

Imperial College London
Department of Electrical and Electronic Engineering

**Extensions of the Theory of Sampling Signals
with Finite Rate of Innovation, Performance
Analysis and an Application to Single Image
Super-Resolution**

Xiaoyao Wei

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy of Imperial College London, 2016

Declaration of Originality

I declare that the intellectual content of this thesis is the product of my own work under the guidance of my supervisor Prof. 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.

Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Abstract

Sampling is the reduction of a continuous-time signal to a discrete sequence. The classical sampling theorem limits the signals that can be perfectly reconstructed to bandlimited signals. In 2002, the theory of finite rate of innovations (FRI) emerged and broadened classical sampling paradigm to classes of signals with finite number of parameters per unit of time, which includes certain classes of non-bandlimited signals. In this thesis we analyse the performance of the FRI reconstruction algorithm and present extensions of the FRI theory. We also extend the FRI theory for the application of image upsampling.

First, we explain the breakdown phenomenon in FRI reconstruction by subspace swap and work out at which noise level FRI reconstruction algorithm is guaranteed to achieve the optimal performance given by the Cramér-Rao bound. Our prediction of the breakdown PSNR is directly related to the distance between adjacent Diracs, sampling rate and the order of the sampling kernel and its accuracy is verified by simulations.

Next, we propose an algorithm that can estimate the rate of innovation of the input signals and this extends the current FRI framework to a universal one that works with arbitrarily unknown rate of innovation.

Moreover, we improve the current identification scheme of “parametrically sparse” systems, i.e. systems that are fully specified by small number of parameters. Inspired by the denoising technique used for FRI signals, we propose the modified Cadzow denoising algorithm which leads to robust system identification. We also show the possibility of perfectly identifying the input signal and the system simultaneously and we also propose reliable algorithm for simultaneous identification of both in the presence of noise.

Lastly, by noting that lines of images can be modelled as piecewise smooth signals, we propose a novel image upsampling scheme based on our proposed method for reconstructing piecewise smooth signals which fuses the FRI method with the classical linear reconstruction method. We further improve our upsampled image by learning from the errors of our upsampled results at lower resolution levels. The proposed algorithm outperforms the state-of-the-art algorithms.

Acknowledgements

Firstly I would like to thank my supervisor Prof. Pier Luigi Dragotti for all his help, support and encouragement during my PhD. I would not be able to finish the research works without his guidance. I am very thankful for all his wise advices in weekly meetings and the countless times of corrections he has made on my papers/reports. Not only being a wise PhD supervisor, he has also been a very responsible lecturer that every student respects and a charismatic speaker whose speeches always attract and inspire the audience. I have learnt a lot from him and it has been an honour to have him as my supervisor.

I am also very grateful to Prof. Thierry Blu, who is my undergraduate final year project supervisor. He helped me to understand many concepts in sampling and also the FRI theory throughout the final year project. Without him I would definitely be less prepared for the PhD study. More importantly, without his mathematical sharpness and expert advices, some of the important results here would not exist.

In addition, I thank my family and friends who have always been extremely supportive. PhD is a long journey and I have had ups and downs during the journey. Life would have been much more difficult without the support of my parents, my friends in London and Hong Kong, and most importantly my husband, who is always by my side and encourage me in all different ways. I am also very thankful for everything the college and the city has brought me, e.g. kind and helpful colleagues who create an excellent working environment, responsible English lecturer who corrects my writings, and street performers whose performance brought me lots of happiness and comforts.

Last but above all, I would like to thank God for His guidance and abundant provisions all the time. As Proverbs 16:9 says, the mind of man plans his way, but the Lord direct.

Abbreviations

1-D	one dimensional
2-D	two dimensional
A-to-D	analogue to digital
B-spline	polynomial spline
CRB	Cramér-Rao bound
dB	Decibel
eMOMS	exponential spline with maximum order and minimum support
E-spline	exponential spline
FRESH	FRI-based single-image super-resolution algorithm
FRI	finite rate of innovation
MSE	mean square error
PSNR	peak signal-to-noise ratio
SNR	signal-to-noise ratio
SSIM	Structural SIMilarity index
SVD	singular value decomposition
TLS	total least square
UWB	ultra-wideband

Notations

continuous-time signals

$f(t)$	continuous-time signal
$f^{(r)}(t)$	r th derivative of $f(t)$. The zero order derivative is the function itself $f^{(0)}(t) = f(t)$.
$F(j\omega)$	Fourier transform of $f(t)$: $\int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt$
$F(s)$	bilateral Laplace transform of $f(t)$: $\int_{-\infty}^{\infty} f(t) e^{-st} dt$
$f(t) * g(t)$	continuous-time convolution: $\int_{-\infty}^{\infty} f(\tau)g(t - \tau)d\tau = \int_{-\infty}^{\infty} f(t - \tau)g(\tau)d\tau$
$\langle f(t), g(t) \rangle$	inner product: $\int_{-\infty}^{\infty} f(t)g^*(t)dt$, where $g^*(t)$ is the complex conjugate of $g(t)$
$\text{sinc}(t)$	$\text{sinc}(t) = \frac{\sin(\pi t)}{\pi t}$
$\delta(t)$	delta Dirac function

discrete-time signals and vectors

$f[n]$ or f_n	discrete-time signal
$f[n] * g[n]$	discrete-time convolution: $\sum_{m=-\infty}^{\infty} f[m]g[n - m] = \sum_{m=-\infty}^{\infty} f[n - m]g[m]$
$F(z)$	z -transform of $f[n]$: $\sum_{-\infty}^{\infty} f[n]z^{-n}$
\mathbf{A}^T	transpose of matrix \mathbf{A}
\mathbf{A}^{-1}	inverse of matrix \mathbf{A}
\mathbf{A}^\dagger	pseudo-inverse of matrix \mathbf{A}
\mathbf{A}^H	conjugate transpose or Hermitian transpose of matrix \mathbf{A}
$\text{Tr}(\mathbf{A})$	trace of matrix \mathbf{A}
$\ \mathbf{a}\ $	Euclidean norm of vector \mathbf{a} : $\ \mathbf{a}\ = \sqrt{\sum_{i=1}^m a_i ^2}$
$\ \mathbf{A}\ _{\text{Fro}}$	Frobenius norm operator: $\ \mathbf{A}\ _{\text{Fro}} = \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{i,j} ^2} = \sqrt{\text{Tr}(\mathbf{A}\mathbf{A}^H)}$
\mathbf{I}	identity matrix

others

$\mathbb{E}\{X\}$	expected value of a random variable X
-------------------	---

Contents

1	Introduction	27
1.1	Motivation and objectives	27
1.2	Outline of the thesis	28
1.3	Publications	29
2	Overview of recent development in sampling theory	31
2.1	Sampling problem and the classical sampling theory	31
2.2	Sampling and perfect reconstruction of FRI signals	33
2.2.1	Signals with finite rate of innovation	33
2.2.2	Sampling kernels	35
2.2.3	Exact sampling using exponential reproducing kernels	37
2.3	Sampling FRI signals in the presence of noise	42
2.3.1	Design of the sampling kernel	43
2.3.2	Robust reconstruction algorithms	44
2.3.3	Performance of the reconstruction algorithms	46
2.4	Approximate Strang-Fix: Sampling FRI signals with arbitrary sampling kernel	47
3	Guaranteed performance of the FRI reconstruction algorithm	51
3.1	A revisit to FRI theory	52

3.2	Necessary condition for subspace swap event	54
3.3	Simulations	57
3.3.1	$K = 1$	58
3.3.2	$K = 2$	58
3.3.3	$K > 2$	59
3.4	Summary	59
4	Universal sampling of signals with finite rate of innovation	61
4.1	Identification of the rate of innovation	62
4.1.1	Discussions	63
4.2	Robust identification algorithm	67
4.2.1	Simulations	67
4.3	Summary	70
5	Simultaneous estimation of sparse linear systems and sparse signals	71
5.1	Introduction	71
5.2	Overview of FRI theory	73
5.3	Identification of parametrically sparse systems	74
5.3.1	Exact identification of E-spline	76
5.3.2	Identification of linear systems with rational Fourier transform	78
5.3.3	Identification of linear systems with rational Fourier transform in noisy scenario	79
5.4	Simultaneous identification of input signals and systems	81
5.5	Simulations	83
5.5.1	System identification using arbitrary known input signal	83
5.5.2	Simultaneous identification of input signals and systems	86
5.6	Summary	88

6 FRESH – FRI-based single-image super-resolution algorithm	89
6.1 Introduction	89
6.2 Sampling and resolution enhancement of 1-D piecewise smooth signals . . .	92
6.2.1 Sampling of piecewise smooth signals	93
6.2.2 Resolution enhancement of 1-D piecewise smooth signal	95
6.3 Image up-sampling	97
6.3.1 Basic image up-sampling algorithm	98
6.3.2 Exploiting cross scale similarities	99
6.4 Simulation results	103
6.4.1 1-D piecewise smooth signal upsampling	103
6.4.2 Image upsampling	104
6.4.3 Computation complexity and discussions	109
6.5 Summary	110
7 Conclusion and Future Work	111
7.1 Conclusion	111
7.2 Future Work	113
A Find the closest matrix in the sense of Frobenius norm	115
B Simultaneous identification of input signals and linear systems	117
B.1 Exact identification of E-spline and input stream of Diracs	117
B.2 Identification of linear circuits and input stream of Diracs	118
Bibliography	119

List of Tables

4.1	The percentage of correct identification of the number (4) of Diracs from 51 samples in different noise level. 10000 realisations for each noise level. . .	69
5.1	System identification in the absence of noise. $\varphi(t)$ is an E-spline and $x(t)$ is a pulse.	84
5.2	Parameter errors of identification of $\Psi(s) = \frac{1}{s^2+1s+1}$ and $\Psi(s) = \frac{10}{s^2+4s+10}$ under different levels of additive noise on the input samples and output samples.	85
5.3	Simultaneous identification of the input signal and the system in the absence of noise. $x(t)$ is a stream of two Diracs. $\varphi(t)$ is an exponential reproducing kernel.	86
6.1	Recovering the high-resolution image scan-line from its approximation coefficients of different wavelet decomposition with different methods.	104
6.2	Comparisons of upsampling results (factor 4) given by different methods in terms of PSNR. Sampling kernel: bior4.4.	105
6.3	Comparisons of upsampling results (factor 4) given by different methods in terms of PSNR. Sampling kernel: linear spline.	106
6.4	Computation cost for upsampling an image using the basic method and FRESH on a Mac mini with 2.6GHz Intel Core i7 CPU and 16GB RAM. .	110

List of Figures

2.1	Sampling set-up. Here $x(t)$ is the input signal, $h(t)$ is the impulse response of the acquisition device and T is the sampling period. The samples are given by $y_n = \langle x(t), \varphi(t/T - n) \rangle$	31
2.2	Linear reconstruction set-up. The reconstruction of the input signal is given by $\hat{x}(t) = \sum_n y_n \tilde{\varphi}(t/T - n)$. Perfect reconstruction is achieved when $x(t) \in \text{span}\{\tilde{\varphi}(t/T - n)\}_{n \in \mathbb{Z}}$	32
2.3	Examples of FRI signals. The shape of the signals are known so they are only characterized by the amplitude and location information.	35
2.4	Weighted sum of shifted versions of a third order E-spline of (a) can reproduce exponentials at three different frequencies, which are shown in (b)(c)(d).	38
2.5	Examples of exponential reproducing functions. (a) E-spline of order $P + 1 = 31$. (b) eMOMS of order $P + 1 = 31$	44
2.6	Standard deviation (averages over 1000 realizations) of the retrieved locations of a FRI signal (a stream of two Diracs) compared to Cramér-Rao lower bounds (CRB). The estimation algorithm (Cadzow+Prony) achieves CRB up to certain peak signal-to-noise ratio (PSNR). The two Diracs are recovered from 29 samples taken by a kernel which can reproduce 29 exponentials.	48
2.7	Approximate reproduction of $e^{-j\frac{3\pi}{16}t}$ with a linear spline.	49
2.8	Approximate reconstruction of a stream of Diracs from samples taken with a linear spline when SNR=10dB.	49
3.1	Necessary condition for subspace swap in the case of single Dirac for different number of samples N	56

3.2	Necessary condition for subspace swap in the case of 2 Diracs for different number of samples N	57
3.3	Subspace swap condition compared to the empirical breakdown points for $K = 1$ with the settings $N = P + 1$, $T = 1/N$ and $\beta = j2\pi/N$	58
3.4	Subspace swap condition compared to the empirical breakdown points for $K = 2$ with the settings $N = P + 1$, $T = 1/N$ and $\beta = j2\pi/N$	58
3.5	Reconstruction of 17 Diracs with minimum Dirac separation $\Delta t \geq 2/(N+1)$ from 41 samples with PSNR = 0dB.	59
4.1	Universal sampling of a stream of unknown number of Diracs using B-spline kernel of order 5 in the absence of noise. (b)(d) The number of Diracs is identified from second derivative of the error function. (a)(c) All the Diracs are retrieved almost perfectly.	64
4.2	Scatterplot of locations retrieved using standard FRI reconstruction algorithm (Cadzow+Prony) compared to 3 times the standard deviation given by Cramér-Rao lower bounds. (a) SNR=10dB (b) SNR=20dB	65
4.3	Universal sampling of a stream of unknown number of Diracs using B-spline kernel of order 5 in the presence of noise. Two close Diracs in the 8 Diracs are recognised as one Dirac in when SNR=10dB and the others are accurately retrieved.	66
4.4	Universal sampling of a stream of unknown number of Diracs using 5th order B-spline in the presence of noise.	69
5.1	A typical system identification set-up. Adapted from [1].	72
5.2	FRI sampling set-up for system identification. Here $x(t)$ is the input signal, $\psi(t)$ is the unknown system and $\varphi(t)$ is the exponential reproducing kernel.	74
5.3	True Transfer Function: $\frac{10}{s^2+4s+10}$ and the Fourier data obtained by taking 50 samples at sampling rate $f = 4$ Hz through an exponential reproducing kernel.	75
5.4	Modified Cadzow denoising on the spectral data of the system: $\Psi(s) = \frac{10}{s^2+4s+10}$. SNR=10dB.	80
5.5	FRI sampling set-up for simultaneous estimation of the input signal and the linear system. Here $x(t)$ is an FRI signal, $\varphi(t)$ is an exponential reproducing kernel and $\psi(t)$ is a linear system of order P	81

5.6	Schematic diagram of the algorithm that recursively estimate the input signal and the system.	82
5.7	Sampling set-up for system identification. The input signal $x(t)$ is a pulse and the sampling kernel $\varphi(t)$ is an exponential reproducing kernel.	83
5.8	Sallen-Key filter topology. The transfer function of a Sallen-Key filter is $\Psi(s) = \frac{\frac{1}{R_1 C_1 R_2 C_2}}{s^2 + s\left(\frac{1}{R_2 C_1} + \frac{1}{R_1 C_1}\right) + \frac{1}{R_1 C_1 R_2 C_2}}$	83
5.9	FRI sampling set-ups for recursive estimation. $\varphi(t)$ is an exponential reproducing kernel.	86
5.10	Simultaneous estimation of both the input signal and the system in the presence of noise. SNR=20dB. (a) The true locations of the input Diracs: $\{0.9875, 8.4875\}$ and the estimations of the locations of the input Diracs after each iteration. (b) Bode diagrams of the true system $\frac{1}{s^2+s+1}$ and the estimated system $\frac{1.000}{s^2+0.979s+1.003}$	87
5.11	Simultaneous estimation of both the input signal and the system in the presence of noise. SNR=20dB. (a) The true locations of the input Diracs: $\{0.3086, 3.4336\}$ and the estimations of the locations of the input Diracs after each iteration. (b) Bode diagrams of the true system $\frac{10}{s^2+4s+10}$ and the estimated system $\frac{9.956}{s^2+4.272s+10.055}$	87
5.12	Simultaneous estimation of both the input signal and the system in the presence of noise. SNR=20dB. (a) The true locations of the input Diracs: $\{8.9766\}$ and the estimations of the locations of the input Diracs after each iteration. (b) Bode diagrams of the true system $\frac{s+2}{s^3+2s^2+2s+1}$ and the estimated system $\frac{1.01s+1.35}{s^3+1.52s^2+1.58s+0.67}$	88
6.1	Image formation process. The incoming irradiance light field is blurred by the lens and sampled by the image sensor.	90
6.2	Natural images and scan-lines of natural images are approximately 2-D and 1-D piecewise smooth functions respectively.	90
6.3	Upsampling results (factor 4) of <i>woman</i> by different methods. The down-sampling kernel is bior4.4.	92
6.4	Enhancing images to infinite resolution can be interpreted as the problem of sampling and reconstructing piecewise smooth functions.	92

6.5	We model piecewise smooth signals as the sum of a piecewise polynomial signal and a globally smooth signal.	94
6.6	The details we need for resolution enhancement are due only to the piecewise polynomial part. (a) The piecewise smooth signal $x(t) = p(t) + r(t)$, the piecewise polynomial part $p(t)$ and the globally smooth part $r(t)$. (b) Wavelet decomposition of $x(t)$, $p(t)$ and $r(t)$ respectively.	94
6.7	Schematic diagram of our proposed sampling and reconstruction strategy for piecewise smooth signals. The blue solid part: FRI reconstruction of the piecewise polynomial function $p(t)$. The black dashed part: linear reconstruction of the smooth residual $r(t)$	95
6.8	K -level biorthogonal filter bank. Given the approximation coefficients $y_J = \langle x(t), \tilde{\varphi}_{J,n} \rangle$ we are looking for a higher resolution version $y_{J-K} = \langle x(t), \tilde{\varphi}_{J-K,n} \rangle$	96
6.9	The schematic diagram of resolution enhancement of a piecewise smooth signal by factor of 2^K using a biorthogonal filter bank. From the given approximation coefficients $y_J = \langle x(t), \tilde{\varphi}_{J,n} \rangle$ and the detail coefficients d_J, \dots, d_{J-K+1} estimated using FRI, we are able to recover a higher resolution version $y_{J-K} = \langle x(t), \tilde{\varphi}_{J-K,n} \rangle$	97
6.10	Assume we only have access to the low-pass subband y_0 of a 2D wavelet transform applied to the high-resolution image y_{-K} . We want to estimate the high-pass coefficients using FRI in order to recover the high-resolution image.	97
6.11	The block diagram of our proposed upsampling scheme.	98
6.12	Upsampling results of a piecewise smooth image by linear reconstruction and by proposed method.	99
6.13	The schematic diagram of how to correct our FRI upsampled image y_{-1}^{FRI}	101
6.14	Our method is able to accurately recover a piecewise smooth signal from its approximation coefficients.	103
6.15	Upsampling results (factor 4) of <i>Zebra</i> by different methods. The down-sampling kernel is bior4.4.	107
6.16	Upsampling results (factor 4) of <i>Comic</i> by different methods. The down-sampling kernel is linear spline.	108

6.17 Upsampling of images taken with Canon 400D. Our upsampling results are sharper than the bicubic interpolation results. (a)(d) original images. (b)(e) bicubic interpolation. (c)(f) our upsampling results. 109

Chapter 1

Introduction

We live in an analogue world. For example, speech signals are continuous-time signals because they are continuous waves of acoustic pressure. In order to store or process the signals digitally, it is necessary to convert them into digital sequences (i.e. recording only values of the signal at certain points in time/space). This conversion process is called sampling. The sampling process by definition is a process of discarding information. However, there are situations where all the information of classes of continuous-time signals can be perfectly recovered from the discrete sequences. This is interesting because it means that specific classes of continuous-time signals can be represented in a very compact form.

The most celebrated and widely applied sampling theorem, which is often attributed to Shannon [2], Whittaker [3], Kotelnikov [4], to name a few, established a sufficient condition for exact sampling and reconstruction for bandlimited signals. It is based on the interpretation that signals are sums of sinusoids of different frequencies and it requires a sampling rate that is at least twice the highest frequency of the input signal for perfect reconstruction. However, real-world signals usually contain quick variations, i.e. possess very high frequency components. In order to recover these signals, the classical sampling theory requires an extremely high sampling rate and this results in power consuming A-to-D systems. For example, UWB (ultra-wideband) communications use ultra-short pulses that are nanoseconds in duration to transmit data. A-to-D converters in UWB receivers usually operate at several gigahertz for accurate channel estimation.

1.1 Motivation and objectives

During the past few years, the theory of sampling signals with finite rate of innovations has emerged. This theory characterises the signals in terms of innovation parameters of their

parametric forms instead of their frequency contents. This theory allows sampling and perfect reconstruction of classes of non-bandlimited parametric signals at a rate characterised by how sparse they are per unit of time rather than the highest frequency component of the signals. This provides new insights into sampling methods for those high-frequency real-life signals that can be accurately described by small number of parameters. It has been shown that FRI theory allows the sampling of specific classes of signals at a low rate, much lower than the Nyquist rate. Successful applications can be found in signal and image processing [5–7], communication systems [8, 9] and biological systems [10–13]. Using the previous example, a UWB signal can be modelled as a sparse parametric expression and channel estimation problem can be regarded as finding the locations and amplitudes of a stream of pulses of known shape. Then the sampling rate can be greatly reduced to the number of degrees of freedom of the received UWB signal.

The FRI sampling theory is a very powerful result which breaks our long-existing understanding of the necessary connection between the sampling rate and the signal bandwidth. It is worth analysing its reconstruction performance and also investigating possible extensions of the FRI sampling theory for other applications. In this thesis, we study the limitations of the FRI reconstruction algorithm in the presence of noise. Moreover, we extend the FRI theory to a universal one that works with unknown rate of innovations. We also extend the FRI sampling theory for identifying systems that are parametrically sparse, and for reconstructing images by modelling them as lines of parametrically sparse signals.

1.2 Outline of the thesis

In Chapter 2 we introduce the sampling problem and the traditional solution to it. We also overview the recent development of sampling theory for signals with finite rate of innovations. Moreover, we explain the algorithms for reconstructing such signals.

In Chapter 3 we explain the breakdown phenomenon observed when reconstructing signals with FRI by subspace swap. We also work out at which noise level the absence of subspace swap is guaranteed and this gives us an accurate prediction of the breakdown PSNR. Simulations verify the accuracy of the prediction.

In Chapter 4 we propose a method to identify the rate of innovation, therefore extend the current framework of sampling FRI signals to a universal one which works with arbitrarily unknown number of Diracs.

In Chapter 5 we improve current system identification schemes by using the FRI framework. We first show the possibility of exact identification of specific systems. We then

propose a modified Cadzow algorithm that effectively denoises the spectral data prior to system identification in the case of noisy measurements, which leads to a robust identification framework for linear systems. When both the input sparse signal and the sparse system are unknown, we show it is possible to exactly identify both the signal and the system when the response of the system is of compact support. We also propose a recursive estimation algorithm that works more reliably in the presence of noise.

Chapter 6 is a novel extension of FRI sampling framework for image upsampling. We model lines of images as piecewise smooth functions and propose a resolution enhancement method for this type of functions. We then apply this method along vertical, horizontal and diagonal directions in an image to obtain a single-image super-resolution algorithm. We also propose a further improvement of the method based on learning from the errors of our super-resolution result at lower resolution levels. Simulation results show that our method outperforms state-of-the-art algorithms under different blurring kernels.

1.3 Publications

Journal paper to be submitted

- X. Wei, H. Akhondi Asl, T. Blu, and P. L. Dragotti, “Simultaneous estimation of sparse linear systems and sparse signals using finite rate of innovation principles,” to be submitted, 2016

Journal papers

- X. Wei and P. L. Dragotti, “FRESH – FRI-based single-image super-resolution algorithm,” *IEEE Transactions on Image Processing*, vol. 25, no. 8, pp. 3723–3735, 2016
- X. Wei and P. L. Dragotti, “Guaranteed performance in the FRI setting,” *IEEE Signal Processing Letters*, vol. 22, no. 10, pp. 1661–1665, 2015

Conference papers

- X. Wei and P. L. Dragotti, “Sampling piecewise smooth signals and its application to image up-sampling,” in *2015 IEEE International Conference on Image Processing*, 2015

- X. Wei and P. L. Dragotti, “Universal sampling of signals with finite rate of innovation,” in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 1803–1807

Chapter 2

Overview of recent development in sampling theory

In order to process continuous-time real-world signals with digital systems, we require a sampling mechanism which converts continuous signals to discrete sequences of numbers, while preserving the information present in those signals. In this chapter, we first formulate the sampling problem mathematically, and introduce the classical solution—Shannon sampling theorem, which is a sufficient condition for unique mapping between the discrete sequence to the original continuous signal of finite bandwidth. Then we introduce the recently developed FRI sampling theory, which overcomes the limitations of the classical sampling theory. More specifically, we explain the exact sampling framework for classes of non-bandlimited signals with specific sampling kernels. We also present robust reconstruction settings and algorithms in the presence of noise and an extension of the exact framework that works with arbitrary sampling kernels.

2.1 Sampling problem and the classical sampling theory

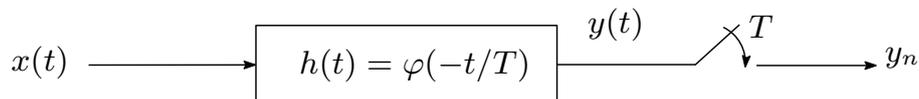


Figure 2.1: Sampling set-up. Here $x(t)$ is the input signal, $h(t)$ is the impulse response of the acquisition device and T is the sampling period. The samples are given by $y_n = \langle x(t), \varphi(t/T - n) \rangle$.

Fig. 2.1 depicts the typical setup for converting a continuous-time signal to discrete sequences. The original continuous-time signal $x(t)$ is filtered with a linear-time invariant

filter with impulse response $h(t)$ and then is sampled with sampling period T . The pre-filtering process may be by design or may be due to the acquisition device. Under this model the samples y_n are given by

$$y_n = \langle x(t), \varphi(t/T - n) \rangle, \quad (2.1)$$

where $\varphi(t)$ is the sampling kernel and is the scaled and time-reversed version of $h(t)$. Note that when no sampling kernel is used, we have $\varphi(t) = \delta(t)$ and $y_n = x(nT)$.

The key issue in sampling theory is to understand whether the set of samples y_n is a faithful representation of the original signal $x(t)$. If so, how can we reconstruct $x(t)$ from the samples? It is well known that Shannon sampling theorem [2] provides an answer to the above questions:

Theorem 2.1 *If a function $x(t)$ contains no frequencies higher than B cycles per second, then samples $y_n = x(nT)$ with $T \leq 1/2B$ are sufficient to reconstruct $x(t)$.*

In this case, the reconstruction process is linear and the reconstruction formulation (see Fig. 2.2 with $\varphi(t) = \delta(t)$ and $\tilde{\varphi}(t) = \text{sinc}(t)$) is given by:

$$x(t) = \sum_{n=-\infty}^{\infty} x(nT) \text{sinc}(t/T - n). \quad (2.2)$$

Shannon sampling theorem is applicable when the input signal $x(t)$ is from the subspace of bandlimited signals. When this is not the case, using a sampling kernel $\varphi(t) = \text{sinc}(t)$ (see Fig. 2.2) allows an approximate reconstruction $\hat{x}(t)$, which is the projection of $x(t)$ onto the subspace of bandlimited signals. Besides the fact that Shannon sampling theorem cannot achieve perfect reconstruction of real-world signals because those signals are never exactly bandlimited [19], we also note that Shannon's reconstruction formula of (2.2) is rarely used in practice because the sinc function has infinite support and slow decay.

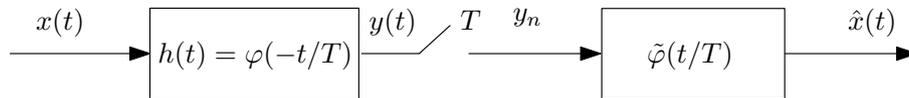


Figure 2.2: Linear reconstruction set-up. The reconstruction of the input signal is given by $\hat{x}(t) = \sum_n y_n \tilde{\varphi}(t/T - n)$. Perfect reconstruction is achieved when $x(t) \in \text{span}\{\tilde{\varphi}(t/T - n)\}_{n \in \mathbb{Z}}$.

In more recent schemes, the space of bandlimited signals in Shannon theory is generalised to the space of shift-invariant signals. See Fig. 2.2 for the sampling and reconstruction

set-up. Here $x(t)$ is reconstructed using a linear filter with response $\tilde{\varphi}(t)$:

$$\hat{x}(t) = \sum_{n=-\infty}^{\infty} y_n \tilde{\varphi}(t/T - n), \quad (2.3)$$

where the pair $\{\tilde{\varphi}(t), \varphi(t)\}$ is chosen so that $\langle \varphi(t - n), \tilde{\varphi}(t - k) \rangle = \delta_{n-k}$. Under this model the sampling and reconstruction process can be interpreted as computing the orthogonal projection of $x(t)$ onto the shift-invariant subspace \mathbf{V} spanned by $\tilde{\varphi}(t)$ and its shifted versions: $\mathbf{V} = \text{span}\{\tilde{\varphi}(t/T - n)\}_{n \in \mathbb{Z}}$. Therefore perfect reconstruction of $x(t)$ is achieved when $x(t) \in \mathbf{V}$.

In this generalised scheme, $\tilde{\varphi}(t)$ does not have to be limited to the sinc function as in the classical Shannon sampling theorem and alternative choices are possible. For example, $\tilde{\varphi}(t)$ could be a polynomial B-spline of a certain order. We also note that slightly more sophisticated forms of linear reconstruction are also possible where, for example, one tries to impose *consistency*. Specifically, the goal is to reconstruct a signal $\hat{x}(t)$ that would lead to the same samples y_n if $\hat{x}(t)$ were to be sampled again. This new constraint leads to a different synthesis filter and to a reconstruction that gives an oblique projection rather than an orthogonal projection of $x(t)$ onto \mathbf{V} . For more details on the topic, we refer to the insightful review [20].

2.2 Sampling and perfect reconstruction of FRI signals

2.2.1 Signals with finite rate of innovation

Signals that are neither bandlimited nor belong to a fixed subspace cannot be reconstructed perfectly using the classical linear reconstruction method discussed in the previous section. However recently it was shown that it is possible to develop sampling schemes for classes of signals having parametric representations with finite number of degrees of freedom. These signals are called signals with finite rate of innovation (FRI) [21] and can be expressed as follows:

$$x(t) = \sum_{k \in \mathbb{Z}} \sum_{r=0}^R a_{r,k} g_r(t - t_k). \quad (2.4)$$

Here $\{g_r(t)\}_{r=0}^R$ is a set of known functions, therefore the only degrees of freedoms in $x(t)$ are the shifts t_k and the amplitudes $a_{r,k}$. If we denote with $C_x(t_a, t_b)$ a function that counts the number of free 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:

$$\rho = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} C_x \left(-\frac{\tau}{2}, \frac{\tau}{2} \right). \quad (2.5)$$

This definition parallels the notion of *information rate* of a source based on the average entropy per unit of time introduced by Shannon in the same 1948 paper [2] where he formulated the sampling theory. The formal definition of signals with finite rate of innovation is given as follows:

Definition 1 *A signal with finite rate of innovation is a signal whose parametric representation is given by (2.4) and with a finite ρ as defined by (2.5).*

In some cases, it is more convenient to consider a local rate of innovation with respect to a moving window of size τ . The local rate of innovation at time t is thus given by

$$\rho_\tau(t) = \frac{1}{\tau} C_x(t - \frac{\tau}{2}, t + \frac{\tau}{2}). \quad (2.6)$$

Note that shift-invariant signals, including bandlimited signals, fall under Definition 1. For example, consider a signal $x(t)$ bandlimited to $[-B/2, B/2]$ which is expressed as (2.2) with $T = 1/B$. Then we can say that it has B degrees of freedom per unit of time (or rate of innovation $\rho = B$) since $x(t)$ is exactly defined by a sequence of real numbers $x(nT)$ spaced $T = 1/B$ seconds apart. Therefore we can say sampling bandlimited signals is possible (Theorem 2.1) because they have finite rate of innovation, rather than because they are bandlimited.

We are particularly interested in the class of signals belonging to (2.4) that are not bandlimited and that cannot be sampled and perfectly reconstructed with the classical sampling theory. Is there a sampling theorem for at least some of these signals? If we stick with the traditional sampling set-up in Fig. 2.1 what kernels $h(t)$ allow for such sampling schemes? What are the reconstruction algorithms?

Classes of FRI signals that can be sampled and reconstructed perfectly include for example streams of Diracs, piecewise polynomial signals and piecewise sinusoids. We show examples of these signals in Fig. 2.3. These signals are fully characterised by some location and amplitude information, therefore reconstruction of these signals is equivalent to recovering these unknown parameters. In a later section we will see that the sequence of samples taken by specific sampling kernel can be transformed to a set of moments τ_m which have a power series form:

$$\tau_m = \sum_{k=0}^{K-1} a_k u_k^m, \quad m = 0, 1, \dots, P \quad (2.7)$$

where u_k, a_k contain the location and amplitude parameters to be estimated respectively. Solving u_k, a_k from $\{\tau_m\}_{m=0}^P$ is a classical problem in spectral estimation [22] and arises in a variety of applications including radar target identification [23], sensor array signal processing [24], signal analysis in power electronics [25], inverse scattering in imaging [26]

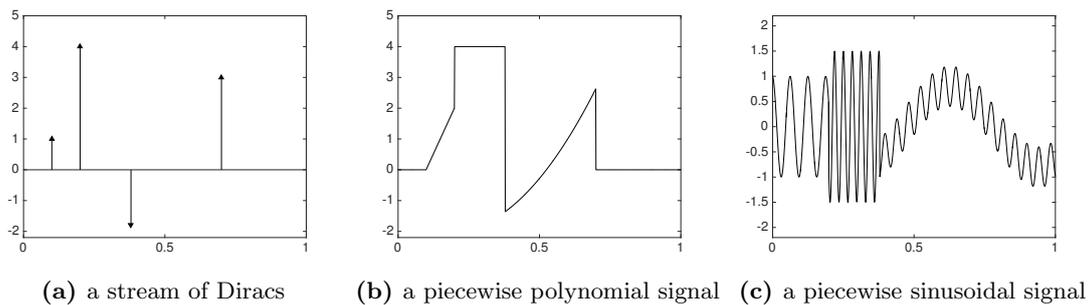


Figure 2.3: Examples of FRI signals. The shape of the signals are known so they are only characterized by the amplitude and location information.

and so forth. There are plenty of ways to solve this problem, for example, Prony’s method or matrix pencil. Reconstruction of the location is non-linear and after the location is recovered, solving for the amplitude is a linear problem. In what follows we discuss different sampling kernels that allow perfect reconstruction, and then we specifically introduce the reconstruction algorithms for recovering stream of Diracs and piecewise polynomial signals from the samples taken with the set-up of Fig. 2.1 because they are useful for the following chapters.

2.2.2 Sampling kernels

There are different sampling kernels that allow perfect reconstruction of FRI signals. In the following we introduce the main two types: the infinite support kernels and the compact support kernels.

The first kernel used for sampling FRI signals is the traditional ideal low-pass filter [21], i.e. the sinc function. Sampling a τ -periodic train of K Diracs with the sinc function with bandwidth B , i.e. $h(t) = \text{sinc}(Bt)$, results in samples y_n and the discrete Fourier coefficients of y_n provide $B\tau$ consecutive Fourier series coefficients of $x(t)$. From at least $2K$ consecutive Fourier series coefficients the locations and amplitudes of K Diracs can be recovered using annihilating method (Prony’s method). However the fundamental limit of the sampling method is that the kernel has infinite support thus is physically non-realizable and it also makes FRI reconstruction algorithms potentially unstable.

Later on, a sampling framework based on compactly supported kernels that are able to reproduce polynomial or exponential functions were presented in [27, 28]. Since these classes of kernels are of compact support, this setup is more stable and practical for sampling finite and infinite duration FRI signals.

A polynomial reproducing kernel of order $P + 1$ is a function that together with its shifted

versions can reproduce polynomials of maximum degree P . That is, any functions $\varphi(t)$ that satisfies:

$$\sum_{n \in \mathbb{Z}} c_{m,n} \varphi(t - n) = t^m, \quad m = 0, 1, \dots, P. \quad (2.8)$$

for proper coefficients $c_{m,n}$. The coefficients $c_{m,n}$ are given by $c_{m,n} = 1/T \int_{-\infty}^{\infty} t^m \tilde{\varphi}(t/T - n) dt$, where $\tilde{\varphi}(t)$ is chosen to form with $\varphi(t)$ a quasi-biorthonormal set [29]. This includes the particular case where $\tilde{\varphi}(t)$ is the dual of $\varphi(t)$. Note that the coefficients can be calculated numerically when the specific kernel that satisfies the polynomial reproduction property is known. This family of kernels includes any functions satisfying the so-called Strang-Fix conditions [30]. Namely, $\varphi(t)$ satisfies (2.8) if and only if

$$\Phi(0) \neq 0 \text{ and } \Phi^{(m)}(j2\pi l) = 0, \quad l \in \mathbb{Z} \setminus \{0\} \text{ and } m = 0, 1, \dots, P, \quad (2.9)$$

where $\Phi(\omega)$ is the Fourier transform of $\varphi(t)$ and here $(\cdot)^{(m)}$ stands for the m -th derivative of (\cdot) .

One important example of functions satisfying Strang-Fix condition is given by the family of B-splines [31]. Zero order B-spline $\beta_0(t)$ is the box function:

$$\beta_0(t) = \begin{cases} 1, & 0 \leq t < 1 \\ 0, & \text{otherwise} \end{cases} \quad \xrightarrow{\mathcal{F}} \quad B_0(j\omega) = \frac{1 - e^{-j\omega}}{j\omega}.$$

A B-spline $\beta_P(t)$ of order P is obtained from the $(P + 1)$ -fold convolution of $\beta_0(t)$:

$$\beta_P(t) = \underbrace{\beta_0(t) * \beta_0(t) \dots * \beta_0(t)}_{P+1 \text{ times}}. \quad (2.10)$$

The B-spline of order P can reproduce polynomials of maximum degree P and the size $(P + 1)$ of its support is the smallest for a function that can achieve that order of approximation. More importantly, it is possible to show that any function $\varphi(t)$ that reproduces polynomials of degree P can be decomposed into a B-spline and a distribution $u(t)$ with $\int u(t) dt \neq 0$, that is, $\varphi(t) = u(t) * \beta_P(t)$ [32–34]. Also note that the B-spline function $\beta_P(t)$ converges to a Gaussian as P tends to infinity [35, Theorem 1].

An exponential reproducing kernel of order $P + 1$ is a function that together with its shifted versions can reproduce complex exponentials of the form $e^{\alpha_m t}$ with $m = 0, 1, \dots, P$. That is, any function that satisfies:

$$\sum_{n \in \mathbb{Z}} c_{m,n} \varphi(t - n) = e^{\alpha_m t}, \quad m = 0, 1, \dots, P \quad (2.11)$$

for proper coefficients $c_{m,n}$. If the exponential reproducing kernel is known, $c_{m,n}$ can be calculated numerically. Note that α_m can be a complex number. It is possible to show that a function satisfies (2.11) if and only if it meet the generalised Strang-Fix conditions [36–38]:

$$\Phi(\alpha_m) \neq 0 \text{ and } \Phi(\alpha_m + j2\pi l) = 0 \quad l \in \mathbb{Z} \setminus \{0\} \quad (2.12)$$

where $\Phi(s)$ is the Laplace transform of $\varphi(t)$.

The theory of exponential reproduction stems from the notion of exponential splines (E-splines) [39]. We denote the E-spline of first order which is able to reproduce the exponential function $e^{\alpha t}$ by β_α :

$$\beta_\alpha(t) = \begin{cases} e^{\alpha t}, & 0 \leq t < 1 \\ 0, & \text{otherwise} \end{cases} \quad \xrightarrow{\mathcal{F}} \quad B_\alpha(j\omega) = \frac{1 - e^{\alpha - j\omega}}{j\omega - \alpha}.$$

Notice that $\beta_\alpha(t)$ reduced to the classical zero-order B-spline when $\alpha = 0$. Higher order E-splines are obtained by successive convolutions of lower-order ones with specific α_m parameters. Therefore the Fourier transform of the E-splines of order $P + 1$ is given by

$$B_\alpha(j\omega) = \prod_{m=0}^P \frac{1 - e^{\alpha_m - j\omega}}{j\omega - \alpha_m}, \quad (2.13)$$

where $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_P)$. An example of exponential reproduction with an e-spline of order $P + 1 = 3$ is shown in Fig. 2.4. Since the exponential reproduction formula is preserved through convolution [39], any composite function of the form $\varphi(t) * \beta_\alpha(t)$ is also able to reproduce exponentials [40]. In the following, we will explain how to sample two specific classes of FRI signals with exponential reproducing kernels.

2.2.3 Exact sampling using exponential reproducing kernels

Consider the sampling set-up in Fig. 2.1 with $\varphi(t)$ being an exponential reproducing kernel. An important characteristic of the exponential reproducing kernel is that it allows us to map the samples y_n with the Laplace or Fourier transform of $x(t)$ at $\{\alpha_m\}_{m=0}^M$ and this is independent of the input signal. Assume that the signal $x(t)$ is of compact support such that it is characterised by only N non-zero samples. By computing weighted sum of these samples, where the weights $c_{m,n}$ are those in (2.11) that reproduce $e^{\alpha_m t}$, we obtain $P + 1$

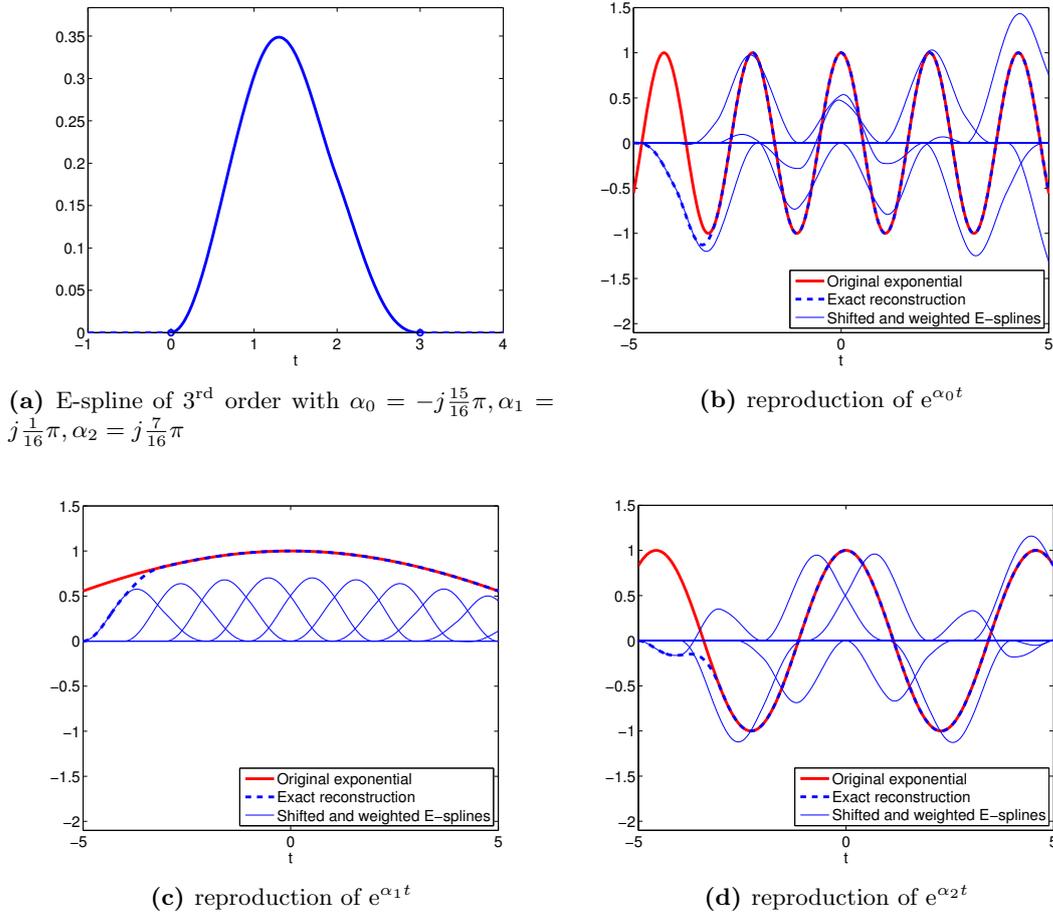


Figure 2.4: Weighted sum of shifted versions of a third order E-spline of (a) can reproduce exponentials at three different frequencies, which are shown in (b)(c)(d).

new measurements, called exponential moments (or e-moments):

$$\begin{aligned} \tau_m &= \sum_n c_{m,n} y_n = \left\langle x(t), \sum_n c_{m,n} \varphi(t/T - n) \right\rangle \\ &= \int_{-\infty}^{\infty} x(t) e^{\alpha_m t/T} dt, \quad m = 0, 1, \dots, P. \end{aligned} \quad (2.14)$$

Note that $\int_{-\infty}^{\infty} x(t) e^{\alpha_m t} dt$ is exactly the bilateral Laplace transform of $x(t)$ evaluated at $\{\alpha_m\}_{m=0}^P$ and denoted by $X(\alpha_m)$. Moreover, when α_m is purely imaginary with $\alpha_m = -j\omega_m$, $X(j\omega_m)$ is the Fourier transform of $x(t)$ at $\omega = \omega_m$.

When $\{\alpha_m\}_{m=0}^P$ are equispaced, i.e. in the form $\alpha_m = \alpha_0 + m\beta$ with $m = 0, \dots, P$, it is possible to establish a one-to-one mapping between $\hat{x}(\alpha_m)$ and $x(t)$ when $x(t)$ is a specific class of signals with FRI. In the following, we discuss two classes of signals allowing the one-to-one mapping—streams of Diracs and piecewise polynomial signals.

Sampling streams of Diracs

Consider a signal $x(t)$ which is a stream of K Diracs located at $\{t_k\}_{k=0}^{K-1}$ (refer to Fig. 2.3a):

$$x(t) = \sum_{k=0}^{K-1} a_k \delta(t - t_k). \quad (2.15)$$

Perfect reconstruction of $x(t)$ from its samples is possible when certain conditions are satisfied:

Theorem 2.2 *Consider a sampling kernel $\varphi(t)$ that can reproduce exponentials $e^{(\alpha_0+m\beta)t}$ with $m = 0, 1, \dots, P$ and $P \geq 2K-1$. Then a stream of K Diracs: $x(t) = \sum_{k=0}^{K-1} a_k \delta(t-t_k)$ is uniquely specified by the samples $y_n = \langle x(t), \varphi(t/T - n) \rangle$.*

The following explains why this is possible. The weighted sum of the samples with weights $c_{m,n}$ in (2.12):

$$\begin{aligned} \tau_m &= \sum_n c_{m,n} y_n \\ &= \int_{-\infty}^{\infty} \sum_{k=0}^{K-1} a_k \delta(t - t_k) e^{\alpha_m t/T} dt \\ &= \sum_{k=0}^{K-1} \underbrace{a_k e^{\alpha_0 t_k/T}}_{\hat{a}_k} \underbrace{e^{(\beta t_k/T)m}}_{u_k^m} \\ &= \sum_{k=0}^{K-1} \hat{a}_k u_k^m, \quad m = 0, 1, \dots, P, \end{aligned} \quad (2.16)$$

is a sum of exponentials.

Retrieving $\{\hat{a}_k, u_k\}_{k=0}^{K-1}$ from $\{\tau_m\}_{m=0}^P$ is a classical problem in spectral estimation and can be solved by annihilating filter method [21, 28] (Prony's method). The key is to note that given a filter $\{h_m\}_{m=0}^K$ such that the roots of its z -transform corresponds to the locations u_k :

$$H(z) = \sum_{m=0}^K h_m z^{-m} = \prod_{k=0}^{K-1} (1 - u_k z^{-1}), \quad (2.17)$$

then this filter can annihilate the moments τ_m :

$$\begin{aligned} h_m * \tau_m &= \sum_{i=0}^K h_i \tau_{m-i} = \sum_{i=0}^K \sum_{k=0}^{K-1} h_i \hat{a}_k u_k^{m-i} \\ &= \sum_{k=0}^{K-1} \hat{a}_k u_k^m \underbrace{\sum_{i=0}^K h_i u_k^{-i}}_{H(u_k)=0} = 0. \end{aligned} \quad (2.18)$$

The zeros of this filter uniquely define the locations t_k of the Diracs. The filter coefficients h_m can be found from the equations $h_m * \tau_m = 0$. The system of equations written in matrix form is as follows:

$$\underbrace{\begin{bmatrix} \tau_K & \tau_{K-1} & \tau_{K-2} & \cdots & \tau_0 \\ \tau_{K+1} & \tau_K & \tau_{K-1} & \cdots & \tau_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \tau_P & \tau_{P-1} & \tau_{P-2} & \cdots & \tau_{P-K} \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} h_0 \\ h_1 \\ \vdots \\ h_K \end{pmatrix}}_{\mathbf{h}} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad (2.19)$$

where \mathbf{A} is a rank-deficient Toeplitz matrix with rank K . Since $h_0 = 1$, we can rearrange (2.19) to a system solving for $\{h_1, \dots, h_K\}$. Then we need at least $2K$ consecutive values of τ_m to solve this system, hence the order $P + 1$ of the sampling kernel must be equal or larger than $2K$, which is exactly the number of degrees of freedom of $x(t)$. After we get h_m , the exact locations t_k are obtained by finding the roots u_k of $H(z)$ and by using the fact that $u_k = e^{\beta t_k/T}$. Finally the exact amplitudes a_k can be retrieved by solving, for example, the first K consecutive equations in (2.16).

Sampling piecewise polynomials

We consider piecewise polynomial functions with K pieces of maximum degree $R - 1$ (refer to Fig. 2.3b):

$$p(t) = \sum_{k=0}^{K-1} \sum_{r=0}^{R-1} a_{k,r} (t - t_k)_+^r, \quad (2.20)$$

where $t_+^r = \max(t, 0)^r$. Sampling and perfect reconstruction of these functions is possible under the following conditions:

Theorem 2.3 *Assume a sampling kernel $\varphi(t)$ that can reproduce exponentials $e^{(\alpha_0 + m\beta)t}$ with $m = 0, 1, \dots, P$ and $P \geq 2KR - 1$. A piecewise polynomial signal with K pieces of maximum degree $R - 1$: $p(t) = \sum_{k=0}^{K-1} \sum_{r=0}^{R-1} a_{k,r} (t - t_k)_+^r$ is uniquely determined from the samples $y_n = \langle x(t), \varphi(t/T - n) \rangle$.*

The following explains why this is possible. We denote the R -th derivative of $p(t)$ by $p^{(R)}(t)$, and note that $p^{(R)}(t)$ is a stream of (differentiated) Diracs. This means that if we are able to relate the samples of $p(t)$ to those of $p^{(R)}(t)$ we can interpret the problem as the one of sampling differentiated Diracs. Indeed this is possible by using the link between discrete differentiation and derivation in continuous domain. More specifically, consider a function $\varphi(t/T)$ with Fourier transform $T\Phi(j\omega T)$ and the difference $\varphi(t/T) - \varphi(t/T - 1)$.

The Fourier transform of $\varphi(t/T) - \varphi(t/T - 1)$ is

$$\begin{aligned} \varphi(t/T) - \varphi(t/T - 1) &\xrightarrow{\mathcal{F}} T\Phi(j\omega T)(1 - e^{-j\omega T}) \\ &= j\omega T\Phi(j\omega T) \cdot T \frac{1 - e^{-j\omega T}}{j\omega T} \\ &= j\omega \cdot T\Phi(j\omega T) \cdot TB_0(j\omega T). \end{aligned} \quad (2.21)$$

Therefore

$$\varphi(t/T) - \varphi(t/T - 1) = \frac{d}{dt}[\varphi(t/T) * \beta_0(t/T)], \quad (2.22)$$

where β_0 is the B-spline of order 0 and its Fourier transform is $\frac{1 - e^{-j\omega}}{j\omega}$. Let $z_n^{(1)}$ denote the finite difference $y_{n+1} - y_n$. It follows that

$$\begin{aligned} z_n^{(1)} = y_{n+1} - y_n &= \langle p(t), \varphi(t/T - n - 1) - \varphi(t/T - n) \rangle \\ &\stackrel{(a)}{=} \left\langle p(t), -\frac{d}{dt}[\varphi(t/T - n) * \beta_0(t/T - n)] \right\rangle \\ &\stackrel{(b)}{=} \left\langle \frac{dp(t)}{dt}, \varphi(t/T - n) * \beta_0(t/T - n) \right\rangle, \end{aligned} \quad (2.23)$$

where (a) follows from (2.22) and (b) is obtained using integration by parts.

Using a similar derivation it is also possible to prove that the R -th finite differences

$$z_n^{(R)} = \sum_{k=0}^R (-1)^{R-k} \binom{R}{k} y_{n+k} \quad \text{with } y_n = \langle p(t), \varphi(t/T - n) \rangle \quad (2.24)$$

is equivalent to the samples $\langle p^{(R)}(t), \varphi_{\text{eq}}(t/T - n) \rangle$ obtained by acquiring $p^{(R)}(t)$ with the new kernel $\varphi_{\text{eq}}(t) = \varphi(t) * \beta_{R-1}(t)$, where $\beta_{R-1}(t)$ is the polynomial B-spline of degree $R - 1$.

The new kernel $\varphi_{\text{eq}}(t) = \varphi(t) * \beta_{R-1}(t)$ also satisfies the generalised Strang-Fix conditions of (2.12) and can reproduce exponentials:

$$\sum_{n \in \mathbb{Z}} c_{m,n} \varphi_{\text{eq}}(t/T - n) = e^{\alpha_m t/T} \quad \text{with } m = 0, 1, \dots, P \quad (2.25)$$

for a proper choice of coefficients $c_{m,n}$.

Because of this connection in what follows we focus on the reconstruction of streams of differentiated Diracs to solve the problem of sampling piecewise polynomials.

The stream of differentiated Diracs $p^{(R)}(t)$ can be expressed as follows:

$$p^{(R)}(t) = \sum_{k=0}^{K-1} \sum_{r=0}^{R-1} a_{k,r} \delta^{(r)}(t - t_k), \quad (2.26)$$

and the samples $z_n^{(R)}$ can be written as:

$$z_n^{(R)} = \left\langle p^{(R)}(t), \varphi_{\text{eq}}(t/T - n) \right\rangle. \quad (2.27)$$

Consider now the following weighted sum of the samples $z_n^{(R)}$, where the weights $c_{m,n}$ are those in (2.25) that reproduce $e^{\alpha_m t/T}$, we have:

$$\begin{aligned} \tau_m &= \sum_n c_{m,n} z_n^{(R)} = \left\langle p^{(R)}(t), \sum_n c_{m,n} \varphi_{\text{eq}}(t/T - n) \right\rangle \\ &= \int_{-\infty}^{\infty} p^{(R)}(t) e^{\alpha_m t/T} dt \\ &= \int_{-\infty}^{\infty} \sum_{k=0}^{K-1} \sum_{r=0}^{R-1} a_{k,r} \delta^{(r)}(t - t_k) e^{\alpha_m t/T} dt \\ &= \sum_{k=0}^{K-1} \sum_{r=0}^{R-1} \hat{a}_{k,r} (\alpha_m)^r u_k^m, \quad m = 0, 1, \dots, P, \end{aligned} \quad (2.28)$$

where $\hat{a}_{k,r} = (-1/T)^r e^{\alpha_0 t_k/T} a_{k,r}$ and $u_k = e^{\beta t_k/T}$.

The locations t_k 's can be retrieved from τ_m using Prony's method (annihilating filter method). The key is to note that given a filter $\{h_m\}_{m=0}^{KR}$ whose z -transform is:

$$H(z) = \sum_{m=0}^{KR} h_m z^{-m} = \prod_{k=0}^{K-1} (1 - u_k z^{-1})^R, \quad (2.29)$$

then this filter can annihilate the sequence τ_m . That is, $h_m * \tau_m = 0$. Similar to the case of recovering stream of Diracs, the KR unknown coefficients of h_m can be found by writing $h_m * \tau_m = 0$ in matrix/vector form using at least $2KR$ consecutive τ_m . From the roots u_k of the annihilating filter $H(z)$ and by using the fact that $u_k = e^{\beta t_k/T}$ we obtain the locations t_k exactly. Then the exact amplitudes can be found by solving, for example, the first KR equations in (2.28).

2.3 Sampling FRI signals in the presence of noise

When there is no noise, incoming FRI signals can be recovered from the moments τ_m regardless of the specific design of the exponential reproducing kernel as long as the con-

dition on the number of exponentials reproduced is satisfied (refer to Theorem 2.2 and Theorem 2.3) and the solution is unique when the locations are distinct.

However when the measurements y_n are corrupted with additive noise, in the case of sampling streams of Diracs, we have only access to:

$$\tilde{y}_n = \underbrace{\sum_{k=0}^{K-1} a_k \varphi(t_k/T - n)}_{y_n} + \epsilon_n, \quad n = 0, 1, \dots, N-1, \quad (2.30)$$

where we assume that ϵ_n are i.i.d. Gaussian random variables with zero mean and standard deviation σ_y . The moments τ_m become noisy as shown below:

$$\tilde{\tau}_m = \sum_n c_{m,n} \tilde{y}_n = \tau_m + \underbrace{\sum_n c_{m,n} \epsilon_n}_{b_m}, \quad m = 0, 1, \dots, P \quad (2.31)$$

or in matrix form:

$$\tilde{\boldsymbol{\tau}} = \mathbf{C} \tilde{\mathbf{y}} = \mathbf{C} \mathbf{y} + \mathbf{C} \boldsymbol{\epsilon}, \quad (2.32)$$

where $\tilde{\boldsymbol{\tau}} = (\tilde{\tau}_0, \tilde{\tau}_1, \dots, \tilde{\tau}_P)^\top$, $\mathbf{y} = (y_0, y_1, \dots, y_{N-1})^\top$, $\boldsymbol{\epsilon} = (\epsilon_0, \epsilon_1, \dots, \epsilon_{N-1})^\top$ and $\mathbf{C} \in \mathbb{C}^{P \times N}$ is a matrix whose entry at location (m, n) is $c_{m,n}$. In order to limit the effect of noise, we need to understand the effect that noise on the samples has on the moments and also how to improve the reconstruction algorithms. Therefore, in the following we investigate both the optimal design of the sampling kernel and the design of good reconstruction algorithms. We then introduce the Cramér-Rao lower bound which can be used to assess the performance of the reconstruction algorithms in noisy settings.

2.3.1 Design of the sampling kernel

From (2.32) we see that the statistics of the noisy moments is highly related to \mathbf{C} . We desire a well conditioned \mathbf{C} in order to have a good $\tilde{\boldsymbol{\tau}}$. Matrix \mathbf{C} is composed of entries $c_{m,n} = c_{m,0} e^{\alpha_m n}$ at position (m, n) . To have a stable \mathbf{C} we want $|c_{m,0}|$ to have same magnitude for all $m = 0, 1, \dots, P$, for example $|c_{m,0}| = 1$. Although the cardinal e-spline does not have this property, given the fact that any composite function of the form $\varphi(t) * \beta_\alpha(t)$ is also able to reproduce exponentials, it is possible to construct exponential reproducing functions satisfying $|c_{m,0}| = 1$. One such family of exponential reproducing kernels are called exponential spline with Maximum Order and Minimum Support (eMOMS) [38, 40]. More details on how to design eMOMS are in [40]. An eMOMS which reproduces same exponentials as $\beta_\alpha(t)$ has minimum support as $\beta_\alpha(t)$ as well. For this reason for the rest of this thesis, we use eMOMS when we mention exponential reproducing kernel. An example

on an eMOMS of order $P + 1 = 31$ is shown in Fig. 2.5(b). Please note that the eMOMS in Fig. 2.5(b) has the same support as the cardinal e-spline in (a).

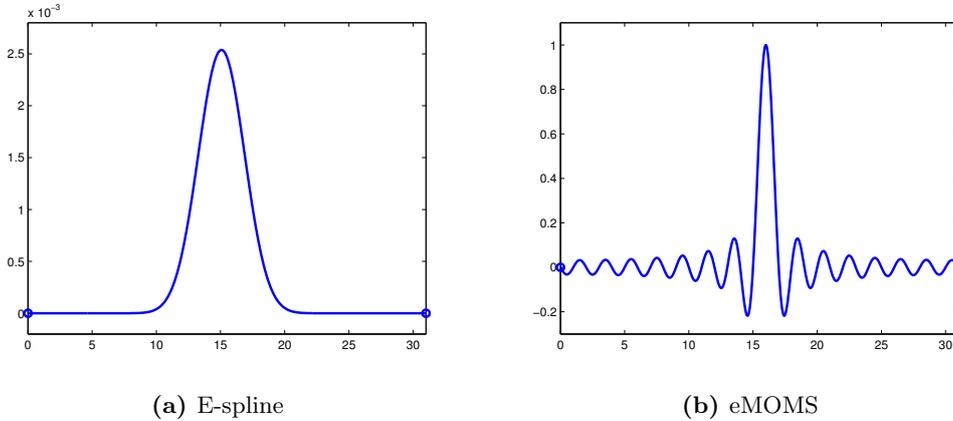


Figure 2.5: Examples of exponential reproducing functions. (a) E-spline of order $P + 1 = 31$. (b) eMOMS of order $P + 1 = 31$.

In addition, in all situations we always want to choose $\alpha_m = m'\beta$, with $m' = m - P/2$, $m = 0, 1, \dots, P$ with $\beta = j \frac{2\pi}{P+1}$ so that the exponentials reproduced by the kernel span the unit circle, which is well known to be the best configuration when recovering the parameters of a power series. At last, whenever possible, we prefer to have a kernel that can produce number of exponentials equalling the number of the samples N , i.e. $P + 1 = N$, this leads to the most stable matrix \mathbf{C} (square and unitary) and the best possible performance for a fixed N [38].

2.3.2 Robust reconstruction algorithms

We have just explained the specific exponential reproducing kernels that gives best moments for reconstruction. We also desire robust reconstruction algorithms which hopefully help us separate the true signal information from the noisy moments. In the following we explain the robust algorithms for recovering streams of Diracs, but note that they are also applicable for piecewise polynomial functions.

TLS solution of Prony's method

When there is noise, \mathbf{A} becomes noisy and no longer rank-deficient and (2.19) is not satisfied exactly. We denote the noisy \mathbf{A} with $\tilde{\mathbf{A}}$. In order to have a more robust reconstruction, we can increase the number of moments by using a sampling kernel that reproduce a number of exponentials $P + 1$ which is greater than the critical number $2K$ (refer to

Theorem 2.2). The total least square (TLS) solution of Prony's method can be found by performing singular value decomposition (SVD) on $\tilde{\mathbf{A}}$. The eigenvector corresponding to the smallest eigenvalue is the desired estimation of \mathbf{h} . Once the t_k are retrieved, the a_k follow from a least square minimization of the difference between the moments $\tilde{\tau}_m$ and the model of (2.16).

Cadzow denoising algorithm

Before Prony's method, we can use an initial denoising (model matching) step, called Cadzow iterative algorithm [41, 42]. It denoises the moments by recovering a closest rank deficient matrix which is Toeplitz. To accelerate the denoising procedure, rather than constructing the Toeplitz matrix \mathbf{A} of size $(P + 1 - K) \times (K + 1)$ built from $P + 1$ consecutive τ_m in (2.19) in the noiseless case, we consider an extended Toeplitz matrix \mathbf{T} of size $(P + 1 - L) \times (L + 1)$:

$$\mathbf{T} = \begin{bmatrix} \tau_L & \tau_{L-1} & \tau_{L-2} & \cdots & \tau_0 \\ \tau_{L+1} & \tau_L & \tau_{L-1} & \cdots & \tau_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \tau_P & \tau_{P-1} & \tau_{P-2} & \cdots & \tau_{P-L} \end{bmatrix} \quad (2.33)$$

with $L \geq K$ constructed using $P + 1$ consecutive moments. In noiseless situation, \mathbf{T} is also of rank K .

The noisy \mathbf{T} , denoted by $\tilde{\mathbf{T}}$, in the presence of noise is full rank and the annihilation in (2.18) is not satisfied any more. The main idea of Cadzow algorithm is to recover the structured low-rank matrix from the full-rank matrix $\tilde{\mathbf{T}}$. We first compute the SVD of $\tilde{\mathbf{T}}$ and we only keep the K largest singular values of $\tilde{\mathbf{T}}$ and set the rest to zero. Now this new matrix is no longer Toeplitz but its best Toeplitz approximation can be obtained by averaging the diagonal elements. A few these iterations lead to a denoised set of moments. Now from the denoised moments we solve the annihilating filter by the total least square solution of Prony's method.

Matrix pencil

An alternative to "Cadzow+Prony" method which achieves comparable denoising performance is given by the matrix pencil method [43–45] which operates as follows.

We perform the SVD of $\tilde{\mathbf{T}}$: $\tilde{\mathbf{T}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$ and keep the K left-singular vectors that correspond to the K largest singular values: $\mathbf{U}_K = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_K]$. \mathbf{U}_K will be good estimates of the singular vectors of the noiseless \mathbf{T} when the presence of additive white noise has

no significant effect on the principal singular vectors. Denote $\overline{\mathbf{U}}_K$ and $\underline{\mathbf{U}}_K$ the matrix \mathbf{U}_K after omission of the first and the last row respectively. It can be shown that when there is no noise, having λ to be one of the parameters $\{u_k\}_{k=1}^K$ in (2.16) would reduce the rank of $\overline{\mathbf{U}}_K - \lambda\underline{\mathbf{U}}_K$, and thus $\{u_k\}_{k=0}^{K-1}$ can be solved by finding the eigenvalues λ of the generalised eigenvalue problem:

$$\overline{\mathbf{U}}_K \mathbf{v} = \lambda \underline{\mathbf{U}}_K \mathbf{v} \iff (\overline{\mathbf{U}}_K - \lambda \underline{\mathbf{U}}_K) \mathbf{v} = \mathbf{0}. \quad (2.34)$$

This generalised eigenvalue problem can be transformed to a simple eigenvalue problem by multiplying both sides by the pseudo-inverse of $\underline{\mathbf{U}}_K$:

$$(\underline{\mathbf{U}}_K^\dagger \overline{\mathbf{U}}_K - \lambda \mathbf{I}) \mathbf{v} = \mathbf{0}. \quad (2.35)$$

2.3.3 Performance of the reconstruction algorithms

Any unbiased estimator of deterministic parameters has a covariance matrix that is lower bounded by the Cramér-Rao bounds [42, 46–48]. An unbiased estimator which achieves this lower bound achieves the lowest possible variance among all unbiased methods. So we could assess the estimation performance of the reconstruction algorithms in the presence of noise by comparing it to the theoretical minimum provided by the Cramér-Rao bound.

In the context of recovering a stream of K Diracs, the estimation problem consists in estimating all the unknown locations and amplitudes $\boldsymbol{\theta} = (t_0, \dots, t_{K-1}, a_0, \dots, a_{K-1})^\top$ from N noisy samples $\tilde{\mathbf{y}} = (\tilde{y}_0, \dots, \tilde{y}_{N-1})^\top$ of (2.30), which is the sum of the deterministic parametric model $f(n, \boldsymbol{\theta}) = \sum_{k=0}^{K-1} a_k \varphi(t_k/T - n)$ and the additive Gaussian noise $\epsilon_n \sim \mathcal{N}(0, \sigma_y^2)$ with covariance matrix $\mathbf{R}_\epsilon = \mathbb{E}\{\boldsymbol{\epsilon}\boldsymbol{\epsilon}^H\} = \sigma_y^2 \mathbf{I}$, where $\boldsymbol{\epsilon}$ is a vector containing values $\{\epsilon_n\}_{n=0}^{N-1}$. We denote any unbiased estimate of the unknown parameters from $\tilde{\mathbf{y}}$ by $\hat{\boldsymbol{\theta}}(\tilde{\mathbf{y}})$.

It is known that the covariance matrix of $\hat{\boldsymbol{\theta}}$ is lower bounded by the inverse of the Fisher information matrix $\mathbf{I}(\boldsymbol{\theta})$:

$$\text{cov}\{\hat{\boldsymbol{\theta}}(\tilde{\mathbf{y}})\} \geq \underbrace{(\mathbf{G}_y^\top \mathbf{R}_\epsilon^{-1} \mathbf{G}_y)^{-1}}_{\mathbf{I}(\boldsymbol{\theta})}, \quad (2.36)$$

where matrix \mathbf{G}_y is a $N \times 2K$ matrix where each row is given by vector $\nabla f(n, \boldsymbol{\theta})$. For simplicity we assume the sampling period T is 1. The expression of \mathbf{G}_y is as follows:

$$\mathbf{G}_y = \left[\begin{array}{ccc|ccc} a_0 \varphi'(t_0) & \dots & a_0 \varphi'(t_{K-1}) & \varphi(t_0) & \dots & \varphi(t_{K-1}) \\ a_0 \varphi'(t_0-1) & \dots & a_0 \varphi'(t_{K-1}-1) & \varphi(t_0-1) & \dots & \varphi(t_{K-1}-1) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_0 \varphi'(t_0-(N-1)) & \dots & a_0 \varphi'(t_{K-1}-(N-1)) & \varphi(t_0-(N-1)) & \dots & \varphi(t_{K-1}-(N-1)) \end{array} \right]. \quad (2.37)$$

In the specific FRI recovery problem where the sampling kernel is exponential reproducing

kernel, the unknown parameters $\boldsymbol{\theta}$ are recovered from the exponential moments $\{\tilde{\tau}_m\}_{m=0}^P$, which is now the sum of the deterministic parametric model $f(m, \boldsymbol{\theta}) = \sum_{k=0}^{K-1} a_k e^{\alpha_m t_k/T}$ and the noise on moments, i.e. $\{b_m\}_{m=0}^P = \sum_n c_{m,n} \epsilon_n$ in (2.31). Therefore, it is of interest to find the Cramér-Rao bound associated to the measurements $\{\tilde{\tau}_m\}_{m=0}^P$ which indicates the best performance that can be achieved when working with $\{\tilde{\tau}_m\}_{m=0}^P$. In this case, the covariance matrix of $\hat{\boldsymbol{\theta}}$ is lower bounded by

$$\text{cov}\{\hat{\boldsymbol{\theta}}(\tilde{\boldsymbol{\tau}})\} \geq \underbrace{(\mathbf{G}_\tau^H \mathbf{R}_b^{-1} \mathbf{G}_\tau)^{-1}}_{\mathbf{I}(\boldsymbol{\theta})}, \quad (2.38)$$

where $\mathbf{R}_b = \mathbb{E}\{\mathbf{b}\mathbf{b}^H\}$, where \mathbf{b} is a vector containing $\{b_m\}_{m=0}^P$. Again for simplicity we assume the sampling period T is 1. Matrix \mathbf{G}_τ takes the form:

$$\mathbf{G}_\tau = \begin{bmatrix} a_0 \alpha_0 e^{\alpha_0 t_0} & \dots & a_{K_1} \alpha_0 e^{\alpha_0 t_{K-1}} & \left| & e^{\alpha_0 t_0} & \dots & e^{\alpha_0 t_{K-1}} \\ a_0 \alpha_1 e^{\alpha_1 t_0} & \dots & a_{K_1} \alpha_1 e^{\alpha_1 t_{K-1}} & \left| & e^{\alpha_1 t_0} & \dots & e^{\alpha_1 t_{K-1}} \\ \vdots & \ddots & \vdots & \left| & \vdots & \ddots & \vdots \\ a_0 \alpha_P e^{\alpha_P t_0} & \dots & a_{K_1} \alpha_P e^{\alpha_P t_{K-1}} & \left| & e^{\alpha_P t_0} & \dots & e^{\alpha_P t_{K-1}} \right. \end{bmatrix}. \quad (2.39)$$

Note that $\mathbf{R}_b = \sigma_y^2 \mathbf{C}\mathbf{C}^H$ and retrieving the unknown parameters from the moments when the number of moments $P+1 < N$ is suboptimal. However, when the number of moments $P+1$ equals the number of samples N , it can be shown that $\mathbf{R}_b = \mathbf{R}_\epsilon$, and the optimal estimation results from the moments equals that from the samples. More details can be found in [38].

Now we want to compare the square root of the theoretical lowest variances of the Dirac locations given by the sample-based/moment-based Cramér-Rao bounds with the standard deviation of the recovered locations $\{\hat{t}_k^{(i)}\}_{i=1}^I$ in I realizations given by $\Delta t_k = \sqrt{\sum_{i=1}^I (t_k - \hat{t}_k^{(i)})^2 / I}$, where t_k is the true location of the Dirac and I is the total number of realisations per noise level. Previous works in FRI, e.g. in [38, 42, 49], have shown that FRI algorithms achieve the Cramér-Rao bound up to certain noise level. Please see Fig. 2.6 for the performance of Cadzow+Prony method in recovering the locations of a stream of two Diracs from its samples taken with an exponential reproducing kernel.

2.4 Approximate Strang-Fix: Sampling FRI signals with arbitrary sampling kernel

We have shown that when the sampling kernel $\varphi(t)$ is an exponential reproducing kernel, we can retrieve the exact moments of the input signals and then recover the unknown parameters of the FRI signals exactly. For any other kernel which does not satisfy the generalised Strang-Fix conditions of (2.12) we want to find a linear combination of $\varphi(t)$

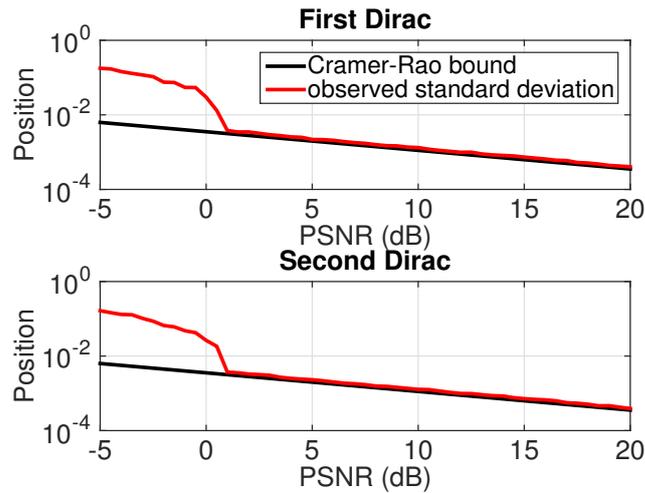


Figure 2.6: Standard deviation (averages over 1000 realizations) of the retrieved locations of a FRI signal (a stream of two Diracs) compared to Cramér-Rao lower bounds (CRB). The estimation algorithm (Cadzow+Prony) achieves CRB up to certain peak signal-to-noise ratio (PSNR). The two Diracs are recovered from 29 samples taken by a kernel which can reproduce 29 exponentials.

with its shifted versions that provides the best approximation to a specific exponential [38]. More precisely, we want to find coefficients $c_{m,n}$ such that:

$$\sum_{n \in \mathbb{Z}} c_{m,n} \varphi(t - n) \approx e^{\alpha_m t}. \quad (2.40)$$

For the sake of clarity, we use $c_{m,n} = c_{m,0} e^{\alpha_m n}$ and then it can be shown that the error in approximating the exponential is [38]:

$$\epsilon_{\text{approx},m}(t) = e^{\alpha_m t} \left[1 - c_{m,0} \sum_{l \in \mathbb{Z}} \Phi(\alpha_m + j2\pi l) e^{j2\pi l t} \right]. \quad (2.41)$$

Note that if the Fourier transform of $\varphi(t)$ decays sufficiently quickly, which is true for any low-pass filter, we can assume the terms $\Phi(\alpha_m + j2\pi l)$ are close to zero for $l \in \mathbb{Z} \setminus \{0\}$. In this case, the approximation error is small and is minimised when

$$c_{m,n} = \Phi(\alpha_m)^{-1} e^{\alpha_m n}. \quad (2.42)$$

Note that when $\varphi(t)$ satisfies the generalised Strang-Fix conditions, this equation gives exactly the $c_{m,n}$ for exact exponential reproduction. Having $c_{m,n}$ we are now able to calculate the approximate moments τ_m of the input signal by $\sum_n c_{m,n} y_n$, where y_n are the samples of the input signal taken by the low-pass filter $\varphi(t)$. From the moments, the input signal can be recovered using the same methods for the case of exponential reproducing kernels in Sec. 2.3.2.

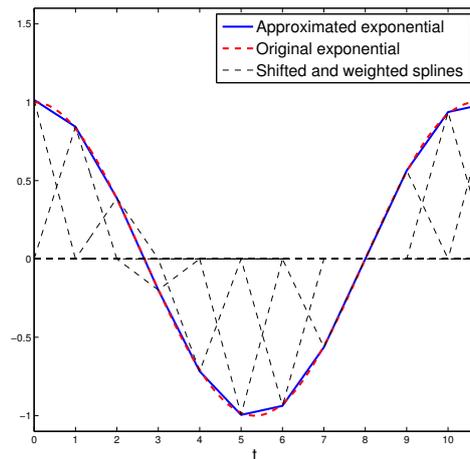


Figure 2.7: Approximate reproduction of $e^{-j\frac{3\pi}{16}t}$ with a linear spline.

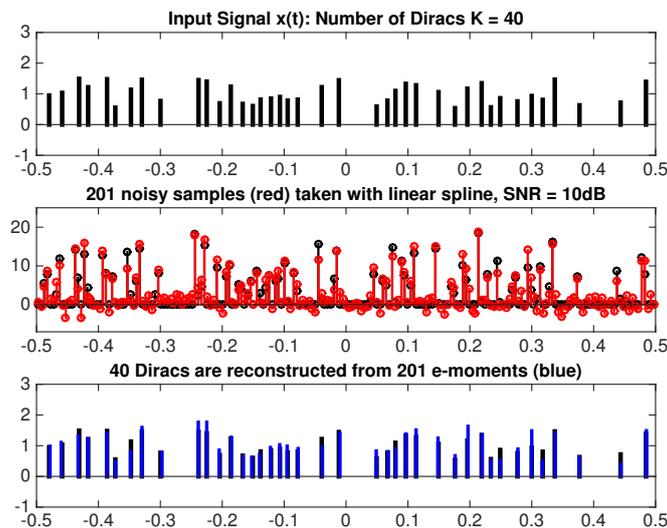


Figure 2.8: Approximate reconstruction of a stream of Diracs from samples taken with a linear spline when SNR=10dB.

This approximate Strang-Fix theory enables accurate reconstruction of the input FRI signal with arbitrary sampling kernels that behave approximately like low-pass filters. An example of approximate reproduction of exponential with a linear spline is shown in Fig. 2.7 and reconstruction of a stream of Diracs with the linear spline when SNR=10dB noise is demonstrated in Fig. 2.8.

This approximate theory will be particularly useful for Chapter 6 on image upscaling since point spread functions are approximately low-pass filters.

A note on selecting frequencies of exponentials As discussed in Section 2.3.1, on the one hand we want the exponentials $e^{\alpha_m t}$ to span the unit circle, on the other hand, we want $|c_{m,0}| = |\Phi(\alpha_m)^{-1}|$ to be as constant as possible for all m . While this can be satisfied at the same time in the exact reproduction framework, for general low-pass filters the two requirements lead to a trade-off in the choice of α_m since the former means spread-out of the exponents while the latter means the frequencies are concentrated around $\omega = 0$. In [38], Urigüen et al. determine the optimal α_m for specific low-pass filters through numerical simulations.

Chapter 3

Guaranteed performance of the FRI reconstruction algorithm

FRI sampling theory has shown that it is possible to sample and perfectly reconstruct classes of non-bandlimited signals such as streams of Diracs. In the case of noisy measurements, we use robust algorithms like Cadzow iterative algorithm [41] or matrix pencil [43], which have been introduced in Section 2.3. They are SVD-based methods, which are centered around splitting the measurement space into an estimated signal-subspace and an orthogonal-subspace. The knowledge of the orthogonal subspace allows the unique reconstruction of the FRI signal. FRI methods achieve the optimal performance given by the Cramér-Rao bound up to a certain PSNR and breaks down for smaller PSNRs (see Fig. 2.6).

It is well known that this breakdown phenomenon appears in non-linear estimation problems, e.g. direction of arrival, and has been studied in the past [50–52]. Knowing the breakdown is crucial for ensuring the estimation algorithm operates in the asymptotic region predicted by the Cramér-Rao bound. To the best of our knowledge, however, precise anticipation of the breakdown in FRI framework is still an open question. In this chapter, we explain the breakdown event by *subspace swap*, which has been broadly recognised as the reason of performance breakdown in SVD-based parameter estimation algorithms. For subspace swap we mean the situation when due to noise the orthogonal subspace mixes with the signal subspace making the retrieval of the signal unreliable. We then work out at which noise level the absence of subspace swap is guaranteed and this gives us an accurate prediction of the breakdown PSNR which we also relate to the sampling rate and the distance between adjacent Diracs.

3.1 A revisit to FRI theory

Now we briefly revisit the FRI sampling method where the emphasis is given only to key aspects of the method which will be used in Section 3.2 to predict when noisy FRI recovery fails.

We consider sampling a stream of Diracs $x(t) = \sum_{k=0}^{K-1} a_k \delta(t - t_k)$ with $t_k \in [0, 1)$, $k = 0, 1, \dots, K - 1$ using the sampling kernel $\varphi(t)$ at sampling rate $1/T$ (refer to Fig. 2.1). Under this model the observed samples are given by

$$y_n = \langle x(t), \varphi(t/T - n) \rangle, \quad n = 0, 1, \dots, N - 1. \quad (3.1)$$

The goal is to estimate the locations $\{t_k\}_{k=0}^{K-1}$ and the amplitudes $\{a_k\}_{k=0}^{K-1}$ from the N samples y_n (sampling rate $T = 1/N$). As discussed in the previous chapter, exact retrieval of these parameters is possible when the sampling kernel $\varphi(t)$ is an exponential reproducing function. Assume we now have a $P + 1$ -th order exponential reproducing kernel $\varphi(t)$, then linear combinations of $\varphi(t)$ with its integer shifts can reproduce $P + 1$ exponentials $e^{\alpha_m t}$, $m = 0, 1, \dots, P$:

$$\sum_{n \in \mathbb{Z}} c_{m,n} \varphi(t - n) = e^{\alpha_m t}, \quad m = 0, 1, \dots, P \quad (3.2)$$

for a proper choice of coefficients $c_{m,n}$.

Here we consider the exponential reproducing kernel which reproduces complex conjugated exponential pairs. That is, α_m in (3.2) have the form $\alpha_m = m' \beta$ with $m' = m - P/2$, $m = 0, \dots, P$ and β is an imaginary number. In the following, we assume P is even and we use $2M$ instead of P (i.e. $M = P/2$) for simplicity and clear presentation of derivations. We use this kernel to take the N samples y_n of $x(t)$. By linearly combining the samples y_n with the coefficients $c_{m,n}$ in (3.2) we obtain the following $2M + 1$ exponential moments:

$$\tau_m = \sum_n c_{m,n} y_n = \sum_{k=0}^{K-1} \hat{a}_k u_k^m, \quad m = 0, 1, \dots, 2M, \quad (3.3)$$

where $\hat{a}_k = e^{-(\beta t_k/T)M}$ and $u_k = e^{\beta t_k/T}$. Recall from Section 2.2.3, retrieving u_k from τ_m can be solved by using the annihilating filter method, which solves for the filter h_m that annihilates τ_m , i.e. $h_m * \tau_m = 0$. It can be written in matrix form of (2.19): $\mathbf{A} \mathbf{h} = \mathbf{0}$.

In noiseless situations, \mathbf{A} is rank deficient (rank K). Therefore, the null space has dimension one and we can solve for \mathbf{h} . Then from the roots of $H(z)$ we retrieve the locations exactly. Given $\{u_k\}_{k=0}^{K-1}$, the amplitudes are retrieved by solving, for example, the first K

consecutive equations in (3.3).

Note that the subspace spanned by the K singular vectors related to the non-zero singular values of \mathbf{A} is called the signal subspace. Also note that an extended Toeplitz matrix \mathbf{T} (see (2.33)), which is built from $\tau_m, m = 0, 1, \dots, 2M$ and is of size $(2M + 1 - L) \times (L + 1)$ with $L \geq K$, is also of rank K and there are $L - K + 1$ independent vectors $\{h_l\}_{l=0}^L$ which can annihilate τ_m . This is shown in matrix form as follows:

$$\mathbf{T}\mathbf{h}_{L+1} = \mathbf{0}, \quad (3.4)$$

Now we assume the measurements y_n are corrupted with additive noise, and we have access to

$$\tilde{y}_n = y_n + \epsilon_n, \quad n = 0, 1, \dots, N - 1, \quad (3.5)$$

where ϵ_n are i.i.d Gaussian random variables with zero mean and standard deviation σ_y . The moments τ_m become noisy as shown below:

$$\tilde{\tau}_m = \sum_n c_{m,n} \tilde{y}_n = \tau_m + \underbrace{\sum_n c_{m,n} \epsilon_n}_{b_m}, \quad m = 0, 1, \dots, 2M. \quad (3.6)$$

The noisy \mathbf{T} , denoted by $\tilde{\mathbf{T}}$, is now full rank and (3.4) is not satisfied any more. We can look for its total least square (TLS) solution that can be found by performing singular value decomposition (SVD) on $\tilde{\mathbf{T}}$ with $L = K$. The eigenvector corresponding to the smallest eigenvalue is the desired estimation of \mathbf{h} . As introduced in the previous chapter, before looking for the TLS solution the moments $\{\tau_m\}_{m=0}^{2M}$ can be first denoised by Cadzow denoising algorithm. It denoises $\tilde{\mathbf{T}}$ by looking for the closest rank deficient matrix which is Toeplitz. First, we consider the Toeplitz matrix $\tilde{\mathbf{T}}$ with $L = M$ for effective denoising and perform a SVD on $\tilde{\mathbf{T}}$: $\tilde{\mathbf{T}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$. Then we truncate it to rank- K approximation \mathbf{T}' by forcing to zero the $M + 1 - K$ smallest singular values. \mathbf{T}' is no longer Toeplitz but its best Toeplitz approximation can be obtained by averaging the diagonals of \mathbf{T}' . A few of these iterations lead to a denoised set of moments.

The above method works under the assumption that \mathbf{T} is well approximated by the K signal-subspace singular vectors of $\tilde{\mathbf{T}}$ and that discarding information in the orthogonal-subspace of $\tilde{\mathbf{T}}$ removes certain amount of noise. However, when the noise on τ_m is very high we may experience a subspace swap. Specifically, at high noise level $\tilde{\mathbf{T}}$ is better approximated by some orthogonal-subspace singular vectors rather than by the weakest signal-subspace singular vector. In this case, the truncation of the $M + 1 - K$ smallest singular values of $\tilde{\mathbf{T}}$ in the first iteration of Cadzow will lead to an unrecoverable removal of information which belongs to the signal-subspace. Consequently we are not able to retrieve the locations by TLS and it is widely recognised that this is when performance

breakdown is observed. Moreover, we conjecture that as long as the subspace swap does not happen in the first iteration, the following iterations would separate the subspaces correctly. Hence the analysis on the SVD of the original noisy $\tilde{\mathbf{T}}$ is sufficient for predicting when subspace swap event happens and is given in the next section.

3.2 Necessary condition for subspace swap event

In this section we are going to work out at which noise level the orthogonal-subspace singular vectors substitute the position of the weakest signal-space singular vector in SVD of the noisy data matrix $\tilde{\mathbf{T}}$ with $L = M$.

We first look at the noiseless $\mathbf{T} \in \mathbb{C}^{(M+1) \times (M+1)}$ and we rewrite it as follows:

$$\mathbf{T} = (M+1)\mathbf{G} \begin{bmatrix} a_1 & & \\ & \ddots & \\ & & a_K \end{bmatrix} \mathbf{G}^H, \quad (3.7)$$

where

$$\mathbf{G} = \frac{1}{\sqrt{M+1}} \begin{bmatrix} \mathbf{g}(t_0) & \dots & \mathbf{g}(t_{K-1}) \end{bmatrix} \quad (3.8)$$

and

$$\mathbf{g}(t) = \begin{bmatrix} e^{\beta t/T} & e^{2\beta t/T} & \dots & e^{(M+1)\beta t/T} \end{bmatrix}^\top. \quad (3.9)$$

\mathbf{G} is Vandermonde and has full rank K since the locations t_k are distinct, hence in the noiseless case \mathbf{T} has rank K . SVD of \mathbf{T} can be written as $\mathbf{T} = \mathbf{U}\boldsymbol{\Sigma}_s\mathbf{V}^H$, where $\boldsymbol{\Sigma}_s = \text{diag}(s_1, s_2, \dots, s_K, \underbrace{0, \dots, 0}_{M+1-K})$, $s_1 \geq s_2 \geq \dots \geq s_K > 0$. The unitary matrix \mathbf{U} has size $(M+1) \times (M+1)$ and can be partitioned into $\mathbf{U}_s = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_K)$ and $\mathbf{U}_0 = (\mathbf{u}_{K+1}, \mathbf{u}_{K+2}, \dots, \mathbf{u}_{M+1})$. The matrix \mathbf{U}_s generates the signal-subspace and the orthogonal-subspace covered by \mathbf{U}_0 completes the space. The subspace \mathbf{U}_0 is used to identify the locations of Diracs as discussed in the previous section.

When there is noise, the new moments $\tilde{\tau}_m$ are noisy as described in (3.6), which can also be written in matrix form as follows:

$$\begin{aligned} \tilde{\boldsymbol{\tau}} &= \mathbf{C}\mathbf{y} + \mathbf{C}\boldsymbol{\epsilon} \\ &= \boldsymbol{\tau} + \mathbf{b}, \end{aligned} \quad (3.10)$$

where \mathbf{C} is composed of entries $c_{m,n} = c_{m,0} e^{(m-M)\beta n}$ at position (m, n) . Note that the statistics of the noise \mathbf{b} depends directly on the distribution of the sample noise $\boldsymbol{\epsilon}$ and on the coefficients \mathbf{C} . A stable matrix \mathbf{C} is therefore desired. As discussed in Sec. 2.3.1, the

most stable \mathbf{C} has condition number one and this can be achieved by choosing $\beta = j\frac{2\pi}{N}$ and the kernel is an eMOMS whose corresponding $|c_{m,0}| = 1/\sqrt{N}$ for $m = 0, 1, \dots, 2M$. Since by construction $\mathbf{C}^H\mathbf{C} = \mathbf{I}$ we have that the covariance matrix of the moment noise equals that of the sample noise: $\mathbf{R}_{\mathbf{b}} = \mathbf{R}_{\boldsymbol{\epsilon}} = \sigma_y^2\mathbf{I}$. Note that for β other than $j\frac{2\pi}{N}$, $\mathbf{R}_{\mathbf{b}}$ is not exactly diagonal, but we have experimental evidence that the derivation in the following part still approximately applies.

The matrix \mathbf{T} becomes $\tilde{\mathbf{T}} = \mathbf{T} + \mathbf{N}$, where $\mathbf{N} \in \mathbb{C}^{(M+1) \times (M+1)}$ is a Toeplitz matrix built with $\{b_m\}_{m=0}^{2M}$ in (3.6). $\tilde{\mathbf{T}}$ has full rank and will not lie entirely within the signal-subspace. When noise is strong, $\tilde{\mathbf{T}}$ will lie far from the signal-subspace and at certain point the subspace swap happens.

Now we describe a worst subspace swap scenario and show that this gives us the necessary condition for a swap event to happen. When this condition is not satisfied, absence of subspace-swap is guaranteed and the standard deviation of the retrieved locations can be predicted by Cramér-Rao bound.

We denote the singular values of \mathbf{N} with $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{M+1}$. The worst situation is when the strongest noise component with strength λ_1 is aligned with one of the elements in \mathbf{U}_0 , and the second strongest noise component with strength λ_2 is aligned with the weakest signal component with strength s_K and sums destructively with it making it become even weaker ($s_K - \lambda_2$). Hence subspace swap happens when

$$\lambda_1 > s_K - \lambda_2. \quad (3.11)$$

This is a *necessary condition* since any other swap event requires stronger conditions on the amplitude of noise singular values.

It is known that the maximum singular value of a $n \times n$ random symmetric Toeplitz matrix whose entries come from a sequence of i.i.d random Gaussian variables with variance σ^2 is smaller than $\sqrt{2\sigma^2 n \ln n}$ [53]. Given this, we can relate the maximum singular value of \mathbf{N} to the order $2M$ of the sampling kernel and the noise level as follows:

$$\lambda_1 < \sqrt{2\sigma_y^2(M+1) \ln(M+1)}. \quad (3.12)$$

Now the least favourable scenario is when both λ_1 and λ_2 equals the maximum: $\lambda_2 = \lambda_1 = \sqrt{2\sigma_y^2(M+1) \ln(M+1)}$ which yields the following necessary condition for a subspace swap:

$$\begin{aligned} \lambda_1 > s_K - \lambda_2 &\iff \lambda_1 > s_K/2 \\ &\iff \sqrt{2\sigma_y^2(M+1) \ln(M+1)} > s_K/2. \end{aligned} \quad (3.13)$$

This can be further derived to:

$$\sigma_y^2 > \frac{s_K^2}{8(M+1)\ln(M+1)}. \quad (3.14)$$

K = 1 :

When we have 1 Dirac with amplitude a , then $s_K = s_1 = |a|(M+1)$. We define $\text{PSNR} = 10 \log_{10} \frac{a^2}{\sigma_y^2}$. Hence we can relate breakdown PSNR to the order $P+1 = 2M+1$ of the kernel as follows:

$$\text{PSNR} < 10 \log_{10} \frac{8 \ln(M+1)}{(M+1)}. \quad (3.15)$$

The condition in (3.15) is shown in Fig. 3.1 by solid curve. Note that (3.15) is a necessary condition for breakdown, hence breakdown may happen for settings in the area below the blue curve and it is *guaranteed* no-breakdown will happen for the area above the curve.

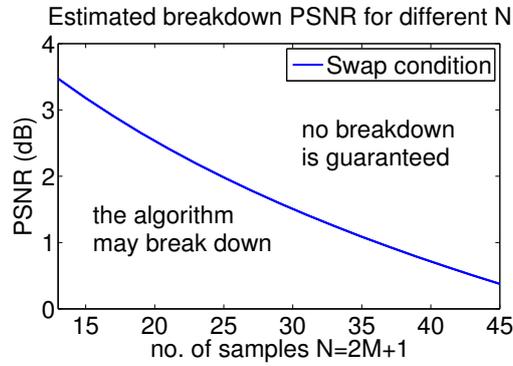


Figure 3.1: Necessary condition for subspace swap in the case of single Dirac for different number of samples N .

K = 2 :

When there are 2 Diracs and we assume both Diracs are with same amplitude a , (3.7) can be written as $\mathbf{T} = a(M+1)\mathbf{G}\mathbf{G}^H$. Its singular values s_1 and s_2 , i.e. the non-zero eigenvalues of \mathbf{T} , can be derived easily by computing the eigenvalues of the two-by-two matrix $a(M+1)\mathbf{G}^H\mathbf{G}$. The expressions are as follows:

$$\begin{aligned} s_1 &= |a|(M+1 + |\langle \mathbf{g}(t_0), \mathbf{g}(t_1) \rangle|) \\ \text{and } s_2 &= |a|(M+1 - |\langle \mathbf{g}(t_0), \mathbf{g}(t_1) \rangle|). \end{aligned} \quad (3.16)$$

Further manipulations gives

$$\begin{aligned} s_2 &= |a|(M+1 - |\langle \mathbf{g}(t_0), \mathbf{g}(t_1) \rangle|) \\ &= |a| \left(M+1 - \left| \frac{\sin(\frac{\beta}{2}(M+1)\Delta t/T)}{\sin(\frac{\beta}{2}\Delta t/T)} \right| \right), \end{aligned} \quad (3.17)$$

where $\Delta t = t_1 - t_0$. By substituting (3.17) to (3.14) and defining PSNR with $10 \log_{10} \frac{a^2}{\sigma_y^2}$, we can relate the breakdown PSNR to the sampling period T , the distance Δt between the two Diracs, the frequency interval β and the order of the kernel $P + 1 = 2M + 1$:

$$\text{PSNR} < 10 \log_{10} \frac{8(M + 1) \ln(M + 1)}{\left(M + 1 - \left| \frac{\sin(\frac{\beta}{2}(M+1)\Delta t/T)}{\sin(\frac{\beta}{2}\Delta t/T)} \right| \right)^2}. \quad (3.18)$$

The threshold PSNR in (3.18) for different number of samples N is illustrated by blue curves in Fig. 3.2, where we use the fact that we set $T = 1/N$.

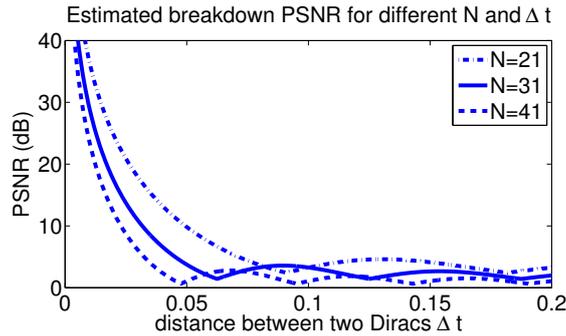


Figure 3.2: Necessary condition for subspace swap in the case of 2 Diracs for different number of samples N .

We can observe that when the relative distance $\Delta t/T$ increases, the level of noise we can handle increases too. Once the PSNR has reached a minimum, it then oscillates near the minimum PSNR. Interestingly, when we use our favourable setting $P + 1 = N$ (i.e. $2M + 1 = N$), then the first local minimum is exactly at $\Delta t = 2/(N + 1)$. This implies that if we guarantee $N + 1$ to be larger than $2/\Delta t$ we can achieve in general the most robust estimation result in the sense of low breakdown PSNR.

When there are more than 2 Diracs, the breakdown point given by subspace swap condition for the case of $K = 2$ still approximately applies when picking the two closest Diracs.

3.3 Simulations

In this section, we verify our proposed no-breakdown condition by comparing it to the empirical breakdown points obtained by the method as in Fig. 2.6. We show that the estimation algorithm¹ breaks down at a PSNR just below our predicted no-breakdown PSNR, confirming the reliability of our derivation. For simplicity, we only show the

¹Note that Cadzow iterative algorithm and matrix pencil lead to very similar performances.

verification result when we use the most robust sampling kernel as mentioned in the previous section, i.e. a kernel of order $P + 1$ (i.e. $2M + 1$) which reproduces $P + 1 = N$ exponentials $\{e^{m'\beta t/T}\}_{m'=-M}^M$ with $\beta = j\frac{2\pi}{N}$ and its exponential reproducing coefficient $|c_{m,0}| = 1/\sqrt{N}$, $m = 0, 1, \dots, 2M$. We emphasize that for other settings our calculated necessary condition for subspace swap (3.15) and (3.18) still accurately anticipate the breakdown event.

3.3.1 $K = 1$

The guaranteed no-breakdown PSNR (3.15) for different kernel order $P + 1 = 2M + 1$ is shown by the blue curve in Fig. 3.3. The measured breakpoints shown by red markers in general agree with our proposed condition.

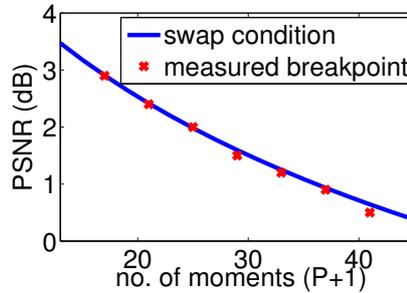


Figure 3.3: Subspace swap condition compared to the empirical breakdown points for $K = 1$ with the settings $N = P + 1$, $T = 1/N$ and $\beta = j2\pi/N$.

3.3.2 $K = 2$

The guaranteed no-breakdown PSNR (3.18) for different number of samples N and Diracs interval Δt is shown by the blue curves in Fig. 3.4. We can see that the observed breakdown points, which are shown by red markers, are well predicted by our proposed condition.

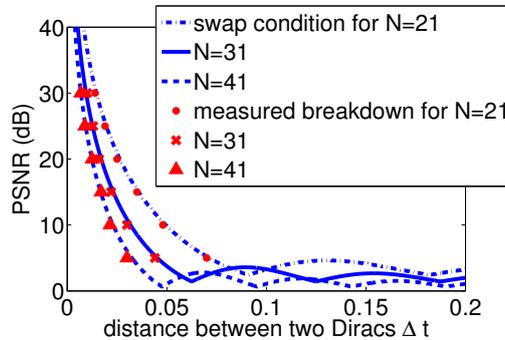


Figure 3.4: Subspace swap condition compared to the empirical breakdown points for $K = 2$ with the settings $N = P + 1$, $T = 1/N$ and $\beta = j2\pi/N$.

3.3.3 $K > 2$

In this simulation, we show that when there are more than $K = 2$ Diracs, the breakpoint given by subspace swap condition for $K = 2$ still approximately applies. For example, from Fig. 3.4 we see that the smallest possible breakdown PSNR for $K = 2$ and $N = 41$ is roughly 0dB and is achieved when $\Delta t \geq 2/(N + 1)$. We now demonstrate that the guideline is reliable even for $K > 2$. We show in Fig. 3.5 that we accurately reconstruct 17 Diracs from $N = 41$ samples in noise of PSNR = 0dB, where the distance between every two Diracs is greater than $2/(N + 1)$.

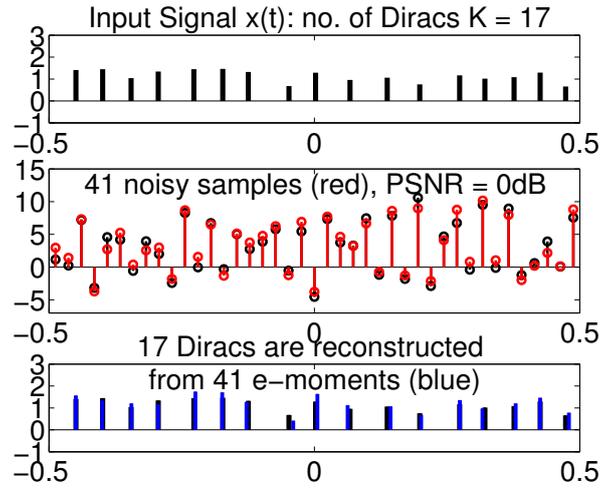


Figure 3.5: Reconstruction of 17 Diracs with minimum Dirac separation $\Delta t \geq 2/(N + 1)$ from 41 samples with PSNR = 0dB.

3.4 Summary

In this chapter, we have studied when FRI estimation algorithms break down in the presence of noise using the subspace swap criterion. We have then derived the breakdown region and related it to the sampling rate, the minimum distance between two Diracs and PSNR. The reliability of our predicted breakdown region has then been confirmed by simulation results.

Chapter 4

Universal sampling of signals with finite rate of innovation

It has been shown that specific classes of non-bandlimited signals known as signals with finite rate of innovation (FRI) can be perfectly reconstructed by using appropriate sampling kernels and reconstruction schemes. This exact FRI framework was later extended to an approximate FRI framework that works with any kernel.

Reconstruction is achieved by recovering all the parameters in the parametric model of the incoming signal, hence it is essential to know the model order (the rate of innovation) to ensure recovery. In view of this, we devise an algorithm for identifying the rate of innovation in order to extend the current sampling scheme to a universal one which enables sampling signals with arbitrary FRI using any acquisition device. Our proposed algorithm can effectively identify the rate of innovation prior to the signal reconstruction using arbitrary kernels and in different noise levels where we show that it achieves perfect identification result in noiseless scenario and achieves 95% correct identification rate even at SNRs as low as 5dB. This identification rate is achieved under the assumption that the algorithm is operating under the guaranteed performance conditions discussed in the previous chapter. This means that a minimum distance condition on the closest Diracs linked to the SNR must be imposed.

Recall that in the exact reproduction framework introduced in Section 2.2, perfect reconstruction of certain classes of FRI signals requires the acquisition device $\varphi(t)$ to behave like an exponential reproducing function and its order $P + 1$ must be equal or larger than the rate of innovation of the signal with FRI. Specifically, in the example of recovering a stream of K Diracs:

$$x(t) = \sum_{k=1}^K a_k \delta(t - t_k), \quad (4.1)$$

from its samples $y_n = \langle x(t), \varphi(t/T - n) \rangle, n = 1, \dots, N$, a condition on the order of the kernel $P + 1 \geq 2K$ must be satisfied. This means that an acquisition device may be no longer usable and perfect reconstruction is no longer possible when the rate of innovation of the incoming signal exceeds $2K$, and this even if we increase the sampling rate. In contrast, the approximate framework introduced in Section 2.4 allows us to use any arbitrary sampling kernel and N samples can give us N approximate exponentials for reconstruction. This property directly relates the highest rate of innovation it can be recovered to the sampling rate rather than the order of the kernel. Hence any acquisition device is always usable for signals with arbitrary rate of innovation below the sampling rate.

4.1 Identification of the rate of innovation

The approximate FRI framework, which allows us to sample FRI signals using any sampling kernel, together with the algorithm we are going to propose for identifying the rate of innovation, enable us to extend the current sampling scheme to a universal one which can recover signals with arbitrarily unknown finite rate of innovation using any sampling kernel.

The general idea behind our algorithm is as follows. Given cN ($c > 1$) samples of the input stream of Diracs y_n , we are able to obtain cN approximated Fourier coefficients $X(\alpha_m), m = 1, \dots, cN$. From these coefficients we estimate at most $N/2$ Diracs. Note that theoretically N samples is enough for recovering $N/2$ Diracs, but in reality we require a slightly higher number of samples per unit time since the Fourier coefficients are all approximated.

We first assume that the number of the Diracs is $p = 1$ and we retrieve the location and amplitude of the Dirac in the parametric model $\sum_{k=1}^p a_k \delta(t - t_k)$. Next we resynthesize the samples to obtain $\tilde{y}_n(p)$ and compute $\epsilon_p = \|\tilde{y}_n(p) - y_n\|$ the error between the resynthesized samples and y_n . Then we repeat this procedure but with assumption that p is 2, 3 up to $N/2$.

We expect that the error on the samples will first decrease gradually when the number of Diracs p we assumed approaches the true number K and will eventually reach nearly zero when p is exactly the number of the Diracs. When we further increase p , the errors will either rise slightly or further decrease with a much slower rate. In either case, the turning point can be recognised from the second derivative of the error. Once the number of Diracs K is known, the input signal $x(t)$ can be recovered using the parametric model with correct order.

We summarize the algorithm as follows:

Algorithm 1: Reconstruction of a stream of unknown number of Diracs

Data: cN samples $y_n = \langle x(t), \varphi(t - n) \rangle$ **Result:** Estimation of the number of Diracs K and corresponding reconstruction of the Diracs $\hat{x}(t)$

- 1 Obtain cN Fourier coefficients $X(\alpha_m)$ from $\{y_n\}_{n=1}^{cN}$;
 - 2 **for** Assumed number of Diracs $p = 1 \dots N/2$ **do**
 - 3 Estimate location(s) \hat{t}_k and amplitude(s) \hat{a}_k of p Diracs from y_n (with Cadzow method or matrix pencil);
 - 4 Resynthesize the samples $\tilde{y}_n(p) = \langle \sum_{k=1}^p \hat{a}_k \delta(t - \hat{t}_k), \varphi(t - n) \rangle$;
 - 5 Compute the error $\epsilon_p = \|\tilde{y}_n(p) - y_n\|$;
 - 6 **end**
 - 7 Compute second derivative ϵ_p'' of the error function interpolated from $\{\epsilon_p\}_{p=1}^P$;
 - 8 Choose for K the number of Diracs p corresponding to the largest ϵ_p'' . Then $\hat{x}(t)$ is the reconstructed stream of Diracs corresponding to the model $\sum_{k=1}^p \hat{a}_k \delta(t - \hat{t}_k)$.
-

4.1.1 Discussions

In this section, we use numerical examples to show that the second derivative is indeed a good indicator for identifying the number of Diracs. We also discuss the limitations of the current algorithm due to the instability of the second derivative method in the noisy scenario and to the fact that the guaranteed recovery conditions of the previous section may not be satisfied.

Universal sampling in the absence of noise

Assume we have a stream of K Diracs with unknown K and we take cN samples with a B-spline of order 5 following the scheme in Fig. 2.1. In the exact framework, this specific acquisition device restricts the number of Diracs we can reconstruct to 3. Thanks to the approximate Strang-Fix framework, with a B-spline of order 5 we can build cN approximated Fourier coefficients which allows us to reconstruct 1 up to $N/2$ Diracs. In Fig. 4.1(a,b) we try to identify the number $K = 31$ of Diracs given the $cN = 99$ samples. In Fig. 4.1(b), we can clearly see a sudden drop in the error function ϵ_p and which position exactly corresponds to the the correct number of Diracs, therefore second derivative at that point is the largest and the number of Diracs can be easily identified. Then all the 31 Diracs are almost perfectly reconstructed in the absence of noise as shown in Fig. 4.1(a).

In Fig. 4.1(c,d), we also highlight the universality of the sampling scheme since even if the number of Diracs of the incoming signal changes from $K = 31$ to $K = 21$, reliable reconstruction can still be achieved without the need to change the set-up.

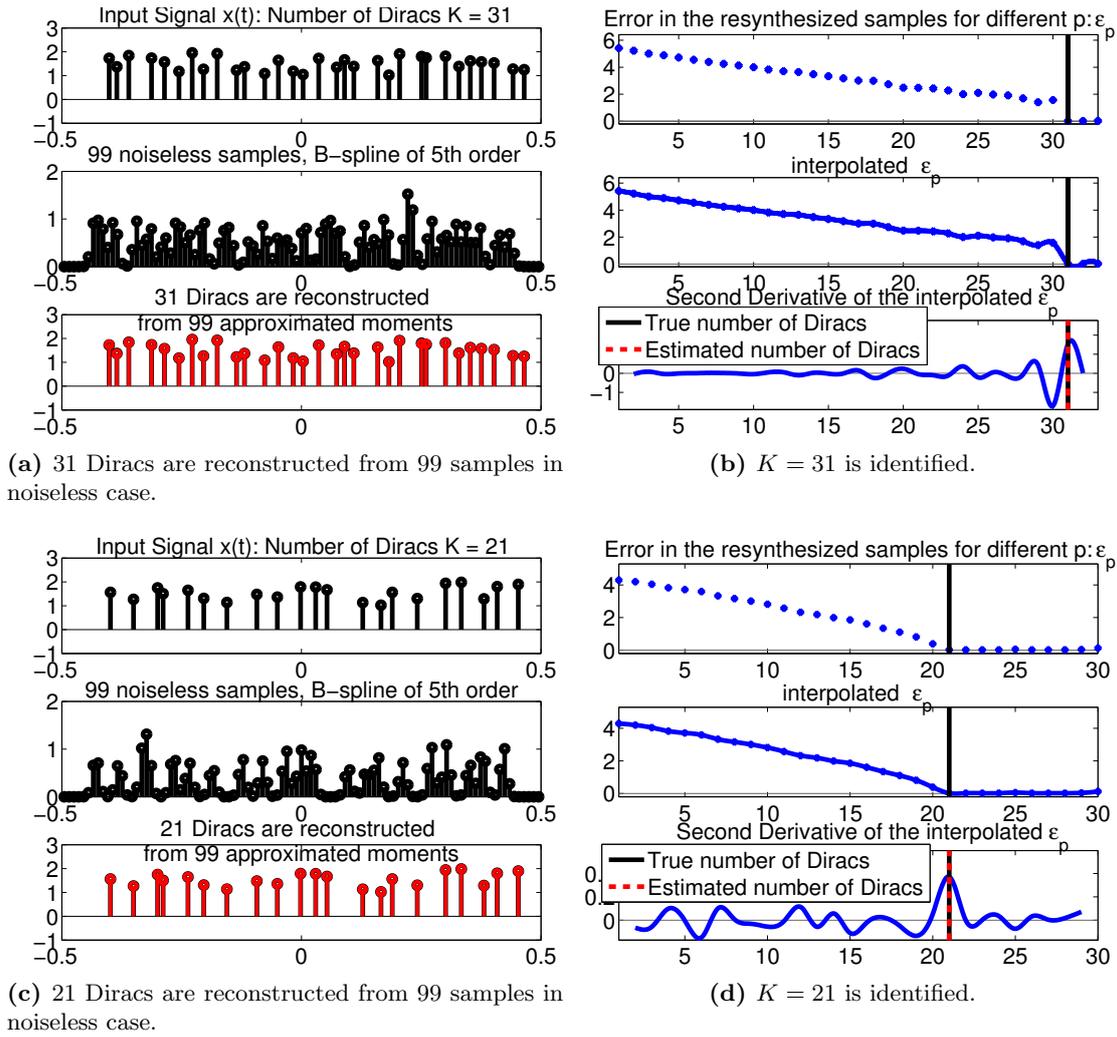


Figure 4.1: Universal sampling of a stream of unknown number of Diracs using B-spline kernel of order 5 in the absence of noise. (b)(d) The number of Diracs is identified from second derivative of the error function. (a)(c) All the Diracs are retrieved almost perfectly.

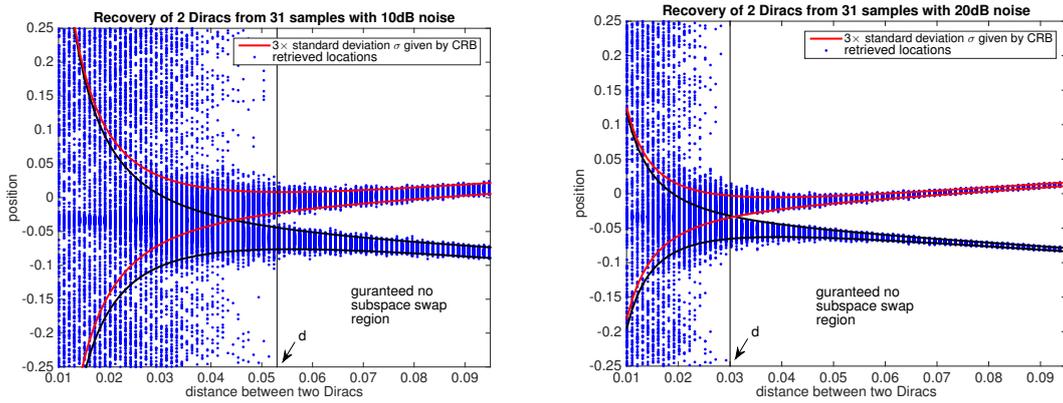
Universal sampling in the presence of noise

In the following we consider the noisy scenario. Here the noise is added to the samples y_n and is white Gaussian noise of variance σ^2 , chosen according to the target signal-to-noise ratio defined as $\text{SNR}(\text{dB}) = 10 \log \frac{\|y\|^2}{N\sigma^2}$.

Not that our proposed algorithm is based on the canonical FRI reconstruction algorithm, which may break down when the necessary condition for subspace swap is satisfied. As a result the performance of our proposed method would also be limited by the subspace swap condition and would not achieve flawless result as shown in noiseless scenario.

In the following, we first briefly discuss the limitation of the FRI location reconstruction

algorithm and then we demonstrate its effects on our proposed universal reconstruction algorithm. As presented in Sec. 2.3.3, we can assess the performance of the location reconstruction algorithm by comparing it to the theoretical best performance provided by the Cramér-Rao bound. In Fig. 4.2 we show the estimated locations and the Cramér-Rao bounds for the situation where there are two Diracs with same amplitude sampled at the rate $1/T = 31$ in noise levels of SNR=10dB and 20dB. It shows that the location reconstruction algorithm in general achieves the Cramér-Rao bounds for distances d beyond some specific critical values.



(a) performance of FRI reconstruction algorithm when SNR=10dB

(b) performance of FRI reconstruction algorithm when SNR=20dB

Figure 4.2: Scatterplot of locations retrieved using standard FRI reconstruction algorithm (Cadzow+Prony) compared to 3 times the standard deviation given by Cramér-Rao lower bounds. (a) SNR=10dB (b) SNR=20dB

For distances d smaller than the critical values, these two Diracs are indistinguishable. More accurately, as we analysed in Chapter 3, the inability of FRI algorithms to achieve the Cramér-Rao bounds for d smaller than specific value can be explained by subspace swap, i.e. the noise is so high that the signal-subspace singular vector with smallest amplitude of \mathbf{A} in (2.19) for reconstruction is replaced by some noise-subspace singular vectors. In this situation, the FRI reconstruction algorithm reconstructs them as one tall Dirac situated in between the true Diracs and one Dirac far away from the true Diracs with negligible amplitude. Consequently, our algorithm for identifying the number of Diracs, which is based on the FRI reconstruction algorithm, will neglect the one with negligible amplitude and identify only one Dirac.

The results in Fig. 4.3 confirm that the performance of our identification algorithm is affected when the FRI location reconstruction algorithm breaks down.

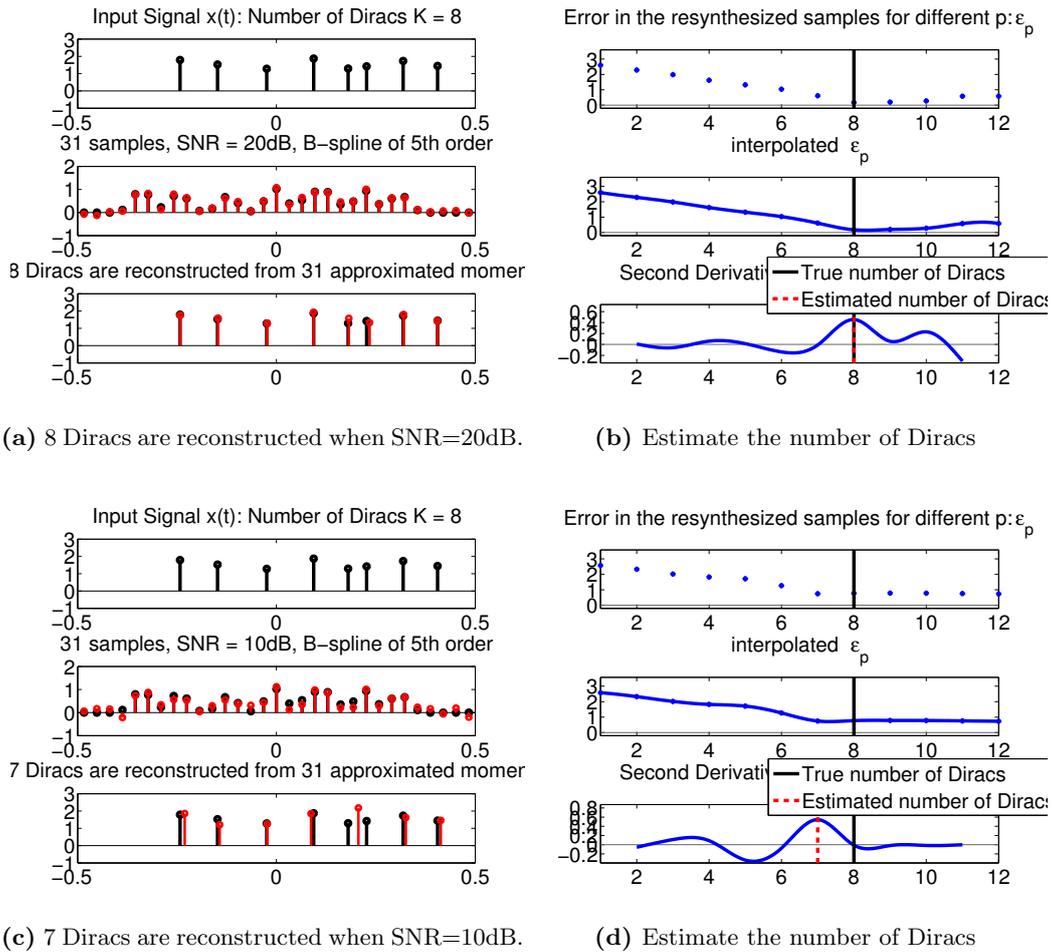


Figure 4.3: Universal sampling of a stream of unknown number of Diracs using B-spline kernel of order 5 in the presence of noise. Two close Diracs in the 8 Diracs are recognised as one Dirac when SNR=10dB and the others are accurately retrieved.

In this example, there are 8 Diracs sampled with 5th order B-spline at rate $1/T = 31$, in which two Diracs are close to each other and the others are well separated. We notice that when SNR=20dB the number of Diracs is identified correctly and all the 8 Diracs are accurately reconstructed. When SNR=10dB, however, the algorithm recognises $K = 7$ in total since the two close Diracs, whose distance, 0.046, is smaller than the guaranteed no-breakdown distance 0.05 (see Fig. 4.2a) are recognised as one Dirac. In such situations, that is, when subspace swap occurs, the misidentification of the number of Diracs is inevitable and is not due to the proposed identification algorithm. Nevertheless, we highlight that this can be easily solved by increasing the sampling rate.

We have just discussed the misidentification of number of Diracs due to subspace swap event. Another reason for misidentification in the presence of noise is that the error ϵ_p

has more variations, and as a result the second derivative method is less effective in noisy scenario. More specifically, the second derivative method is able to correctly identify the number of Diracs when the rate of decay of the error ϵ_p at the correct number K is the highest compared to that of other possible values of K . However this is not always true when the noise level is moderate to high. For example, there are situations where the rate of decay of the error is slow and identification based only on second derivative is then affected by noise.

4.2 Robust identification algorithm

The previous section has shown the use of second derivatives to identify the number of Diracs, and this method resulted in the publication [18]. Here we propose a more robust and faster scoring system to determine the number of Diracs given the noisy samples. This new algorithm is different from Algorithm 1 in the following ways. First, while the previous version of algorithm tests all possible number of Diracs, this new version would immediately stop when the resynthesis error increased, i.e. when $\epsilon_p - \epsilon_{p-1} > 0$ is detected. We denote the stopping number by K_{stop} . Moreover, instead of deciding the number of Diracs by looking for the highest second derivative of resynthesis error ϵ_p , we propose a scoring system that gives scores to all the numbers $p = 1, \dots, K_{\text{stop}}$ based on resynthesis error ϵ_p on samples, first derivative of ϵ_p and second derivative of ϵ_p . More specifically, the score S_p for $p = 1, \dots, K_{\text{stop}}$ is as follows:

$$S_p = \epsilon_p'' - 2|\epsilon_p'| - 2\epsilon_p - (\epsilon_1 - \epsilon_2)p. \quad (4.2)$$

The term ϵ_p'' in S_p rewards p which is a sudden turning point in ϵ_p , and $-2|\epsilon_p'|$ penalizes p where the speed of decreasing in the resynthesis error is comparatively high. $-\epsilon_p$ penalizes p with large resynthesis error. Finally, $-(\epsilon_p + (\epsilon_1 - \epsilon_2)p)$ penalizes large p that cannot reach the expected decreasing rate of the resynthesis error. The estimated number of Diracs with highest score will be our chosen number of Diracs. Note that this scoring system decides the number of Diracs by choosing the p with highest possibility among all instead of an absolute numerical criterion, it works regardless of the number of Diracs, signal to noise ratio, amplitude of Diracs, etc. We summarize the algorithm in Algorithm 2.

4.2.1 Simulations

In this part, we want to test the performance of the proposed algorithm. We want to show that our algorithm can identify the number of Diracs when the sampling setting is

Algorithm 2: Reconstruction of a stream of unknown number of Diracs**Data:** cN samples $y_n = \langle x(t), \varphi(t - n) \rangle$ **Result:** Estimation of the number of Diracs K and corresponding reconstruction of the stream of Diracs $\hat{x}(t)$

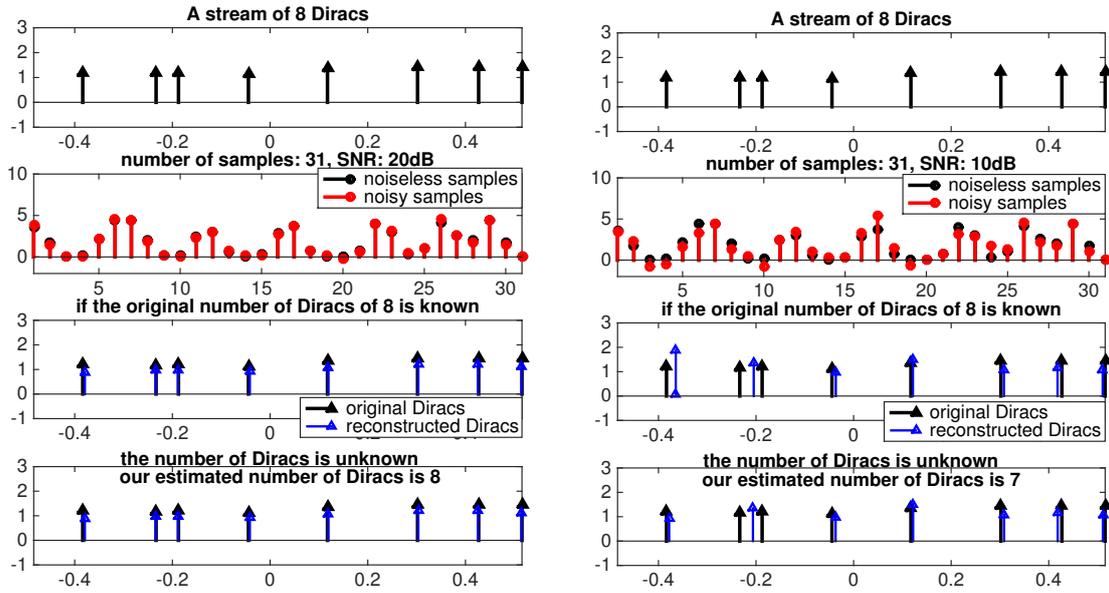
-
- 1 Obtain cN Fourier coefficients $X(\alpha_m)$ from $\{y_n\}_{n=1}^{cN}$;
 - 2 $p = 0$;
 - 3 **while** *assumed number of Diracs* $p < N/2$ **and** *errDecreaseDetector* **do**
 - 4 $p = p + 1$;
 - 5 Estimate location(s) \hat{t}_k and amplitude(s) \hat{a}_k of p Diracs from y_n (with Cadzow method or matrix pencil);
 - 6 Resynthesize the samples $\tilde{y}_n(p) = \langle \sum_{k=1}^p \hat{a}_k \delta(t - \hat{t}_k), \varphi(t - n) \rangle$;
 - 7 Compute the error $\epsilon_p = \|\tilde{y}_n(p) - y_n\|$;
 - 8 **if** $p > 1$ **then**
 - 9 $\text{errDecreaseDetector} = (\epsilon_p - \epsilon_{p-1}) \leq 0$;
 - 10 $K_{\text{stop}} = p$;
 - 11 **end**
 - 12 **end**
 - 13 Compute the scores $S_p = \epsilon_p'' - 2|\epsilon_p'| - 2\epsilon_p - (\epsilon_1 - \epsilon_2)p$ for $p = 1, \dots, K_{\text{stop}}$;
 - 14 Choose for K the number of Diracs p corresponding to the largest S_p . Then $\hat{x}(t)$ is the reconstructed stream of Diracs corresponding to the model $\sum_{k=1}^p \hat{a}_k \delta(t - \hat{t}_k)$.
-

in the no-breakdown region, and when it is not the algorithm is able to choose the most reasonable number of Diracs.

In the example of Fig. 4.4, there are 8 Diracs sampled at $1/T = 31$, in which two close-by Diracs have a distance satisfying the 20dB no-breakdown condition but not the 10dB no-breakdown condition. We notice that when SNR=20dB the number of Diracs is identified correctly and all the 8 Diracs are accurately reconstructed. When SNR=10dB subspace swap happens, FRI reconstruction methods are not able to resolve the two Diracs no matter what the number of Diracs we assumes, but $K = 7$ results in a better reconstruction. Our proposed algorithm makes the right selection of the number of Diracs when there is subspace swap event.

Now we do a more systematic test where we generate a stream of 4 Diracs with random locations and random amplitudes between 1 to 1.5, satisfying the minimum distance required by the no-breakdown condition. Note that the results in Chapter 3 are a bit optimistic when compared to the current setting since the kernel in Chapter 3 reproduces exponential exactly whereas the kernels in this chapter achieves approximate reproduction only. In this setting, we estimate the breakdown distance empirically, for example, by looking at breakdown distance d in Fig. 4.2.

We use our method to reconstruct the unknown number of Diracs from 51 samples taken with the kernel of 5th order B-spline. There are 10000 realisations for each noise level.



(a) SNR=20dB, no subspace swap happens and we identified the number of Diracs correctly.

(b) SNR=10dB, two close-by Diracs are not resolvable when subspace swap happens. The best choice is 7 Diracs instead of 8 and we identified it.

Figure 4.4: Universal sampling of a stream of unknown number of Diracs using 5th order B-spline in the presence of noise.

The percentage of correct estimations for different level of noise is shown in Table 4.1. The result is compared with our previous version of identification algorithm, i.e. Algorithm 1, which is slower and only based on the second derivative of resynthesis errors.

Table 4.1: The percentage of correct identification of the number (4) of Diracs from 51 samples in different noise level. 10000 realisations for each noise level.

noise level (SNR)	no noise	20dB	15dB	10dB	5dB
identif. % of Alg. 1	100%	99.92%	97.38%	89.32%	63.55%
identif. % of Alg. 2	100%	100%	100%	99.99%	95.72%

Both versions of our proposed algorithms achieve flawless identification in the absence of noise. In the presence of noise, the old version is still able to identify the model order in medium to high SNRs and the updated one is more robust since it works well in all levels of noise. We can observe a decreasing identification rate when the noise level increases. This is not surprising because for high noise levels, the turning point of the resynthesis error is more shallow, and as a result the scores for different points are similar and this situation leads to misidentification.

Note that as long as the model order is identified correctly, the FRI reconstruction algo-

rithm will then achieve the best possible result indicated by the Cramér-Rao bounds. So we can draw the conclusion that in the presence of noise, our proposed algorithm is able to robustly recover streams of any unknown number of Diracs and its performance in general achieves the Cramér-Rao bounds, with only very low failure rate.

4.3 Summary

In this chapter we have shown how to sample FRI signals with arbitrary kernels and that a novel algorithm can identify the model order accurately prior to reconstruction. Simulation results have confirmed the effectiveness of the proposed method.

Chapter 5

Simultaneous estimation of sparse linear systems and sparse signals using FRI principles

In Sec. 2.2.3 we have shown that with FRI sampling theory, we are able to obtain exact Fourier coefficients of specific classes of FRI signals and that by using this information we are able to achieve perfect reconstruction at low sampling rate. There are classes of linear systems that can also be fully specified by a small number of parameters (“parametrically sparse” system) and are fully characterized by their Fourier samples. We expect that these systems can be identified from samples at a low rate instead of the Nyquist sampling rate as in current system identification scheme. Therefore in this chapter, we explore the possibility of extending the framework for sampling FRI signals to achieve perfect system identification at a low rate.

5.1 Introduction

A classical set-up for identifying systems that are fully characterized by finite number of parameters is shown in Fig. 5.1. In the first channel, the discrete Fourier transform of an excitation signal $x(t)$ is obtained. In the second channel, $x(t)$ is first fed to an unknown system with impulse response $\psi(t)$ then its discrete Fourier transform is obtained. The aim of system identification is to determine $\psi(t)$ from the two set of Fourier data at specific frequencies.

This chapter is a joint work with Hojjat Akhondi Asl and Thierry Blu.

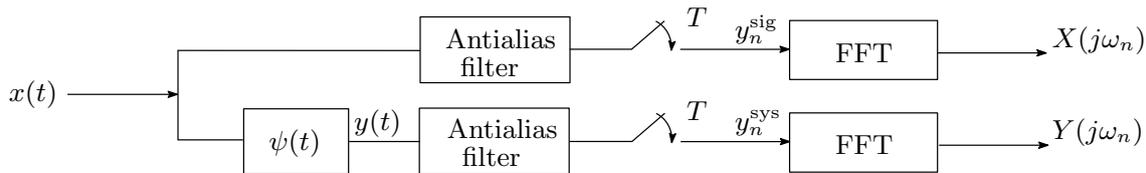


Figure 5.1: A typical system identification set-up. Adapted from [1].

Since the input signals $x(t)$ and the convolution of $x(t)$ and $\psi(t)$ are normally non-bandlimited, the exact Fourier transform of these signals can never be computed and a faithful approximation is achieved only when anti-aliasing filters are used and the requirement of Nyquist sampling rate is satisfied [1, 54]. So perfect identification of an unknown system is in general not possible.

However, by replacing the anti-aliasing filter in Fig. 5.1 with exponential reproducing kernels, we show that when both the input signals and the system are of compact support it is possible to retrieve the exact continuous Fourier transform of $\psi(t)$ at specific frequencies and hence to perfectly reconstruct the unknown system.

The fact that our approach is universal in that it uses the classical set-up with anti-aliasing filter replaced with an exponential reproducing kernel but can be used with any finite $x(t)$ at low sampling rate makes it valuable when considering real-life applications, for example, a system identification problem where the users prefer to apply random excitation signal for psychological or technological reasons [55], or the situation where the maximum sampling rate (or the sampling duration) is restricted by various environmental conditions [56].

We then focus our attention on linear circuits and while their impulse responses have infinite support, our set-up is still able to yield high quality discrete spectral data and rather good identification results. When the samples are corrupted with noise, we propose a modified Cadzow iterative algorithm which is proved to be effective for denoising the set of spectral data which significantly improve the identification accuracy.

We finally show (Theorem 5.2, Section 5.4) that even when we have no access to the samples of the input signal and only have y_n^{sys} we can still perfectly reconstruct the input signal and the unknown system $\psi(t)$ if the input signal is a certain type of FRI signal and the response system is of compact support. We also propose a recursive algorithm that reliably identify the unknown input signal and the system in the presence of noise.

We note that system and channel estimation from low-rate samples has already been considered in [57–60], but the systems considered there are in discrete-time and are different from the continuous-time systems studied here.

The outline of this chapter is as follows. In Section 5.2 we revisit the theory of finite rate

of innovation that will assist us developing methods for system identification in the rest of the content. In Section 5.3, we present our sampling set-up for system identification and discuss the method to obtain the Fourier data for system identification in the frequency domain. We also propose the modified Cadzow iterative algorithm to handle additive noise introduced by the acquisition device. In Section 5.4, we show perfect identification of both the unknown input signal and the system is possible and then propose a recursive algorithm to simultaneously estimate them in the presence of noise. In Section 5.5, we show simulation results on the nearly perfect reconstruction of an unknown system when there is no noise in the Fourier data, and also on the data denoising performance using the modified Cadzow iterative algorithm. Moreover, we show the possibility of simultaneous estimation of the unknown signal and system. Finally we conclude in Section 5.6.

5.2 Overview of FRI theory

Given that a system is fully specified by finite free parameters (“parametrically sparse” system), the system identification problem is nothing but estimating the unknown free parameters with the prior knowledge of the system’s model. The difficulties in identifying unknown systems is similar to that in reconstruction FRI signals, which are raised from the fact that real-life signals are usually non-bandlimited and the impossibility of getting an ideal low pass filter. In fact, perfect reconstruction of FRI signals and perfect identification of “parametrically sparse” system are impossible by classical sampling theory and in practice they require extremely high sampling rate. However, FRI theory makes sampling non-bandlimited signal at low rate possible. It also provides three ideas for handling the system identification problem:

First, if we sample using the setup of Fig. 2.1 a compactly supported signal $x(t)$ with an exponential reproducing kernel $\varphi(t)$ satisfying:

$$\sum_{n \in \mathbb{Z}} c_{m,n} \varphi(t/T - n) = e^{\alpha_m t/T} \quad \text{with } m = 0, 1, \dots, M, \quad (5.1)$$

then from its samples $y_n = \langle x(t), \varphi(t/T - n) \rangle$, $n = 0, 1, \dots, N - 1$ we would be able to compute the e-moments of $x(t)$ by:

$$\tau_m = \sum_{n \in \mathbb{Z}} c_{m,n} y_n \quad (5.2)$$

$$= \int_{-\infty}^{\infty} x(t) e^{\alpha_m t/T} dt, \quad m = 0, 1, \dots, M. \quad (5.3)$$

Note that when α_m is purely imaginary and is given by $\alpha_m = -j\omega T$, the e-moments

are exactly the *continuous-time Fourier transform* of $x(t)$ at $\omega = \omega_m, m = 0, 1, \dots, M$, denoted by $X(j\omega_m)$.

Furthermore, we can perfectly reconstruct classes of FRI signals, for example, stream of Diracs, piecewise polynomials, which are all fully characterized by its Fourier data, using annihilating filter method which is described in detail in Section 2.2.3 and 2.2.3.

Lastly, it is possible to use Cadzow—a structured low-rank approximation method to denoise the Fourier data if the matrix composed by this data is rank deficient and structured (e.g. Toeplitz) before being corrupted by noise. Recall that the Cadzow denoises the noisy matrix by iteratively and alternatively finding the rank-deficient approximation and closest Toeplitz approximation of the noisy matrix.

5.3 Identification of parametrically sparse systems

In this section, we show how the sampling scheme explained can be employed to estimate an unknown system $\psi(t)$. The sampling set-up for addressing the system identification problem is shown in Fig. 5.2. This is similar to the classical set-up in Fig. 5.1 except the anti-aliasing filter is replaced with an exponential reproducing kernel, which is of compact support. The general idea of identifying $\psi(t)$ is similar to reconstructing FRI signals: we

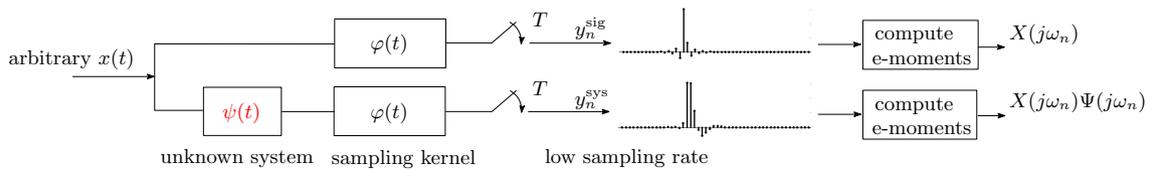


Figure 5.2: FRI sampling set-up for system identification. Here $x(t)$ is the input signal, $\psi(t)$ is the unknown system and $\varphi(t)$ is the exponential reproducing kernel.

first compute the Fourier coefficients of the unknown system $\psi(t)$ using samples taken with an exponential reproducing kernel $\varphi(t)$, then we estimate the free parameters in $\psi(t)$'s frequency-domain model. The details of how to obtain the Fourier coefficients of $\psi(t)$ are described as follows:

First, we compute the $M + 1$ exponential moments of the signal $x(t)$ as described in (5.3) yielding (refer to channel 1 in Fig. 5.2):

$$\tau_m^{\text{sig}} = X(j\omega_m), \quad m = 0, \dots, M. \quad (5.4)$$

Next, we compute the $M + 1$ exponential moments of $x(t) * \psi(t)$ as follows (refer to channel 2 in Fig. 5.2):

$$\begin{aligned}
 \tau_m^{\text{sys}} &= \sum_n c_{m,n} y_n^{\text{sys}} \\
 &= \left\langle x(t) * \psi(t), \sum_n c_{m,n} \varphi(t/T - n) \right\rangle \\
 &= \int_{-\infty}^{\infty} [x(t) * \psi(t)] e^{-j\omega_m t} dt \\
 &= X(j\omega_m) \cdot \Psi(j\omega_m), \quad m = 0, \dots, M.
 \end{aligned} \tag{5.5}$$

where we have used the fact that convolution in the time domain corresponds to multiplication in Fourier domain.

Then $\Psi(j\omega_m)$ can be obtained by dividing τ_m^{sys} by τ_m^{sig} :

$$\begin{aligned}
 \frac{\tau_m^{\text{sys}}}{\tau_m^{\text{sig}}} &= \frac{X(j\omega_m) \cdot \Psi(j\omega_m)}{X(j\omega_m)} \\
 &= \Psi(j\omega_m), \quad m = 0, \dots, M.
 \end{aligned} \tag{5.6}$$

Here we assume $X(j\omega_m) \neq 0$.

The above derivation leads to an important fact: we can retrieve some Fourier coefficients $\Psi(j\omega_0), \Psi(j\omega_1), \dots, \Psi(j\omega_M)$ of the unknown system $\psi(t)$ from the samples taken with an exponential reproducing kernel of order $M + 1$, and this is independent of the choice of input signal. Fig. 5.3 shows that we are able to retrieve accurate Fourier data of a system with transfer function $\frac{10}{s^2+4s+10}$ by taking 50 samples at sampling rate of 4Hz through an exponential reproducing kernel. Note that when both the input signal $x(t)$ and the system response $x(t) * \psi(t)$ are of compact support, the Fourier coefficients $\{X(j\omega_m)\}_{m=0}^M$ and $\{X(j\omega_m) \cdot \Psi(j\omega_m)\}_{m=0}^M$ we retrieve are exact.

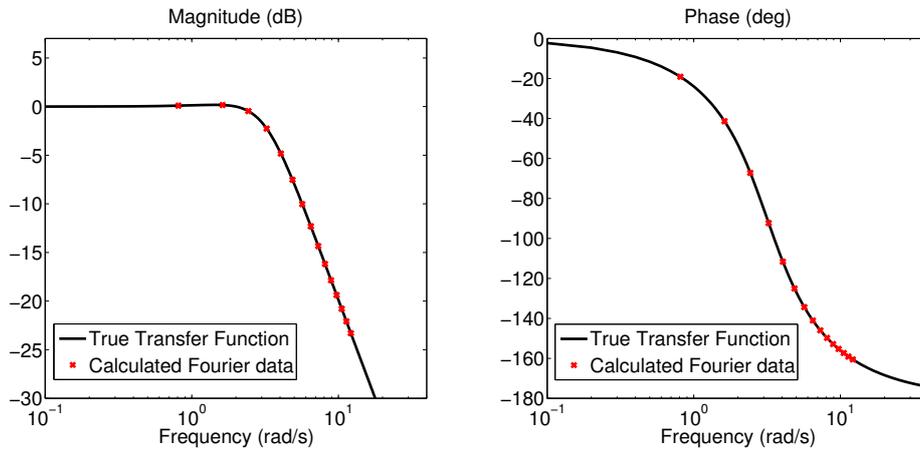


Figure 5.3: True Transfer Function: $\frac{10}{s^2+4s+10}$ and the Fourier data obtained by taking 50 samples at sampling rate $f = 4$ Hz through an exponential reproducing kernel.

In what follows we show that this partial knowledge of the spectrum together with the Fourier transform function model of the unknown system is in some cases sufficient to solve for the unknown parameters in the model. In other words, a specific system $\psi(t)$ will be completely determined by $\Psi(j\omega_0), \Psi(j\omega_1), \dots, \Psi(j\omega_M)$, where the frequency samples need not to be uniform.

5.3.1 Exact identification of E-spline

We have shown we are able to obtain Fourier coefficients of a system $\psi(t)$ with our sampling set-up, any $\psi(t)$ that is fully characterized by its Fourier coefficients can be reconstructed. For example, streams of Diracs, E-splines, etc. In particular when $\psi(t)$ is of compact support, we are able to obtain its exact Fourier coefficients and it is possible to establish a one-to-one mapping between y_n^{sys} and $\psi(t)$. Now we demonstrate how to perfectly identify an E-spline system $\psi(t)$.

Assume $\psi(t)$ is an E-spline of known order P . Its Fourier transform is:

$$B_\gamma(j\omega) = \prod_{p=0}^{P-1} \frac{1 - e^{\gamma_p - j\omega}}{j\omega - \gamma_p}, \quad (5.7)$$

where $\gamma = (\gamma_0, \gamma_1, \dots, \gamma_{P-1})$ are the unknown parameters to be retrieved. Based on the observation that dividing τ_m^{sys} by τ_m^{sig} yields the exact Fourier transform of $\psi(t)$ at ω_m , we are able to estimate the unknown parameters γ using the following approach:

We have the relationship:

$$\Psi(j\omega_m) = \prod_{p=0}^{P-1} \frac{1 - e^{\gamma_p - j\omega_m}}{j\omega_m - \gamma_p}, \quad m = 0, 1, \dots, M. \quad (5.8)$$

By rearranging the equation we get:

$$\Psi(j\omega_m) \underbrace{\prod_{p=0}^{P-1} (j\omega_m - \gamma_p)}_{\text{polynomial } p(j\omega_m)} = \underbrace{\prod_{p=0}^{P-1} (1 - e^{\gamma_p - j\omega_m})}_{\text{polynomial } q(e^{-j\omega_m})} \quad (5.9)$$

Then by expanding the two products of sums we have the following:

$$\Psi(j\omega_m) \sum_{p=0}^P a_p (j\omega_m)^p = \sum_{p=0}^P b_p (e^{-j\omega_m})^p. \quad (5.10)$$

The $M+1$ equations above form a linear system with unknown parameters $\sigma = (a_0, \dots, a_P,$

b_0, b_1, \dots, b_P):

$$\begin{aligned}
& \begin{bmatrix} \Psi(j\omega_0) & 0 & \dots & 0 \\ 0 & \Psi(j\omega_1) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Psi(j\omega_M) \end{bmatrix} \begin{bmatrix} 1 & j\omega_0 & \dots & (j\omega_0)^P \\ 1 & j\omega_1 & \dots & (j\omega_1)^P \\ \vdots & \vdots & \vdots & \vdots \\ 1 & j\omega_M & \dots & (j\omega_M)^P \end{bmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_P \end{pmatrix} \\
&= \begin{bmatrix} 1 & e^{-j\omega_0} & \dots & (e^{-j\omega_0})^P \\ 1 & e^{-j\omega_1} & \dots & (e^{-j\omega_1})^P \\ \vdots & \vdots & \vdots & \vdots \\ 1 & e^{-j\omega_M} & \dots & (e^{-j\omega_M})^P \end{bmatrix} \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_P \end{pmatrix}.
\end{aligned} \tag{5.11}$$

Since $a_P = 1$ and $b_0 = 1$, all the unknown coefficients of the polynomials p and q , ($a_0, \dots, a_{P-1}, b_1, \dots, b_P$), can be obtained by solving the equivalent system (5.12).

$$\begin{aligned}
& \underbrace{\begin{bmatrix} \Psi(j\omega_0) & \Psi(j\omega_0)j\omega_0 & \dots & \Psi(j\omega_0)(j\omega_0)^{P-1} & -e^{-j\omega_0} & \dots & -(e^{-j\omega_0})^P \\ \Psi(j\omega_1) & \Psi(j\omega_1)j\omega_1 & \dots & \Psi(j\omega_1)(j\omega_1)^{P-1} & -e^{-j\omega_1} & \dots & -(e^{-j\omega_1})^P \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \Psi(j\omega_M) & \Psi(j\omega_M)j\omega_M & \dots & \Psi(j\omega_M)(j\omega_M)^{P-1} & -e^{-j\omega_M} & \dots & -(e^{-j\omega_M})^P \end{bmatrix}}_{\text{size}(M+1) \times (2P)} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{P-1} \\ b_1 \\ \vdots \\ b_P \end{pmatrix} \\
&= \begin{pmatrix} 1 - \Psi(j\omega_0)(j\omega_0)^P \\ 1 - \Psi(j\omega_1)(j\omega_1)^P \\ \vdots \\ 1 - \Psi(j\omega_M)(j\omega_M)^P \end{pmatrix}.
\end{aligned} \tag{5.12}$$

Since we need $M + 1 \geq 2P$ to solve this system, we require the sampling kernel to be able to reproduce exponentials $\{e^{\alpha_m t/T}\}_{m=0}^M$ with $M \geq 2P - 1$. Once all the polynomial coefficients have been found, the unknown parameters $\{\gamma_p\}_{p=0}^P$ can be retrieved by looking for the roots of either polynomial. As a result, the E-spline system is fully determined. The above steps show that it is possible to identify an unknown E-spline system of order P from its samples. Hence we have the following result.

Theorem 5.1 *Assume a sampling kernel $\varphi(t)$ that can reproduce exponentials $e^{\alpha_m t}$ with $m = 0, 1, \dots, M$. An E-spline system $\psi(t)$ of order P , is uniquely defined by the samples $y_n^{\text{sig}} = \langle x(t), \varphi(t/T - n) \rangle$ together with the samples $y_n^{\text{sys}} = \langle x(t) * \psi(t), \varphi(t/T - n) \rangle$ if $M \geq 2P - 1$ and $x(t)$ is of compact support.*

5.3.2 Identification of linear systems with rational Fourier transform

Linear systems with rational Fourier transform can be characterised through their transfer functions in the frequency domain and are entirely determined by their Fourier coefficients. In the previous subsection, the possibility of perfect identification of an unknown E-spline system is partly due to the fact that the E-spline function $\psi(t)$ is of compact support. When a system is of compact support, the weighted sum $\sum_n c_{m,n} y_n^{\text{sys}}$ yields $X(j\omega_m) \cdot \Psi(j\omega_m)$ exactly. When the transient response of the system is infinite, we can still reasonably extend the results to this case. Since we can only access the truncated version of $\psi(t)$ and of $x(t) * \psi(t)$ in the case of linear circuits, $\sum_n c_{m,n} y_n^{\text{sys}}$ is an approximation of $X(j\omega_m) \cdot \Psi(j\omega_m)$. Nevertheless, assuming enough samples are taken, that is, the number of samples is large enough to retain most of the energy of the impulse response, the error in this approximation is negligible. This is later verified by the simulation results in Table 5.1.

Notice that a P -th order LTI system can be characterized by the transfer function:

$$\Psi(s) = \frac{\sum_{q=0}^Q b_q s^q}{\sum_{p=0}^P a_p s^p}. \quad (5.13)$$

Taking $s = j\omega_m$, we have $M + 1$ equations:

$$\underbrace{\Psi(j\omega_m) \sum_{p=0}^P a_p (j\omega_m)^p}_{\text{polynomial } p(j\omega_m)} = \underbrace{\sum_{q=0}^Q b_q (j\omega_m)^q}_{\text{polynomial } q(j\omega_m)}, \quad m = 0, 1, \dots, M. \quad (5.14)$$

Therefore, a linear system with unknowns $\boldsymbol{\theta} = (a_0, a_1, \dots, a_P, b_0, b_1, \dots, b_Q)$ can be constructed:

$$\underbrace{\begin{bmatrix} \Psi(j\omega_0) & 0 & \dots & 0 \\ 0 & \Psi(j\omega_1) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Psi(j\omega_M) \end{bmatrix}}_{\boldsymbol{\Delta}_{M+1}} \underbrace{\begin{bmatrix} 1 & j\omega_0 & \dots & (j\omega_0)^P \\ 1 & j\omega_1 & \dots & (j\omega_1)^P \\ \vdots & \vdots & \vdots & \vdots \\ 1 & j\omega_M & \dots & (j\omega_M)^P \end{bmatrix}}_{\mathbf{V}_{P+1}} \underbrace{\begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_P \end{pmatrix}}_{\mathbf{p}} = \underbrace{\begin{bmatrix} 1 & j\omega_0 & \dots & (j\omega_0)^Q \\ 1 & j\omega_1 & \dots & (j\omega_1)^Q \\ \vdots & \vdots & \vdots & \vdots \\ 1 & j\omega_M & \dots & (j\omega_M)^Q \end{bmatrix}}_{\mathbf{W}_{Q+1}} \underbrace{\begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_Q \end{pmatrix}}_{\mathbf{q}}. \quad (5.15)$$

From (5.15) we can directly solve for all the unknown parameters to determine the transfer function in much the same way as in the case of identifying an E-spline system from (5.11).

5.3.3 Identification of linear systems with rational Fourier transform in noisy scenario

In practical situations, the measurements are corrupted with noise: $\hat{y}_n = y_n + \epsilon_n$ and (5.15) is not satisfied exactly. We therefore propose a modified Cadzow algorithm to denoise the spectral data $\{\Psi(j\omega_m)\}_{m=0}^M$ of the unknown system before solving for the unknown parameters in (5.13). The method we propose exploits the prior knowledge of the system model (or the structure of the noiseless system matrix (5.15)) and yields denoised spectral data that well matches the model. The algorithm we propose is in spirit similar to the one in [61, 62].

First of all, we intend to better condition the system of equations in (5.15). Since \mathbf{W}_{Q+1} is an $(M+1) \times (Q+1)$ matrix with $M > Q$, we can find an orthonormal set of $M-Q$ vectors, ordered in an $(M+1) \times (M-Q)$ orthonormal matrix \mathbf{U}_{M-Q} such that $\mathbf{U}_{M-Q}^H \mathbf{W}_{Q+1} = \mathbf{0}$. Hence, by eliminating \mathbf{q} , (5.15) is equivalent to

$$\mathbf{U}_{M-Q}^H \mathbf{\Delta}_{M+1} \mathbf{V}_{P+1} \mathbf{p} = \mathbf{0}. \quad (5.16)$$

We can further perform the QR decomposition of $\mathbf{V}_{P+1} = \mathbf{Q}_{P+1} \mathbf{R}$ where \mathbf{Q}_{P+1} is an orthonormal $(M+1) \times (P+1)$ matrix leading to

$$\underbrace{\mathbf{U}_{M-Q}^H \mathbf{\Delta}_{M+1} \mathbf{Q}_{P+1}}_{\mathbf{A}_{P+1}} \mathbf{R} \mathbf{p} = \mathbf{0}. \quad (5.17)$$

Here, both \mathbf{U}_{M-Q} and \mathbf{Q}_{P+1} are orthonormal matrices which provide the stability in the computations that we were looking for. We note that for a transfer function with P poles, \mathbf{A}_{P+1} has size $(M-Q) \times (P+1)$ and is a rank-deficient matrix with rank P . We call \mathbf{A}_{P+1} the annihilation matrix since $\mathbf{A}_{P+1} \mathbf{R} \mathbf{p} = \mathbf{0}$.

Moreover, for the same unknown system of order P , we notice that if we extend its corresponding linear system of equations (5.14) to:

$$\Psi(j\omega_m) \cdot p_{\text{ext}}(j\omega_m) = q_{\text{ext}}(j\omega_m), \quad \text{for } m = 1, 2, \dots, M \quad (5.18)$$

as if we were looking for an unknown system of higher-order P_{ext} rather than P , the solution is no longer unique. The possible solutions are $p_{\text{ext}}(j\omega) = p(j\omega)r(j\omega)$ and $q_{\text{ext}}(j\omega) = q(j\omega)r(j\omega)$, where r is any arbitrary polynomial of degree equal to or smaller than $P_{\text{ext}} - P$. The solution space is thus of dimension $P_{\text{ext}} - P$. Therefore, the extension of the annihilation matrix in (5.17), denoted as $\mathbf{A}_{P_{\text{ext}}+1}$, is still of rank P if no noise is present. Note that the extension does not change the property of the annihilation matrix and working on it can accelerate the denoising process.

When there is noise, however, the annihilation matrix \mathbf{A} or its extension is full rank.

Given the rank-deficiency property of the annihilation matrix \mathbf{A} in the noiseless situation, we intend to recover the structured low-rank matrix from the noisy \mathbf{A} for the purpose of denoising. However, the canonical Cadzow algorithm is not directly applicable in this case since the matrix \mathbf{A} is not Toeplitz. Therefore, we propose a modified Cadzow iterative algorithm which enables denoising \mathbf{A} with the specific structure $\mathbf{U}^H \mathbf{\Delta} \mathbf{Q}$. More specifically, from \mathbf{A} we find a rank-deficient matrix that belongs to the matrix space spanned by the matrices $\mathbf{U}^H \mathbf{e}_k \mathbf{e}_k^T \mathbf{Q}$ for $k = 1, \dots, M + 1$, where \mathbf{e}_k is the canonical vector $(0, \dots, 1, 0, \dots, 0)^T$ with element 1 at index k . At each iteration of this algorithm, a low-rank approximation of the annihilation matrix, \mathbf{A}' , is obtained by keeping the P largest singular values. Notice that the low-rank approximation is no longer consistent with the annihilation matrix's structure $\mathbf{U}^H \mathbf{\Delta} \mathbf{Q}$, we therefore look for a closest matrix in the form of $\mathbf{U}^H \mathbf{\Delta}' \mathbf{Q}$ to the rank- P approximation \mathbf{A}' (see Appendix A for more details). This routine is iterated until the annihilation matrix \mathbf{A} becomes effectively of rank- P . The algorithm is summarized in the inset called ‘Iterative Denoising Algorithm’.

Algorithm 3: Iterative Denoising Algorithm

input : The moments $\Psi(j\omega_m)$
output: The denoised moments $\tilde{\Psi}(j\omega_m)$
1 repeat
2 From the moments $\Psi(j\omega_m)$, build the modified $(M - Q_{\text{ext}}) \times (P_{\text{ext}} + 1)$ annihilation matrix: $\mathbf{A}_{P_{\text{ext}}+1} = \mathbf{U}_{M-Q_{\text{ext}}}^H \mathbf{\Delta}_{M+1} \mathbf{Q}_{P_{\text{ext}}+1}$.

3 Perform the SVD of $\mathbf{A}_{P_{\text{ext}}+1}$: $\mathbf{A}_{P_{\text{ext}}+1} = \mathbf{U} \mathbf{S} \mathbf{V}^H$ and keep only the P largest singular values in $\mathbf{S} \rightarrow \mathbf{S}'$, to build a rank- P matrix $\mathbf{A}' = \mathbf{U} \mathbf{S}' \mathbf{V}^H$.

4 Find the closest matrix to \mathbf{A}' that belongs to the matrix space spanned by the matrices $\mathbf{U}_{M-Q_{\text{ext}}}^H \mathbf{e}_k \mathbf{e}_k^T \mathbf{Q}_{P_{\text{ext}}+1}$ for $k = 1, 2, \dots, M + 1$. This results in a “denoised” annihilation matrix $\tilde{\mathbf{A}}$ and its corresponding moments $\tilde{\Psi}(j\omega_m)$.

5 until $\mathbf{S}(P + 1, P + 1) < \delta$, where δ is a predefined threshold

The denoising performance is exemplified in Fig. 5.4.

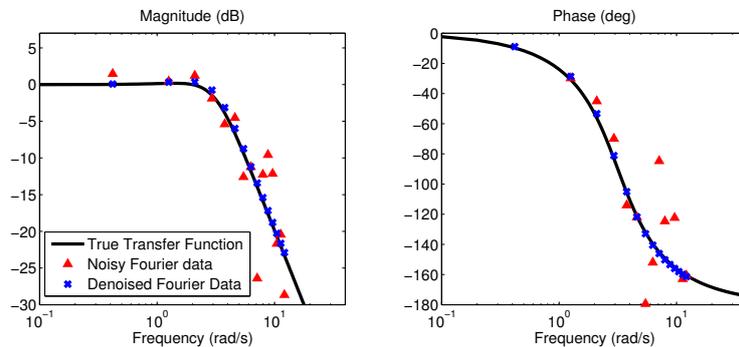


Figure 5.4: Modified Cadzow denoising on the spectral data of the system: $\Psi(s) = \frac{10}{s^2 + 4s + 10}$. SNR=10dB.

5.4 Simultaneous identification of input signals and systems

Consider now the scenario shown in Fig. 5.5 where both $x(t)$ and $\psi(t)$ need to be estimated.

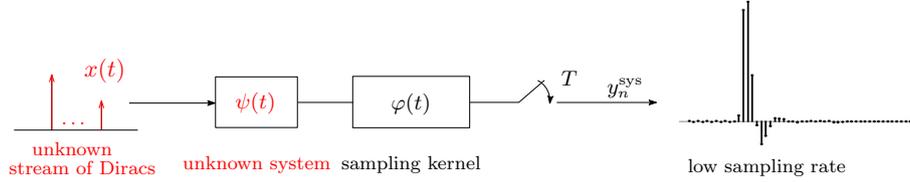


Figure 5.5: FRI sampling set-up for simultaneous estimation of the input signal and the linear system. Here $x(t)$ is an FRI signal, $\varphi(t)$ is an exponential reproducing kernel and $\psi(t)$ is a linear system of order P .

The problem is more difficult since both the input signal and the system are unknown and we only have access to the output samples y_n^{sys} . We can show we are able to use an extension of the annihilating filter for FRI signals (e.g. filter of (2.17) for streams of Diracs) to identify both the input signal $x(t)$ and the system $\psi(t)$ *perfectly* when $x(t)$ is an FRI signal and the response of the linear system $\psi(t)$ is of compact support. In particular, if $x(t)$ is a stream of K Diracs we can state the following:

Theorem 5.2 *Assume a sampling kernel $\varphi(t)$ that can reproduce exponentials $e^{(\alpha_0+m\lambda)t}$ with $m = 0, 1, \dots, M$. An E-spline system of order P , $\psi(t)$, and an unknown stream of input K Diracs, $x(t)$, are uniquely defined by the samples $y_n^{\text{sys}} = \langle x(t) * \psi(t), \varphi(t/T - n) \rangle$ if $M \geq 2K(P + 1)$.*

See Appendix B.1 for the proof. Here the exact identification is possible because the response of the system is of compact support and the Fourier coefficients we obtain are exact. When the system is a linear system characterized by a transfer function, there is still an annihilating filter method to solve for both the input signal and the system (see Appendix B.2 for more details). Remember that its response has infinite support, consequently the Fourier coefficients and the identification results are not exact. In such practical situations when the Fourier coefficients obtained are not exact due to for example the infinite response or noise, we propose a recursive version of the identification method in Section 5.3 to reliably estimate $x(t)$ and $\psi(t)$ simultaneously.

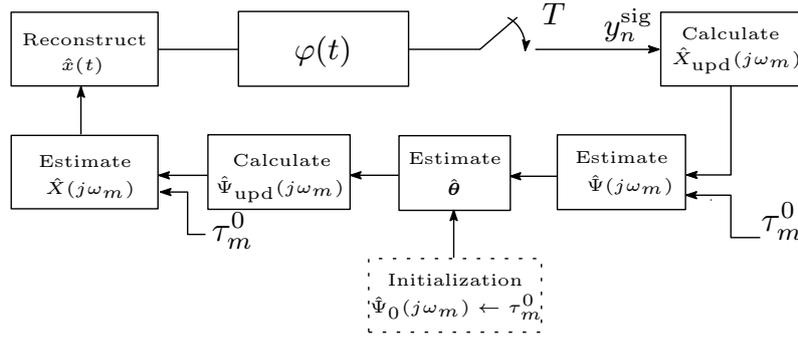


Figure 5.6: Schematic diagram of the algorithm that recursively estimate the input signal and the system.

The schematic diagram for the recursive algorithm is shown in Fig. 5.6. The moments can be calculated by $\tau_m^0 = \sum_n c_{m,n} y_n^{\text{sys}}$, and $\tau_m^0 = X(j\omega_m) \cdot \Psi(j\omega_m)$ as derived in (5.5). We first let the initial guess of the spectral data $\Psi(j\omega_m)$ of the unknown system be the moments τ_m^0 , and estimate the parameters in the transfer function from the spectral data $\Psi(j\omega_m)$. Next, by using the estimated parameters we are able to re-synthesize $\Psi(j\omega_m)$. Now we can calculate $X(j\omega_m)$ given the relationship $\tau_m^0 = X(j\omega_m) \cdot \Psi(j\omega_m)$. Since we know that the input signal $x(t)$ is a specific type of FRI signal, we are able to estimate it using the annihilating filter method from $X(j\omega_m)$ as described in Section 2.2. Given an estimation of the input signal, we can re-synthesize the samples y_n^{sig} and subsequently $X(j\omega_m)$, and then obtain $\Psi(j\omega_m)$ by using the relationship $\tau_m^0 = X(j\omega_m) \cdot \Psi(j\omega_m)$. By updating $X(j\omega_m)$ and $\Psi(j\omega_m)$ alternatively and iteratively, this process converges to an estimation of the input signal and the unknown system. The algorithm is summarized as follows:

Algorithm 4: Simultaneous estimation of the input signal and the system

input : the moments τ_m^0 , which are equivalent to $X(j\omega_m) \cdot \Psi(j\omega_m)$, $m = 0, 1, \dots, M$

output: the estimation of $x(t)$ and of the system parameters θ

- 1 set $\hat{\Psi}(j\omega_m) = \tau_m^0$ in order to get an initial $\hat{\theta}$
 - 2 **while** *iteration* < a finite number **do**
 - 3 estimate unknown parameters $\hat{\theta}$ of $\hat{\Psi}(j\omega_m)$ by solving (5.15) (apply the modified Cadzow denoising algorithm)
 - 4 calculate $\hat{\Psi}_{\text{upd}}(j\omega_m)$ from the estimated parameters $\hat{\theta}$
 - 5 estimate $\hat{X}(j\omega_m)$ from $\hat{X}(j\omega_m) = \frac{\tau_m^0}{\hat{\Psi}_{\text{upd}}(j\omega_m)}$
 - 6 calculate $\hat{x}(t)$ by applying FRI reconstruction methods on $\hat{X}(j\omega_m)$
 - 7 calculate $\hat{X}_{\text{upd}}(j\omega_m)$ by re-synthesizing y_n^{sys} from the estimated $\hat{x}(t)$
 - 8 estimate $\hat{\Psi}(j\omega_m)$ from $\hat{\Psi}(j\omega_m) = \frac{\tau_m^0}{\hat{X}_{\text{upd}}(j\omega_m)}$
 - 9 **end**
-

5.5 Simulations

5.5.1 System identification using arbitrary known input signal

In Section 5.3, we proved (Theorem 5.1) the possibility of *exact* reconstruction of a “parametrically sparse” system with compact support using any arbitrary finite-length input signal. Here we evaluate our proposed scheme on identifying linear circuits, whose response is infinite. We will first show the nearly error-free identification performance in the noiseless scenario to prove that the spectral data of linear circuits obtained by our proposed set-up is of extremely high accuracy. We also evaluate the performance of the modified Cadzow denoising algorithm we proposed and show that it yields reliable identification results in the presence of noise. The experimental set-up is shown in Fig. 5.7. The input signal is a pulse, and the sampling kernel is an exponential reproducing kernel.

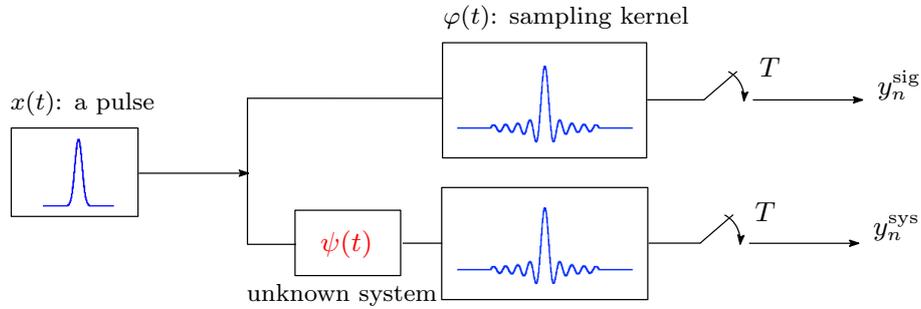


Figure 5.7: Sampling set-up for system identification. The input signal $x(t)$ is a pulse and the sampling kernel $\varphi(t)$ is an exponential reproducing kernel.

The system to be estimated are two Sallen-Key low pass filters (see Fig. 5.8) with transfer functions $\Psi(s) = \frac{1}{s^2+s+1}$ and $\Psi(s) = \frac{10}{s^2+4s+10}$, and a higher order system $\frac{s+2}{s^3+2s^2+2s+1}$. The sampling rates for the three unknown systems are respectively 1 Hz, 4 Hz and 2 Hz. The number of samples N are 50, 50 and 100 respectively.

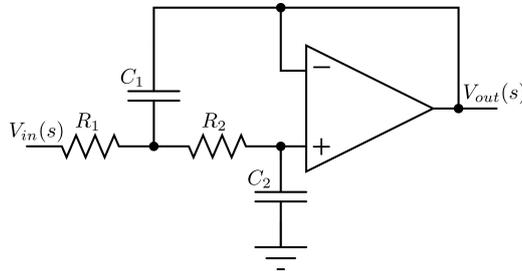


Figure 5.8: Sallen-Key filter topology. The transfer function of a Sallen-Key filter is $\Psi(s) = \frac{1}{s^2+s\left(\frac{1}{R_2C_1} + \frac{1}{R_1C_1}\right) + \frac{1}{R_1C_1R_2C_2}}$.

Noiseless samples

In this part, we assess the quality of the Fourier data obtained with our sampling scheme using noiseless measurements. We use our computed Fourier data to estimate the unknown parameters directly by solving the linear system of equations in (5.15).

As shown in Table 5.1, our estimated transfer functions are nearly error-free which confirms that our sampling method provides high-quality Fourier data even at low sampling rate and when $\psi(t)$ has infinite support. In comparison, when we input the low-rate samples to the widely-used MATLAB system identification toolbox which is based on prediction error method [63], it fails to identify the system with satisfactory accuracy.

Table 5.1: System identification in the absence of noise. $\varphi(t)$ is an E-spline and $x(t)$ is a pulse.

True Transfer Function	Method	Estimation (ideal set-up)
$\frac{1}{s^2+s+1}$	Our method	$\frac{1.000}{s^2+1.000s+1.000}$
	Toolbox	$\frac{2.00}{s^2+1.89s+1.99}$
$\frac{10}{s^2+4s+10}$	Our method	$\frac{10.000}{s^2+4.000s+10.000}$
	Toolbox	$\frac{20.46}{s^2+7.91s+20.32}$
$\frac{s+2}{s^3+2s^2+2s+1}$	Our method	$\frac{1.000s+2.000}{s^3+2.000s^2+2.000s+1.000}$
	Toolbox	$\frac{695.3s+3200}{s^3+1901s^2+2024s+1710}$

Noisy samples

In the following simulations, we show the robustness of our proposed sampling scheme and denoising scheme. We assume we have full control of the input signal. This means that we have access to the noiseless samples of the input signal y_n^{sig} . The system samples \hat{y}_n^{sys} are corrupted with additive white Gaussian noise of variance σ^2 . Signal-to-noise ratio (SNR) in our case is defined as:

$$\text{SNR(dB)} = 10 \log \frac{\|\mathbf{y}\|^2}{N\sigma^2}, \quad (5.19)$$

where N is the number of samples.

By comparing the identification results obtained by using the following two data sets, we evaluate the effectiveness of our proposed denoising scheme.

1. Raw spectral data $\hat{\Psi}(j\omega_m) = \frac{\sum_n c_{m,n} \hat{y}_n^{\text{sys}}}{\sum_n c_{m,n} y_n^{\text{sig}}}$.
2. Denoised spectral data $\tilde{\Psi}(j\omega_m)$ by performing our proposed Cadzow routine on $\hat{\Psi}(j\omega_m)$.

Table 5.2 shows the average estimation errors (200 realizations) on all the free parameters in different levels of noise.

The results in the table shows that the modified Cadzow algorithm can effectively denoise the spectral data, and the identification accuracy is satisfactory in low to medium level of noise. A higher order exponential reproducing kernel would give more Fourier coefficients and therefore further improve the identification accuracy.

Table 5.2: Parameter errors of identification of $\Psi(s) = \frac{1}{s^2+1s+1}$ and $\Psi(s) = \frac{10}{s^2+4s+10}$ under different levels of additive noise on the input samples and output samples.

True Transfer function: $\frac{b_0}{a_2s^2+a_1s+a_0} = \frac{1}{s^2+s+1}$		
Fourier coefficients used	SNR	Errors on parameters b_0, a_1, a_0
Noisy data	20dB	0.034 0.040 0.020
Denosed data		0.025 0.029 0.017
Noisy data	15dB	0.057 0.070 0.034
Denosed data		0.044 0.048 0.032
Noisy data	10dB	0.124 0.152 0.069
Denosed data		0.081 0.093 0.052
True Transfer function: $\frac{b_0}{a_2s^2+a_1s+a_0} = \frac{10}{s^2+4s+10}$		
Fourier coefficients used	SNR	Errors on parameters b_0, a_1, a_0
Noisy data	20dB	0.493 0.235 0.281
Denosed data		0.288 0.127 0.225
Noisy data	15dB	0.863 0.400 0.485
Denosed data		0.527 0.236 0.371
Noisy data	10dB	1.588 0.813 0.882
Denosed data		0.941 0.420 0.733
True Transfer function: $\frac{b_1s+b_0}{a_3s^3+a_2s^2+a_1s+a_0} = \frac{s+2}{s^3+2s^2+2+1}$		
Fourier coefficients used	SNR	Errors on parameters b_1, b_0, a_2, a_1, a_0
Noisy data	20dB	0.658 4.264 3.641 2.627 2.541
Denosed data		0.146 0.676 0.397 0.433 0.347
Noisy data	15dB	1.308 7.556 5.056 7.063 3.248
Denosed data		0.265 1.332 0.775 0.853 0.679
Noisy data	10dB	2.192 10.61 6.628 10.58 4.114
Denosed data		0.324 1.717 0.998 1.103 0.867

5.5.2 Simultaneous identification of input signals and systems

In Section 5.3, we showed identification of systems with arbitrary known input signals. Then in Section 5.4 we proved (Theorem 5.2) that when the input signal is an unknown FRI signal, we can identify both the input signal and the unknown system *perfectly*. We also proposed a more robust recursive algorithm. Here we evaluate the identification accuracy of the recursive algorithm. The experimental set-up is shown in Fig. 5.9.

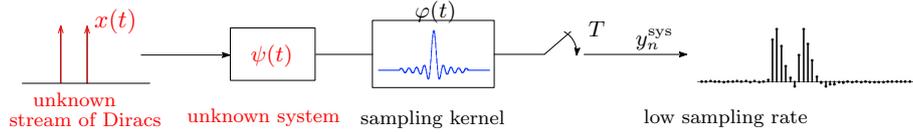


Figure 5.9: FRI sampling set-ups for recursive estimation. $\varphi(t)$ is an exponential reproducing kernel.

Noticing an ambiguity between the amplitude of the Dirac and the magnitude of the unknown system, we assume we know the amplitude of the input Dirac.

Noiseless samples

When there is no noise, the algorithm gives nearly exact estimations of both the unknown input signal and the unknown system within few iterations. The identification results are shown in Table 5.3.

Table 5.3: Simultaneous identification of the input signal and the system in the absence of noise. $x(t)$ is a stream of two Diracs. $\varphi(t)$ is an exponential reproducing kernel.

	Ground Truth	Estimation
Transfer Function of $\psi(t)$	$\frac{1}{s^2+s+1}$	$\frac{1.00}{s^2+1.00s+1.00}$
Locations of input Diracs $x(t)$	6.2344, 12.4844	6.2344, 12.4844
Transfer Function of $\psi(t)$	$\frac{10}{s^2+4s+10}$	$\frac{10.00}{s^2+4.00s+10.00}$
Locations of input Diracs $x(t)$	0.5039, 1.8711	0.5039, 1.8711
Transfer Function of $\psi(t)$	$\frac{s+2}{s^3+2s^2+2s+1}$	$\frac{1.00s+2.00}{s^3+2.00s^2+2.00s+1.00}$
Locations of input Diracs $x(t)$	8.9766	8.9766

Noisy samples

In this part we demonstrate three simultaneous estimation results when the samples y_n^{sys} are corrupted with noise of SNR=20dB. In Fig. 5.10 the input signal is a stream of two

Diracs of amplitudes 1 at locations $\{0.9875, 8.4875\}$ and the system to be estimated is $\frac{1}{s^2+s+1}$. In Fig. 5.11 the input signal is a stream of two Diracs of amplitudes 1 at locations $\{0.3086, 3.4336\}$ and the system to be estimated is $\frac{10}{s^2+4s+10}$. Moreover in Fig. 5.12, the input signal is one Dirac of amplitude 1 at location 8.9766 and the system to be estimated is $\frac{s+2}{s^3+2s^2+2s+1}$. For each simulation, we plot the estimation of the location(s) of the Dirac(s) after each iteration, and the Bode diagram of the estimated system after 10 iterations. From the results we see that both the location(s) of the Dirac(s) and the unknown systems are estimated accurately.

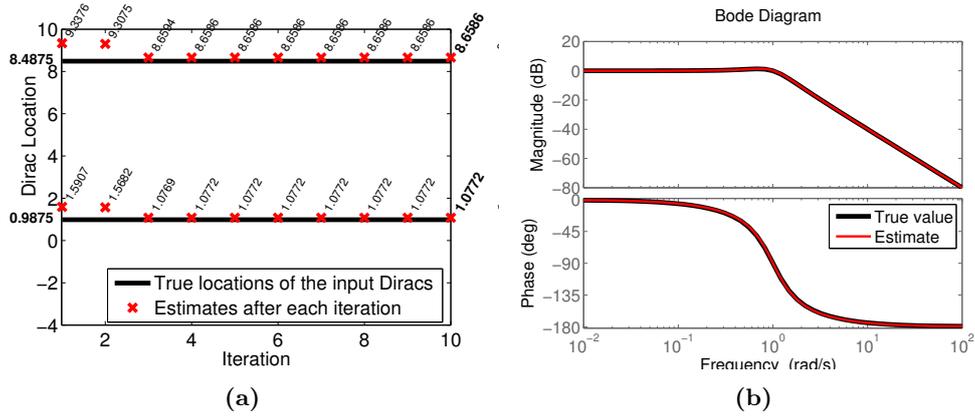


Figure 5.10: Simultaneous estimation of both the input signal and the system in the presence of noise. SNR=20dB. (a) The true locations of the input Diracs: $\{0.9875, 8.4875\}$ and the estimations of the locations of the input Diracs after each iteration. (b) Bode diagrams of the true system $\frac{1}{s^2+s+1}$ and the estimated system $\frac{1.000}{s^2+0.979s+1.003}$.

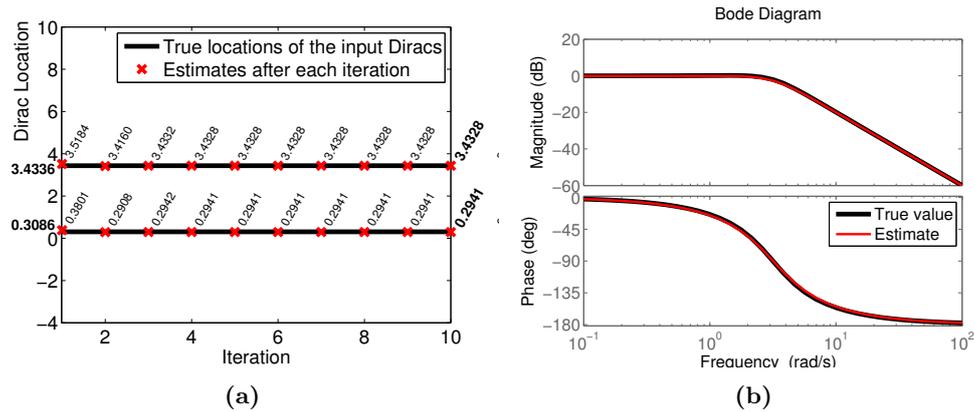


Figure 5.11: Simultaneous estimation of both the input signal and the system in the presence of noise. SNR=20dB. (a) The true locations of the input Diracs: $\{0.3086, 3.4336\}$ and the estimations of the locations of the input Diracs after each iteration. (b) Bode diagrams of the true system $\frac{10}{s^2+4s+10}$ and the estimated system $\frac{9.956}{s^2+4.272s+10.055}$.

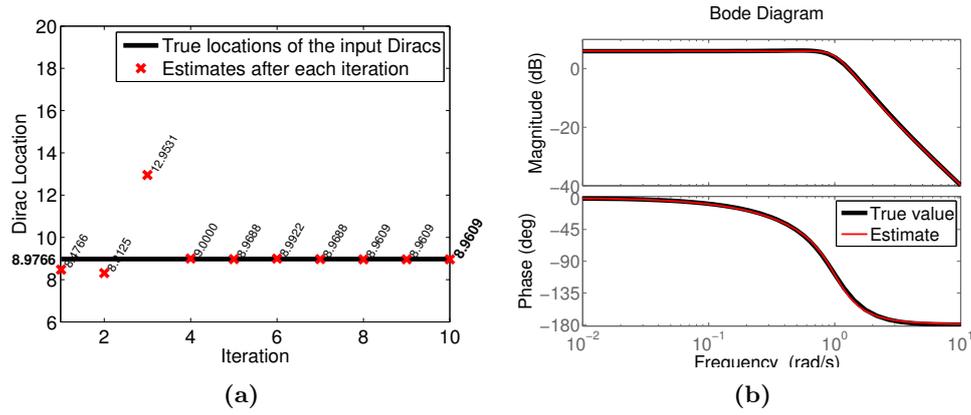


Figure 5.12: Simultaneous estimation of both the input signal and the system in the presence of noise. SNR=20dB. (a) The true locations of the input Diracs: $\{8.9766\}$ and the estimations of the locations of the input Diracs after each iteration. (b) Bode diagrams of the true system $\frac{s+2}{s^3+2s^2+2s+1}$ and the estimated system $\frac{1.01s+1.35}{s^3+1.52s^2+1.58s+0.67}$.

5.6 Summary

In this chapter we have used the FRI sampling framework to develop new methods for system identification. We have shown that exact identification of specific classes of sparse system is possible when using exponential reproducing kernels. We have then extended this approach to the estimation of linear circuits. We also proposed a corresponding denoising technique when the samples are corrupted with additive white Gaussian noise, and the effectiveness of the denoising scheme is verified by the simulations. Finally, an iterative algorithm has been introduced for the simultaneous estimation of the FRI input signal and the sparse system. Since there is an approximate framework, which is introduced in Section 2.4, that enables computation of approximations of Fourier data with an arbitrary sampling kernel, the proposed method for system identification in this chapter works also with general low-pass filters but with less accuracy .

Chapter 6

FRESH – FRI-based single-image super-resolution algorithm

In this chapter we consider the problem of single image super-resolution and propose a novel algorithm that outperforms state-of-the-art methods without the need of learning patches pairs from external datasets. We have the observation that images can be modelled by lines of piecewise smooth functions, although non-bandlimited, can be seen as the sum of a piecewise polynomial signal that can be recovered by FRI and a slowly varying part that can be recovered by classical linear reconstruction. Based on this model, we first propose a resolution enhancement method for this type of functions and then its extension to 2-D images. We also propose a further improvement of the method based on learning from the errors of our super-resolution result at lower resolution levels.

6.1 Introduction

Single-image super-resolution refers to the problem of obtaining a high-resolution (HR) version of a single low-resolution (LR) image. This differs from the more traditional multi-frame super-resolution problem where one has access to multiple shifted versions of the LR image and tries to estimate a single HR image from these multiple images (see [64] for a nice overview). The single image super-resolution problem is highly ill-posed since it is possible to find many high-resolution images that can lead to the same low-resolution one. Thus prior knowledge of the properties of natural images has to be used to regularize the problem.

Strategies to solve the resolution enhancement problem are typically categorized into three broad methods: interpolation based methods, constrained reconstruction based methods,

and learning based methods. Interpolation based techniques [65,66], e.g. bilinear, bicubic interpolation algorithms have their roots in sampling theory and their essence is to recover the continuous-time signal from the given discrete pixels. They are computationally simple, however they typically are based on a slow-varying image model (sum of weighted and shifted versions of a basis function) so often produce images without high frequency details. Reconstruction based approaches (e.g., [67–73]) define constraints for the target high-resolution image. Commonly used priors includes statistical prior of natural images [67,68], total-variation prior [69], gradient-profile prior [71]. There are also learning based algorithms which infer missing high frequency information based on a dictionary containing pairs of low-resolution and high-resolution patches. The dictionary is either trained externally using a database of low-resolution (LR) and high-resolution (HR) image pairs [74–79] or internally using self-similarities of the image at different scales [80–83].

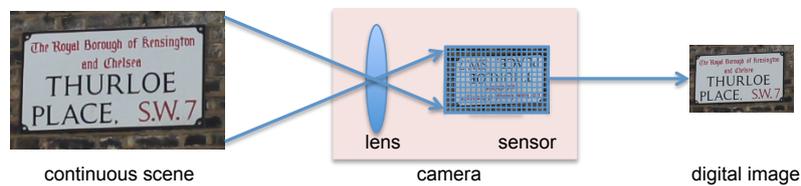


Figure 6.1: Image formation process. The incoming irradiance light field is blurred by the lens and sampled by the image sensor.

In this chapter we connect the single image super-resolution problem to the one of sampling and reconstructing piecewise regular functions. We note that the high-resolution, high-frequency information of an image is lost during the acquisition process (see Fig. 6.1) because of lens blur (usually modelled by the point spread function) and limited density of imaging sensors. This process is very similar to the way acquisition is modelled in traditional sampling theory where the analogue signal is low-pass filtered (equivalent to the blurring due to lenses) and then sampled (equivalent to the sensor grid in a digital camera). Linear interpolation methods have the merit of making this connection explicit. However, they are not effective in practice because they can only recover globally smooth functions, whereas images and scan-lines of images are piecewise regular (see Fig. 6.2).

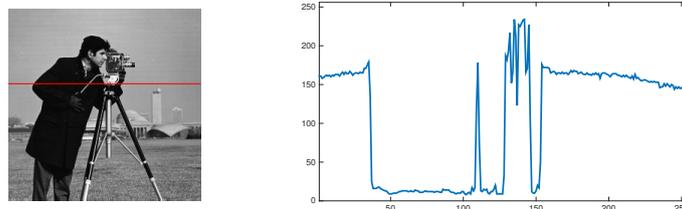


Figure 6.2: Natural images and scan-lines of natural images are approximately 2-D and 1-D piecewise smooth functions respectively.

The recently developed FRI theory overcomes in some cases the limitation of the linear interpolation methods by exploiting the fact that many signals, like piecewise polynomial functions, are fully specified by a finite number of parameters. In Chapter 2, we have shown that these signals, although non-bandlimited, can be uniquely reconstructed by only a small number of samples taken with specific acquisition devices. In other words, there is a unique mapping between a specific low-resolution version to the infinite-resolution version for these signals and there is a constructive way for recovery. We have also shown in Sec. 2.4 that this exact framework is later extended to the approximate FRI framework that works with any sampling kernel [38].

This insight inspires a novel method for sampling continuous-time image scan-lines or 1-D piecewise smooth functions: a piecewise smooth signal can be modelled as the sum of a piecewise polynomial and a globally smooth part and we propose a hybrid reconstruction method based on classical linear recovery of the smooth part and non-linear recovery of the piecewise polynomial part using FRI on the same set of samples. We then leverage from wavelet theory and the corresponding multi-resolution analysis [84] to adapt this hybrid reconstruction method to the resolution enhancement problem. In particular, enhancing the resolution of a signal is equivalent to finding the detail wavelet coefficients at finer scales. We do this using FRI and we combine the details with the coarse linear approximation. Because of the connection with wavelet theory this can be achieved using filter banks. This leads to a fast and extremely effective algorithm to enhance the resolution of 1-D piecewise smooth functions.

We extend this approach to images by approximating the point-spread-function with a scaling function in the wavelet theory (typically a spline of a certain order) and apply the 1-D method along vertical, horizontal and diagonal directions. These reconstruction are then combined with the low-resolution version of the image using a 2-D filter-bank.

Finally, inspired by the works in single image super-resolution based on self-learning, we propose correcting the error in our FRI upsampling result by learning from the pair of input LR image and the corresponding FRI image of same size recovered from an even lower scale. Instead of using simple bicubic interpolation as “LR image” in self-learning algorithms, we are using FRI reconstruction which is much sharper, and as a result we do not require learning at every small increment of scales. The end result is an FRI-based single-image Super-resolution algorithm - FRESH, which outperforms state-of-the-art methods in most situations. Fig. 6.3 shows a comparison on one example.

The chapter is organized as follows. In Section 6.2, we first show our interpretation of sampling from multi-resolution property of wavelet transform and propose how to sample piecewise smooth signals, then by relating the resolution enhancement problem to sampling problem we propose a method for enhancing the resolution of piecewise smooth signals

using filter-banks. In Section 6.3, we show how the 1-D upsampling method is extended to enhance the resolution of 2-D images. We then propose an improved upsampling method which corrects errors in the FRI upsampled image by exploiting the similarities between images at different resolutions. We show simulation results in Section 6.4 and conclude in Section 6.5.

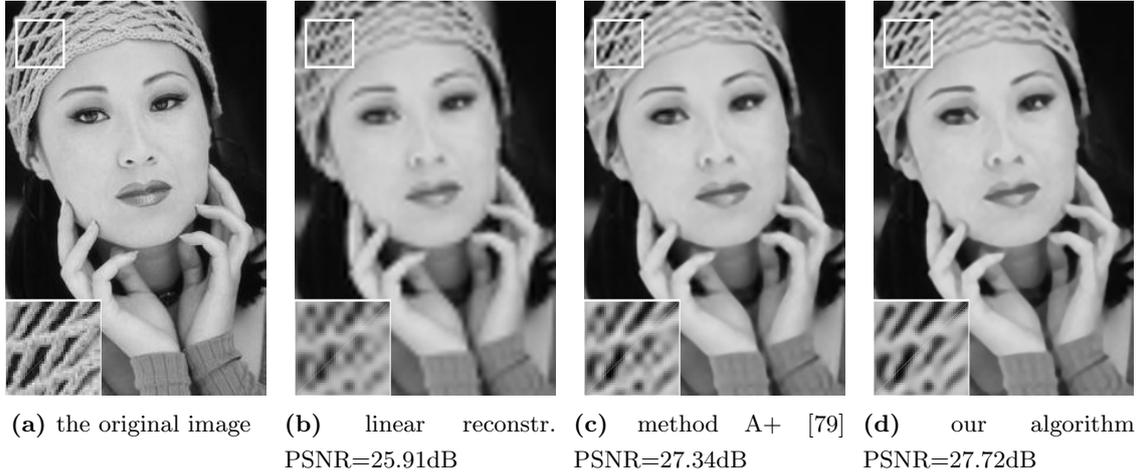


Figure 6.3: Upsampling results (factor 4) of *woman* by different methods. The downsampling kernel is *bior4.4*.

6.2 Sampling and resolution enhancement of 1-D piecewise smooth signals

We have the following observations for the problem of enhancing the resolution of images. First of all we note that the image formation process in a digital camera (refer to Fig. 6.1) can be seen as a 2-D version of the classical sampling set-up of Fig. 6.4 where the sampling kernel now is the point spread function of the camera. Moreover, images are piecewise regular functions (see Fig. 6.2), therefore, enhancing images to infinite resolution can be interpreted as the problem of sampling and reconstructing 2-D piecewise smooth functions.

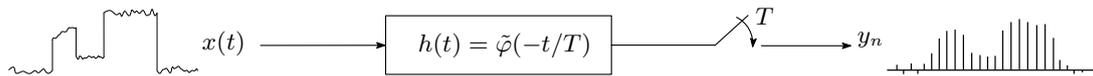


Figure 6.4: Enhancing images to infinite resolution can be interpreted as the problem of sampling and reconstructing piecewise smooth functions.

In this section, we consider a 1-D version of this problem and discuss the 2-D case in Sec. 6.3. We consider the sampling and reconstruction of 1-D continuous-time functions

first and then the resolution enhancement of 1-D discrete-time signals.

6.2.1 Sampling of piecewise smooth signals

We consider the case where the sampling kernel $\varphi(t)$ in the setup of Fig. 6.4 (or point spread function in 2-D case) is the scaling function of a wavelet transform because the wavelet framework naturally relates the linear and FRI non-linear reconstruction methods to the notion of resolution enhancement and therefore provides a proper way to combine them. Here we emphasize that the FRI sampling method itself is universal since it works with any kernel. The wavelet scheme may seem to limit the classes of acquisition filters we are able to handle, however it is worth mentioning that the point spread function of a camera can be accurately modelled by splines which are valid scaling functions.

Now we provide an interpretation of the problem of sampling an input signal $x(t)$ from the multi-resolution representation of $x(t)$. Denote with $\varphi(t)$ and $\psi(t)$ the scaling and wavelet functions respectively, and with $\varphi_{J,n}(t) = 2^{-J/2}\varphi(2^{-J}t-n)$ and $\psi_{m,n}(t) = 2^{-m/2}\psi(2^{-m}t-n)$, $J, m, n \in \mathbb{Z}$ the set of dilated and shifted versions of the scaling and wavelet function. Consider the following multi-resolution representation of a signal $x(t)$ in terms of the scaling and wavelet functions:

$$x(t) = \underbrace{\sum_{n=-\infty}^{\infty} y_{J,n}\varphi_{J,n}(t)}_{x_J(t)} + \sum_{m=-\infty}^J \sum_{n=-\infty}^{\infty} d_{m,n}\psi_{m,n}(t), \quad (6.1)$$

where $y_{J,n} = \langle x(t), \tilde{\varphi}_{J,n} \rangle$ and $d_{m,n} = \langle x(t), \tilde{\psi}_{m,n} \rangle$. Here $\tilde{\varphi}_{J,n}$, $\tilde{\psi}_{m,n}$ are the dual bases of $\varphi_{J,n}$ and $\psi_{m,n}$ respectively. We also note that $x_J(t)$ in (6.1) represents an approximation of $x(t)$ at resolution 2^J . Adding more and more levels of details $\sum_{n=-\infty}^{\infty} d_{m,n}\psi_{m,n}(t)$ to the coarse version $x_J(t)$ gives finer and finer resolution approximations and eventually the original signal $x(t)$.

The inner products $y_{J,n} = \langle x(t), \tilde{\varphi}_{J,n} \rangle$ are equivalent to the samples obtained by sampling $x(t)$ with sampling kernel $\tilde{\varphi}_{J,n}$ and sampling period $T = 2^J$ (see Fig. 6.4). Moreover, the coarse approximation $x_J(t) = \sum_{n=-\infty}^{\infty} y_{J,n}\varphi_{J,n}(t)$ in (6.1) corresponds to the linear reconstruction process in classical sampling theory discussed in Sec. 2.1 (see also Fig. 2.2), which finds projection of $x(t)$ onto the shift-invariant subspace spanned by $\{\varphi_{J,n}(t)\}_{n \in \mathbb{Z}}$ [20]. However, for the purpose of resolution enhancement, we are after an algorithm which is able to recover details from $y_{J,n}$ which are beyond the coarse approximation.

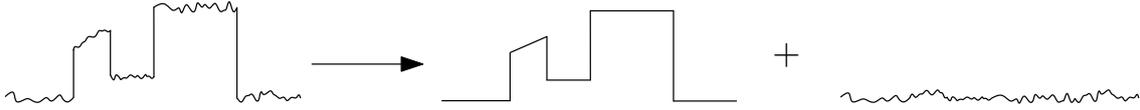


Figure 6.5: We model piecewise smooth signals as the sum of a piecewise polynomial signal and a globally smooth signal.

We assume $x(t)$ is piecewise smooth and model piecewise smooth functions as the combination of a piecewise polynomial signal $p(t)$ and a globally smooth function $r(t)$ (see Fig. 6.5). We assume the smooth part $r(t)$ lives in the shift-invariant subspace generated by integer shifts of $\varphi(2^{-J}t)$. The piecewise smooth function can then be expressed as:

$$\begin{aligned}
 x(t) &= p(t) + r(t) \\
 &= \underbrace{\sum_{n=-\infty}^{\infty} y_{J,n}^p \varphi_{J,n}(t)}_{p(t)} + \sum_{m=-\infty}^J \sum_{n=-\infty}^{\infty} d_{m,n}^p \psi_{m,n}(t) + \underbrace{\sum_{n=-\infty}^{\infty} y_{J,n}^r \varphi_{J,n}(t)}_{r(t)} \\
 &= \sum_{n=-\infty}^{\infty} \underbrace{(y_{J,n}^p + y_{J,n}^r)}_{y_{J,n}} \varphi_{J,n}(t) + \sum_{m=-\infty}^J \sum_{n=-\infty}^{\infty} d_{m,n}^p \psi_{m,n}(t). \tag{6.2}
 \end{aligned}$$

By comparing (6.2) with (6.1), we notice that the details $d_{m,n}^p$ we need for resolution enhancement are due only to the piecewise polynomial part $p(t)$ (see also Fig. 6.6), and the remaining part can simply be obtained through linear recovery using coefficients $y_{J,n}$.

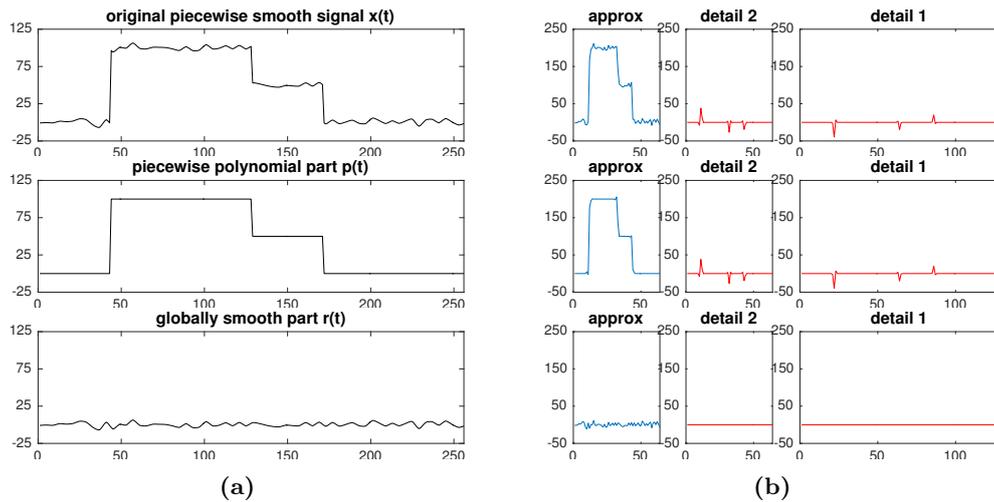


Figure 6.6: The details we need for resolution enhancement are due only to the piecewise polynomial part. (a) The piecewise smooth signal $x(t) = p(t) + r(t)$, the piecewise polynomial part $p(t)$ and the globally smooth part $r(t)$. (b) Wavelet decomposition of $x(t)$, $p(t)$ and $r(t)$ respectively.

This observation together with the fact that FRI theory enables accurate reconstruction of piecewise polynomial function using any kernel $\tilde{\varphi}(t)$ (refer to Section 2.2.3 for the reconstruction algorithm and Section 2.4 for reconstruction with an arbitrary kernel), leads to the proposed hybrid reconstruction strategy highlighted in Fig. 6.7, which recovers $p(t)$ using FRI method and $r(t)$ using the traditional linear reconstruction approach.

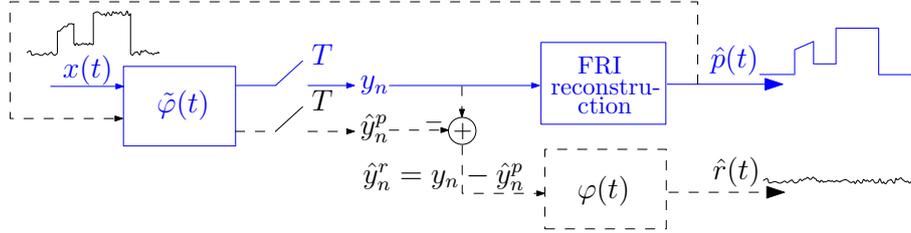


Figure 6.7: Schematic diagram of our proposed sampling and reconstruction strategy for piecewise smooth signals. The blue solid part: FRI reconstruction of the piecewise polynomial function $p(t)$. The black dashed part: linear reconstruction of the smooth residual $r(t)$.

Specifically, our proposed scheme first reconstructs the piecewise polynomial part $p(t)$ using the approximate Strang-Fix theory from the samples $y_{J,n}$ by treating the globally smooth residual as noise (the blue solid part in Fig. 6.7). Given the estimated piecewise polynomial $\hat{p}(t)$, we compute $\hat{y}_{J,n}^p = \langle \hat{p}(t), \tilde{\varphi}_{J,n}(t) \rangle$ which can then be removed from the samples $y_{J,n}$ to obtain the contribution $\hat{y}_{J,n}^r = y_{J,n} - \hat{y}_{J,n}^p$ due to the smooth residual $r(t)$. Then $r(t)$ can be reconstructed by classical linear method using the dual of the sampling kernel, i.e. $r(t) = \sum_{n=-\infty}^{\infty} \hat{y}_{J,n}^r \varphi_{J,n}(t)$ (the black dashed part in Fig. 6.7). The estimation of $x(t)$ is then the summation of estimated piecewise polynomial and estimated smooth part.

6.2.2 Resolution enhancement of 1-D piecewise smooth signal

Given the discrete-time sequence $y_{J,n} = \langle x(t), \tilde{\varphi}_{J,n} \rangle$, rather than trying to reconstruct the original continuous-time signal $x(t)$, one could be more interested in just trying to enhance the resolution of $y_{J,n}$ and the multi-resolution decomposition of (6.1) provides the right framework to achieve this goal. Assume that $\tilde{\varphi}(t)$ is a valid scaling function satisfying the two-scale relation:

$$\tilde{\varphi}(t) = \sqrt{2} \sum h_0[n] \tilde{\varphi}(2t - n), \quad (6.3)$$

and that we aim to enhance the resolution of $y_{J,n}$ by a factor 2^K for some positive integer K . It is then natural to seek for the signal $y_{J-K,n}$ which corresponds to the sequence obtained by sampling $x(t)$ with scaling function $\tilde{\varphi}_{J-K}(t)$ at finer scale 2^{J-K} .

Because of the two-scale equation (6.3) we can relate $y_{J,n}$ to $y_{J-K,n}$ using the K -level filter

bank of Fig. 6.8. More precisely,

$$y_J = (y_{J-K} * h_0^{(K)}) \downarrow^{2^K}, \quad (6.4)$$

where $h_0^{(K)}$ in z -domain is $H_0^{(K)}(z) = H_0(z)H_0(z^2) \dots H_0(z^{2^{K-1}})$ and is derived based on the fact that the cascade of K analysis filters $H_0(z)$ each followed by subsampling by 2 is equivalent to one filter $H_0^{(K)}(z) = H_0(z)H_0(z^2) \dots H_0(z^{2^{K-1}})$ followed by subsampling by 2^K . Here $H_0(z)$ is the z -transform of h_0 in (6.3).

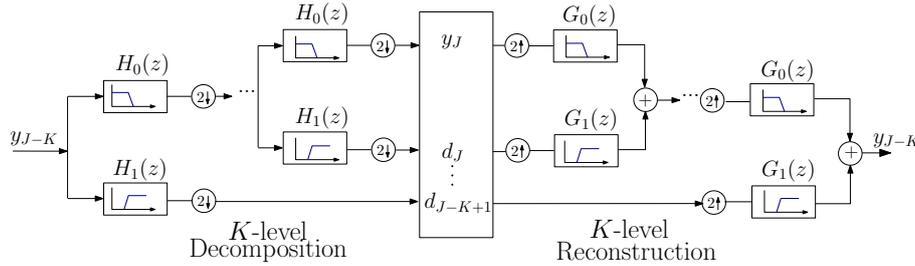


Figure 6.8: K -level biorthogonal filter bank. Given the approximation coefficients $y_J = \langle x(t), \tilde{\varphi}_{J,n} \rangle$ we are looking for a higher resolution version $y_{J-K} = \langle x(t), \tilde{\varphi}_{J-K,n} \rangle$.

Therefore, one reasonable linear upsampling of $y_{J,n}$ is the version obtained by simply feeding $y_{J,n}$ to the wavelet reconstruction stage, and can be expressed as:

$$\hat{y}_{J-K} = y_J \uparrow^{2^K} * g_0^{(K)}, \quad (6.5)$$

where $g_0^{(K)}$ in z -domain is $G_0^{(K)}(z) = G_0(z)G_0(z^2) \dots G_0(z^{2^{K-1}})$ and \hat{y}_{J-K} is the projection of y_{J-K} to the subspace spanned by $\{g_0^{(K)}[n - 2^k k]\}_{k \in \mathbb{Z}}$. Here $G_0(z)$ is the synthesis low-pass filter.

However, as Fig. 6.8 also indicates, this linear reconstruction does not allow us to retrieve the missing detail coefficients $d_J \dots d_{J-K+1}$, so in order to get a better estimate of y_{J-K} we estimate the detail coefficients using FRI.

More specifically, knowing the low-pass filter $\tilde{\varphi}_J(t)$ we first apply FRI method in Sec. 2.2.3 to estimate the piecewise polynomial part $p(t)$ from the approximation coefficients $y_{J,n}$, with the assumption that the contribution to $y_{J,n}$ only comes from $p(t)$ and that the smooth part $r(t)$ is noise. We call this estimated piecewise polynomial signal $\hat{p}^{\text{FRI}}(t)$. We then put $\hat{p}^{\text{FRI}}(t)$ onto a grid of resolution 2^{J-K-L} with $L > 0$, which is a grid finer than the resolution 2^{J-K} we are looking for, and we denote this discretized polynomial with $\hat{p}_{J-K-L}^{\text{FRI}}[n]$. Recall that in our piecewise smooth model the detail coefficients are due only to the piecewise polynomial, we therefore obtain the detail coefficients at resolution 2^J to 2^{J-K+1} from $\hat{p}_{J-K-L}^{\text{FRI}}[n]$ through $(K+L)$ -level filter-bank decomposition. Finally the estimation of $y_{J-K,n}$ is obtained by computing K -level wavelet reconstruction

of Fig. 6.8 using the approximation coefficients $y_{J,n}$ and the estimated detail coefficients $d_J[n] \dots d_{J-K+1}[n]$. We summarize this resolution enhancement method in Fig. 6.9.

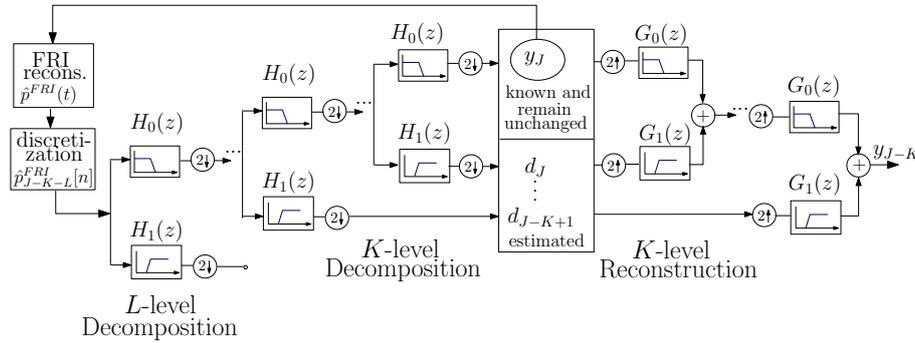
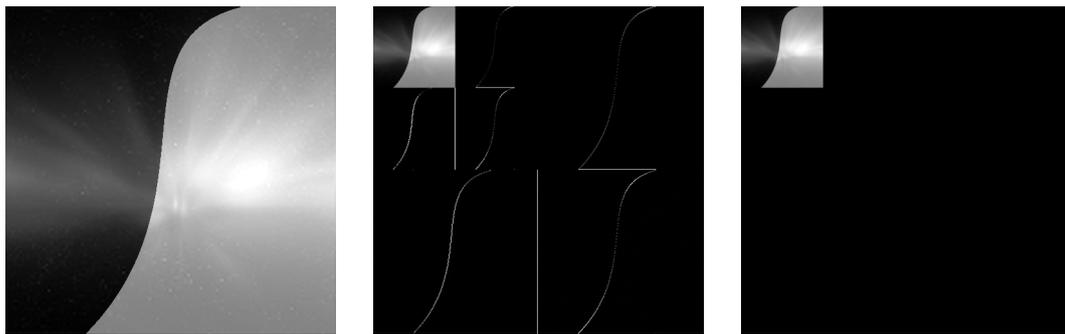


Figure 6.9: The schematic diagram of resolution enhancement of a piecewise smooth signal by factor of 2^K using a biorthogonal filter bank. From the given approximation coefficients $y_J = \langle x(t), \tilde{\varphi}_{J,n} \rangle$ and the detail coefficients d_J, \dots, d_{J-K+1} estimated using FRI, we are able to recover a higher resolution version $y_{J-K} = \langle x(t), \tilde{\varphi}_{J-K,n} \rangle$.

6.3 Image up-sampling

Equipped with the resolution enhancement method of the previous section, we now approach image upsampling by modelling lines (along different directions) of images as 1-D piecewise smooth functions and extend the method of 1-D case to 2-D images. For clarity and simplicity, we denote the image at original low-resolution with y_0 and its upsampled version by factor 2^K with y_{-K} . The low-resolution image y_0 of size $N \times N$ is the low-pass version of a K -level 2D wavelet transform applied to the high-resolution image y_{-K} of size $2^K N \times 2^K N$ with all the high-pass coefficients discarded (see Fig. 6.10).



(a) The high-resolution piecewise smooth image

(b) Low-pass and high-pass subbands of a 2-level 2D wavelet transform of (a)

(c) We only have access to the low-pass subband in (b)

Figure 6.10: Assume we only have access to the low-pass subband y_0 of a 2D wavelet transform applied to the high-resolution image y_{-K} . We want to estimate the high-pass coefficients using FRI in order to recover the high-resolution image.

The 2D wavelet decomposition leads to a set of approximation coefficients and three types of high-pass coefficients, representing horizontal, vertical and diagonal high-frequency details respectively. Similar to the 1D case, a simple linear reconstruction of y_{-K} from approximation coefficients y_0 merely increases the size of y_0 without increasing the resolution. The goal is to estimate the high-pass coefficients for the purpose of resolution enhancement.

In part A, we introduce the basic FRI image upsampling algorithm, and we explain in part B an improved method which uses the basic algorithm and also exploits patch repetitions across scales to correct errors of the basic upsampling method.

6.3.1 Basic image up-sampling algorithm

The basic idea is that high-pass coefficients can be estimated from the set of FRI recovered images (stacks of high-resolution piecewise polynomials) along horizontal, vertical and diagonal directions. The proposed image upsampling method is summarized in the block diagram of Fig. 6.11 and described in further detail in this section.

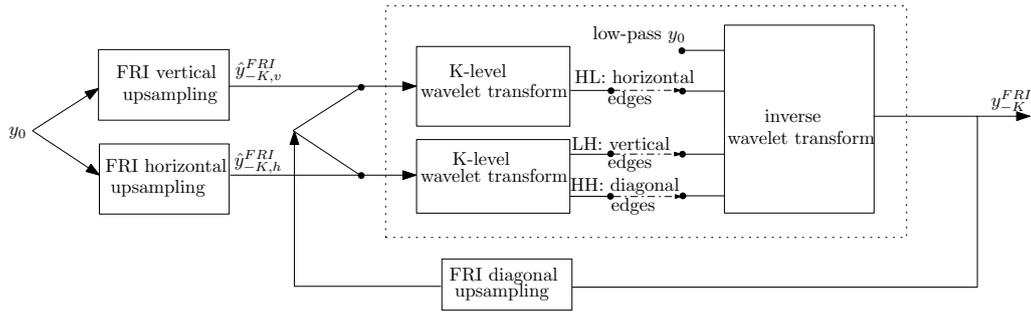


Figure 6.11: The block diagram of our proposed upsampling scheme.

We first linearly interpolate column by column the low-resolution image y_0 to size $2^K N \times N$ using the synthesis filter $g_0^{(K)}$ (refer to (6.5)). Then from the N samples of each horizontal line we reconstruct a piecewise polynomial of length $2^K N$ using the resolution enhancement method described in Sec. 6.2.2. In this way we get an image $\hat{y}_{-K,h}^{FRI}$ of size $2^K N \times 2^K N$ which contains estimates of the vertical edges of y_{-K} . Similarly, we operate on the other coordinate of y_0 to recover an image $\hat{y}_{-K,v}^{FRI}$ that contains horizontal edges.

Then we obtain a first reconstruction \hat{y}_{-K}^{FRI} by feeding $\hat{y}_{-K,h}^{FRI}$ to the decomposition channels that capture vertical edges, $\hat{y}_{-K,v}^{FRI}$ to the channels that capture horizontal edges and either one of them to the channels that capture diagonal edges.

We then improve this reconstruction by operating diagonally. We do this to remove jaggies in \hat{y}_{-K}^{FRI} . We first downsample each 45 degree diagonal line of \hat{y}_{-K}^{FRI} by factor 2, and then

recover it with FRI and we call the recovered image $\hat{y}_{-K,d1}^{\text{FRI}}$. We then do the same on -45 degree diagonal lines of $\hat{y}_{-K,d1}^{\text{FRI}}$ and obtain the image $\hat{y}_{-K,d2}^{\text{FRI}}$. Then we reconstruct a final image by selecting patch by patch (size 4×4 with 1-pixel overlap) from $\hat{y}_{-K,d1}^{\text{FRI}}$ and $\hat{y}_{-K,d2}^{\text{FRI}}$. For patch with dominant gradient direction closer to 45 degrees we use the patch from $\hat{y}_{-K,d1}^{\text{FRI}}$ and otherwise from $\hat{y}_{-K,d2}^{\text{FRI}}$. Then we ensure the consistency between our reconstruction and the input data y_0 by replacing the approximation coefficients of our reconstruction with y_0 . The final unsampled image with upsampling factor 2^K is denoted with y_{-K}^{FRI} .

Fig. 6.12 demonstrates that our proposed method is able to upsample a piecewise smooth image with sharp edges. It is evident that we gain by adding our estimated high-frequency information.

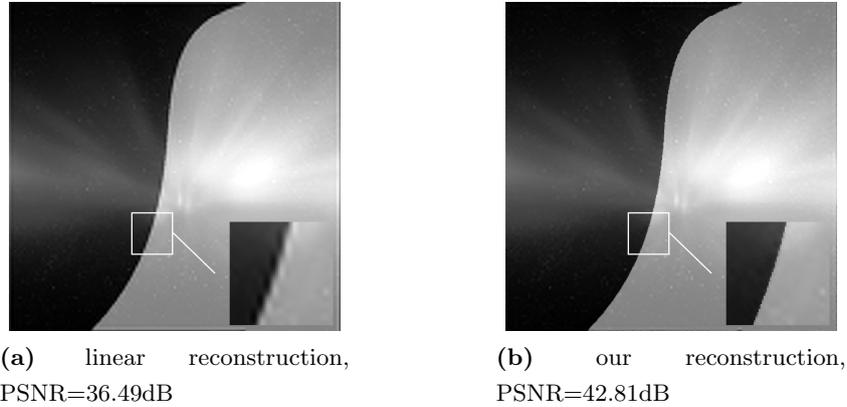


Figure 6.12: Upsampling results of a piecewise smooth image by linear reconstruction and by proposed method.

6.3.2 Exploiting cross scale similarities

Inevitably, there are errors in our FRI reconstruction y_{-K}^{FRI} , some are due to imperfection of the piecewise smooth model—for example not all edges behave like an immediate transition between two nearby pixels. Inspired by the idea of deriving a HR patch from an input LR patch with a linear transformation learnt from internal LR and HR dictionary patches [82], we propose estimating and correcting the error in upsampled FRI image by learning the relationship between the ground truth input LR image and our FRI reconstruction recovered from an even lower resolution version of the input LR image. More precisely, a specific patch in y_{-1}^{FRI} is expected to be corrected by a linear transformation \mathbf{M} which transforms its similar patches in y_{-1+m}^{FRI} to corresponding patches in y_{-1+m} , where y_{-1+m} is an intermediate scale 1.25 times smaller than y_{-1} (1.6 times larger than y_0). This is possible because there are patch repetitions across small-scale factors (typically

1.25). This is also the prior typically used in most single-image super-resolution techniques without external dictionary [80–82].

Now we explain in detail our proposed algorithm for upsampling by 2. Its main idea is depicted in Fig. 6.13. For an upsampling factor 2^K ($K > 1$), we iterate it K times.

First of all, we use the basic FRI upsampling method explained in Section 6.3.1 to upsample input low-resolution image y_0 to y_{-1}^{FRI} (see Fig. 6.13a).

Then for updating the y_{-1}^{FRI} we try to create the FRI image y_{-1+m}^{FRI} and ground truth image y_{-1+m} pair, whose resolution is slightly lower than y_{-1} . A temporary version of y_{-1+m} is obtained by bicubic interpolation of y_0 by factor 1.6 (see Fig. 6.13a), and its corresponding FRI image y_{-1+m}^{FRI} is obtained by downsampling y_{-1+m} by factor 2 followed by upsampling by 2 using basic FRI image upsampling method of Section 6.3.1 (see Fig. 6.13b).

Having obtained the first version of y_{-1+m} and y_{-1+m}^{FRI} pair, we use them to update y_{-1}^{FRI} as follows (see Fig. 6.13c): first, for each patch $b_{-1,i}^{\text{FRI}}$ in y_{-1}^{FRI} (size 5×5 with 1-pixel shift each time), we search for P (e.g. 4) similar patches $\{b_{-1+m,i,j}^{\text{FRI}}\}_{j=1}^P$ in y_{-1+m}^{FRI} . We do the search locally within a small window of 25 by 25 pixels centred around the relative center-coordinates of $b_{-1,i}^{\text{FRI}}$. Now we compute the transformation $\mathbf{M}_i \in \mathbb{R}^{25 \times 25}$ that maps the P vectorized patches in y_{-1+m}^{FRI} to the corresponding P vectorized patches in y_{-1+m} . This transformation \mathbf{M}_i is expected to correct $b_{-1,i}^{\text{FRI}}$ and the way to compute it is explained in detail later. We apply the \mathbf{M}_i learnt to correct $b_{-1,i}^{\text{FRI}}$:

$$b_{-1,i}^{\text{corrected}}(\cdot) = \mathbf{M}_i b_{-1,i}^{\text{FRI}}(\cdot), \quad (6.6)$$

where $b(\cdot)$ denote the vectorized version of patch b . All the corrected patches $b_{-1,i}^{\text{corrected}}$ are then combined to obtain a corrected high-resolution image $y_{-1}^{\text{corrected}}$ by averaging contributing patch values at each pixel. Then we ensure data fidelity by replacing the low-pass coefficients of $y_{-1}^{\text{corrected}}$ with the ground-truth y_0 .

Lastly, we want to update y_{-1+m} , y_{-1+m}^{FRI} and subsequently $y_{-1}^{\text{corrected}}$ because the current y_{-1+m} we learnt from is simply a bicubic interpolation of y_0 which is blurred and is not well served as the ground truth image at resolution 2^{-1+m} . So we update it (see Fig. 6.13d) by downsampling $y_{-1}^{\text{corrected}}$ by factor 1.25 using bicubic interpolation. Its corresponding FRI version y_{-1+m}^{FRI} is updated by downsampling y_{-1+m} by 2 followed by upsampling by 2 using the basic FRI image upsampling method. Then given the new y_{-1+m} and y_{-1+m}^{FRI} pair we re-calculate the transformation \mathbf{M}_i for each patch $b_{-1,i}^{\text{FRI}}$ in y_{-1}^{FRI} and apply the new \mathbf{M}_i to $b_{-1,i}^{\text{FRI}}$ as in (6.6). Again, we combine all the patches and ensure the low-pass version of $y_{-1}^{\text{corrected}}$ is y_0 . We note that we could further repeat this updating step. However, we have numerical evidence that further iterations would not improve the result significantly, so we stick with one updating iteration.

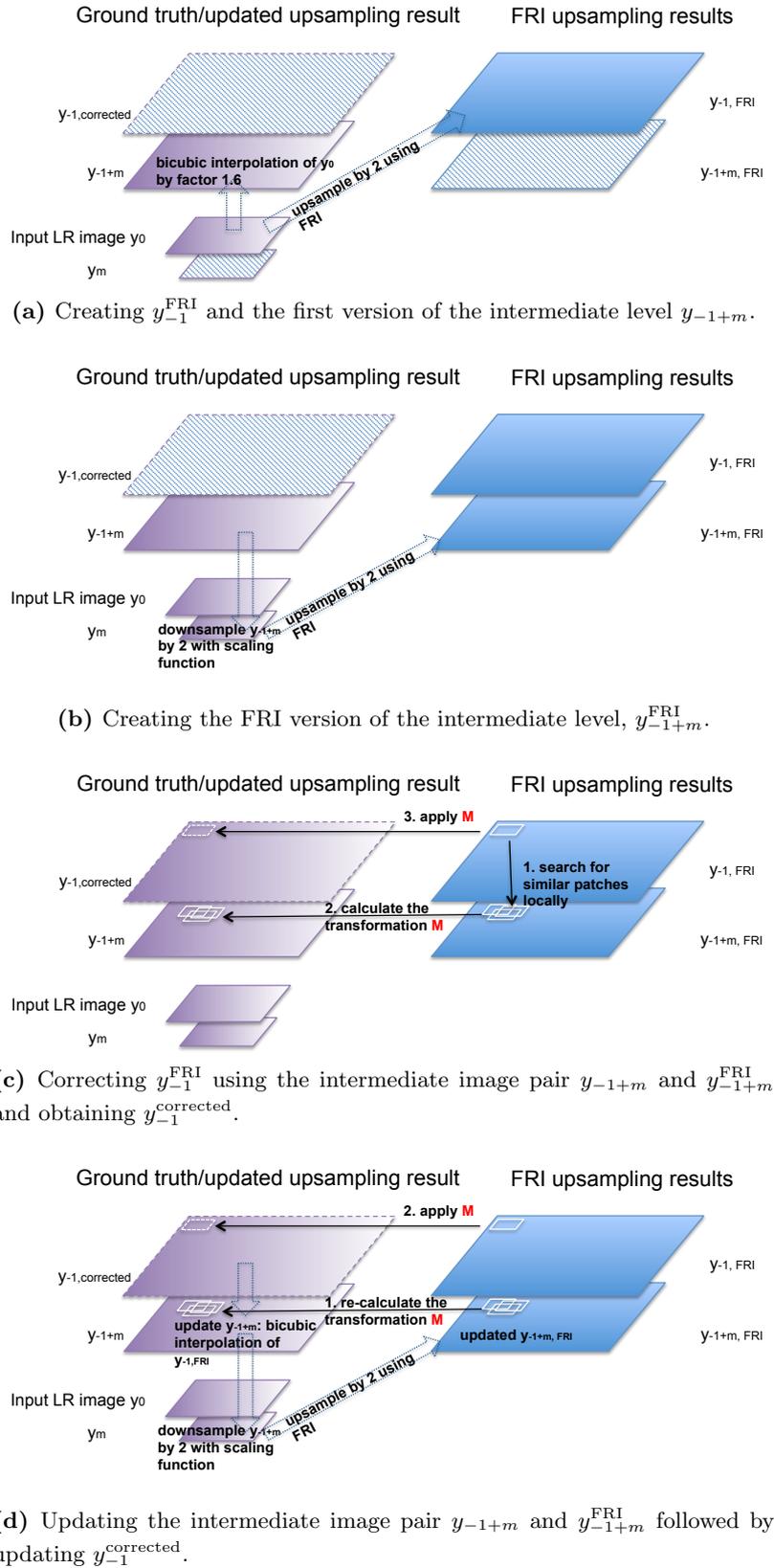


Figure 6.13: The schematic diagram of how to correct our FRI upsampled image y_{-1}^{FRI} .

The linear transformation $\mathbf{M}_i \in \mathbb{R}^{D^2 \times D^2}$ that maps the P FRI patches $\{b_{-1,i,j}^{\text{FRI}} \in \mathbb{R}^{D \times D}\}_{j=1}^P$ to ground truth patches $\{b_{-1,i,j} \in \mathbb{R}^{D \times D}\}_{j=1}^P$ can be found by minimizing the empirical fitting error between all P pairs of examples. However, this problem is underdetermined and Tikhonov regularization is added to solve it stably:

$$\begin{aligned} \mathbf{M}_i &= \underset{\mathbf{M}_i \in \mathbb{R}^{D^2 \times D^2}}{\operatorname{argmin}} \sum_{j=1}^P \|b_{i,j}(\cdot) - \mathbf{M}_i b_{i,j}^{\text{FRI}}(\cdot)\|^2 + \lambda \|\mathbf{M}_i\|_F^2 \\ &= \underset{\mathbf{M}_i \in \mathbb{R}^{D^2 \times D^2}}{\operatorname{argmin}} \|\mathbf{B}_i - \mathbf{M}_i \mathbf{B}_i^{\text{FRI}}\|_F^2 + \lambda \|\mathbf{M}_i\|_F^2, \end{aligned} \quad (6.7)$$

where λ is a regularization parameter, \mathbf{B}_i and $\mathbf{B}_i^{\text{FRI}}$ are matrices with $\{b_{i,j}(\cdot)\}_{j=1}^P$ and $\{b_{i,j}^{\text{FRI}}(\cdot)\}_{j=1}^P$ as their columns respectively. The solution of (6.7) can be written in a close-form as follows:

$$\mathbf{M}_i = \mathbf{B}_i \mathbf{B}_i^{\text{FRI}\top} \left(\mathbf{B}_i^{\text{FRI}} \mathbf{B}_i^{\text{FRI}\top} + \lambda \mathbf{I} \right)^{-1}, \quad (6.8)$$

where \mathbf{I} is the identity matrix.

To conclude, we summarize the complete method in Algorithm 5.

Algorithm 5: FRI-based single image super-resolution algorithm

input : y_0

output: y_{-1} : upsampled version of y_0

- 1 Upsample y_0 to y_{-1}^{FRI} by the basic FRI algorithm in part A.
 - 2 Create the intermediate level y_{-1+m} by upsampling y_0 by a factor 1.6 using bicubic interpolation
 - 3 Create the corresponding intermediate FRI level y_{-1+m}^{FRI} by first downsampling y_{-1+m} by 2 followed by upsampling it using the basic FRI algorithm.
 - 4 **for** patch $b_{-1,i}^{\text{FRI}}$ in y_{-1}^{FRI} **do**
 - 5 Search for P similar patches $\{b_{-1+m,i,j}^{\text{FRI}}\}_{j=1}^P$ in y_{-1+m}^{FRI} , locally within the small window centred around the relative center-coordinates of $b_{-1,i}^{\text{FRI}}$.
 - 6 Calculate the linear transformation \mathbf{M}_i that transforms the patches $\{b_{-1+m,i,j}^{\text{FRI}}\}_{j=1}^P$ to the corresponding ground truth patches $\{b_{-1+m,i,j}\}_{j=1}^P$ in y_{-1+m} (refer to (6.8)).
 - 7 Correct $b_{-1,i}^{\text{FRI}}$ by applying the linear transformation \mathbf{M}_i to it (refer to (6.6)).
 - 8 **end**
 - 9 Combine all the corrected patches and replace the low-pass coefficients of the corrected y_{-1}^{FRI} with the ground-truth y_0 , and call the image $y_{-1}^{\text{corrected}}$.
 - 10 Update the intermediate level y_{-1+m} by downsampling $y_{-1}^{\text{corrected}}$ with bicubic interpolation.
 - 11 Update y_{-1+m}^{FRI} by downsampling the updated y_{-1+m} by 2 followed by upsampling by 2 with the basic FRI algorithm.
 - 12 Repeat step 4 to 9 with the updated pair of intermediate images except step 5 need not to be recalculated.
-

6.4 Simulation results

6.4.1 1-D piecewise smooth signal upsampling

In the following simulations, we show the resolution enhancement results using our novel hybrid reconstruction method introduced in Sec. 6.2.2. In this section we assume our samples y_0 (refer to Fig. 6.8) are the low-pass coefficients of 2-level wavelet decomposition applied to a high-resolution piecewise smooth signal and we want to recover it to its original resolution.

First, we demonstrate in Fig. 6.14 that when the signal is exactly the discrete-time version of model (6.2), our method, compared to the linear reconstruction and the total variation method, is able to achieve nearly perfect reconstruction.

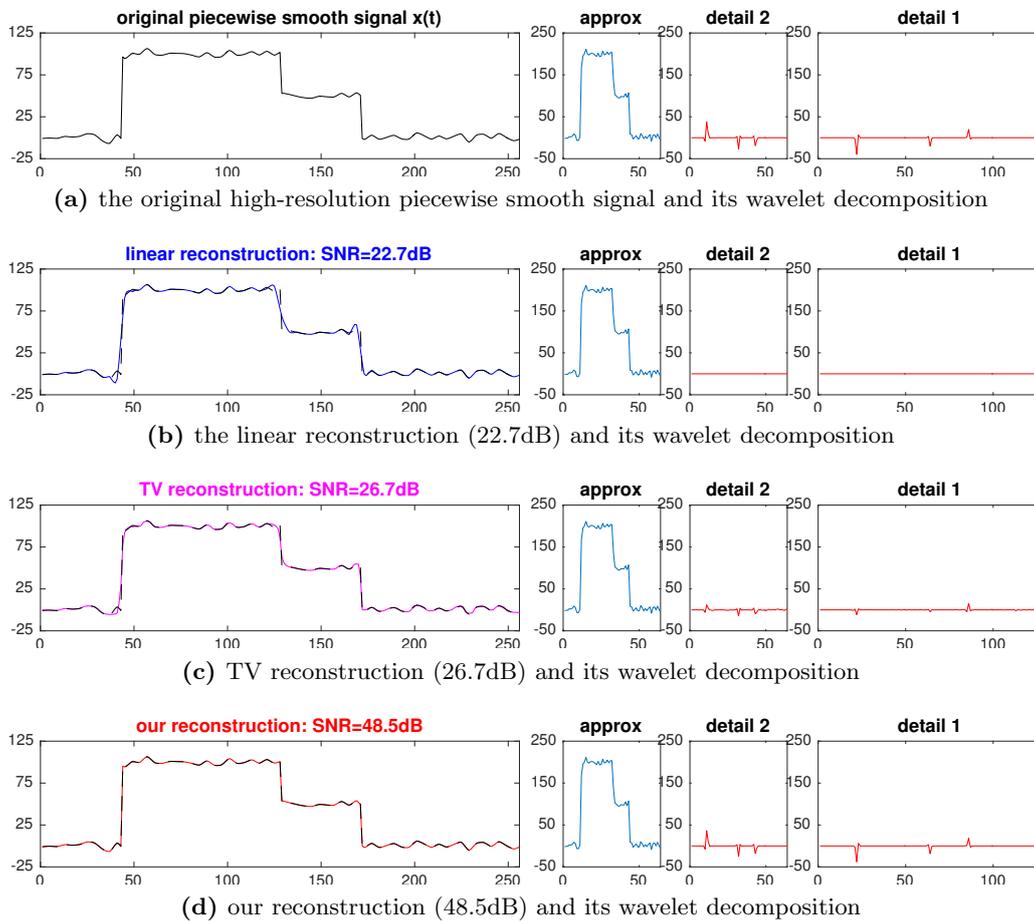


Figure 6.14: Our method is able to accurately recover a piecewise smooth signal from its approximation coefficients.

Then we also test our algorithm in the case where the high-resolution signal is not exactly in our model but is a scan-line of an image (see Fig. 6.2). The result of comparisons

between our reconstruction, the linear reconstruction, and the total variation method in Table 6.1 shows that our model is an accurate representation of the scan-lines such that our proposed method is still efficient in this case. The result also proves the universality of our method in the sense that it works robustly with different downsampling kernels.

Table 6.1: Recovering the high-resolution image scan-line from its approximation coefficients of different wavelet decomposition with different methods.

scaling function	linear reconstr.	TV reconstr.	our reconstr.
linear spline	21.7 dB	23.1 dB	24.2 dB
cubic spline	22.2 dB	23.1 dB	24.4 dB
bior4.4	23.9 dB	23.0 dB	25.0 dB

6.4.2 Image upsampling

Upsampling of artificially downsampled images

In this part, we test upsampling of low-resolution images obtained by downsampling the original ones by a factor 4 using the 2-level 2-D wavelet decomposition with biorthogonal 4.4 filter, and those obtained by downsampling with a linear spline. We compare our basic upsampling method in Sec. 6.3.1 and our improved method in Sec. 6.3.2 with the linear reconstruction method and some of the state-of-the-art algorithms, we show the upsampling results of the two different downsampling kernels in terms of PSNR and SSIM (structural similarity index [85]) in Table 6.2 and Table 6.3 respectively. Visual comparison on one of the test images ‘Zebra’ (downsampling kernel of bior4.4) and ‘Comic’ (downsampling kernel of linear spline) are shown in Fig. 6.15 and Fig. 6.16 respectively. Note that for self-learning method [80], we use a third-party implementation [86] and we cannot guarantee the implementation duplicates the original results. For other methods we ensure the comparison is fair because we modified the blurring kernel in the source codes to the specific kernel we use (bior4.4/linear spline), and for dictionary-based methods of [78, 79], the dictionary was re-trained with the same kernel used in upsampling process.

Our proposed method is universal in the sense that it works with different blurring kernels. The results demonstrate that our proposed basic (fast) method, with no learning involved, outperforms other reconstruction-based algorithms, e.g. total variation [69], contourlet [70] and even one of the dictionary learning methods [78]. The improved method – FRESH is robust and outperforms state-of-the-art methods in different categories. Moreover, our method leads to visually pleasant edges.

Table 6.2: Comparisons of upsampling results (factor 4) given by different methods in terms of PSNR. Sampling kernel: bior4.4.

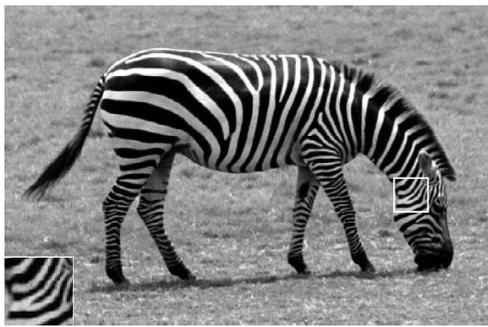
PSNR (dB) & SSIM [85]	linear	TV [69]	con- tourlet ^a [70]	sparse	A+ [79]	self- learn [80]	our basic method	FRESH
				cod- ing [78]				
Peppers	29.91	30.98	30.19	31.03	31.73	31.31	31.13	31.85
	0.823	0.835	0.824	0.837	0.846	0.837	0.836	0.845
Lena	29.49	29.97	29.82	30.17	30.64	29.99	30.16	30.55
	0.835	0.837	0.839	0.844	0.853	0.843	0.841	0.849
Cameraman	28.35	28.85	28.59	29.21	29.63	29.04	29.16	29.82
	0.872	0.883	0.872	0.885	0.892	0.886	0.886	0.894
Butterfly	21.50	23.76	21.66	22.37	23.19	24.41	23.40	24.24
	0.744	0.853	0.736	0.796	0.836	0.859	0.833	0.864
Bird	29.37	29.99	29.75	30.23	30.94	30.04	30.29	31.04
	0.877	0.887	0.881	0.893	0.905	0.888	0.891	0.904
Comic	20.83	21.10	–	21.07	21.29	21.00	21.13	21.32
	0.621	0.647	–	0.636	0.654	0.654	0.642	0.659
Zebra	23.61	24.34	–	24.38	24.71	24.44	24.50	25.11
	0.711	0.721	–	0.726	0.731	0.732	0.726	0.737
Woman	25.91	26.71	–	26.66	27.34	27.36	26.91	27.72
	0.842	0.857	–	0.861	0.876	0.872	0.859	0.879
Average	26.12	26.96	–	26.89	27.43	27.20	27.09	27.71
	0.791	0.815	–	0.810	0.824	0.821	0.814	0.829

^a contourlet interpolation software does not support upsampling of non-square images.

Table 6.3: Comparisons of upsampling results (factor 4) given by different methods in terms of PSNR. Sampling kernel: linear spline.

PSNR (dB) & SSIM [85]	linear	TV [69]	con- tourlet ^a [70]	sparse	A+ [79]	self- learn [80]	our basic method	FRESH
				cod- ing [78]				
Peppers	29.95	31.09	30.25	30.77	31.63	31.49	31.28	31.95
	0.821	0.835	0.824	0.831	0.844	0.836	0.836	0.845
Lena	29.53	30.02	29.90	29.84	30.58	30.13	30.28	30.66
	0.834	0.838	0.84	0.835	0.850	0.837	0.842	0.850
Cameraman	28.40	28.89	28.66	28.88	29.54	29.27	29.34	30.05
	0.870	0.884	0.871	0.877	0.890	0.881	0.888	0.896
Butterfly	21.55	23.91	21.72	22.08	23.03	24.22	23.52	24.61
	0.737	0.856	0.736	0.782	0.830	0.860	0.835	0.870
Bird	29.42	30.05	29.82	29.86	30.73	30.23	30.43	31.26
	0.876	0.887	0.881	0.883	0.901	0.888	0.892	0.906
Comic	20.85	21.15	–	20.92	21.27	21.09	21.22	21.41
	0.619	0.648	–	0.616	0.647	0.632	0.644	0.662
Zebra	23.67	24.41	–	24.13	24.61	24.65	24.61	25.30
	0.711	0.722	–	0.709	0.728	0.728	0.728	0.741
Woman	25.95	26.78	–	26.36	27.22	27.23	27.04	27.86
	0.838	0.858	–	0.851	0.872	0.867	0.861	0.881
Average	26.17	27.04	–	26.60	27.33	27.29	27.22	27.89
	0.788	0.816	–	0.798	0.820	0.816	0.816	0.831

^a contourlet interpolation software does not support upsampling of non-square images.



(a) the original image

(b) linear reconstruction
PSNR=23.61dB, SSIM=0.711(c) total variation [69]
PSNR=24.34dB, SSIM=0.721(d) sparse coding [78]
PSNR=24.38dB, SSIM=0.726(e) A+ [79]
PSNR=24.71dB, SSIM=0.731(f) self-learning [80]
PSNR=24.44dB, SSIM=0.732(g) our basic algorithm
PSNR=24.50dB, SSIM=0.726(h) our improved algorithm
PSNR=25.11dB, SSIM=0.737

Figure 6.15: Upsampling results (factor 4) of *Zebra* by different methods. The downsampling kernel is bior4.4.



Figure 6.16: Upsampling results (factor 4) of *Comic* by different methods. The downsampling kernel is linear spline.

Upsampling of images taken with a camera

Finally, we show that the proposed algorithm is also able to upsample the images taken with a real camera, where the blurring due to lens is not exactly a scaling function as assumed previously but can still be modelled as a spline. We demonstrate in Fig. 6.17 that the algorithm achieves visually good performance for upsampling factor of 4. In the following result, the original photographs are taken with Canon 400D, and its point spread function is modelled by the fifth order spline. The upsampling is performed only on the luminance component of the input image and the chrominance component are simply upscaled by bicubic interpolation.

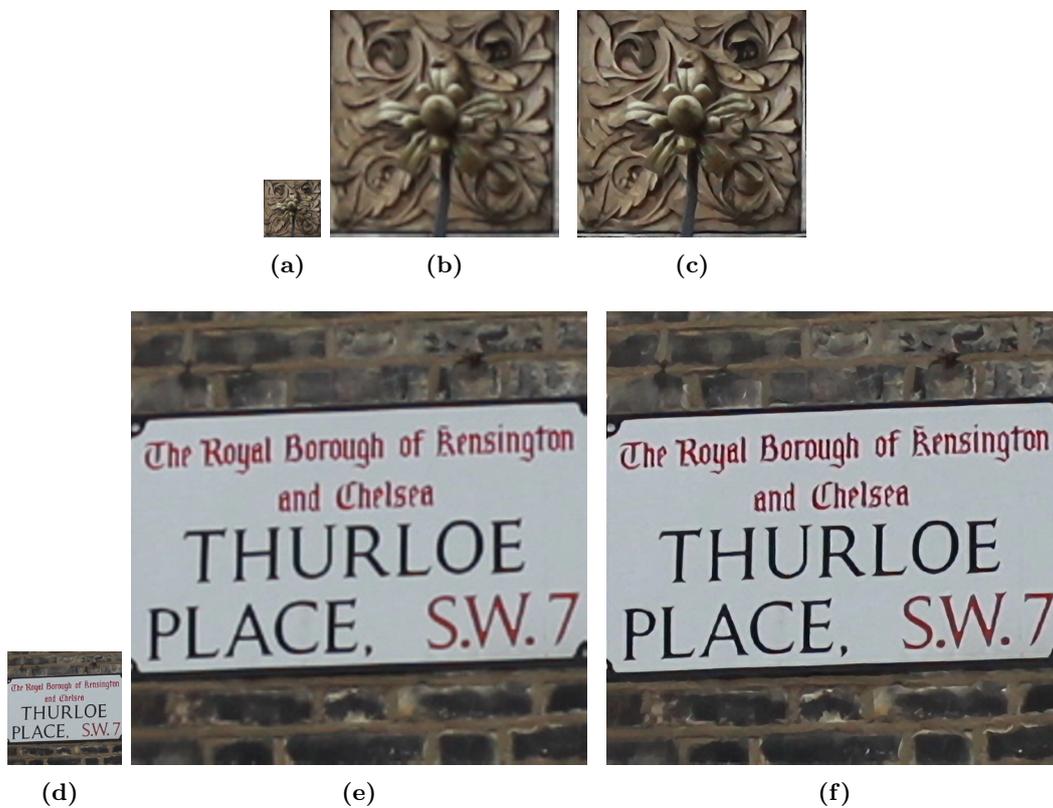


Figure 6.17: Upsampling of images taken with Canon 400D. Our upsampling results are sharper than the bicubic interpolation results. (a)(d) original images. (b)(e) bicubic interpolation. (c)(f) our upsampling results.

6.4.3 Computation complexity and discussions

Upsampling an image of size $N \times N$ to $2^K N \times 2^K N$ with the basic algorithm proposed in Sec. 6.3.1 requires number of line upsampling operations in the order of $2^K N$ and also $2^K N \times 2^K N$ block selecting operations.

Unlike the basic algorithm, the improved algorithm proposed in Sec. 6.3.2 requires up-sampling one level per time (factor 2 each time) and involves searching for similar blocks, which is computational more expensive.

Table 6.4 shows the execution time (averaged over 5 test images) of the C++ implementations¹ of the basic method and FRESH on a Mac mini with 2.6GHz Intel Core i7 CPU and 16GB RAM.

Table 6.4: Computation cost for upsampling an image using the basic method and FRESH on a Mac mini with 2.6GHz Intel Core i7 CPU and 16GB RAM.

time (seconds)	upsample 64×64 images by 2	upsample 64×64 images by 4
basic algorithm	0.45	2.68
FRESH	1.65	8.95

Since we are using a wavelet scheme to handle the upsampling, the scale factor can only be powers of 2. Moreover, the point spread function of the camera needs to be fairly close to the scaling function of a wavelet transform. The wavelet scheme may seem to limit the classes of acquisition filters we are able to handle, nevertheless we want to emphasize that the point spread function of cameras in general can be accurately modelled by splines which are valid scaling functions and numerical results on images taken with real cameras confirm the universality of our approach.

6.5 Summary

In this chapter, we have proposed a scheme for upsampling piecewise smooth signals and its extension to images by modelling images as lines of piecewise smooth signals. We show that the method proposed improves classical linear reconstruction results by making use of an additional non-linear reconstruction method based on FRI theory. The method is further improved by using a self-learning approach which also makes use of FRI. The resulting algorithm outperforms state-of-the-art methods and does not require the use of external datasets.

¹C++ implementations of our algorithms were created by Matteo Maggioni.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

In this thesis we have studied sampling and reconstruction scheme for signals with finite rate of innovations. This sampling scheme breaks the limitation of classical sampling theorem where only bandlimited signals can be recovered. We have studied the phenomenon of performance breakdown of the FRI reconstruction algorithms when the noise level is higher than a threshold and have provided an accurate prediction of the breakdown PSNR in terms of settings like sampling rate and the distance between adjacent Diracs. We have also extended the current FRI sampling theory to a universal one that works with unknown rate of innovations. Moreover, we have explored possible applications of FRI sampling theory. More precisely, we have extended the FRI sampling theory for identifying systems that are parametrically sparse, and also for reconstructing images that can be modelled by parametrically sparse signals.

In Chapter 2, we have revisited the classical theory for sampling bandlimited signals and discussed its limitations. Then we have introduced FRI theory, which provides a more general representation of signals, and we have shown FRI theory is able to perfectly recover classes of signals that cannot be recovered by classical sampling theory. Specifically, we have explained the reconstruction algorithms of streams of Diracs and piecewise polynomial signals with the sampling kernel being the exponential reproducing functions. Moreover, we have described how to choose the most stable sampling kernel and what algorithms can be used for robust reconstruction when the samples are noisy. We have explained the method to access performance of FRI reconstruction algorithms by using Cramér-Rao bounds, and shown that the widely used FRI reconstruction algorithm, i.e. Cadzow algorithm with Prony's method, in general is able to achieve the best possible performance given by the Cramér-Rao bound above a certain level of signal-to-noise ratio.

Lastly, we have revisited the generalised sampling scheme that works with any arbitrary sampling kernel where the exponential reproducing properties need not to be satisfied.

In Chapter 3, we have explained the breakdown phenomenon observed in the FRI reconstruction algorithms by subspace swap, which refers to the event where noise subspace singular vectors are so large that replace one or more of the signal subspace singular vectors. By analysing the signal and noise singular values of the data matrix on which Cadzow denoising algorithm is operated, we have derived the necessary conditions for subspace swap to happen, and from there we have established an explicit connection between breakdown point (PSNR) to kernel settings and distance between input streams of Diracs. Specifically, the closer the adjacent Diracs or the lower the order of the kernel, the higher the breakdown PSNR. Simulation results have proven accurate prediction of the breakdown point in different setups.

In Chapter 4, we have proposed a robust algorithm to estimate the number of Diracs and thus extend the current approximate framework to one that works with arbitrary and unknown numbers of Diracs. The algorithm tests all possible number of Diracs and decides the number of Diracs by choosing the turning point of the resynthesis error function. The old version of the algorithm finds the turning point by computing the second derivative of the resynthesis error. The new version, instead of relying only on second derivatives, calculates a score which dependent on the resynthesis error and also the first and second derivative of it for each possible number. The number with highest score will be the estimated number of Diracs. Simulation results have shown that our newly proposed method is robust in that it achieves the Cramér-Rao bound with only a very low failure rate.

In Chapter 5, we have shown the possibility of perfect identification of linear systems that are fully characterized by a small number of its Fourier coefficients. The set-up is a multi-channel version of the current FRI sampling set-up with one channel taking samples of the input signal only using an exponential reproducing kernel and the other one taking samples of the output of the input signal through the unknown system using the same kernel. In order to have the scheme work robustly in noisy scenario, we have proposed a modified Cadzow algorithm that can efficiently denoise the Fourier coefficients for identification. This algorithm, similar to the Cadzow denoising algorithm, makes use of the fact that the Fourier data matrix is structured and rank-deficient when there is no noise. The algorithm gradually and iteratively recovers a rank-deficient matrix which also follows the structure of the noiseless matrix. Moreover, we have shown that in certain scenarios, even when the sampling setup has only one channel, i.e. we have no direct access to the samples of the input signal, we can still perfectly identify the unknown system without iterations. In noisy scenario, we have proposed a robust algorithm that identify the input signal and

unknown system iteratively.

In Chapter 6, we have investigated the extensions of FRI theory for image upsampling. Since we have the observation that images can be modelled by lines of piecewise smooth signals, we first approach the problems of sampling and resolution enhancement piecewise smooth signals before solving the image upsampling problem. Firstly, we have proposed a scheme that combines the classical linear reconstruction and the non-linear methods of FRI for sampling piecewise smooth signals. We have also considered the case where the input signal is discrete and proposed resolution enhancement of piecewise smooth signals by relating the notion of multi-resolution in the wavelet transform to sampling. Then based on the idea of our proposed 1-D resolution enhancement scheme, we have applied the 1-D upsampling scheme along vertical, horizontal and diagonal directions in an image to obtain an image upsampling algorithm. We have also proposed a further improvement of the algorithm, which improves our upsampled image by learning from the errors of our upsampled results at lower resolution levels. This leads to our robust algorithm for image upsampling called FRESH. Simulation results have shown that FRESH outperforms state-of-the-art algorithms without the need of external dictionaries. Moreover, numerical results on images taken with real cameras have confirmed the universality of our approach.

7.2 Future Work

In what follows, we propose some future research topics:

- We have explained the breakdown phenomenon of FRI reconstruction algorithm in high level of noise by subspace swap in Chapter 3 and we have provided an accurate prediction of the breakdown PSNR in terms of the sampling rate and the distance between adjacent Diracs for exponential reproducing kernels. Extending this performance analysis for arbitrary low-pass filters would be interesting because it would give us a better understanding on the performance of approximate Strang-Fix. This work would require studying how the response of the low-pass filter affects the noise on the exponential moments, and then the orthogonal subspace singular values.
- In Chapter 6 we have proposed an image upsampling algorithm based on the idea that lines of images can be modelled by piecewise smooth signals. One crucial step of the algorithm is to recover lines of piecewise polynomial signals. In the proposed algorithm we set the number of discontinuities of the piecewise polynomial signals to a predefined number. This implies that for all lines in all kinds of input images, we are assuming a piecewise smooth signal model with only freedom in locations

of discontinuities and amplitude information for each piece, but no freedom in the number of discontinuities. We could imagine there should be a better choice than using a predefined number of discontinuities. We may try to apply the method for estimating number of Diracs proposed in Chapter 4 for estimating the number of discontinuities.

- Piecewise smooth signal we have discussed in Chapter 6 actually is a very important class of sparse signals. We can find them in images and many punctuated real-world phenomena [87]. In Chapter 6 we show that we can recover such a signal from its samples accurately by recovering the smooth part and the piecewise polynomial part separately. When recovering the piecewise polynomial part, we regard the smooth part as noise. This means, as long as there is a smooth part the proposed method cannot recover the piecewise smooth signal perfectly, and certainly the larger the amplitudes of the smooth part, the larger the error on the reconstruction of the piecewise polynomial part. To solve the problem, we may try to develop an exact sampling framework for piecewise smooth signals. In order to perfectly recover the piecewise polynomial part, we would need a sampling setup which can not only annihilate the smooth part and output a set of samples that are only due to the piecewise polynomial signal, but also satisfy the Strang-Fix condition. Moreover, the sampling setup have to be able to output another set of samples which are due to both the piecewise polynomial part and the smooth part so that we may deduce the contribution from the recovered piecewise polynomial part from the samples and recover the smooth part linearly.

Appendix A

Find the closest matrix in the sense of Frobenius norm

To find the closest matrix to \mathbf{A}' that belongs to the matrix spanned by the matrices $\mathbf{U}_{M-Q_{\text{ext}}}^{\text{H}} \mathbf{e}_k \mathbf{e}_k^{\text{T}} \mathbf{Q}_{P_{\text{ext}}+1}$ for $k = 1, 2, \dots, M+1$, we look for a diagonal matrix $\mathbf{\Delta}$ that sets to zero the differential-infinitesimal-variation of the Frobenius distance between \mathbf{A}' and $\mathbf{U}^{\text{H}} \mathbf{\Delta} \mathbf{Q}$:

$$d \left\{ \|\mathbf{U}^{\text{H}} \mathbf{\Delta} \mathbf{Q} - \mathbf{A}'\|_{\text{Fro}}^2 \right\} = 0, \quad \text{for all possible variations } d\mathbf{\Delta}. \quad (\text{A.1})$$

Since the square of the Frobenius norm of a matrix \mathbf{A} is equal to the trace of the matrix $\mathbf{A}\mathbf{A}^{\text{H}}$, denoted $\text{Tr}(\mathbf{A}\mathbf{A}^{\text{H}})$, (A.1) can be rewritten as follows:

$$\begin{aligned} 0 &= d \left\{ \text{Tr} \left((\mathbf{U}^{\text{H}} \mathbf{\Delta} \mathbf{Q} - \mathbf{A}) (\mathbf{U}^{\text{H}} \mathbf{\Delta} \mathbf{Q} - \mathbf{A})^{\text{H}} \right) \right\} \\ &= \text{Tr} \left(d \left\{ (\mathbf{U}^{\text{H}} \mathbf{\Delta} \mathbf{Q} - \mathbf{A}) (\mathbf{U}^{\text{H}} \mathbf{\Delta} \mathbf{Q} - \mathbf{A})^{\text{H}} \right\} \right) \\ &= \text{Tr} \left(d\mathbf{\Delta} \left(\mathbf{U}\mathbf{U}^{\text{H}} \mathbf{\Delta} \mathbf{Q} \mathbf{Q}^{\text{H}} - \mathbf{U} \mathbf{A} \mathbf{Q}^{\text{H}} \right)^{\text{H}} \right) \end{aligned} \quad (\text{A.2})$$

for all possible variations $d\mathbf{\Delta}$. This implies that (A.1) is satisfied only when all the diagonal elements of the matrix $\mathbf{U}\mathbf{U}^{\text{H}} \mathbf{\Delta} \mathbf{Q} \mathbf{Q}^{\text{H}} - \mathbf{U} \mathbf{A} \mathbf{Q}^{\text{H}}$ vanish, which amounts to a system of linear equations:

$$\sum_{k'} uu_{k,k'} qq_{k,k'} d_{k'} = uaq_{k,k}, \quad \text{for all possible values of } k, \quad (\text{A.3})$$

where d_k refers to the diagonal entries of $\mathbf{\Delta}$, and $uu_{k,l}$, $qq_{k,l}$ and $uaq_{k,l}$ refer to the entries of the matrices $[\mathbf{U}\mathbf{U}^{\text{H}}]_{k,l}$, $[\mathbf{Q}\mathbf{Q}^{\text{H}}]_{k,l}$ and $[\mathbf{U}\mathbf{A}\mathbf{Q}^{\text{H}}]_{k,l}$ respectively. After (A.3) is solved, we obtain a “denoised” diagonal matrix $\mathbf{\Delta}$. The corresponding $\mathbf{U}^{\text{H}} \mathbf{\Delta} \mathbf{Q}$ is the closest matrix to \mathbf{A}' in the Frobenius sense.

Appendix B

Simultaneous identification of input signals and linear systems

B.1 Exact identification of E-spline and input stream of Diracs

Assume the linear system is an E-spline of order P with Fourier transform:

$$\Psi_{\gamma}(j\omega) = \prod_{p=0}^{P-1} \frac{1 - e^{\gamma_p - j\omega}}{j\omega - \gamma_p} \quad (\text{B.1})$$

and the sampling kernel can reproduce $M + 1$ exponentials $e^{(\alpha_0 + m\lambda)t/T}$, $m = 0, \dots, M$. By computing the weighted sum of the samples y_n^{sys} taken by this kernel, we have:

$$\begin{aligned} \tau_m^{\text{sys}} &= \sum_n c_{m,n} y_n^{\text{sys}} \\ &= X(j\omega_m) \cdot \Psi(j\omega_m) \\ &= \frac{\sum_{k=0}^{K-1} \hat{a}_k u_k^m q(e^{-j\omega_m})}{p(j\omega_m)}, \end{aligned} \quad (\text{B.2})$$

where $\hat{a}_k = a_k e^{\alpha_0 t_k/T}$, $u_k = e^{\lambda t_k/T}$, $p(j\omega_m) = \sum_{p=0}^P b_p (j\omega_m)^p$ and $q(e^{-j\omega_m}) = \sum_{q=0}^P c_q (e^{-j\omega_m})^q$ are polynomials of degree P (the order of the E-spline). Multiplying both sides of (B.2) by $p(j\omega_m)$ gives us:

$$p(j\omega_m) \tau_m^{\text{sys}} = \sum_{k=0}^{K-1} \sum_{q=0}^P f_{k,q} (e^{m\lambda(t_k - q)/T}). \quad (\text{B.3})$$

Therefore a filter with z -transform:

$$H(z) = \prod_{k=0}^{K-1} \prod_{q=0}^P (1 - e^{\lambda(t_k - q)/T} z^{-1}) = \sum_{d=0}^D h[d] z^{-d} \quad (\text{B.4})$$

with $D = K(P + 1)$ will annihilate (B.3). It follows that

$$\sum_{d=0}^D h[d] p(j\omega_{m-d}) \tau_{m-d} = 0, \quad m = D, D + 1, \dots, M. \quad (\text{B.5})$$

By replacing $p(j\omega_m)$ with $\sum_{p=0}^P b_p(j\omega_m)^p$ in (B.5) we get:

$$\sum_{d=0}^D h[d] \sum_{p=0}^P b_p(j\omega_{m-d})^p \tau_{m-d} = 0, \quad m = D, D + 1, \dots, M. \quad (\text{B.6})$$

We can write this system in matrix form as follows:

$$\begin{bmatrix} \tau_D & \dots & (j\omega_D)^P \tau_D & \dots & \tau_0 & \dots & (j\omega_0)^P \tau_0 \\ \tau_{D+1} & \dots & (j\omega_{D+1})^P \tau_{D+1} & \dots & \tau_1 & \dots & (j\omega_1)^P \tau_1 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \tau_M & \dots & (j\omega_M)^P \tau_M & \dots & \tau_{M-D} & \dots & (j\omega_{M-D})^P \tau_{M-D} \end{bmatrix} \begin{pmatrix} h[0]b_0 \\ \vdots \\ h[0]b_P \\ \vdots \\ h[D]b_0 \\ \vdots \\ h[D]b_P \end{pmatrix} = \mathbf{0}. \quad (\text{B.7})$$

By solving this linear system of equations we can then find (b_0, \dots, b_P) by setting $h[0] = 1$, and we are able to solve for the unknown parameters $\{\gamma_p\}_{p=0}^P$ by finding the roots of $p(j\omega)$. Having (b_0, \dots, b_P) also means we can retrieve $h[1], \dots, h[D]$. Now by finding the roots of $H(z)$ we are able to retrieve t_k exactly. The solution is unique and we need at least $2D + 1$ values of τ_m to solve this system, so we need the sampling kernel $\varphi(t)$ to be able to reproduce $\{e^{(\alpha_0 + m\lambda)t}\}_{m=0}^M$ with $M \geq 2D = 2K(P + 1)$.

B.2 Identification of linear circuits and input stream of Diracs

Similaly, when we try to identify a P -th order linear circuit, which is specified by:

$$\Psi(j\omega) = \frac{\sum_{q=0}^Q b_q(j\omega)^q}{\sum_{p=0}^P a_p(j\omega)^p}, \quad (\text{B.8})$$

we first compute the weighted sum of the samples as follows:

$$\begin{aligned}\tau_m^{\text{sys}} &= \sum_n c_{m,n} y_n^{\text{sys}} \\ &= X(j\omega_m) \cdot \Psi(j\omega_m) \\ &= \frac{\sum_{k=0}^{K-1} \hat{a}_k u_k^m q(j\omega_m)}{p(j\omega_m)},\end{aligned}\tag{B.9}$$

where $\hat{a}_k = a_k e^{\alpha_0 t_k/T}$, $u_k = e^{\lambda t_k/T}$, $p(j\omega_m) = \sum_{p=0}^P b_p (j\omega)^p$ is a polynomial of degree P (the order of the LTI system) and $q(j\omega_m) = \sum_{q=0}^Q c_q (j\omega)^q$ is a polynomial of degree Q . Multiplying both sides of (B.9) by $p(j\omega_m)$ we obtain:

$$p(j\omega_m) \tau_m^{\text{sys}} = \sum_{k=0}^{K-1} \sum_{q=0}^Q f_{k,q} m^q (e^{m\lambda t_k/T}).\tag{B.10}$$

Then a filter with z -transform:

$$H(z) = \prod_{k=0}^{K-1} (1 - e^{\lambda t_k/T} z^{-1})^{Q+1} = \sum_{d=0}^D h[d] z^{-d}\tag{B.11}$$

with $D = K(Q + 1)$ will annihilates (B.10). Now we can establish a linear system of equations as the one in (B.7). Again we find (b_0, \dots, b_P) and coefficients of $H(z)$, and then the locations of the Diracs t_k by looking for the roots of $H(z)$. Finally finding the remaining unknown parameters in the transfer function is straightforward. Here in order to solve the linear system we require a sampling kernel $\varphi(t)$ that can reproduce $\{e^{(\alpha_0 + m\lambda)t}\}_{m=0}^M$ with $M \geq 2D = 2K(Q + 1)$.

Bibliography

- [1] R. Pintelon and J. Schoukens, *System Identification: A Frequency Domain Approach*, Wiley-IEEE Press, 2004.
- [2] C. E. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.
- [3] E. T. Whittaker, “On the functions which are represented by the expansion of interpolating theory,” in *Proceedings of the Royal Society of Edinburgh*, 1915, vol. 35, pp. 181–194.
- [4] V. A. Kotelnikov, “On the carrying capacity of the ether and wire in telecommunications,” in *Material for the First All-Union Conference on Questions of Communication, Izd. Red. Upr. Svyazi RKKA, Moscow*, 1933, vol. 1.
- [5] 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.
- [6] C. Chen, P. Marziliano, and A. C. Kot, “2D finite rate of innovation reconstruction method for step edge and polygon signals in the presence of noise,” *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 2851–2859, 2012.
- [7] H. Pan, T. Blu, and P. L. Dragotti, “Sampling curves with finite rate of innovation,” *IEEE Transactions on Signal Processing*, vol. 62, no. 2, pp. 458–471, 2014.
- [8] I. Maravic, J. Kusuma, and M. Vetterli, “Low-sampling rate UWB channel characterization and synchronization,” *Journal of Communications and Networks*, vol. 5, no. 4, pp. 319–327, 2003.
- [9] I. Maravic, M. Vetterli, and K. Ramchandran, “Channel estimation and synchronization with sub-Nyquist sampling and application to ultra-wideband systems,” in *Proceedings of the 2004 International Symposium on Circuits and Systems, 2004*. IEEE, 2004, vol. 5, pp. V–381.

- [10] Y. Hao, P. Marziliano, M. Vetterli, and T. Blu, "Compression of ECG as a signal with finite rate of innovation," in *27th Annual International Conference of the Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005.*, Jan 2005, pp. 7564–7567.
- [11] 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, no. 4, pp. 1827–1842, 2011.
- [12] J. Onativia, S. R. Schultz, and P. L. Dragotti, "A finite rate of innovation algorithm for fast and accurate spike detection from two-photon calcium imaging," *Journal of Neural Engineering*, vol. 10, no. 4, pp. 046017, 2013.
- [13] G. Baechler, N. Freris, R. F. Quick, and R. E. Crochiere, "Finite rate of innovation based modeling and compression of ECG signals," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2013, pp. 1252–1256.
- [14] X. Wei, H. Akhondi Asl, T. Blu, and P. L. Dragotti, "Simultaneous estimation of sparse linear systems and sparse signals using finite rate of innovation principles," to be submitted, 2016.
- [15] X. Wei and P. L. Dragotti, "FRESH – FRI-based single-image super-resolution algorithm," *IEEE Transactions on Image Processing*, vol. 25, no. 8, pp. 3723–3735, 2016.
- [16] X. Wei and P. L. Dragotti, "Guaranteed performance in the FRI setting," *IEEE Signal Processing Letters*, vol. 22, no. 10, pp. 1661–1665, 2015.
- [17] X. Wei and P. L. Dragotti, "Sampling piecewise smooth signals and its application to image up-sampling," in *2015 IEEE International Conference on Image Processing*, 2015.
- [18] X. Wei and P. L. Dragotti, "Universal sampling of signals with finite rate of innovation," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 1803–1807.
- [19] D. Slepian, "On bandwidth," *Proceedings of the IEEE*, vol. 64, no. 3, pp. 292–300, 1976.
- [20] M. Unser, "Sampling-50 years after shannon," *Proceedings of the IEEE*, vol. 88, no. 4, pp. 569–587, 2000.
- [21] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Transactions on Signal Processing*, vol. 50, no. 6, pp. 1417–1428, 2002.

- [22] P. Stoica and R. L. Moses, *Spectral analysis of signals*, Prentice Hall, 2005.
- [23] R. Carriere and R. L. Moses, "High resolution radar target modeling using a modified prony estimator," *IEEE Transactions on Antennas and Propagation*, vol. 40, no. 1, pp. 13–18, 1992.
- [24] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, 1996.
- [25] Z. Leonowicz, T. Lobos, and J. Rezmer, "Advanced spectrum estimation methods for signal analysis in power electronics," *IEEE Transactions on Industrial Electronics*, vol. 50, no. 3, pp. 514–519, 2003.
- [26] L. Borcea, G. Papanicolaou, C. Tsogka, and J. Berryman, "Imaging and time reversal in random media," *Inverse Problems*, vol. 18, no. 5, pp. 1247, 2002.
- [27] P. L. Dragotti, M. Vetterli, and T. Blu, "Exact sampling results for signals with finite rate of innovation using strang-fix conditions and local kernels," in *2005 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, 2005, vol. 4, pp. iv–233.
- [28] 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, no. 5, Part 1, pp. 1741–1757, 2007.
- [29] 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.
- [30] G. Strang and G. Fix, "A Fourier analysis of the finite element variational method," in *Constructive Aspects of Functional Analysis*, pp. 793–840. Springer, 1971.
- [31] I. J. Schoenberg, "Contributions to the problem of approximation of equidistant data by analytic functions, part b: On the problem of osculatory interpolation, a second class of analytic approximation formulae," *Quarterly of Applied Mathematics*, vol. 4, no. 2, pp. 112–141, 1946.
- [32] A. Ron, "Factorization theorems for univariate splines on regular grids," *Israel Journal of Mathematics*, vol. 70, no. 1, pp. 48–68, 1990.
- [33] T. Blu, P. Thévenaz, and M. Unser, "MOMS: Maximal-order interpolation of minimal support," *IEEE Transactions on Image Processing*, vol. 10, no. 7, pp. 1069–1080, 2001.

- [34] T. Blu, P. Thévenaz, and M. Unser, “Complete parameterization of piecewise-polynomial interpolation kernels,” *IEEE Transactions on Image Processing*, vol. 12, no. 11, pp. 1297–1309, 2003.
- [35] M. Unser, A. Aldroubi, and M. Eden, “On the asymptotic convergence of b-spline wavelets to gabor functions,” *IEEE transactions on information theory*, vol. 38, no. 2, pp. 864–872, 1992.
- [36] I. Khalidov, T. Blu, and M. Unser, “Generalized L-spline wavelet bases,” in *Optics & Photonics 2005*. International Society for Optics and Photonics, 2005, pp. 59140F–59140F.
- [37] C. Vonesch, T. Blu, and M. Unser, “Generalized Daubechies wavelet families,” *IEEE Transactions on Signal Processing*, vol. 55, no. LIB-ARTICLE-2007-013, pp. 4415–4429, 2007.
- [38] J. A. Urigüen, T. Blu, and P. L. Dragotti, “FRI sampling with arbitrary kernels,” *IEEE Transactions on Signal Processing*, vol. 61, no. 21, pp. 5310–5323, 2013.
- [39] M. Unser and T. Blu, “Cardinal exponential splines: Part I-theory and filtering algorithms,” *IEEE Transactions on Signal Processing*, vol. 53, no. 4, pp. 1425–1438, 2005.
- [40] J. A. Urigüen, P. L. Dragotti, and T. Blu, “On the exponential reproducing kernels for sampling signals with finite rate of innovation,” *Proceedings of Sampling Theory and Applications (SampTA)*, 2011.
- [41] J. A. Cadzow, “Signal enhancement-a composite property mapping algorithm,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, no. 1, pp. 49–62, 1988.
- [42] 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.
- [43] 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, no. 5, pp. 814–824, 1990.
- [44] T. K. Sarkar and O. Pereira, “Using the matrix pencil method to estimate the parameters of a sum of complex exponentials,” *IEEE Antennas and Propagation Magazine*, vol. 37, no. 1, pp. 48–55, Feb 1995.

- [45] I. Maravic and M. Vetterli, “Sampling and reconstruction of signals with finite rate of innovation in the presence of noise,” *Signal Processing, IEEE Transactions on*, vol. 53, no. 8, pp. 2788–2805, 2005.
- [46] H. Cramér, *Mathematical methods of statistics*, vol. 9, Princeton university press, 1945.
- [47] C. R. Rao, “Information and the accuracy attainable in the estimation of statistical parameters,” vol. 37, pp. 81–91. 1945.
- [48] B. Porat and B. Friedlander, “Computation of the exact information matrix of gaussian time series with stationary random components,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, no. 1, pp. 118–130, 1986.
- [49] P. L. Dragotti and F. Homann, “Sampling signals with finite rate of innovation in the presence of noise,” in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2009, pp. 2941–2944.
- [50] J. K. Thomas, L. L. Scharf, and D. W. Tufts, “The probability of a subspace swap in the SVD,” *IEEE Transactions on Signal Processing*, vol. 43, no. 3, pp. 730–736, 1995.
- [51] D. W. Tufts, A. C. Kot, and R. J. Vaccaro, “The threshold analysis of SVD-based algorithms,” in *1988 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Apr 1988, pp. 2416–2419 vol.4.
- [52] M. Hawkes, A. Nehorai, and P. Stoica, “Performance breakdown of subspace-based methods: Prediction and cure,” in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, 2001, vol. 6, pp. 4005–4008.
- [53] M. W. Meckes, “On the spectral norm of a random Toeplitz matrix,” *Electronic Communications in Probability*, vol. 12, pp. 315–325, 2007.
- [54] M. Richardson, “Fundamentals of the discrete Fourier transform,” *Sound and Vibration Magazine*, pp. 40–46, 1978.
- [55] J. Schoukens, Y. Rolain, and R. Pintelon, “Analysis of windowing/leakage effects in frequency response function measurements,” *Automatica*, vol. 42, no. 1, pp. 27–38, 2006.
- [56] C. R. Pickrel and P. J. White, “Flight flutter testing of transport aircraft: In-flight modal analysis,” in *Proceedings of IMAC 21, the International Modal Analysis Conference*, 2003.

- [57] B. M. Sanandaji, T. L. Vincent, K. Poolla, and M. B. Wakin, "A tutorial on recovery conditions for compressive system identification of sparse channels," in *2012 IEEE 51st Annual Conference on Decision and Control (CDC)*. IEEE, 2012, pp. 6277–6283.
- [58] M. McCormick, Y. M. Lu, and M. Vetterli, "Learning sparse systems at sub-Nyquist rates: A frequency-domain approach," in *2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, 2010, pp. 4018–4021.
- [59] Y. Chen, Y. Gu, and A. O. Hero, "Sparse LMS for system identification," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2009, pp. 3125–3128.
- [60] O. Taheri and S. A. Vorobyov, "Sparse channel estimation with lp-norm and reweighted l1-norm penalized least mean squares," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2011, pp. 2864–2867.
- [61] D. Kandaswamy, T. Blu, and D. Van De Ville, "Analytic sensing: Noniterative retrieval of point sources from boundary measurements," *SIAM Journal on Scientific Computing*, vol. 31, no. 4, pp. 3179–3194, 2009.
- [62] Z. Dogan, V. Tsiminaki, I. Jovanovic, T. Blu, and D. Van De Ville, "Localization of point sources for systems governed by the wave equation," in *Proceedings of SPIE*, 2011, vol. 8138, pp. 81380P–1.
- [63] L. Ljung, "Experiments with identification of continuous time models," *IFAC Proceedings Volumes*, vol. 42, no. 10, pp. 1175–1180, 2009.
- [64] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Advances and challenges in super-resolution," *International Journal of Imaging Systems and Technology*, vol. 14, no. 2, pp. 47–57, 2004.
- [65] R. G. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 29, no. 6, pp. 1153–1160, 1981.
- [66] P. Thévenaz, T. Blu, and M. Unser, "Interpolation revisited," *IEEE Transactions on Medical Imaging*, vol. 19, no. 7, pp. 739–758, 2000.
- [67] M. F. Tappen, B. C. Russell, and W. T. Freeman, "Exploiting the sparse derivative prior for super-resolution and image demosaicing," in *In IEEE Workshop on Statistical and Computational Theories of Vision*. Citeseer, 2003.
- [68] R. Fattal, "Image upsampling via imposed edge statistics," in *ACM Transactions on Graphics (TOG)*. ACM, 2007, vol. 26, p. 95.

- [69] J. M. Bioucas-Dias and M. A. Figueiredo, “A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration,” *IEEE Transactions on Image Processing*, vol. 16, no. 12, pp. 2992–3004, 2007.
- [70] N. Mueller, Y. Lu, and M. N. Do, “Image interpolation using multiscale geometric representations,” in *Proc. SPIE Conf. on Electronic Imaging*, San Jose, USA, 2007.
- [71] J. Sun, J. Sun, Z. Xu, and H.-Y. Shum, “Image super-resolution using gradient profile prior,” in *IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008*. IEEE, 2008, pp. 1–8.
- [72] Q. Shan, Z. Li, J. Jia, and C.-K. Tang, “Fast image/video upsampling,” in *ACM Transactions on Graphics (TOG)*. ACM, 2008, vol. 27, p. 153.
- [73] P. Sen and S. Darabi, “Compressive image super-resolution,” in *2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers*. IEEE, 2009, pp. 1235–1242.
- [74] W. T. Freeman, T. R. Jones, and E. C. Pasztor, “Example-based super-resolution,” *Computer Graphics and Applications, IEEE*, vol. 22, no. 2, pp. 56–65, 2002.
- [75] H. Chang, D.-Y. Yeung, and Y. Xiong, “Super-resolution through neighbor embedding,” in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004*. IEEE, 2004, vol. 1, pp. I–I.
- [76] K. I. Kim and Y. Kwon, “Single-image super-resolution using sparse regression and natural image prior,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 6, pp. 1127–1133, 2010.
- [77] J. Yang, J. Wright, T. S. Huang, and Y. Ma, “Image super-resolution via sparse representation,” *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861–2873, 2010.
- [78] R. Zeyde, M. Elad, and M. Protter, “On single image scale-up using sparse-representations,” in *Curves and Surfaces*, pp. 711–730. Springer, 2012.
- [79] R. Timofte, V. De Smet, and L. Van Gool, “A+: Adjusted anchored neighborhood regression for fast super-resolution,” in *Computer Vision—ACCV 2014*, pp. 111–126. Springer, 2014.
- [80] D. Glasner, S. Bagon, and M. Irani, “Super-resolution from a single image,” in *2009 IEEE 12th International Conference on Computer Vision*. IEEE, 2009, pp. 349–356.
- [81] G. Freedman and R. Fattal, “Image and video upscaling from local self-examples,” *ACM Transactions on Graphics (TOG)*, vol. 30, no. 2, pp. 12, 2011.

- [82] M. Bevilacqua, A. Roumy, C. Guillemot, and M.-L. Alberi Morel, “Single-image super-resolution via linear mapping of interpolated self-examples,” *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5334–5347, 2014.
- [83] J.-B. Huang, A. Singh, and N. Ahuja, “Single image super-resolution from transformed self-exemplars,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5197–5206.
- [84] S. G. Mallat, “A theory for multiresolution signal decomposition: the wavelet representation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.
- [85] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [86] C.-Y. Yang, C. Ma, and M.-H. Yang, “Single-image super-resolution: A benchmark,” in *Computer Vision–ECCV 2014*, pp. 372–386. Springer, 2014.
- [87] M. F. Duarte, M. B. Wakin, and R. G. Baraniuk, “Fast reconstruction of piecewise smooth signals from incoherent projections,” *SPARS05*, 2005.