processing. *IEEE Signal Processing Magazine*, 16(6):22–38, August 1999.

[74] M. Unser. Sampling-50 years after shannon. *Proceedings of the IEEE*, 88(4):569–587, April 2000.

[75] M. Unser. Cardinal exponential splines: part II-think analog, act digital. *IEEE Transactions on Signal Processing*, 53(4):1439–1449, March 2005.

[76] M. Unser and A. Aldroubi. A general sampling theory for nonideal acquisition devices. *IEEE Transactions on Signal Processing*, 42(11):2915–2925, November 1994.

[77] M. Unser, A. Aldroubi, and M. Eden. B-spline signal processing. I- theory. *IEEE Transactions on Signal Processing*, 41(2):821 –833, February 1993.

[78] M. Unser, A. Aldroubi, and M. Eden. B-spline signal processing. II- efficiency design and applications. *IEEE Transactions on Signal Processing*, 41(2):834 –848, February 1993.

[79] M. Unser and T. Blu. Cardinal exponential splines: Part I-theory and filtering algorithms. *IEEE Transactions on Signal Processing*, 53(4):1425–1438, April 2005.

[80] M. Unser and J. Zerubia. Generalized sampling: Stability and performance analysis. *IEEE Transactions on Signal Processing*, 45(12):2941–2950, December 1997.

[81] M. Unser and J. Zerubia. A generalized sampling theory without bandlimiting constraints. *IEEE Transactions on Circuits Syst. II*, 45(8):959–969, August 1998.

[82] F. Vanpoucke, M. Moonen, and Y. Berthoumieu. Efficient subspace algorithm for 2-D harmonic retrieval. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages 461–464, Toulouse (France), April 1994.

[83] K. Vetterli and J. Kovacevic. *Wavelets and Subband Coding*. Prentice-Hall, 1995.

[84] 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.

[85] J.M. Whittaker. The Fourier theory of the cardinal function. *Proceedings of the Edinburgh Mathematical Society (Series 2)*, 1(03):169–176, 1928.

[86] Z. Yang and F. Cohen. Cross-weighted moments and affine invariants for image registration and matching. *IEEE Transactions on Pattern Analysis and Machine Iintelligence*, 21(8):804–814, August 1999.