Exact and approximate Strang-Fix conditions to reconstruct signals with Finite Rate of Innovation from samples taken with arbitrary kernels

by
Jose Antonio Urigüen Garaizabal

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

Communications & Signal Processing Group
Department of Electrical & Electronic Engineering
Imperial College London
2013

Statement of originality

I declare that the intellectual content of this thesis is the product of my own research work under the guidance of my thesis advisor, 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.

Abstract

In the last few years, several new methods have been developed for the sampling and exact reconstruction of specific classes of non-bandlimited signals known as signals with finite rate of innovation (FRI). This is achieved by using adequate sampling kernels and reconstruction schemes. An example of valid kernels, which we use throughout the thesis, is given by the family of exponential reproducing functions. These satisfy the generalised Strang-Fix conditions, which ensure that proper linear combinations of the kernel with its shifted versions reproduce polynomials or exponentials exactly.

The first contribution of the thesis is to analyse the behaviour of these kernels in the case of noisy measurements in order to provide clear guidelines on how to choose the exponential reproducing kernel that leads to the most stable reconstruction when estimating FRI signals from noisy samples. We then depart from the situation in which we can choose the sampling kernel and develop a new strategy that is universal in that it works with any kernel. We do so by noting that meeting the exact exponential reproduction condition is too stringent a constraint. We thus allow for a controlled error in the reproduction formula in order to use the exponential reproduction idea with arbitrary kernels and develop a universal reconstruction method which is stable and robust to noise.

Numerical results validate the various contributions of the thesis and in particular show that the approximate exponential reproduction strategy leads to more stable and accurate reconstruction results than those obtained when using the exact recovery methods.

Acknowledgement

Completing a PhD is not an easy task. Research in itself is normally not very rewarding until you get to really understand the subject and are able to contribute to the community by publishing novel work. However, then, the entire learning and thinking process starts all over again. The key to success is being motivated and keeping momentum knowing that, eventually, so much effort is compensated with good results. The overall endeavour is challenging but, thankfully, fun too. For the last three years I have had ups and downs, and I admit I would not have been able to finish my PhD just on my own. There are a number of important persons that have helped me achieving my goals and I want to dedicate this introduction to them.

First of all, I would like to thank Dr. Pier Luigi Dragotti for his supervision and guidance throughout the entire duration of my PhD. Without his clear mind and wise advices I would not have been able to complete my thesis, let alone on time. Also, I would like to thank Dr. Thierry Blu, who has been a very close collaborator, without whose expert knowledge the current work would not have been possible. I hope to be able to be in touch with both of them in my new stage as investigator in Spain.

In addition, I thank my family and friends who have always been extremely supportive with every choice I have made in life. My mother Pilar – she has never needed to understand engineering to be happy for me as long as I was happy myself with my decisions. My sister Elena – she has given me many reasons to move on and to be proud of myself. And my friends in Bilbao and London: Jon Ander, Besart, Andryi, Sira or Jon to name just a few. Life in between Bilbao and London during these years would not have been the same without you all.

Finally I want to thank Amaia, my girlfriend and soon to be wife, thanks to whom I have found the inspiration I was lacking just before we met to complete the final stage of my PhD. To me, a fulfilling personal life is just as important, if not more, than a successful career. Without a doubt my PhD would not have ended so satisfactorily had I not met her at the right time.

Contents

List of Figures

List of Tables

Abbreviations

AWGN	additive white Gaussian noise
B-Spline	polynomial spline
CRB	CramérRao lower Bound
CS	Compressed Sensing
dB	decibel
eMOMS	exponential MOMS
E-Spline	exponential spline
GSVD	generalised singular value decomposition
IDFT	inverse discrete Fourier transform
FT	Fourier transform
FRI	finite rate of innovation
MOMS	maximum order, minimum support
MSE	mean square error
PSNR	peak signal-to-noise ratio
QSVD	quotient singular value decomposition
RMSE	root mean square error
SNR	signal-to-noise error
SVD	singular value decomposition
TLS	total least square

Notations and definitions

Tables 1, 2 and 3 provide notations, definitions and symbols that we use throughout the rest of the thesis. We summarise them here for ease of read and also for the reader to have a compact reference.

Table 1: Notations

Continuous time signals	
$f(t) \in L_2$	Real or complex-valued continuously defined signals with $t \in \mathbb{R}$, typically included in $L_2(\mathbb{R})$, which is the Hilbert space of finite-energy functions [1].
$f^{(r)}(t)$	rth derivative of $f(t)$. We note that the zero order derivative coincides with the function itself $f^{(0)}(t) = f(t)$. We may equivalently use $\frac{\mathrm{d}^{(R)}}{\mathrm{d}t^{(R)}}(f(t))$.
$f(t)^{-1}$	Multiplicative inverse or reciprocal of $f(t)$, i.e. $f(t)^{-1} = \frac{1}{f(t)}$. This should not be confused with the inverse function $f^{-1}(t)$.
$\langle f(\cdot), g(\cdot)\rangle$	Inner product in $L_2(\mathbb{R})$, defined as $\langle f(\cdot), g(\cdot)\rangle = \int_{-\infty}^{\infty} f(t)g^*(t)\mathrm{d}t$, where $g^*(t)$ is the complex conjugate of $g(t)$.
$\|f\|_{L_2}$	L_2-norm of $f(t)$, defined based on the inner product as $\|f\|_{L_2} = \sqrt{\langle f, f\rangle}$. When the context is clear we may simply use $\|f\|$.
$(f * g)(t)$	The convolution of two continuous-time functions $f(t)$ and $g(t)$ is $(f*g)(t) = \int_{-\infty}^{\infty} f(x)g^*(t-x)\mathrm{d}x$ which is equal to the inner product $\langle f(\cdot), g(t-\cdot)\rangle$.
$\hat{f}(\omega)$	Fourier transform of $f(t)$, given by $\hat{f}(\omega) = \int_{-\infty}^{\infty} f(t)\mathrm{e}^{-j\omega}\mathrm{d}t$ for $f(t)$ absolutely- and square- integrable. The inverse Fourier transform is hence defined as $x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(\omega)\mathrm{e}^{j\omega}\mathrm{d}t$.
$\hat{f}(s)$	Bilateral Laplace transform, defined as $\hat{f}(s) = \int_{-\infty}^{\infty} f(t)\mathrm{e}^{-s}\mathrm{d}t$. When $\hat{f}(s)$ is analytical along $s = j\omega$ then the bilateral Laplace transform at $s = j\omega$ coincides with the Fourier transform.

Discrete time signals and vectors	
$a_n \in \ell_2$	Real or complex-valued discrete time signals with $n \in \mathbb{Z}$, included in ℓ_2, which is the Hilbert space of square-summable sequences. [1]. We sometimes use $a[n]$ which is the same as a_n.
$(a \star b)[n]$	The convolution of two sequences a_n and b_n is defined as $(a \star b)[n] = \sum_{k\in\mathbb{Z}} a[k]b[n-k]$.
$\hat{a}(z)$	z-transform of the sequence a_n defined as $\hat{a}(z) = \sum_{n\in\mathbb{Z}} a_n z^{-n}$. Its discrete Fourier transform is obtained by setting $z = \mathrm{e}^{j\omega}$.
$\vec{\alpha}$, $\mathbf{u}$, $\mathbf{S}$	We mark vectors with an arrow to represent N-tuples, i.e., $\vec{\alpha} = (\alpha_1, \ldots, \alpha_N)$. Also, we write them in boldface lowercase, such as $\mathbf{u}$, whereas we use boldface uppercase to indicate matrices, $\mathbf{S}$. We usually work with column vectors.
$\langle \mathbf{u}, \mathbf{v} \rangle$	Inner product in ℓ_2, defined as $\langle \mathbf{u}, \mathbf{v} \rangle = \sum_n u_n v_n^*$, where v_n^* is the complex conjugate of v_n.
$\|\mathbf{u}\|_{\ell_2}$	ℓ_2-norm of $\mathbf{u}$, defined based on the inner product as $\|\mathbf{u}\|_{\ell_2} = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}$. When the context is clear we may simply use $\|\mathbf{u}\|$.
$(\cdot)^T$, …, $(\cdot)^\dagger$	$(\cdot)^T$ indicates transpose, $(\cdot)^*$ represents element-wise conjugate, $(\cdot)^H$ means Hermitian or conjugate transpose, $(\cdot)^{-1}$ refers to the inverse and $(\cdot)^\dagger$ to the Moore-Penrose pseudo-inverse.
$\mathrm{diag}(\cdot)$	Diagonal operator. It transforms a vector $(\cdot)$ into a diagonal matrix with elements $(\cdot)$ in its main diagonal.
$\mathbf{I}$	Identity matrix. If the size of the matrix is not clear from the context we will denote by $\mathbf{I}_N$ the identity matrix of size $N \times N$.

Table 2: Definitions

Functions	
$\mathrm{sinc}(t)$	We use the sinc function with the definition $\mathrm{sinc}(t) = 1$ for $t = 0$ and $\mathrm{sinc}(t) = \frac{\sin(\pi t)}{\pi t}$ elsewhere, with Fourier transform $\hat{f}(\omega) = \mathrm{rect}\left(\frac{\omega}{2\pi}\right)$.
$\mathrm{rect}(t)$	We define the rectangular function as $\mathrm{rect}(t) = 1$ for $\|t\| < \frac{1}{2}$, also $\mathrm{rect}(t) = \frac{1}{2}$ at $\|t\| = \frac{1}{2}$ and 0 elsewhere.
$\delta(t)$	The delta Dirac $\delta(t)$ is a distribution function that satisfies $\int_{-\infty}^{\infty} f(t)\delta^{(r)}(t - t_0)\mathrm{d}t = (-1)^r f^{(r)}(t_0)$, where $f(t)$ it r times continuously differentiable [2].

Useful formulae		
Poisson summation	$\sum_{n\in\mathbb{Z}} f(t+nT) = \frac{1}{T}\sum_{k\in\mathbb{Z}} \hat{f}\left(\frac{2\pi k}{T}\right) \mathrm{e}^{j2\pi k\frac{t}{T}}$	(1)

.

Table 3: Symbols

$x(t)$	Continuous-time input signal.
K	Number of degrees of freedom.
(t_k, a_k)	Innovation parameters of a train of Diracs.
$h(t)$	Sampling filter.
$\varphi(t)$	Sampling kernel.
$y(t)$	Filtered input.
y_n	Samples.
N	Number of samples.
T	Sampling period.
f_s	Sampling frequency.
τ	Sampling interval.
s_m	Exponential moments.
α_m	Exponential parameters.
$P+1$	Number of moments.
(u_k, x_k)	Parameters of the power sum sequence.
h_m	Annihilating filter.
$c_{m,n}$	Coefficients for the exponential reproducing property.
$\mathbf{y}$	Vector of samples $\mathbf{y} = (y_0, \ldots, y_{N-1})^T$.
$\mathbf{s}$	Vector of moments $\mathbf{s} = (s_0, \ldots, s_P)^T$.
$\mathbf{C}$	Matrix of coefficients $\mathbf{C} = [c_{m,n}]_{m=0,n=0}^{P,N-1}$.
ϵ_n	Additive white Gaussian noise samples.
b_n	Noise in the moments domain.
$\tilde{y}_n$	Noisy samples.
$\tilde{s}_m$	Noisy moments.
$(\hat{t}_k, \hat{a}_k)$	Estimate of (t_k, a_k).
$\mathbf{R}$	Covariance matrix of the additive white Gaussian noise vector $\mathbf{R} = \mathrm{E}\{\epsilon\epsilon^H\} = \sigma^2\mathbf{I}$.

σ^2	Additive white Gaussian noise variance.
$\mathbf{R_b}$	Covariance matrix of the noise vector $\mathbf{b}$, i.e. $\mathbf{R_b} = \mathrm{E}\{\mathbf{b}\mathbf{b}^H\}$
$\mathbf{R_B}$	Covariance matrix of the Toeplitz noise $\mathbf{B}$, i.e. $\mathbf{R_B} = \mathrm{E}\{\mathbf{B}^H\mathbf{B}\}$
ω	Angular frequency (radians per second).
B	Bandwidth [Hz].
dB	Decibels.
Hz	Hertz.
j	Imaginary unit $j = \sqrt{-1}$.
kHz	Kilohertz.
ρ	Rate of innovation.
s	Seconds.
V	Volts.
$\mathbb{C}$	Set of complex numbers.
$\mathbb{Z}$	Set of integer numbers.
$\mathbb{Z}\backslash\{0\}$	Set of integer numbers except for 0.
$\mathbb{R}$	Set of real numbers.
$\lvert\cdot\rvert$	Absolute value.
$\binom{n}{k}$	Binomial coefficient (n choose k).
$\lceil\cdot\rceil$	Ceil function.
$\mathrm{E}\{\cdot\}$	Expectation operator.
$(\cdot)!$	Factorial of the non-negative integer $(\cdot)$.
$\lfloor\cdot\rfloor$	Floor function.
$\mathrm{Im}\{\cdot\}$	Imaginary part.
$\mathrm{Re}\{\cdot\}$	Real part.
$[\cdot]$	Round function.

Chapter 1

Introduction

1.1 Motivation

Sampling, or the conversion of real-life continuous signals into discrete sequences of numbers that represent the original signals, plays a vital role in signal processing. Consider the typical sampling setup shown in Figure 1.1 where the original continuous-time phenomenon $x(t)$ is filtered before being (uniformly) sampled with sampling rate $f_{\mathrm{s}} = \frac{1}{T}$. The filtering may be a design choice or may be due to the acquisition device. If we denote with $y(t) = h(t) * x(t)$ the filtered version of $x(t)$, the samples y_n are given by

$$y_n = \left\langle x(t), \varphi\left(\frac{t}{T} - n\right)\right\rangle = \int_{-\infty}^{\infty} x(t)\varphi\left(\frac{t}{T} - n\right) \mathrm{d}t, \tag{1.1}$$

where the *sampling kernel* $\varphi(t)$ is the scaled and time-reversed version of $h(t)$.

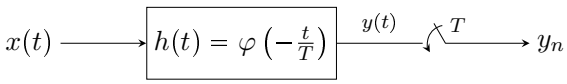


Figure 1.1: ***Traditional Sampling scheme.*** **The continuous-time input signal $x(t)$ is filtered with $h(t)$ and sampled every T seconds. The samples are then given by $y_n = (x * h)(t)|_{t=nT}$.**

Two basic questions arise in the context of the sampling scheme of Figure 1.1. First, under what conditions is there a one-to-one mapping between the measurements y_n and the original signal $x(t)$? Second, assuming such a mapping exists and given the samples y_n, how can a practical algorithm recover the original signal?

Sampling is a typical ill-posed problem in that one can construct an infinite number of signals that lead to the same samples y_n. To make the problem tractable one then has to impose some constraints on the choice of $x(t)$. Typically, the assumption made is that the bandwidth of $x(t)$ is limited to a maximum known frequency. In this case it is well known that a proper choice of the sampling kernel leads to a unique reconstruction formula (for instance by Shannon [3,4] Whittaker [5] or Kotelnikov [6,7]) In fact, the whole

sampling process can be interpreted as an approximation procedure in which the original signal is projected onto the shift-invariant subspace of bandlimited functions and only this projection can be reconstructed. This subspace interpretation has then been used to extend Shannon's theorem to classes of signals that belong to shift-invariant subspaces, such as uniform splines [8].

More recently, more general classes of signals have been considered and this includes signals that belong to union of subspaces. In particular in [9], the authors provide a general formulation of the multiple-subspace interpretation. The theories of *Finite Rate of Innovation* [2, 10] and *Compressed Sensing* [11, 12] are specific examples of complete procedures to sample and perfectly reconstruct some classes of signals living in a union of subspaces. The former provides an effective way to reconstruct parametric continuous-time signals from their samples, using adequate sampling kernels. The latter asserts that it is possible to recover certain types of discrete-time signals from less samples than traditional methods, relying on two principles: sparsity of the signals of interest, and incoherence of the sampling [13]. Interestingly, in both cases, the sampling procedure remains linear, but at the expense of non-linear reconstruction algorithms.

Even though finite rate of innovation (FRI) theory has evolved considerably over the last few years, its potential remains to be fully exploited. This is likely due to the fact that the reconstruction of these types of signals is a non-linear problem and some scenarios are potentially unstable in the presence of noise. Interestingly, the finite rate of innovation framework not only offers a sub-Nyquist alternative to the sampling paradigm, it also provides a parametric signal modelling that can accommodate many existing signals. However, finite rate of innovation theory has found its place only for a few applications. Again, improvements in the stability of the reconstruction are key to the development of the theory for real life scenarios.

1.2 Problem Statement

In this thesis we present an in-depth study of sampling and reconstruction of finite rate of innovation signals in the presence of noise. Specifically, we address the problem of making the FRI recovery stage as accurate and stable as possible in two different scenarios: First, when we have full control on the design of the sampling kernel and second when the sampling kernel is fixed but we have enough information about its shape to reconstruct the signal.

For the first part of the thesis we consider that the sampling kernel belongs to the family of exponential reproducing functions. Within this context, we analyse the behaviour of these kernels in the case of noisy measurements and provide clear guidelines on how to choose the exponential reproducing kernel that leads to the most stable reconstruction when estimating FRI signals from noisy samples. We then depart from the situation in which we can choose the sampling kernel and develop a new strategy that is universal in that it works with any kernel. We do so by noting that meeting the exact exponential

reproduction condition is too stringent a constraint. We thus allow for a controlled error in the reproduction formula in order to use the exponential reproduction idea with arbitrary kernels and develop a universal reconstruction method which is stable and robust to noise. Specifically, we compare our recovery method with the current state-of-the-art prior to our work for polynomial reproducing kernel and Gaussian kernels.

1.3 Outline of the thesis

The outline of the thesis is as follows. In Chapter 2 we review the noiseless scenario. We begin the chapter by formalising the notion of signals with finite rate of innovation. We then give some examples of FRI signals and explain the various types of sampling kernels used in the literature. We mainly concentrate on exponential reproducing kernels and introduce the generalised Strang-Fix conditions, for which we provide a simple proof. Finally, we describe how to sample and perfectly reconstruct the prototypical FRI signal: a train of Diracs. Moreover, we also explain how to sample and perfectly reconstruct other types of FRI signals using exponential reproducing kernels.

In Chapter 3 we treat the more realistic setup where noise is present in the acquisition process. Here, we describe practical techniques to retrieve a train of Diracs from samples obtained by an exponential reproducing kernel. We then adapt the main algebraic methods explained in the literature to work with coloured noise, which appears in the recovery process when working with exponential reproducing kernels. In addition, we present the Cramér–Rao bound (CRB) for the estimation problem related to the retrieval of the parameters of the input from the noisy samples. We also introduce a CRB formulation based on the exponential moments of the input that is better suited to measure the accuracy of the reconstruction for exponential reproducing kernels.

In Chapter 4 we design a family of exponential reproducing kernels that is most resilient to noise. We begin the chapter by considering the main sources of instability for FRI recovery. Moreover, we provide a practical method to select the proper parameters in order to design exponential reproducing kernels that are robust to noise. This new family of kernels extends the types of exponential reproducing kernels that have been used in the FRI literature. To end, we validate the stability and accuracy of these kernels with simulations.

In Chapter 5 we elaborate on the approximate FRI framework and develop the basic ideas to sample FRI signals with any kernel. The new approach is universal since it can be used with any sampling kernel. Furthermore, even though the recovery of FRI signals using this method is by definition only approximate, we show how to make the reconstruction error arbitrary small. Interestingly, we also show that with the new approximate framework we can improve the accuracy of the reconstruction associated to sampling kernels for which existing exact recovery methods become unstable in the presence of noise. In this chapter we provide extensive sets of simulations to demonstrate the potential of our method.

Finally, in Chapter 6 we present an application of FRI theory in Neuroscience. We first propose a simple parametric model for the neuronal activity signals. Based on this, we design an iterative reconstruction algorithm that can estimate a neuronal signal from FRI samples that have been obtained using an exponential reproducing kernel at reduced sampling rates. Our main contribution is that we show that state-of-the-art spike sorting performances can be reached with from recovered signals from samples taken at sub-Nyquist sampling rates.

1.4 Original Contributions and Publications

The material presented in this thesis has resulted in the following publications:

Journal papers

– J. A. Urigüen, P.L. Dragotti and T. Blu. "*FRI Sampling with Arbitrary Kernels*," to be published in IEEE Transactions on Signal Processing, 2013.

Book chapter

– J. A. Urigüen, P.L. Dragotti, Y. C. Eldar and Z. Ben-Haim. "*Sampling at the Rate of Innovation: Theory and Applications*" in "*Compressed Sensing: Theory and Applications*", edited by Yonina C. Eldar and published by Cambridge University Press, 1st edition (June 29, 2012).

Conference papers

– J. A. Urigüen, P.L. Dragotti and T. Blu. "*Approximate FRI with arbitrary kernels*," in Proceedings of the Tenth International Conference on Sampling Theory and Applications (SampTA'13), Bremen (Germany), July 1–5, 2013.

– J. Oñativia, J. A. Urigüen and P.L. Dragotti. "*Sequential local FRI sampling of infinite streams of Diracs*," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2013. Vancouver (Canada).

– J. Caballero, J. A. Urigüen, S. R. Schultz and P. L. Dragotti. "*Spike Sorting at Sub-Nyquist Rates*," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), March 2012. Japan.

– J. A. Urigüen, P.L. Dragotti and T. Blu. "*On the Exponential Reproducing Kernels for Sampling Signals with Finite Rate of Innovation*," in Proceedings of the Ninth International Workshop on Sampling Theory and Applications (SampTA'11), Singapore, May 2–6, 2011.

Conference Abstracts

- S. R. Schultz, J. Oñativia, J. A. Urigüen and P. L. Dragotti. "*A Finite Rate of Innovation Algorithm for Spike Detection from Two-Photon Calcium Imaging*," in Neuroscience 2012, October.

- J. A. Urigüen, P. L. Dragotti and T. Blu. "*Exponential Reproducing Kernels for Sparse Sampling*," in Signal Processing with Adaptive Sparse Structured Representations (SPARS) conference 2011. June 27-30. Edinburgh.

Chapter 2

Sampling signals with finite rate of innovation

In this chapter we review the theory of sampling and reconstructing finite rate of innovation in the absence of noise. We begin the chapter by introducing the notion of signals with finite rate of innovation (FRI) in Section 2.1. In addition, we provide some examples of FRI signals that can be sampled and perfectly reconstructed at their rate of innovation. In Section 2.2 we give an overview of the history of FRI, concentrating mostly on the noiseless FRI setting. We explain the more realistic noisy setup in Chapter 3. In Section 2.3 we describe the main types of sampling kernels that can be used for FRI. We end the section explaining exponential reproducing kernels in detail, since the theory of exponential reproduction is at the heart of our work. Then, in Section 2.4 we give a brief overview of the main techniques that have been developed in spectral estimation and related fields since the first high resolution method appeared. The reason, as explained in the following section, is the direct relation between spectral estimation and algebraic retrieval of signals with FRI. In Section 2.5 we review the canonical setting of sampling and perfectly reconstructing a train of K Diracs, from which many other sampling results can be derived. We also explain how to sample and perfectly reconstruct other types of FRI signals using exponential reproducing kernels. To conclude, in Section 2.6 we give some remarks for the specific exponential reproducing kernels that we use in the thesis, in Section 2.7 we anticipate some results to motivate our research further, and we end the chapter in Section 2.8.

2.1 Signals with finite rate of innovation

A signal bandlimited to $\left[-\frac{B}{2}, \frac{B}{2}\right]$ can be expressed as an infinite sum of properly weighted and shifted versions of the sinc function:

$$x(t) = \sum_{n \in \mathbb{Z}} x[n] \operatorname{sinc}(Bt - n), \tag{2.1}$$

where $x[n] = \langle x(t), B \operatorname{sinc}(Bt - n) \rangle = x(\frac{n}{B})$. We may also say that the signal $x(t)$ has B degrees of freedom per second, since it is exactly defined by a sequence of numbers $\{x[n]\}_{n \in \mathbb{Z}}$ spaced $T = B^{-1}$ seconds apart, given that the basis function sinc is known [14]. Equivalently, the signal $x(t)$ has a *rate of innovation* $\rho = B$. This idea can be generalised by replacing the sinc with the generating function $\varphi(t)$ that defines an approximation space by linear combinations of its shifted versions [8]. More specifically, the set of signals

$$x(t) = \sum_{n \in \mathbb{Z}} x[n] \varphi(Bt - n), \tag{2.2}$$

defines a shift-invariant subspace, which is not necessarily bandlimited, but that again has a rate of innovation $\rho = B$. Such functions can be efficiently sampled and reconstructed using linear methods [8]. Note that now $x_n = x[n]$ is the discrete representation of $x(t)$ in the approximation space, however it may differ from the samples in (2.1) since $\varphi(t)$ can be quite different from $\operatorname{sinc}(t)$.

Consider now the prototypical continuous-time sparse signal: a sum of Diracs located at instants of time $\{t_k\}_{k \in \mathbb{Z}}$ and weighted with amplitudes $\{a_k\}_{k \in \mathbb{Z}}$. Moreover, suppose that the average distance between consecutive Diracs is $\frac{1}{\lambda}$. This happens for example when the signal is generated from a Poisson process where the inter-arrival time is exponentially distributed with parameter λ [14]. Then, since the only degrees of freedom are the amplitudes and locations of the Diracs, we can calculate the rate of innovation of the signal as $\rho = 2\lambda$ [14]. Note that this class of signals no longer belongs to a single subspace, but rather to a union of subspaces [9]. In fact, observe that once we fix the delay values t_k (i.e. they are known), but let the amplitudes be unknown, then the train of Diracs lives in a linear subspace, spanned by $\{\delta(t - t_k)\}_{k \in \mathbb{Z}}$. Therefore, the entire signal class can be modelled as a union of subspaces, each of which corresponds to a set of possible delays $\{t_k\}_{k \in \mathbb{Z}}$. The estimation of these types of signals is clearly a non-linear problem, hence traditional sampling theory does not hold any more.

Is there a sampling theorem for this type of sparse signals? That is, can we acquire such signal by taking about ρ samples per unit of time, and perfectly reconstruct the original signal, just as the Shannon sampling procedure does? As it turns out, the answer is yes. Finite rate of innovation theory demonstrates that there are many types of signals that belong to a union of subspaces which can be sampled and perfectly reconstructed at the rate of innovation [2, 10, 14, 15]. In the rest of the section we give a formal definition of signals with FRI and provide several examples of such signals.

To easily navigate through the thesis, the reader can find the most frequent notations that are used in the sequel in Table 2.1.

Table 2.1: Frequently used notations

Symbol	Meaning
$x(t)$, τ	train of Diracs, sampling interval
t_k, a_k, K	innovation parameters of the train of Diracs, number of Diracs
$\varphi(t)$	antialiasing filter prior to sampling (typically an exponential reproducing kernel)
y_n, N, T	samples, number of samples, sampling period
s_m, $P+1$	exponential moments, number of moments (normally order of the kernel) $N \geqslant P+1 \geqslant 2K$
h_m, $M+1$	annihilating filter of length $K+1$, extended length of the filter

2.1.1 Formal definition of FRI signals

The concept of FRI is intimately related to parametric signal modelling. If a signal variation depends only on a few unknown parameters, then we can see it as having finite rate of innovation. Consider a signal of the form:

$$x(t) = \sum_{k\in\mathbb{Z}} \sum_{r=0}^{R-1} \gamma_{k,r} g_r(t - t_k). \tag{2.3}$$

If the set of functions $\{g_r(t)\}_{r=0}^{R-1}$ is known, then clearly the only degrees of freedom of the signal are the arbitrary shifts t_k and amplitudes $\gamma_{k,r}$. It is convenient to introduce a counting function $C_x(t_a, t_b)$ that counts the number of parameters of $x(t)$ over an interval of time $[t_a, t_b]$. Then the rate of innovation of the signal $x(t)$ is defined as [2, 10]:

$$\rho = \lim_{\tau\to\infty} \frac{1}{\tau} C_x\left(-\frac{\tau}{2}, \frac{\tau}{2}\right). \tag{2.4}$$

A signal with *finite rate of innovation* is a signal whose parametric representation is given by (2.3) and with a finite ρ as defined by (2.4). Given a signal $x(t)$ with finite rate of innovation ρ, we expect to be able to recover $x(t)$ from ρ measurements per unit time.

Another useful concept is that of a *local* rate of innovation over a sliding window of

size τ. The local rate of innovation at time t is given by [2, 10]:

$$\rho_\tau(t) = \frac{1}{\tau} C_x \left(t - \frac{\tau}{2}, t + \frac{\tau}{2} \right).$$

Note that $\rho_\tau(t)$ tends to ρ as τ tends to infinity.

2.1.2 Examples of FRI signals

As a motivation for the forthcoming analysis, we illustrate several examples of finite duration FRI signals in Figure 2.1. For simplicity, these examples are signals defined over the range $[0, 1)$, but the extension to other intervals and FRI models is straightforward.

An important example that we have introduced before is the *stream of Diracs* of Figure 2.1 (a). A stream of K Diracs with amplitudes $\{a_k\}_{k=0}^{K-1}$ located at different instants of time $\{t_k\}_{k=0}^{K-1}$, can be written as

$$x(t) = \sum_{k=0}^{K-1} a_k \delta(t - t_k), \tag{2.5}$$

where $t_k \in [0, 1)$ for all k. This signal has $2K$ degrees of freedom in total. The train of Diracs (2.5) can be easily generalised as a *stream of pulses* by simply replacing the Dirac shape $\delta(t)$ by a known pulse shape $p(t)$. An example of this more realistic signal is depicted in Figure 2.1 (e).

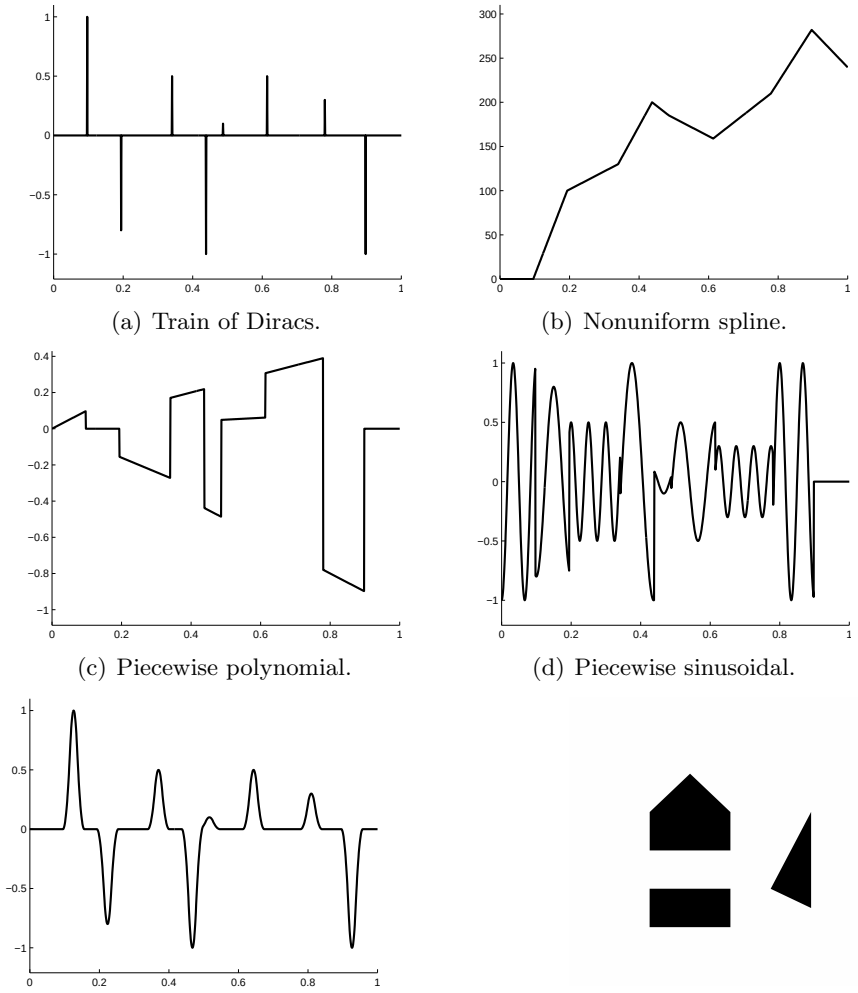
A signal is a *nonuniform spline* of degree R with amplitudes $\{a_k\}_{k=0}^{K-1}$ and knots at locations $\{t_k\}_{k=0}^{K-1}$ where $t_k \in [0, 1)$, for all k, if and only if its $(R+1)$th derivative is a stream of K weighted Diracs (2.5). Consequently, a nonuniform spline of order R consists of $K + 1$ segments (K transitions), each of which is a polynomial of degree R, such that the entire function is continuously differentiable $R - 1$ times. The $(R + 1)$th derivative then turns the knots into Diracs. This signal also has $2K$ degrees of freedom: the K amplitudes and K locations of the Diracs. An example is shown in Figure 2.1 (b) and its second derivative is the train of Diracs illustrated in (a).

We now recall the definition of derivative of Diracs, which is useful to introduce piecewise polynomial signals. The Dirac function is a distribution function whose rth derivative satisfies $\int_{-\infty}^{\infty} f(t)\delta^{(r)}(t - t_0)\mathrm{d}t = (-1)^r f^{(r)}(t_0)$, where $f(t)$ is r times continuously differentiable. A *stream of differentiated Diracs* with amplitudes $\{a_{k,r}\}_{k=0,r=0}^{K-1,R_k}$ and time locations $\{t_k\}_{k=0}^{K-1}$ is a linear combination of properly displaced and weighted differentiated Diracs, i.e.:

$$x(t) = \sum_{k=0}^{K-1} \sum_{r=0}^{R_k-1} a_{k,r} \delta^{(r)}(t - t_k).$$

In this case, the number of degrees of freedom of the signal is determined by K locations and $\tilde{K} = \sum_{k=0}^{K-1} R_k$ different weights.

A signal $x(t)$ is a *piecewise polynomial* with K segments of maximum degree $R - 1$



(a) Train of Diracs.

(b) Nonuniform spline.

(c) Piecewise polynomial.

(d) Piecewise sinusoidal.

(e) Stream of pulses.

(f) 2D set of bilevel polygons.

Figure 2.1: ***Examples of FRI signals that can be sampled and perfectly reconstructed at their rate of innovation.***

($R > 0$) if and only if its Rth derivative is a stream of differentiated Diracs, that is

$$x^{(R)}(t) = \sum_{k=0}^{K-1} \sum_{r=0}^{R-1} a_{k,r} \delta^{(r)}(t - t_k).$$

The signal has $K + \tilde{K} = K + RK = K(R + 1)$ degrees of freedom. An example of this type of signal is shown in Figure 2.1 (c). The difference with a nonuniform spline is that the piecewise polynomial is not differentiable at the knots.

Piecewise sinusoidal functions are a linear combination of truncated sinusoids, with unknown amplitudes a_{kd}, angular frequencies ω_{kd} and phases θ_{kd} [16]. Mathematically:

$$x(t) = \sum_{k=0}^{K-1} \sum_{d=0}^{D-1} a_{kd} \cos(\omega_{kd} t - \theta_{kd}) \xi_d(t),$$

with $\xi_d(t) = u(t - t_d) - u(t - t_{d+1})$, where t_d are locations to be determined, and $u(t)$ is the Heaviside step function. Fig. 2.1(d) shows an example of such a signal.

Finally, it is also possible to consider FRI signals in higher dimensions. For instance, a *2D stream of Diracs* can be written as

$$f(x, y) = \sum_{k=0}^{K-1} a_k \delta(x - x_k, y - y_k). \tag{2.6}$$

In Fig. 2.1(f) we show another type of two-dimensional signal, a *2D set of bilevel polygons*.

As shown in [2, 10, 16] all these are FRI signals for which we can derive exact sampling and reconstruction results.

2.2 History of FRI

Sampling signals with finite rate of innovation was first proposed by Vetterli et al. in [2]. The main focus of the paper was on sampling and perfectly reconstructing τ-periodic streams of K Diracs using a sinc kernel in the absence of noise. The authors show that recovery of the input is equivalent to finding the frequencies and amplitudes of a sum of complex exponentials. The latter is a standard problem in spectral analysis [17] and can be solved using conventional techniques, such as the annihilating filter method [14, 17], as long as the number of samples is at least $2K + 1$. Other classes of FRI signals are treated, such as nonuniform splines, streams of differentiated Diracs and piecewise polynomials.

Even though periodic inputs are convenient to analyse, in practice signals can often be modelled as finite or infinite streams of pulses. The first treatment of finite streams of Diracs appears in [2], using a Gaussian sampling kernel. Perfect reconstruction is achieved in the noiseless setting, but this approach is subject to numerical instability caused by the exponential decay of the kernel. A different method, based on calculating the moments of the input signal, was developed in [10], where the sampling kernels have

compact support in time and are able to reproduce polynomials or exponentials. The input can be determined from the signal moments by using again standard spectral estimation tools. In [15] the authors propose a similar scheme using a more complicated, but stable, exponential reproducing sampling kernel [18, 19]. Infinite streams of pulses with finite local rate of innovation have also been considered in the literature. For instance, in [10] sufficient conditions are derived to recover an infinite stream from a sequence of separate finite problems. A similar technique is employed in [15].

There has also been some work on FRI setups departing from the simple single channel one-dimensional scheme of Figure 1.1. Multichannel setups are considered in [20–22]. Some forms of distributed sampling have been studied in [23]. There has also been work on multidimensional FRI signals, such as images [24, 25]. To conclude, the applications of FRI theory include image super-resolution [26, 27], ultrasound imaging [15], multipath medium identification [28], super-resolution radar [28] and wideband communications [29, 30], among others.

A somewhat related field that has gained even greater attention in the last few years is Compressed Sensing (CS). Initially developed in parallel by Candes et al. [11] and Donoho [12], CS asserts that it is possible to recover certain types of discrete-time signals from less samples than traditional methods, relying on two principles: sparsity of the signals of interest, and incoherence of the sampling [13]. Interestingly, in both CS and FRI, the sampling procedure remains linear, but at the expense of non-linear reconstruction algorithms. In CS, the retrieval problem is addressed using convex relaxation techniques or greedy methods and, in the absence of noise, the signal is recovered with "overwhelming" probability [13] using M measurements, where M satisfies $M \gtrsim \mathcal{O}(K \log(N/M))$ [31].

Compressed sensing has recently been extended to working with continuous-time signals, for instance by Romberg, with the random convolution [32]: convolution with a random waveform followed by random time domain subsampling. The technique has been further elaborated from the perspective of random filtering in [33]. The random demodulator [34] and the modulated wideband converter (MWC) [35] are two recently proposed CS techniques for the acquisition of continuous-time spectrally-sparse signals [36]. Moreover, Xampling [37] is a design methodology for sub-Nyquist sampling of continuous-time analog signals that extends the MWC by proposing a digital algorithm which extracts each band of the signal from the compressed measurements.

Interestingly, neither CS or FRI are limited to uniform measurements (for instance, see [32] for CS and [38] for FRI) and either approach can potentially accommodate arbitrary sampling kernels. In particular, we present novel theory for FRI sampling and reconstruction with arbitrary kernels in this thesis [19].

2.3 Sampling kernels

The FRI acquisition process of Figure 1.1 is usually modelled as a filtering stage of the input $x(t)$ with a smoothing function $\varphi(t)$ (or sampling kernel), followed by uniform sam-

pling at a rate $f_s = \frac{1}{T}$[Hz]. According to this setup, the measurements are given by $y_n = \langle x(t), \varphi\left(\frac{t}{T} - n\right)\rangle$. The sampling kernel $\varphi(t)$ is therefore central to signal acquisition and, as we show in Section 2.5, also to signal reconstruction. The fundamental problem in sampling theory is to recover the original waveform $x(t)$ from the samples y_n. Equivalently, this is to say that the input $x(t)$ can be completely characterised by the set of samples y_n.

In the literature of FRI there exist two main types of filters: kernels of infinite duration [2], such as the sinc and the Gaussian functions, and kernels of compact support [10] like polynomial and exponential reproducing functions. In this section we review these kernels and highlight their main characteristics when used for FRI sampling.

2.3.1 Kernels of infinite support

The first kernel used for FRI was the traditional ideal low-pass filter of bandwidth B, i.e. $h(t) = \text{sinc}(Bt)$. In [2] and [14] the authors show that sampling a τ-periodic train of K Diracs with a sinc kernel of bandwidth B[Hz] leads to samples y_n that are the inverse discrete Fourier transform (IDFT) of $2M + 1 = B\tau$ consecutive Fourier series coefficients $\hat{x}_m$ of the input. And these coefficients can in turn be expressed as a power sum series, from which the $2K$ parameters of the Diracs can be estimated using the annihilating filter [14, 17]. The scheme is equivalent to that of sampling a finite stream of K Diracs with a τ-periodic sinc function (or Dirichlet kernel). This setup has shown to be very stable in noisy conditions [14].

The fundamental limit of the above sampling methods, as well as of the classical Shannon reconstruction scheme, is that the choice of the sampling kernel is very limited and the required kernels are of infinite support. As a consequence, the reconstruction algorithm is usually physically non-realisable (e.g., realisation of an ideal low-pass filter) or, in the case of FRI signals, becomes complex and unstable. The complexity is in fact influenced by the global rate of innovation of $x(t)$.

2.3.2 Kernels of compact support

Building upon previous work, new kernels of finite duration in time where presented by Dragotti et al. [10], to provide more practical and stable setups for sampling finite and infinite duration FRI signals. These kernels are known as polynomial and exponential reproducing kernels.

An exponential reproducing kernel is any function $\varphi(t)$ that, together with a linear combination of its shifted versions, can reproduce functions of the form $e^{\alpha_m t}$, with parameters α_m that can be complex valued. This can be expressed mathematically as follows:

$$\sum_{n\in\mathbb{Z}} c_{m,n}\varphi(t-n) = e^{\alpha_m t}, \tag{2.7}$$

for proper coefficients $c_{m,n}$ and where $m = 0, \ldots, P$ and $\alpha_m \in \mathbb{C}$. Exponential reproducing

kernels for which (2.7) is true satisfy the so-called generalised Strang-Fix conditions [39] (see also Appendix A.1 for a proof). Specifically, Equation (2.7) holds if and only if

$$\hat{\varphi}(\alpha_m) \neq 0 \text{ and } \hat{\varphi}(\alpha_m + 2j\pi l) = 0, \text{ for } m = 0, \dots, P \text{ and } l \in \mathbb{Z}\backslash\{0\}, \tag{2.8}$$

where $\hat{\varphi}(\alpha_m)$ represents the double-sided Laplace transform of $\varphi(t)$ evaluated at α_m.

Moreover, the coefficients $c_{m,n}$ in (2.7) are given by

$$c_{m,n} = \left\langle \mathrm{e}^{\alpha_m t}, \tilde{\varphi}(t-n) \right\rangle = \int_{-\infty}^{\infty} \mathrm{e}^{\alpha_m t} \tilde{\varphi}\left(t-n\right) \mathrm{d}t, \tag{2.9}$$

where the function $\tilde{\varphi}(t)$ forms a quasi-biorthonormal set [10,40] with $\varphi(t)$. One particular case of this is when $\tilde{\varphi}(t)$ is the dual of $\varphi(t)$, i.e. $\langle \tilde{\varphi}(t-n), \varphi(t-m) \rangle = \delta_{m-n}$. We also note that the coefficients $c_{m,n}$ are discrete-time exponentials. This fact can be shown by making a change of variable in (2.9):

$$c_{m,n} = \int_{-\infty}^{\infty} \mathrm{e}^{\alpha_m t} \tilde{\varphi}\left(t-n\right) \mathrm{d}t = \int_{-\infty}^{\infty} \mathrm{e}^{\alpha_m x} \mathrm{e}^{\alpha_m n} \tilde{\varphi}(x) \mathrm{d}x = \mathrm{e}^{\alpha_m n} c_{m,0}, \tag{2.10}$$

where $c_{m,0} = \int_{-\infty}^{\infty} \mathrm{e}^{\alpha_m x} \tilde{\varphi}(x) \mathrm{d}x$. We now show how to find a closed form expression for $c_{m,n}$. We first combine (2.10) and (2.7) to get

$$\sum_{n \in \mathbb{Z}} \mathrm{e}^{\alpha_m n} c_{m,0} \varphi(t-n) = \mathrm{e}^{\alpha_m t},$$

which can also be written as

$$c_{m,0} \sum_{n \in \mathbb{Z}} \mathrm{e}^{-\alpha_m (t-n)} \varphi(t-n) = 1.$$

Consequently the coefficients take the form:

$$c_{m,n} = \mathrm{e}^{\alpha_m n} c_{m,0} = \frac{\mathrm{e}^{\alpha_m n}}{\sum_{k \in \mathbb{Z}} \mathrm{e}^{-\alpha_m (t-k)} \varphi(t-k)}. \tag{2.11}$$

An alternative way of calculating the coefficients can be derived from the generalised Strang-Fix conditions. We do this in Chapter 4 and conclude that $c_{m,n} = \mathrm{e}^{\alpha_m n} \hat{\varphi}(\alpha_m)^{-1}$, i.e. $c_{m,0}^{-1}$ is the Laplace transform of the kernel evaluated at α_m for $m = 0, \dots, P$.

Cardinal exponential splines

A function $\beta_\alpha(t)$ with Fourier transform $\hat{\beta}_\alpha(\omega) = \frac{1-\mathrm{e}^{\alpha - j\omega}}{j\omega - \alpha}$ is called cardinal exponential spline of first order, with $\alpha \in \mathbb{C}$ [1]. The time domain representation of such function is $\beta_\alpha(t) = \mathrm{e}^{\alpha t}$ for $t \in [0,1)$ and zero elsewhere. The function $\beta_\alpha(t)$ is, therefore, of compact support: $\beta_\alpha(t) = 0$ for $t \notin [0,1)$, and a linear combination of its shifted versions reproduces the exponential $\mathrm{e}^{\alpha t}$.

Higher order cardinal exponential splines (E-Splines) can be obtained through convo-

lution of first order ones, so that for instance $\beta_{\vec{\alpha}}(t) = (\beta_{\alpha_0} * \beta_{\alpha_1} * \ldots * \beta_{\alpha_P})(t)$, where $\vec{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_P)$, is an E-Spline of order $P+1$. This can also be written in the Fourier domain as follows:

$$\hat{\beta}_{\vec{\alpha}}(\omega) = \prod_{m=0}^{P} \frac{1 - \mathrm{e}^{\alpha_m - j\omega}}{j\omega - \alpha_m}. \tag{2.12}$$

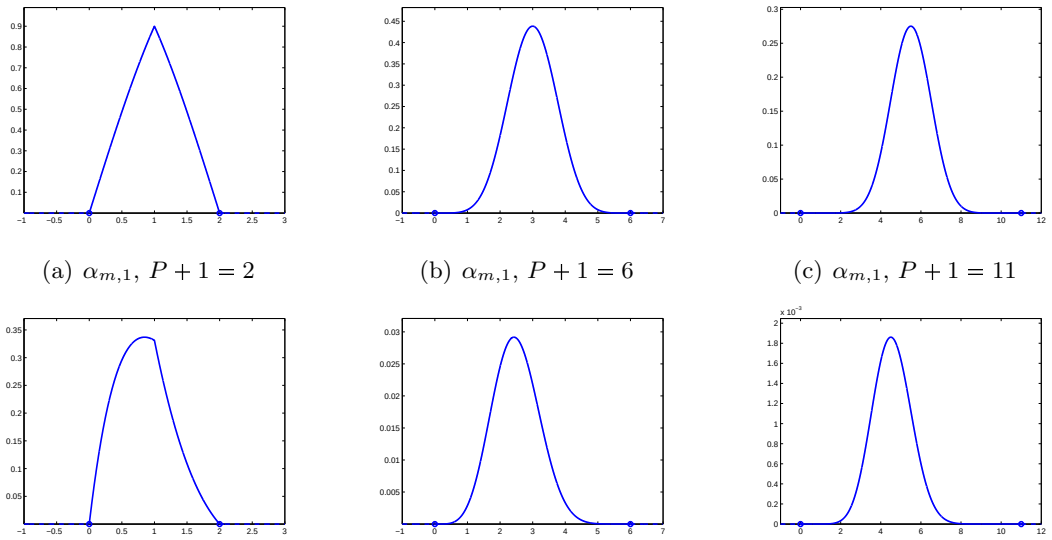
The higher order E-Splines are of compact support $P+1$ and their regularity increases with P (i.e. they have $P-1$ derivatives). These functions, combined with their shifted versions, can reproduce any exponential in the subspace spanned by $\{\mathrm{e}^{\alpha_0 t}, \mathrm{e}^{\alpha_1 t}, \ldots, \mathrm{e}^{\alpha_P t}\}$ [1, 10].

We also note that the family of E-Splines is a generalisation of that of B-Splines, in that when $\alpha_m = 0$ for $m = 0, \ldots, P$, the function $\beta_{\vec{\alpha}}(t)$ reduces to a B-Spline and no longer reproduces exponentials but polynomials up to order P. Furthermore, the exponential reproduction property is preserved through convolution [1, 10]. Thus, any function

$$\varphi(t) = \gamma(t) * \beta_{\vec{\alpha}}(t), \tag{2.13}$$

where $\gamma(t)$ is an arbitrary function, even a distribution, and $\beta_{\vec{\alpha}}(t)$ is an E-Spline, is still able to reproduce the exponentials in the subspace spanned by $\{\mathrm{e}^{\alpha_0 t}, \mathrm{e}^{\alpha_1 t}, \ldots, \mathrm{e}^{\alpha_P t}\}$.

We conclude by showing some examples of real valued E-spline functions of orders $P+1$ equal to 2, 6 and 11 in Figure 2.2. Kernels illustrated in (a, b, c) are built with $\alpha_{m,1} = j\frac{\pi}{2(P+1)}(2m - P)$ and those illustrated in (d, e, f) with $\alpha_{m,2} = -1 + \alpha_{m,1}$, $m = 0, \ldots, P$. Note that, contrary to the former, the latter are not symmetric around the middle point of the support $\frac{P+1}{2}$ due to the constant real part of the exponential parameters.



(a) $\alpha_{m,1}$, $P+1=2$ (b) $\alpha_{m,1}$, $P+1=6$ (c) $\alpha_{m,1}$, $P+1=11$

(d) $\alpha_{m,2}$, $P+1=2$ (e) $\alpha_{m,2}$, $P+1=6$ (f) $\alpha_{m,2}$, $P+1=11$

Figure 2.2: ***Examples of E-Splines.*** **In (a,b,c) we show real valued E-Spline kernels built using** (2.12) **with** $\alpha_{m,1} = j\frac{\pi}{2(P+1)}(2m - P)$ **for** $m = 0, \ldots, P$**. In (d,e,f) we show real valued E-Spline kernels built using** (2.12) **with** $\alpha_{m,2} = -1 + \alpha_{m,1}$ **for** $m = 0, \ldots, P$**.**

Generalised cardinal E-Splines

In [41] Unser defines a more general form of the E-Spline functions introduced so far. Consider the Fourier domain expression:

$$\hat{\beta}_{\vec{\alpha}}(\omega) = \prod_{m=0}^{P} \frac{1 - \mathrm{e}^{\alpha_m - j\omega}}{j\omega - \alpha_m} \prod_{\ell=0}^{M} (j\omega - \gamma_\ell), \tag{2.14}$$

which characterises a generalised E-Spline kernel, that is well defined as long as $M < P$ and $\gamma_\ell \neq \alpha_m$ for all ℓ and m. This is a much richer class than the one introduced before and it includes some known functions such as the family of maximum order, minimum support (MOMS) kernels that has been characterised in [42]. These correspond to the parametrisation $\alpha_m = 0$ for $m = 0, \ldots, P$, and $\gamma_\ell \in \mathbb{R}$ for $\ell = 0, \ldots, M$. In addition, generalised cardinal E-Splines also include the family of exponential MOMS (eMOMS) of [18,19], which correspond to α_m existing in complex conjugate pairs and being different from each other, and $\gamma_\ell \in \mathbb{R}$ for $\ell = 0, \ldots, M$. We study eMOMS in detail in Chapter 4.

Generality of exponential reproducing kernels

We conclude by highlighting the generality of exponential reproducing kernels. First, when $\alpha_m = 0$, any exponential reproducing kernel reduces to a kernel satisfying the Strang-Fix conditions [43]. These are still valid sampling kernels but reproduce polynomials rather than exponentials. Moreover, functions satisfying Strang-Fix conditions are extensively used in wavelet theory and the above result provides a connection between sampling of FRI signals and wavelets. In addition, it is possible to show that any device whose input and output are related by linear differential equations can be turned into an exponential reproducing kernel and can therefore be used to sample FRI signals [10]. This includes for example any linear electrical circuit. Given the ubiquity of such devices and the fact that in many cases the sampling kernel is given and cannot be modified, FRI theory with exponential reproducing kernels becomes even more relevant in practical scenarios.

2.4 A note on spectral estimation

As we will show in the next section, the parameters that characterise FRI signals that have been sampled by appropriate sampling kernels can be recovered by transforming the set of measurements y_n into a sequence of moments s_m equivalent to:

$$s_m = \sum_{k=0}^{K-1} x_k u_k^m, \tag{2.15}$$

for $m = 0, \ldots, P$. Here we are interested in retrieving x_k and u_k given only s_m and the knowledge that they are in a power series form (2.15). Equation (2.15) is common to problems such as decomposing a signal built from a linear mixture of complex exponential

(spectral estimation), estimating the direction of arrival of sources (array processing) or obtaining a polygonal shape from its complex moments (computed tomography, geophysical inversion, and thermal imaging) [44]. In the context of this thesis we always assume K is known beforehand, the same as for the survey of the current section. The case with K unknown is a related but completely different problem, the model order selection, that is dealt with in a wide variety of papers (see Appendix C in [17] and also [45] and references therein). Evidently, not knowing K would deteriorate the parametric estimation, however we do not delve into the problem since it is as hard as that of solving (2.15) and the interested reader can easily find considerable literature on the subject.

The first solution to (2.15) was given by Gaspard de Prony in 1795 [46]. His formulation is also known as the annihilating filter method, which is a standard tool in high-resolution spectral estimation [17]. In short, Prony's method reformulates (2.15) into:

$$\mathbf{Sh} = 0 \tag{2.16}$$

where $\mathbf{S}$ is a matrix built with the moments s_m and presents either Toeplitz or Hankel structure, and where the elements of $\mathbf{h}$ are the annihilating filter coefficients. Using Prony's method one can find the values u_k of (2.15) exactly using the minimum number of measurements $P + 1 = 2K$. Then, x_k are obtained directly from (2.15) in a variety of ways, such as ordinary or total least-squares [44].

When using noisy moments, one should expect to deviate from the exact relationship and having a greater number of moments $P+1 > 2K$ should improve the estimation of the parameters. Then, the simplest idea for solving the set of equations (2.16) given by Prony's method is to use least-squares [44]. Nevertheless, a total-least-squares (TLS) alternative should be preferred since all complex moments are perturbed and error in the equations appear on both sides. Numerically, the TLS problem is solved using the singular value decomposition (SVD). In any of the aforementioned variants of Prony's method, as well as in the original approach, the solution of the proposed system of equations leads to the coefficients of the annihilating filter. Then, the values u_k can be obtained by calculating the roots of the filter. An efficient and stable method for finding the roots is the companion matrix method [44], that converts a root-finding problem into an eigenvalue one.

Applying the SVD directly is suboptimal because we are in fact interested in a constrained SVD operation that reduces the rank of the matrix but also retains its Toeplitz structure. In his paper, Cadzow [47] suggested a simple numerical algorithm that attempts to solve this structured SVD problem by alternatively reducing the rank of $\mathbf{S}$ and then imposing the Toeplitz structure of the resulting matrix. The problem can be thought of as a structured-total least-squares one, which has been extensively analysed in [48, 49]. To conclude the methods based on Prony, we remark that the estimation problem can be made consistent by posing it as a statistical estimation one, for instance in terms of maximum-likelihood [44].

Starting again from (2.15), another useful relation leads to the matrix pencil method [50,

51]. It is possible to show that the following pencil relation holds:

$$\mathbf{S}_0\mathbf{v} = \lambda\mathbf{S}_1\mathbf{v}, \tag{2.17}$$

from which we obtain the values u_k as the eigenvalues of (2.17). Here, $\mathbf{S}_1$ is matrix $\mathbf{S}$ with either the first row or column removed and $\mathbf{S}_0$ is matrix $\mathbf{S}$ with either the last row or column removed. The matrix pencil method hence consists in build matrices $\mathbf{S}_1$ and $\mathbf{S}_0$ from the measurements and then finding the values u_k directly by solving the pencil (2.17). Note that this relation is true for the noiseless moments, and even weak noise may lead to no direct solution for the pencil.

The method "generalized pencil of function" (GPOF) [52] is designed for the noisy context. A perturbation analysis seems to indicate its near-optimality with respect to the Cramér Rao Bound [44]. Interestingly, a relationship between this method and several variants of the ESPRIT method [53, 54] is derived in [50]. ESPRIT is an advanced technique for solving the problem of direction of arrival (DOA) estimation, with performance and computational advantages over the well known MUSIC approach. Later work further improved the GPOF results by forcing the Hankel structure [55].

To conclude, we note that the subspace estimator method that we explain in Chapter 3 is based on a key algebraic property derived from the pencil (2.17) that is also used for the ESPRIT algorithm. Specifically, the noiseless matrices $\mathbf{S}_1$ and $\mathbf{S}_0$ satisfy the shift invariance property, i.e. $\mathbf{S}_1 = \mathbf{S}_0\mathbf{\Phi}$ where $\mathbf{\Phi}$ is a $K \times K$ diagonal matrix with u_k in its main diagonal. The matrices $\mathbf{U}$ and $\mathbf{V}$ containing the left or right singular vectors of the SVD decomposition of $\mathbf{S}$ also satisfy such property. Consequently, the values u_k can also be obtained as the eigenvalues of an operator that maps $\mathbf{U}_0$ onto $\mathbf{U}_1$ or $\mathbf{V}_0$ onto $\mathbf{V}_1$.

2.5 Sampling FRI signals in the noise-free setting

In this section, we first present the canonical setup of sampling and perfectly reconstructing a finite stream of Diracs with an exponential reproducing kernel of compact support, as discussed in [10]. In general, recovery is achieved by linearly combining the samples in order to obtain a new set of measurements, and then by retrieving the FRI signal parameters from the new sequence. The second stage is always equivalent to the problem of determining the amplitudes and frequencies of a signal formed by a sum of complex exponentials. This problem has been treated extensively in the array processing literature, and can be solved using conventional tools from spectral estimation theory [17] such as the annihilating filter [14] and subspace-based methods [50, 56–58].

Then, we explain how to sample and reconstruct other types of FRI signals treated in [2] using the Dirichlet kernel (or periodic sinc), and in [10] with polynomial reproducing kernels. These are nonuniform splines, streams of differentiated Diracs and piecewise polynomials. In [10] the authors provide sampling results for these signals, however no detailed derivations like the ones we include in the present section have appeared so far.

2.5.1 Sampling a train of K Diracs

Assume that the input signal $x(t)$ is a stream of K Diracs (2.5) and that we want to retrieve the innovation parameters $\{t_k, a_k\}_{k=0}^{K-1}$ from the samples

$$y_n = \left\langle x(t), \varphi\left(\frac{t}{T} - n\right)\right\rangle = \sum_{k=0}^{K-1} a_k \varphi\left(\frac{t_k}{T} - n\right), \tag{2.18}$$

where $n = 0, 1, \ldots, N-1$. Here, $\varphi(t)$ is an exponential reproducing kernel. The acquisition process is modelled in Figure 2.3.

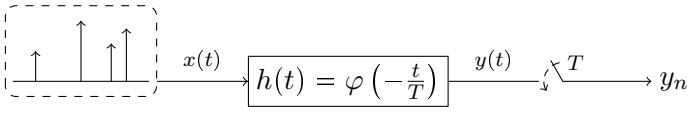


Figure 2.3: ***Sampling a train of Diracs.*** **The continuous-time input signal $x(t)$, a train of Diracs, is filtered with $\varphi\left(-\frac{t}{T}\right)$ and sampled every T seconds. The samples are then given by $y_n = \left\langle x(t), \varphi\left(\frac{t}{T} - n\right)\right\rangle$.**

Now, assume that all the locations satisfy $t_k \in [0, \tau)$, that they are different from each other and that the interval τ is an integer multiple of sampling period T, i.e. $\tau = NT$. Throughout the thesis we always restrict our analysis to parameters of the form

$$\alpha_m = \alpha_0 + m\lambda \tag{2.19}$$

for $m = 0, \ldots, P$, where $\alpha_0, \lambda \in \mathbb{C}$.

In order to show that the input can be unambiguously retrieved from the set of N samples y_n, we first linearly combine them with the coefficients $c_{m,n}$ of (2.7), to obtain the new measurements:

$$s_m = \sum_{n=0}^{N-1} c_{m,n} y_n, \tag{2.20}$$

for $m = 0, \ldots, P$. Then, combining (2.20) with (2.18) and taking into account the exponential reproducing property (2.7), we have [10]:

$$\begin{aligned} s_m &= \left\langle x(t), \sum_{n=0}^{N-1} c_{m,n} \varphi\left(\frac{t}{T} - n\right)\right\rangle = \int_{-\infty}^{\infty} x(t) \mathrm{e}^{\alpha_m \frac{t}{T}} \mathrm{d}t \\ &= \int_{-\infty}^{\infty} \sum_{k=0}^{K-1} a_k \delta(t - t_k) \mathrm{e}^{\alpha_m \frac{t}{T}} \mathrm{d}t \\ &= \sum_{k=0}^{K-1} a_k \mathrm{e}^{\alpha_m \frac{t_k}{T}} = \sum_{k=0}^{K-1} x_k u_k^m, \end{aligned} \tag{2.21}$$

for $m = 0, \ldots, P$ and with $x_k = a_k \mathrm{e}^{\alpha_0 \frac{t_k}{T}}$ and $u_k = \mathrm{e}^{\lambda \frac{t_k}{T}}$ for $k = 0, \ldots, K-1$. Here, it is the choice (2.19) that makes s_m have a power sum series form, which is key to the recovery of the innovation parameters of the input. The values s_m are precisely the (exponential)

moments of the signal $x(t)$, and are equivalent to the projection of $x(t)$ onto the subspace spanned by $\{\mathrm{e}^{\alpha_m \frac{t}{T}}\}_{m=0}^{P}$. Notice also that s_m represents the bilateral Laplace transform of $x(t)$ at α_m.

The new pairs of unknowns $\{u_k, x_k\}_{k=0}^{K-1}$ can then be retrieved from (2.21) by using conventional tools from spectral analysis [17] such as Prony's method or the annihilating filter method [2, 14]. More specifically, in order to find the values u_k in (2.21), let h_m for $m = 0, \ldots, K$ denote the filter whose z-transform is $\hat{h}(z) = \sum_{m=0}^{K} h_m z^{-m} = \prod_{m=0}^{K-1} \left(1 - u_k z^{-1}\right)$. That is, the roots of $\hat{h}(z)$ equal the unknown values u_k to be found. Then, it follows that:

$$h_m \star s_m = \sum_{i=0}^{K} h_i s_{m-i} = \sum_{i=0}^{K} \sum_{k=0}^{K-1} x_k h_i u_k^{m-i} = \sum_{k=0}^{K-1} x_k u_k^m \underbrace{\sum_{i=0}^{K} h_i u_k^{-i}}_{\hat{h}(u_k)=0} = 0 \tag{2.22}$$

where the last equality is due to the fact that we evaluate the z-transform of the filter at its zeros. The filter h_m is called an annihilating filter, since it nulls out the signal s_m. Its roots uniquely define the set of values u_k, provided that the locations t_k are distinct. The identity (2.22) can be written in matrix-vector form as:

$$\mathbf{Sh} = 0 \tag{2.23}$$

which reveals that the Toeplitz matrix $\mathbf{S}$ is rank deficient. Assuming without loss of generality that $h_0 = 1$ we may also write

$$\begin{pmatrix} s_K & s_{K-1} & \cdots & s_0 \\ s_{K+1} & s_K & \cdots & s_1 \\ \vdots & \vdots & \ddots & \vdots \\ s_P & s_{P-1} & \cdots & s_{P-K} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_K \end{pmatrix} = - \begin{pmatrix} s_0 \\ s_1 \\ \vdots \\ s_{K-1} \end{pmatrix} \tag{2.24}$$

which is a Yule-Walker system of equations that reveals we need at least $2K$ consecutive values of s_m to solve the above system. This implies that $P + 1 \geqslant 2K$, which indicates that the order $P + 1$ of the exponential reproducing kernel depends on the number of degrees of freedom of the input signal $x(t)$. Once the filter has been found, the locations t_k are retrieved from the zeros u_k of the z-transform of h_m. These in turn are given by root finding for the polynomial generated by $\hat{h}(z)$. Given the locations, the weights x_k can then be obtained by considering K consecutive equations in (2.21). For example, if we use the coefficients for $k = 0, 1, \ldots, K-1$, then we can write (2.21) in matrix-vector form as follows:

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ u_0 & u_1 & \cdots & u_{K-1} \\ \vdots & \vdots & \ddots & \vdots \\ u_0^{K-1} & u_1^{K-1} & \cdots & u_{K-1}^{K-1} \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_{K-1} \end{pmatrix} = \begin{pmatrix} s_0 \\ s_1 \\ \vdots \\ s_{K-1} \end{pmatrix}. \tag{2.25}$$

This is a Vandermonde system of equations that yields a unique solution for the weights x_k since the u_ks are distinct. We thus conclude that the original signal $x(t)$ is completely determined by the knowledge of $2K$ consecutive values s_m. We detail the main steps of the annihilating filter method in Algorithm 1.

Algorithm 1 Annihilating filter method.

Retrieve the parameters $\{t_k, a_k\}_{k=0}^{K-1}$ of a train of Diracs (2.5) from the samples (2.18) taken by an exponential reproducing kernel.

1: Calculate the sequence $s_m = \sum_{n=0}^{N-1} c_{m,n} y_n$ for $m = 0, \ldots, P$, from the N samples y_n of (2.18).

2: Build the system of equations (2.24) using the exponential moments s_m.

3: Retrieve the annihilating filter coefficients h_m, for $m = 0, \ldots, K$, by performing the singular value decomposition (SVD) [59] of the Toeplitz matrix $\mathbf{S}$ of (2.23) and choosing the singular vector corresponding to the zero singular value.

4: Compute the roots $u_k = \mathrm{e}^{\lambda \frac{t_k}{T}}$ of the z-transform $\hat{h}(z) = \sum_{k=0}^{K} h_m z^{-m}$ and obtain $\{t_k\}_{k=0}^{K-1}$.

5: Calculate $\{a_k\}_{k=0}^{K-1}$ as the least square solution of the N equations

$$y_n - \sum_{k=0}^{K-1} a_k \varphi\left(\frac{t_k}{T} - n\right) = 0.$$

If the measurements y_n are noisy, then it is necessary to denoise them by using the methods of Chapter 3.

The case of sampling and perfectly reconstructing streams of Diracs can be easily extended to any pulse satisfying $\hat{p}(s) \neq 0$ for $s = \alpha_m$, where $\hat{p}(s)$ is the Laplace transform of the pulse. This is due to the fact that sampling a stream of pulses with the kernel $\varphi(t)$ is equivalent to sampling a stream of Diracs with $\phi(t) = p(t) * \varphi(t)$. The above condition guarantees that the resulting kernel $\phi(t)$ is still able to reproduce the original set of exponentials.

We end the above discussion by noting that all FRI reconstruction setups proposed so far ([2,10,14,15]) can be unified as shown in Figure 2.4. Here, the samples are represented with the vector $\mathbf{y} = (y_0, y_1, \ldots, y_{N-1})^T$ and the moments are given by $\mathbf{s} = \mathbf{C}\mathbf{y}$. The matrix $\mathbf{C}$, of size $(P+1) \times N$ with coefficients $c_{m,n}$ at position (m, n), depends on the sampling kernel and its role becomes pivotal in noisy scenarios as discussed throughout the thesis. Techniques that are better suited to dealing with noise are discussed in Chapter 3.

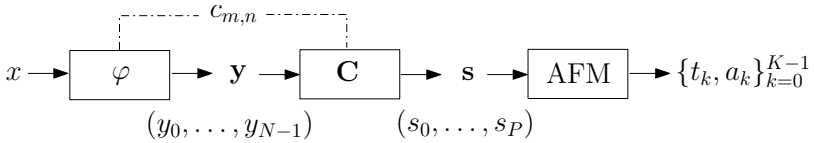


Figure 2.4: ***Unified FRI sampling and reconstruction.*** **The continuous-time input signal x is filtered with φ and uniformly sampled. Then, the vector of samples $\mathbf{y}$ is linearly combined to obtain the moments $\mathbf{s} = \mathbf{C}\mathbf{y}$. Finally, the parameters of the input are retrieved from $\mathbf{s}$ using the annihilating filter method (AFM).**

More on annihilation

We now briefly revisit (2.22). First, notice that any filter $\{h_m\}_{m=0}^{M}$ with $M \geqslant K$ having u_k as zeros satisfies the equation and, conversely, any filter that annihilates the coefficients s_m is also such that the values u_k are among its zeros. Now, using the fact that we can calculate s_m for $m = 0, \ldots, P$, we can rewrite (2.22):

$$\sum_{k=0}^{M} h_k s_{m-k} = 0, \quad m = 0, \ldots, P \tag{2.26}$$

where the sum runs from 0 to M. Expanding the equation in matrix form, we have:

$$\begin{pmatrix} s_M & s_{M-1} & \cdots & s_0 \\ s_{M+1} & s_M & \cdots & s_1 \\ \vdots & \vdots & \ddots & \vdots \\ s_P & s_{P-1} & \cdots & s_{P-M} \end{pmatrix} \begin{pmatrix} h_0 \\ h_1 \\ \vdots \\ h_M \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{2.27}$$

which has $P - M + 1$ equations and $M + 1$ unknowns, where $M \geqslant K$. We may also write $\mathbf{S}\mathbf{h} = \mathbf{0}$ and we do not require $h_0 = 1$ anymore. Moreover, when $M > K$ there are $M - K + 1$ independent polynomials of degree M with zeros at $\{u_k\}_{k=0}^{K-1}$. Thus, there are $M - K + 1$ vectors $\mathbf{h}$ that satisfy (2.27). Interestingly, the rank of $\mathbf{S}$ never exceeds K, which is key to solve (2.27) using the total least squares [14] or subspace estimator methods [60] that we present in Chapter 3.

2.5.2 Nonuniform splines

We now consider a nonuniform spline of order R with knots at $\{t_k\}_{k=0}^{K-1} \in [0, \tau)$ characterised by its $(R+1)$th derivative being a stream of K weighted Diracs:

$$x^{(R+1)}(t) = \sum_{k=0}^{K-1} a_k \delta(t - t_k). \tag{2.28}$$

In this case, we need to be able to relate the moments $s_m = \sum_n c_{m,n} y_n = \left\langle x(t), \mathrm{e}^{\alpha_m \frac{t}{T}} \right\rangle$

with the sequence

$$s_m^{(R+1)} = \left\langle x^{(R+1)}(t), \mathrm{e}^{\alpha_m \frac{t}{T}} \right\rangle = \sum_{k=0}^{K-1} x_k u_k^m,$$

for $m = 0, \ldots, P$, since the latter can be expressed in a power sum series form. Here, again, we have that $x_k = a_k \mathrm{e}^{\alpha_0 \frac{t_k}{T}}$ and $u_k = \mathrm{e}^{\lambda \frac{t_k}{T}}$. In order to do so, we note that

$$\begin{aligned} s_m^{(R+1)} &= \left\langle x^{(R+1)}(t), \mathrm{e}^{\alpha_m \frac{t}{T}} \right\rangle \\ &\overset{(a)}{=} \left\langle x(t), \left(\frac{\alpha_m}{T}\right)^{(R+1)} \mathrm{e}^{\alpha_m \frac{t}{T}} \right\rangle \\ &= \left(\frac{\alpha_m}{T}\right)^{(R+1)} s_m, \end{aligned} \tag{2.29}$$

for $m = 0, \ldots, P$, where (a) is due to the fact that the Dirac delta function satisfies $x(t) * \delta^{(R+1)}(t) = (-1)^{R+1} x^{(R+1)}(t)$, and explains the equivalence of both inner products. Therefore, we may obtain $s_m^{(R+1)}$ from s_m by using (2.29). Since $s_m^{(R+1)}$ is in a power sum series form, the retrieval of the innovation parameters is now straightforward using the annihilating filter method.

2.5.3 Streams of differentiated Diracs

Consider a finite stream of K differentiated Diracs with amplitudes $\{a_{k,r}\}_{k=0,r=0}^{K-1,R_k-1}$ and time locations $\{t_k\}_{k=0}^{K-1}$, which can be expressed as:

$$x(t) = \sum_{k=0}^{K-1} \sum_{r=0}^{R_k-1} a_{k,r} \delta^{(r)}(t - t_k). \tag{2.30}$$

The number of degrees of freedom of the signal is $K + \tilde{K}$, determined by K locations and $\tilde{K} = \sum_{k=0}^{K-1} R_k$ different weights. Assume that the signal is filtered with an exponential reproducing kernel to obtain measurements (2.18) for $n = 0, \ldots, N-1$. In this case, the exponential moments of the signal $x(t)$ can be simplified as follows:

$$\begin{aligned} s_m &= \sum_{m \in \mathbb{Z}} c_{m,n} y_n \overset{(a)}{=} \int_{-\infty}^{\infty} x(t) \mathrm{e}^{\alpha_m \frac{t}{T}} \mathrm{d}t \\ &\overset{(b)}{=} \sum_{k=0}^{K-1} \sum_{r=0}^{R_k-1} a_{k,r} \int_{-\infty}^{\infty} \mathrm{e}^{\alpha_m \frac{t}{T}} \delta^{(r)}(t - t_k) \mathrm{d}t \\ &\overset{(c)}{=} \sum_{k=0}^{K-1} \sum_{r=0}^{R_k-1} a_{k,r} (-1)^r \left(\frac{\alpha_m}{T}\right)^r \mathrm{e}^{\alpha_m \frac{t_k}{T}} \\ &= \sum_{k=0}^{K-1} \sum_{r=0}^{R_k-1} x_{k,r} (\alpha_m)^r u_k^m \end{aligned} \tag{2.31}$$

for $m = 0, \ldots, P$, where (a) follows from the linearity of the inner product and the exponential reproduction formula (2.20), (b) is due to the definition of $x(t)$, and (c) follows

from the application of the rth derivative property of the Dirac distribution function. Here, we have defined $u_k = \mathrm{e}^{\lambda \frac{t_k}{T}}$ and $x_{k,r} = a_{kr}(-T)^{-r}\mathrm{e}^{\alpha_0 \frac{t_k}{T}}$.

The annihilating filter method can be applied to the sequence s_m. We are looking for the filter h_m with z-transform $\hat{h}(z) = \prod_{k=0}^{K-1}(1-u_k z^{-1})^{R_k}$ that can annihilate $(\alpha_m)^r u_k^m = (\alpha_0 + \lambda m)^r u_k^m$ for $r = 0, \ldots, R_k - 1$ [2] (see also Appendix A.2). Now, the K coefficients of h_m can be found solving a Yule-Walker system of equations, by considering the annihilating filter equation $(h \star s)[m] = 0$. We need at least K equations to find the coefficients h_m (hence $N \geqslant P + 1 \geqslant 2K - 1$), from which the roots u_k and thus the time locations t_k can be calculated. Finally, the generalised Vandermonde system obtained using $\tilde{K}$ equations in (2.31) yields the amplitudes $x_{k,r}$. The solution is unique provided the time locations are different from each other.

2.5.4 Piecewise Polynomials

A signal $x(t)$ is a piecewise polynomial, with K pieces of maximum degree $R-1$ ($R > 0$) if and only if its Rth derivative is a stream of differentiated Diracs:

$$x^{(R)}(t) = \sum_{k=0}^{K-1} \sum_{r=0}^{R-1} a_{kr} r! \delta^{(R-r-1)}(t - t_k).$$

The number of degrees of freedom for this signal is $K + \tilde{K} = K + RK$. Assume that the signal is filtered with an exponential reproducing kernel to obtain measurements (2.18) for $n = 0, \ldots, N-1$.

The retrieval of the input is more involved than for the examples treated so far. However, using the Rth finite difference $z_n^{(R)}$, which is defined as:

$$z_n^{(R)} = \sum_{k=0}^{R} (-1)^k \binom{R}{k} y_{n+k}$$

we show that it is possible to relate the piecewise polynomial case with the derivative of Diracs scenario explained before. We also need to consider that, for any function $\varphi(t)$ with Fourier transform $\hat{\varphi}(\omega)$, the following is true:

$$\sum_{k=0}^{R} (-1)^k \binom{R}{k} \varphi\left(\frac{t}{T} - k\right) \xleftrightarrow{\mathrm{FT}} \tag{2.32}$$
$$\sum_{k=0}^{R} (-1)^k \binom{R}{k} T\hat{\varphi}\left(T\omega\right) \mathrm{e}^{-j\omega kT} \overset{(a)}{=} T(j\omega T)^R \hat{\varphi}\left(T\omega\right) \left(\frac{1 - \mathrm{e}^{-j\omega T}}{j\omega T}\right)^R$$

where (a) follows from Pascal's rule, which states $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k$, using $x = 1$ and $y = -\mathrm{e}^{-j\omega T}$. As a consequence, we have that

$$\sum_{k=0}^{R} (-1)^k \binom{R}{k} \varphi\left(\frac{t}{T} - k\right) = T^{R-1} \frac{\mathrm{d}^{(R)}}{\mathrm{d}t^{(R)}} \left[\varphi\left(\frac{t}{T}\right) * \beta_{R-1}\left(\frac{t}{T}\right)\right], \tag{2.33}$$

by just calculating the inverse Fourier transform of the right hand side of (2.32). Here $\beta_{R-1}(t) \leftrightarrow \left(\frac{1-\mathrm{e}^{-j\omega}}{j\omega}\right)^R$ is the B-spline of order $R-1$.

Then, we can relate the Rth finite difference $z_n^{(R)}$ with the Rth derivative of the signal $x^{(R)}(t)$ as follows:

$$\begin{aligned} z_n^{(R)} &= \sum_{k=0}^{R} (-1)^k \binom{R}{k} y_{n+k} \\ &\overset{(a)}{=} \left\langle x(t), \sum_{k=0}^{R} (-1)^k \binom{R}{k} \varphi\left(\frac{t}{T} - n - k\right) \right\rangle \\ &\overset{(b)}{=} \left\langle x(t), T^{R-1} \frac{\mathrm{d}^{(R)}}{\mathrm{d}t^{(R)}} \left[\varphi\left(\frac{t}{T} - n\right) * \beta_{R-1}\left(\frac{t}{T}\right) \right] \right\rangle \\ &\overset{(c)}{=} (-1)^R T^{R-1} \left\langle x^{(R)}(t), \varphi\left(\frac{t}{T} - n\right) * \beta_{R-1}\left(\frac{t}{T}\right) \right\rangle \end{aligned}$$

where (a) comes from the definition of the samples y_n and the linearity of the inner product, (b) is due to equation (2.33), and (c) follows from integration by parts. Thus, the Rth finite difference of the samples y_n, that we call $z_n^{(R)}$, is equivalent to the measurements that would be obtained sampling $x^{(R)}(t)$ with the kernel $\phi(t) = \varphi(t) * \beta_{R-1}(t)$. The advantage of this equivalent formulation is that the Rth derivative of the input is a stream of differentiated Diracs, which we showed how to sample and perfectly reconstruct in the previous section. Note, moreover, that $\phi(t)$ is able to reproduce the original set of exponentials $\{\mathrm{e}^{\alpha_m t}\}_{m=0}^{P}$, and is of compact support $S+R$, provided $\varphi(t)$ is of support S.

2.6 Remarks on real valued exponential reproducing kernels

In this thesis we work with real valued sampling kernels $\varphi(t) = \gamma(t) * \beta_{\vec{\alpha}}(t)$. Therefore, we require that $\gamma(t)$ and $\beta_{\vec{\alpha}}(t)$ be real valued. This second condition (i.e. $\beta_{\vec{\alpha}}(t)$ real valued) is satisfied when the parameters α_m are real or appear in complex conjugate pairs. We also restrict the exponential parameters to be of the form $\alpha_m = \alpha_0 + m\lambda$ with $m = 0, \ldots, P$. Therefore, we require that $\lambda = \frac{\alpha_0^* - \alpha_0}{P} = -2j\frac{\mathrm{Im}(\alpha_0)}{P}$ for the kernels to be real. As a consequence, the parameter λ is always purely imaginary and the roots $u_k = \mathrm{e}^{\lambda \frac{t_k}{T}}$ are periodic and lie on the unit circle. Since λ is purely imaginary and the values u_k are periodic, then it is necessary that t_k satisfy $0 \leqslant t_k < 2\pi \frac{T}{|\mathrm{Im}\{\lambda\}|}$ for $k = 0, \ldots, K-1$ in order to retrieve all the locations unambiguously.

Consider now the following equivalent way of writing the exponential parameters in order for them to exist in complex conjugate pairs:

$$\alpha_m = \alpha + j\frac{\pi}{L}(2m - P), \qquad m = 0, \ldots, P. \tag{2.34}$$

This is like saying $\alpha_0 = \alpha - j\frac{\pi P}{L}$ and $\lambda = j\frac{2\pi}{L}$. In this situation, the locations need to be such that $0 \leqslant t_k < LT$ for $k = 0, \ldots, K-1$, and we can control the retrieval interval by

simply modifying L in (2.34). For instance, if we want to sample a train of Diracs in an interval of duration τ, such that $t_k \in [0, \tau)$ for all k then we need that $L \geqslant \frac{\tau}{T}$.

2.7 Preliminary results

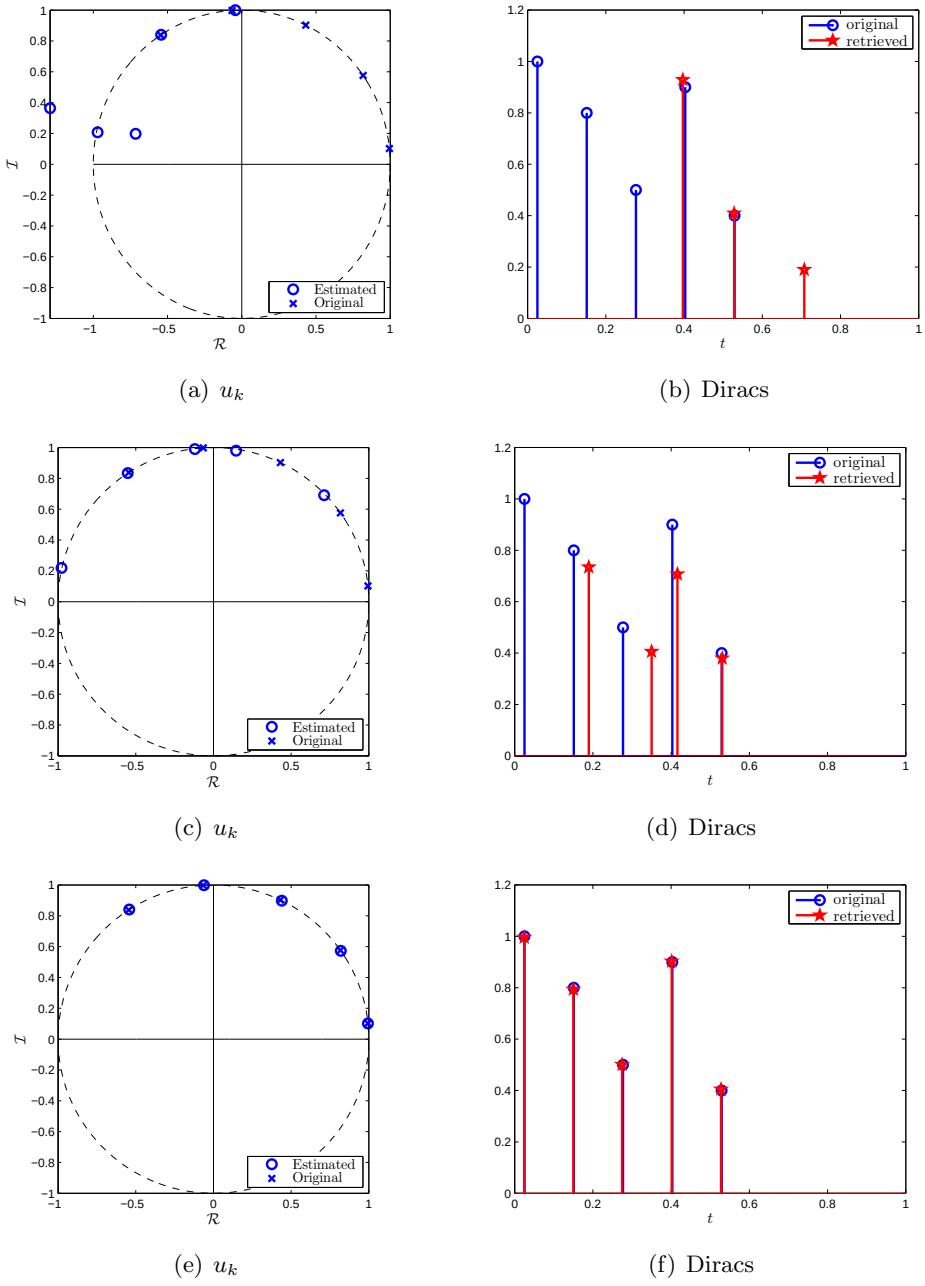
In this section we present a simple example that anticipates some results derived from the techniques we have developed in this thesis. Without entering into the details of the sampling and reconstruction setup, which will be described in the following chapters, in Figure 2.5 we illustrate the values u_k and Diracs that are reconstructed from noisy samples taken by an E-Spline kernel of parameters $\alpha_m = 1 + j\frac{2m-P}{2(P+1)}$ with $m = 0, \ldots, P$ and $P = 15$. The SNR is 50dB, which is very high, but already shows the instability of the retrieval based on the state-of-the art [10].

In Fig. 2.5 (a, b) we use the annihilating filter method of Section 2.5 with no additional filter coefficients or preconditioning step of any kind to provide robustness with noise. In Fig. 2.5 (c, d) we employ the modified subspace estimator of Section 3.3.2 which already improves the results notably. Finally, in Fig. 2.5 (e, f) we use the approximate FRI recovery of Section 5.2 to further improve the stability of the reconstruction and the accuracy of the retrieval.

2.8 Summary

In this chapter we have introduced the notion of signals with finite rate of innovation (FRI) as those characterised by a finite number of degrees of freedom. We have also given a brief overview of how noiseless FRI was developed, and we have presented various types of FRI signals and kernels. Then, we have focused on exponential reproducing kernels and shown how to perfectly reconstruct various FRI signals from noiseless samples. In the remaining of the thesis, we still mainly concentrate on exponential reproducing kernels. We no longer consider other FRI signals, but only the canonical case of a stream of Diracs, the retrieval of which we have linked to the classical problem of estimating the parameters of a sum of complex exponentials.

Exponential reproducing kernels are important for us for various reasons: (i) Any other kernel of compact support that has been used in the FRI literature is an instance of an exponential reproducing kernel. (ii) They are physically realisable and can accommodate many existing filtering devices. (iii) As we show in Chapter 4, other stable FRI setups such as the original one based on the Dirichlet kernel can be reduced to the case of using adequate exponential reproducing kernels and periodic sampling. (iv) Finally, as we show in Chapter 5, the exponential reproducing property becomes key in the generalisation of the FRI framework to sampling with arbitrary kernels.



(a) u_k (b) Diracs

(c) u_k (d) Diracs

(e) u_k (f) Diracs

Figure 2.5: ***Comparison of the methods proposed in the thesis with the state-of-the-art.*** **(a,b) Basic annihilating filter method based on the state-of-the art. (c,d) Modified subspace estimator of Section 3.3.2. (e,f) Approximate FRI recovery of Section 5.2. The SNR is 50dB in all cases.**

Chapter 3

Sampling and recovery of FRI signals in the presence of noise

In the previous chapter we have introduced the basic ideas on FRI, mainly the definition of FRI signals and sampling kernels, and the noiseless setup for various types of signals. In this chapter, we study the noisy scenario and review existing techniques to deal with non-ideal measurements. Moreover, we adapt the main subspace based techniques used in the FRI literature [14,60] to work with coloured noise, which appears when sampling with exponential reproducing kernels. Our formulation is general, so that the methods can be easily used for other types of sampling kernels. In addition, in this chapter we employ a standard tool for evaluating parametric estimation problems [17]: the Cramér–Rao lower bound (CRB) [61, 62]. We not only use the CRB to determine the best accuracy with which the train of Diracs may be recovered from the samples y_n, but we also introduce a new bound based on the sum of exponentials model of equation (2.21). This chapter should also serve to identify the main sources of instability in the retrieval process of FRI signals. Equipped with the conclusions of the current chapter, we elaborate denoising further in Chapters 4 and 5 in an attempt to optimise FRI reconstruction in the presence of noise when using exponential reproducing kernels.

We should note that the way we modify the TLS-Cadzow routine of [14] has not been used for FRI before, but is common in the spectral estimation literature (for example [63, 64]). On the other hand, the way we modify the subspace estimator of [60] is a unique contribution of this thesis. Also, the CRB for the estimation of the innovation parameters from the samples has already been used in the FRI literature [14], but the new bound based on the sum of exponentials model is novel.

To begin, we explain the noisy setting in Section 3.1. Next, in Section 3.2 we review existing methods in the FRI literature that have been successfully used to combat additive white Gaussian noise. Then, in Section 3.3 we present modifications of some of the methods in order to properly deal with coloured noise. To conclude, in Section 3.4 we discuss the Cramér–Rao lower bound associated to the samples and the moments. Finally, we validate the algorithms discussed in Section 3.5 to then conclude the chapter in Section 3.6.

3.1 The noisy FRI setting

"Noise", or more generally model mismatch, is present in data acquisition, making the solution presented in Chapter 2 valid only in ideal circumstances. In the presence of noise the acquisition process may be modelled as illustrated in Figure 3.1. In general, we may assume the signal is corrupted in the continuous-time domain by analogue noise and in the discrete-time domain by digital noise.

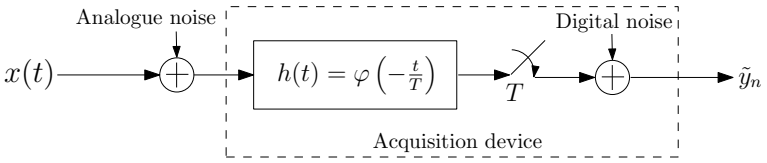


Figure 3.1: ***Noise perturbations in the sampling set-up.*** **The continuous-time signal $x(t)$ can be corrupted either in the analog or the digital paths. In this thesis we consider only the perturbation due to digital noise.**

For simplicity, we assume that the noiseless samples y_n are corrupted only by digital additive noise and that we have access to the measurements

$$\tilde{y}_n = y_n + \epsilon_n = \sum_{k=0}^{K-1} a_k \varphi\left(\frac{t_k}{T} - n\right) + \epsilon_n, \tag{3.1}$$

with $n = 0, \ldots, N-1$. We further assume ϵ_n are i.i.d. Gaussian random variables of zero mean and standard deviation σ. Equation (3.1) is valid when the input $x(t)$ is a finite stream of K Diracs (2.5) and the noise is additive. When the samples are corrupted by noise, the set of measurements s_m of Equation (2.20) changes, and perfect reconstruction is no longer possible. Specifically we now have the noisy (exponential) moments:

$$\begin{aligned} \tilde{s}_m = \sum_{n=0}^{N-1} c_{m,n}\tilde{y}_n &= \underbrace{\sum_{n=0}^{N-1} c_{m,n} y_n}_{s_m} + \underbrace{\sum_{n=0}^{N-1} c_{m,n}\epsilon_n}_{b_m} \\ &= \sum_{k=0}^{K-1} x_k u_k^m + b_m, \end{aligned} \tag{3.2}$$

for $m = 0 \ldots, P$ and where $x_k = a_k \mathrm{e}^{\alpha_0 \frac{t_k}{T}}$ and $u_k = \mathrm{e}^{\lambda \frac{t_k}{T}}$ with $k = 0, \ldots, K-1$.

Note that the retrieval procedure for FRI signals is based on calculating measurements (2.20) from the samples y_n. This is the case for any FRI setup, only that the values $c_{m,n}$ change with the choice of the sampling kernel. For instance, if the sampling kernel is the τ-periodic sinc or Dirichlet kernel [2, 14], the coefficients $c_{m,n}$ take the simple form of being the entries of the inverse discrete Fourier transform (IDFT) matrix. Another example is the Gaussian function [2, 29], for which the coefficients $c_{m,n}$ are equal to the diagonal entries of a matrix with $\mathrm{e}^{\alpha n^2}$ along its main diagonal and zeros elsewhere. When the samples are corrupted by noise, the ideal measurements s_m of Equation (2.20) be-

come (3.2). Moreover, the statistics of the noise measurements b_m depend directly on the original noise distribution and on the coefficients $c_{m,n}$.

Note also that equation (3.2) may be written in matrix-vector as follows:

$$\tilde{\mathbf{s}} = \underbrace{\mathbf{C}\mathbf{y}}_{\mathbf{s}} + \underbrace{\mathbf{C}\epsilon}_{\mathbf{b}}, \tag{3.3}$$

where we remind the reader that matrix $\mathbf{C}$, as introduced in Chapter 2, is a matrix of dimensions $(P+1) \times N$ with coefficients $c_{m,n}$ at location (m,n); $\mathbf{y}$ and ϵ are vectors of length N with the noiseless samples and the noise measurements respectively; and $\mathbf{s}$, $\mathbf{b}$ and $\tilde{\mathbf{s}}$ are vectors of length $P+1$ containing the noiseless moments, the noise measurements and the noisy moments, respectively.

3.2 Retrieval of FRI signals in the presence of AWGN

As we already highlighted in the previous chapter, the FRI recovery problem parallels that of line spectra estimation. Consequently, most of the methods used in the spectral estimation literature may be adapted to the FRI setting. A comprehensive review of the spectral estimation literature can be found in the book [17]. We also refer the reader to other relevant publications on the subject such as [44, 45].

In this section we first review some methods that have been used for FRI and that allow to retrieve the innovation parameters of a train of K Diracs "optimally" in the presence of additive white Gaussian noise (AWGN). Optimality should be understood in the sense that the methods reach the Cramér–Rao lower Bound (CRB), which indicates the best achievable performance of an unbiased estimator. We explain how to deal with non-white additive Gaussian noise later on in the chapter. We conclude the section with a brief literature review including other approaches that deal with the noisy FRI setting.

3.2.1 Total least squares and Cadzow algorithm

When noise is present in the acquisition process, the annihilating equation (2.27) is not satisfied exactly. The reason is that, now, we have the following Toeplitz matrix of measurements:

$$\tilde{\mathbf{S}} = \begin{pmatrix} \tilde{s}_M & \tilde{s}_{M-1} & \cdots & \tilde{s}_0 \\ \tilde{s}_{M+1} & \tilde{s}_M & \cdots & \tilde{s}_1 \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{s}_P & \tilde{s}_{P-1} & \cdots & \tilde{s}_{P-M} \end{pmatrix},$$

where each $\tilde{s}_m = s_m + b_m$. Equivalently, we may write that $\tilde{\mathbf{S}} = \mathbf{S} + \mathbf{B}$ and therefore (2.27) becomes

$$\tilde{\mathbf{S}}\mathbf{h} = (\mathbf{S} + \mathbf{B})\mathbf{h} \approx \mathbf{0}. \tag{3.4}$$

Here, $\mathbf{B}$ is also Toeplitz, and its elements are the noise components associated to each element of $\mathbf{S}$. Note that the matrices in (3.4) are all of size $(P-M+1)\times(M+1)$, subject to $P+1 \geqslant 2M \geqslant 2K$. Typically, the noise matrix $\mathbf{B}$ is full rank, which makes the noisy matrix of measurements $\tilde{\mathbf{S}}$ become full rank. This is in contrast to the original matrix $\mathbf{S}$ that is rank deficient and of rank K. To be more precise, the conditioning of $\mathbf{B}$ is related to that of $\mathbf{C}$, even though the relation in between both matrices is rather cumbersome (see Appendix C.5). If $\mathbf{C}$ is close to rank deficient so is $\mathbf{B}$, fact that emphasises the importance for having a well conditioned matrix of coefficients $\mathbf{C}$.

Even though (2.27) no longer holds, it is still possible to obtain the annihilating filter by solving (3.4) approximately. One way of doing so is to assume $h_0 = 1$ and $M = K$, then move the first column of $\tilde{\mathbf{S}}$ to the right-hand-side of the equation. The solution to (3.4) is the vector $\mathbf{h}$ that minimises the ℓ_2 norm of the error, but considering the only source of noise appears on the vector of the right-hand-side. This is the least-squares solution, that can be computed using the pseudoinverse of the matrix of the left-hand-side times the vector of the right-hand-side.

However, since noise is present in all measurements, it is more appropriate to search for the solution that minimises the norm $\|\tilde{\mathbf{S}}\mathbf{h}\|^2$ under the constrain $\|\mathbf{h}\|^2 = 1$. This is a classical total least square (TLS) problem that can be solved using singular value decomposition (SVD) [14]. In order to do so, first the matrices $\mathbf{U}$, $\boldsymbol{\Delta}$ and $\mathbf{V}$ that satisfy $\tilde{\mathbf{S}} = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^H$, have to be found. Then, the annihilating filter coefficients $\mathbf{h}$ are given by the right singular vector corresponding to the smallest singular value of $\tilde{\mathbf{S}}$ or, in other words, the last column of $\mathbf{V}$. When $M = K$ and since $\mathbf{V}$ is of size $(M+1)\times(M+1)$, then there is only one vector in the noise space of $\tilde{\mathbf{S}}$ and using SVD we directly determine it is the last column of $\mathbf{V}$. Also, this vector is of length $M+1 = K+1$ and its roots correspond to the estimated values for $\hat{u}_k$. However, whenever $M > K$ then there are $M-K+1$ vectors that form the noise space of $\tilde{\mathbf{S}}$ and they are the last $M-K+1$ columns of $\mathbf{V}$. In this case, it is possible to select $\mathbf{h}$ as any single one of the vectors of the noise subspace, or even combine all of them [65]. The solution is now of length $M+1$ and the K estimated values for $\hat{u}_k$ are among the M roots. We know from Chapter 2 that the noiseless u_k lie on the unit circle, therefore we may choose the K out of the M possible roots that lie inside or on the unit circle [45, 66].

By only applying TLS we are not solving the noisy system of equations (3.4) optimally, since if we use SVD and keep the last right singular vector we are equivalently finding $\mathbf{h}$ such that $\tilde{\mathbf{S}}_K\mathbf{h} = \mathbf{0}$, where $\tilde{\mathbf{S}}_K$ is the rank-K approximation of $\tilde{\mathbf{S}}$. This implies that $\tilde{\mathbf{S}}_K$ does not maintain the original Toeplitz form of $\tilde{\mathbf{S}}$. In fact, we really want to solve what is known as the structured low-rank approximation problem [67] that attempts at finding the rank-K Toeplitz matrix $\hat{\mathbf{S}}$ that is closest to $\tilde{\mathbf{S}}$ in the sense that it minimises the distance $\|\hat{\mathbf{S}} - \tilde{\mathbf{S}}\|$, for some matrix norm. This problem is linked to structured total least squares [48, 49].

Therefore we do better by solving a constrained SVD problem that retains the Toeplitz structure of $\tilde{\mathbf{S}}$ and at the same time reduces its rank. In his paper, Cadzow [47] suggests

a simple approach that attempts at refining the solution given by the TLS method by solving the structured SVD problem. The idea is to first apply a rank-reduction operation by keeping the K largest singular values and replacing the smallest singular values by zero. Then, the Toeplitz structure is forced by making the diagonals entries equal to the average value along the elements of the diagonals of the low rank approximation. Iterating between these two operations, Cadzow's algorithm converges to a local minimum of the function defined by the distance to the closest reduced-rank Toeplitz matrix [44]. This procedure works best when $\tilde{\mathbf{S}}$ is as close as possible to a square matrix [14], and so a good choice would be using $M \approx \frac{P}{2}$. Even though this algorithm works well in practice [44], De Moor showed that the process does not guarantee the optimum rank-K Toeplitz matrix is obtained [67, 68]. To conclude, we summarise the TLS-Cadzow routine in the box Algorithm 2.

3.2.2 The subspace estimator method

The subspace estimator is a robust parametric estimation algorithm that directly estimates u_k without the need of calculating the annihilating filter coefficients h_m. As such, it achieves comparable denoising performance to the TLS-Cadzow routine, but it is not iterative. The subspace estimator is based on the shift-invariance property derived from the pencil (2.17) that also holds for the matrices $\mathbf{U}$ and $\mathbf{V}$ containing the left or right singular vectors of the SVD decomposition of $\mathbf{S}$ in (2.17). The method obtains the SVD decomposition of such matrix and estimates u_k as the eigenvalues of an operator that maps either $\mathbf{U}_0$ onto $\mathbf{U}_1$ or $\mathbf{V}_0$ onto $\mathbf{V}_1$.

To be more specific, consider the Hankel (equivalent to Toeplitz) matrix of ideal measurements

$$\mathbf{S} = \begin{pmatrix} s_0 & s_1 & \cdots & s_M \\ s_1 & s_2 & \cdots & s_{M+1} \\ \vdots & \vdots & \ddots & \vdots \\ s_{P-M} & s_{P-M-1} & \cdots & s_P \end{pmatrix} \tag{3.5}$$

where s_m are as in (2.20) for $m = 0, \ldots, P$. Then, this matrix can be decomposed as

$$\mathbf{S} = \underbrace{\begin{pmatrix} 1 & 1 & \cdots & 1 \\ u_0 & u_1 & \cdots & u_{K-1} \\ \vdots & \vdots & \ddots & \vdots \\ u_0^{P-M} & u_1^{P-M} & \cdots & u_{K-1}^{P-M} \end{pmatrix}}_{\mathbf{U}} \underbrace{\begin{pmatrix} x_0 & 0 & \cdots & 0 \\ 0 & x_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_{K-1} \end{pmatrix}}_{\mathbf{\Delta}} \underbrace{\begin{pmatrix} 1 & u_0 & \cdots & u_0^M \\ 1 & u_1 & \cdots & u_1^M \\ \vdots & \vdots & \ddots & \vdots \\ 1 & u_{K-1} & \cdots & u_{K-1}^M \end{pmatrix}}_{\mathbf{V}^H}.$$

This factorization is not unique, in fact $\mathbf{S} = (\mathbf{UP})(\mathbf{P}^{-1}\mathbf{\Delta}\mathbf{Q})(\mathbf{Q}^{-1}\mathbf{V}^H)$ is another possibility, for any $\mathbf{P}$ and $\mathbf{Q}$ of size $K \times K$ that are invertible.

In order to be precise, we introduce some notation that we use throughout the section. Assume the noiseless matrix $\mathbf{S}$ is as in (3.5) of dimensions $(P-M+1) \times (M+1) = d_1 \times d_2$,

Algorithm 2 TLS-Cadzow algorithm.

Retrieve the parameters $\{t_k, a_k\}_{k=0}^{K-1}$ of a train of Diracs (2.5) from noisy moments using TLS and Cadzow's iterative routine.

1: Calculate the sequence

$$\tilde{s}_m = \sum_{n=0}^{N-1} c_{m,n}\tilde{y}_n \quad \text{for } m = 0, \dots, P,$$

from the N noisy samples $\tilde{y}_n$ given by (3.1).

2: Form the Toeplitz matrix (3.4) using the measurements $\tilde{s}_m$. The matrix is of dimensions $(P - M + 1) \times (M + 1)$ where $P + 1 \geqslant 2M \geqslant 2K$.

3: Calculate the SVD decomposition of the matrix $\tilde{\mathbf{S}} = \mathbf{U}_{\tilde{\mathbf{S}}}\mathbf{\Delta}_{\tilde{\mathbf{S}}}\mathbf{V}_{\tilde{\mathbf{S}}}^H$. We can write the SVD components:

$$\tilde{\mathbf{S}} = \begin{pmatrix}\mathbf{U}_K & \mathbf{U}_{M-K+1}\end{pmatrix} \cdot \begin{pmatrix}\mathbf{\Delta}_K & \mathbf{0} \\ \mathbf{0} & \mathbf{\Delta}_{M-K+1}\end{pmatrix} \cdot \begin{pmatrix}\mathbf{V}_K^H \\ \mathbf{V}_{P-M-K+1}^H\end{pmatrix}.$$

4: Truncate $\tilde{\mathbf{S}}$ to rank K by using $\tilde{\mathbf{S}}_K = \mathbf{U}_K\mathbf{\Delta}_K\mathbf{V}_K^H$. There are more sophisticated versions of this operation, see [64] for further details.

5: Compute the Toeplitz approximation $\tilde{\mathbf{S}}_{\text{av}}$ to $\tilde{\mathbf{S}}_K$ by averaging over the diagonals. Make this new matrix be $\tilde{\mathbf{S}}$ and repeat from step 3 until some optimality criterion is met. For example the criterion may be that the ratio between the K^{th} and the $(K+1)^{\text{th}}$ singular values of $\tilde{\mathbf{S}}$ is above a certain threshold.

6: Obtain the annihilating filter coefficients by solving the system $\tilde{\mathbf{S}}_{\text{av}}\mathbf{h} \approx \mathbf{0}$.

7: Calculate the roots of the polynomial formed by the filter coefficients $\hat{h}(z) = \sum_{k=0}^{M} h_m z^{-m}$. These yield directly the estimated values $\{\hat{u}_k\}_{k=0}^{K-1}$ by imposing $M = K$. Otherwise, select the K roots inside or on the unit circle for $M > K$.

8: Compute the estimated time locations of the Diracs by using $\hat{t}_k = \frac{T}{\lambda}\ln(\hat{u}_k)$.

9: Finally, obtain the estimated amplitudes $\hat{a}_k$ as the least square solution of the N equations

$$\tilde{y}_n - \sum_{k=0}^{K-1} \hat{a}_k \varphi\left(\frac{\hat{t}_k}{T} - n\right) \approx 0 \qquad n = 0, \dots, N-1.$$

The algorithm assumes the noise added to the moments is white and Gaussian, otherwise see Algorithm 3.

then the SVD decomposition can be written as follows:

$$\mathbf{S} = \underbrace{\begin{pmatrix}\mathbf{U}_K & \mathbf{U}_{M-K+1}\end{pmatrix}}_{\mathbf{U}_\mathbf{S}(d_1\times d_2)} \cdot \underbrace{\begin{pmatrix}\mathbf{\Delta}_K & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}}_{\mathbf{\Delta}_\mathbf{S}(d_2\times d_2)} \cdot \underbrace{\begin{pmatrix}\mathbf{V}_K^H \\ \mathbf{V}_{P-M-K+1}^H\end{pmatrix}}_{\mathbf{V}_\mathbf{S}^H(d_2\times d_2)} \tag{3.6}$$

since $\mathbf{S}$ is of rank K. Here $\mathbf{U}_K$ is $d_1 \times K$, $\mathbf{\Delta}_K$ is $K \times K$ and $\mathbf{V}_K^H$ is $K \times d_2$. The columns of $\mathbf{U}_K$ and of $\mathbf{V}_K$ span the signal subspace. The remaining vectors characterise the orthogonal complement of the signal subspace (or the null space of $\mathbf{S}$). Note also that the noiseless matrix satisfies $\mathbf{S} = \mathbf{U}_K\mathbf{\Delta}_K\mathbf{V}_K^H$.

Importantly, $\mathbf{U}$ and $\mathbf{V}$ satisfy the shift-invariance subspace property [60]. This means that $\overline{\mathbf{U}} = \underline{\mathbf{U}}\mathbf{\Phi}$ and $\overline{\mathbf{V}} = \underline{\mathbf{V}}\mathbf{\Phi}^H$, where the operations $\overline{(\cdot)}$ and $\underline{(\cdot)}$ are for omitting the first and last rows of $(\cdot)$ respectively [1], and $\mathbf{\Phi}$ is a $K \times K$ diagonal matrix with u_k in its main diagonal. This is also true for any of the matrices $\mathbf{UP}$ and $\mathbf{VQ}$ of the other possible factorisations of $\mathbf{S}$. Specifically, $\overline{\mathbf{UP}} = \underline{\mathbf{UP}}\mathbf{P}^{-1}\mathbf{\Phi}\mathbf{P}$ and $\overline{\mathbf{VQ}} = \underline{\mathbf{VQ}}\mathbf{Q}^{-1}\mathbf{\Phi}^H\mathbf{Q}$. Since $\mathbf{P}^{-1}\mathbf{\Phi}\mathbf{P}$ is a similarity transformation on $\mathbf{\Phi}$, the new matrix has the same eigenvalues u_k[2].

Both noiseless matrices $\mathbf{U}_K$ and $\mathbf{V}_K$ are of rank K, the same as $\mathbf{U}$ and $\mathbf{V}$, hence there exist $K \times K$ non-singular matrices $\mathbf{P}$ and $\mathbf{Q}$ such that $\mathbf{U}_K = \mathbf{UP}$ and $\mathbf{V}_K = \mathbf{VQ}$. Thus, matrices $\mathbf{U}_K$ and $\mathbf{V}_K$ satisfy the shift-invariance property, and the values u_k are the eigenvalues of an operator that maps $\underline{\mathbf{U}}_K$ onto $\overline{\mathbf{U}}_K$ or $\underline{\mathbf{V}}_K$ onto $\overline{\mathbf{V}}_K$. This operator is either $\mathbf{Z} = \underline{\mathbf{U}}_K^+\overline{\mathbf{U}}_K$ or $\mathbf{Z} = (\underline{\mathbf{V}}_K^+\overline{\mathbf{V}}_K)^*$. The minimum required size of the data matrix $\mathbf{S}$ in the noiseless case is $(K+1) \times (K+1)$.

We have already mentioned that a key feature of matrix $\mathbf{S}$ is that it is of rank K, property that changes in the presence of noise. Therefore the components of the SVD decomposition of the noisy matrix $\tilde{\mathbf{S}}$ do not satisfy the shift-invariance property. However, consider the presence of additive white Gaussian noise (AWGN) on the measurements. Since this has little effect on the principal singular vectors [69, 70], those corresponding to the K dominant singular values (that is $\tilde{\mathbf{U}}_K$ and $\tilde{\mathbf{V}}_K$) will be good estimates of the singular vectors of the original, noiseless matrix $\mathbf{S}$ ($\mathbf{U}_K$ and $\mathbf{V}_K$). In the presence of noise, the pole estimating operator becomes either $\mathbf{Z} = \underline{\tilde{\mathbf{U}}}_K^+\overline{\tilde{\mathbf{U}}}_K$ or $\mathbf{Z} = (\underline{\tilde{\mathbf{V}}}_K^+\overline{\tilde{\mathbf{V}}}_K)^*$.

Furthermore, the subspace estimator can be used to improve the accuracy on the estimation of the locations $t_k = \frac{T\ln(u_k)}{\lambda}$. The Vandermonde structure of $\mathbf{U}$ and $\mathbf{V}$ allows for a more general version of the shift-invariance property [60], since it is also true that $\overline{\mathbf{U}}^p = \underline{\mathbf{U}}_p\mathbf{\Phi}^p$ and $\overline{\mathbf{V}}^p = \underline{\mathbf{V}}_p(\mathbf{\Phi}^H)^p$, with p a positive integer and where $\overline{(\cdot)}^p$ and $\underline{(\cdot)}_p$ are for omitting the first and last p rows of $(\cdot)$. The main difference is that now matrix $\mathbf{\Phi}^p$ has elements $u_k^p = \mathrm{e}^{\lambda\frac{pt_k}{T}}$ on its main diagonal. Therefore, the advantage of using values of p larger than one is that the separation among the estimated time delays is increased p times. This enhances the resolution capabilities of the original method. However, note

[1] Here $\overline{\mathbf{U}}$ is exactly $\mathbf{U}_1$ when the first *row* is removed, and $\underline{\mathbf{U}}$ is exactly $\mathbf{U}_0$ when the last *row* is removed. The same would apply if columns were removed, but we keep derivations consistent with [60].

[2] From the original relation $\mathbf{\Phi}\mathbf{x} = \lambda\mathbf{x}$ it is also true that $\mathbf{\Phi}\mathbf{P}\mathbf{x}' = \lambda\mathbf{P}\mathbf{x}'$ where $\mathbf{x} = \mathbf{P}\mathbf{x}'$ with $\mathbf{P}$ invertible. Finally, $\mathbf{P}^{-1}\mathbf{\Phi}\mathbf{P}\mathbf{x}' = \lambda\mathbf{x}'$, which means $\mathbf{\Phi}$ are the eigenvalues of $\mathbf{P}^{-1}\mathbf{\Phi}\mathbf{P}$.

that for each computed eigenvalue u_k^p, there exists a set of p possible time delays $t_k + \frac{lT}{p}$, where $l = 0, 1, \ldots, p-1$. This ambiguity may be removed by successive approximations of the locations given by using increasing values of p [60].

3.2.3 Other approaches

Noisy FRI has also been treated in [10]. The authors present a simple algorithm that reduces the noise in the moments and, hence, the estimation error by oversampling. Their idea consists in first separating the samples into their polyphase components, to then calculate the moments of the input from each component independently and average the obtained values. This method is adapted by Berent et al. in [16] to the case of sampling piecewise sinusoidal signals, but without the explicit need for oversampling.

The problem has also been examined from a stochastic modelling perspective by Tan and Goyal [71], using Gibbs sampling. Unlike previous approaches, the sampling kernel plays no fundamental role in the reconstruction algorithm. The authors use Gibbs sampling in order to extract the innovation parameters from their posterior distribution, assuming the noise can be fully modelled, with the goal of minimising the mean square error of the estimated parameters. Their motivation is that their approach effectively circumvents the ill-conditioning of the problem that algebraic methods do not. Erdozain and Crespo build upon Tan and Goyal's work in [72] by embedding Gibbs sampling within the framework of a genetic algorithm, the rationale being that genetic algorithms bring mechanisms to escape local minima in optimisation problems.

Up to our work, using state-of-the-art algebraic techniques [2, 60] for recovering a stream of Diracs from noisy samples taken by a Gaussian kernel led to a reconstruction stage that was very unstable. This normally resulted in inaccurate estimation of the innovation parameters via algebraic methods, even with the help of the preconditioning matrices proposed in [60]. The inaccuracy of the estimation was in part alleviated by the stochastic procedures explained before, but at the expense of requiring much longer execution times. With the use of a better form of prewhitening that we introduce in the next sections and, moreover, with the use of the universal FRI recovery of Chapter 5, we are able to achieve optimal accuracy as well as minimal execution times given by algebraic solutions. Note also that the simple alternatives by Tan and Goyal [71] and Erdozain and Crespo [72] are difficult to generalised for kernels other than the Gaussian function and noises other than additive, white and Gaussian.

3.3 FRI reconstruction in the presence of coloured noise

Consider again Equation (3.2) and assume that the noise measurements b_m are not i.i.d. Gaussian random variables. As a consequence, the entries of the Toeplitz noise matrix $\mathbf{B}$ may no longer have the same variance and may even be correlated. That is, the covariance matrix $\mathbf{R_B} = \mathrm{E}\{\mathbf{B}^H\mathbf{B}\}$ is not a multiple of the identity. In this situation,

SVD is not guaranteed to correctly separate the signal and noise subspaces. Nevertheless, we may still solve the problem if we know or can estimate the covariance matrix of the noise $\mathbf{R_B}$. The reason is that we can recover the appropriate subspaces by considering the SVD of the matrix $\tilde{\mathbf{S}}' = \tilde{\mathbf{S}}\mathbf{W}$, where $\mathbf{W}$ is a whitening matrix such that $\mathbf{W}^H\mathbf{R_B}\mathbf{W} = \mathbf{I}$. "Pre-whitening" is a well known approach proposed by various authors in the spectral estimation community (for instance by De Moor in [63] or Jensen et al. in [64]). A form of preconditioning was also successfully used in the context of FRI by Maravic and Vetterli [60].

There exist various ways to modify the original covariance matrix of the coloured noise. Provided $\mathbf{R_B}$ is positive definite, we can factor it using Cholesky decomposition: $\mathbf{R_B} = \mathbf{Q}^T\mathbf{Q}$. This gives a weighting matrix equal to $\mathbf{W} = \mathbf{Q}^{-1}$ [64]. If the matrix is not positive definite, or is almost rank deficient, we may use a weight equal to the square root of the pseudoinverse of the covariance matrix, i.e. $\mathbf{W} = \mathbf{R}_\mathbf{B}^{\dagger/2}$ [73]. We can perform the operation via the SVD of the noise covariance matrix $\mathbf{R_B}$ by calculating the square root of the inverse of the values (that are greater than zero) of the diagonal matrix of the decomposition. To end, another possibility consists in looking for two matrices $\mathbf{F}$ and $\mathbf{G}$ such that $\tilde{\mathbf{S}}' = \mathbf{FSG}$ is contaminated by noise $\tilde{\mathbf{B}}' = \mathbf{FBG}$ with a covariance matrix with constant diagonal elements [60]. This means that we can think of the entries of the weighted noise $\mathbf{FBG}$ as being samples of Gaussian noise, with the same variance, but not necessarily uncorrelated. The authors use diagonal matrices so that their method is simple: it involves only averaging along rows/columns and avoids matrix inversion, while achieving the desired goal. We do not deal with this approach any further for various reasons: It works well for the Gaussian kernel of [60], since the diagonal matrices are easy to find, however it is not straightforward to use for exponential reproducing kernels. The noise is not de-correlated in general, hence we cannot guarantee that FRI algorithms work optimally on $\tilde{\mathbf{S}}' = \mathbf{FSG}$. And last, TLS-Cadzow and the subspace estimator are more easily adapted to the case of having only a post-multiplication weight: $\tilde{\mathbf{S}}' = \tilde{\mathbf{S}}\mathbf{W}$, as shown next.

3.3.1 Modified TLS-Cadzow algorithm

Suppose we use TLS and Cadzow on the weighted matrix $\tilde{\mathbf{S}}' = \tilde{\mathbf{S}}\mathbf{W}$. Then, once we modify the singular values of $\tilde{\mathbf{S}}'$, we need to revert the effect of weighting. If we need to obtain an approximation to $\tilde{\mathbf{S}}$, we can reconstruct $\tilde{\mathbf{S}} = \tilde{\mathbf{S}}'\mathbf{W}^{-1}$. If, on the other hand, we just want to find the vector $\mathbf{h}$ solution to the minimisation problem, we can find the vector $\mathbf{h_0}$ of the null-space of $\tilde{\mathbf{S}}'$ and then compute $\mathbf{h} = \mathbf{W}^{-1}\mathbf{h_0}$. In any case, the explicit use of $\mathbf{W}$ and its inverse may result in inaccurate data calculations [63]. This can be avoided by using the quotient singular value decomposition (QSVD) of the pair $(\tilde{\mathbf{S}}, \mathbf{W}^{-1})$ [63, 64]. This decomposition is also known as generalised singular value decomposition (GSVD) [59].

QSVD has the advantage that it allows us to use the TLS-Cadzow routine such that the use of the weighting matrix itself $\mathbf{W}$ is not needed. In fact, if we perform the QSVD

on the pair $(\tilde{\mathbf{S}}, \mathbf{W}^{-1})$ we obtain unitary matrices $\mathbf{U}$ and $\mathbf{V}$ and also a non-singular matrix $\mathbf{X}$ so that the following holds [64]:

$$\begin{aligned} \tilde{\mathbf{S}} &= \mathbf{U}\boldsymbol{\Delta}\mathbf{X}^H \\ \mathbf{W}^{-1} &= \mathbf{V}\boldsymbol{\Sigma}\mathbf{X}^H \end{aligned} \tag{3.7}$$

where $\boldsymbol{\Delta} = \text{diag}(\delta_1, \ldots, \delta_{M+1})$ and $\boldsymbol{\Sigma} = \text{diag}(\sigma_1, \ldots, \sigma_{M+1})$, with $\delta_i \geqslant \delta_{i+1}, i = 1, \ldots, M+1$ and $\sigma_i \leqslant \sigma_{i+1}, i = 1, \ldots, M+1$. Then, we can show that formulation (3.7) is equivalent to calculating the SVD of $\tilde{\mathbf{S}}' = \tilde{\mathbf{S}}\mathbf{W}$, by noting that:

$$\tilde{\mathbf{S}}\mathbf{W} = \mathbf{U}\boldsymbol{\Delta}\mathbf{X}^H(\mathbf{V}\boldsymbol{\Sigma}\mathbf{X}^H)^{-1} = \mathbf{U}\boldsymbol{\Delta}\mathbf{X}^H(\mathbf{X}^H)^{-1}\boldsymbol{\Sigma}^{-1}\mathbf{V}^H = \mathbf{U}(\boldsymbol{\Delta}\boldsymbol{\Sigma}^{-1})\mathbf{V}^H. \tag{3.8}$$

Here, we have used $\mathbf{V}^{-1} = \mathbf{V}^H$. The matrices $\mathbf{U}$, $\boldsymbol{\Delta}\boldsymbol{\Sigma}^{-1}$ and $\mathbf{V}$ are identical to those of the SVD of $\tilde{\mathbf{S}}' = \mathbf{U}\mathbf{D}\mathbf{V}^H$, with $\mathbf{D} = \boldsymbol{\Delta}\boldsymbol{\Sigma}^{-1}$. Thus, working with the QSVD of $(\tilde{\mathbf{S}}, \mathbf{W}^{-1})$ is mathematically equivalent to working with the SVD of $\tilde{\mathbf{S}}'$. In fact, the whitening operation is now an integral part of the algorithm, by means of the QSVD operation, and we can apply TLS-Cadzow to $\tilde{\mathbf{S}} = \mathbf{U}\boldsymbol{\Delta}\mathbf{X}^H$ as explained in the box Algorithm 3.

3.3.2 The modified subspace estimator method

We begin the section by noting that, to the best of our knowledge, the way we now modify the subspace estimator of Section 3.2.2 to include prewhitening is a novel contribution to FRI and spectral estimation. We apply a prewhitening transform to the Hankel matrix of moments $\tilde{\mathbf{S}}$ and show that finding the values u_k can be performed directly without the need of reverting prewhitening. It is important to highlight that this modification can be used only when we post-multiply (or pre-multiply) $\tilde{\mathbf{S}}$ by a whitening transform. It can thus *not* be used with the preconditioning step of [60] since the authors employ a modification equal to $\tilde{\mathbf{S}}' = \mathbf{FSG}$.

Moreover, according to our experience, the modified versions of the TLS-Cadzow routine and of the subspace estimator method are similarly effective in the noisy FRI setting, the advantage of the latter that it is not iterative. Therefore, we choose the modified subspace estimator as our retrieval method for the rest of the thesis.

We have already described the subspace method when the measurements are corrupted by AWGN: we obtain the SVD decomposition of $\tilde{\mathbf{S}} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^H$, keep the K columns of $\mathbf{U}$ corresponding to the dominant singular values and compute u_k as the eigenvalues of $\underline{\mathbf{U}}_K^+\overline{\mathbf{U}}_K$, where $\underline{(\cdot)}$ and $\overline{(\cdot)}$ are operations to omit the last and first rows of $(\cdot)$.

Consider now the more general case of $\mathbf{B}$ being due to non-white Gaussian noise. In such situation, we have seen that it is necessary to work with a pre-whitened version of the noisy matrix, i.e. $\tilde{\mathbf{S}}' = \tilde{\mathbf{S}}\tilde{\mathbf{W}}$. We note that we can still recover the appropriate signal poles u_k by applying the subspace estimator directly to $\tilde{\mathbf{S}}'$. In order to verify this fact, first note that the matrix $\mathbf{S}' = \mathbf{SW}$ is still of rank K, provided $\mathbf{W}$ is full-rank. Second,

Algorithm 3 Modified TLS-Cadzow algorithm.

Retrieve the parameters $\{t_k, a_k\}_{k=0}^{K-1}$ of a train of Diracs (2.5) from noisy moments using TLS and Cadzow's iterative routine. This algorithm assumes the noise added to the moments is coloured.

1: Calculate the sequence

$$\tilde{s}_m = \sum_{n=0}^{N-1} c_{m,n}\tilde{y}_n \quad \text{for } m = 0, \dots, P,$$

from the N noisy samples $\tilde{y}_n$ given by (3.1).

2: Choose $M \in [K, P]$ and build the rectangular Toeplitz matrix (3.4) using the measurements $\tilde{s}_m$.

2: Calculate or estimate the noise covariance matrix $\mathbf{R_B}$. Then choose a weighting matrix such as the inverse of the Cholesky factor $\mathbf{W} = \mathbf{Q}^{-1}$ of $\mathbf{R_B} = \mathbf{Q}^T\mathbf{Q}$ or the pseudo-inverse of the square root of the covariance matrix $\mathbf{W} = \mathbf{R_B}^{\dagger/2}$.

3: Perform the QSVD decomposition of $(\tilde{\mathbf{S}}, \mathbf{W}^{-1})$, to obtain matrices $\mathbf{U}$, $\mathbf{\Delta}$, $\mathbf{\Sigma}$, $\mathbf{V}$ and $\mathbf{X}$ as in (3.7). Build the diagonal matrix $\mathbf{\Delta}_K$ keeping only the K largest elements of $\mathbf{\Delta}$, and deduce the total least-squares approximation $\tilde{\mathbf{S}}_K = \mathbf{U}_K\mathbf{\Delta}_K\mathbf{X}_K^H$.

4: Construct the best Toeplitz approximation $\tilde{\mathbf{S}}_{\text{av}}$ by averaging the diagonals of $\tilde{\mathbf{S}}_K$. At any iteration, this new matrix gives a denoised set of moments, $\tilde{s}'_m$. Make $\tilde{\mathbf{S}} = \tilde{\mathbf{S}}_{\text{av}}$ and repeat from step 3 until, for instance, the $(K+1)$th largest diagonal element of $\mathbf{D}$ is smaller than the Kth element by a predefined factor.

5: Obtain the annihilating filter coefficients by solving the system $\tilde{\mathbf{S}}_{\text{av}}\mathbf{h} \approx \mathbf{0}$.

6: Calculate the roots of the polynomial formed by the filter coefficients $\hat{h}(z) = \sum_{k=0}^{M} h_m z^{-m}$. These yield directly the estimated values $\{\hat{u}_k\}_{k=0}^{K-1}$ by imposing $M = K$. Otherwise, select the K roots inside or on the unit circle for $M > K$.

7: Compute the estimated time locations of the Diracs by using $\hat{t}_k = \frac{T}{\lambda}\ln(\hat{u}_k)$.

8: Finally, obtain the estimated amplitudes $\hat{a}_k$ as the least mean square solution of the N equations

$$\tilde{y}_n - \sum_{k=0}^{K-1} \hat{a}_k \varphi\left(\frac{\hat{t}_k}{T} - n\right) \approx 0 \qquad n = 0, \dots, N-1.$$

we have seen that matrix $\mathbf{S}$ can be decomposed as

$$\mathbf{S} = \underbrace{\begin{pmatrix} 1 & 1 & \cdots & 1 \\ u_0 & u_1 & \cdots & u_{K-1} \\ \vdots & \vdots & \ddots & \vdots \\ u_0^{P-M} & u_1^{P-M} & \cdots & u_{K-1}^{P-M} \end{pmatrix}}_{\mathbf{U}} \underbrace{\begin{pmatrix} x_0 & 0 & \cdots & 0 \\ 0 & x_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_{K-1} \end{pmatrix}}_{\mathbf{\Delta}} \underbrace{\begin{pmatrix} 1 & u_0 & \cdots & u_0^M \\ 1 & u_1 & \cdots & u_1^M \\ \vdots & \vdots & \ddots & \vdots \\ 1 & u_{K-1} & \cdots & u_{K-1}^M \end{pmatrix}}_{\mathbf{V}^H},$$

then, $\mathbf{S}' = \mathbf{U}\mathbf{\Delta}\mathbf{V}^H\mathbf{W} = \mathbf{U}\mathbf{\Delta}\mathbf{V}'^H$, where $\mathbf{V}' = \mathbf{W}^H\mathbf{V}$. Moreover, since $\mathbf{S}'$ is still of rank K, then its SVD should look like:

$$\mathbf{S}' = \underbrace{\begin{pmatrix} \mathbf{U}'_K & \mathbf{U}'_{M-K+1} \end{pmatrix}}_{\mathbf{U}_{\mathbf{S}'}} \cdot \underbrace{\begin{pmatrix} \mathbf{\Delta}'_K & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}}_{\mathbf{\Delta}_{\mathbf{S}'}} \cdot \underbrace{\begin{pmatrix} \mathbf{V}'^H_K \\ \mathbf{V}'^H_{P-M-K+1} \end{pmatrix}}_{\mathbf{V}^H_{\mathbf{S}'}}. \tag{3.9}$$

Equivalently, the SVD may be written in an "economic" form as $\mathbf{S}' = \mathbf{U}'_K\mathbf{\Delta}'_K\mathbf{V}'^H_K$.

Then, it is possible to find matrices $\mathbf{P}$ and $\mathbf{Q}$ of size $K \times K$ that are invertible such that the following holds:

$$\mathbf{S}' = \underbrace{(\mathbf{U}\mathbf{P})}_{\mathbf{U}'_K} \underbrace{(\mathbf{P}^{-1}\mathbf{\Delta}\mathbf{Q})}_{\mathbf{\Delta}'_K} \underbrace{(\mathbf{Q}^{-1}(\mathbf{W}^H\mathbf{V})^H)}_{\mathbf{V}'^H_K}. \tag{3.10}$$

Note that now it is only $\mathbf{U}$ that still satisfies the shift-invariance property: $\overline{\mathbf{U}} = \underline{\mathbf{U}}\mathbf{\Phi}$. The same relation does not apply to $\mathbf{V}' = \mathbf{W}^H\mathbf{V}$ any more.[3] Consequently, also $\mathbf{U}_K = \mathbf{U}\mathbf{P}$ satisfies the property, i.e. $\overline{\mathbf{U}\mathbf{P}} = \underline{\mathbf{U}\mathbf{P}}\mathbf{P}^{-1}\mathbf{\Phi}\mathbf{P}$.

To conclude, then, matrix $\mathbf{U}'_K$ satisfies the shift-invariance property, and the values u_k are the eigenvalues of an operator that maps $\overline{\mathbf{U}'_K}$ to $\underline{\mathbf{U}'_K}$. This operator is $\mathbf{Z} = \underline{\mathbf{U}'_K}^{+}\overline{\mathbf{U}'_K}$. The fact that the parameters u_k can be obtained without inverting the whitening transform $\mathbf{W}$ is a very useful feature of the subspace estimator method. We summarise the main steps of this routine in Algorithm 4.

3.4 Measuring the performance: The Cramér–Rao lower bound

In order to analyse the effect of noise on the accuracy with which FRI signals can be recovered we may use the Cramér–Rao lower bound (CRB). This is a lower bound on the mean square error (MSE) achievable by any unbiased estimator [74]. As such, it provides a measure of the difficulty of a given estimation problem, and can indicate whether or

[3] In [60] the authors prove the shift-invariance property for $\mathbf{U}\mathbf{P}$. From this proof, it immediately follows that only when $\mathbf{U}$ satisfies the shift-invariance property and is *post*-multiplied by an invertible matrix, then $\mathbf{U}\mathbf{P}$ satisfies the property too. If, on the other hand, the matrix is *pre*-multiplied by an invertible matrix, the property does not hold.

Algorithm 4 Modified subspace estimator method.

Retrieve the parameters $\{t_k, a_k\}_{k=0}^{K-1}$ of a train of Diracs (2.5) from noisy moments using TLS and Cadzow's iterative routine. This algorithm assumes the noise added to the moments is coloured.

1: Calculate the sequence

$$\tilde{s}_m = \sum_{n=0}^{N-1} c_{m,n}\tilde{y}_n \quad \text{for } m = 0, \dots, P,$$

from the N noisy samples $\tilde{y}_n$ given by (3.1).

2: Then, choose $M \in [K, P]$ and build the rectangular Toeplitz $\tilde{\mathbf{S}}$ with the sequence $\tilde{s}_m$. Here $\tilde{\mathbf{S}} = \mathbf{S} + \mathbf{B}$.

2: Estimate $\mathbf{R_B} = \mathrm{E}\{\mathbf{B}^H\mathbf{B}\}$ and define the new matrix $\tilde{\mathbf{S}}' = \tilde{\mathbf{S}}\mathbf{W}$, where $\mathbf{W} = \mathbf{R}_{\mathbf{B}}^{-\dagger/2}$.

3: Apply the subspace estimator method to $\tilde{\mathbf{S}}'$: Obtain the decomposition $\tilde{\mathbf{S}}' = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H$, keep the K columns of $\mathbf{U}$ corresponding to the K dominant singular values and estimate u_k as the eigenvalues of $\underline{\mathbf{U}}_K^{+}\overline{\mathbf{U}}_K$. Here, $\underline{(\cdot)}$ and $\overline{(\cdot)}$ are operations to omit the last and first rows of $(\cdot)$.

4: Compute the estimated time locations of the Diracs by using $\hat{t}_k = \frac{T}{\lambda}\ln(\hat{u}_k)$.

5: Finally, obtain the estimated amplitudes $\hat{a}_k$ as the least mean square solution of the N equations

$$\tilde{y}_n - \sum_{k=0}^{K-1} \hat{a}_k \varphi\left(\frac{\hat{t}_k}{T} - n\right) \approx 0 \qquad n = 0, \dots, N-1.$$

not existing techniques come close to optimal. It can also be used to measure the relative merit of different types of estimation algorithms.

FRI signals are completely characterised by their innovation parameters. For instance, a stream of K Diracs is completely specified from the locations $\{t_k\}_{k=0}^{K-1}$ and amplitudes $\{a_k\}_{k=0}^{K-1}$. Consider the vector $\Theta = (t_0, \dots, t_{K-1}, a_0, \dots, a_{K-1})^T$; then the goal is to estimate Θ from the vector of N noisy samples $\tilde{\mathbf{y}} = (\tilde{y}_0, \dots, \tilde{y}_{N-1})^T$ given by (3.1). For simplicity we assume the sampling period is $T = 1$. A way to determine the CRB of this estimation problem was given in [14] assuming ϵ_n is a zero-mean Gaussian noise with covariance matrix $\mathbf{R} = \mathrm{E}\{\mathbf{e}\mathbf{e}^H\}$, where $\mathbf{e}$ is a vector of length N with values ϵ_n. In the set-up of [14] it was shown that any unbiased estimate of the unknown parameters $\hat{\Theta}(\tilde{\mathbf{y}}) = (\hat{t}_0, \dots, \hat{t}_{K-1}, \hat{a}_0, \dots, \hat{a}_{K-1})^T$ has a covariance matrix that is lower bounded by

$$\mathrm{cov}(\hat{\Theta}) \geqslant (\Phi_{\mathbf{y}}^T \mathbf{R}^{-1} \Phi_{\mathbf{y}})^{-1}, \tag{3.11}$$

where the matrix $\Phi_{\mathbf{y}}$ is given by[4]

$$\Phi_{\mathbf{y}} = \left(\begin{array}{ccc|ccc} a_0\varphi'(t_0) & \dots & a_{K-1}\varphi'(t_{K-1}) & \varphi(t_0) & \dots & \varphi(t_{K-1}) \\ a_0\varphi'(t_0-1) & \dots & a_{K-1}\varphi'(t_{K-1}-1) & \varphi(t_0-1) & \dots & \varphi(t_{K-1}-1) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_0\varphi'(t_0-(N-1)) & \dots & a_{K-1}\varphi'(t_{K-1}-(N-1)) & \varphi(t_0-(N-1)) & \dots & \varphi(t_{K-1}-(N-1)) \end{array}\right). \tag{3.12}$$

While this is one possible way to measure the performance of various FRI recovery techniques [14, 60] given the noisy samples $\tilde{\mathbf{y}}$, we also note that if we use kernels that reproduce exponentials, we need the sequence of moments $\tilde{\mathbf{s}} = \mathbf{C}\tilde{\mathbf{y}}$. Then the goal is to estimate Θ from the vector of $P+1$ noisy measurements $\tilde{\mathbf{s}} = (\tilde{s}_0, \dots, \tilde{s}_{N-1})^T$ given by (3.2). It is therefore of interest to find the CRB associated to the measurements $\tilde{\mathbf{s}}$ of (3.2), since this will indicate the best performance that can be achieved when working with $\tilde{\mathbf{s}}$. In this context, however, expression (3.11) needs to be generalised [75]:

$$\mathrm{cov}(\hat{\Theta}) \geqslant (\Phi_{\mathbf{s}}^H \mathbf{R}_{\mathbf{b}}^{-1} \Phi_{\mathbf{s}})^{-1}, \tag{3.13}$$

where $\mathbf{R}_{\mathbf{b}} = \mathrm{E}\{\mathbf{b}\mathbf{b}^H\}$ and now we use the Hermitian transpose of matrix $\Phi_{\mathbf{s}}$. Here, $\mathbf{b}$ is the vector of $P+1$ noisy samples b_m and now the matrix $\Phi_{\mathbf{s}}$ takes the form:

$$\Phi_{\mathbf{s}} = \left(\begin{array}{ccc|ccc} a_0\alpha_0 \mathrm{e}^{\alpha_0 t_0} & \dots & a_{K-1}\alpha_0 \mathrm{e}^{\alpha_0 t_{K-1}} & \mathrm{e}^{\alpha_0 t_0} & \dots & \mathrm{e}^{\alpha_0 t_{K-1}} \\ a_0\alpha_1 \mathrm{e}^{\alpha_1 t_0} & \dots & a_{K-1}\alpha_1 \mathrm{e}^{\alpha_1 t_{K-1}} & \mathrm{e}^{\alpha_1 t_0} & \dots & \mathrm{e}^{\alpha_1 t_{K-1}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_0\alpha_P \mathrm{e}^{\alpha_P t_0} & \dots & a_{K-1}\alpha_P \mathrm{e}^{\alpha_P t_{K-1}} & \mathrm{e}^{\alpha_P t_0} & \dots & \mathrm{e}^{\alpha_P t_{K-1}} \end{array}\right). \tag{3.14}$$

Note that for general complex values $\tilde{\mathbf{s}}$, the covariance matrix $\mathbf{R}_{\mathbf{b}}$ may not contain all the information regarding the real and imaginary parts of $\mathbf{b}$. This can be addressed by using

[4] The matrix can be obtained calculating the derivative of $\tilde{y}_n$ with respect to each parameter in Θ. That is, the columns of $\Phi_{\mathbf{y}}$ to the left of | are $\frac{\partial \tilde{y}_n}{\partial t_k}$ and the columns of $\Phi_{\mathbf{y}}$ to the right of | are $\frac{\partial \tilde{y}_n}{\partial a_k}$.

augmented vectors formed by stacking $\tilde{\mathbf{s}}$ and its complex conjugate [76, 77]. But this is not an issue in our case, since we use exponential parameters α_m that exist in conjugate pairs, such that $\tilde{\mathbf{s}}$ is formed with values that are complex conjugate of each other.

Furthermore, since we have assumed that the noise ϵ_n added to the samples is additive, white and Gaussian, with variance σ^2, the covariance matrix of the filtered noise is $\mathbf{R_b} = \mathrm{E}\{\mathbf{b}\mathbf{b}^H\} = \sigma^2\mathbf{C}\mathbf{C}^H$. Consequently, there is a direct relation between (3.11) and (3.13), that can be expressed through the matrix $\mathbf{C}$. In order to see this, note that $\tilde{\mathbf{s}} = \mathbf{C}(\mathbf{y}+\mathbf{e}) = \mathbf{s}+\mathbf{b}$ and also that $\Phi_\mathbf{s} = \mathbf{C}\Phi_\mathbf{y}$. Moreover, we know that

$$\mathrm{cov}(\hat{\Theta}(\tilde{\mathbf{y}})) \geqslant (\Phi_\mathbf{y}^H\mathbf{R}^{-1}\Phi_\mathbf{y})^{-1} = \sigma^2(\Phi_\mathbf{y}^H\Phi_\mathbf{y})^{-1},$$

since the noise ϵ_n is AWGN. Thus, it is true that

$$\mathrm{cov}(\hat{\Theta}(\tilde{\mathbf{s}})) \geqslant (\Phi_\mathbf{s}^H\mathbf{R}_\mathbf{b}^{-1}\Phi_\mathbf{s})^{-1} = \sigma^2(\Phi_\mathbf{y}^H\mathbf{C}^H(\mathbf{C}\mathbf{C}^H)^{-1}\mathbf{C}\Phi_\mathbf{y})^{-1} = \sigma^2(\Phi_\mathbf{y}^H\mathbf{C}^\dagger\mathbf{C}\Phi_\mathbf{y})^{-1},$$

where $\mathbf{C}^\dagger = \mathbf{C}^H(\mathbf{C}\mathbf{C}^H)^{-1}$ is the left pseudo-inverse of $\mathbf{C}$. When the number of moments $P+1$ equals the number of samples N, then both formulations are equivalent. This is easily seen since in such case $\mathbf{C}^\dagger\mathbf{C} = \mathbf{C}^{-1}\mathbf{C} = \mathbf{I}_N$. Intuitively, this is the optimal configuration because no linear combination of the samples should improve the estimation based on the samples themselves. Retrieving the innovation parameters through the moments when $P+1 < N$ is instead suboptimal.

In Appendix B.1 we provide closed form expressions of the moment-based bound for the case of a single exponential. We compare the case of having additive white Gaussian noise with the case of having uncorrelated noise. We compare these since it is easy to find closed form expressions, but also because they can be obtained from the same set of parameters $\alpha_m = j\frac{\pi}{N}(2m-P)$ for $m = 0, \ldots, P$, the former being characterised by $|c_{m,0}| = 1$ for all m (see Appendix C.5).

3.5 Simulations

We now present simulation results for exponential reproducing kernels in the presence of noise. More specifically, we analyse the performance of E-Splines of various orders for retrieving a train of Diracs using the subspace estimator method of Section 3.2, compared to the modified version of Section 3.3 that uses prewhitening. We show that the latter improves the results of the former for any experiment and, moreover, is optimal according to the bound predicted by (3.13). Therefore, our goals are to show that our version of the subspace estimator adapted to sampling with exponential reproducing kernels is the best we can do to reach the moments based CRB.

We do not present simulations using the TLS-Cadzow routine or its modified version, since results are equivalent to those given by the subspace estimator.

3.5.1 The experimental setup

We take $N = 31$ samples following Figure 1.1 by computing $y_n = \sum_{k=0}^{K-1} a_k \varphi\left(\frac{t_k}{T} - n\right)$ for $n = 0, \ldots, N-1$ and $k = 0, \ldots, K-1$, since the input is a train of K Diracs. The sampling period is $T = \frac{1}{N}$. We corrupt the vector of samples $\mathbf{y}$ with additive white Gaussian noise of fixed variance σ^2. This is related to the signal-to-noise ratio for the samples, defined as $\text{SNR(dB)} = 10 \log \frac{\|\mathbf{y}\|^2}{N\sigma^2}$. We finally calculate the noisy $P+1$ moments (3.2) and retrieve the innovation parameters $\{a_k, t_k\}_{k=0}^{K-1}$ of the input, using the subspace estimator method of Sections 3.2 and 3.3.

We are mainly interested in obtaining the error in the estimation of the time locations, since these are the most challenging parameters to retrieve. For each Dirac, we show the standard deviation of this error:

$$\Delta t_k = \sqrt{\frac{\sum_{i=0}^{I-1} (\hat{t}_k^{(i)} - t_k)^2}{I}} \qquad k = 0, \ldots, K-1, \tag{3.15}$$

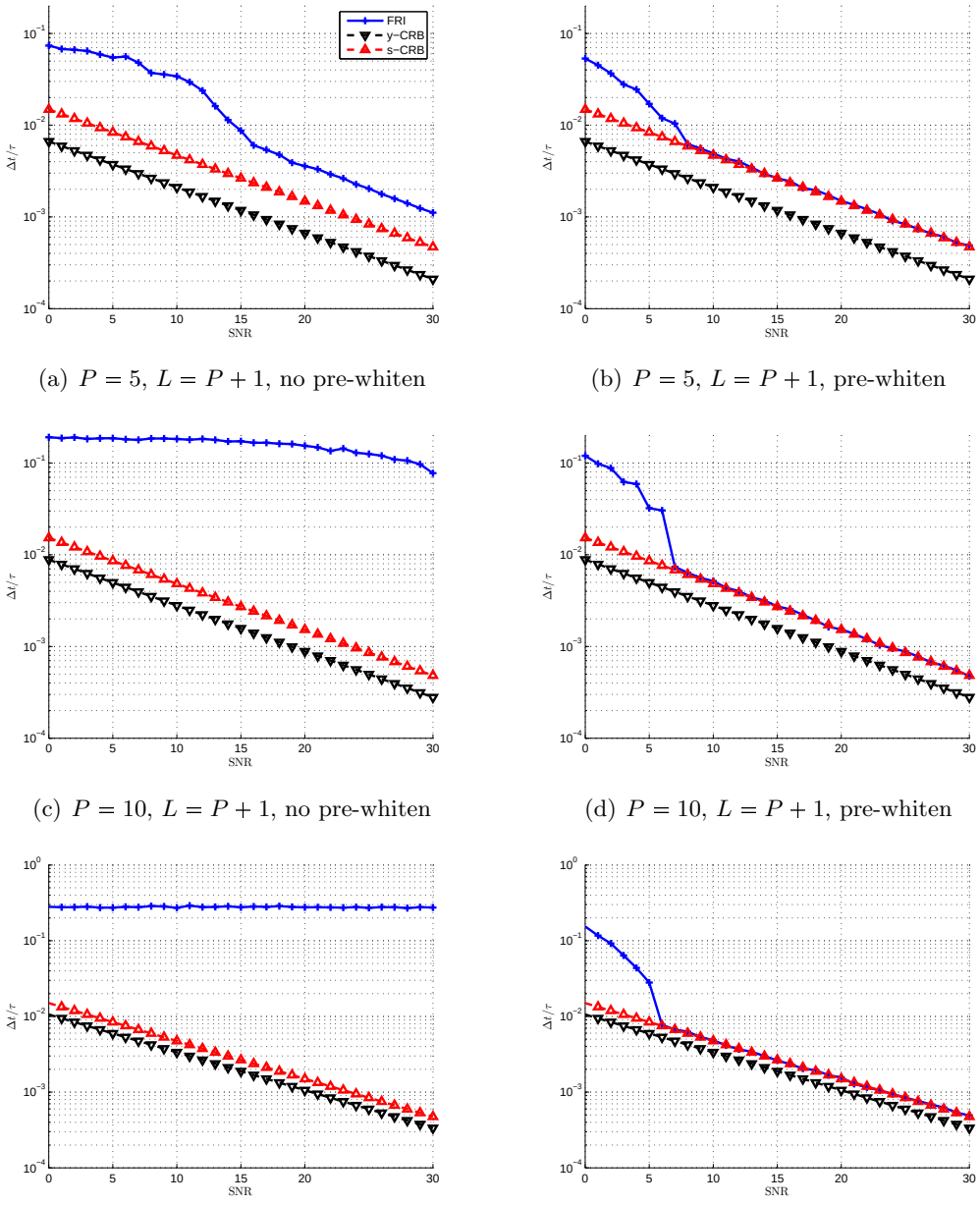
where $\hat{t}_k^{(i)}$ are the estimated time locations at experiment i and I is the total number of realisations. We calculate (3.15) for a range of fixed signal-to-noise ratios and average the effects using $I = 1000$ noise realisations for each SNR. We compare the performance (3.15) with the square root of the variance predicted by the two different Cramér–Rao bounds (CRB) of Section 3.4: the sample-based CRB (3.11) and the moment-based CRB (3.13).

3.5.2 Results

In Figure 3.2 (a, c, e) we present simulation results when we retrieve $K = 1$ Diracs using E-Spline kernels and the subspace estimator method of Section 3.2. Then, in Figure 3.2 (b, d, f) we present the results obtained when we retrieve the same Dirac using the modified method of Section 3.3. For any order $P+1$ the latter method, which uses prewhitening, improves the accuracy of the estimated location for every SNR, and reaches the moment-based CRB predicted by (3.13) (in red, denoted s-CRB in the legend). Note that both the estimation error and the moment-based CRB get closer to the sample-based CRB (in black, and denoted y-CRB in the legend) as the value of $P+1$ increases. The exponential parameters are equal to $\alpha_m = j\frac{\pi}{P+1}(2m - P)$ with $m = 0, \ldots, P$.

3.6 Summary

In this chapter we have examined various FRI procedures to estimate the innovation parameters (t_k, a_k) of a train of K Diracs in the presence of noise. We have limited our analysis to the scenario of adding white Gaussian noise to the samples y_n taken by the sampling kernel. Due to the fact that we mainly use exponential reproducing kernels, we have presented modified versions of existing denoising methods that appeared before in the literature. Specifically, we have adapted the TLS-Cadzow routine of [14] and the



(a) $P = 5$, $L = P + 1$, no pre-whiten

(b) $P = 5$, $L = P + 1$, pre-whiten

(c) $P = 10$, $L = P + 1$, no pre-whiten

(d) $P = 10$, $L = P + 1$, pre-whiten

(e) $P = 15$, $L = P + 1$, no pre-whiten

(f) $P = 15$, $L = P + 1$, pre-whiten

Figure 3.2: ***Performance of E-Spline kernels.*** **We show the performance of E-Spline kernels of parameters $\alpha_m = j\frac{\pi}{P+1}(2m - P)$ with $m = 0, \dots, P$ for different orders $P + 1$ when noise is added to the samples. (a, c, e) are the errors in the estimation of the time location of $K = 1$ Dirac with the subspace estimator method of Section 3.2. (b, d, f) are for the modified method of Section 3.3. For any order the latter method improves the accuracy of the estimated location, and reaches the moment-based CRB predicted by** (3.13)**.**

subspace estimator method suggested in [60] to working with coloured noise added to the moments s_m.

The way we modify the TLS-Cadzow routine of [14] is new to FRI, but is well known in the spectral estimation literature [63, 64]. On the other hand, the way we modify the subspace estimator of [60] is, to the best of our knowledge, a unique contribution of this thesis to FRI and spectral estimation.

In addition, we have used the CRB to measure the difficulty of estimating the innovation parameters of the input from the samples y_n, as suggested in [14]. Moreover, we have introduced a performance measure that is better suited to our retrieval scenario: the CRB lower bound associated to the exponentials that compose the power sum series (2.21), which is a novel contribution to FRI. With such bound, we are able to determine whether the FRI retrieval procedures presented throughout the chapter come close to optimal, i.e. if they reach the CRB.

We have validated our proposed algorithms through simulations and found that it is only the modified versions of the state-of-the-art denoising methods that are able to reach the CRB.

Chapter 4

Optimising noisy FRI recovery

In the previous chapter we have studied the noisy FRI scenario and reviewed the main techniques used in the FRI literature to deal with measurements that are not ideal. We have also adapted the two main subspace based techniques, i.e. the total least squares and Cadzow approach and the subspace estimator method, to behave optimally in the presence of coloured noise. Optimality is in the sense that the modified techniques reach the moment-based CRB whereas the original methods do not. In this chapter we use a different perspective to analyse the noisy scenario even further: we optimise the sampling kernel in order to make the retrieval of the parameters of a train of Diracs most accurate and stable. We restrict our attention to exponential reproducing kernels, and determine the best kernels in this family for noisy FRI retrieval.

We begin the chapter in Section 4.1 with an analysis of the various sources of instability for FRI recovery. We end the section with a practical method to select the best exponential parameters for our FRI problem, which translates into selecting a proper matrix of coefficients $\mathbf{C}$. Then, in Section 4.2 we apply the conclusions of the previous section to design exponential reproducing kernels of maximum order and minimum support (eMOMS) that are most robust to white Gaussian noise added to the samples. We then derive a more general family of eMOMS that may be built from other non-optimal matrices $\mathbf{C}$. We conclude the section by computing the CRB associated to the estimation problem of retrieving the parameters of a single Dirac from the samples taken by these kernels. In Section 4.3 we provide simulations that prove these new kernels outperform E-Splines and are always able to reach the moment-based CRB and then conclude the chapter in Section 4.4.

4.1 Sources of instability for FRI reconstruction

The FRI recovery process is in all equivalent to finding the parameters x_k and u_k of the power sum series

$$\tilde{s}_m = \sum_{k=0}^{K-1} x_k u_k^m + b_m, \tag{4.1}$$

for $m = 0 \ldots, P$ and where $x_k = a_k \mathrm{e}^{\alpha_0 \frac{t_k}{T}}$ and $u_k = \mathrm{e}^{\lambda \frac{t_k}{T}}$ with $k = 0, \ldots, K-1$. In what follows we analyse the sensitivity of the problem under noise b_m that is not necessarily white and establish the parameters that lead to most stable reconstruction.

4.1.1 Parametric retrieval from the noisy series

In the previous chapters we have established that in order to obtain the parameters x_k and u_k from the above series, we may construct a system of equations $\tilde{\mathbf{S}}'\mathbf{h} \approx \mathbf{0}$ and solve the system using for instance the annihilating filter or the subspace estimator methods. Here, $\tilde{\mathbf{S}}' = (\mathbf{S} + \mathbf{B})\mathbf{W}$ where $\mathbf{S}$ and $\mathbf{B}$ are Hankel. Note that $\mathbf{S}$ can be decomposed as follows: $\mathbf{S} = \mathbf{U} \cdot \mathbf{\Delta} \cdot \mathbf{V}^H$, where $\mathbf{U}$ and $\mathbf{V}$ are Vandermonde matrices with nodes u_k and $\mathbf{\Delta}$ is diagonal with elements x_k. Complex square Vandermonde matrices are perfectly conditioned when the generating elements u_k are uniformly spread on the unit circle [78]. Rectangular matrices, on the other hand, may be well conditioned provided the nodes are close to the unit circle, sufficiently separated from each other and when the number of rows is large enough [79]. In addition, the matrix of coefficients $\mathbf{C}$ plays a fundamental role since the whitening transform $\mathbf{W}$ is related to the covariance matrix of $\mathbf{B}$, for instance $\mathbf{W} = \mathbf{R}_{\mathbf{B}}^{-1/2}$, and this is turn is directly related to $\mathbf{C}$ (see Appendix C.5).

The sensitivity of the estimation problem can be derived by inspecting $\tilde{\mathbf{S}}'\mathbf{h} \approx \mathbf{0}$. Whenever any of the matrices that compose $\mathbf{S}'$ are ill-conditioned, then the estimation problem becomes ill-conditioned as well. The condition numbers of $\mathbf{U}$ and $\mathbf{V}$ grow exponentially large with K when the nodes $u_k = \mathrm{e}^{\lambda \frac{t_k}{T}}$ are not on the unit circle [51]. This source of instability is inherent to the exponential parameters α_m. When we use parameters that exist in complex conjugate pairs:

$$\alpha_m = \alpha + j\frac{\pi}{L}(2m - P), \qquad m = 0, \ldots, P, \tag{4.2}$$

for the kernel to be real valued, then $\alpha_0 = \alpha - j\frac{\pi P}{L}$ and $\lambda = j\frac{2\pi}{L}$, that is, λ is purely imaginary. This implies that the nodes u_k always lie on the unit circle. Ideally, the nodes should be such that they span the unit circle, however their distance in the complex plane is determined by the time separation of the Diracs, the sampling period T and the parameter L. As such, this distance is normally fixed for a given sampling setup, the only free parameter being L. The condition number of $\mathbf{\Delta}$ is related to the ratio between the largest and the smallest coefficients $|x_k| = |a_k \mathrm{e}^{\alpha_0 \frac{t_k}{T}}|$. This translates into the fact that the more similar the coefficients, the better the conditioning of $\mathbf{\Delta}$. Thus, if α_0 is also purely

imaginary we are guaranteed that the original ratio $\frac{|a_{\max}|}{|a_{\min}|}$ is preserved. The condition of $\mathbf{W}$ is not straightforward to determine but we know that it directly depends on the condition of matrix of coefficients $\mathbf{C}$. Therefore, we now study the form of $\mathbf{C}$ in more detail.

4.1.2 Choice of matrix C

The first step in the FRI reconstruction stage is to transform the vector of samples $\mathbf{y}$ into the vector of moments $\mathbf{s} = \mathbf{C}\mathbf{y}$, therefore, our first aim is to get a well conditioned $\mathbf{C}$.

Matrix $\mathbf{C}$ is composed of elements $c_{m,n} = c_{m,0}\mathrm{e}^{\alpha_m n}$ at position (m, n), where $n = 0, \ldots, N-1$ and $m = 0, \ldots, P$. Therefore, we may decompose it as follows:

$$\underbrace{\begin{pmatrix} c_{0,0} & c_{0,0}\mathrm{e}^{\alpha_0} & \cdots & c_{0,0}\mathrm{e}^{\alpha_0(N-1)} \\ c_{1,0} & c_{1,0}\mathrm{e}^{\alpha_1} & \cdots & c_{1,0}\mathrm{e}^{\alpha_1(N-1)} \\ \vdots & \vdots & \ddots & \vdots \\ c_{P,0} & c_{P,0}\mathrm{e}^{\alpha_P} & \cdots & c_{P,0}\mathrm{e}^{\alpha_P(N-1)} \end{pmatrix}}_{\mathbf{C}} = \underbrace{\begin{pmatrix} c_{0,0} & 0 & \cdots & 0 \\ 0 & c_{1,0} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & c_{P,0} \end{pmatrix}}_{\mathbf{D}} \underbrace{\begin{pmatrix} 1 & \mathrm{e}^{\alpha_0} & \cdots & \mathrm{e}^{\alpha_0(N-1)} \\ 1 & \mathrm{e}^{\alpha_1} & \cdots & \mathrm{e}^{\alpha_1(N-1)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \mathrm{e}^{\alpha_P} & \cdots & \mathrm{e}^{\alpha_P(N-1)} \end{pmatrix}}_{\mathbf{V}},$$

where $\mathbf{D}$ is diagonal and $\mathbf{V}$ Vandermonde. Hence, to have a stable $\mathbf{C}$ we want the absolute values of the diagonal elements of $\mathbf{D}$ to be the same, for instance $|c_{m,0}| = 1$. Moreover, we want the elements in $\mathbf{V}$ to lie on the unit circle, hence we select:

$$\mathrm{e}^{\alpha_m n} = \mathrm{e}^{j\frac{\pi}{L}(2m-P)n} \qquad \text{for } m = 0, \ldots, P, \text{ i.e. } \alpha = 0, \tag{4.3}$$

where L is a free parameter that allows us to control the separation in between consecutive elements of $\mathbf{V}$.

Purely imaginary α_m make the Vandermonde matrix $\mathbf{V}$ better conditioned [78]. We are therefore only left with the problem of finding the best L in (4.3). Since we have experimentally seen that FRI algorithms are able to reach the moment CRB (3.13) if $\mathbf{C}$ is well conditioned, one way to determine L is to choose the value that minimises (3.13) for the location of a single Dirac. We observe that the minimum is achieved when $L = P+1$, as shown in Figure 4.1 for various choices of P and L, given $|c_{m,0}| = 1$ for all m. Even though we have no mathematical proof, we believe this to be a general phenomenon. The reason is that, when $|c_{m,0}| = 1$ for all m, the noise added to the moments is approximately white regardless of L (but more exactly the closer L is to N) and, as we show in Appendix B.1, the moment-based CRB is then minimised when $L = P + 1$.

In addition, this choice ensures that the exponentials span the entire unit circle, which is well known to be the best configuration when recovering the parameters of a power series [51]. Finally, if we impose $P + 1 = N$, besides minimising (3.13), we also ensure that the moment-based CRB in (3.13) matches the sample-based bound in (3.11), leading to the best possible performance. In this situation, the matrix $\mathbf{C}$ is square and unitary. This is the most stable numerical transformation since its condition number is one.

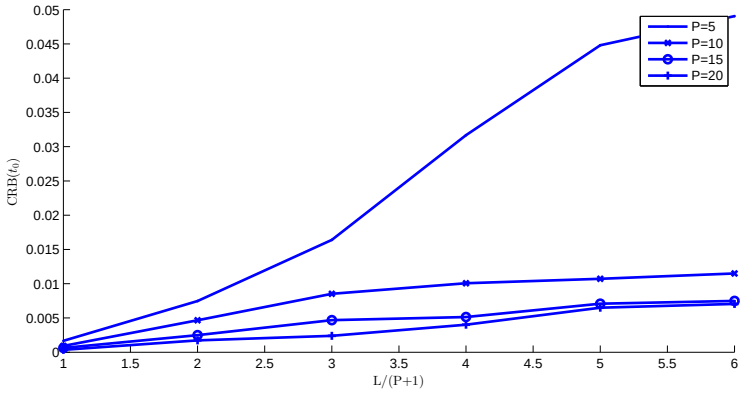


Figure 4.1: *CRB vs.* L. Here we plot various CRB values (3.13) **($\sigma = 1$) for coefficients satisfying $|c_{m,0}| = 1$, $m = 0, \dots, P$ when we vary L in equation** (4.2)**, $\alpha = 0$. For any value of P the CRB is minimised when $L = P + 1$ (note that all the lines are monotonically increasing).**

In summary, the best exponential reproducing kernels should reproduce exponentials with exponents of the form $\alpha_m = j\frac{\pi}{P+1}(2m - P)$ provided $|c_{m,0}| = 1$ for $m = 0, \dots, P$. Finally, whenever possible, the order of the kernel (which equals the number of moments) should be $P + 1 = N$. However, there are instances in which the condition $P + 1 = N$ cannot be imposed. Thus, in the next section we show how to obtain the best kernels but we require no constraint on $P + 1$.

4.2 Exponential MOMS

Equipped with the analysis of the previous section we now design optimal exponential reproducing kernels of maximum-order and minimum-support (eMOMS). As discussed before, we require $|c_{m,0}| = 1$ for $m = 0, \dots, P$ and the exponential parameters to be of the form:

$$\alpha_m = j\omega_m = j\frac{\pi}{P+1}(2m - P) \qquad m = 0, \dots, P. \tag{4.4}$$

To start, we note that by using (2.7) we have

$$\begin{aligned} 1 &= c_{m,0} \sum_{n\in\mathbb{Z}} \mathrm{e}^{\alpha_m(n-t)}\varphi(t-n) \\ &\overset{(a)}{=} c_{m,0} \sum_{k\in\mathbb{Z}} \hat{\varphi}(\alpha_m + j2\pi k)\mathrm{e}^{j2\pi kt} \\ &\overset{(b)}{=} c_{m,0}\hat{\varphi}(\alpha_m), \end{aligned}$$

where (a) follows from Poisson summation formula (1) and (b) from the application of the generalised Strang-Fix conditions (2.8). Therefore, we have that for any exponential reproducing kernel $c_{m,0} = \hat{\varphi}(\alpha_m)^{-1}$, where $\hat{\varphi}(\alpha_m)$ is the Laplace transform of $\varphi(t)$ evaluated at α_m, and $c_{m,n} = c_{m,0}\mathrm{e}^{\alpha_m n}$.

By taking into account that an exponential reproducing kernel $\varphi(t)$ can be written as $\varphi(t) = \gamma(t) * \beta_{\vec{\alpha}}(t)$, we design $\gamma(t)$ so that $|c_{m,0}| = 1$ is satisfied for $m = 0, \dots, P$. Since we know that $c_{m,0} = \hat{\varphi}(\alpha_m)^{-1}$, we then realise that imposing $|c_{m,0}| = 1$ is equivalent to requiring $|\hat{\varphi}(\alpha_m)| = 1$. Finally, by using $\hat{\varphi}(\alpha_m) = \hat{\gamma}(\alpha_m)\hat{\beta}_{\vec{\alpha}}(\alpha_m)$ and evaluating the Laplace transforms at $\alpha_m = j\omega_m$, we arrive at the following condition on $\hat{\gamma}(\omega_m)$:

$$|\hat{\varphi}(\omega_m)| = |\hat{\gamma}(\omega_m)\hat{\beta}_{\vec{\alpha}}(\omega_m)| = 1 \quad \leftrightarrow \quad |\hat{\gamma}(\omega_m)| = |\hat{\beta}_{\vec{\alpha}}(\omega_m)|^{-1}, \tag{4.5}$$

where we now work with the Fourier transform of each function (we have used $\alpha_m = j\omega_m$).

Among all the admissible kernels satisfying (4.5), we are interested in the one with the shortest support $P+1$. We thus consider the kernels given by a linear combination of various derivatives of the original E-Spline $\beta_{\vec{\alpha}}(t)$, i.e.:

$$\varphi(t) = \sum_{\ell=0}^{P} d_\ell \beta_{\vec{\alpha}}^{(\ell)}(t), \tag{4.6}$$

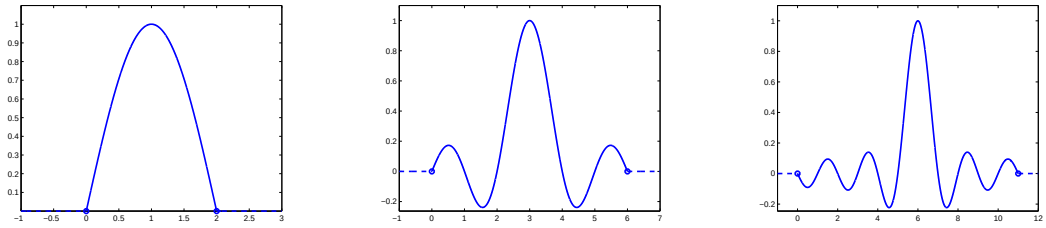
where $\beta_{\vec{\alpha}}^{(\ell)}(t)$ is the ℓth derivative of $\beta_{\vec{\alpha}}(t)$, with $\beta_{\vec{\alpha}}^{(0)}(t) = \beta_{\vec{\alpha}}(t)$, and d_ℓ is a set of coefficients. This is like saying that $\gamma(t)$ is a linear combination of the Dirac delta function and its derivatives, up to order P [80]. These kernels are still able to reproduce the exponentials $\mathrm{e}^{\alpha_m t}$ and are a variation of the maximum-order minimum-support (MOMS) kernels introduced in [42]. This is why we call them exponential MOMS (or eMOMS). They are also a specific case of the broader family of generalised E-Splines presented in [41] (and that we briefly reviewed in Chapter 2). The advantage of this formulation is twofold: First, the modified kernel $\varphi(t)$ is of minimum support $P+1$, the same as that of $\beta_{\vec{\alpha}}(t)$. Second, we only need to find the coefficients d_ℓ that meet the constraint (4.5), in order to achieve $|c_{m,0}| = 1$. Using the Fourier transform of (4.6), which is given by:

$$\hat{\varphi}(\omega) = \hat{\beta}_{\vec{\alpha}}(\omega) \sum_{\ell=0}^{P} d_\ell (j\omega)^\ell,$$

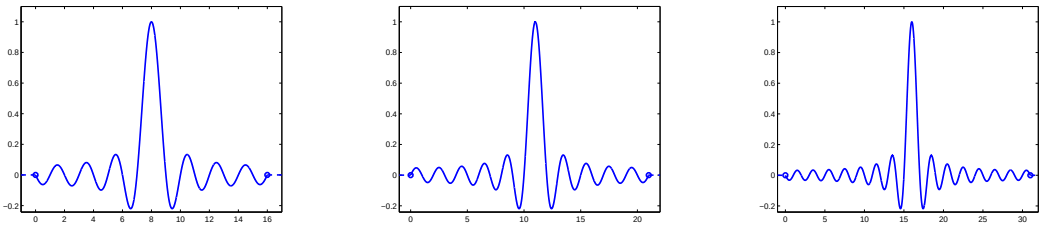
we can satisfy (4.5) by choosing d_ℓ so that the resulting polynomial $\hat{\gamma}(\omega) = \sum_\ell d_\ell (j\omega)^\ell$ interpolates the set of points $(\omega_m, |\hat{\beta}_{\vec{\alpha}}(\omega_m)|^{-1})$ for $m = 0, 1, \dots, P$.

Once we have designed the kernels such that $c_{m,0}$ has modulus one for all m, we are left with a phase ambiguity, since we may write $c_{m,0} = |c_{m,0}|\mathrm{e}^{j\omega_m \Delta}$ for $m = 0, \dots, P$. The form of the phase comes from the fact that it is equivalent to introducing a time shift $\Delta \in \mathbb{R}$ for the E-Spline in (4.6), as we show in the next subsection. The phase gives an additional degree of freedom and we may obtain its value by imposing, for instance, that the function be continuous. In order for the exponential MOMS with $|c_{m,0}| = 1$ and parameters (4.4) to be continuous-time functions we need that Δ be an integer greater than or equal to 1 and smaller than or equal to P. In Figure 4.2 we present some of the kernels obtained by implementing the procedure explained above. Interestingly, as shown in Appendix C.1, these specific functions always equal one period of the Dirichlet kernel. We also point out

that when $P + 1 = N$ the scenario derived using this family of exponential reproducing kernels converges to the original FRI formulation of [2] when we periodise the input or, equivalently, the sampling kernel.



(a) $P + 1 = 2$ (b) $P + 1 = 6$ (c) $P + 1 = 11$



(d) $P + 1 = 16$ (e) $P + 1 = 21$ (f) $P + 1 = 31$

Figure 4.2: ***Examples of exponential MOMS.*** **These are 6 of the 30 possible kernels with support $P + 1 \leqslant N = 31$ samples. They coincide with one period of the Dirichlet kernel of period $P + 1$ for P even or $2(P + 1)$ for P odd (see Appendix C.1). All of them are built selecting the phase of $c_{m,0}$ such that they are continuous-time functions centred around $\Delta = \lceil \frac{P+1}{2} \rceil$.**

To summarise, we have explained how to build exponential reproducing kernels of maximum order and minimum support characterised by $|c_{m,0}| = 1$ for $m = 0, \dots, P$. The kernels consist of the linear combination (4.6) of the E-Spline $\beta_{\vec{\alpha}}^{(\ell)}(t)$ of parameters (4.4) and its P derivatives. The coefficients d_ℓ are such that the polynomial $\hat{\gamma}(\omega) = \sum_\ell d_\ell (j\omega)^\ell$ interpolates the set of points $(\omega_m, |\hat{\beta}_{\vec{\alpha}}(\omega_m)|^{-1})$ for $m = 0, 1, \dots, P$. Note that when $P+1 = N$, then $\mathbf{C}$ is the inverse discrete Fourier transform (IDFT) matrix of size $N \times N$. In order to determine the phase of the coefficients, we introduce an additional degree of freedom Δ and impose continuity of the kernel. We achieve this by using $c_{m,0} = |c_{m,0}| \mathrm{e}^{j\omega_m \Delta}$ for $m = 0, \dots, P$ and an integer shift $\Delta \in [1, P]$.

4.2.1 More general exponential MOMS

The family of eMOMS is, however, not limited to the specific kernels derived so far. For example, in Appendix C.1 we prove that the SoS family of kernels [15] is a specific instance of eMOMS, obtained by relaxing condition (4.5) but still using parameters (4.4). Another example are the E-Spline kernels of Section 2.3.2, which are simply eMOMS generated by imposing $\hat{\beta}_{\vec{\alpha}}(\alpha_m) = c_{m,0}^{-1}$ (or equivalently $\hat{\gamma}(\alpha_m) = 1$) for all m. This means that $d_0 = 1$ and $d_\ell = 0$ for any other ℓ in (4.6).

In order to design more general eMOMS, we begin by introducing an additional degree

of freedom that removes the phase ambiguity: a time shift Δ such that (4.6) becomes

$$\phi(t) = \varphi(t+\Delta) = \sum_{\ell=0}^{P} d_\ell \frac{\mathrm{d}^\ell \beta_{\vec{\alpha}}(t+\Delta)}{\mathrm{d}t^\ell}. \tag{4.7}$$

This time domain expression is characterised by the double-sided Laplace transform

$$\hat{\phi}(s) = \sum_{\ell=0}^{P} d_\ell s^\ell \hat{\beta}_{\vec{\alpha}}(s)\mathrm{e}^{s\Delta} = D(s)\hat{\beta}_{\vec{\alpha}}(s)\mathrm{e}^{s\Delta},$$

where the double-sided Laplace transform of an E-Spline is:

$$\hat{\beta}_{\vec{\alpha}}(s) = \prod_{m=0}^{P} \frac{1-\mathrm{e}^{\alpha_m - s}}{s-\alpha_m}. \tag{4.8}$$

Assume that we want the exponential MOMS to satisfy the $P+1$ conditions (more general than (4.5)) given by the following system of equations:

$$\hat{\phi}(\alpha_m) = D(\alpha_m)\hat{\beta}_{\vec{\alpha}}(\alpha_m)\mathrm{e}^{\alpha_m\Delta} = |c_{m,0}^{-1}| = \eta_m, \qquad m = 0,\dots,P, \tag{4.9}$$

where $\eta_m \in \mathbb{R}$ and where now α_m may have a real part and $L \neq P+1$ as in (4.2). Note that defining the eMOMS as (4.7) and imposing the conditions (4.9) is equivalent to using the original eMOMS equation (4.6) and imposing $\hat{\varphi}(\alpha_m) = c_{m,0}^{-1} = |c_{m,0}|^{-1}\mathrm{e}^{-\alpha_m\Delta}$.

In any case, this is to say that the polynomial $D(s)$ interpolates the set of points $(\alpha_m, \eta_m\hat{\beta}_{\vec{\alpha}}(\alpha_m)^{-1}\mathrm{e}^{-\alpha_m\Delta})$ for $m = 0,\dots,P$. We can find the polynomial via Lagrange interpolation, by first defining:

$$\ell_i(s) = \prod_{\substack{m=0 \\ m\neq i}}^{P} \frac{s-\alpha_m}{\alpha_i - \alpha_m}$$

with which we can define:

$$D(s) = \sum_{i=0}^{P} (\eta_i \hat{\beta}_{\vec{\alpha}}(\alpha_i)^{-1}\mathrm{e}^{-\alpha_i\Delta})\ell_i(s),$$

since $\ell_i(\alpha_m) = \delta_{m,n}$. Then, using (4.8) for $\hat{\beta}_{\vec{\alpha}}(s)$ we conclude that

$$D(s) = \sum_{i=0}^{P} \eta_i \mathrm{e}^{-\alpha_i\Delta} \prod_{\substack{m=0 \\ m\neq i}}^{P} \frac{s-\alpha_m}{1-\mathrm{e}^{\alpha_m-\alpha_i}}. \tag{4.10}$$

Finally, we may find Δ by imposing continuity of the kernel (4.7). The kernel is continuous when the polynomial (4.10) is of one degree less than the maximum, hence we may obtain Δ by making the coefficient for s^P be equal to zero. For instance, in the special case we use exponential parameters $\alpha_m = j\omega_m = j\frac{\pi}{P+1}(2m-P)$, for $m = 0,\dots,P$,

as in (4.4) and $\eta_i = 1$ for all i, then the interpolating polynomial (4.10) becomes:

$$D(s) = \sum_{i=0}^{P} \frac{\mathrm{e}^{-\alpha_i \Delta}}{P+1} \prod_{\substack{m=0 \\ m \neq i}}^{P} (s - \alpha_m).$$

Here we have used the fact that $\prod_{m=0, m\neq i}^{P} (1 - \mathrm{e}^{\alpha_m - \alpha_i}) = P + 1$ when the exponential parameters α_m are as in (4.4). We prove this equality in Appendix C.4. Therefore, for this particular choice of exponential parameters continuity of the kernel implies that

$$\begin{aligned} 0 = \sum_{m=0}^{P} \mathrm{e}^{-j\omega_m \Delta} &= \mathrm{e}^{j\frac{\pi P \Delta}{P+1}} \sum_{m=0}^{P} \mathrm{e}^{-j\frac{2\pi m \Delta}{P+1}} \\ &= \mathrm{e}^{j\frac{\pi P \Delta}{P+1}} \frac{1 - \mathrm{e}^{-j2\pi\Delta}}{1 - \mathrm{e}^{-j\frac{2\pi\Delta}{P+1}}} \quad \leftrightarrow \quad \mathrm{e}^{-j2\pi\Delta} = 1 \end{aligned}$$

which is equivalent to saying that Δ is an integer. We thus conclude that for this specific case, which corresponds to the eMOMS kernels of the previous subsection, $\Delta = 1, 2, \ldots, P$.

4.2.2 Cramér–Rao bound for exponential MOMS

To conclude the section on eMOMS we provide the CRB associated to the problem of estimating the parameters (t_0, a_0) of a single Dirac that has been sampled with the kernel in (4.6). The kernel reproduces exponentials of parameters (4.4), satisfies that $c_{m,0}^{-1} = \hat{\varphi}(\omega_m) = \hat{\varphi}_m$ where $\alpha_m = j\omega_m$, and is such that $N = P + 1$[1] and P even. This is a simple case for which it is possible to derive closed form expressions for the deviation of the location and the amplitude. The proof can be found in Appendix C.2.

The uncertainty in the location satisfies:

$$\frac{\Delta t_0}{\tau} \geqslant \frac{1}{2\pi} \sqrt{\frac{N}{\sum_{k\in\mathcal{K}} k^2 |\hat{\varphi}_k|^2}} \mathrm{PSNR}^{-\frac{1}{2}},$$

where we have defined the peak signal-to-noise ratio as $\mathrm{PSNR} = \left(\frac{a_0}{\sigma}\right)^2$, and the uncertainty in the amplitude satisfies:

$$\frac{\Delta a_0}{|a_0|} \geqslant \sqrt{\frac{N}{\sum_{k\in\mathcal{K}} |\hat{\varphi}_k|^2}} \mathrm{PSNR}^{-\frac{1}{2}}.$$

Note that when $|\hat{\varphi}_k| = 1$ for all k, and if we denote $\mathcal{K} = \{k : k = -M, \ldots, M\}$, the

[1] We observe that the condition $P + 1 = N$ can be imposed only for blockwise sampling, e.g. when sampling periodic signals using N samples. This condition cannot be imposed on infinite length signals since the number of samples is in this case infinite, and sequential reconstruction algorithms should operate on blocks with possibly varying number of samples.

above expressions simplify to:

$$\frac{\Delta t_0}{\tau} \geqslant \frac{1}{2\pi}\sqrt{\frac{3N}{M(M+1)(2M+1)}}\mathrm{PSNR}^{-\frac{1}{2}}, \tag{4.11}$$

$$\frac{\Delta a_0}{|a_0|} \geqslant \sqrt{\frac{N}{2M+1}}\mathrm{PSNR}^{-\frac{1}{2}}.$$

We now compare (C.11) with the uncertainty obtained for the CRB associated to a sum of exponentials in AWGN (B.2), which can be found in Appendix B.1. The two expressions can be compared by simply multiplying (B.2) by $\sqrt{N}$, which is like saying that the noise covariance matrix of the noise added to the exponentials is $\mathbf{R_b} = \sigma^2 N\mathbf{I}$ (this is needed for the scenarios to be equivalent). The equations should match since the FRI problem described in this section is the same as the problem of estimating the parameters of the linear combination of the $P+1$ exponentials $a_0 \mathrm{e}^{\alpha_m t_0}$ for $m = 0, \ldots, P$. By simple manipulations of both expressions, we conclude that in either case:

$$\frac{\Delta t_0}{\tau} \geqslant \frac{1}{\pi}\sqrt{\frac{3N}{P(P+1)(P+2)}}\mathrm{PSNR}^{-\frac{1}{2}}.$$

We end by noting that for more general cases, i.e. any (generalised) exponential reproducing kernel (4.7) of support $P+1 \leqslant N$ and exponential parameters (4.2) the above expressions can be adapted to provide the uncertainties in the location and amplitude of a single Dirac. We provide a formal derivation of this in Appendix C.3.

4.3 Simulations

We now present the performance of the exponential MOMS kernels compared to the performance of the E-Splines of Chapter 3. We further show the stability of eMOMS kernels when retrieving a high number of Diracs.

4.3.1 The experimental setup

The setup is the same as that of Chapter 3: We take N samples by computing $y_n = \sum_{k=0}^{K-1} a_k \varphi\left(\frac{t_k}{T} - n\right)$ for $n = 0, \ldots, N-1$. The sampling period is $T = \frac{1}{N}$ unless specified otherwise. We then either use the noiseless samples or corrupt them with additive white Gaussian noise of variance σ^2, according to the target signal-to-noise ratio $\mathrm{SNR(dB)} = 10\log\frac{\|\mathbf{y}\|^2}{N\sigma^2}$. We finally calculate the noisy $P+1$ moments and then retrieve the innovation parameters $\{a_k, t_k\}_{k=0}^{K-1}$ of the input. We always use the subspace estimator method as our recovery algorithm, as described in Sections 3.2 and 3.3. eMOMS usually require no whitening transform, whereas for E-Splines we use $\mathbf{W} = \mathbf{R}_{\mathbf{B}}^{\dagger/2}$.

We present results for single realisations of the sampling and reconstruction process or for average performance over multiple trials. For the latter, we show the root mean

square error (RMSE) of the locations:

$$e_{t_k} = \sqrt{\frac{\sum_{i=0}^{I-1}(\hat{t}_k^{(i)} - t_k)^2}{I}},$$

where $\hat{t}_k^{(i)}$ is the i-th estimated time location, I is the total number of realisations and e_{t_k} the error for each of the K Diracs. We do this for a range of fixed signal-to-noise ratio (SNR) values and average the effects using $I = 1000$ noise realisations at each SNR. We compare the RMSE with the square root of the variance predicted by the sample-based CRB (3.11) and the moment-based CRB (3.13).

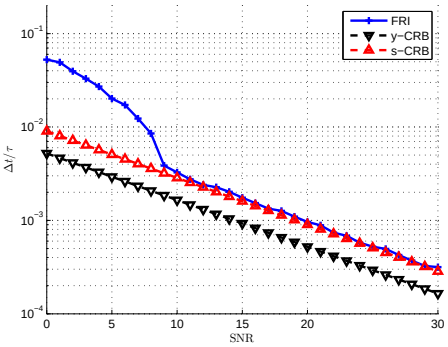
4.3.2 Results

In Figure 4.3 (a-d) we present simulation results when we retrieve $K = 2$ Diracs from $N = 31$ samples using the exponential MOMS kernels of Section 4.2. We specifically show the deviation in the location of the first Dirac, the deviation of the second Dirac being very similar. We see that for any order $P + 1$ eMOMS achieve the moment-based CRB (in red and denoted s-CRB in the legend), even without the use of a whitening transform. This bound gets closer to the sample-based CRB (in black and denoted y-CRB in the legend) as the value of $P + 1$ increases and as expected matches it when $P + 1 = N$. To further illustrate the stability of eMOMS, in Figure 4.3 (e) we show the retrieval of $K = 20$ Diracs randomly spaced over $\tau = NT = 1$ and with arbitrary amplitudes. The signal-to-noise ratio is 15dB, and we use $N = 61$ samples and $P + 1 = N = 61$ moments.

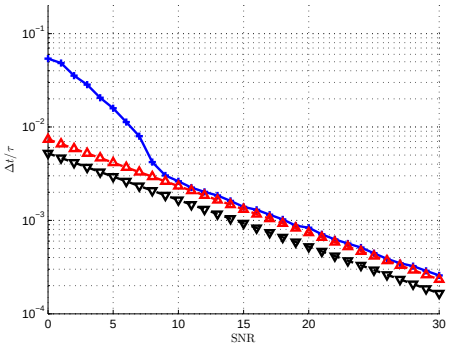
In Figure 4.4 (a-b) we present simulation results when we retrieve $K = 2$ Diracs from $N = 31$ samples using a standard E-Spline and the exponential MOMS kernels of Section 4.2. The former is characterised by purely imaginary exponents $\alpha_m = j\frac{\pi}{2(P+1)}(2m-P)$ for $m = 0, \ldots, P$, since this set guarantees a stable and accurate recovery for all orders $P + 1$ we use. We can see that eMOMS outperform E-splines for any order we consider, which is as expected since they are optimal for the proposed setup used for the simulations. We end by highlighting that, contrary to E-Splines, the sample-based CRB for eMOMS remains unchanged when the order $P + 1$ increases. The effect we see is that the moment-based CRB and the performance improve constantly until $P + 1$ equals N, when they both match the sample-based CRB. This is in contrast with E-Splines for which the moment-based CRB and the performance improve but at the same time the sample-based CRB worsens. Therefore, when $P + 1 = N$ they all coincide, but the bound may not be as low as the one for lower E-Spline orders.

4.4 Summary

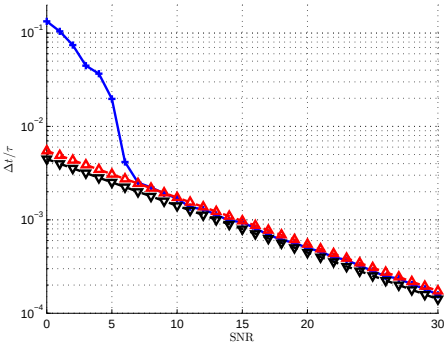
In this chapter we have further analysed the noisy scenario introduced in Chapter 3 by studying the main sources of instability for FRI recovery. Specifically, we have shown that the retrieval from the power sum (4.1) can be quite unstable, in particular when



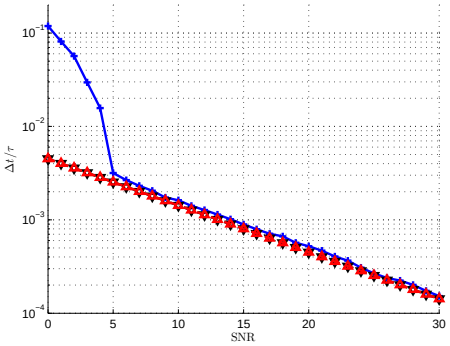
(a) eMOMS $P + 1 = 11$



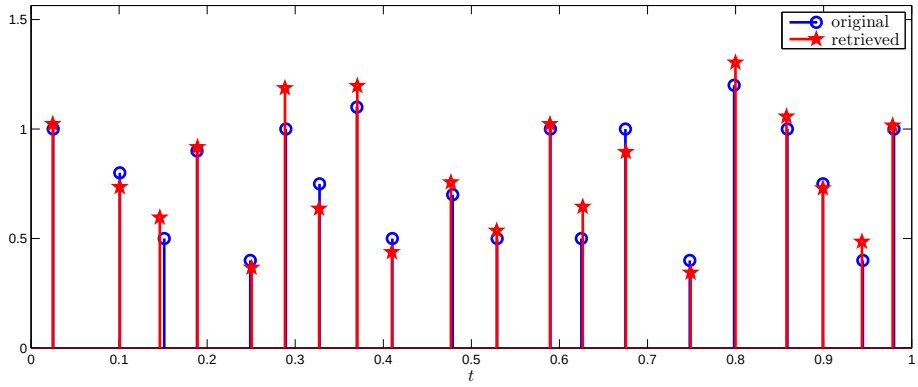
(b) eMOMS $P + 1 = 16$



(c) eMOMS $P + 1 = 21$

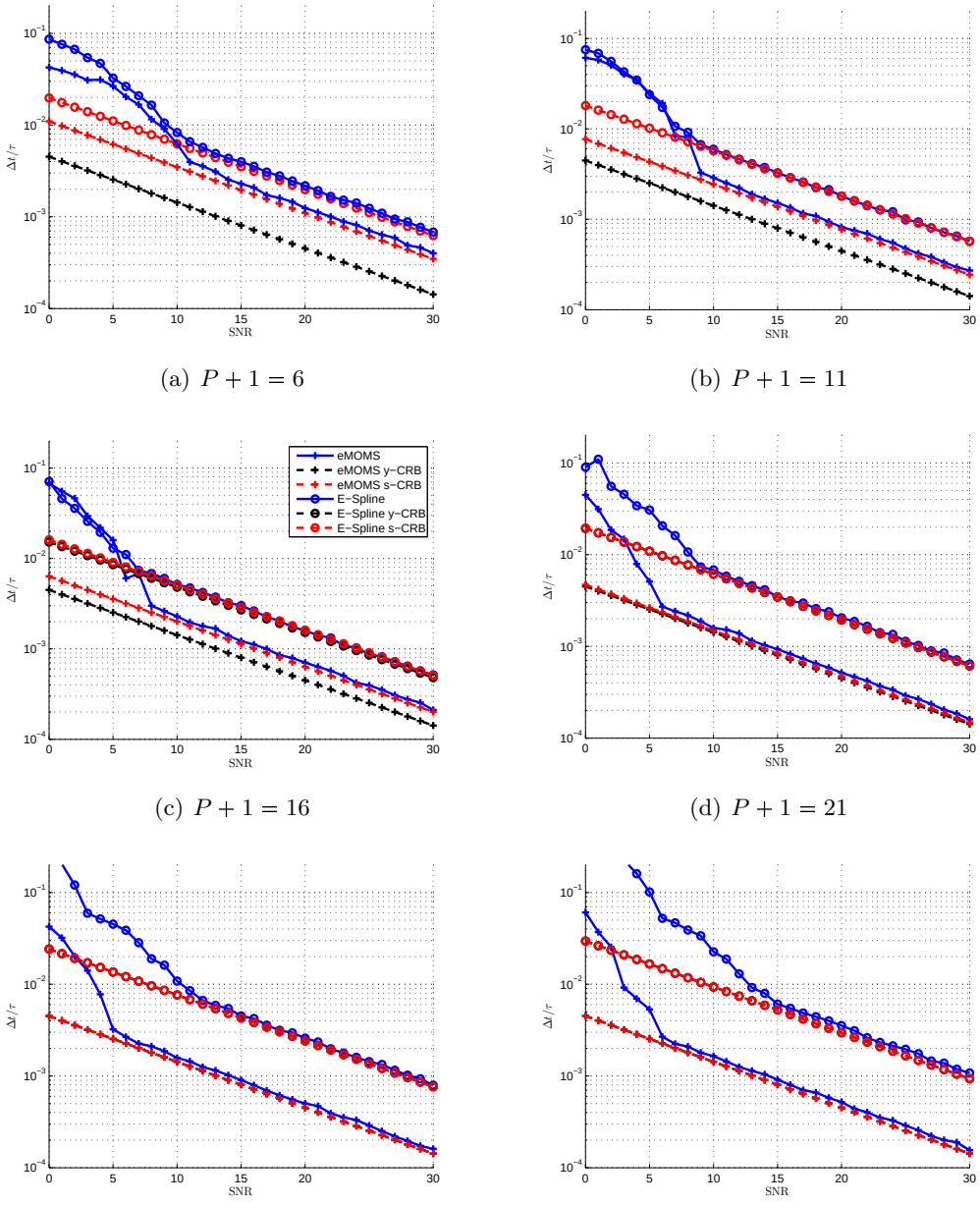


(d) eMOMS $P + 1 = 31$



(e) Retrieval of $K = 20$ Diracs

Figure 4.3: ***Performance of exponential MOMS kernels.*** **(a-d) show the performance of exponential MOMS kernels of different orders $P + 1$ when white Gaussian noise is added to the $N = 31$ samples. We show the recovery of the first of $K = 2$ Diracs. eMOMS always reach the moment-based CRB (s-CRB), even though pre-whitening is not utilised. This bound gets closer to the sample-based CRB (y-CRB) as the value of $P + 1$ increases and as expected matches it when $P + 1 = N$. Finally, (e) shows the retrieval of $K = 20$ Diracs randomly spaced over $\tau = NT = 1$ when doing τ-periodic sampling. The signal-to-noise ratio is 15dB, and we use $N = 61$ samples and $P + 1 = N = 61$ moments.**



(a) $P+1=6$ (b) $P+1=11$

(c) $P+1=16$ (d) $P+1=21$

(e) $P+1=26$ (f) $P+1=31$

Figure 4.4: ***Performance of exponential MOMS vs. E-Spline kernels.*** **We compare the performance of E-Splines vs. exponential MOMS kernels of different orders $P+1$ when noise is added to $N=31$ samples. We show the recovery of the first of $K=2$ Diracs. We note that eMOMS always outperform E-splines even though both achieve the moment-based CRB (s-CRB). Prewhitening is only needed for E-Splines.**

the nodes u_k are not on the unit circle or when the amplitudes x_k span several orders of magnitude. We have also established that through a carefully chosen matrix $\mathbf{C}$ we can make the retrieval problem tractable.

In addition, we have determined the optimal parameters that characterise exponential reproducing kernels so they are most stable and have the best possible performance. Equipped with such analysis, we have designed optimal exponential reproducing kernels of maximum order and minimum support (eMOMS). We have shown that these kernels are superior to carefully chosen E-Splines throughout the simulations.

Moreover, we have extended optimal eMOMS to an even broader family that is also a particularisation of the general E-Splines of [41]. Interestingly, we have derived a closed form expression for the CRB associated to the retrieval of the innovation parameters of $K = 1$ Dirac from the samples taken by the optimal eMOMS. We have seen that, under certain circumstances, it equals the CRB associated to the retrieval of the parameters from the noisy power sum series as developed in Appendix B.1.

Chapter 5

Universal sampling of signals with finite rate of innovation

In the previous chapter we have shown how to design optimal exponential reproducing kernels for FRI sampling, when the input is a train of Diracs and we contaminate the samples with AWGN. In many practical circumstances, however, the freedom to choose the sampling kernel $\varphi(t)$ is a luxury that we may not have.

Essential in the FRI setting is the ability of $\varphi(t)$ to reproduce exponential functions, because this allows us to map the signal reconstruction problem to Prony's method in line-spectral estimation theory. In this chapter we relax this condition and consider any function $\varphi(t)$ for which the exponential reproduction property (2.7) does not necessarily hold. For these functions it is still possible to find coefficients $c_{m,n}$ such that the reproduction of exponentials is approximate rather than exact. We propose to use this approximate reproduction and the corresponding coefficients $c_{m,n}$ to retrieve FRI signals from the samples obtained using these kernels. This new approach has several advantages: First, it is universal in that it can be used with any kernel $\varphi(t)$. In fact, as we shall show in the following sections, this new formulation does not even require an exact knowledge of the kernel. Second, while reconstruction of FRI signals with this new method is not going to be exact, we show that in many cases a proper iterative algorithm can make the reconstruction error arbitrary small. Finally, this new approach can be used to increase the resiliency to noise of some unstable kernels proposed in the FRI literature. For example, kernels like polynomial splines or the Gaussian function lead to very ill-conditioned reconstruction procedures. We show that by replacing the original $\mathbf{C}$ with the one formed from properly chosen coefficients $c_{m,n}$, based on approximate reproduction of exponentials, we achieve a much more stable reconstruction with the same kernels.

The chapter is organised as follows: In Section (5.1) we formalise the notion of approximation of exponentials. Specifically, we explain how to choose the coefficients $c_{m,n}$ according to the type of approximation we want to achieve. Then, in Section 5.2 we use the property of approximation of exponentials to recover a train of K Diracs from the samples taken by the kernel $\varphi(t)$ in the absence of noise. We also propose an iterative

method to refine the reconstruction and make its error arbitrarily small. Finally, we simulate the approximate FRI technique in Section 5.3 and show the results obtained when retrieving trains of Diracs in the presence of white Gaussian noise added to the samples. We then conclude the chapter in Section 5.4.

5.1 Approximate reproduction of exponentials

Assume we want to use a function $\varphi(t)$ and its integer shifts to approximate the exponential $\mathrm{e}^{\alpha t}$. Specifically, we want to find the coefficients c_n such that:

$$\sum_{n\in\mathbb{Z}} c_n \varphi(t-n) \approxeq \mathrm{e}^{\alpha t}. \tag{5.1}$$

This approximation becomes exact only when $\varphi(t)$ satisfies the generalized Strang-Fix conditions (2.8). For any other function it is of particular interest to find the coefficients c_n that best fit (5.1). In order to do so, we directly use[1] $c_n = c_0 \mathrm{e}^{\alpha n}$ and introduce the 1-periodic function

$$g_\alpha(t) = c_0 \sum_{n\in\mathbb{Z}} \mathrm{e}^{-\alpha(t-n)} \varphi(t-n). \tag{5.2}$$

We then find that approximating the exponential $\mathrm{e}^{\alpha t}$ with integer shifts of $\varphi(t)$ can be transformed into approximating $g_\alpha(t)$ by the constant value 1. The reason is that we can rewrite (5.1) in the form of the right-hand side of (5.2) by substituting $c_n = c_0 \mathrm{e}^{\alpha n}$ and moving $\mathrm{e}^{\alpha t}$ to the left-hand side.

As a consequence of Poisson summation formula (1), we have that the Fourier series expansion of $g_\alpha(t)$ is given by

$$g_\alpha(t) = \sum_{l\in\mathbb{Z}} g_l \mathrm{e}^{j2\pi l t} = \sum_{l\in\mathbb{Z}} c_0 \hat{\varphi}(\alpha + j2\pi l) \mathrm{e}^{j2\pi l t}.$$

More specifically we have used:

$$\sum_{n\in\mathbb{Z}} \mathrm{e}^{-\alpha(t-n)} \varphi(t-n) = \sum_{l\in\mathbb{Z}} \hat{\varphi}\left(\alpha + j2\pi l\right) \mathrm{e}^{j2\pi l t},$$

where in order to calculate the Fourier transform of $\phi(t) = \mathrm{e}^{-\alpha x}\varphi(x)$ we can evaluate its Laplace transform $\hat{\phi}(s) = \hat{\varphi}(s+\alpha)$ at $s = j\omega$. As a consequence, our approximation problem reduces to:

$$g_\alpha(t) = \sum_{l\in\mathbb{Z}} c_0 \hat{\varphi}(\alpha + j2\pi l) \mathrm{e}^{j2\pi l t} \approxeq 1. \tag{5.3}$$

This shows more deeply the relation between the generalised Strang-Fix conditions (2.8) and the approximation of exponentials. If $\varphi(t)$ satisfies the generalised Strang-Fix conditions (2.8) then $\hat{\varphi}(\alpha + j2\pi l) = 0$ for $l \in \mathbb{Z}\backslash\{0\}$ and (5.3) holds exactly for $c_0\hat{\varphi}(\alpha) = 1$.

[1] The exact exponential reproducing coefficients always satisfy $c_n = c_0 \mathrm{e}^{\alpha n}$. We now anticipate that different sets of approximation coefficients we derive throughout the section also have the same form.

Otherwise, the terms $\hat{\varphi}(\alpha + j2\pi l)$ for $l \in \mathbb{Z}\backslash\{0\}$ do not vanish, and we can only find the coefficient c_0 so that $g_\alpha(t) \approxeq 1$. However, the closer the values $\hat{\varphi}(\alpha + j2\pi l)$ are to zero, the better the approximation in (5.1) is.

In general $\varphi(t)$ can be any function and we can find different sets of coefficients c_n in order for the approximation (5.1) to hold. Regardless of the coefficients we use, we can determine the accuracy of our approximation by using the Fourier series expansion of $g_\alpha(t)$. In fact, the error of approximating $f(t) = \mathrm{e}^{\alpha t}$ by the function $s(t) = \sum_{n\in\mathbb{Z}} c_n \varphi(t-n)$ with coefficients $c_n = c_0 \mathrm{e}^{\alpha n}$ is equal to:

$$\begin{aligned} \varepsilon(t) = f(t) - s(t) &= \mathrm{e}^{\alpha t}\left[1 - g(t)\right] \\ &= \mathrm{e}^{\alpha t}\left[1 - c_0 \sum_{l\in\mathbb{Z}} \hat{\varphi}(\alpha + j2\pi l)\mathrm{e}^{j2\pi l t}\right]. \end{aligned} \tag{5.4}$$

Note that, if the Laplace transform of $\varphi(t)$ decays sufficiently quickly, very few terms of the Fourier series expansion are needed to have an accurate bound for the error.

A natural choice of the coefficients $c_n = c_0 \mathrm{e}^{\alpha n}$ is the one given by the *least-squares approximation*. Despite the fact that $f(t)$ is not square-integrable, we can still obtain the coefficients by computing the orthogonal projection of $f(t)$ onto the subspace spanned by $\varphi(t-n)$ [81]. In Appendix D.1 we show that these coefficients take the form

$$c_n = \frac{\hat{\varphi}(-\alpha)}{\hat{a}_\varphi(\mathrm{e}^\alpha)}\mathrm{e}^{\alpha n},$$

where $\hat{a}_\varphi(\mathrm{e}^\alpha) = \sum_{l\in\mathbb{Z}} a_\varphi[l]\mathrm{e}^{-\alpha l}$ is the z-transform of $a_\varphi[l] = \langle \varphi(t-l), \varphi(t)\rangle$, evaluated at $z = \mathrm{e}^\alpha$.

The least-squares approximation has the disadvantage that it requires the exact knowledge of $\varphi(t)$. However, as we stated before, if the Laplace transform of $\varphi(t)$ decays sufficiently quickly, we can assume the terms $\hat{\varphi}(\alpha + j2\pi l)$ are close to zero for $l \in \mathbb{Z}\backslash\{0\}$. In this case we have that the error in (5.4) is easily minimised by choosing $c_0 = \hat{\varphi}(\alpha)^{-1}$. We denote this second type of approximation *constant least-squares*. Besides its simplicity, a second advantage of choosing $c_n = \hat{\varphi}(\alpha)^{-1}\mathrm{e}^{\alpha n}$ is that it requires only the knowledge of the Laplace transform of $\varphi(t)$ at α. If we put ourselves in the FRI setting where we require the approximate reproduction of the exponentials $\mathrm{e}^{\alpha_m t}$ with $m = 0, \ldots, P$, then this simplified formulation needs only the knowledge of the Laplace transform of $\varphi(t)$ at α_m, $m = 0, \ldots, P$.

Finally, a third interesting choice of coefficients is the one that ensures that $s(t)$ *interpolates* $f(t)$ exactly at $t = \ell \in \mathbb{Z}$ [1, 8]. These coefficients, as we prove in Appendix D.1, are as follows:

$$c_n = \frac{1}{\sum_{l\in\mathbb{Z}} \mathrm{e}^{-\alpha l}\varphi(l)}\mathrm{e}^{\alpha n}.$$

Note that in order to use the interpolation coefficients we only need information on $\varphi(t)$ at integer instants of time. We summarise the previous results in Table 5.1.

Table 5.1: Coefficients for the approximate reproduction (5.1)

Type	Coefficients
Least-squares approximation	$c_n = \dfrac{\hat{\varphi}(-\alpha)}{\hat{a}_\varphi(\mathrm{e}^{\alpha})}\mathrm{e}^{\alpha n}$
Constant least-squares	$c_n = \hat{\varphi}(\alpha)^{-1}\mathrm{e}^{\alpha n}$
Interpolation	$c_n = \dfrac{1}{\sum_{l\in\mathbb{Z}} \mathrm{e}^{-\alpha l}\varphi(l)}\mathrm{e}^{\alpha n}$

According to our experience, the least-square coefficients provide a smaller error (5.4) when the exponential $\mathrm{e}^{\alpha t}$ to approximate has an exponent α that is not purely imaginary. Otherwise, the constant least-squares coefficients are just as good. Interpolation coefficients are very easy to compute given the values of the kernel at integer points in time. However, they always provide a worse approximation quality.

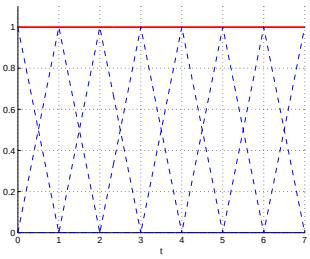
We show an example of the above analysis when the sampling kernel is a linear spline and we want to use the linear combinations of its shifted versions to reproduce exponentials. The linear spline reproduces polynomials of orders 0 and 1 exactly, as illustrated in Figure 5.1 (a-b). Now, with the same function, we address the problem of approximately reproducing the 4 complex exponentials $\mathrm{e}^{\alpha_m} = \mathrm{e}^{j\frac{\pi}{16}(2m-7)t}$ for $m = 0, \ldots, 3$. The interval of approximation depends on the support of the spline $M+1$ and the number of samples. If, for instance, we define an approximation interval $[0, \tau)$, with $\tau \in \mathbb{Z}$, then we should employ indices $-M \leqslant n \leqslant \tau - 1$. We show the approximation of the real part of the exponentials obtained by using the constant least-squares coefficients

$$c_{m,n} = \hat{\varphi}\left(\alpha_m\right)^{-1} \mathrm{e}^{\alpha_m n}, \qquad m = 0, \ldots, 3$$

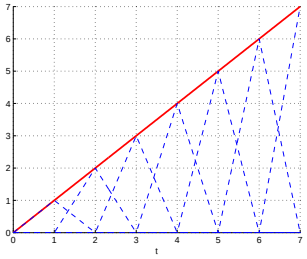
where $\alpha_m = j\frac{\pi}{16}(2m-7)$, in Figure 5.1 (c, d, e, i). In addition, we show the interpolation of the real part of the exponentials obtained by using the coefficients

$$c_{m,n} = \frac{1}{\sum_{\ell=0}^{M+1} \mathrm{e}^{\alpha_m \ell}\varphi(\ell)}\mathrm{e}^{\alpha_m n}, \qquad m = 0, \ldots, 3$$

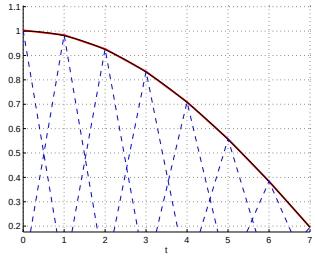
in Figure 5.1 (f, g, h, j). Some exponentials are better reproduced than others, in this example the ones with lower frequency. We have seen in practice that higher order splines tend to improve the approximation quality of the reproduction, however, we have chosen a linear spline for illustration purposes because it makes clear that the constant least-squares approximations are superior to those obtained using the interpolation coefficients. Also note that the number of exponentials that can be approximated is arbitrary, and is independent of the order of the spline.



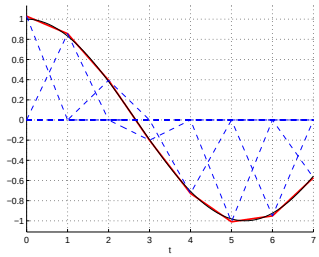
(a) Reproduction of 1



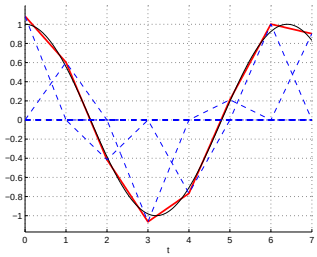
(b) Reproduction of t



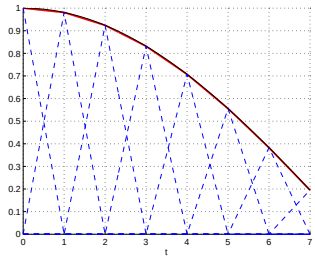
(c) Approximation of $\mathrm{Re}\{\mathrm{e}^{-j\frac{\pi}{16}t}\}$



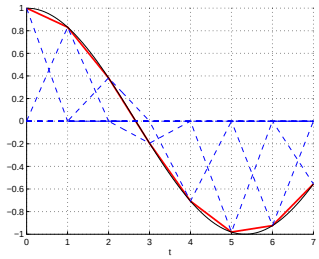
(d) Approximation of $\mathrm{Re}\{\mathrm{e}^{-j\frac{3\pi}{16}t}\}$



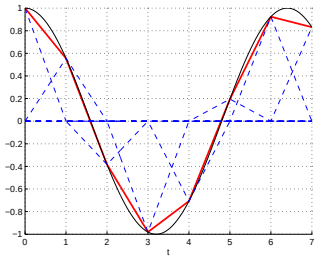
(e) Approximation of $\mathrm{Re}\{\mathrm{e}^{-j\frac{5\pi}{16}t}\}$



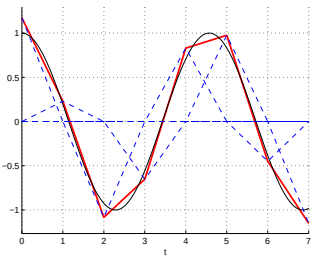
(f) Interpolation of $\mathrm{Re}\{\mathrm{e}^{-j\frac{\pi}{16}t}\}$



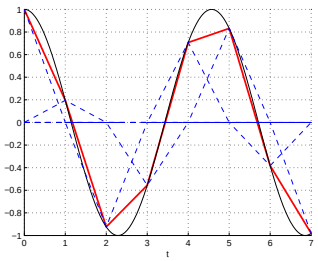
(g) Interpolation of $\mathrm{Re}\{\mathrm{e}^{-j\frac{3\pi}{16}t}\}$



(h) Interpolation of $\mathrm{Re}\{\mathrm{e}^{-j\frac{5\pi}{16}t}\}$



(i) Approximation of $\mathrm{Re}\{\mathrm{e}^{-j\frac{7\pi}{16}t}\}$



(j) Interpolation of $\mathrm{Re}\{\mathrm{e}^{-j\frac{7\pi}{16}t}\}$

Figure 5.1: ***B-Spline kernel reproduction and approximation capabilities.*** **Figures (a-b) show the exact reconstruction of polynomials of orders 0 and 1. Figures (c-j) show the constant least-squares approximation and the interpolation of the real parts of 4 complex exponentials: $\mathrm{e}^{j\frac{\pi}{16}(2m-7)t}$ for $m = 0, \ldots, 3$. We plot the weighted and shifted versions of the splines with dashed blue lines, the reconstructed polynomials and exponentials with red solid lines, and the exact functions to be reproduced with solid black lines.**

5.1.1 Further remarks

We end the above analysis by noting that there are two main ways of introducing the sampling period T in FRI sampling schemes that translate into different types of shifts for the kernel $\varphi(t)$ to reproduce exponentials. In the first case, compact support kernels usually obtain samples $y_n = \left\langle x(t), \varphi\left(\frac{t}{T} - n\right)\right\rangle$, which means that the linear combination of integer shifts of the time-scaled version of the kernel $\varphi(\frac{t}{T})$ needs to satisfy:

$$\mathrm{e}^{\alpha \frac{t}{T}} \approx \sum_{n\in\mathbb{Z}} c_n \varphi\left(\frac{t}{T} - n\right).$$

The coefficients c_n for the integer shifts of $\varphi(t)$ and $\varphi(\frac{t}{T})$ to reproduce $\mathrm{e}^{\alpha t}$ and $\mathrm{e}^{\alpha \frac{t}{T}}$ respectively are identical. Therefore Table 5.1 can be used straight away.

In the second case, other types of kernels $h(t)$ such as the sinc and the Gaussian functions take samples $y_n = \langle x(t), h(nT - t)\rangle$, so that the linear combination of integer shifts of the kernel $h(t)$ satisfy:

$$\mathrm{e}^{\alpha t} \approx \sum_{n\in\mathbb{Z}} c_n h(t - nT). \tag{5.5}$$

We may derive the equivalent coefficients for this scenario by just rewriting (5.5) as follows:

$$\mathrm{e}^{\alpha' \frac{t}{T}} \approx \sum_{n\in\mathbb{Z}} c_n \varphi\left(\frac{t}{T} - n\right),$$

where $\alpha' = \alpha T$ and $\varphi(t) = h(Tt)$. As a consequence, we can directly use the coefficients of Table 5.1 with α' instead of α. Then, after some simple manipulations and since $\hat{\varphi}(s) = T\hat{h}(\frac{s}{T})$ we get Table 5.2.

Table 5.2: Coefficients for the approximate reproduction (5.5)

Type	Coefficients
Least-squares approximation	$c_n = \dfrac{T\hat{h}(-\alpha)}{\hat{a}_h(\mathrm{e}^{\alpha T})}\mathrm{e}^{\alpha n T}$
Constant least-squares	$c_n = T\hat{h}(\alpha)^{-1}\mathrm{e}^{\alpha n T}$
Interpolation	$c_n = \dfrac{1}{\sum_{l\in\mathbb{Z}} \mathrm{e}^{-\alpha l T} h(-lT)}\mathrm{e}^{\alpha n T}$

In Appendix D.2 we show how to obtain the different types of coefficients for B-Splines and for Gaussian kernels. We also illustrate the approximation capabilities of Gaussian kernels with an example similar to that of Figure 5.1.

5.2 Approximate FRI recovery

Consider again the stream of Diracs $x(t) = \sum_{k=0}^{K-1} a_k \delta(t - t_k)$ and the samples

$$y_n = \left\langle x(t), \varphi\left(\frac{t}{T} - n\right)\right\rangle = \sum_{k=0}^{K-1} a_k \varphi\left(\frac{t_k}{T} - n\right). \tag{5.6}$$

We want to retrieve the locations and amplitudes of the Diracs from (5.6), but now we make no assumption on the sampling kernel. We instead find proper coefficients for $\varphi(t)$ to approximate the exponentials $\mathrm{e}^{\alpha_m t}$, where $m = 0, \dots, P$, $\alpha_m = \alpha_0 + m\lambda$ and $\alpha_0, \lambda \in \mathbb{C}$. From the previous section we know that a good quality of the reproduction is achieved if we choose the constant least-squares coefficients

$$c_{m,n} = c_{m,0}\mathrm{e}^{\alpha_m n}, \qquad \text{with } c_{m,0} = \hat{\varphi}(\alpha_m)^{-1}.$$

We thus only need to know the Laplace transform of $\varphi(t)$ at α_m, $m = 0, \dots, P$. Also, note that P no longer needs to be related to the support of $\varphi(t)$, but we can use any value subject to $P \geqslant 2K - 1$.

In order to retrieve the innovation parameters (t_k, a_k), we proceed in the same way as in the case of exact reproduction of exponentials, but now we have that the moments are

$$\begin{aligned} s_m &= \sum_{n=0}^{N-1} c_{m,n} y_n = \left\langle x(t), \sum_{n=0}^{N-1} c_{m,n} \varphi\left(\frac{t}{T} - n\right)\right\rangle \\ &= \sum_{k=0}^{K-1} x_k u_k^m - \underbrace{\sum_{k=0}^{K-1} a_k \varepsilon_m\left(\frac{t_k}{T}\right)}_{\zeta_m} \end{aligned} \tag{5.7}$$

where $x_k = a_k \mathrm{e}^{\alpha_0 \frac{t_k}{T}}$ and $u_k = \mathrm{e}^{\lambda \frac{t_k}{T}}$. There is a model mismatch due to the approximation error $\varepsilon_m(t)$ of (5.4), equal to ζ_m. We treat it as noise, and retrieve the parameters of the signal using the methods of Chapter 3. The model mismatch depends on the quality of the approximation, dictated by the coefficients $c_{m,n}$, the values α_m and P, and the kernel $\varphi(t)$. If ζ_m is negligible when compared to other forms of noise then the procedure is sufficiently good. In close-to-noiseless settings, however, the estimation of the Diracs can be refined using the iterative method of Algorithm 5. The basic idea of the algorithm is that, given an estimate of the locations of the Diracs, we can compute an approximation of ζ_m and use it to refine the computation of the moments s_m.

We conclude by highlighting that when $K = 1$ Diracs we can analyse the convergence by writing the solution in a fixed-point iteration form, as we show in Appendix D.3. In this way, we may establish sufficient conditions for Algorithm 5 to converge in the simple case of recovering $K = 1$ Diracs.

Algorithm 5 Recovery of a train of K Diracs using approximation of exponentials

1: Compute the moments $s_m^0 = \sum_n c_{m,n} y_n$, from the original data and set $s_m^i = s_m^1 = s_m^0$.

2: Build the system of equations (2.24) using s_m^i and retrieve the annihilating filter coefficients h_m, for $m = 0, \dots, M$, where $M \geqslant K$.

3: Calculate the values u_k^i from the roots of h_m, and determine the locations t_k^i, for the ith iteration.

4: Find the amplitudes a_k^i from x_k^i, obtained by solving the first K consecutive equations in (2.21).

5: Recalculate the moments for the next iteration $i+1$ by removing the model mismatch from the moments calculated from the original data using

$$s_m^{i+1} = s_m^0 + \sum_{k=0}^{K-1} a_k^i \varepsilon_m(t_k^i),$$

for $m = 0, \dots, P$ and where $\varepsilon_m(t)$ is the error of the approximation (5.4).

6: Repeat steps 2 to 5 until convergence of the values (a_k^i, t_k^i).

5.2.1 How to select the exponents α_m

In Chapter 4 we have determined that, if we have full control on the design of the sampling kernel, we should use as many moments as samples, $P + 1 = N$, the exponents should be purely imaginary and of the form:

$$\alpha_m = j\omega_m = j\frac{\pi}{L}(2m - P) \quad m = 0, \dots, P, \tag{5.8}$$

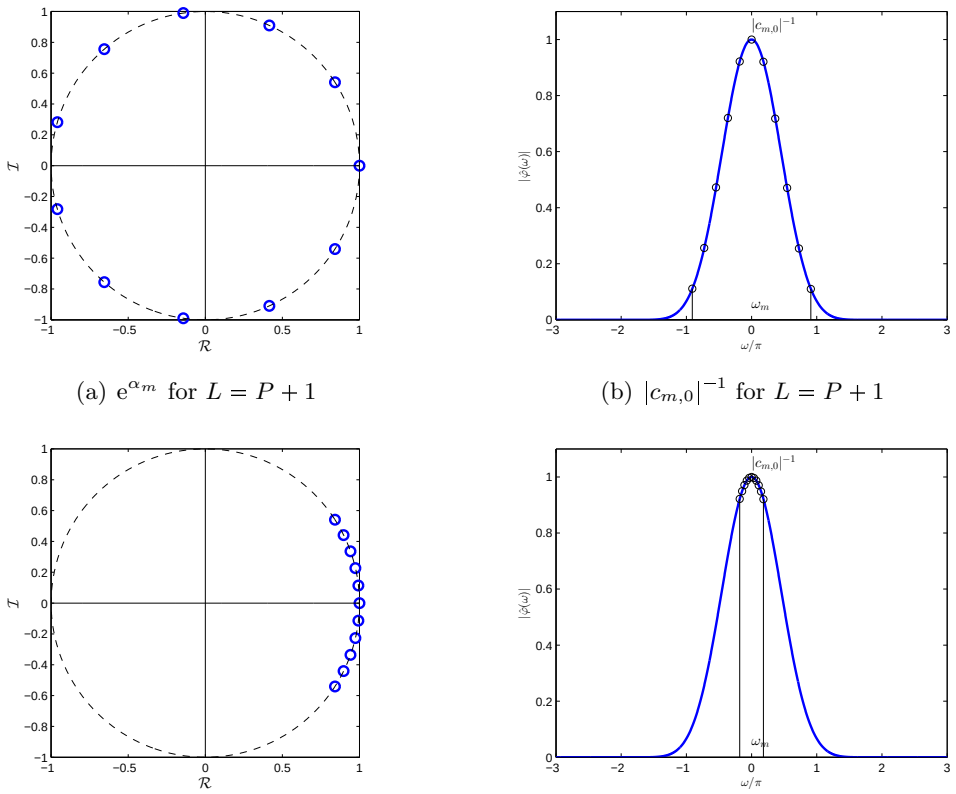
where $L = P + 1$, and the coefficients $c_{m,n}$ should be such that $|c_{m,0}| = 1$ for $m = 0, \dots, P$. This type of construction led to optimal kernels.

However, in the approximated FRI scenario, the sampling kernel is fixed and we can only choose the number of moments $P+1$ and the values $\alpha_m = j\omega_m$ but we cannot impose $|c_{m,0}| = 1$. If we follow the rules used for eMOMS, on the one hand we want that e^{α_m} span the unit circle and, on the other hand, we want $|c_{m,0}|$ to be as close as possible to 1. These requirements lead to a tradeoff in the choice of α_m since the former means that $L = P + 1$, whereas the latter occurs when all ω_m are very close to each other [2], which means that L should be as large as possible. On way for solving the tradeoff is to use exponents of the form (5.8) and then use an optimisation criterion to determine the only remaining free parameters P and L that optimise the above tradeoff. The criterion we

[2] For example, for coefficients such that $c_{m,0} = \hat{\varphi}(\omega_m)^{-1}$, and a filter that is approximately lowpass, then when all ω_m are close to zero then the absolute values of the coefficients are approximately constant.

follow is to choose the values of P and L that minimise the CRB (3.13) when retrieving the location of a single Dirac. We show examples of this procedure in the simulations.

We conclude the section by showing an example of how the roots e^{α_m} change in the unit circle compared to the values $c_{m,0} = \hat{\varphi}(\omega_m)^{-1}$ in order to better understand the tradeoff. In Figure 5.2 (a, b) we plot how with $L = P + 1$ we span the unit circle but make the values $|c_{m,0}|^{-1}$ too different from each other, worsening the condition of $\mathbf{C}$. In Figure 5.2 (c, d) we show that when L is quite large $|c_{m,0}|^{-1}$ become similar to each other but at the expense of concentrating the roots e^{α_m} on the unit circle, which also worsens the condition of $\mathbf{C}$. In all cases the kernel is a B-Spline of order $M + 1 = 6$ and the exponents are of the form (5.8) with $P + 1 = 11$.



(a) e^{α_m} for $L = P + 1$

(b) $|c_{m,0}|^{-1}$ for $L = P + 1$

(c) e^{α_m} for $L = 5(P + 1)$

(d) $|c_{m,0}|^{-1}$ for $L = 5(P + 1)$

Figure 5.2: ***Tradeoff for the choice of*** L**. In (a,b) we plot how** $L = P+1$ **spans the unit circle but widens the values** $|c_{m,0}|^{-1}$**. In (c,d) we show that a large** L **makes** $|c_{m,0}|^{-1}$ **similar to each other but concentrates the roots** e^{α_m}**. The kernel is a B-Spline with** $M + 1 = 6$ **and the exponents** (5.8) **with** $P + 1 = 11$**.**

5.3 Simulations

In Section 5.2 we have presented a method to recover a train of K Diracs from the samples (5.6) taken uniformly by any kernel $\varphi(t)$ that can approximately reproduce the set of exponentials $e^{\alpha_m t}$, $m = 0, \ldots, P$. We have assumed no other sources of error than the

model mismatch due to the approximation. Now we consider the scenario of adding white Gaussian noise to the samples (5.6), and we show that it is still possible to reconstruct the train of K Diracs using the techniques of Chapter 3. We concentrate on case studies for B-Splines and Gaussian kernels.

We compare the state-of-the-art algebraic techniques developed in [10] for B-Splines and in [60] for the Gaussian kernels with our universal reconstruction. Note that for existing methods the former is aided by a preconditioning step explained in Appendix D.4 and further stabilised with prewhitening (Chapter 3), and the latter is aided by the preconditioning step explained in [60].

5.3.1 The experimental setup

The setup is the same as that of Chapters 3 and 4: We take N samples by computing $y_n = \sum_{k=0}^{K-1} a_k \varphi\left(\frac{t_k}{T} - n\right)$ for $n = 0, \dots, N-1$. The sampling period is $T = \frac{1}{N}$ unless specified otherwise. We then either use the noiseless samples or corrupt them with additive white Gaussian noise of variance σ^2, according to the target signal-to-noise ratio $\text{SNR(dB)} = 10 \log \frac{\|\mathbf{y}\|^2}{N\sigma^2}$. We finally calculate the noisy $P+1$ moments and then retrieve the innovation parameters $\{a_k, t_k\}_{k=0}^{K-1}$ of the input. We always use the subspace estimator method as our recovery algorithm, as described in Section 3.3 with $\mathbf{W} = \mathbf{R}_{\mathbf{B}}^{\dagger/2}$ for whitening the data.

We present results for single experiments or for average performance over multiple trials. For the latter, we show the root mean square error (RMSE) of the locations:

$$\Delta t_k = \sqrt{\frac{\sum_{i=0}^{I-1} (\hat{t}_k^{(i)} - t_k)^2}{I}},$$

where $\hat{t}_k^{(i)}$ is the i-th estimated time location, I is the total number of realisations and e_{t_k} the error for each of the K Diracs. We do this for a range of fixed signal-to-noise ratio (SNR) values and average the effects using $I = 1000$ noise realisations at each SNR. We compare the RMSE with the square root of the variance predicted by the sample-based CRB (3.11) and the moment-based CRB (3.13). In order to anticipate the behaviour of our algorithm, we must take into account that the CRB associated to the samples depends directly on the sampling kernel, whereas the CRB associated to the moments depends on the exponential functions we *approximate*.

5.3.2 Case study 1: Universal FRI reconstruction with B-Spline kernels

First of all, in Appendix D.4 we provide a summary of the exact recovery scheme for polynomial reproducing kernels, as described in [10]. We compare this algebraic method aided by preconditioning (see Appendix D.4) and prewhitening to our universal reconstruction.

For the recovery based on approximation of exponentials we have to choose the exponential parameters α_m given the B-Spline kernel of order $M+1$ and the number of moments $P+1$ we want to generate. In order to do this, we use parameters of the form (5.8) and

then determine P and L that minimise the CRB (3.13) when retrieving the location of a single Dirac. Among the various types of coefficients summarised in Table 5.1 we concentrate on constant-least squares since they provide very good reproduction of exponential functions with parameters (5.8). These are $c_{m,n} = \hat{\varphi}(\alpha_m)^{-1} \mathrm{e}^{\alpha_m n}$ for $m = 0, \ldots, P$.

We have seen experimentally that $P+1$ can be chosen arbitrarily but is generally greater or equal than the support of the kernel $M+1$. Once $P+1$ is selected, experimental evidence also suggests that the best L is normally in the range $P+1 \leqslant L \leqslant 4(P+1)$ [3]. For the rest of this setup we work with a B-Spline of either order $M+1 = 6$ or order $M+1 = 16$ and decide the value of $P+1$ depending on the simulation. We show an example of the choice of L in Figure 5.3 when the kernel is a B-Spline of order $M+1 = 16$. Here, for $N = 31$ we fix $P+1$, vary L, calculate α_m and $c_{m,n}$, to then plot the CRB (3.13) (with $\sigma = 1$) for a single Dirac. The minima in this example are always around $L = 1.5(P+1)$.

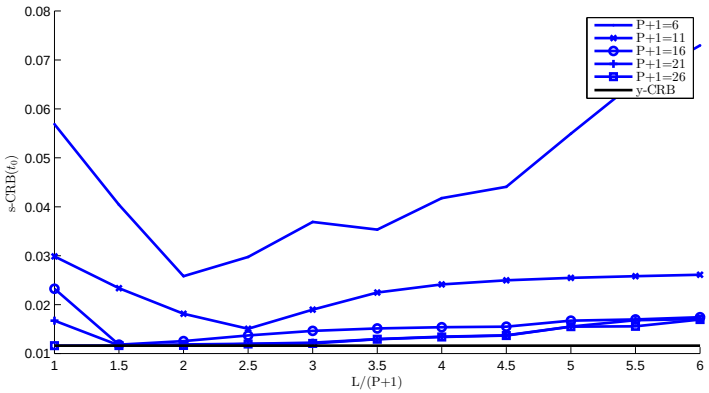


Figure 5.3: *CRB vs.* L. Here we plot moment-based CRB values (3.13) **($\sigma = 1$) for exponential parameters** (5.8) **for different values of P when we vary L. We use the constant least-squares coefficients and a B-Spline of order $M+1 = 16$. Note that the minima are always obtained around $L = 1.5(P+1)$.**

We now show two different sets of simulations. The first compares the retrieval methods based on exact reproduction of polynomials and approximate reconstruction of exponentials in Figure 5.4. The second, is an example of how the retrieval based on approximate FRI can be finely tuned to reach the sample-based CRB in Figure 5.5.

In Figure 5.4 (a-b) we show the deviation in the location for $K = 1$ Dirac that has been sampled using a B-Spline kernel of order $M+1 = 16$. We compare the performance (a) when we use the retrieval technique based on reproduction of polynomials with (b) the retrieval method when we use approximation of exponentials. Here, $\alpha_m = j\frac{\pi}{2(P+1)}(2m-P)$ with $m = 0, \ldots, P$. Both recovery methods are applied to $N = 31$ noisy samples, generate $M+1 = P+1 = 16$ moments and are aided by pre-whitening. As shown in the figure it is only in the latter case that the kernel is able to reach the sample-based CRB.

In Figure 5.4 (c-d) we show a single realisation of the recovery of $K = 6$ Diracs that

[3] We notice that $L = P+1$ can in practice only be used for the eMOMS kernels of Chapter 4. Also note that the higher the value of L, the worse conditioned $\mathbf{C}$ becomes, reason why the experimental upper bound $L = 4(P+1)$ makes sense.

have been sampled with a B-Spline of order $M+1=16$. The Diracs are located at random over $\tau = NT = 1$ and have arbitrary amplitudes. We compare the results of (c) when we use the recovery scheme based on reproduction of polynomials and $M+1=16$ moments, with (d) when we apply the retrieval based on approximation of exponentials, with $P+1=24$ moments and $\alpha_m = j\frac{\pi}{2(P+1)}(2m-P)$. The number of samples is $N=31$ and the signal-to-noise ratio is 20dB in both cases. Once more only the latter method is able to recover all the Diracs and with much better accuracy than the former method.

To end the comparison, in Figure 5.4 (e-f) we show the retrieval of $K=4$ Diracs from $N=31$ noiseless samples taken by a B-Spline of order $M+1=6$. The order is not sufficient to apply the exact retrieval since $M+1<2K$. Thus, in (e) only 3 Diracs are retrieved with this method, being their locations and amplitudes "averages" of the real ones. On the contrary, in (f) the approximate FRI method can retrieve all the Diracs correctly by using $P+1 \geqslant 2K$ moments. The locations and amplitudes are estimated with an error, due to the approximation of exponentials, that can be removed by using the iterative procedure of Algorithm 5. The reason for the approximate FRI method to recover all the Diracs is that $N>2K$.

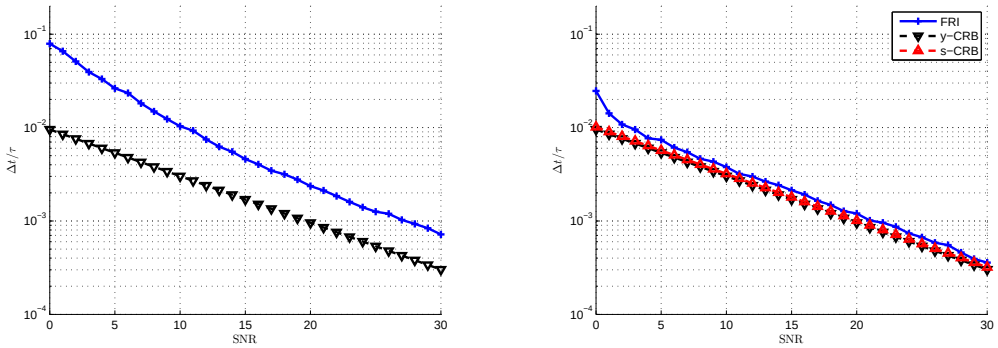
We end the simulations results for B-Splines in Figure 5.5 showing how the accuracy of the retrieval can improve by generating more moments $P+1$ from a fixed set of N samples taken by a sampling kernel of fixed order $M+1$. we use the approximate method to retrieve $K=2$ Diracs from $N=31$ noisy samples taken by a B-Spline kernel of order $M+1=6$. We use exponential parameters $\alpha_m = j\frac{\pi}{L}(2m-P)$ with $m=0,\dots,P$ and $L=1.5(P+1)$. In Figure 5.5 (a-d) we show that, even though the order of the kernel is fixed at $M+1=6$, we improve the performance by generating more moments, that is, by choosing $P>M$. As the number of moments increases, the performance improves to eventually reach the sample-based CRB as shown in Figure 5.5 (d).

5.3.3 Case study 2: Universal FRI reconstruction with Gaussian kernels

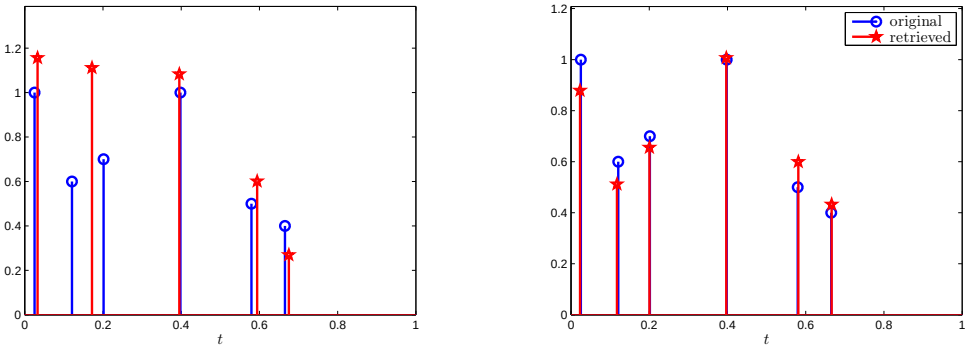
In Appendix D.4 we provide a summary of the exact recovery scheme for Gaussian kernels, as described in [2]. We compare this algebraic method aided by preconditioning (see [2]) with our universal reconstruction.

For the recovery based on approximation of exponentials, we have to choose the exponential parameters α_m given the Gaussian kernel of standard deviation γ and the number of moments $P+1$ we want to generate. In order to do this, we use parameters of the form (5.8) and then determine P and L that minimise the CRB (3.13) when retrieving the location of a single Dirac. We concentrate on the constant-least squares coefficients which, according to Table 5.2 now take the form $c_{m,n} = T\hat{h}_\gamma(\alpha_m)^{-1}\mathrm{e}^{\alpha_m nT}$ for $m=0,\dots,P$.

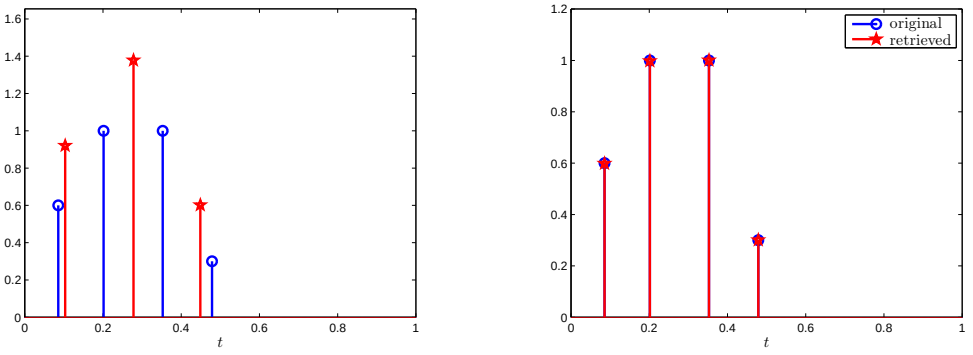
To get an idea of the potential of our algorithm, we use an experimental setup similar to that of Tan and Goyal's for their Gibbs algorithm [71]. We first show how to choose L in Figure 5.6 for a Gaussian kernel with standard deviation $\gamma=1$ that we use in the simulations afterwards. To begin, we fix $N=31$ and $T=\frac{2}{3}$. Then, we choose $P+1$,



(a) Default (b) Approximation

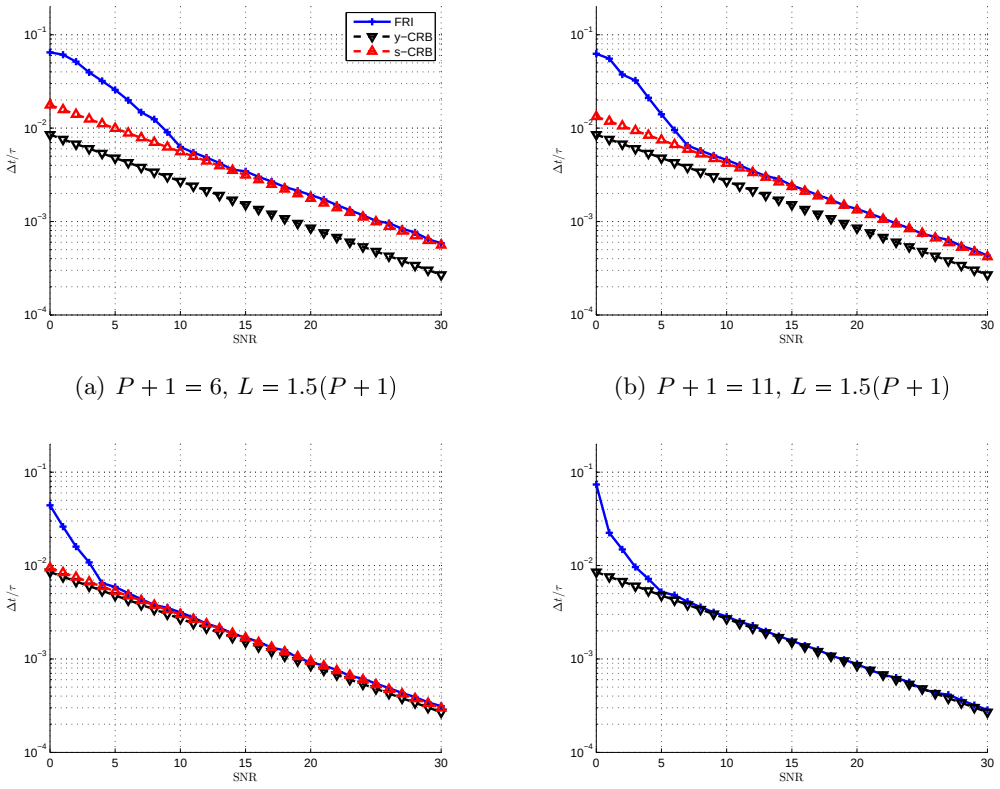


(c) Default (d) Approximation



(e) Default (f) Approximation

Figure 5.4: ***Exact vs. approximated FRI with B-Splines.*** **1) Deviation in the location for $K = 1$ Dirac that has been sampled using a B-Spline kernel of order $M + 1 = 16$. (a) is for the recovery based on polynomial reproduction, enhanced using pre-whitening. (b) is for the retrieval based on approximate reproduction of exponentials with $\alpha_m = \frac{\pi}{2(P+1)}(2m - P)$, $m = 0, \ldots, P$ and $P + 1 = 16$. Only the latter case reaches the CRB. 2) Reconstruction of $K = 6$ Diracs sampled with a B-Spline of order $M + 1 = 16$ from $M + 1 = P + 1 = 16$ moments. (c) illustrates the recovery based on reproduction of polynomials for and (d) shows the reconstruction based on approximation of exponentials. Only the latter is able to retrieve all the Diracs. The SNR for is 20dB. 3) Recovery of $K = 4$ Diracs in the absence of noise, sampled with a B-Spline of order $M + 1 = 6$. (e) is for the polynomial based method for which the number of moments is not sufficient to retrieve the Diracs ($M + 1 < 2K$). (f) is for the approximate FRI method that can generate $P + 1 \geqslant 2K$ moments to retrieve all the Diracs. The number of samples is $N = 31$ for all the simulations.**



(a) $P+1=6$, $L=1.5(P+1)$

(b) $P+1=11$, $L=1.5(P+1)$

(c) $P+1=21$, $L=1.5(P+1)$

(d) $P+1=31$, $L=1.5(P+1)$

Figure 5.5: ***Approximated FRI with B-Splines.*** **These figures show the error in the estimation of the first Dirac out of $K=2$ by using the approximated FRI recovery. The error for the second Dirac is very similar. We show how, even when we fix the order of the kernel $M+1=6$, we can reconstruct any number of moments $P+1$ and improve the performance. By properly selecting the exponential parameters the performance improves until it (d) eventually reaches the sample-based CRB.**

vary L and calculate α_m and $c_{m,n}$. Finally, we plot the CRB (3.13) (with $\sigma = 1$) for a single Dirac. The minima for the s-CRB in this example are when L is in the range $1.5 \cdot T(P+1) \leqslant L \leqslant 4 \cdot T(P+1)$. We may choose any pair $(P+1), L$ such that the condition number for $\mathbf{C}$ is low enough to apply prewhitening successfully. For instance, when we use $P+1 = 16$ and $L = 2.0 \cdot T(P+1)$ the condition number of $\mathbf{C}$ is less than 13. Other good choices are $P+1 = 21$ and $L = 1.5 \cdot T(P+1)$ with a condition number of less than 91.

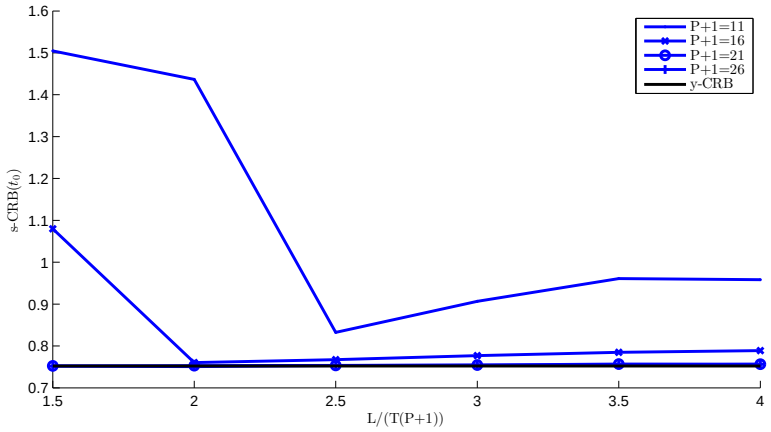
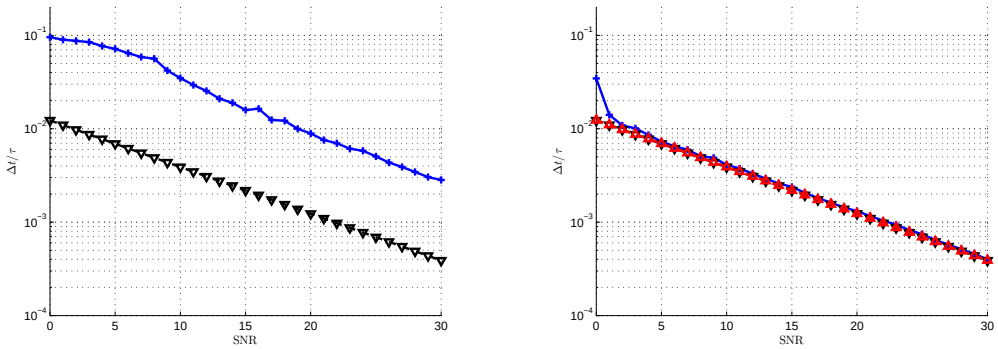


Figure 5.6: ***CRB vs.*** L**. Here we plot the CRB values** (3.13) ($\sigma = 1$) **for exponential parameters** (5.8) **when we vary** P **and** L **given** $N = 31$ **samples and** $\gamma = 1$**. Note that the minima are for** $1.5 \cdot T(P+1) \leqslant L \leqslant 4 \cdot T(P+1)$**.**

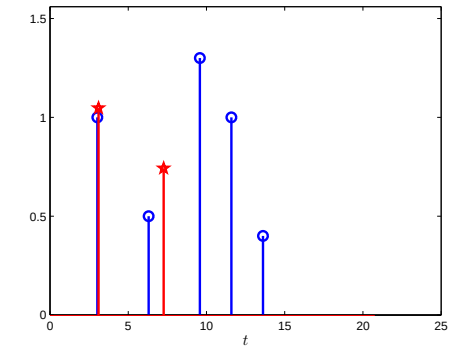
In Figure 5.7 (a-b) we show the deviation in the location for $K = 1$ Dirac that has been sampled using a Gaussian kernel of $\gamma = 1$. We compare the performance (a) when we use the retrieval technique of [60] with (b) the retrieval method when we use approximation of exponentials. Here, $\alpha_m = j\frac{\pi}{2(P+1)}(2m - P)$ with $m = 0, \ldots, P$ and $P + 1 = 16$. Both recovery methods are applied to $N = 31$ noisy samples, taken with $T = \frac{2}{3}$, and are aided by pre-whitening. It is only in the latter case that the kernel is able to reach the sample-based CRB.

In Figure 5.7 (c-d) we show a single realisation of the recovery of $K = 5$ Diracs that have been sampled with the same Gaussian kernel. The Diracs are located at random over $[0, 15]$ and have arbitrary amplitudes. We compare the results of (c) when we use the original recovery scheme with (d) when we apply the retrieval based on approximation of exponentials, with $P + 1 = 16$ moments and $\alpha_m = j\frac{\pi}{2T(P+1)}(2m - P)$. The number of samples is $N = 31$ and the signal-to-noise ratio is 20dB in both cases. Once more only the latter method is able to recover all the Diracs and with much better accuracy than the former method.

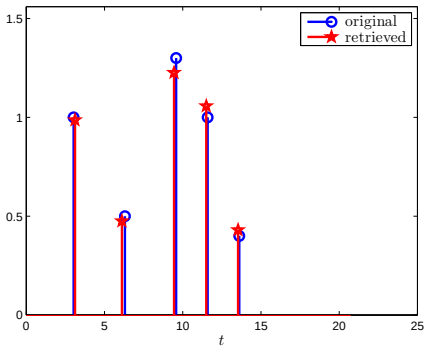


(a) Default

(b) Approximation



(c) Default



(d) Approximation

Figure 5.7: ***Gaussian kernel behaviour.*** **1) Deviation in the location of a single Dirac retrieved from $N = 31$ samples taken with period $T = \frac{2}{3}$ by a Gaussian kernel with $\gamma = 1$. (a) reconstruction based on the exact recovery scheme and in (b) results for the approximated retrieval. 2) Recovery of $K = 5$ Diracs from $N = 31$ samples taken by a Gaussian kernel of standard deviation $\gamma = 1$. (e) shows the results of the original technique and (f) the results of the retrieval based on approximation of exponentials, both for** SNR $= 20$**dB.**

5.3.4 Effect of the approximation error on the accuracy of the reconstruction

In this example we test the hypothesis that better approximation of exponentials leads to more accurate reconstruction of Diracs. Suppose we sample $K = 1$ Diracs with a linear spline and we want to recover its location using approximation of exponentials. In Figure 5.1 we have shown that the linear spline can approximate complex exponentials of lower frequencies better than those with higher frequencies. We now generate four moments s_m using the constant least-squares coefficients that are associated to the same exponentials $\mathrm{e}^{\alpha_m t} = \mathrm{e}^{j\frac{\pi}{16}(2m-7)t}$ for $m = 0, \ldots, 3$ of Figure 5.1. Finally, we compare the estimation of the location of the Dirac obtained from the moments associated to the higher frequencies (HF) s_0 and s_1 to the estimation obtained from the moments associated to the lower frequencies (LF) s_2 and s_3.

In Table 5.3 we summarise the results of the root mean squared error of the estimation obtained from either pair of moments. The error is averaged over 100 realisations each of which corresponds to placing the Dirac at a location $t_0 = 0.15(i-1)$ for $i = 1, \ldots, 100$. As expected the approximation with lower frequency achieves a better performance.

Table 5.3: Accuracy of the reconstruction

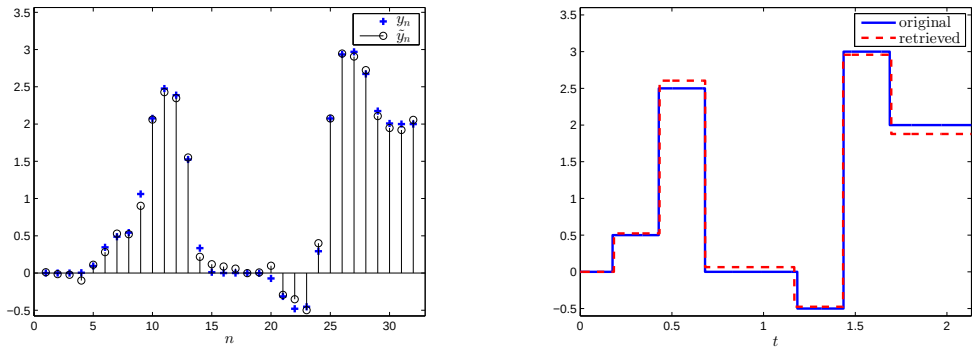
	HF moments		LF moments	
	s_0	s_1	s_2	s_3
Approximation error	0.061	0.028	0.0093	0.00098
Reconstruction error	0.0018		0.00019	

5.3.5 Alternative FRI signals

We conclude the simulations by showing that it is possible to adapt the approximate FRI framework to sample and reconstruct alternative FRI signals. For example, sampling a piecewise constant function with a kernel $\varphi(t)$ and calculating the first finite difference of the samples $z_n = y_n - y_{n+1}$ yields the same measurements as sampling the derivative of the signal with $\phi(t) = \varphi(t) * \beta_0(t)$, where $\beta_0(t)$ is a box function [10]. The derivative of the signal is a train of K Diracs. Consequently, we may recover the signal by calculating $c_{m,n}$ for the linear combination of shifted versions of $\phi(t)$ to approximate exponentials and then applying the annihilating filter method to the moments $s_m = \sum_n c_{m,n} z_n$.

We illustrate the process in Figure 5.8. Here, we sample a piecewise constant function with $K = 6$ discontinuities using a B-Spline kernel of order $M + 1 = 6$. The sampling period is $T = \frac{1}{15}$. In Figure 5.8 we (a) contaminate the $N = 31$ samples with additive white Gaussian noise and calculate their first order difference. Then, we generate $P + 1 = 21$ moments using the constant least-squares coefficients from exponential parameters $\alpha_m = j\frac{\pi}{L}(2m - P)$ with $m = 0, \ldots, P$ and $L = 1.4(P + 1)$. The signal-to-noise ratio is 25dB. Note that the order of the spline is not sufficient to apply the retrieval method based on

reproduction of polynomials of [10]. On the contrary, we can use the method based on approximation of exponentials as long as $P + 1 \geqslant 2K$. The original and reconstructed signal are shown in Figure 5.8 (b).



(a) Samples

(b) Retrieved signal

Figure 5.8: ***Piecewise constant functions and B-Splines.*** **These figures show the sampling and retrieval process, based on approximation of exponentials, for a piecewise constant function with $K = 6$ discontinuities in the presence of noise of 25dB.**

5.4 Summary

In this chapter we have extended the results of FRI reconstruction by allowing for the linear combination of integer shifts of arbitrary sampling kernels to approximate exponential functions. This allows us to always map the signal reconstruction problem to Prony's method in line-spectral estimation theory, regardless of the sampling kernel. We must note that the property of reproducing exponential functions is common to many FRI recovery procedures.

We have shown that for kernels that approximately satisfy the generalised Strang-Fix conditions it is possible to find coefficients $c_{m,n}$ such that the reproduction of exponentials is approximate rather than exact. We have used these coefficients $c_{m,n}$, along with carefully chosen exponential parameters α_m, to retrieve FRI signals from the samples obtained using these kernels. This new approach is universal since it can be used with any kernel $\varphi(t)$. In addition, we have proposed an iterative algorithm that is able to make the reconstruction error of FRI signals, due to the model mismatch, arbitrary small in the absence of other sources of noise. Finally, we have proved that this new approach can be used to increase the resiliency to noise of some unstable kernels proposed in the FRI literature. Specifically of polynomial splines and Gaussian kernels, for which the original setups lead to very ill-conditioned reconstruction.

Chapter 6

Spike sorting at sub-Nyquist rates

Communication between neurons is carried out by action potentials (spikes) propagating as electrochemical impulses along the nervous system. Much is to be learnt from the way neurons are interconnected and how they communicate in response to external stimuli to the body. Understanding the neuronal code could provide invaluable medical information on neurological diseases such as epilepsy or Alzheimer's disease and our knowledge on the physiological structure of the brain. Spike sorting analyses the brain activity at neuron resolution and relies on the ability to detect the temporal occurrence of action potentials and their relation to specific neurons, helping the analysis of brain activity.

Spike sorting has been shown to be successful at monitoring a limited number of neurons, for instance by using a microwire implanted in the brain. However, substantial information in order to study how communication inside the brain occurs requires the problem to be scaled up. According to Shannon, since spikes typically contain frequencies up to 8kHz, sampling rates of at least 16kHz are normally required. This poses fundamental problems for simultaneous multichannel spike sorting in terms of energy consumption, computational complexity and hardware demands.

The activity of a neuron can be viewed as a temporal point process of identical spikes. Furthermore, the firing rate of neurons is by nature very low and action potentials can be shown to be approximately sparse in the wavelet domain. These conditions make neuronal information suitable to modern sampling techniques, such as finite rate of innovation or compressed sensing, advocating for an economic acquisition of information.

In this chapter we propose a novel algorithm capable of sampling and reconstructing neuronal data at sub-Nyquist rates, preserving enough features of the original signal so that spike sorting is performed equally reliably. This was joint work with Jose Caballero. The results obtained during the development of the algorithm led to the paper [82].

The chapter is organised as follows: In Section 6.1 we introduce the concept of spike sorting and explain the motivation for the development of our algorithm. In Section 6.2 we describe the modules of our proposed algorithm for low-sampling-rate acquisition and reconstruction of neuronal activity signals. We then show the simulation results in Section 6.3 and conclude in Section 6.5.

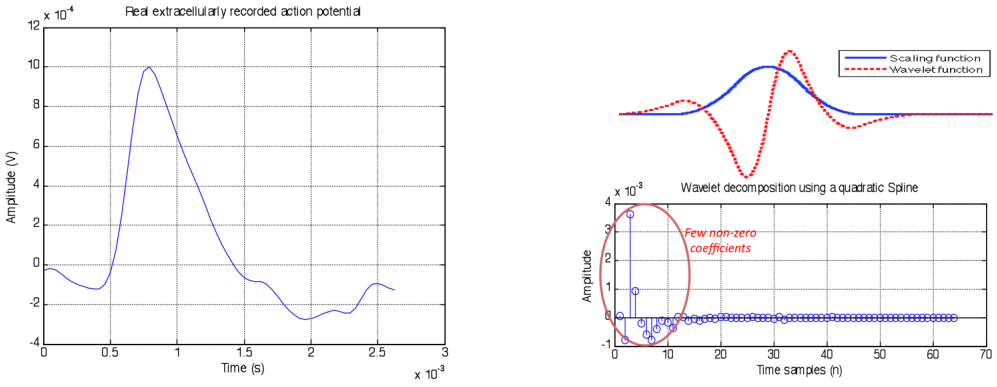
6.1 Spike sorting

A spike is the electrochemical action potential fired by a neuron for data transmission through the nervous system and their distinctive signature shapes largely depend on their morphology and the recording process.

Since all spikes of a given neuron look alike, the form of the action potential appears not to carry any information regarding the stimulus that caused the spike [83]. Rather, it is the frequency with which they are fired, and the distribution and number of neurons that generate them in a given nerve that have a useful meaning [84]. Therefore, spike sorting algorithms allow to study neuronal populations, because action potential shapes are believed to be useful to distinguish among neurons and classes of neurons (shapes look very much the same for the same neuron, similar for the same type of neurons, and dissimilar among different classes of neurons [85]).

The objective of spike sorting algorithms is to detect action potentials and identify which neuron generated them. The vast majority follow three basic steps. They begin with a spike detection stage, mainly achieved by voltage thresholding with respect to an estimation of the noise amplitude in the signal. Then, a feature extraction step characterises detected spikes, the main property looked for among these features being that they present a multimodal distribution that ideally allows to separate spikes fired by different neurons. Principal Component Analysis (PCA) and wavelet decomposition have widely been used in the literature for feature extraction [86–89]. To end, and based on these features, a clustering step is necessary to relate each spike to a particular neuron.

Existing algorithms suffer from scalability issues due to high sampling rates. Neural activity from one neuron has been shown to be compressible [90, 91], and as such it is suitable for sparse sampling. In fact, it has been empirically shown [90] that about 1/6 of wavelet transform coefficients of a spike gather around 99% of the signal energy. We show an example of the wavelet decomposition of a real spike in Figure 6.1.



(a) Action potential

(b) Wavelet decomposition

Figure 6.1: ***Sparsity in the wavelet domain.*** **These figure shows the a recorded real neuronal action potential and its wavelet decomposition using a quadratic spline. It is clear that only a few of the wavelets coefficients are representative of the spike shape.**

Reducing the sampling frequency would imply that large simultaneous extracellular recordings could be obtained and the additional reliability provided by multi-channel recordings would be exploited in the sorting. Improvement in sampling techniques is also crucial to scaling up the recording technologies to hundreds or thousands of neurons, i.e. to the population sizes really necessary to understand brain function and to control the next generation of neuroprostheses.

Since spike sorting is not the main purpose of the algorithm developed in this chapter, we do not deal with the subject any further. For a comprehensive review of spike sorting techniques the reader is referred to the surveys in [92–94].

6.2 Design of the algorithm

In this section we propose the algorithm that we use to sample the neuronal data below Nyquist rate. A relatively realistic signal modelling allows to split the input data into simpler units, each of which consists in the convolution of two basic signals. Then, the problem is transformed into a sequential signal and system estimation problem.

We first propose the signal model, then describe how to separately estimate each of the two basic signals and finally merge everything together into an iterative algorithm.

6.2.1 Modelling the neuronal signal

There are multiple mathematical models that describe how action potentials in neurons are initiated and propagated. A well known and comprehensive example is the Hodgkin–Huxley model, from which many others are derived. Several simplifications have been proposed over the years, among which the Spike Response Model (SMR) is one of the most commonly used. Therefore, we consider that a single action potential can be described as follows [83]:

$$u(t) - u_{\text{rest}} \approx ap(t - t_0) \tag{6.1}$$

where only the spike amplitude a, its shape $p(t)$ and the moment when it is triggered t_0 are relevant. Here, $u(t)$ refers to the measured voltage signal and u_{rest} is the resting potential of the neuron.

The activity of a neuron can be interpreted as a point process in which roughly the same spike is fired at different instants of time. Assuming stationary neurons and no bursting exists, the amplitudes of the spikes are constant for the same neuron. Thus, the neuronal signal to be sampled from neuron i can be thought of as the result of the convolution of its spike shape with a train of Diracs, i.e.

$$x_i(t) = p_i(t) * \sum_{k=0}^{K-1} a_i \delta(t - t_{i,k}),$$

where $p_i(t)$ is the spike, a_i is the constant amplitude and $t_{i,k}$ are the firing instants. If we consider the contribution of I different neurons, which add up linearly, we can write the

complete model for the neuronal signal as

$$\tilde{x}(t) = x(t) + n(t) = \sum_{i=0}^{I-1} x_i(t) + n(t),$$

where $n(t)$ accounts for any source of noise.

For the sake of clarity we assume that contributions do not overlap. In such a case, we can sequentially retrieve one pulse after the other and simply write that $x(t) = p(t) * d(t)$, where $p(t)$ is the pulse shape of the neuron and $d(t)$ is one Dirac that represents the amplitude and location of the spike.

6.2.2 How to extract the information on $d(t)$ given $p(t)$

The signal $x(t)$ is acquired using an exponential reproducing kernel as discussed in Chapter 2 leading to the N samples y_n as in (2.18). The Dirac signal $d(t)$ can be retrieved from prior knowledge on the pulse shape $p(t)$ and the samples y_n. If $p(t)$ is known exactly, then we can write

$$\begin{aligned} y_n &= \left\langle x(t), \varphi\left(\frac{t}{T} - n\right)\right\rangle = \left\langle p(t) * d(t), \varphi\left(\frac{t}{T} - n\right)\right\rangle \\ &= \left\langle d(t), \varphi_p\left(\frac{t}{T} - n\right)\right\rangle, \end{aligned} \tag{6.2}$$

where now we have an equivalent exponential reproducing kernel $\varphi_p(t) = \varphi(t) * p^*(-t)$.

Then the location and amplitude that characterise $d(t)$ can be retrieved using the annihilating filter procedure with the new set of moments $s_m^p = \sum_n c_{m,n}^p y_n$, where $c_{m,n}^p$ are the coefficients so that $\varphi_p(t)$ satisfies the exponential reproducing formula (2.7).

6.2.3 How to extract the information on $p(t)$ given $d(t)$

In this chapter we make use of the eMOMS kernels proposed in Chapter 4 to sample the neuronal signals. We choose exponential parameters $\alpha_m = j\omega_m = j\frac{\pi}{N}(2m - P)$, $m = 0, \ldots, P$ and $P + 1$ odd.

With this information, it is also possible to estimate the pulse shape $p(t)$ from the samples y_n, given prior knowledge on $d(t)$. Consider the exponential moments s_m, for which the following holds

$$\begin{aligned} s_m &= \sum_{n=0}^{N-1} c_{m,n} y_n = \sum_{n=0}^{N-1} c_{m,n} \left\langle x(t), \varphi\left(\frac{t}{T} - n\right)\right\rangle \\ &= \left\langle x(t), \mathrm{e}^{\alpha_m \frac{t}{T}}\right\rangle = \int_{-\infty}^{\infty} x(t)\mathrm{e}^{\alpha_m t}\mathrm{d}t, \qquad m = 0, \ldots, P. \end{aligned} \tag{6.3}$$

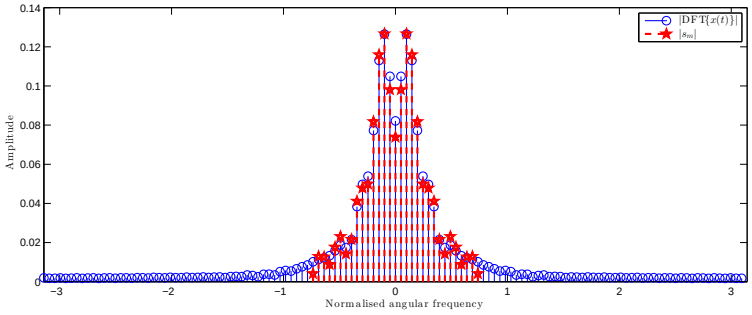
Thanks to our choice of parameters α_m, (6.3) is precisely the Fourier transform (FT)

of $x(t)$ at $\omega = \omega'_m = \frac{\omega_m}{T}$. Therefore, the moments s_m satisfy

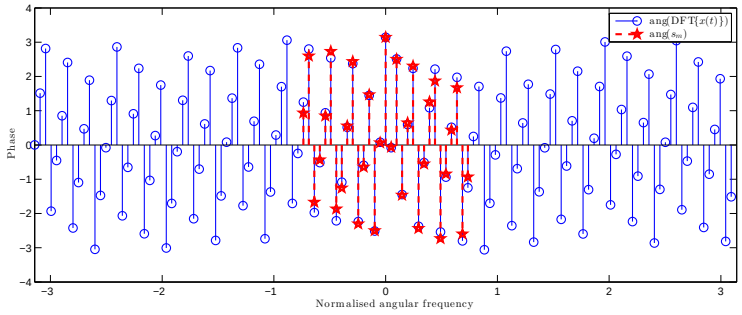
$$s_m = \hat{x}\left(\omega'_m\right) = \hat{p}\left(\omega'_m\right)\hat{d}\left(\omega'_m\right), \qquad m = 0, \ldots, P, \tag{6.4}$$

where $\omega'_m = \frac{\pi}{\tau}(2m - P)$ and $\tau = NT$. If $d(t)$ is known, it is straightforward to obtain the FT of $p(t)$ at ω'_m from the moments s_m. We then retrieve the pulse shape via an ℓ_1 minimisation as explained next.

To end the section, in Figure 6.2 we show an example of the DFT of a single spike $x(t)$ and the moments obtained after sampling the pulse and combining the samples y_n with the coefficients $c_{m,n}$. We see that the amplitude and phase of the DFT of the spike coincide with the amplitude and phase of the moments for the range $m = 0, \ldots, P$. Here we have used $N = 31$ samples, $P + 1 = 31$ moments and a DFT of 128 points.



(a) Amplitude



(b) Phase

Figure 6.2: ***DFT of the spike and its exponential moments.*** **These figure shows the DFT of a single spike and the moments obtained after sampling the pulse and combining the samples y_n with the coefficients $c_{m,n}$.**

6.2.4 Spike shape recovery

The aim of this processing block is to reconstruct a finely discretised version $\tilde{\mathbf{p}}$ of $p(t)$ from $\hat{p}\left(\omega_m\right)$, $m = 0, \ldots, P$.

Assume $\tilde{\mathbf{p}}$ is the pulse shape discretised to a vector of length L ($L \gg P$) and denote with $\tilde{\hat{\mathbf{p}}}$ the approximated discrete Fourier transform (DFT) of $\tilde{\mathbf{p}}$ obtained from (6.4). We

can write:

$$\tilde{\hat{\mathbf{p}}} = \mathbf{F}\tilde{\mathbf{p}} + \mathbf{n} \tag{6.5}$$

where $\mathbf{F}$ is the DFT matrix of size $(P+1) \times L$, $\tilde{\mathbf{p}}$ is known and is obtained from (6.4), and $\mathbf{n}$ is additive noise used to account for any model mismatch. We know that neuronal pulses can be sparsely described in the wavelet domain. We therefore rewrite (6.5) as

$$\tilde{\hat{\mathbf{p}}} = \mathbf{F}\mathbf{W}^{-1}\mathbf{w} + \mathbf{n} = \mathbf{A}\mathbf{w} + \mathbf{n}, \tag{6.6}$$

where $\mathbf{W}$ is the $L \times L$ matrix representing the wavelet transform and $\mathbf{w} = \mathbf{W}\tilde{\mathbf{p}}$ is the wavelet representation of $\tilde{\mathbf{p}}$.

The above system is underdetermined but we only need to search for a sparse vector $\mathbf{w}$ that satisfies (6.6). This modelling is reminiscent of the traditional CS framework where, in our context, the acquisition matrix is a 'fat' Fourier matrix rather than a more conventionally used random matrix. We therefore assume a sufficiently large P and a sufficiently sparse vector $\mathbf{w}$ in order to solve for $\tilde{\mathbf{p}}$ using an ℓ_1 minimisation technique such as Basis Pursuit (BP).

6.2.5 Complete algorithm

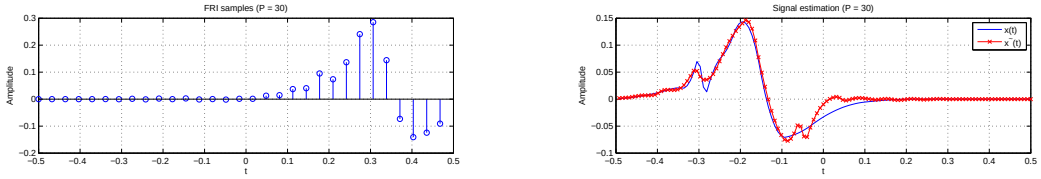
In the previous sections it has been shown how N samples are enough to recover $x(t)$ by breaking down the problem into estimating $d(t)$ and $p(t)$ separately. An iterative algorithm can thus be applied to retrieve $x(t)$ without any prior knowledge.

Assume we initialise the algorithm by setting $p(t) = \delta(t)$, meaning that at the first iteration the kernel $\varphi_p(t)$ coincides with $\varphi(t)$. The $d(t)$ estimation module will look for the location and amplitude of the Dirac, although the signal is actually a spike. The first estimation of $d(t)$ will therefore be inaccurate, but it is enough to obtain a good estimation of $p(t)$ using the recovery technique of Sections 6.2.3 and 6.2.4. Once there is useful information of $p(t)$, it can be used to update $\varphi_p(t)$ and the new set of coefficients $c^p_{m,n}$ to compute s^p_m. Using the updated moments the process can be repeated again, the convergence criterion being a maximum number of iterations or that a solution within a predefined tolerance is reached.

6.3 Results

The algorithm has proven to converge experimentally to the sought pulse shape at the desired location in about 5 iterations for a mean square error (MSE) convergence tolerance of 10^{-5}. The estimation of $p(t)$ is however suboptimal from the point of view of sparsity in the wavelet domain. The reason for this is probably that the algorithm is able to find a solution out of various stable regions in the solution space. One example of the estimation of $x(t)$ can be seen in Fig. 6.3.

The order $P+1$ of the eMOMS is a relevant design choice that influences the perfor-



(a) Samples

(b) Original $x(t)$ and estimated $\hat{x}(t)$

Figure 6.3: ***Reconstruction of*** $x(t)$ ***using the proposed sampling algorithm.*** **In this figure we show (b) an example of the reconstruction of a neuronal spike from the (a) samples obtained using the FRI sampling scheme.**

mance of the algorithm. Low values provide better estimation accuracy but high values are able to capture high resolution details of the spike shapes. We choose the latter because fine details are relevant for sorting.

Finally, we have assessed the suitability of the sub-Nyquist sampling algorithm for spike sorting, working with surrogate data available from the NeuroEngineering Lab at the University of Leicester. Spikes are simulated using a database of 594 different average action potentials recorded in the neocortex and basal ganglia. Three distinct spike shapes are placed at arbitrary times with normalised peak amplitude of 1 and background noise is generated with a standard deviation relative to 1 from superimposed spikes selected at random. Difficulties for sorting mainly come from similarities among spike shapes, realistic background noise and overlapping spikes susceptible of generating errors.

We compare the performance achieved feeding the original data (@24kHz) and an FRI subsampled version to the spike sorting algorithm "Wave_Clus" [89], estimating a total of 1000 action potentials. We use $N = 31$ FRI samples to represent pulses of length $L = 128$, and an E-Spline of order $P = 30$, to achieve a sampling rate reduction by a factor 4. The reconstruction of individual spikes assumes that their location is known a priori. The results are presented in Table 6.1, and they show that our method is able to preserve the performance achieved by traditional spike sorting algorithms. We have noticed that there is a decrease in missed spikes and an increase in false positives. We believe this is due to the detection threshold value, chosen proportional to the median of the absolute value of the recording [89], which is lowered since the reconstruction process slightly smoothes spikes out.

6.4 Other applications in Neuroscience

Prior to the development of the theoretical work presented in Chapters 2 to 5 of this thesis we also applied FRI recovery algorithms to the problems of detection of voltage neuronal spikes and calcium transients. Accurate time detection of action potentials is a key step needed for their posterior sorting and classification, since it allows to determine the precise occurrence of spikes. FRI can be applied to this task by considering that the input to the sampling process is a train of pulses, with possibly different shapes, contaminated with noise. Even though preliminary work showed the validity of the FRI setting, there exist

		Missed spikes		False positives		Misclassified spikes		Unclassified spikes		Success Rate	
Spike set	**Noise s.d.**	**24K C**	**5.8K F**	**24K C**	**5.8K F**	**24K C**	**5.8K F**	**24K C**	**5.8K F**	**24K C**	**5.8K F**
Easy (1)	0.05	111	135	0	2	22	21	30	20	83.7	82.2
	0.1	93	91	6	9	29	34	9	4	86.3	86.2
	0.15	143	129	7	21	50	56	1	2	79.9	79.2
	0.2	248	216	1	18	37	44	1	2	71.3	72
Difficult (2)	0.05	140	149	0	0	17	7	70	71	77.3	77.3
	0.1	101	80	0	16	418	199	0	16	48.1	69.9
	0.15	115	86	1	20	346	454	0	0	53.8	44
	0.2	160	108	3	19	441	420	0	0	39.6	45.3
(Av.)	0.125	138.88	124.25	2.24	13.13	170	154.38	13.88	14.38	67.5	69.51

Table 6.1: ***Spike sorting comparison for datasets acquired at different rates.*** **C and F stand for classical and FRI sampling respectively. We measure (i) undetected spikes, (ii) noise detected as spikes, (iii) spikes in the wrong cluster, and (iv) spikes that cannot be identified.**

robust alternative techniques to solve the problem and FRI would only contribute further if sampling schemes could be implemented at sub-Nyquist rate.

On the other hand, inferring the times of sequences of action potentials from two-photon imaging of calcium signals is an open problem whose optimal solution remains unsolved. The detection of action potentials (APs) from calcium transients offers certain advantages over traditional electrophysiological approaches, since up to thousands of spatially and immunohistochemically defined neurons can be recorded simultaneously [95]. However, due to noise, dye buffering and the limited sampling rates in common microscopy congurations, accurate detection of APs from calcium time series has proved to be a difficult problem. However, in the FRI setting and for calcium transients well fit by a single exponential, the problem is reduced to reconstructing a stream of decaying exponentials. In [95] the authors built upon our preliminary work, which showed the validity of the FRI setting, and introduce a novel approach that provides additional robustness to noise with double consistency spike search using a sliding window. The final algorithm is fast, non-iterative and parallelisable, such that spike inference can be performed in real-time. The algorithm has been reported to outperform several recently proposed methods for spike train inference from calcium imaging data.

6.5 Summary

One of the most plausible improvements for spike sorting algorithms is in their data acquisition methods, due to the natural suitability of extracellular recorded data towards sparse acquisition methods. It is enough to know the spike shape from a given neuron and the locations of occurrence (with their individual amplitudes at most if non-stationary data is considered) in order to completely define the activity of a given neuron. If an acquisition process is designed to look for this information already at the sampling stage instead of going through a classical Nyquist sampling process, acquisition could be simplified and made more economic. Moreover, the huge dimensionality reduction of data needed for feature extraction makes it also intuitive that distinctive features could probably be extracted already during sampling. That is, sampling could be interpreted as a useful tool for ad-hoc data analysis instead of only as a means to represent continuous-time signals as faithfully as possible. Wireless recording electrodes for instance could then transmit information at much lower rates to a local processing machine, reducing energy consumption.

In this chapter, we have proposed an iterative reconstruction algorithm that can estimate a neuronal signal from FRI samples that have been obtained using an exponential MOMS at reduced sampling rates. The design of the algorithm is motivated by the sparse representation of the neuronal activity signal. Our main contribution is that we show that state-of-the-art spike sorting performances can be reached with a reduction in the sampling rate of a factor 4 compared to traditional methods.

Chapter 7

Conclusions

7.1 Main contributions

In this thesis we have studied two key aspects of the reconstruction of signals with finite rate of innovation in the presence of noise: stability and accuracy. The traditional algebraic approaches provide an exact mathematical framework to sample and perfectly reconstruct various types of FRI signals. However, noise is generally present in data acquisition, making some of these methods very unstable. For instance, the use of Gaussian kernels, polynomial reproducing kernels and some exponential reproducing is potentially ill-conditioned when noise is present.

To address the aforementioned problems, we have first provided improved versions of the main algebraic methods for FRI reconstruction that take into account how the noise may become colored when we work with exponential reproducing kernels. The idea is to apply a whitening transform to the noisy data that is able to decorrelate noise samples and make them be characterised by a uniform variance. Our formulation is general, hence the methods we have described can be easily adapted to other types of sampling kernels.

In some circumstances, however, simply applying a whitening transform does not completely solve the stability problem. We have therefore used a different perspective to analyse the noisy scenario further, which consisted in optimising the exponential reproducing sampling kernel in order to make the retrieval of the parameters of a train of Diracs most accurate and stable. Selecting the best exponential parameters for our FRI problem has translated into selecting a proper matrix of coefficients $\mathbf{C}$, which is key to the stability of the reconstruction. Based on this analysis, we have proposed a new family of kernels that is most resilient to additive white Gaussian noise added to the samples. We have termed this class of kernels eMOMS, because they are part of the family of generalised E-Splines and they are of maximum order and minimum support.

Moreover, and partly based on the stability analysis, we have proposed a generalisation of the FRI framework that applies to any sampling kernel by relaxing the exponential reproducing property. The new approach is more general because it is based on approximation of exponentials, which may be achieved in several ways by just finding appropriate

coefficients. The traditional FRI framework can be viewed as a particular case in which reproduction is exact. The advantage of our formulation is that it may be used with any fixed sampling kernel, as long as enough information is available of its transfer function at just a few specific frequencies. Moreover, while reconstruction of FRI signals with this new method is not going to be exact, we have also presented an iterative algorithm that can make the error arbitrary small in the absence of noise. Finally, this new approach can be used to increase the resiliency to noise of some unstable kernels proposed in the FRI literature, such as polynomial splines or the Gaussian function which lead to very ill-conditioned reconstruction procedures.

In addition, we have introduced a Cramér–Rao lower bound formulation related to the exponential moments of the input. This bound is useful because it allows us to determine whether each of the techniques designed throughout the thesis come close to optimal. Once more, optimality should be understood in the sense that the various techniques reach the moment-based CRB. We have successfully compared our recovery method to the current state-of-the-art techniques developed prior to our work for polynomial reproducing kernel and Gaussian kernels and we have determined they behave optimally with respect to the CRB.

To conclude, we have presented an application of the theory of FRI in the field of Neuroscience. Specifically, we have proposed an algorithm capable of sampling and reconstructing neuronal data at sub-Nyquist rates, preserving enough representative features of the original signal so that spike sorting can be performed equally reliably.

7.2 Extensions and applications

Broadening the FRI paradigm to kernels that only need to approximately satisfy the generalised Strang-Fix conditions has the consequence that FRI theory may be applied in many other scenarios. For instance, situations in which the strict constraints imposed on sampling kernels by the original formulation are not satisfied, but for which the input signal can be modelled as having finite rate of innovation. Thanks to our approach, now only proper modeling of the input signal in parametric form is needed, along with information of the sampling kernel that is easy to obtain with simple calibration.

Extensions of the current work include multidimensional FRI, multichannel setups and non-uniform sampling. If the approximate FRI framework is used for two dimensional signals, such as images, then a straightforward application is on time-of-flight cameras. These are aimed at range acquisition by measuring the time difference of arrival between a transmitted pulse and the scene reflection. The advantage of using the approximate FRI framework is that there is no need to design cameras that have an FRI based acquisition system. It is the reconstruction stage that can accommodate the characteristics of existing cameras and exploit the sparsity of the Laplacian of the depth map of a typical scene.

In the thesis of Lo ic Baboulaz [96] the author used the shape of the acquisition device as an advantage to obtain superresolved images. He, however, points out that the knowledge

of the sampling kernel may be seen as a constraint in the design of the device. Therefore, his work can fully benefit from the theory of approximated FRI, which only needs the knowledge of the kernel at certain frequencies.

7.3 Future work

In Chapters 4 and 5 we provide a method to select the exponentials that should be reproduced for a given setup with the goal of optimising the recovery performance. This is done in an experimental way by minimising the CRB associated to the estimation of the innovation parameters of the input from its exponential moments. The design of exponential MOMS and the approximate FRI framework would benefit from an analytical expression for selecting the exponential parameters, the former since it would prove optimality, the latter since it would make the algorithm more robust.

In addition, in Chapter 5 we have seen in practice that the better the approximation of the selected exponentials is the better the retrieval in the absence of noise becomes. This is an intuitive result for which a mathematical derivation is difficult to obtain. The relation between the model mismatch and the error of the approximation is not simple, due to the fact that the model mismatch is calculated as a linear combination of the error evaluated at the Diracs' locations. However, we believe that by obtaining such derivation, other interesting properties of the approximate FRI scenario may be found. One possible line of research that could be followed to solve the problem is related to studying the effects of having a constant bias term in the measurements taken as a power sum, and determine how this bias affects the accuracy of the parametric estimation.

Throughout the thesis we have assumed the number of Diracs K is known before we recover the input from the given set of samples. The case with K unknown, the model order selection, is a related but quite different problem which is just as hard to solve as the parametric estimation given K. Only a few FRI publications consider the case of K unknown, for instance [97], however the authors do so from a theoretical perspective and they present their parametric estimation method given the true K. Our experiments suggest that it is rather easy to overmodel by selecting a value for K that is larger than the true one, accounting for part of the noise. This introduces spurious spikes not present in the original signal, which contribute to degrade the parametric estimation. The theory of FRI would therefore benefit from additional results for the model order selection order and in particular from robust algorithms that could simultaneously estimate K and then the input paramters.

Finally, we believe the theory of approximate FRI goes one step forward into understanding the connections in between FRI and CS. In [14] the authors already establish a preliminary connection in between both theories. Nevertheless, they remark that compressed sensing could potentially accommodate arbitrary sampling kernels and not only the ones that satisfy an annihilation property, whereas FRI could not. This situation changes due to the approximate FRI framework, which seems to indicate there is an even

tighter connection in between the two FRI and CS. Sparse signal processing would clearly benefit from the knowledge of such link and from the combination of both theories.

Appendices

Appendix A

Chapter 2

A.1 Generalised Strang-Fix conditions

An exponential reproducing kernel is any function $\varphi(t)$ that, together with a linear combination of its shifted versions, can generate exponential polynomials of the form $t^r \mathrm{e}^{\alpha_m t}$ [1,41] for $m = 0, \ldots, P$ and $r = 0, \ldots, R$. The parameters α_m are in general complex valued. In this Appendix we prove that exponential reproducing kernels satisfy the generalised Strang-Fix conditions. More specifically, a kernel $\varphi(t)$ is able to reproduce exponential polynomials, i.e.:

$$t^r \mathrm{e}^{\alpha_m t} = \sum_{n \in \mathbb{Z}} c_{m,n,r} \varphi(t-n),$$

if and only if

$$\hat{\varphi}^{(r)}(\alpha_m) \neq 0 \text{ and } \hat{\varphi}^{(r)}(\alpha_m + 2j\pi l) = 0$$

for $l \neq 0$, $r = 0, \ldots, R$ and $m = 0, \ldots, P$. Here, $\hat{\varphi}^{(r)}(s)$ represents the rth order derivative of the double-sided Laplace transform of $\varphi(t)$.

The proof is obtained from the Strang-Fix conditions for polynomial reproducing kernels, by considering the function $\psi(t) = \mathrm{e}^{-\alpha_m t}\varphi(t)$ that clearly reproduces polynomials of the form t^r for $r = 0, \ldots, R$. The Strang-Fix conditions [10, 43] state that a kernel $\psi(t)$ is able to reproduce polynomials, i.e.:

$$t^r = \sum_{n \in \mathbb{Z}} c_{r,n} \psi(t-n),$$

if and only if

$$\hat{\psi}(0) \neq 0 \text{ and } \hat{\psi}^{(r)}(2\pi l) = 0$$

for $l \neq 0$ and $r = 0, \ldots, R$. Here, $\hat{\psi}(\omega)$ is the Fourier transform of $\psi(t)$, and $\hat{\psi}^{(r)}(\omega)$ represents its rth order derivative. Then, by taking into account that the Fourier transform of $\psi(t)$ is related to the Laplace transform of $\varphi(t)$ through $\hat{\psi}(\omega) = \hat{\varphi}(\alpha_m + j\omega)$, the above

equation turns into the generalised Strang-Fix conditions for $\varphi(t)$:

$$\hat{\varphi}(\alpha_m) \neq 0 \text{ and } \hat{\varphi}^{(r)}(\alpha_m + j2\pi l) = 0$$

for $l \neq 0$, $r = 0, \ldots, R$ and $m = 0, \ldots, P$. Now, $\hat{\varphi}^{(r)}(s)$ represents the rth order derivative of the double-sided Laplace transform of $\varphi(t)$. This proves that a kernel that reproduces exponential polynomials satisfies the generalised Strang-Fix conditions.

The converse is also true. Consider a kernel $\varphi(t)$ that satisfies the generalised Strang-Fix conditions. Then, a kernel $\psi(t)$ with Fourier transform $\hat{\psi}(\omega) = \hat{\varphi}(\alpha_m + j\omega)$ is guaranteed to satisfy the Strang-Fix conditions and, consequently, reproduces polynomials t^r for $r = 0, \ldots, R$. Finally, due to the relation of the kernels in the Laplace domain it is necessary that $\psi(t) = \mathrm{e}^{-\alpha_m t}\varphi(t)$. This implies that $\varphi(t)$ reproduces exponential polynomials $t^r \mathrm{e}^{\alpha_m t}$ for $l \neq 0$, $r = 0, \ldots, R$ and $m = 0, \ldots, P$, which completes the proof.

A.2 Annihilating other sequences

The signal $s_m = m^r u^m$ is annihilated by a filter with R poles, where $r \leqslant R - 1$ [2]. Consider the filter

$$\hat{h}(z) = (1 - uz^{-1})^R = \sum_{\ell=0}^{R} h[\ell] z^{-\ell}, \tag{A.1}$$

and compute the rth derivative evaluated at $z = u$. We have:

$$R(R-1)\ldots(R-r+1)(1-uz^{-1})^{R-r}|_{z=u} = 0 = \sum_{\ell=0}^{R} -(1)^r \ell(\ell-1)\ldots(\ell-r+1)h[\ell]u^{-\ell+r}, \tag{A.2}$$

which is true for $r = 0, \ldots, R-1$. Therefore, by properly combining weighted versions of (A.2) we obtain

$$\sum_{\ell=0}^{R} h[\ell]p[\ell]u^{-\ell} = 0, \tag{A.3}$$

for any polynomial $p[\ell]$ of degree less than or equal to $R-1$. As a consequence, it is easy to see that the signal s_m is annihilated by h_m, since the following holds:

$$(h \star s)[m] = \sum_{\ell=0}^{K} h[\ell]s[m-\ell] = \sum_{\ell=0}^{K} h[\ell](m-\ell)^r u^{m-\ell} = 0. \tag{A.4}$$

This is because (A.4) is just (A.3) with $R = K$ and $p[\ell] = (m-\ell)^r$. Moreover, the filter $\hat{h}(z) = \prod_{k=0}^{K-1}(1-u_k z^{-1})^{R_k}$ can annihilate $(\alpha_m)^r u_k^m = (\alpha_0 + \lambda m)^r u_k^m$ for $r = 0, \ldots, R_k - 1$. Here, for each k we have that $R = R_k$ and also that $p[\ell] = (\alpha_0 + \lambda(m-\ell))^r$ in (A.3).

Appendix B

Chapter 3

B.1 CRB derivations for power sum series

B.1.1 CRB when AWGN is added to the moments

Consider the set of measurements (3.2) for $K = 1$:

$$\tilde{s}_m = s_m + b_m = a_0 \mathrm{e}^{\alpha_m t_0} + b_m, \qquad m = 0, \ldots, P, \tag{B.1}$$

where b_m are i.i.d. Gaussian random variables of zero mean and variance σ^2. Any unbiased estimate $\Theta(\hat{\mathbf{s}})$ of the unknown parameters (t_0, a_0) has a covariance matrix that is lower bounded by $\sigma^2(\Phi_{\mathbf{s}}^H \Phi_{\mathbf{s}})^{-1}$ (see (3.13)), since $\mathbf{R_b} = \sigma^2 \mathbf{I}$, where $\Phi_{\mathbf{s}}$ is given by

$$\Phi_{\mathbf{s}} = \left(\begin{array}{c|c} a_0\alpha_0 \mathrm{e}^{\alpha_0 t_0} & \mathrm{e}^{\alpha_0 t_0} \\ a_0\alpha_1 \mathrm{e}^{\alpha_1 t_0} & \mathrm{e}^{\alpha_1 t_0} \\ \vdots & \vdots \\ a_0\alpha_P \mathrm{e}^{\alpha_P t_0} & \mathrm{e}^{\alpha_P t_0} \end{array} \right).$$

In order to calculate (3.13) we first derive the simpler case of purely imaginary parameters that appear in complex conjugate pairs, i.e. $\alpha_m = j\omega_m = j\omega_0(2m - P)$ for $m = 0, \ldots, P$. Then, we have

$$\Phi_{\mathbf{s}}^H \Phi_{\mathbf{s}} = \left(\begin{array}{cc} |a_0|^2 \sum_{l=0}^{P} |\alpha_l|^2 |\mathrm{e}^{\alpha_l t_0}|^2 & a_0^* \sum_{l=0}^{P} \alpha_l^* |\mathrm{e}^{\alpha_l t_0}|^2 \\ a_0 \sum_{l=0}^{P} \alpha_l |\mathrm{e}^{\alpha_l t_0}|^2 & \sum_{l=0}^{P} |\mathrm{e}^{\alpha_l t_0}|^2 \end{array} \right) = \left(\begin{array}{cc} |a_0|^2 \sum_{l=0}^{P} |\omega_l|^2 & 0 \\ 0 & P+1 \end{array} \right),$$

because $|\mathrm{e}^{\alpha_l t_0}|^2 = |\mathrm{e}^{j\omega_l t_0}|^2 = 1$ and also $\sum_{l=0}^{P} \alpha_l^* = \sum_{l=0}^{P} \alpha_l = 0$. The uncertainty in the location is given by the square root of the first element of $\sigma^2(\Phi_{\mathbf{s}}^H \Phi_{\mathbf{s}})^{-1}$:

$$\frac{\Delta t_0}{\tau} \geqslant \frac{1}{\tau} \sqrt{\sigma^2 \left(|a_0|^2 \sum_{l=0}^{P} |\omega_l|^2 \right)^{-1}} = \frac{1}{\tau} \sqrt{\frac{1}{\sum_{l=0}^{P} |\omega_l|^2}} \mathrm{PSNR}^{-\frac{1}{2}}. \tag{B.2}$$

With our assumption that $\omega_l = \omega_0(2l - P)$ for $l = 0, \ldots, P$ we know that

$$\sum_{l=0}^{P} |\omega_l|^2 = |\omega_0|^2 \frac{P(P+1)(P+2)}{3} \text{ for either } P \text{ even or odd.} \tag{B.3}$$

We see from (B.2) and (B.3) that the uncertainty in the location decreases linearly with $|\omega_0| = \frac{\pi}{L}$. We may therefore choose the smallest value of L to have the best uncertainty. This is achieved when $L = P+1$ since a smaller value would make exponential reproducing kernel have a non-valid basis [1, 98].

In the more general case of having parameters with a real part α, that is $\alpha_m = \alpha + j\omega_m$ for $m = 0, \ldots, P$, we need to invert $\Phi_{\mathbf{s}}^H \Phi_{\mathbf{s}}$, which now has the form:

$$(\Phi_{\mathbf{s}}^H \Phi_{\mathbf{s}})^{-1} = \frac{1}{|\Phi_{\mathbf{s}}^H \Phi_{\mathbf{s}}|} \begin{pmatrix} \sum_{l=0}^{P} |\mathrm{e}^{\alpha_l t_0}|^2 & -a_0^* \sum_{l=0}^{P} \alpha_l^* |\mathrm{e}^{\alpha_l t_0}|^2 \\ -a_0 \sum_{l=0}^{P} \alpha_l |\mathrm{e}^{\alpha_l t_0}|^2 & |a_0|^2 \sum_{l=0}^{P} |\alpha_l|^2 |\mathrm{e}^{\alpha_l t_0}|^2 \end{pmatrix},$$

where

$$|\Phi_{\mathbf{s}}^H \Phi_{\mathbf{s}}| = |a_0|^2 \sum_{l=0}^{P} |\alpha_l|^2 |\mathrm{e}^{\alpha_l t_0}|^2 \sum_{l=0}^{P} |\mathrm{e}^{\alpha_l t_0}|^2 - a_0^* \sum_{l=0}^{P} \alpha_l^* |\mathrm{e}^{\alpha_l t_0}|^2 a_0 \sum_{l=0}^{P} \alpha_l |\mathrm{e}^{\alpha_l t_0}|^2.$$

Since $\alpha_m = \alpha + j\omega_m$ and we choose $\omega_m = \omega_0(2m - P)$ with $m = 0, \ldots, P$ for the parameters to exist in complex conjugate pairs, we have that

$$(\Phi_{\mathbf{s}}^H \Phi_{\mathbf{s}})^{-1} = \frac{1}{|\Phi_{\mathbf{s}}^H \Phi_{\mathbf{s}}|} \begin{pmatrix} \mathrm{e}^{2\alpha t_0}(P+1) & -a_0^* \mathrm{e}^{2\alpha t_0}(P+1)\alpha \\ -a_0 \mathrm{e}^{2\alpha t_0}(P+1)\alpha & |a_0|^2 \mathrm{e}^{2\alpha t_0} \sum_{l=0}^{P} |\alpha_l|^2 \end{pmatrix},$$

and also

$$|\Phi_{\mathbf{s}}^H \Phi_{\mathbf{s}}| = |a_0|^2 \mathrm{e}^{4\alpha t_0}(P+1) \sum_{l=0}^{P} |\omega_l|^2,$$

where we have used $\sum_l |\alpha_l|^2 = (P+1)\alpha^2 + \sum_l |\omega_l|^2$. In total, then, the uncertainty in the location can be calculated as follows:

$$\frac{\Delta t_0}{\tau} = \frac{1}{\tau} \sqrt{\sigma^2 \left(|a_0|^2 \mathrm{e}^{2\alpha t_0} \sum_{l=0}^{P} |\omega_l|^2 \right)^{-1}} = \frac{1}{\tau} \sqrt{\frac{\mathrm{e}^{2\alpha t_0}}{\sum_{l=0}^{P} |\omega_l|^2}} \mathrm{PSNR}^{-\frac{1}{2}}. \tag{B.4}$$

This may suggest that having $\alpha \neq 0$ could improve the uncertainty (B.4) compared to (B.2). But this is not true, since if we assume that the location t_0 is uniformly distributed in an interval, say, $[0, \tau)$, then the mean of the squared uncertainty (B.4) is:

$$\mathrm{E}\{C\mathrm{e}^{-2\alpha t_0}\} = C \int_0^\tau \frac{1}{\tau} \mathrm{e}^{-2\alpha t_0} \mathrm{d}t_0 = \frac{C}{-2\alpha\tau}(\mathrm{e}^{-2\alpha\tau} - 1),$$

where $C = \frac{1}{\tau^2}\frac{1}{\sum_{l=0}^{P}|\omega_l|^2}\text{PSNR}^{-1}$. This expression is minimised with respect to α for $\alpha = 0$. We therefore conclude that in order to minimise the CRB associated to the power sum series (2.21) when the measurements are contaminated by AWGN, the best exponential parameters are of the form $\alpha_m = j\omega_m = j\frac{\pi}{P+1}(2m-P)$ for $m = 0, \dots, P$.

We conclude by noting that the scenario of exponential parameters with a real part $\alpha \neq 0$ is incompatible with that of having AWGN on the moments for FRI setups in which we add AWGN on the samples (3.1) (see Appendix C.5). We have obtained the proof to show that, even in the case in which both conditions could be satisfied, the uncertainty in the location would be minimised for $\alpha = 0$.

B.1.2 CRB when uncorrelated noise is added to the moments

We now derive a closed form expression for (B.1) when b_m are samples of uncorrelated noise, but with different variance among samples. This is to say that $\mathbf{R_b} = \text{E}\{\mathbf{b}\mathbf{b}^H\} = \sigma^2\mathbf{C}\mathbf{C}^H = \sigma^2\text{diag}(|c_{m,0}|^2)$ for $m = 0, \dots, P$. This is a valid FRI scenario when have N samples (3.1) contaminated by AWGN and we have purely imaginary parameters $\alpha_m = j\omega_m = j\omega_0(2m-P)$ for $m = 0, \dots, P$ with $\omega_0 = \frac{\pi}{N}$ (see Appendix C.5).

Any unbiased estimate $\Theta(\hat{\mathbf{s}})$ of the unknown parameters (t_0, a_0) has a covariance matrix that is lower bounded by $(\Phi_\mathbf{s}^H\mathbf{R}_\mathbf{b}^{-1}\Phi_\mathbf{s})^{-1}$ (see (3.13)), where $\mathbf{R_b} = \sigma^2\text{diag}(|c_{m,0}|^2)$ and the matrix $\Phi_\mathbf{s}$ is given by:

$$\begin{aligned}\Phi_\mathbf{s}^H\mathbf{R}_\mathbf{b}^{-1}\Phi_\mathbf{s} &= \sigma^{-2}\begin{pmatrix} |a_0|^2\sum_{l=0}^{P}|c_{l,0}|^{-2}|\alpha_l|^2|\text{e}^{\alpha_l t_0}|^2 & a_0^*\sum_{l=0}^{P}|c_{l,0}|^{-2}\alpha_l^*|\text{e}^{\alpha_l t_0}|^2 \\ a_0\sum_{l=0}^{P}|c_{l,0}|^{-2}\alpha_l|\text{e}^{\alpha_l t_0}|^2 & \sum_{l=0}^{P}|c_{l,0}|^{-2}|\text{e}^{\alpha_l t_0}|^2 \end{pmatrix} \\ &= \sigma^{-2}\begin{pmatrix} |a_0|^2\sum_{l=0}^{P}|c_{l,0}|^{-2}|\omega_l|^2 & 0 \\ 0 & \sum_{l=0}^{P}|c_{l,0}|^{-2} \end{pmatrix},\end{aligned}$$

because $|\text{e}^{\alpha_l t_0}|^2 = |\text{e}^{j\omega_l t_0}|^2 = 1$ and also $\sum_{l=0}^{P}|c_{l,0}|^{-2}\alpha_l^* = \sum_{l=0}^{P}|c_{l,0}|^{-2}\alpha_l = 0$. The latter is true for exponential parameters that exist in complex conjugate pairs since in such case it follows that $c_{m,0} = c_{P-m,0}$. The uncertainty in the location is given by

$$\begin{aligned}\frac{\Delta t_0}{\tau} &= \frac{1}{\tau}\sqrt{\sigma^2\left(|a_0|^2\sum_{l=0}^{P}|c_{l,0}|^{-2}|\omega_l|^2\right)^{-1}} \\ &= \frac{1}{\tau}\sqrt{\frac{1}{\sum_{l=0}^{P}|c_{l,0}|^{-2}|\omega_l|^2}}\text{PSNR}^{-\frac{1}{2}} \geqslant \frac{1}{\tau}\sqrt{\frac{1}{\sum_{l=0}^{P}|\omega_l|^2}}\text{PSNR}^{-\frac{1}{2}}. \end{aligned} \tag{B.5}$$

For the last inequality we have assumed that $|c_{l,0}|^{-1} \leqslant 1$ for the case of exponential reproducing kernels. This comes from the fact that these kernels satisfy the generalised Strang-Fix conditions (see Appendix A.1) and, as a consequence, it is true that $c_{l,0}^{-1} = \hat{\varphi}(\alpha_l)$

for any l. Then, for Laplace transforms that are normalised to $|\hat{\varphi}(0)| = 1$ it follows that $|c_{l,0}|^{-1} \leqslant 1$. It is a relevant fact that, according to (B.5), the uncertainty in the location will be larger (and hence *worse*) when the measurements are affected by uncorrelated noise than when they are contaminated by white noise.

Consider the more general case of parameters having a real part $\alpha_m = \alpha + j\omega_m$ for $m = 0, \dots, P$. Assume also that the noise is uncorrelated, i.e. $\mathbf{R_b} = \mathrm{E}\{\mathbf{b}\mathbf{b}^H\} = \sigma^2\mathbf{C}\mathbf{C}^H = \sigma^2\mathrm{diag}(|c_{m,0}|^2)$ and that $\omega_m = \omega_0(2m - P)$ for $m = 0, \dots, P$. Then, we have to invert $\Phi_\mathbf{s}^H\mathbf{R}_\mathbf{b}^{-1}\Phi_\mathbf{s}$, which now has the form:

$$(\Phi_\mathbf{s}^H\mathbf{R}_\mathbf{b}^{-1}\Phi_\mathbf{s})^{-1} = \frac{1}{|\Phi_\mathbf{s}^H\mathbf{R}_\mathbf{b}^{-1}\Phi_\mathbf{s}|}\begin{pmatrix} \sum_{l=0}^{P}|c_{l,0}|^{-2}|\mathrm{e}^{\alpha_l t_0}|^2 & -a_0^*\sum_{l=0}^{P}|c_{l,0}|^{-2}\alpha_l^*|\mathrm{e}^{\alpha_l t_0}|^2 \\ -a_0\sum_{l=0}^{P}|c_{l,0}|^{-2}\alpha_l|\mathrm{e}^{\alpha_l t_0}|^2 & |a_0|^2\sum_{l=0}^{P}|c_{l,0}|^{-2}|\alpha_l|^2|\mathrm{e}^{\alpha_l t_0}|^2 \end{pmatrix},$$

where

$$\begin{aligned}|\Phi_\mathbf{s}^H\mathbf{R}_\mathbf{b}^{-1}\Phi_\mathbf{s}| = {} & |a_0|^2\sum_{l=0}^{P}|c_{l,0}|^{-2}|\alpha_l|^2|\mathrm{e}^{\alpha_l t_0}|^2\sum_{l=0}^{P}|c_{l,0}|^{-2}|\mathrm{e}^{\alpha_l t_0}|^2 \\ & - a_0^*\sum_{l=0}^{P}|c_{l,0}|^{-2}\alpha_l^*|\mathrm{e}^{\alpha_l t_0}|^2 a_0\sum_{l=0}^{P}|c_{l,0}|^{-2}\alpha_l|\mathrm{e}^{\alpha_l t_0}|^2\end{aligned}$$

Since $\alpha_m = \alpha + j\omega_m$ and we choose $\omega_m = \omega_0(2m - P)$ with $m = 0, \dots, P$ for the parameters to exist in complex conjugate pairs, we have that

$$(\Phi_\mathbf{s}^H\mathbf{R}_\mathbf{b}^{-1}\Phi_\mathbf{s})^{-1} = \frac{1}{|\Phi_\mathbf{s}^H\mathbf{R}_\mathbf{b}^{-1}\Phi_\mathbf{s}|}\begin{pmatrix} \mathrm{e}^{2\alpha t_0}\sum_{l=0}^{P}|c_{l,0}|^{-2} & -a_0^*\mathrm{e}^{2\alpha t_0}\sum_{l=0}^{P}|c_{l,0}|^{-2}\alpha \\ -a_0\mathrm{e}^{2\alpha t_0}\sum_{l=0}^{P}|c_{l,0}|^{-2}\alpha & |a_0|^2\mathrm{e}^{2\alpha t_0}\sum_{l=0}^{P}|c_{l,0}|^{-2}|\alpha_l|^2 \end{pmatrix},$$

and also

$$|\Phi_\mathbf{s}^H\mathbf{R}_\mathbf{b}^{-1}\Phi_\mathbf{s}| = |a_0|^2\mathrm{e}^{4\alpha t_0}\sum_{l=0}^{P}|c_{l,0}|^{-2}|\omega_l|^2\sum_{l=0}^{P}|c_{l,0}|^{-2},$$

where we have used $|\alpha_l|^2 = \alpha^2 + \omega_l^2$. In total, the uncertainty in the location can be calculated as follows:

$$\frac{\Delta t_0}{\tau} = \frac{1}{\tau}\sqrt{\sigma^2\left(|a_0|^2\mathrm{e}^{2\alpha t_0}\sum_{l=0}^{P}|c_{l,0}|^{-2}|\omega_l|^2\right)^{-1}} = \frac{1}{\tau}\sqrt{\frac{\mathrm{e}^{2\alpha t_0}}{\sum_{l=0}^{P}|c_{l,0}|^{-2}|\omega_l|^2}}\mathrm{PSNR}^{-\frac{1}{2}}.$$

We end by noting that when $\alpha_m = \alpha + j\omega_m$ for $m = 0, \dots, P$ uncorrelated noise is in fact not possible for FRI setups in which white Gaussian noise is added to the samples (3.1) (see Appendix C.5).

Appendix C

Chapter 4

C.1 eMOMS include the Dirichlet and SoS kernels

Let us consider the exponential reproducing kernel $\varphi_0(t) = \varphi\left(t + \frac{P+1}{2}\right)$ of support $P+1$ and centred in zero, with $\varphi(t) = \gamma(t) * \beta_{\vec{\alpha}}(t)$, where $\beta_{\vec{\alpha}}(t)$ is an E-Spline. We restrict our analysis to P being even and we use exponential parameters

$$\alpha_m = j\omega_m = j\frac{\pi}{P+1}(2m - P), \tag{C.1}$$

where $m = 0, \ldots, P$. We next use the $P+1$–periodic extension of $\varphi_0(t)$, that is $\varphi_{P+1}(t) = \sum_{l\in\mathbb{Z}} \varphi_0(t + l(P+1))$, which is equivalent to:

$$\varphi_{P+1}(t) = \frac{1}{P+1} \sum_{k\in\mathbb{Z}} \hat{\varphi}_0\left(j\frac{2\pi k}{P+1}\right) \mathrm{e}^{j\frac{2\pi k}{P+1}t}, \tag{C.2}$$

from the application of Poisson summation formula (1). The case of P being odd can be derived likewise, but by periodising over $2(P+1)$. Also note that the Fourier transform of the shifted kernel $\varphi_0(t)$ is equal to:

$$\hat{\varphi}_0(j\omega) = \gamma(j\omega) \prod_{m=0}^{P} \mathrm{sinc}\left(\frac{\omega - \omega_m}{2}\right). \tag{C.3}$$

The set of equations

$$\hat{\varphi}_0(j\omega_m) = |\hat{\varphi}(j\omega_m)| = |\hat{\gamma}(j\omega_m)\hat{\beta}_{\vec{\alpha}}(j\omega_m)| = \eta_m, \tag{C.4}$$

lead to design exponential reproducing kernels of maximum order and minimum support (eMOMs), different from those of Section 4.2, but that still correspond to a specific subfamily of the generalised exponential reproducing kernels of [41].

In (C.2) the Fourier transform $\hat{\varphi}_0(j\omega)$ is evaluated at $j\omega_k = j\frac{2\pi k}{P+1}$. Taking into account (C.4), we know that $\hat{\varphi}_0(j\omega_k) = \eta_k$ for $k = -\frac{P}{2}, \ldots, \frac{P}{2}$. We also have that $\hat{\varphi}_0(j\omega_k) = 0$ for any other k, because we can find a term in the product (C.3) equal to $\mathrm{sinc}(\ell\pi) = 0$,

$\ell \in \mathbb{Z}$. Therefore, (C.2) can be reduced to:

$$\varphi_{P+1}(t) = \frac{1}{P+1} \sum_{k=-\frac{P}{2}}^{\frac{P}{2}} \eta_k \mathrm{e}^{j\frac{2\pi k}{P+1}t}. \tag{C.5}$$

Note that when the values $\eta_k = 1$ for all k, then (C.5) reduces to one period of the Dirichlet kernel of period $P+1$:

$$\varphi_{P+1}(t) = \frac{1}{P+1} \sum_{k=-\frac{P}{2}}^{\frac{P}{2}} \mathrm{e}^{j\frac{2\pi k}{P+1}t} = \frac{1}{P+1} \frac{\sin(\pi t)}{\sin(\frac{\pi t}{P+1})}.$$

And this is precisely the $P+1$–periodic extension of the eMOMs kernels of Section 4.2.

To end, we now consider one period of (C.5) and denote $t = \frac{x}{T}$, $N = P+1$ and $\tau = NT = (P+1)T$. Then we get the time domain definition of the SoS kernel [15]:

$$g(x) = \mathrm{rect}\left(\frac{x}{\tau}\right) \varphi_{P+1}\left(\frac{x}{T}\right) = \mathrm{rect}\left(\frac{x}{\tau}\right) \frac{1}{N} \sum_{k\in\mathcal{K}} \eta_k \mathrm{e}^{j\frac{2\pi k}{\tau}x}.$$

Here, the number of samples N needs to be odd, since P is even, and the set of indices $\mathcal{K} = \{-\frac{N-1}{2}, \ldots, \frac{N-1}{2}\}$.

C.2 Analysis of the Cramér–Rao bound for eMOMS

Let us consider eMOMS kernels (4.6) that reproduce exponentials of parameters (4.4), where $N = P+1$ and P even. Moreover, assume the kernel satisfies that $c_{m,0}^{-1} = \hat{\varphi}(\omega_m) = \hat{\varphi}_m$ where $\omega_m = \frac{\pi}{N}(2m-P)$. We use the general form of the coefficients $c_{m,0} = |c_{m,0}|\mathrm{e}^{\alpha_m \Delta}$.

We want to find the Cramér–Rao bound associated to the estimation of the innovation parameters (t_0, a_0) of $K = 1$ Dirac, directly from the N noisy samples

$$\tilde{y}_n = a_0 \psi(t_0 - nT) + \epsilon_n, \quad n = 0, \ldots, N-1. \tag{C.6}$$

Here, $\psi(t) = \sum_{\ell\in\mathbf{Z}} \varphi(\frac{t-\ell\tau}{T})$ is the τ-periodic extension of the eMOMS kernel $\varphi(t)$, $\tau = NT$ and ϵ_n are i.i.d. Gaussian random variables, of zero mean and standard deviation σ. In order to evaluate the minimum deviations of the amplitude a_0 and time location t_0 that an unbiased estimator may achieve in the presence of noise, we need to calculate the covariance matrix (3.11) for $K = 1$. As a consequence, we have that $\mathrm{CRB}(\Theta) = (\Phi_{\mathbf{y}}^T \mathbf{R}^{-1} \Phi_{\mathbf{y}})^{-1}$ with $\mathbf{R} = \mathrm{E}\{\epsilon\epsilon^H\} = \sigma^2 \mathbf{I}_N$ and where matrix $\Phi_{\mathbf{y}}$ is as follows:

$$\Phi_{\mathbf{y}} = \begin{pmatrix} a_0\psi'(t_0) & \psi(t_0) \\ a_0\psi'(t_0 - T) & \psi(t_0 - T) \\ \vdots & \vdots \\ a_0\psi'(t_0 - (N-1)T) & \psi(t_0 - (N-1)T) \end{pmatrix}. \tag{C.7}$$

Therefore, the CRB is given by the following square and size 2×2 matrix:

$$\mathrm{CRB}(\Theta) = \sigma^2 \begin{pmatrix} \sum_{n=0}^{N-1} \left(a_0\psi'(t_0 - nT)\right)^2 & \sum_{n=0}^{N-1} a_0\psi'(t_0 - nT)\psi(t_0 - nT) \\ \sum_{n=0}^{N-1} \psi(nT - t_0)a_0\psi'(t_0 - nT) & \sum_{n=0}^{N-1} \left(\psi(t_0 - nT)\right)^2 \end{pmatrix}^{-1}. \quad \text{(C.8)}$$

To compute the summations, it is convenient to use the Fourier series representations of $\psi(t)$ and of its derivative $\psi'(t)$, because the inner product of the sequences $f(nT)$ and $g(nT)$, obtained by sampling functions $f(t)$ and $g(t)$ at $t = nT$, satisfies [28, 99]:

$$\begin{aligned} \sum_{n=0}^{N-1} f(nT)g^*(nT) &\overset{(a)}{=} \sum_{n=0}^{N-1} \left(\sum_k \hat{f}_k \mathrm{e}^{j2\pi kn\frac{T}{\tau}}\right)\left(\sum_{k'} \hat{g}^*_{k'} \mathrm{e}^{-j2\pi k'n\frac{T}{\tau}}\right) \\ &= \sum_k \hat{f}_k \sum_{k'} \hat{g}^*_{k'} \frac{1 - \mathrm{e}^{j2\pi(k-k')N\frac{T}{\tau}}}{1 - \mathrm{e}^{-j2\pi(k-k')\frac{T}{\tau}}} \\ &\overset{(b)}{=} \sum_k \hat{f}_k \sum_{k'} \hat{g}^*_{k'} N\delta_{k,k'} = N\sum_k \hat{f}_k \hat{g}^*_k, \end{aligned} \quad \text{(C.9)}$$

where in (a) we have assumed $f(t)$ and $g(t)$ are τ-periodic and we use their Fourier series expansions, and in (b) we apply $\tau = NT$, hence, the sum is only non-zero when $k = k'$.

Furthermore, if we call $\hat{\psi}_k$ the coefficients for the expansion of $\psi(t)$, then $\hat{\psi'}_k = j2\pi\frac{k}{\tau}\hat{\psi}_k$ are the coefficients for the expansion of its derivative $\psi'(t)$; and $\hat{\psi}^{(t_0)}_k = \mathrm{e}^{-j2\pi k\frac{t_0}{\tau}}\hat{\psi}_k$ the coefficients for the expansion of its shifted version $\psi(t - t_0)$. By using these equivalences and equation (C.9) it is easy to obtain the sums in (C.8). We begin by highlighting that the function $\psi(t)$ is characterized by the Fourier series coefficients $\hat{\psi}_k = \frac{1}{\tau}\int_0^\tau \psi(t)\mathrm{e}^{-j\frac{2\pi k}{\tau}}\mathrm{d}t = \frac{1}{N}\hat{\varphi}\left(\frac{2\pi k}{N}\right) = \frac{1}{N}\hat{\varphi}_k$, for $k \in \mathcal{K} = \{k : k = \frac{2m-P}{2}, m = 0, \ldots, P\}$[1] and $\hat{\psi}_k = 0$ otherwise. Then the first diagonal element in (C.8) before inverting can be obtained as follows:

$$\begin{aligned} \sigma^{-2}\sum_{n=0}^{N-1}\left(a_0\psi'(t_0 - nT)\right)^2 &= \sigma^{-2}a_0^2 N \sum_{k\in\mathcal{K}} \hat{\psi}'_k \mathrm{e}^{-j2\pi k\frac{t_0}{\tau}} \hat{\psi}'^*_k \mathrm{e}^{j2\pi k\frac{t_0}{\tau}} \\ [\mathrm{I}(\Theta)]_{11} &= \frac{1}{N}\left(\frac{a_0}{\sigma}\frac{2\pi}{\tau}\right)^2 \sum_{k\in\mathcal{K}} k^2|\hat{\varphi}_k|^2 \end{aligned}$$

and the second can be derived likewise. In addition, when we compute the elements of the anti-diagonal we find a factor of the form $\sum_{k\in\mathcal{K}} k|\hat{\varphi}_k|^2$, which is equal to zero as long as $|\hat{\varphi}_k| = |\hat{\varphi}_{-k}|$. This is true, for instance, if we want to design real filters, since they satisfy $\hat{\varphi}_k = \hat{\varphi}^*_{-k}$. Thus, in total, we have that

$$\mathrm{CRB}(\Theta) = \begin{pmatrix} N\left(\frac{\sigma}{a_0}\right)^2 \left(\frac{\tau}{2\pi}\right)^2 \frac{1}{\sum_{k\in\mathcal{K}} k^2|\hat{\varphi}_k|^2} & 0 \\ 0 & \sigma^2 N \frac{1}{\sum_{k\in\mathcal{K}} |\hat{\varphi}_k|^2} \end{pmatrix}. \quad \text{(C.10)}$$

We determine the uncertainties in the location and the amplitude from the CRB (C.10). We know that the diagonal values are lower bounds for the variances of t_0 and a_0 respec-

[1] We have assumed P is even, therefore $k = -\frac{P}{2}, \ldots, \frac{P}{2}$ is a valid set of consecutive integers.

tively and, since we are interested in unbiased estimators, the variances equal the MSE of the estimation of each unknown. The uncertainty in the location satisfies that:

$$\frac{\Delta t_0}{\tau} \geqslant \frac{1}{2\pi}\sqrt{\frac{N}{\sum_{k\in\mathcal{K}} k^2|\hat{\varphi}_k|^2}}\mathrm{PSNR}^{-\frac{1}{2}},$$

where we have defined the peak signal-to-noise ratio as $\mathrm{PSNR} = \left(\frac{a_0}{\sigma}\right)^2$, and the uncertainty in the amplitude satisfies:

$$\frac{\Delta a_0}{|a_0|} \geqslant \sqrt{\frac{N}{\sum_{k\in\mathcal{K}} |\hat{\varphi}_k|^2}}\mathrm{PSNR}^{-\frac{1}{2}}.$$

Note that when $|\hat{\varphi}_k| = 1$ for all k, and if we denote $\mathcal{K} = \{k : k = -M, \ldots, M\}$, the above expressions simplify to:

$$\frac{\Delta t_0}{\tau} \geqslant \frac{1}{2\pi}\sqrt{\frac{3N}{M(M+1)(2M+1)}}\mathrm{PSNR}^{-\frac{1}{2}}, \tag{C.11}$$

$$\frac{\Delta a_0}{|a_0|} \geqslant \sqrt{\frac{N}{2M+1}}\mathrm{PSNR}^{-\frac{1}{2}}.$$

The above expressions can easily be shown to be equal to the uncertainties derived in [14] for the periodic sinc.

C.3 Generic CRB for eMOMS

We now show that equation (C.10) is in fact valid for any eMOMS kernel, including E-Splines, by assuming k is no longer restricted to the set $\mathcal{K}$. Consider a kernel of the form

$$\varphi(t) = \sum_{\ell=0}^{P} d_\ell \frac{\mathrm{d}^\ell \beta_{\vec{\alpha}}(t)}{\mathrm{d}t^\ell},$$

such that, given the coefficients $c_{m,0} = |c_{m,0}|\mathrm{e}^{\alpha_m \Delta}$, proper linear combinations of the kernels can reproduce the exponential functions $\mathrm{e}^{\alpha_m t}$, where

$$\alpha_m = \alpha + j\frac{\pi}{L}(2m - P) \qquad m = 0, \ldots, P.$$

The key idea to prove that (C.10) holds is to note that the N noiseless samples

$$y_n = a_0\varphi\left(\frac{t_0}{T} - n\right), \quad n = 0, \ldots, N-1$$

are identical to

$$y_n = a_0\psi(t_0 - nT), \quad n = 0, \ldots, N-1$$

if we use $\psi(t) = \sum_{\ell\in\mathbf{Z}} \varphi(\frac{t-\ell\tau}{T})$, which is the τ-periodic extension of the eMOMS kernel

$\varphi(t)$, and provided that $\tau = NT$ and $N \geqslant P+1$. For these sets of samples to be equal, we need the Dirac to be located in a time position where the kernel is able to reproduce exponentials exactly. For instance, we may have $0 \leqslant t_0 < T$ and take samples with indices $n = -P, \ldots, N-P-1$. Then, the first $P+1$ samples in either case are different from zero and the other $N-P-1$ are zeros.

When these conditions are satisfied, we may proceed exactly in the same way as in Appendix C.2 to conclude that, in general, the following is true:

$$\mathrm{CRB}(\Theta) = \mathrm{I}(\Theta)^{-1} = \begin{pmatrix} N\left(\frac{\sigma}{a_0}\right)^2 \left(\frac{\tau}{2\pi}\right)^2 \frac{1}{\sum_{k\in\mathbb{Z}} k^2 |\hat{\varphi}_k|^2} & 0 \\ 0 & \sigma^2 N \frac{1}{\sum_{k\in\mathbb{Z}} |\hat{\varphi}_k|^2} \end{pmatrix}, \tag{C.12}$$

where

$$\hat{\varphi}_k = \hat{\varphi}\left(\frac{2\pi k}{N}\right) = \sum_{\ell=0}^{P} d_\ell (j\omega)^\ell \prod_{m=0}^{P} \frac{1 - \mathrm{e}^{\alpha_m - j\omega}}{j\omega - \alpha_m}\bigg|_{\omega=\frac{2\pi k}{N}},$$

is the Fourier transform of $\varphi(t)$ at $\omega = \frac{2\pi k}{N}$. To end, the uncertainty in the location satisfies:

$$\frac{\Delta t_0}{\tau} \geqslant \frac{1}{2\pi} \sqrt{\frac{N}{\sum_{k\in\mathbb{Z}} k^2 |\hat{\varphi}_k|^2}} \mathrm{PSNR}^{-\frac{1}{2}}, \tag{C.13}$$

where the peak signal-to-noise ratio is $\mathrm{PSNR} = \left(\frac{a_0}{\sigma}\right)^2$, and the uncertainty in the amplitude satisfies:

$$\frac{\Delta a_0}{|a_0|} \geqslant \sqrt{\frac{N}{\sum_{k\in\mathbb{Z}} |\hat{\varphi}_k|^2}} \mathrm{PSNR}^{-\frac{1}{2}}. \tag{C.14}$$

Even though the finite length summations over n in Equation (C.8) are now infinite length summations over k in (C.12), the latter expressions are independent of t_0 and are easier to compare for different kernels. In practice, due to the absolute value of the Fourier transform $\hat{\varphi}(\omega)$ being low-pass and normally fast decaying, just a few terms of k for either (C.13) or (C.14) provide good estimates of the uncertainties.

C.4 Polynomial with roots spanning the unit circle

Consider the polynomial $x^{P+1} - 1$, with zeros being the $P+1$ roots of unity $x_m = \mathrm{e}^{j\frac{2\pi m}{P+1}} = \mathrm{e}^{\alpha'_m}$. We define $\alpha'_m = j\frac{2\pi m}{P+1} = \alpha_m + \alpha$, where $\alpha_m = j\frac{\pi}{P+1}(2m - P)$ and $\alpha = j\frac{\pi P}{P+1}$.

This polynomial can be written as $x^{P+1} - 1 = \prod_m (x - \mathrm{e}^{\alpha'_m})$. If we then use $x = \mathrm{e}^s$ we have that:

$$\mathrm{e}^{s(P+1)} - 1 = \prod_m (\mathrm{e}^s - \mathrm{e}^{\alpha'_m}) = \mathrm{e}^{s(P+1)} \prod_m (1 - \mathrm{e}^{\alpha'_m - s})$$

and by multiplying with $\mathrm{e}^{-s(P+1)}$ on both sides, we obtain:

$$1 - \mathrm{e}^{-s(P+1)} = \prod_m (1 - \mathrm{e}^{\alpha'_m - s}) = \prod_m (1 - \mathrm{e}^{\alpha_m - (s-\alpha)})$$

which finally yields

$$\prod_{m=0}^{P}(1-\mathrm{e}^{\alpha_m-s}) = 1-\mathrm{e}^{-(P+1)(s+j\frac{\pi P}{P+1})}$$

Then, if we use $s=\alpha_i$ in the previous equation, we have an indeterminate form that we solve using L'Hôpital's rule:

$$\begin{aligned}\prod_{\substack{m=0\\m\neq i}}^{P}(1-\mathrm{e}^{\alpha_m-\alpha_i}) &= \lim_{s\to\alpha_i}\frac{\prod_{m=0}^{P}(1-\mathrm{e}^{\alpha_m-s})}{1-\mathrm{e}^{\alpha_i-s}}\\ &= \lim_{s\to\alpha_i}\frac{1-\mathrm{e}^{-(P+1)s}(-1)^P}{1-\mathrm{e}^{\alpha_i-s}} \overset{(a)}{=} \lim_{s\to\alpha_i}\frac{-(P+1)\mathrm{e}^{-(P+1)s}(-1)^P}{-\mathrm{e}^{\alpha_i-s}} = P+1,\end{aligned}$$

where in (a) we differentiate the numerator and denominator, i.e. we apply L'Hôpital's rule, converting the indeterminate form $\frac{0}{0}$ into a determinate form.

C.5 Types of noise in the moments domain

Consider the noisy samples (3.1) where ϵ_n are i.i.d. Gaussian random variables of zero mean and standard deviation σ. These samples lead to the set of moments (3.2) where b_m are Gaussian random variables of zero mean but not i.i.d. any more.

We begin by calculating the covariance matrix of the noise $\mathbf{R_B} = \mathrm{E}\{\mathbf{B}^H\mathbf{B}\}$, as "seen" by the subspace estimator method, where

$$\mathbf{B} = \begin{pmatrix} b_M & b_{M-1} & \cdots & b_0\\ b_{M+1} & b_M & \cdots & b_1\\ \vdots & \vdots & \ddots & \vdots\\ b_P & b_{P-1} & \cdots & b_{P-M}\end{pmatrix}, \tag{C.15}$$

with $b_m = \sum_n c_{m,n}\epsilon_n$ and $\mathrm{E}\{\epsilon_n\epsilon_{n'}\} = \sigma^2\delta_{n-n'}$. The resulting elements are

$$[\mathbf{R_B}]_{k,l} = \sigma^2\left\langle \mathbf{C}_{[M-k:P-k,:]}, \mathbf{C}_{[M-l:P-l,:]}\right\rangle,$$

where $k,l = 0,\ldots,M$, $\mathbf{C}_{[a:b,:]}$ denotes the $\mathbf{C}$-submatrix composed of rows a to b and all the columns, we use entries from 0 to P and 0 to $N-1$ for the rows and columns of $\mathbf{C}$ respectively, and we define the inner product $\langle\mathbf{X},\mathbf{Y}\rangle$ of two matrices $\mathbf{X}$ and $\mathbf{Y}$ as[2]

$$\langle\mathbf{X},\mathbf{Y}\rangle = \sum_{m,n}\mathbf{X}^*_{m,n}\mathbf{Y}_{m,n} = \mathrm{tr}(\mathbf{X}^H\mathbf{Y}).$$

[2] this is sometimes referred as the Frobenius inner product $\mathbf{X}:\mathbf{Y}$

We may equivalently write that

$$[\mathbf{R_B}]_{k,l} = \sigma^2 \left\langle \mathbf{C}_{[M-k:P-k,:]}, \mathbf{C}_{[M-l:P-l,:]} \right\rangle = \sigma^2 \sum_{m=0}^{P-M} \underbrace{\sum_{n=0}^{N-1} c^*_{m+M-k,n} c_{m+M-l,n}}_{[\mathbf{C}^*\mathbf{C}^T]_{(m+M-k,m+M-l)}},$$

for $k, l = 0, \ldots, M$. This derivation is useful to calculate the whitening matrices of Chapter 3. We are now ready to define the various types of noise that can be present in FRI scenarios when white Gaussian noise is added to the samples y_n.

Noise on the moments is limited to a few specific configurations, which are the following:

1. *Correlated noise.* This is the most general form of noise we may encounter. It is characterised by a covariance matrix $\mathbf{R_b} = \mathrm{E}\{\mathbf{b}\mathbf{b}^H\}$ that is not diagonal. The random variable $\mathbf{b}$ is therefore not proper [77], however the distribution remains Gaussian. Hence, it can be completely characterised by its mean (which is zero), and its covariance matrix, which takes the form:

$$\mathbf{R_b} = \mathrm{E}\{\mathbf{b}\mathbf{b}^H\} = \sigma^2 \mathbf{C}\mathbf{C}^H, \tag{C.16}$$

with terms equal to σ^2 times

$$[\mathbf{C}\mathbf{C}^H]_{a,b} = \sum_n c_{a,0} c^*_{b,0} \mathrm{e}^{j(\omega_a - \omega_b)n}$$

for $a \geqslant b$, in case the exponential parameters are of the form $\alpha_m = j\omega_m$ (or as in (4.2) with $\alpha = 0$), and then $\mathbf{C}\mathbf{C}^H = (\mathbf{C}\mathbf{C}^H)^H$. Note that the diagonal terms ($a = b$) are simply $[\mathbf{C}\mathbf{C}^H]_{a,a} = N|c_{a,0}|^2$. The noise in Toeplitz form has covariance matrix with elements:

$$[\mathbf{R_B}]_{k,l} = \sigma^2 \sum_{m=0}^{P-M} [\mathbf{C}^*\mathbf{C}^T]_{(m+M-k,m+M-l)} = \sum_{m=0}^{P-M} \sum_n c_{a,0} c^*_{b,0} \mathrm{e}^{j(\omega_a - \omega_b)n}, \tag{C.17}$$

for $k, l = 0, \ldots, M$ and where $a = m + M - l$ and $b = m + M - k$.

2. *Uncorrelated noise* This type of noise happens when $\mathbf{R_b} = \mathrm{E}\{\mathbf{b}\mathbf{b}^H\}$ is diagonal. Since we have restricted our analysis to parameters of the form (4.2), we need that $\alpha = 0$ and $L = N$. The random variable $\mathbf{b}$ is now proper [77] and Gaussian, of zero mean, and covariance matrix:

$$\mathbf{R_b} = \mathrm{E}\{\mathbf{b}\mathbf{b}^H\} = \sigma^2 \mathrm{diag}(|c_{m,0}|^2) \qquad m = 0, \ldots, P.$$

The signal in Toeplitz form is affected by noise of covariance matrix with entries:

$$[\mathbf{R_B}]_{k,k} = \sigma^2 \sum_{m=0}^{P-M} |c_{m+M-k,0}|^2,$$

for $k = 0, \ldots, M$ and zero otherwise (it is also diagonal).

3. *Correlated noise of constant diagonal terms.* This type of noise takes place when $|c_{m,0}| = 1$ and $\alpha_m = j\frac{\pi}{L}(2m - P)$ for $m = 0, \ldots, P$ when $L \neq N$. In this case, $\mathbf{b}$ is characterised by a covariance matrix like (C.16) but with diagonal terms all equal to N. The noise in Toeplitz form has a covariance matrix (C.17) but with constant diagonal terms equal to $N(P - M + 1)$.

4. *Circular white Gaussian noise* [77] (which implies the random vector $\mathbf{b}$ is zero-mean and proper). This type of noise occurs when $|c_{m,0}| = 1$ and $\alpha_m = j\frac{\pi}{N}(2m - P)$ for $m = 0, \ldots, P$. In this case, $\mathbf{b}$ is characterised by a covariance matrix:

$$\mathbf{R_b} = \mathrm{E}\{\mathbf{b}\mathbf{b}^H\} = \sigma^2 N \mathbf{I}_{P+1},$$

and the signal in Toeplitz form is affected by noise of covariance matrix:

$$\mathbf{R_B} = \mathrm{E}\{\mathbf{B}^H\mathbf{B}\} = \sigma^2 N(P - M + 1)\mathbf{I}_{M+1}.$$

Appendix D

Chapter 5

D.1 Coefficients for approximate exponential reproduction

In this appendix we derive the optimal coefficients in the least-squares sense and of exact interpolation at integer points in time for the approximate reproduction of exponentials introduced in Section 5.1. Given an arbitrary function $f(t)$, we can determine the coefficients c_n for the linear combination $s(t) = \sum_{n\in\mathbb{Z}} c_n \varphi(t-n)$ to approximate $f(t)$ in the least-square sense by computing its orthogonal projection onto the subspace spanned by $\varphi(t-n)$ [81]. Therefore, we know that the error $f(t) - s(t)$ needs to be orthogonal to any function in the approximation subspace. This means that:

$$\begin{aligned} \langle f(t) - s(t), \varphi(t-k)\rangle &= 0 \leftrightarrow \\ \langle f(t), \varphi(t-k)\rangle &= \langle s(t), \varphi(t-k)\rangle \leftrightarrow \\ d_k &= \sum_{l\in\mathbb{Z}} c_l \langle \varphi(t-l), \varphi(t-k)\rangle \leftrightarrow \\ d_k &= a_\varphi[k] \star c_k \end{aligned}$$

where we have defined the sequence $d_k = \langle f(t), \varphi(t-k)\rangle$. Also $a_\varphi[k] = \langle \varphi(t-l), \varphi(t-k)\rangle$ represents the sampled autocorrelation of $\varphi(t)$ and the operation $(\star)$ denotes the discrete convolution of two sequences. Therefore, we can calculate the optimal approximation coefficients in the least-squares sense by simply using a filtering operation:

$$c_k = d_k \star a_\varphi^{-1}[k] \tag{D.1}$$

where $a_\varphi^{-1}[k]$ indicates that the filter should be inverted in the z-transform domain. Moreover, when we consider the approximation of $f(t) = \mathrm{e}^{\alpha t}$ we can find a closed form expression for (D.1). We have, from the definition of d_k and the function $f(t)$ to approximate, that the following holds:

$$d_k = \langle f(t), \varphi(t-k)\rangle = \int_{-\infty}^{\infty} \mathrm{e}^{\alpha t}\varphi(t-k)\mathrm{d}t = \mathrm{e}^{\alpha k}\hat{\varphi}(-\alpha)$$

where, $\hat{\varphi}(s)$ indicates the Laplace transform of $\varphi(t)$. Now, calling $q_k = a_\varphi^{-1}[k]$, we conclude that the least-squares coefficients are:

$$\begin{aligned} c_k &= d_k \star q_k = \sum_{l\in\mathbb{Z}} d_{k-l} q_l \\ &= \mathrm{e}^{\alpha k}\hat{\varphi}(-\alpha)\sum_{l\in\mathbb{Z}} \mathrm{e}^{-\alpha l} q_l = \mathrm{e}^{\alpha k}\hat{\varphi}(-\alpha) Q(\mathrm{e}^{\alpha}) \\ &= \frac{\mathrm{e}^{\alpha k}\hat{\varphi}(-\alpha)}{\hat{a}_\varphi(\mathrm{e}^{\alpha})} \end{aligned}$$

where we have used the fact that $Q(z) = \sum_{l\in\mathbb{Z}} q_l z^{-l}$, and where $\hat{a}_\varphi(\mathrm{e}^{\alpha}) = \sum_{l\in\mathbb{Z}} a_\varphi[l]\mathrm{e}^{-\alpha l}$.

It is also possible to find the coefficients c_k for $s(t)$ to interpolate $f(t)$ at integer values of time exactly. The solution is given by $c_k = f_k \star q_k$, where the filter q_k is defined through its z-transform [1, 8]:

$$Q(z) = \frac{1}{\sum_{k\in\mathbb{Z}} \varphi(k) z^{-k}}$$

Consider again that the input signal is an exponential $f(t) = \mathrm{e}^{\alpha t}$. In this situation, it is particularly simple to obtain the coefficients c_k so that $s(t)$ interpolates $f_k = f(t)|_{t=k}$ as follows:

$$\begin{aligned} c_k &= \sum_{l\in\mathbb{Z}} q_l f_{k-l} = \sum_{l\in\mathbb{Z}} q_l \mathrm{e}^{\alpha(k-l)} \\ &= \mathrm{e}^{\alpha k}\sum_{l\in\mathbb{Z}} q_l \mathrm{e}^{-\alpha l} = \mathrm{e}^{\alpha k} Q(\mathrm{e}^{\alpha l}) \\ &= \frac{\mathrm{e}^{\alpha k}}{\sum_{l\in\mathbb{Z}} \mathrm{e}^{-\alpha l}\varphi(l)} \end{aligned}$$

where we have used the fact that $Q(z) = \sum_{l\in\mathbb{Z}} q_l z^{-l}$.

D.2 Approximation of exponentials with other FRI kernels

D.2.1 Case study 1: B-Spline kernels

In this case study we find the various types of coefficients c_n for the following relation to hold:

$$\sum_{n\in\mathbb{Z}} c_n \beta_{M+1}(t-n) \approx \mathrm{e}^{\alpha t},$$

where $\beta_{M+1}(t)$ is the B-Spline of order $M+1$. To begin, we define the B-Spline of order $M+1$ as the convolution of $M+1$ box functions[1] $\beta_1(t)$ [100] characterised by a Fourier

[1] In fact, most of the literature uses the definition of order M for the convolution of $M+1$ box functions $\beta_1(t)$. However we use order $M+1$ to be consistent with the definition given for E-Splines.

transform $\hat{\beta}_1(\omega) = \frac{1-\mathrm{e}^{j\omega}}{j\omega}$, that is:

$$\varphi(t) = \beta_{M+1}(t) = \underbrace{(\beta_1 * \beta_1 \cdots * \beta_1)}_{M+1 \text{ times}}(t). \tag{D.2}$$

The B-Spline functions are of support $M+1$ since the box function is of support one. The double-sided Laplace transform of the B-Spline of order $M+1$ is:

$$\hat{\varphi}(s) = \hat{\beta}_{M+1}(s) = \prod_{m=0}^{M} \frac{1-\mathrm{e}^{-s}}{s}. \tag{D.3}$$

Hence, the constant least-squares and the interpolation coefficients are straightforward to compute by evaluating the expressions given in Table 5.1 with the definitions (D.2) and (D.3). In practice, when we obtain the interpolation coefficients we do not need to evaluate $l \in \mathbb{Z}$, but only $l \in [0, M+1]$ due to the support of the kernel.

On the other hand, the least-squares coefficients involve evaluating the z-transform of the sampled autocorrelation $a_\varphi[l] = \langle \varphi(t-l), \varphi(t) \rangle$ at $z = \mathrm{e}^{\alpha}$. A nice feature of B-Splines is that the convolution of two B-Splines of orders M_1 and M_2 is another B-Spline of order $M_1 + M_2$ (this follows immediately from the fact that a B-Spline of order $M+1$ is defined as the convolution of $M+1$ box functions). In other words, we may find the autocorrelation by noting that:

$$\begin{aligned} a_\varphi(x) &= \int_{-\infty}^{\infty} \varphi(t-x)\varphi(x)\mathrm{d}t = \varphi(x) * \varphi(-x) \\ &= \beta_{2(M+1)}(x - (M+1)), \end{aligned} \tag{D.4}$$

which is a B-Spline of double order centred in zero. The sampled autocorrelation follows from evaluating (D.4) at $x = l \in \mathbb{Z}$. Then, the least-squares coefficients are obtained from the expression of Table 5.1 by calculating the z-transform of the sampled autocorrelation at $z = \mathrm{e}^{\alpha}$.

We have shown an example of the above analysis in Figure 5.1 of Section 5.1. We have used a linear combinations of shifted versions of a linear spline to reproduce exponentials using the constant least-squares and the interpolation coefficients.

D.2.2 Case study 2: Approximation with Gaussian kernels

We now turn our attention to the approximation capabilities of Gaussian functions, by examining how linear combinations of shifted versions of these kernels reproduce exponential functions. That is, we find the various types of coefficients c_n for the following relation to hold:

$$\sum_{n \in \mathbb{Z}} c_n h_\gamma(t-n) \approx \mathrm{e}^{\alpha t},$$

where $h_\gamma(t)$ is the Gaussian kernel of variance γ^2 centred in zero. We use the time domain definition of the kernel as given in [2, 60], specifically

$$\varphi(t) = h_\gamma(t) = \mathrm{e}^{-\frac{t^2}{2\gamma^2}}. \tag{D.5}$$

Gaussian functions are of infinite length, but they can be characterised by an "effective" support due to their exponential decay. In fact, they are almost negligible beyond 3.5 times their standard deviation [97] on both sides. The double-sided Laplace transform of the Gaussian kernel (D.5) is:

$$\hat{\varphi}(s) = \hat{h}_\gamma(s) = \sqrt{2\pi} \cdot \gamma \cdot \mathrm{e}^{\frac{\gamma^2 s^2}{2}}. \tag{D.6}$$

Therefore, the constant least-squares and interpolation coefficients are obtained simply by evaluating the expressions given in Table 5.2 with the time domain characterisation (D.5) and the Laplace transform (D.6).

The least-squares coefficients are more difficult to derive, since we need to evaluate the z-transform of the sampled autocorrelation $a_\varphi[l] = \langle \varphi(t-l), \varphi(t) \rangle$ at $z = \mathrm{e}^\alpha$. The convolution of two Gaussian functions is another Gaussian with mean the sum of the original means and variance the sum of the original variances[2]. Then it is possible to find a closed form expression for the autocorrelation:

$$\begin{aligned} a_\varphi(x) &= \int_{-\infty}^{\infty} \varphi(t-x)\varphi(x)\mathrm{d}t = h_\gamma(x) * h_\gamma(-x) \\ &= (h_\gamma * h_\gamma)(x) = \mathrm{e}^{-\frac{x^2}{4\gamma^2}}. \end{aligned} \tag{D.7}$$

Hence, the autocorrelation is a Gaussian function of variance four times γ^2. Finally, the least-squares coefficients are obtained from the expression of Table 5.2 and calculating the z-transform of the autocorrelation (D.7) sampled at $x = l \in \mathbb{Z}$.

We show an example of the above analysis in Figure D.1. We approximate exponentials $\mathrm{e}^{\alpha_m} = \mathrm{e}^{j\frac{\pi}{16}(2m-7)t}$ for $m = 0, \ldots, 3$ using linear combinations of a Gaussian function (D.5) with standard deviation $\gamma = 0.63$. The interval of approximation for the Gaussian function depends on the "effective" support of the kernel (we use 7γ) and the number of samples. If, for instance, we define an approximation interval $[0, \tau)$, with $\tau \in \mathbb{Z}$, then we should employ indices $-\lceil S \rceil \leqslant n \leqslant \tau + \lceil S \rceil - 1$, where $S = 3.5\gamma$. We show the approximation of the real part of the exponentials obtained by using the constant least-squares coefficients

$$\left(\sqrt{2\pi} \cdot \gamma \cdot \mathrm{e}^{\frac{\gamma^2 \alpha_m^2}{2}}\right)^{-1} \mathrm{e}^{\alpha_m n}, \qquad m = 0, \ldots, 3$$

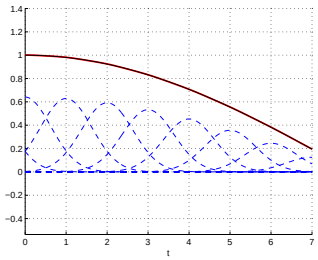
where $\alpha_m = j\frac{\pi}{16}(2m-7)$, in Figure D.1 (c, d, e, i). In addition, we show the interpolation

[2] This can be easily proved in the frequency domain.

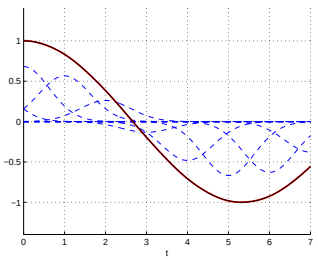
of the real part of the exponentials obtained by using the coefficients

$$c_{m,n} = \left(\sum_{\ell=-S}^{S} \mathrm{e}^{\alpha_m \ell} \mathrm{e}^{-\frac{\ell^2}{2\gamma^2}} \right)^{-1} \mathrm{e}^{\alpha_m n}, \qquad m = 0, \ldots, 3$$

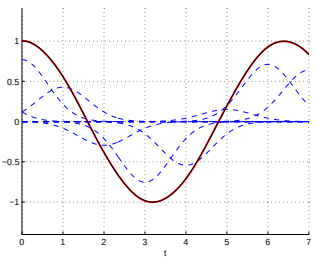
in Figure D.1 (f, g, h, j). In this example the results obtained by either set of coefficients appear identical even though the constant-least squares coefficients are again superior. We notice that the quality of the reproduction is better than that obtained in the example of Figure 5.1, however the effective support of the Gaussian kernel is much bigger than the support of the linear spline. Higher order B-Splines provide similar accuracy to the Gaussian kernel. The number of exponentials that can be approximated is also arbitrary but the quality of the approximation depends on γ.



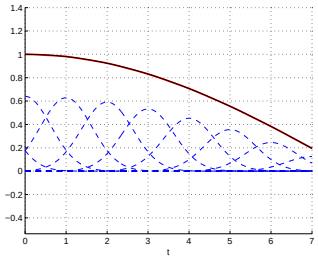
(a) Approximation of $\mathrm{Re}\{\mathrm{e}^{-j\frac{\pi}{16}t}\}$



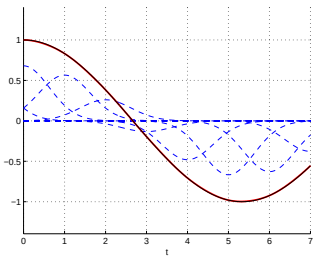
(b) Approximation of $\mathrm{Re}\{\mathrm{e}^{-j\frac{3\pi}{16}t}\}$



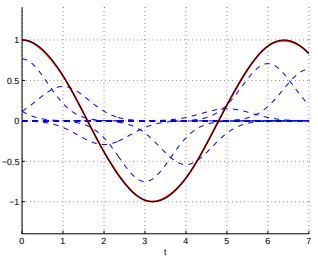
(c) Approximation of $\mathrm{Re}\{\mathrm{e}^{-j\frac{5\pi}{16}t}\}$



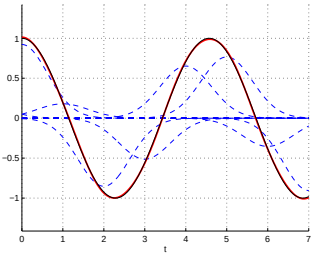
(d) Interpolation of $\mathrm{Re}\{\mathrm{e}^{-j\frac{\pi}{16}t}\}$



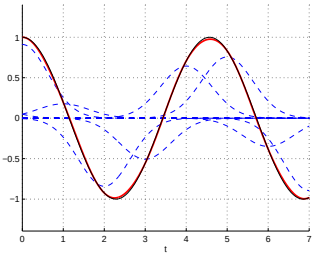
(e) Interpolation of $\mathrm{Re}\{\mathrm{e}^{-j\frac{3\pi}{16}t}\}$



(f) Interpolation of $\mathrm{Re}\{\mathrm{e}^{-j\frac{5\pi}{16}t}\}$



(g) Approximation of $\mathrm{Re}\{\mathrm{e}^{-j\frac{7\pi}{16}t}\}$



(h) Interpolation of $\mathrm{Re}\{\mathrm{e}^{-j\frac{7\pi}{16}t}\}$

Figure D.1: ***Gaussian kernel approximation capabilities.*** **Figures (a-h) show the constant least-squares approximation and the interpolation of the real parts of 4 complex exponentials:** $\mathrm{e}^{j\frac{\pi}{16}(2m-7)t}$ **for** $m = 0, \ldots, 3$**. We plot the weighted and shifted versions of the splines with dashed blue lines, the reconstructed polynomials and exponentials with red solid lines, and the exact functions to be reproduced with solid black lines.**

D.3 Fixed point iteration for the reconstruction of one Dirac

In this section we study whether the iterative method of Algorithm 5 may converge after sufficient iterations for the simple case of $K = 1$ and $T = 1$. Without loss of generality we further assume $a_0 = 1$. For this situation we can explicitly write the non-linear relation between the values s_m and the unknown t_0. The annihilating equation (2.22) for $K = 1$ is simply

$$s_1 h_0 + s_0 h_1 = 0,$$

with the filter

$$\hat{h}(z) = h_0 + h_1 z^{-1} = 1 - u_0 z^{-1} = 1 - \mathrm{e}^{\lambda t_0} z^{-1}.$$

Therefore, the filter coefficients are $h_0 = 1$ and $h_1 = -\mathrm{e}^{\lambda t_0}$. Consequently, the annihilating equation becomes $s_1 - s_0 \mathrm{e}^{\lambda t_0} = 0$, and from this expression it follows that

$$t_0^i = \frac{1}{\lambda} \ln \frac{s_1^i}{s_0^i} = \frac{1}{\lambda} \ln \frac{s_1^0 + \varepsilon_1(t_0^{i-1})}{s_0^0 + \varepsilon_0(t_0^{i-1})} = f(t_0^{i-1}), \tag{D.8}$$

which is in the form of a fixed point iteration $t_0^i = f(t_0^{i-1})$ [101]. The value t_0 is in fact a fixed point, since it satisfies that $t_0 = f(t_0)$ [101]. Note that the error ζ_m of (5.7) reduces to a single term $\varepsilon_m(t_0)$ in (D.8) when $K = 1$ and $a_0 = 1$. In addition, $\varepsilon_m(t)$ can be calculated using (5.4), which states:

$$\varepsilon_m(t) = \mathrm{e}^{\alpha_m t} \left[1 - c_{m,0} \sum_{l \in \mathbb{Z}} \hat{\varphi}(\alpha_m + j2\pi l) \mathrm{e}^{j2\pi l t} \right],$$

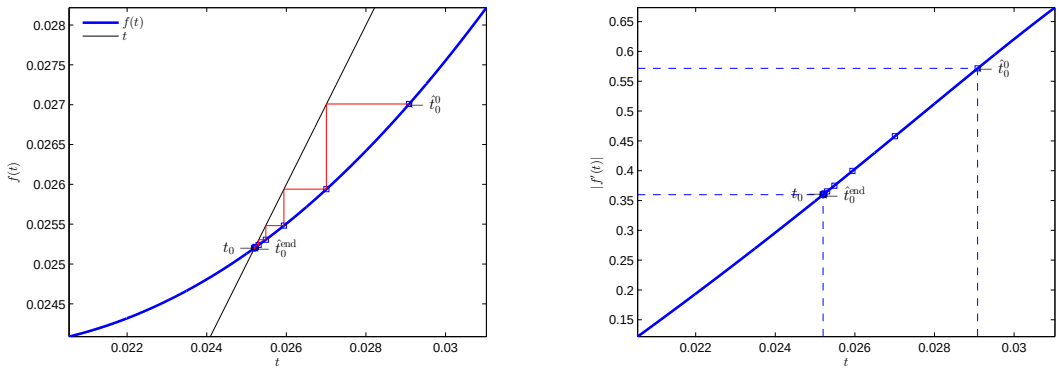
that is, the error depends on $c_{m,0}$, α_m and $\hat{\varphi}(s)$.

We now establish a condition for the iteration of equation (D.8) to converge. Consider an interval $[t_a, t_b]$ such that $f([t_a, t_b]) \in [t_a, t_b]$ and also where $f(t)$ is continuous. Then, we are guaranteed there must exist a fixed point $t_0 = f(t_0)$ in that interval [101]. Moreover, suppose that $f'(t)$ is defined over $[t_a, t_b]$ and there exists a constant $C < 1$ such that $|f'(t)| \leqslant C$ for any $t \in [t_a, t_b]$. Then, the fixed point t_0 is unique in the interval $[t_a, t_b]$ and the sequence $\{t_0^i\}_{i=0}^{\infty}$ defined as $t_0^{i+1} = f(t_0^i)$ will converge to the fixed point t_0 [101]. Even though the aforementioned result is general, evaluating $|f'(t)|$ with $f(t)$ as in (D.8) has to be done in a case by case basis. That $|f'(t)| \leqslant C < 1$ depends on the kernel $\varphi(t)$, the parameters α_m and the coefficients $c_{m,0}$.

We demonstrate the above analysis with an example. Assume we use (D.8) to refine the estimation of the location t_0 of a Dirac that has been sampled by a B-Spline kernel of order $M + 1 = 4$. We use the constant least-squares coefficients $c_{m,n} = \hat{\varphi}(\alpha_m)^{-1} \mathrm{e}^{\alpha_m n}$ for the B-Spline to approximate exponentials of parameters $\alpha_m = j \frac{\pi}{1.4(P+1)}(2m - P)$ for $m = 0, \ldots, P$ with $P + 1 = 16$. In this scenario, and for a Dirac with amplitude $a_0 = 1$ and location $t_0 = \frac{50}{1984}$ the process converges in 20 iterations to a solution exact to numerical precision (MSE $\propto 10^{-30}$). Other configurations converge much faster, but this example

shows the convergence process very clearly, as illustrated in Figure D.2.

In Figure D.2 we show (a) $f(t)$ and (b) $|f'(t)|$ in an interval around the location to be estimated t_0. The derivative is always smaller than 1, which explains convergence. In addition, the rate of convergence is proportional to $|f'(t_0)|$ [101]. In either figure we indicate $\hat{t}_0^0$ for the first estimation and $\hat{t}_0^{\text{end}}$ for the last, which coincides with t_0.



(a) Function around t_0

(b) Derivative of the function

Figure D.2: ***Representation of the fixed point iteration equation*** (D.8)**. We show (a)** $f(t)$ **and (b)** $|f'(t)|$ **in an interval around the location to be estimated** t_0**. The derivative is always smaller than 1, which explains convergence.**

We conclude by saying that whenever the input has more than one Diracs, then there is no simple fixed-point equation analogous to (D.8) to determine the locations. Only when we can write (2.23) as an homogeneous system of K equations and K time values $\mathbf{G}(t_0, t_1, \ldots, t_{K-1}) = \mathbf{G}(\mathbf{t}) = \mathbf{0}$ and transform it into an equivalent system of the form $\mathbf{t} = \mathbf{F}(\mathbf{t})$, then we have the fixed-point iteration of several variables

$$\mathbf{t}^i = \mathbf{F}(\mathbf{t}^{i-1}).$$

The existence and uniqueness of fixed points of vector-valued functions of several variables can be described in a similar manner to the single-variable case [102]. The function $\mathbf{F}$ has a fixed point in a domain $D \in \mathbb{R}^K$ if $\mathbf{F}$ maps D into D. Furthermore, if there exists a constant $C < 1$ such that, in some natural matrix norm, $\|J_{\mathbf{F}}\| < C$ for $\mathbf{t} \in D$, where $J_{\mathbf{F}}$ is the Jacobian matrix of first partial derivatives of $\mathbf{F}$ evaluated at $\mathbf{t}$, then $\mathbf{F}$ has a unique fixed point $\mathbf{t}$ in D. In addition the fixed-point iteration is guaranteed to converge to $\mathbf{t}$ for any initial guess in D.

D.4 Exact FRI recovery schemes for other kernels

D.4.1 Polynomial reproducing kernels

We now briefly summarise the exact reconstruction of a train of K Diracs that has been sampled by a polynomial reproducing kernel $\varphi(t)$ in the absence of noise. We follow the

same steps as in [10]. Consider $x(t) = \sum_{k=0}^{K-1} a_k \delta(t - t_k)$ and the samples

$$y_n = \left\langle x(t), \varphi\left(\frac{t}{T} - n\right)\right\rangle = \sum_{k=0}^{K-1} a_k \varphi\left(\frac{t_k}{T} - n\right),$$

taken by a kernel $\varphi(t)$ that is able to reproduce polynomials of maximum degree $M \geqslant 2K - 1$. That is

$$\sum_{n\in\mathbb{Z}} c_{m,n}\varphi\left(\frac{t}{T} - n\right) = \left(\frac{t}{T}\right)^m, \qquad m = 0, \ldots, M. \tag{D.9}$$

In order to obtain the locations and amplitudes of the Diracs $\{a_k, t_k\}_{k=0}^{K-1}$ from y_n, we begin by computing the first $M + 1$ moments of the input (see [10] for the exact derivations):

$$\tau_m = \sum_n c_{m,n} y_n = \sum_{k=0}^{K-1} a_k \left(\frac{t_k}{T}\right)^m, \qquad m = 0, \ldots, M. \tag{D.10}$$

Then we define the filter $\hat{h}(z) = \sum_{k=0}^{K-1} h_m z^m = \prod_{k=0}^{K-1}(1 - \frac{t_k}{T} z^{-1})$ that is able to annihilate the sequence τ_m. In other words:

$$h_m \star \tau_m = \sum_{i=0}^{K} h_i \tau_{m-i} = 0. \tag{D.11}$$

The zeros of the filter are unique provided the locations t_k are different from each other. Finally, we may write the annihilating identity (D.11) in matrix form which leads to a Yule-Walker system of equations like (2.24). This system can be solved whenever $M \geqslant 2K - 1$. The weights a_k may be retrieved from the moments expression (D.10) which in matrix form is a Vandermonde system similar to (2.25).

Just like with exponential reproducing kernels, when $M + 1$ is strictly larger than the minimum number of moments $2K$ we may solve the problem using least-squares. Moreover, in the presence of white Gaussian noise added to the samples, the sequence of moments is contaminated by colored noise. Hence, it is possible to enhance the accuracy of the recovery by employing the denoising methods described in Chapter 3. The recovery may be further enhanced by improving the conditioning or the problem as suggested in [51]. Whenever $T \ll 1$ then the Yule-Wlaker and Vandermonde systems of equations are badly conditioned. If, on the other hand, we transform (D.9) into

$$\sum_{n\in\mathbb{Z}} \underbrace{(c_{m,n}T^m)}_{c'_{m,n}} \varphi\left(\frac{t}{T} - n\right) = t^m, \qquad m = 0, \ldots, M,$$

then (D.10) becomes

$$\tau_m = \sum_n c'_{m,n} y_n = \sum_{k=0}^{K-1} a_k t_k^m, \qquad m = 0, \ldots, M,$$

yielding much more stable systems of equations. To end, note that the coefficients $c_{m,n}$ are not straightforward to calculate. In the next subsection we explain a way to dealing with the problem.

D.4.2 Coefficients for the polynomial reproduction property

This derivation was proposed to our group by Jon Oñativia Bravo. I include it in the appendices with his consent.

The coefficients for the polynomial reproducing formula:

$$\sum_{n\in\mathbb{Z}} c_{m,n}\varphi(t-n) = t^m, \qquad m = 0,\dots,P, \tag{D.12}$$

to hold may be obtained using the dual of the polynomial reproducing kernel, $\tilde{\varphi}(t)$, as follows:

$$\begin{aligned} c_{m,n} &= \int_{-\infty}^{\infty} t^m \varphi(t-n)\mathrm{d}t = \int_{-\infty}^{\infty} (t+n)^m \varphi(t)\mathrm{d}t = \\ &= \sum_{k=0}^{m} \binom{m}{k} n^{m-k} \int_{-\infty}^{\infty} t^k \varphi(t)\mathrm{d}t = \sum_{k=0}^{m} \binom{m}{k} n^{m-k} c_{k,0}. \end{aligned} \tag{D.13}$$

Then, in order to obtain the values $c_{k,0}$ for $k = 0,\dots,P$ we substitute (D.13) into (D.12) to obtain:

$$\begin{aligned} t^m &= \sum_{n\in\mathbb{Z}} \sum_{k=0}^{m} \binom{m}{k} n^{m-k} c_{k,0}\varphi(t-n) \\ &= c_{m,0} \sum_{n\in\mathbb{Z}} \varphi(t-n) + \sum_{k=0}^{m-1} \binom{m}{k} c_{k,0} \sum_{n\in\mathbb{Z}} n^{m-k}\varphi(t-n), \end{aligned}$$

equation from which we may obtain the coefficients recursively by using:

$$c_{m,0} = \frac{t^m - \sum_{k=0}^{m-1} \binom{m}{k} c_{k,0} \sum_{n\in\mathbb{Z}} n^{m-k}\varphi(t-n)}{\sum_{n\in\mathbb{Z}} \varphi(t-n)}. \tag{D.14}$$

For instance, the first coefficients are:

$$c_{0,0} = \frac{1}{\sum_{n\in\mathbb{Z}} \varphi(t-n)},$$

$$c_{1,0} = \frac{t - c_{0,0} \sum_{n\in\mathbb{Z}} n\varphi(t-n)}{\sum_{n\in\mathbb{Z}} \varphi(t-n)},$$

$$c_{2,0} = \frac{t^2 - c_{0,0} \sum_{n\in\mathbb{Z}} n^2\varphi(t-n) - 2c_{1,0} \sum_{n\in\mathbb{Z}} n\varphi(t-n)}{\sum_{n\in\mathbb{Z}} \varphi(t-n)}.$$

D.4.3 Gaussian kernels

We now turn our attention to FRI reconstruction based on Gaussian kernels, as described in [2, 60], for the noiseless scenario. A finite stream of K Diracs $x(t) = \sum_{k=0}^{K-1} a_k \delta(t - t_k)$ sampled with a Gaussian kernel $h_\gamma(t) = \mathrm{e}^{-t^2/(2\gamma^2)}$ produces samples:

$$y_n = \langle x(t), h_\gamma(nT - t) \rangle = \sum_{k=0}^{K-1} a_k \mathrm{e}^{-\frac{(t_k - nT)^2}{2\gamma^2}}, \qquad n = 0, \ldots, N-1.$$

In order to obtain the locations and amplitudes of the Diracs $\{a_k, t_k\}_{k=0}^{K-1}$ from y_n, we only need multiply above expression by $w_n = \mathrm{e}^{(nT)^2/(2\gamma^2)}$ to obtain:

$$u_n = y_n w_n = \sum_{k=0}^{K-1} a_k \mathrm{e}^{-\frac{t_k^2}{2\gamma^2}} \mathrm{e}^{\frac{n t_k T}{\gamma^2}} = \sum_{k=0}^{K-1} x_k z_k^n, \qquad n = 0, \ldots, N-1, \tag{D.15}$$

where $x_k = a_k \mathrm{e}^{-t_k^2/(2\gamma^2)}$ and $z_k = \mathrm{e}^{t_k T/\gamma^2}$. Hence, the sequence u_n is a power sum series and it is possible to retrieve the parameters x_k and z_k by using the annihilating filter.

Here, γ needs to be carefully chosen in order for the exact FRI recovery of [2] to be well conditioned. In [60] the authors show there is an optimum γ related to the minimum spacing between consecutive Diracs t_s. This value is normally just a fraction of t_s. When the number of Diracs K increases, the ratio $\frac{\sigma}{t_s}$ increases as well. The performance in turn worsens since the system becomes badly conditioned. In case the average distance is not known, then γ must be set according to the required resolution [97]. Moreover, if we fix $\tau = NT$ then the time interval for perfect reconstruction is smaller than τ. In fact, due to the effective support $2S = 7\gamma$ of the kernel, the N samples with indices $n = -\lceil S \rceil, \ldots, N - 1 - \lceil S \rceil$ cover the time interval $0 \leqslant t \leqslant (N-1)T - 2S = \tau - T - 2S$.

Whenever N is strictly larger than the minimum number of moments $2K$ then we may solve the problem using least-squares. Moreover, in the presence of white Gaussian noise added to the samples, the new sequence u_n is contaminated by colored noise, the effects of which may be alleviated by using the pre-conditioning procedure explained in [60]. We conclude by highlighting that the FRI scenario based on the Gaussian kernel is very unstable, specially for high values of N and small values of the standard deviation γ and of the sampling period T.

Bibliography

[1] M. Unser and T. Blu, "Cardinal Exponential Splines: Part I—Theory and Filtering Algorithms," *IEEE Transactions on Signal Processing*, vol. 53, pp. 1425–1438, April 2005.

[2] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Transactions on Signal Processing*, vol. 50, pp. 1417–1428, June 2002.

[3] C. E. Shannon, "Communication in the presence of noise," *Proceedings of the IRE*, vol. 37, pp. 10–21, 1949.

[4] C. E. Shannon, "Classic paper: Communication in the presence of noise," *Proceedings of the IEEE*, vol. 86, no. 2, pp. 447–457, 1998.

[5] J. M. Whittaker, "The Fourier theory of the cardinal functions," *Proceedings of the Edinburgh Mathematical Society*, vol. 1, pp. 169–176, 1929.

[6] V. A. Kotelnikov, "On the transmission capacity of "ether" and wire in electrocommunications," *Izd. Red. Upr. Svyazzi RKKA (Moscow)*, 1933.

[7] V. A. Kotelnikov, "Reprint: On the transmission capacity of "ether" and wire in electrocommunications," in *Modern Sampling Theory: Mathe- matics and Applications* (J. J. Benedetto and P. J. S. G. Ferreira, eds.), Boston: Birkhauser, 2000.

[8] M. Unser, "Sampling-50 years after Shannon," *Proceedings of the IEEE*, pp. 569–587, April 2000.

[9] Y. M. Lu and M. N. Do, "A Theory for Sampling Signals From a Union of Subspaces," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2334–2345, 2008.

[10] 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*, vol. 55, pp. 1741–1757, May 2007.

[11] E. J. Candès, J. K. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, pp. 489–509, February 2006.

[12] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, pp. 1289–1306, April 2006.

[13] E. J. Candès and M. B. Waking, "An Introduction To Compressive Sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[14] T. Blu, P. L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, "Sparse Sampling of Signal Innovations," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 31–40, 2008.

[15] R. Tur, Y. C. Eldar, and Z. Friedman, "Innovation Rate Sampling of Pulse Streams with Application to Ultrasound Imaging," *IEEE Transactions on Signal Processing*, vol. 59, pp. 1827–1842, April 2011.

[16] J. Berent, P. L. Dragotti, and T. Blu, "Sampling Piecewise Sinusoidal Signals With Finite Rate of Innovation Methods," *IEEE Transactions on Signal Processing*, vol. 58, pp. 613–625, February 2010.

[17] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 2005.

[18] J. A. Urigüen, P. L. Dragotti, and T. Blu, "On the exponential reproducing kernels for sampling signals with finite rate of innovation," in *Proceedings of the Ninth International Workshop on Sampling Theory and Applications (SampTA'11), Singapore*, May 2011.

[19] J. A. Urigüen, P. L. Dragotti, and T. Blu, "Approximate Strang-Fix: FRI Sampling with Arbitrary Kernels," *Submitted to IEEE Transactions on Signal Processing*, 2013.

[20] K. Gedalyahu and Y. C. Eldar, "Time Delay Estimation from Low Rate Samples: A Union of Subspaces Approach," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3017–3031, 2010.

[21] H. Akhondi Asl, P. L. Dragotti, and L. Baboulaz, "Multichannel Sampling of Signals with Finite Rate of Innovation," *IEEE Signal Processing Letter*, vol. 17, pp. 762–765, August 2010.

[22] K. Gedalyahu, R. Tur, and Y. C. Eldar, "Multichannel Sampling of Pulse Streams at the Rate of Innovation," *IEEE Transactions on Signal Processing*, vol. 59, April 2011.

[23] A. Hormati, O. Roy, Y. M. Lu, and M. Vetterli, "Distributed sampling of signals linked by sparse filtering: Theory and applications," *IEEE Transactions on Signal Processing*, vol. 58, pp. 1095–1109, February 2010.

[24] I. Maravic and M. Vetterli, "Exact Sampling Results for Some Classes of Parametric Non-Bandlimited 2-D Signals," *IEEE Transactions on Signal Processing*, vol. 52, pp. 175–189, January 2004.

[25] P. Shukla and P. L. Dragotti, "Sampling Schemes for Multidimensional Signals with Finite Rate of Innovation," *IEEE Transactions on Signal Processing*, 2006.

[26] L. Baboulaz and P. L. Dragotti, "Distributed acquisition and image super-resolution based on continuous moments from samples," in *Proceedings of IEEE International Conference on Image Processing (ICIP)*, pp. 3309–3312, 2006.

[27] 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*, vol. 18, no. 2, pp. 281–298, 2009.

[28] J. A. Urigüen, Y. C. Eldar, P. L. Dragotti, and Z. Ben-Haim, "4. Sampling at the Rate of Innovation: Theory and Applications," in *Compressed Sensing: Theory and Applications* (Y. C. Eldar and G. Kutyniok, eds.), Englewood Cliffs, NJ: Cambridge University Press, 2012.

[29] I. Maravic, J. Kusuma, and M. Vetterli, "Low-Sampling Rate UWB Channel Characterization and Synchronization," *Journal of Communications and Networks KOR, special issue on ultra-wideband systems*, vol. 5, no. 4, pp. 319–327, 2003.

[30] I. Maravic, M. Vetterli, and K. Ramchandran, "Channel Estimation and Synchronization with Sub-Nyquist Sampling and Application to Ultra-Wideband Systems," in *Proceedings IEEE International Symposium on Circuits and Systems*, vol. 5, pp. 381–384, 2004.

[31] D. L. Donoho and J. Tanner, "Precise Undersampling Theorems," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 913–924, 2010.

[32] J. Romberg, "Compressive sensing by random convolution," *SIAM J. Img. Sci.*, vol. 2, pp. 1098–1128, November 2009.

[33] L. Li, Y. Xiang, and F. Li, "Theoretical analysis of compressive sensing via random filter," *CoRR*, vol. abs/0811.0152, 2008.

[34] J. Tropp, J. Laska, M. Duarte, J. Romberg, and R. Baraniuk, "Beyond Nyquist: Efficient sampling of sparse bandlimited signals," *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 520–544, 2010.

[35] M. Mishali and Y. Eldar, "From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals," *IEEE of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 375–391, 2010.

[36] M. A. Lexa, M. Davies, and J. Thompson, "Reconciling Compressive Sampling Systems for Spectrally Sparse Continuous-Time Signals," *IEEE Transactions on Signal Processing*, vol. 60, pp. 155–171, January 2012.

[37] M. Mishali and Y. C. Eldar, "3. Xampling: compressed sensing of analog signals," in *Compressed Sensing: Theory and Applications* (Y. C. Eldar and G. Kutyniok, eds.), Englewood Cliffs, NJ: Cambridge University Press, 2012.

[38] X. Wei, T. Blu, and P. L. Dragotti, "Finite rate of innovation with non-uniform samples." 2012.

[39] I. Khalidov, T. Blu, and M. Unser, "Generalized L-Spline Wavelet Bases," in *Proceedings of the SPIE Conference on Mathematical Imaging: Wavelet XI*, vol. 5914, (San Diego CA, USA), pp. 59140F–1–59140F–8, July 31-August 3, 2005.

[40] T. Blu and M. Unser, "Approximation Error for Quasi-Interpolators and (Multi-) Wavelet Expansions," *Applied and Computational Harmonic Analysis*, vol. 6, no. 2, pp. 219–251, 1999.

[41] M. Unser, "Cardinal Exponential Splines: Part II—Think Analog, Act Digital," *IEEE Transactions on Signal Processing*, vol. 53, pp. 1439–1449, April 2005.

[42] T. Blu, P. Thevenaz, and M. Unser, "MOMS: maximal-order interpolation of minimal support," *IEEE Transactions on Image Processing*, vol. 10, pp. 1069–1080, July 2001.

[43] G. Strang and G. Fix, "Fourier analysis of the finite element variational method," *Constructive Aspect of Functional Analysis*, pp. 796–830, 1971.

[44] M. Ela, P. Milanfar, and G. H. Golub, "Shape from Moments — An Estimation Theory Perspective," *IEEE Transactions on Signal Processing*, vol. 52, pp. 1814–1829, July 2004.

[45] Q. Cheng and H. Yingbo, "A review of Parametric high-resolution methods," in *High-resolution and robust signal processing* (H. Yingbo, A. Gershman, and Q. Cheng, eds.), Marcel Dekker, 2003.

[46] G. C. F. M. R. de Prony, "Essai expérimental et analytique sur les lois de la dilatabilité des fluides élastiques, et sur celles de la force expansive de la vapeur de leau et de la vapeur de lalkool, á différentes températures.," *Journal de l'Ecole Polytechnique*, vol. 1, pp. 24–76, 1795.

[47] J. A. Cadzow, "Signal Enhancement – A Composite Property Mapping Algorithm," *Acoustics, Speech and IEEE Transactions on Signal Processing*, vol. 36, pp. 49–62, January 1988.

[48] B. De Moor, "Structured total least squares and L_2 approximation problems," *Linear Algebra Applications*, vol. 188-189, pp. 163–205, July–August 1993.

[49] P. Lemmerling, L. Vanhamme, S. V. Huffel, and B. D. Moor, "IQML-like algorithms for solving structured total least squares problems: A unified view," *Signal Processing*, vol. 81, pp. 1935–1945, September 2001.

[50] 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*, vol. 38, pp. 814–824, May 1990.

[51] G. H. Golub, P. Milanfar, and J. Varah, "A stable numerical method for inverting shape from moments," *SIAM Journal on Scientific Computing (SISC)*, vol. 21, pp. 1222–1243, December 1999.

[52] Y. Hua and T. K. Sarkar, "Generalized pencil-of-function method for extracting poles of an EM system from its transient response," *IEEE Transactions on Antennas Propagation*, pp. 229–234, February 1988.

[53] R. Roy, A. Paulraj, and T. Kailath, "ESPRIT: A subspace rotation approach to estimation of parameters of cissoids in noise," *IEEE Transactions on Acoustic, Speech and Signal Processing*, vol. 34, pp. 1340–1342, May 1986.

[54] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustic, Speech and Signal Processing*, vol. 37, pp. 984–995, July 1989.

[55] Y. Li, K. J. R. Liu, and J. Razavilar, "A parameter estimation scheme for damped sinusoidal signals based on low-rank Hankel approximation," *IEEE Transactions on Signal Processing*, vol. 45, pp. 481–486, 1997.

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

[57] B. D. Rao, "Model based processing of signals: a state space approach," *Proceedings of the IEEE*, vol. 80, pp. 283–309, February 1992.

[58] Y. Hua and T. K. Sarkar, "On SVD for estimating generalized eigenvalues of singular matrix pencil in noise," *IEEE Transactions on Signal Processing*, vol. 39, pp. 892–900, April 1991.

[59] G. H. Golub and C. Van Loan, *Matrix Computations.* Baltimore: Johns Hopkins University Press, Third ed., 1996.

[60] 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*, vol. 53, pp. 2788–2805, August 2005.

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

[62] C. R. Rao, "nformation and the accuracy attainable in the estimation of statistical parameters," *Bulletin of the Calcutta Mathematical Society*, vol. 37, pp. 81–89, 1945.

[63] B. De Moor, "The Singular Value Decomposition and Long and Short Spaces of Noisy Matrices," *IEEE Transactions on Signal Processing*, vol. 41, pp. 2826–2838, September 1993.

[64] S. H. Jensen, P. C. Hansen, S. D. Hansen, and J. A. Sørensen, "Reduction of Broad-Band Noise in Speech by Truncated QSVD," *IEEE Transactions on Speech and Audio Processing*, vol. 3, pp. 439–448, November 1995.

[65] D. Kundu and A. Mitra, "Consistent method of estimating superimposed exponential signals," *Scandinavian Journal of Statistics*, vol. 22, pp. 73–83, March 1995.

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

[67] J. W. Gillard, "Cadzow's basic algorithm, alternating projections and singular spectrum analysis," *Statistics and Its Interface*, vol. 3, no. 3, pp. 333–343, 2010.

[68] B. De Moor, "Total least squares for affinely structured matrices and the noisy realization problem," *IEEE Transactions on Signal Processing*, vol. 42, pp. 3104–3113, November 1994.

[69] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1990.

[70] G. W. Stewart, "Perturbation Theory for the Singular Value Decomposition," tech. rep., University of Maryland, September 1990.

[71] V. Y. F. Tan and V. K. Goyal, "Estimating signals with finite rate of innovation from noisy samples: A stochastic algorithm," *IEEE Transactions on Signal Processing*, vol. 56, pp. 5135–5146, October 2008.

[72] A. Erdozain and P. M. Crespo, "A new stochastic algorithm inspired on genetic algorithms to estimate signals with finite rate of innovation from noisy samples," *Signal Processing*, vol. 90, pp. 134–144, January 2010.

[73] Y. C. Eldar and A. V. Oppenheim, "MMSE Whitening and Subspace Whitening," *IEEE Transactions on Signal Processing*, vol. 49, pp. 1846–1851, July 2003.

[74] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice Hall, 1993.

[75] E. Ollila, "On the Cramér-Rao bound for the constrained and unconstrained complex parameters," *Sensor Array and Multichannel Signal Processing Workshop*, pp. 414–418, July 2008.

[76] P. J. Schreier and L. L. Scharf, "Second-order analysis of improper complex random vectors and processes," *IEEE Transactions on Signal Processing*, vol. 51, pp. 714–725, March 2003.

[77] T. Adali, P. Schreier, and L. Scharf, "Complex-Valued Signal Processing: The Proper Way to Deal With Impropriety," *IEEE Transactions on Signal Processing*, vol. 59, November 2011.

[78] L. Berman and A. Feuer, "On perfect conditioning of vandermonde matrices on the unit circle," *Electronic Journal of Linear Algebra*, vol. 16, pp. 157–161, July 2007.

[79] F. S. V. Bazán, "Conditioning of Rectangular Vandermonde Matrices with Nodes in the Unit Disk," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, no. 2, pp. 679–693, 2000.

[80] A. Ron, "Factorization theorems for univariate splines on regular grids," *Israel Journal of Mathematics*, vol. 70, no. 1, pp. 48–68, 1990.

[81] M. Unser, A. Aldroubi, and M. Eden, "Polynomial Spline Signal Approximations: Filter Design and Asymptotic Equivalence with Shannon's Sampling Theorem," *IEEE Transactions on Information Theory*, vol. 38, pp. 95–103, January 1992.

[82] J. Caballero, J. A. Urigüen, P. L. Dragotti, and S. R. Schultz, "Spike Sorting at Sub-Nyquist Rates," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2012.

[83] W. Gerstner and W. M. Kistler, *Spiking Neuron Models*. Cambridge University Press, First ed., 2002.

[84] M. F. Bear, B. Connors, and M. Paradiso, *Neuroscience: Exploring the Brain*. Lippincott Williams & Wilkins, February 2006.

[85] R. Q. Quiroga, "What is the real shape of extracellular spikes?," *Journal of Neuroscience Methods*, vol. 177, no. 1, pp. 194–198, 2009.

[86] D. A. Adamos, E. K. Kosmidis, and G. Theophilidis, "Performance evaluation of PCA-based spike sorting algorithms," *Computer methods and programs in Biomedicine*, vol. 91, no. 3, pp. 232–244, 2008.

[87] J. C. Letelier and P. P. Weber, "Spike sorting based on discrete wavelet transform coefficients," *Journal of neuroscience methods*, vol. 101, no. 2, pp. 93–106, 2000.

[88] E. Hulata, R. Segev, and E. Ben-Jacob, "A method for spike sorting and detection based on wavelet packets and Shannon's mutual information," *Journal of neuroscience methods*, vol. 117, no. 1, pp. 1–12, 2004.

[89] R. Q. Quiroga, Z. Nadasdy, and Y. B. Shaul, "Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering," *Neural Computation*, vol. 16, pp. 1661–1687, 2004.

[90] Z. Charbiwala, V. Karkare, S. Gibson, D. Markovic, and M. B. Srivastava, "Compressive Sensing of Neural APs Using a Learned Union of Supports," *International Workshop on WIBSN*, pp. 53–58, 2011.

[91] S. Lakshminarayan, L. R. Varshney, and J. Kusuma, "Acquisition of Action Potentials with Ultra-Low Sampling Rates," in *Annual International Conference of the IEEE EMBS*, pp. 4213–4216, 2010.

[92] M. Lewicki, "A review of methods for spike sorting: the detection and classification of neural action potentials," *Network: Computation in Neural Systems*, vol. 9, no. 4, pp. R53–R78, 1998.

[93] M. Abeles and M. G. Jr, "Multispike train analysis," *Proceedings of the IEEE*, vol. 65, pp. 762–773, 1997.

[94] E. M. Schmidt, "Computer separation of multi-unit neuro electric data: a review," *Journal of Neuroscience Methods*, vol. 12, pp. 95–111, 1984.

[95] Jon Oñativia and Simon R. Schultz and Pier Luigi Dragotti, "A finite rate of innovation algorithm for fast and accurate spike detection from two-photon calcium imaging," *Journal of Neural Engineering*, vol. 10, July 2013.

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

[97] A. Erdozain and P. M. Crespo, "Reconstruction of aperiodic FRI signals and estimation of the rate of innovation based on the state space method," *Signal Processing*, vol. 91, no. 8, pp. 1709–1718, 2011.

[98] F. J. Homann and P. L. Dragotti, "Robust Sampling of "Almost" Sparse Signals," tech. rep., Imperial College London, 2008.

[99] L. Coulot, M. Vetterli, T. Blu, and P. L. Dragotti, "Sampling Signals with Finite Rate of Innovation in the Presence of Noise," tech. rep., Ecole Polytechnique Federale de Lausanne, Switzerland, 2007.

[100] M. Unser, "Splines: a perfect fit for signal and image processing," *IEEE Signal Processing Magazine*, vol. 16, no. 6, pp. 22–38, 1999.

[101] R. L. Burden and J. D. Faires, "2.2 Fixed-Point Iteration," in *Numerical Analysis (7th ed.)*, Brooks Cole, 2000.

[102] J. Lambers, "Fixed Points for Functions of Several Variables." University Lecture, 2009.