

1 Sampling at the Rate of Innovation: Theory and Applications

Jose Antonio Urigüen¹, Yonina C. Eldar^{2,3}, Pier Luigi Dragotti¹ & Zvika Ben-Haim²

¹Imperial College, London, ²Technion, Israel Institute of Technology, Haifa, Israel, ³Visiting Professor at Stanford, CA, USA

Parametric signals, such as streams of short pulses, appear in many applications including bio-imaging, radar, and spread-spectrum communication. The recently developed finite rate of innovation (FRI) framework, has paved the way to low rate sampling of such signals, by exploiting the fact that only a small number of parameters per unit of time are needed to fully describe them. For example, a stream of pulses can be uniquely defined by the time-delays of the pulses and their amplitudes, which leads to far fewer degrees of freedom than the signal's Nyquist rate samples. This chapter provides an overview of FRI theory, algorithms and applications. We begin by discussing theoretical results and practical algorithms allowing perfect reconstruction of FRI signals from a minimal number of samples. We then turn to treat recovery from noisy measurements. Finally, we overview a diverse set of applications of FRI theory, in areas such as superresolution, radar and ultrasound.

1.1 Introduction

We live in an analog world, but we would like our digital computers to interact with it. For example, sound is a continuous-time phenomenon, which can be characterized by the variations in air pressure as a function of time. For digital processing of such real-world signals to be possible, we require a sampling mechanism which converts continuous signals to discrete sequences of numbers, while preserving the information present in those signals.

In classical sampling theory, which dates back to the beginning of the 20th century [1–3], a bandlimited signal whose maximum frequency is f_{\max} is sampled at or above the Nyquist rate $2f_{\max}$. It is well known that the signal can then be perfectly reconstructed from its samples. Unfortunately, real-world signals are rarely truly bandlimited, if only because most signals have finite duration in time. Even signals which are approximately bandlimited often have to be sampled at a fairly high Nyquist rate, requiring expensive sampling hardware and high-throughput digital machinery.

Classical sampling theory necessitates a high sampling rate whenever a signal has a high bandwidth, even if the actual information content in the signal is low.

For instance, a piecewise linear signal is non-differentiable; it is therefore not bandlimited, and moreover, its Fourier transform decays at the fairly low rate $O(\frac{1}{f^2})$. However, the signal is completely described by the positions of knots (transitions between linear segments) and the signal values at those positions. Thus, as long as the knots are known to have a minimum separation, this signal has a finite information rate. It seems wasteful to sample such signals at the Nyquist rate. It would be more efficient to have a variety of sampling techniques, tailored to different signal models, such as bandlimited or piecewise linear signals. Such an approach echoes the fundamental quest of compressive sampling, which is to capture only the essential information embedded in a signal. This chapter, together with Chapter 3, on Xampling, apply the idea of compressed sensing to certain classes of analog signals. While the focus of Xampling is on signals lying in unions of subspaces and on developing a unified architecture for efficient sampling of various classes of signals, here we concentrate on a comprehensive review of finite rate of innovation (FRI) theory.

To be specific, suppose that a function $x(t)$ has the property that any finite duration segment of length τ is completely determined by no more than K parameters. In this case, the function $x(t)$ is said to have a local rate of innovation equal to $\frac{K}{\tau}$ [4], because it has no more than K degrees of freedom every τ seconds. In general, a signal is said to have FRI if its local rate of innovation is finite for a sufficiently large τ . For example, the aforementioned piecewise linear signal has this property. Many important signal models, such as splines and pulse streams, also satisfy the FRI property, and will be explored in depth later in this chapter.

An elegant and powerful result is that, in many cases, certain types of FRI signals can be reconstructed without error from samples taken at the rate of innovation [4]. The advantage of this result is self-evident: FRI signals need not be bandlimited, and even if they are, the Nyquist frequency can be much higher than the rate of innovation. Thus, by using FRI techniques, the sampling rate required for perfect reconstruction can be lowered substantially. However, exploiting these capabilities requires careful design of the sampling mechanism and of the digital post-processing. The purpose of this chapter is to review the theory, recovery techniques, and applications of the FRI model.

1.1.1 The sampling scheme

Consider the sampling setup shown in Fig. 1.1, where the original continuous-time signal $x(t)$ is filtered before being uniformly sampled at a rate of $f_s = \frac{1}{T}$. The filtering may be a design choice or may be due to the acquisition device. If we denote the filtered version of $x(t)$ by $y(t) = h(t) * x(t)$, then the samples $\{y_n\}$ are given by

$$y_n = y(nT) = \left\langle x(t), \varphi\left(\frac{t}{T} - n\right) \right\rangle = \int_{-\infty}^{\infty} x(t) \varphi\left(\frac{t}{T} - n\right) dt, \quad (1.1)$$

where the *sampling kernel* $\varphi(t)$ is the scaled and time-reversed version of $h(t)$. For example, the previously discussed classical sampling setup often incorporates an anti-aliasing low-pass filter $h(t) = \text{sinc}(t)$, which eliminates any signal components having frequencies above $\frac{f_s}{2}$.

Changing the sampling kernel $\varphi(t)$ provides considerable flexibility in the information transferred to the samples $\{y_n\}$. Indeed, many modern sampling techniques, such as sampling in shift-invariant spaces, rely on an appropriate choice of the sampling kernel [5, 6]. As we will see, the model of Fig. 1.1 with adequate sampling kernels also provides the basis for most FRI sampling techniques. On the other hand, FRI recovery methods are typically more elaborate, and involve nonlinear digital processing of the samples. This is an important practical aspect of FRI techniques: the sampling hardware is simple, linear, and easy to implement, but it is followed by nonlinear algorithms in the digital stage, since this is typically easier and cheaper to customize.

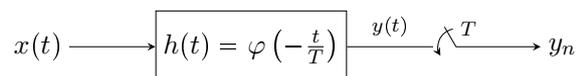


Figure 1.1 *Traditional sampling scheme.* The continuous-time input signal $x(t)$ is filtered with $h(t)$ and sampled every T seconds. The samples are then given by $y_n = (x * h)(t)|_{t=nT}$.

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

Sampling is a typical ill-posed problem in that one can construct an infinite number of signals that lead to the same samples $\{y_n\}$. To make the problem tractable one then has to impose some constraints on the choice of $x(t)$. Bandlimited signals are the prototypical example of such a constraint, and yield both a one-to-one mapping and a practical recovery technique. The set of bandlimited signals also happens to form a shift-invariant subspace of the space of continuous-time functions. As it turns out, the classical sampling theorem can be extended to signals belonging to arbitrary shift-invariant subspaces, such as splines having uniformly spaced knots [5, 6].

In many cases, however, requiring that a signal belongs to a subspace is too strong a restriction. Consider the example of piecewise linear functions. Is the set of all such functions a subspace? Indeed it is, since the sum of any two piecewise linear signals is again piecewise linear, as is the product of a piecewise linear function with a scalar. However, the sum of two such functions will usually contain a knot wherever either of the summands has a knot. Repeatedly summing piecewise linear signals will therefore lead to functions containing an infinite number of infinitesimally spaced knots; these contain an infinite amount

of information per time unit and clearly cannot be recovered from samples taken at a finite rate.

To avoid this difficulty, we could consider uniform piecewise linear signals, i.e., we could allow knots only at predetermined, equally-spaced locations. This leads to the shift-invariant subspace setting mentioned above, for which stable recovery techniques exist [6]. However, instead of forcing fixed knot positions, one could merely require, for example, a combination of a finite number of piecewise linear signals with arbitrary known locations. In many cases, such a restriction better characterizes real-world signals, although it can no longer be modeled as a linear subspace. Rather, this is an instance of a union of subspaces [7, 8]: each choice of valid knot positions forms a subspace, and the class of allowed signals is the union of such subspaces. The minimum separation model also satisfies the FRI property, and can be recovered efficiently from samples taken at the rate of innovation. The union of subspaces structure, which is explored in more detail in Chapter 3, is useful in developing a geometrical intuition of FRI recovery techniques. There is, however, a distinction between the union of subspaces and FRI models. In particular, there exist FRI settings which cannot be described in terms of unions of subspaces, for example, when the signal parameters do not include an amplitude component. There are also unions of subspaces which do not conform to the FRI scenario, in particular when the parameters affect the signal in a non-local manner, so that finite-duration segments are not determined by a finite number of parameters.

Our discussion thus far has concentrated on perfect recovery of FRI signals in the absence of noise. However, empirical observations indicate that, for some noisy FRI signals, substantial performance improvements are achievable when the sampling rate is increased beyond the rate of innovation [9–12]. This leads to two areas of active research on FRI: first, the development of algorithms with improved noise robustness [9–11, 13–17], and, second, the derivation of bounds on the best possible performance at a given noise level [12, 16]. By comparing FRI techniques with performance bounds, we will demonstrate that while noise treatment has improved in recent years, there remain cases in which state-of-the-art techniques can still be enhanced.

1.1.2 History of FRI

The idea of analyzing FRI signals was first proposed by Vetterli et al. [4]. Although the minimal sampling rate required for such settings has been derived, no generic reconstruction scheme exists for the general problem. Nonetheless, some special cases have been treated in previous work, including streams of pulses, which will be our focus in this chapter.

A stream of pulses can be viewed as a parametric signal, uniquely defined by the time-delays of the pulses and their amplitudes. An efficient sampling scheme for *periodic* streams of impulses, having K impulses in each period, was proposed in [4]. Using this technique, one obtains a set of Fourier series coefficients of the

periodic signal. Once these coefficients are known, the problem of determining the time-delays and amplitudes of the pulses becomes that of finding the frequencies and amplitudes of a sum of sinusoids. The latter is a standard problem in spectral analysis [18] which can be solved using conventional approaches, such as the annihilating filter method [18,19], as long as the number of samples is no smaller than $2K$. This result is intuitive since $2K$ is the number of degrees of freedom in each period: K time-delays and K amplitudes.

Periodic streams of pulses are mathematically convenient to analyze, but not very practical. By contrast, *finite* streams of pulses are prevalent in applications such as ultrasound imaging [10]. The first treatment of finite Dirac streams appears in [4], in which a Gaussian sampling kernel was proposed. The time-delays and amplitudes are then estimated from the samples. However, this approach is subject to numerical instability, caused by the exponential decay of the kernel. A different approach, based on moments of the signal, was developed in [9], where the sampling kernels have compact time support. This method treats streams of Diracs, differentiated Diracs, and short pulses with compact support. The moments characterize the input akin to the Fourier coefficients used in [4]. In fact, the time delays and pulses can again be determined from the moments by using standard spectral estimation tools. Another technique that utilizes finite-support sampling kernels, was proposed in [10]. This approach has improved numerical stability, thanks to the choice of the sampling kernel, especially for high rates of innovation. The method was then generalized in [11].

Infinite streams of pulses arise in applications such as ultra-wideband (UWB) communications, where the communicated data changes frequently. Using a compactly supported filter [9], and under certain limitations on the signal, the infinite stream can be divided into a sequence of separate finite problems. The individual finite cases may be treated using methods for the finite setting; however, this leads to a sampling rate that is higher than the rate of innovation. A technique achieving the rate of innovation was proposed in [11], based on a multichannel sampling scheme which uses a number of sampling kernels in parallel.

In related work, a *semi-periodic* pulse model was proposed in [17], wherein the pulse time delays do not change from period to period, but the amplitudes vary. This is a hybrid case in which the number of degrees of freedom in the time-delays is finite, but there is an infinite number of degrees of freedom in the amplitudes. Therefore, the proposed recovery scheme generally requires an infinite number of samples.

The effect of digital noise on the recovery procedure was first analyzed by Maravic and Vetterli [13], where an improved model-based approach was proposed. In this technique, known as the subspace estimator, proper use of the algebraic structure of the signal subspace is exploited, leading to improved noise robustness. An iterative version of the subspace estimator was later proposed by Blu et al. [19]. This approach is optimal for the sinc sampling kernel of [4], but can also be adapted to compactly supported kernels. FRI recovery in the presence of noise was also examined from a stochastic modeling perspective by Tan

and Goyal [14] and by Erdozain and Crespo [15]. The performance in the presence of analog noise has been recently examined in [12]. Treating analog noise allows to analyze the interaction between FRI techniques and the underlying sampling methods. In particular, bounds are obtained which are independent of the sampling method. For different classes of FRI signals, this allows to identify an optimal sampling approach that achieves the bound. In addition, it is shown that under certain scenarios the sampling schemes of [11] are optimal in the presence of analog noise. This framework can also be used to identify FRI settings in which noise-free recovery techniques deteriorate substantially under slight noise levels.

There has also been some work on FRI setups departing from the simple one-dimensional scheme of Fig. 1.1. We have already mentioned multichannel setups, in which sampling is performed simultaneously using several distinct kernels, but with a lower total sampling rate [11, 17, 20]. The problem of recovering an FRI pulse stream in which the pulse shape is unknown was examined in [21]. Some forms of distributed sampling have been studied in [22]. There has also been work on multidimensional FRI signals, i.e., signals which are a function of two or more parameters (such as images) [23, 24]. The many applications of FRI theory include image superresolution [25, 26], ultrasound imaging [10], radar [27], multipath identification [17], and wideband communications [28, 29].

1.1.3 Chapter outline

Throughout the rest of the chapter, we treat the basic concepts underlying FRI theory in greater detail. We mainly focus on FRI pulse streams, and consider in particular the cases of periodic, finite, infinite, and semi-periodic pulse streams. In Section 1.2, we provide a general definition and some examples of FRI signals. In Section 1.3, we treat the problem of recovering FRI signals from noiseless samples taken at the rate of innovation. Specifically, we concentrate on a pulse stream input signal and develop recovery procedures for various types of sampling kernels. Modifications of these techniques when noise is present in the system are discussed in Section 1.4. Simulations illustrating the ability to recover FRI signals are provided in Section 1.5. We conclude the chapter in Section 1.6 with several extensions of the FRI model and a brief discussion of some of its practical application areas.

1.1.4 Notation and conventions

The following notation will be used throughout the chapter. \mathbb{R} , \mathbb{C} and \mathbb{Z} denote the sets of real, complex and integer numbers, respectively. Boldface uppercase letters \mathbf{M} denote matrices, while boldface lowercase letters \mathbf{v} indicate vectors. The identity matrix is denoted \mathbf{I} . The notation $\mathbf{M}_{a \times b}$ explicitly indicates that the matrix is of dimensions $a \times b$. The superscripts $(\cdot)^T$, $(\cdot)^*$, $(\cdot)^{-1}$ and $(\cdot)^\dagger$, when referring to operations on matrices or vectors, mean the transpose, Hermitian

conjugate, inverse, and Moore-Penrose pseudoinverse respectively. Continuous-time functions are denoted $x(t)$, whereas discrete-time sequences are denoted x_n or $x[n]$. The expectation operator is $\mathbb{E}(\cdot)$. The box function $\text{rect}(t)$ equals 1 in the range $[-\frac{1}{2}, \frac{1}{2}]$ and 0 elsewhere. The Heaviside or step function $u(t)$ is 0 for $t < 0$ and 1 for $t \geq 0$.

The continuous-time Fourier transform $\hat{x}(\omega)$ of the function $x(t)$ is defined as

$$\hat{x}(\omega) \triangleq \int_{-\infty}^{\infty} x(t)e^{-jt\omega} dt, \quad (1.2)$$

while the discrete-time Fourier transform (DTFT) of a sequence $a[n]$ is given by

$$\hat{a}(e^{j\omega T}) \triangleq \sum_{n \in \mathbb{Z}} a[n]e^{-j\omega n T}. \quad (1.3)$$

The Fourier series $\{\hat{x}_m\}_{m \in \mathbb{Z}}$ of a τ -periodic function is defined as

$$\hat{x}_m \triangleq \frac{1}{\tau} \int_0^{\tau} x(t)e^{-j2\pi m \frac{t}{\tau}} dt. \quad (1.4)$$

We will also use the Fourier series (1.4) for finite-duration signals, i.e., signals whose support is contained in $[0, \tau]$.

We conclude the section with some identities which will be used in several proofs throughout the chapter. These are the Poisson summation formula [30]

$$\sum_{n \in \mathbb{Z}} x(t + nT) = \frac{1}{T} \sum_{k \in \mathbb{Z}} \hat{x} \left(\frac{2\pi k}{T} \right) e^{j2\pi k \frac{t}{T}} \quad (1.5)$$

and Parseval's theorem for the equivalence of the inner product [30, 31]

$$\langle x(t), y(t) \rangle = \frac{1}{2\pi} \langle \hat{x}(\omega), \hat{y}(\omega) \rangle, \quad (1.6)$$

where $\langle x(t), y(t) \rangle = \int_{-\infty}^{\infty} x^*(t)y(t)dt$.

1.2 Signals with finite rate of innovation

As explained in the beginning of this chapter, FRI signals are those that can be described by a finite number of parameters per time unit. In this section we introduce the original definition as stated by Vetterli et al. in [4]. In addition, we provide some examples of FRI signals that can be sampled and perfectly reconstructed at their rate of innovation using the techniques of [4, 9–11, 17]. We also formally define periodic, semi-periodic and finite duration signals.

1.2.1 Definition of signals with FRI

The concept of FRI is intimately related to parametric signal modeling. If a signal variation depends on a few unknown parameters, then we can see it as having a limited amount of degrees of freedom per unit time.

More precisely, given a set of known functions $\{g_r(t)\}_{r=0}^{R-1}$, arbitrary shifts t_k and amplitudes $\gamma_{k,r}$, consider a signal of the form:

$$x(t) = \sum_{k \in \mathbb{Z}} \sum_{r=0}^{R-1} \gamma_{k,r} g_r(t - t_k). \quad (1.7)$$

Since the set of functions $\{g_r(t)\}_{r=0}^{R-1}$ is known, the only free parameters of the signal are the coefficients $\gamma_{k,r}$ and the time shifts t_k . Consider a counting function $C_x(t_a, t_b)$ that is able to compute the number of parameters over a time interval $[t_a, t_b]$. The rate of innovation is defined as follows

$$\rho = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} C_x \left(-\frac{\tau}{2}, \frac{\tau}{2} \right). \quad (1.8)$$

Definition 1.1. [4] *A signal with Finite Rate of Innovation can be defined as a signal with a parametric representation such as that given by (1.7), and with a finite ρ given by (1.8).*

Another useful concept is that of a *local* rate of innovation over a window of size τ , defined as:

$$\rho_\tau(t) = \frac{1}{\tau} C_x \left(t - \frac{\tau}{2}, t + \frac{\tau}{2} \right), \quad (1.9)$$

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

Given an FRI signal with a rate of innovation ρ , we expect to be able to recover $x(t)$ from ρ samples (or parameters) per unit time. The rate of innovation turns out to have another interesting interpretation in the presence of noise: it is a lower bound on the ratio between the average mean-squared error (MSE) achievable by any unbiased estimator of $x(t)$ and the noise variance, regardless of the sampling method [12].

1.2.2 Examples of FRI signals

It is well-known from classical sampling theory that a signal bandlimited to $[-\frac{B}{2}, \frac{B}{2}]$ can be expressed as an infinite sum of properly weighted and shifted versions of the sinc function:

$$x(t) = \sum_{n \in \mathbb{Z}} x[n] \text{sinc}(Bt - n), \quad (1.10)$$

where $x[n] = \langle x(t), B \text{sinc}(Bt - n) \rangle$. Comparing equations (1.10) and (1.7) immediately reveals that a bandlimited signal can be interpreted as having finite rate of innovation. In this case, we can say that the signal $x(t)$ has B degrees of freedom per second, since it is exactly defined by a sequence of numbers $\{x[n]\}_{n \in \mathbb{Z}}$ spaced $T = B^{-1}$ seconds apart, given that the basis function sinc is known.

This idea can be generalized by replacing the sinc basis function with any other function $\varphi(t)$. The set of signals

$$x(t) = \sum_{n \in \mathbb{Z}} x[n] \varphi(Bt - n), \quad (1.11)$$

defines a shift-invariant subspace, which is not necessarily bandlimited, but that again has a rate of innovation $\rho = B$. Such functions can be efficiently sampled and reconstructed using linear methods [5, 6], and thus typically do not require the more elaborate techniques of FRI theory. However, many FRI families of signals form a union of subspaces [7, 8], rather than a subspace, and can still be sampled and perfectly reconstructed at the rate of innovation. As a motivation for the forthcoming analysis, several examples of such signals are plotted in Fig. 1.2 and described below. For simplicity, these examples describe finite-duration FRI signals defined over the range $[0, 1]$, but the extension to infinite or periodic FRI models is straightforward.

- (i) The first signal of interest is a *stream of K Diracs* with amplitudes $\{a_k\}_{k=0}^{K-1}$ and time locations $\{t_k\}_{k=0}^{K-1}$. Mathematically, $x(t)$ can be written as

$$x(t) = \sum_{k=0}^{K-1} a_k \delta(t - t_k). \quad (1.12)$$

The signal has $2K$ degrees of freedom, because it has K amplitudes and K locations that are unknown. A typical realization of such a signal can be seen in Fig. 1.2(a).

- (ii) A signal $x(t)$ is a *nonuniform spline of order R* with amplitudes $\{a_k\}_{k=0}^{K-1}$ and knots at $\{t_k\}_{k=0}^{K-1} \in [0, 1]$ if and only if its $(R + 1)$ th derivative is a stream of K weighted Diracs. Equivalently, such a signal consists of $K + 1$ segments, each of which is a polynomial of degree R , such that the entire function is differentiable R times. This signal also has $2K$ degrees of freedom, because it is only the K amplitudes and K locations of the Diracs that are unknown. An example is the piecewise linear signal described in Section 1.1 and shown in Fig. 1.2(b). The second derivative of this signal is the train of Diracs shown in (a).
- (iii) A *Stream of K differentiated Diracs* with amplitudes $\{a_{kr}\}_{k=0, r=0}^{K-1, R_k-1}$ and time locations $\{t_k\}_{k=0}^{K-1}$ is similar to the stream of Diracs, but combining linearly a set of properly displaced and weighted differentiated Diracs, $\delta^{(r)}(t)$. Mathematically, we can write:

$$x(t) = \sum_{k=0}^{K-1} \sum_{r=0}^{R_k-1} a_{kr} \delta^{(r)}(t - t_k). \quad (1.13)$$

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

- (iv) A signal $x(t)$ is a *piecewise polynomial* with K segments of maximum degree $R - 1$ ($R > 0$) if and only if its R th derivative is a stream of differentiated

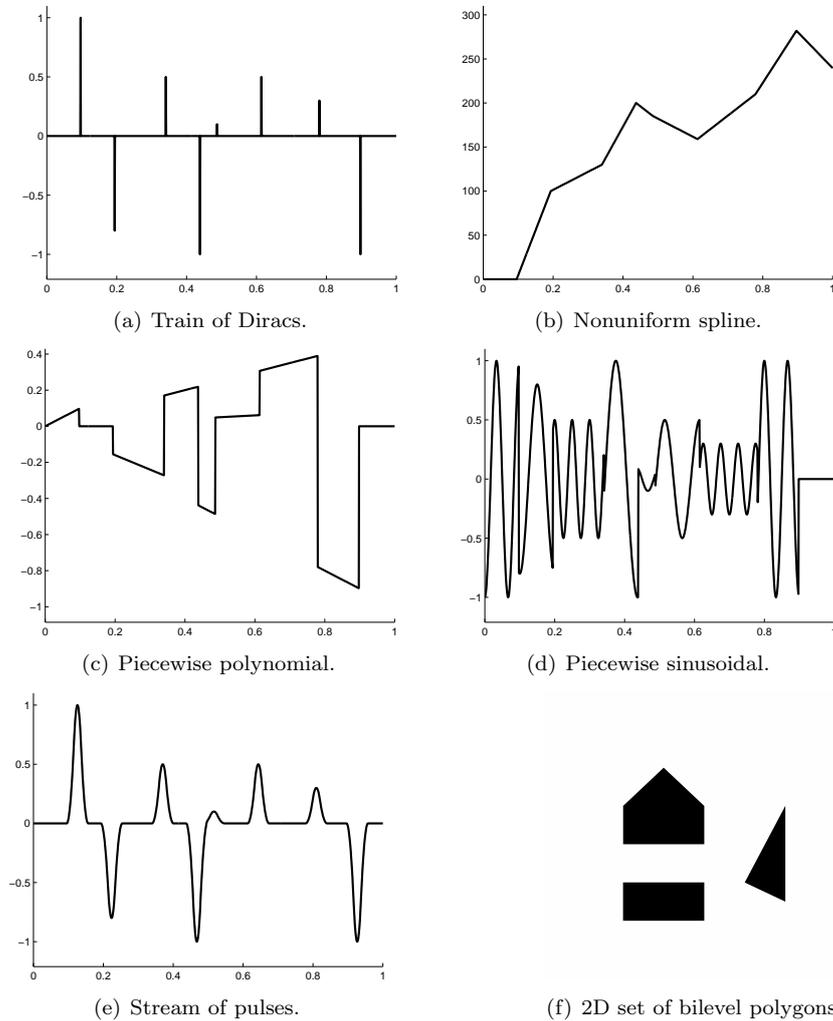


Figure 1.2 Examples of FRI signals that can be sampled and perfectly reconstructed at the rate of innovation.

Diracs. The signal again has $K + \tilde{K}$ degrees of freedom. An example is shown in Fig. 1.2(c). The difference between a piecewise polynomial and a spline is that the former is not differentiable at the knots.

- (v) Another family of signals, considered in [16], are *piecewise sinusoidal* functions. These are a linear combination of truncated sinusoids, with unknown amplitudes a_{kd} , angular frequencies ω_{kd} and phases θ_{kd} , so that

$$x(t) = \sum_{k=0}^{K-1} \sum_{d=0}^{D-1} a_{kd} \cos(\omega_{kd}t - \theta_{kd}) \xi_d(t), \quad (1.14)$$

with $\xi_d(t) = u(t - t_d) - u(t - t_{d+1})$, where t_d are locations to be determined, and $u(t)$ is the Heaviside step function. Fig. 1.2(d) shows an example of such a signal.

- (vi) An important example we focus on in this chapter is a *stream of pulses*, which is uniquely defined by a known pulse shape $p(t)$ and the unknown locations $\{t_k\}_{k=0}^{K-1}$ and amplitudes $\{a_k\}_{k=0}^{K-1}$ that characterize the pulses. The signal can thus be expressed mathematically as

$$x(t) = \sum_{k=0}^{K-1} a_k p(t - t_k). \quad (1.15)$$

The stream of pulses has $2K$ degrees of freedom. A realization of a train of pulses is shown in Fig. 1.2(e).

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

$$f(x, y) = \sum_{k=0}^{K-1} a_k \delta(x - x_k, y - y_k). \quad (1.16)$$

In Fig. 1.2(f) we show another type of two-dimensional signal, a *2D set of bilevel polygons*.

We conclude this section by focusing on streams of pulses, which are the prototypical signals we use from now on in the remaining of the chapter. We thus assume for simplicity a single pulse shape $p(t)$ in (1.7), and describe an infinite-length stream of pulses as

$$x(t) = \sum_{k \in \mathbb{Z}} a_k p(t - t_k). \quad (1.17)$$

Periodic FRI signals turn out to be particularly convenient for analysis, and will be discussed in depth in Section 1.3.1. If we assume that there are only K different time locations $\{t_k\}$ and amplitudes $\{a_k\}$ in (1.17), and that they are repeated every τ , we have

$$x(t) = \sum_{m \in \mathbb{Z}} \sum_{k=0}^{K-1} a_k p(t - t_k - m\tau). \quad (1.18)$$

The total number of parameters determining the signal for each period is thus $2K$, leading to a rate of innovation given by $\frac{2K}{\tau}$.

Another variant is a finite-duration pulse stream, which consists of K pulses, whose shifts $\{t_k\}$ are known to be located within a finite segment of length τ . Under the assumption of a single pulse shape $p(t)$, we can express finite-duration FRI signals as

$$x(t) = \sum_{k=0}^{K-1} a_k p(t - t_k). \quad (1.19)$$

Such signals are of practical relevance, since it is unrealistic to expect any measured signal to continue indefinitely. Here again, a finite number of parameters determines $x(t)$ entirely. In this case we are interested in the local rate of innovation $\rho_\tau = \frac{2K}{\tau}$.

We will also consider semi-periodic signals, which we define as signals of the form

$$x(t) = \sum_{k=0}^{K-1} \sum_{m \in \mathbb{Z}} a_k[m] p(t - t_k - m\tau). \quad (1.20)$$

Such signals are similar to the periodic pulse stream (1.18), with amplitudes that vary from period to period. Signals from this class can be used, for example, to describe the propagation of a pulse with known shape $p(t)$ which is transmitted at a constant rate $\frac{1}{\tau}$ through a medium consisting of K paths. Each path has a constant delay t_k and a time-varying gain $a_k[n]$ [17]. Due to the delays being repeated over the subsequent periods of the signal, estimation in this model is simpler than in the finite or infinite signal cases [11, 12].

1.3 Sampling and recovery of FRI signals in the noise-free setting

In this section, we present the basic mechanisms for reconstruction of pulse stream FRI signals from their low-rate samples. Recovery is achieved by first linearly combining the samples in order to obtain a new set of measurements $\{\hat{x}_m\}$, which represent the Fourier transform of $x(t)$, and then recovering the FRI signal parameters from $\{\hat{x}_m\}$. The latter stage is equivalent to the problem of determining the frequencies of a signal formed by a sum of complex exponentials. This problem has been treated extensively in the array processing literature, and can be solved using conventional tools from spectral estimation theory [18] such as the matrix pencil [32], subspace-based estimators [33, 34], and the annihilating filter [19].

Recovery of FRI signals is most readily understood in the setting of a periodic stream of pulses given by (1.18), and this is therefore the first scenario we explore. We later discuss recovery techniques that use finite-support sampling kernels. These can be used in the finite setting of equation (1.19) as well as the original infinite FRI model of equation (1.7). Finally, we also discuss a technique for recovering semi-periodic signals of the form (1.20).

1.3.1 Sampling using the sinc kernel

Consider a τ -periodic stream of K pulses $p(t)$ at locations $\{t_k\}_{k=0}^{K-1}$ and with amplitudes $\{a_k\}_{k=0}^{K-1}$, as defined in (1.18). The pulse shape is known a-priori, and therefore the signal has only $2K$ degrees of freedom per period.

Since $x(t)$ is periodic it can be represented in terms of its Fourier series coefficients \hat{x}_m as

$$\begin{aligned} x(t) &= \sum_{k=0}^{K-1} a_k \sum_{m \in \mathbb{Z}} p(t - t_k - m\tau) \\ &\stackrel{(a)}{=} \sum_{k=0}^{K-1} a_k \frac{1}{\tau} \sum_{m \in \mathbb{Z}} \hat{p}\left(\frac{2\pi m}{\tau}\right) e^{j2\pi m \frac{t-t_k}{\tau}} \\ &= \sum_{m \in \mathbb{Z}} \hat{x}_m e^{j2\pi m \frac{t}{\tau}}, \end{aligned} \quad (1.21)$$

where in (a) we used Poisson summation formula (1.5), and

$$\hat{x}_m = \frac{1}{\tau} \hat{p}\left(\frac{2\pi m}{\tau}\right) \sum_{k=0}^{K-1} a_k e^{-j2\pi m \frac{t_k}{\tau}}, \quad (1.22)$$

are the Fourier series coefficients of $x(t)$.

If we have direct access to a set \mathcal{K} of M consecutive Fourier coefficients for which $\hat{p}\left(\frac{2\pi m}{\tau}\right) \neq 0$, and $M \geq 2K$, then it is possible to retrieve the $2K$ free parameters $\{a_k, t_k\}$, $k = 0, 1, \dots, K-1$ by using conventional tools from spectral analysis [18] such as Prony's method or the annihilating filter method [18, 19]. To show this fact we first write (1.22) as

$$\hat{x}_m \hat{p}^{-1}\left(\frac{2\pi m}{\tau}\right) = \frac{1}{\tau} \sum_{k=0}^{K-1} a_k u_k^m, \quad (1.23)$$

where $u_k = e^{-j2\pi \frac{t_k}{\tau}}$ and \hat{p}^{-1} denotes the multiplicative inverse of p . Since $p(t)$ is known a-priori, we assume for simplicity of notation that $\hat{p}\left(\frac{2\pi m}{\tau}\right) = 1$ for $m \in \mathcal{K}$; this happens for example when $x(t)$ is a stream of Diracs. Otherwise one must simply divide each measurement by the corresponding value of $\hat{p}\left(\frac{2\pi m}{\tau}\right)$.

In order to find the values u_k in (1.23), let $\{h_m\}_{m=0}^{K-1}$ denote the filter whose z -transform is

$$\hat{h}(z) = \sum_{m=0}^{K-1} h_m z^{-m} = \prod_{m=0}^{K-1} (1 - u_k z^{-1}). \quad (1.24)$$

That is, the roots of $\hat{h}(z)$ equal the values u_k to be found. Then, it follows that:

$$h_m * \hat{x}_m = \sum_{i=0}^K h_i \hat{x}_{m-i} = \sum_{i=0}^K \sum_{k=0}^{K-1} a_k h_i u_k^{m-i} = \sum_{k=0}^{K-1} a_k u_k^m \underbrace{\sum_{i=0}^K h_i u_k^{-i}}_{=0} = 0 \quad (1.25)$$

where the last equality is due to the fact that $\hat{h}(u_k) = 0$. The filter $\{h_m\}$ is called an annihilating filter, since it zeroes the signal \hat{x}_m . Its roots uniquely define the set of values u_k , provided that the locations t_k are distinct.

Assuming without loss of generality that $h_0 = 1$, the identity in (1.25) can be written in matrix/vector form as

$$\begin{pmatrix} \hat{x}_{-1} & \hat{x}_{-2} & \cdots & \hat{x}_{-K} \\ \hat{x}_0 & \hat{x}_{-1} & \cdots & \hat{x}_{-K+1} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{x}_{K-2} & \hat{x}_{K-3} & \cdots & \hat{x}_{-1} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_K \end{pmatrix} = - \begin{pmatrix} \hat{x}_0 \\ \hat{x}_1 \\ \vdots \\ \hat{x}_{K-1} \end{pmatrix} \quad (1.26)$$

which reveals that we need at least $2K$ consecutive values of \hat{x}_m to solve the above system. Once the filter has been found, the locations t_k are retrieved from the zeros u_k of the z -transform in (1.24). Given the locations, the weights a_k can then be obtained by considering for instance K consecutive Fourier-series coefficients in (1.23). For example, if we use the coefficients for $k = 0, 1, \dots, K-1$, then we can write (1.23) in matrix/vector form as follows:

$$\frac{1}{\tau} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ u_0 & u_1 & \cdots & u_{K-1} \\ \vdots & \vdots & \ddots & \vdots \\ u_0^{K-1} & u_1^{K-1} & \cdots & u_{K-1}^{K-1} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{K-1} \end{pmatrix} = \begin{pmatrix} \hat{x}_0 \\ \hat{x}_1 \\ \vdots \\ \hat{x}_{K-1} \end{pmatrix}. \quad (1.27)$$

This is a Vandermonde system of equations that yields a unique solution for the weights a_k since the u_k s are distinct. We thus conclude that the original signal $x(t)$ is completely determined by the knowledge of $2K$ consecutive Fourier coefficients.

However, the Fourier coefficients are not readily available, rather they need to be determined from the samples $y_n = \langle x(t), \varphi(\frac{t}{\tau} - n) \rangle$ (see also Fig. 1.1). In [4], the sampling kernel considered is the sinc function of bandwidth B , where $B\tau$ is assumed to be an *odd* integer. We denote this kernel by $\phi_B(t)$. In this case, the Fourier coefficients can be related to the samples as follows:

$$\begin{aligned} y_n &= \langle x(t), \phi_B(nT - t) \rangle & (1.28) \\ &\stackrel{(a)}{=} \sum_{m \in \mathbb{Z}} \hat{x}_m \langle e^{j2\pi m \frac{t}{\tau}}, \phi_B(nT - t) \rangle \\ &\stackrel{(b)}{=} \frac{1}{2\pi} \sum_{m \in \mathbb{Z}} \hat{x}_m \left\langle \delta\left(\omega - \frac{2\pi m}{\tau}\right), \hat{\phi}_B(\omega) e^{j\omega n T} \right\rangle \\ &= \sum_{m \in \mathbb{Z}} \hat{x}_m \hat{\phi}_B\left(\frac{2\pi m}{\tau}\right) e^{j2\pi n \frac{\tau}{N} \frac{m}{\tau}} \\ &= \frac{1}{B} \sum_{|m| \leq M = \lfloor \frac{B\tau}{2} \rfloor} \hat{x}_m e^{j2\pi \frac{m n}{N}} \end{aligned}$$

where in (a), (1.21) and the linearity of the inner product have been used, and for (b) Parseval's theorem (1.6) has been applied. Equation (1.28) relates the samples y_n and the Fourier series coefficients \hat{x}_m by means of the inverse discrete Fourier transform (IDFT). Thus, calculating the DFT of the samples would directly

yield \hat{x}_m for $|m| \leq M$. Since we need $2K$ consecutive Fourier coefficients and we require $B\tau$ to be an odd number we obtain the requirement $B\tau \geq 2K + 1$.

We summarize the above sampling and recovery discussion by highlighting the main steps necessary for the retrieval of $x(t)$:

- (1) Obtain the Fourier series coefficients \hat{x}_m for $|m| \leq M$. This can be done by calculating the DFT coefficients of the samples using $\hat{y}_m = \sum_{n=0}^{N-1} y_n e^{-j2\pi \frac{nm}{N}}$ and the fact that they relate through $\hat{x}_m = B\hat{y}_m$, $|m| \leq M$.
- (2) Retrieve the coefficients of the filter that annihilates \hat{x}_m . These coefficients can be found writing down (1.25) as a linear system of equations of the form (1.26), which has K equations and K unknowns. There is only one solution to the system, since the filter h_m is unique for the given signal \hat{x}_m .
- (3) Obtain the roots of the filter $\hat{h}(z)$, which yield the values u_k and, therefore, the locations t_k .
- (4) Find the amplitudes a_k using the first K consecutive equations in (1.23). This yields the Vandermonde system of equations (1.27), which also has a unique solution for different values of the locations t_k .

We note that while the mechanism described above correctly identifies the signal parameters in the present setting, it becomes inaccurate if noise is added to the system. Techniques which are better suited to dealing with noise will be discussed in Section 1.4.

For the sake of brevity we have concentrated on the annihilating filter method for retrieving the signal parameters. However, other techniques exist such as the matrix pencil method [32] as well as subspace-based estimators [33, 34]. In the presence of noise the latter methods can provide improved performance compared to the annihilating filter approach [12, 13].

1.3.2 Sampling using the sum of sines kernel

While the above procedure has shown that it is indeed possible to reconstruct exactly a periodic stream of pulses, it has the disadvantage that it uses a sampling kernel of infinite support and slow decay. It is thus natural to investigate whether a similar procedure can be used with alternative, possibly compactly supported, kernels. As we will see shortly, another important advantage of compactly supported kernels is that the resulting methods can be used in conjunction with finite and infinite FRI signals, rather than periodic signals as was the case in Section 1.3.1. Essentially, we are looking for alternative kernels that can still be used to relate the samples y_n to the Fourier coefficients of $x(t)$. This is because we have seen that, given the Fourier coefficients, $x(t)$ can be retrieved using spectral estimation techniques.

Consider for now a periodic FRI signal (1.18). Assuming a generic sampling kernel $g(t)$, we have that [10]

$$\begin{aligned}
 y_n &= \langle x(t), g(t - nT) \rangle & (1.29) \\
 &= \left\langle \sum_{m \in \mathbb{Z}} \hat{x}_m e^{j2\pi m \frac{t}{\tau}}, g(t - nT) \right\rangle \\
 &\stackrel{(a)}{=} \sum_{m \in \mathbb{Z}} \hat{x}_m e^{j2\pi m \frac{nT}{\tau}} \left\langle e^{j2\pi m \frac{t}{\tau}}, g(t) \right\rangle \\
 &\stackrel{(b)}{=} \sum_{m \in \mathbb{Z}} \hat{x}_m e^{j2\pi m \frac{nT}{\tau}} \hat{g}^* \left(\frac{2\pi m}{\tau} \right),
 \end{aligned}$$

where (a) follows from the linearity of the inner product and a change of variable, and (b) is due to the definition (1.2) of the Fourier transform.

Having control over the filter $g(t)$, we now impose the following condition on its Fourier transform:

$$\hat{g}^*(\omega) = \begin{cases} 0, & \omega = \frac{2\pi m}{\tau}, \quad m \notin \mathcal{K}, \\ \text{non-zero}, & \omega = \frac{2\pi m}{\tau}, \quad m \in \mathcal{K}, \\ \text{arbitrary}, & \text{otherwise,} \end{cases} \quad (1.30)$$

where \mathcal{K} is a set of coefficients which will be determined shortly. Then, we have

$$y_n = \sum_{m \in \mathcal{K}} \hat{x}_m e^{j2\pi m \frac{nT}{\tau}} \hat{g}^* \left(\frac{2\pi m}{\tau} \right). \quad (1.31)$$

In general, the system in (1.31) has a unique solution provided the number of samples N is no smaller than the cardinality of \mathcal{K} , which we will call $M = |\mathcal{K}|$. The reason is that, in this case, the matrix defined by the elements $e^{j2\pi m \frac{nT}{\tau}}$ is left-invertible. The idea is that each sample y_n is a combination of the elements \hat{x}_m , and the kernel $g(t)$ is designed to pass the coefficients for $m \in \mathcal{K}$ and suppress those for $m \notin \mathcal{K}$. Note that for any real filter satisfying (1.30), we have that if $m \in \mathcal{K}$, then $-m \in \mathcal{K}$, since by conjugate symmetry $\hat{g} \left(\frac{2\pi m}{\tau} \right) = \hat{g}^* \left(-\frac{2\pi m}{\tau} \right)$.

In the particular situation in which the number of samples N equals M , and when the sampling period T is related to the total period τ by $T = \frac{\tau}{N}$, we can write

$$y_n = \sum_{m \in \mathcal{K}} \hat{g}_m^* \hat{x}_m e^{j \frac{2\pi m n}{N}} \quad (1.32)$$

where $\hat{g}_m^* = \hat{g}^* \left(\frac{2\pi m}{\tau} \right)$. This equation relates the samples y_n and the Fourier coefficients of the input \hat{x}_m through a “weighted” IDFT. This means that calculating the DFT of the samples yields each of the weighted Fourier series coefficients $\text{DFT}\{y_n\} = \hat{y}_m = \hat{g}_m^* \hat{x}_m$ or, equivalently, the coefficients themselves by inversion of each equation, $\hat{x}_m = \hat{g}_m^{*-1} \hat{y}_m$. Thus, sampling with a filter that satisfies (1.30) allows to obtain the Fourier coefficients \hat{x}_m in a simple manner.

It is straightforward to see that one particular case of a filter obeying (1.30) is the sinc function $g(t) = \text{sinc}(Bt)$ with $B = \frac{M}{\tau}$. A family of alternative kernels

satisfying (1.30) was introduced in [10] and is known as the family of Sum of Sincs (SoS). This class of kernels is defined in the frequency domain as

$$\hat{g}(\omega) = \tau \sum_{m \in \mathcal{K}} b_m \operatorname{sinc} \left(\frac{\omega}{\frac{2\pi}{\tau}} - m \right), \quad (1.33)$$

where $b_m \neq 0$ for $m \in \mathcal{K}$. The resulting filter is real valued if $m \in \mathcal{K}$ implies $-m \in \mathcal{K}$ and $b_m = b_{-m}^*$. In the time domain, the sampling kernel is of compact support, and can be written as

$$g(t) = \operatorname{rect} \left(\frac{t}{\tau} \right) \sum_{m \in \mathcal{K}} b_m e^{j2\pi m \frac{t}{\tau}}. \quad (1.34)$$

The filter can be further generalized when using a function $\phi(t)$ instead of the sinc in (1.33). This could be useful when we need a smoother time implementation than the one involving the rect function as in (1.34). A key feature of $g(t)$ is that it is compactly supported in time. This will become important when sampling finite-length FRI signals.

One interesting set of coefficients is $b_m = 1$ for $m = -p, \dots, p$, so that the filter in (1.34) becomes:

$$g(t) = \operatorname{rect} \left(\frac{t}{\tau} \right) \sum_{m=-p}^p e^{j2\pi m \frac{t}{\tau}} = \operatorname{rect} \left(\frac{t}{\tau} \right) D_p \left(\frac{2\pi t}{\tau} \right) \quad (1.35)$$

where $D_p(t)$ is the Dirichlet kernel. It is shown in [10] that under certain conditions this choice is optimal in the presence of noise. Fig. 1.3 shows this kernel together with the one obtained when the coefficients b_m form a Hamming window [10]. Here M is the cardinality of the set \mathcal{K} and $M \geq 2K$. In general, the free parameters $\{b_k\}_{k \in \mathcal{K}}$ may be optimized for different goals.

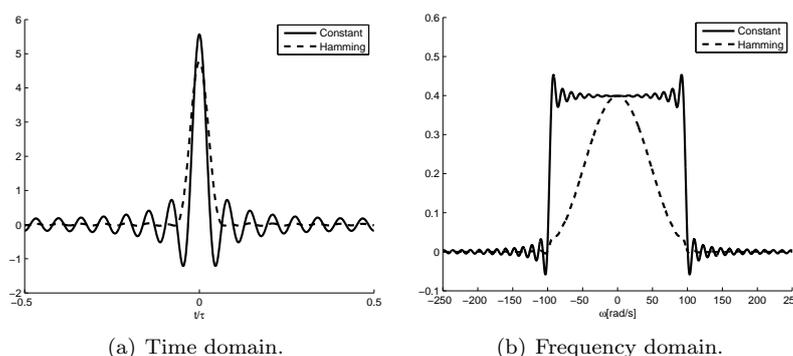


Figure 1.3 *SoS sampling kernels.* The figures show the time and frequency domain representations of the SoS family of kernels given by (1.34) and (1.33) for $b_m = 1, \forall m$ and when the coefficients b_k follow a Hamming window pattern.

To summarize, given the samples y_n , we need to obtain their DFT, and the resulting sequence is related to the Fourier series coefficients \hat{x}_m through (1.32)

(we use $N = M$ and $\tau = NT$). We can then build a system of equations as in (1.26) to determine the annihilating filter coefficients, from which the locations t_k are found calculating by its roots. Finally we build another system of equations like (1.27) to determine the amplitudes a_k , using (1.32).

The fact that the SoS kernels have compact support allows us to depart from the case of periodic signals, facilitating sampling finite and infinite length FRI signals, as discussed below.

Sampling finite streams of pulses

Finite streams of pulses can be processed based on the above analysis for the periodic case. For the finite length scenario, we need to relate the samples obtained from the finite stream of pulses to those of the periodic stream. Let $\tilde{x}(t)$ be a finite FRI signal of the form (1.19). It is shown in [10] that

$$y_n = \langle \tilde{x}(t), \tilde{g}(t - nT) \rangle = \langle x(t), g(t - nT) \rangle \quad (1.36)$$

where $x(t)$ is the periodic continuation of the finite stream $\tilde{x}(t)$, and where we have defined the periodic extension of the filter $g(t)$ as $\tilde{g}(t) = \sum_{m \in \mathbb{Z}} g(t - m\tau)$. Therefore, the set of samples $y_n = \langle x(t), g(t - nT) \rangle$, which uniquely represent a τ -periodic stream of pulses, are equivalent to those that could be obtained by sampling the finite length signal $\tilde{x}(t)$ with the τ -periodic extension of the filter, $\tilde{g}(t - nT)$.

However, it is not practical to use an infinitely long sampling kernel. Assume the pulse $p(t)$ is equal to zero for any $|t| \geq \frac{R}{2}$. Then, the samples have the form [10]

$$y_n = \left\langle \tilde{x}(t), \sum_{m=-r}^r g(t - nT - m\tau) \right\rangle, \quad (1.37)$$

where $r = \left\lceil \frac{\frac{R}{\tau} + 3}{2} \right\rceil - 1$. The advantage of this approach is that we can immediately follow the same retrieval procedure as with the periodic stream of pulses. The reason is that now we obtain the same set of samples given by equation (1.36) sampling the finite length signal $\tilde{x}(t)$ with the finite support kernel

$$g_r(t) = \sum_{m=-r}^r g(t - nT - m\tau). \quad (1.38)$$

Moreover, if the support of $p(t)$ satisfies $R \leq \tau$, then $r = 1$, and the extension of $g(t)$ will contain only three repetitions, i.e. $g_r(t) = g(t) + g(t + \tau) + g(t - \tau)$.

The multichannel sampling scheme of [11] can also be used to sample finite FRI signals. As we will see in Section 1.3.4, the use of a filter (or modulator) bank allows us to avoid forming a delayed pulse as in $g_r(t)$. In cases in which such delays are difficult to implement in hardware, it may be advantageous to use multiple channels without the need for delays.

Sampling infinite-length streams of pulses

A similar technique may also be used to sample and recover infinite length FRI pulse streams of the form

$$x(t) = \sum_{k \in \mathbb{Z}} a_k p(t - t_k). \quad (1.39)$$

Concretely, in this case, we assume the signal is characterized by bursts of maximal duration τ which contain at most K pulses, separated by quiet phases of a certain length. This separation depends on the support of the sampling kernel which, in turn, is related to the pulse shape $p(t)$. For example, in order to sample a finite length stream of Diracs we showed that the filter $g_{3p}(t) = g(t) + g(t \pm \tau)$ was capable of sampling the signal leading to its perfect reconstruction. The support of the filter is 3τ and we then know that, if we want to use a sequential retrieval algorithm for the infinite-length input signal case, the separation of consecutive bursts has to be at least $\frac{3\tau}{2}$. However, this technique requires a sampling rate which is higher than the rate of innovation. Achieving perfect reconstruction for infinite FRI signals from samples taken at the rate of innovation requires a multichannel sampling scheme, and is the subject of Section 1.3.4.

1.3.3 Sampling using exponential reproducing kernels

Another important class of compact support kernels that can be used to sample FRI signals is given by the family of exponential reproducing kernels.

An exponential reproducing kernel is any function $\varphi(t)$ that, together with its shifted versions, can generate complex exponentials of the form $e^{\alpha_m t}$. Specifically,

$$\sum_{n \in \mathbb{Z}} c_{m,n} \varphi(t - n) = e^{\alpha_m t} \quad (1.40)$$

where $m = 0, 1, \dots, P$ and $\alpha_0, \lambda \in \mathbb{C}$. The coefficients are given by $c_{m,n} = \langle e^{\alpha_m t}, \tilde{\varphi}(t - n) \rangle$, where $\tilde{\varphi}(t)$ is the dual of $\varphi(t)$, that is, $\langle \varphi(t - n), \tilde{\varphi}(t - k) \rangle = \delta_{n,k}$. When we use these kernels in the FRI process, the choice of the exponents in (1.40) is restricted to $\alpha_m = \alpha_0 + m\lambda$ with $\alpha_0, \lambda \in \mathbb{C}$ and $m = 0, 1, \dots, P$. This is done to allow the use of the annihilating filter method at the reconstruction stage. This point will be more evident later on.

The theory related to the reproduction of exponentials relies on the concept of E-splines [35]. A function $\beta_\alpha(t)$ with Fourier transform $\hat{\beta}_\alpha(\omega) = \frac{1 - e^{\alpha - j\omega}}{j\omega - \alpha}$ is called an E-spline of first order, with $\alpha \in \mathbb{C}$. The time domain representation of such a function is $\beta_\alpha(t) = e^{\alpha t} \text{rect}(t - \frac{1}{2})$. The function $\beta_\alpha(t)$ is of compact support, and a linear combination of its shifted versions $\beta_\alpha(t - n)$ reproduces the exponential $e^{\alpha t}$. Higher order E-splines can be obtained through convolution of first order ones, e.g., $\beta_{\vec{\alpha}}(t) = (\beta_{\alpha_0} * \beta_{\alpha_1} * \dots * \beta_{\alpha_P})(t)$, where $\vec{\alpha} = (\alpha_0, \alpha_1, \dots, \alpha_P)$. This can also be written in the Fourier domain as follows:

$$\hat{\beta}_{\vec{\alpha}}(\omega) = \prod_{k=0}^P \frac{1 - e^{\alpha_k - j\omega}}{j\omega - \alpha_k}. \quad (1.41)$$

Higher order E-splines are also of compact support and, combined with their shifted versions, $\beta_{\bar{\alpha}}(t-n)$, can reproduce any exponential in the subspace spanned by $\{e^{\alpha_0}, e^{\alpha_1}, \dots, e^{\alpha_P}\}$ [9, 35]. Notice that the exponent α_m can be complex, which indicates that E-splines need not be real. However, this can be avoided by choosing complex conjugate exponents. Fig. 1.4 shows examples of real E-spline functions of orders one to four. Finally, note that the exponential reproduction property is preserved through convolution [9, 35] and, therefore, any function $\varphi(t) = \psi(t) * \beta_{\bar{\alpha}}(t)$, combined with its shifted versions, is also able to reproduce the exponentials in the subspace spanned by $\{e^{\alpha_0}, e^{\alpha_1}, \dots, e^{\alpha_P}\}$.

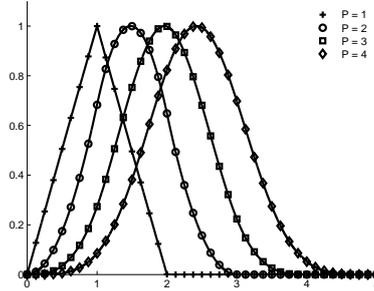


Figure 1.4 Example of exponential reproducing kernels. The shortest function shown is obtained by convolving two first order splines with complex parameters $\pm j\omega_0 = \pm j\frac{2\pi}{N}$ and $N = 32$ samples, resulting in a real function. The successive E-splines, shown in order from left to right, are obtained by convolving kernels with parameters $\alpha_m = j\omega_0(2m - P)$, $m = 0, \dots, P$.

Reconstruction of FRI signals using exponential reproducing kernels is better understood in the time domain. For simplicity, we assume that $p(t)$ is a Dirac function, even though other types of pulses can be sampled and perfectly recovered. In fact, any pulse satisfying $\hat{p}(\omega) \neq 0$ for $\omega = \alpha_m$ can be used. Here α_m , $m = 0, 1, \dots, P$ are the exponents of the exponentials reproduced by the kernel. This is due to the fact that sampling a stream of pulses with the kernel $\varphi(t)$ is equivalent to sampling a stream of Diracs with the kernel $p(t) * \varphi(t)$. The above condition guarantees that $p(t) * \varphi(t)$ is still able to reproduce exponentials.

Consider a finite-duration FRI signal of length τ :

$$x(t) = \sum_{k=0}^{K-1} a_k \delta(t - t_k). \quad (1.42)$$

Assuming a sampling period of $T = \frac{\tau}{N}$, the measurements are

$$y_n = \left\langle x(t), \varphi\left(\frac{t}{T} - n\right) \right\rangle = \sum_{k=0}^{K-1} a_k \varphi\left(\frac{t_k}{T} - n\right), \quad (1.43)$$

for $n = 0, 1, \dots, N - 1$. The E-spline reconstruction scheme, first proposed in [9], operates as follows. The samples are first linearly combined with the coefficients

$c_{m,n}$ of (1.40) to obtain the new measurements

$$s_m = \sum_{n=0}^{N-1} c_{m,n} y_n, \quad m = 0, 1, \dots, P. \quad (1.44)$$

Then, using (1.43), we have that

$$\begin{aligned} s_m &= \left\langle x(t), \sum_n c_{m,n} \varphi\left(\frac{t}{T} - n\right) \right\rangle = \int_{-\infty}^{\infty} x(t) e^{\alpha_m t} dt \\ &= \sum_{k=0}^{K-1} \hat{a}_k u_k^m, \quad m = 0, 1, \dots, P \end{aligned} \quad (1.45)$$

where $\hat{a}_k = a_k e^{\alpha_0 \frac{t_k}{T}}$ and $u_k = e^{\lambda \frac{t_k}{T}}$. Here we have used the fact that $\alpha_m = \alpha_0 + m\lambda$. Note that the new measurements s_m represent the bilateral Laplace transform of $x(t)$ at locations α_m , $m = 0, 1, \dots, P$. These measurements are again in a power sum series form as those discussed in the previous sections. Therefore the pairs of unknowns $\{\hat{a}_k, u_k\}$ can be retrieved from $s_m = \sum_{k=0}^{K-1} \hat{a}_k u_k^m$ using the annihilating filter method. Consequently, the main steps in the reconstruction of FRI signals with exponential reproducing kernels are the same as those discussed previously. The only difference is that the samples were previously combined using a weighted DFT, whereas in this case the linear combination is dictated by the coefficients $c_{m,n}$. Since $2K$ consecutive coefficients s_m are needed to run the annihilating filter method, we have the condition $P \geq 2K - 1$.

We conclude by highlighting the generality of exponential reproducing kernels. First, when the exponent α_m is purely imaginary, that is, when $\alpha_m = j\omega_m$, then $s_m = \hat{x}(\omega_m)$ is precisely the Fourier transform of $x(t)$ at ω_m . Since $x(t)$ is time-limited, this can be thought of as the Fourier series coefficients of the signal. In this case, and for a proper choice of the parameters N and P , it can be shown [36] that the coefficients $c_{m,n}$ constitute a DFT. For this situation the above analysis converges to the one of Section 1.3.1. Moreover, the SoS sampling kernel introduced in Section 1.3.2 is an exponential reproducing kernel of this type. Second, when $\alpha_m = 0$, $m = 0, 1, \dots, P$, the E-spline becomes a polynomial spline (or B-spline). In general, when $\alpha_m = 0$, any exponential reproducing kernel reduces to a kernel satisfying the Strang-Fix conditions [37]. These are still valid sampling kernels but reproduce polynomials rather than exponentials. Functions satisfying Strang-Fix conditions are extensively used in wavelet theory and the above result provides an intriguing connection between sampling of FRI signals and wavelets. This connection allows to combine FRI theory with wavelets to develop efficient centralized and distributed algorithms for the compression of piecewise smooth functions [38,39]. Finally, it is possible to show that any device whose input and output are related by linear differential equations can be turned into an exponential reproducing kernel and can therefore be used to sample FRI signals [9]. This includes, for example, any linear electrical circuit. Given the ubiquity of such devices and the fact that in many cases the sampling kernel is

given and cannot be modified, FRI theory with exponential reproducing kernels becomes even more relevant in practical scenarios.

1.3.4 Multichannel sampling

The techniques discussed so far were based on uniform sampling of the signal $x(t)$ convolved with a single kernel $h(t)$ (see Fig. 1.1). While this is the simplest possible sampling scheme, improved performance and lower sampling rates can be achieved at the cost of slightly more complex hardware. In particular, one can consider a multichannel sampling setup, in which the signal $x(t)$ is convolved with P different kernels $s_1^*(-t), \dots, s_P^*(-t)$, and the output of each channel is sampled at a rate $\frac{1}{T}$ [11, 17]. The set of samples in this case is given by

$$c_\ell[m] = \langle s_\ell(t - mT), x(t) \rangle, \quad \ell = 1, \dots, P, \quad m \in \mathbb{Z}. \quad (1.46)$$

The system is said to have a total sampling rate of $\frac{P}{T}$. Note that the standard (single-channel) scenario is a special case of this scheme, which can be obtained either by choosing $P = 1$ sampling channels, or with $P > 1$ copies of the sampling kernel $h(t)$ which are shifted in time.

An alternative multichannel structure can be obtained in which the filter is replaced by a modulator (i.e. multiplier) followed by an integrator. In this case the output of each branch is given by

$$c_\ell[m] = \int_{(m-1)T}^{mT} x(t) s_\ell(t), \quad \ell = 1, \dots, P, \quad m \in \mathbb{Z}, \quad (1.47)$$

where $s_\ell(t)$ is the modulating function on the ℓ th branch. This scheme is particularly simple, and as we show below, can be used to treat all classes of FRI signals: periodic, finite, infinite, and semi-periodic, under the assumption that the pulse $p(t)$ is compactly supported. In contrast, the filterbank approach is beneficial in particular for semi-periodic pulse streams and can accommodate arbitrary pulse shapes $p(t)$, including infinite-length functions. Furthermore, the multichannel filter bank structure can often be collapsed to a single sampling channel followed by a serial to parallel converter, in order to produce the parallel sampling sequences in (1.46). Thus, when applicable, this scheme may lead to savings in hardware over the modulator based approach, while still retaining the benefits of low sampling rate.

Due to its generality and simplicity, we begin by discussing the modulator-based multichannel structure. The merits of this approach are best exposed by first considering a τ -periodic stream of K pulses.

Before proceeding we note that alternative multichannel systems have been proposed in the literature. In [20] a multichannel extension of the method in [9] was presented. This scheme allows reduced sampling rate in each channel, but the overall sampling rate is similar to [9] and therefore does not achieve the rate of innovation. Two alternative multichannel methods, were proposed in [40] and [41]. These approaches, which are based on a chain of integrators [40] and

exponential filters [41], allow only sampling of infinite streams of Diracs at the rate of innovation. In addition, we show in the simulation section, that these methods are unstable, especially for high rates of innovation.

Periodic FRI signals

Consider a τ -periodic stream of K pulses, as in (1.18). Recall from Section 1.3.1 that if the Fourier coefficients of this signal are available, then standard techniques of spectral analysis can be used to recover the unknown pulse shifts and amplitudes. The multichannel setup provides a simple and intuitive method for obtaining these Fourier coefficients by correlating the signal $x(t)$ with the Fourier basis functions

$$s_\ell(t) = \begin{cases} e^{j\frac{2\pi}{\tau}\ell t}, & t \in [0, \tau], \\ 0, & \text{elsewhere,} \end{cases} \quad (1.48)$$

for $\ell \in \mathcal{L}$, where \mathcal{L} is a set of $2K$ contiguous integers. We set the sampling interval T to be equal to the signal period τ , yielding a total sampling rate of $\frac{2K}{T}$ for all channels. Thus we have a sampling scheme functioning at the rate of innovation, and yielding $2K$ Fourier coefficients of $x(t)$. These can then be used to recover the original signal, for example using the annihilating filter method discussed in Section 1.3.1. An additional advantage of this approach is that the kernels have compact support; indeed, the support corresponds to precisely one period of the FRI signal, which is smaller than the support of the kernel proposed in Section 1.3.2. This property will facilitate the extension of the multichannel system to infinite FRI signals.

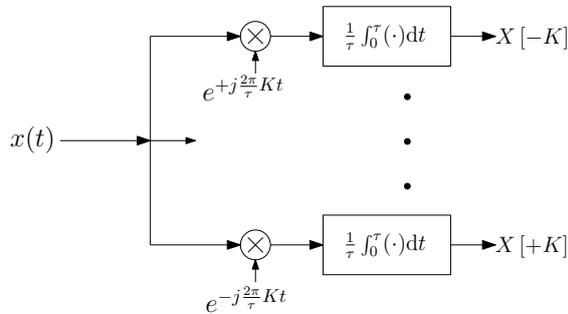


Figure 1.5 Multichannel sampling scheme for periodic FRI signals. The resulting samples are the Fourier series coefficients of $x(t)$. Note that we only sample once every period, thus $T = \tau$.

Instead of functions of the form (1.48), one can just as well use sampling kernels which are a linear combination of these sinusoids, as in Fig. 1.6. This can be advantageous from a hardware point of view, since it may be difficult in practice to implement accurate sinusoids. On the other hand, by allowing such linear combinations, the modulating functions $s_\ell(t)$ can be chosen to have a simple form, such as lowpassed versions of binary sequences [11]. These sequences were shown

to be advantageous in other sub-Nyquist configurations as well, such as the modulated wideband converter, designed to sample wideband signals at sub-Nyquist rates [42, 43], and sampling of stream of pulses with unknown shapes [21]. In addition, in real-life scenarios one or more channels might fail, due to malfunction or noise corruption, and therefore we lose the information stored in that channel. By mixing the coefficients we distribute the information about each Fourier coefficient among several sampling channels. Consequently, when one or more channels fail, the required Fourier coefficients may still be recovered from the remaining operating channels.

When using a mixture of sinusoids, a linear operation is needed to recover the Fourier coefficients from the resulting samples. Specifically, denoting by \mathbf{x} the vector of Fourier coefficients of $x(t)$, the output of Fig. 1.6 is given by $\mathbf{S}\mathbf{x}$ where \mathbf{S} is the matrix of elements s_{ik} . As long as \mathbf{S} has full column rank, we can recover \mathbf{x} from the samples and then proceed using, e.g., the annihilating method to recover the delays and amplitudes. The new kernels retain the desirable property of compact support with length equal to a single signal period. It is also interesting to note that by proper choice of these linear combinations, the modulator bank can implement the SoS filters [11]. This offers an alternative implementation for finite-length FRI signals that avoids the need to form delayed versions of the SoS kernel at the expense of more complex hardware.

Connection to the modulated wideband converter

The concept of using modulation waveforms, is based on ideas which were presented in [42–44]. We now briefly review the sampling problem treated in [43] and its relation to our setup. We also show the practical hardware implementation of both systems is similar. For a more detailed description of this scheme see Chapter 3.

The model in [43] is of multiband signals: signals whose CTFT is concentrated on N_{bands} frequency bands, and the width of each band is no greater than B . The location of the bands is unknown in advance. A low rate sampling scheme allowing recovery of such signals at a rate of $4BN_{\text{bands}}$ was proposed in [45]. This scheme exploits the sparsity of multiband signals in the frequency domain, to reduce the sampling rate well below the Nyquist rate. In [42, 43], this approach was extended to a more practical sampling scheme, which uses a modulation stage and referred to as the modulated wideband converter (MWC). In each channel of the MWC, the input is modulated with some periodic waveform, and then sampled using a lowpass filter (LPF) followed by a low rate uniform sampler. The main idea is that in each channel, the spectrum of the signal is shuffled, such that a portion of the energy of all bands appears at baseband. Mixing the frequency bands in [43] is analogous to mixing the Fourier coefficients in Fig. 1.6.

We note here some differences between the methods. First, following the mixing stage, we use an integrator in contrast to the LPF used in [43]. This is a result of the different signal quantities measured: Fourier coefficients in our work as opposed to the frequency bands content in [43]. The second difference is in the

purpose of the mixing procedure. In [43] mixing is performed to reduce the sampling rate relative to the Nyquist rate. In our setting, the mixing is used in order to simplify hardware implementation and to improve robustness to failure in one of the sampling channels.

Nonetheless, the hardware considerations in the mixing stage in both systems are similar. Recently, a prototype of the MWC has been implemented in hardware [42]. This design is composed of $P = 4$ sampling channels, where the repetition rate of the modulating waveforms is $\frac{1}{T} \approx 20$ MHz. In each period there are 108 rectangular pulses. This prototype, with certain modifications, can be used to implement our sampling scheme as well. These modifications mainly include adding shaping filters on modulating waveforms lines, and reducing the number of rectangular pulses in each period.

Infinite FRI signals

Consider an infinite-duration FRI signal of the form (1.39), where we use $T = \tau$. Furthermore, suppose that the T -local rate of innovation is $\frac{2K}{T}$, for some specified value T . Thus, there are no more than K pulses in any interval of size T , i.e. $I_m = [(m-1)T, mT]$. Assume further that the pulses do not overlap interval boundaries, i.e., if $t_k \in I_m$ then $p(t - t_k) = 0$ for all $t \notin I_m$. Such a requirement automatically holds if $p(t)$ is a Dirac, and will hold with high probability as long as the support of $p(t)$ is substantially smaller than T .

The signal parameters in each interval can now be treated separately. Specifically, consider the T -periodic signal obtained by periodic continuation of the values of $x(t)$ within a particular interval I_m . This periodic signal can be recovered by obtaining $2K$ of its Fourier coefficients. As explained above, these coefficients can be determined using sampling kernels of the form (1.48), whose support is limited to the interval I_m itself (rather than its periodic continuation).

This precise technique can thus be used directly on the non-periodic signal $x(t)$, since the portion of the periodic signal which is sampled includes only the interval I_m [11]. Specifically, this requires obtaining a sample from each of the channels once every T seconds, and using $P \geq 2K$ channels. The resulting procedure is equivalent to a multichannel sampling scheme with rate $\frac{1}{T}$, as depicted in Fig. 1.6. Observe that the success of this technique hinges on the availability of sampling kernels whose support is limited to a single period of the periodic waveform. The output of the channel is equal to $\mathbf{c}[m] = \mathbf{S}\mathbf{x}[m]$ where \mathbf{S} is the matrix of elements s_{ik} , and $\mathbf{x}[m]$ are the Fourier coefficients of $x(t)$ over the interval I_m . We can then invert \mathbf{S} to obtain the Fourier coefficients over each interval.

Semi-periodic FRI signals

The multichannel scheme is also effective for reconstructing FRI signals having the semi-periodic structure of (1.20). That is, signals consisting of K pulses occurring at repeated intervals T , with amplitudes $a_k[m]$ which vary from one period to the next. The modulator approach can be used as in the infinite case,

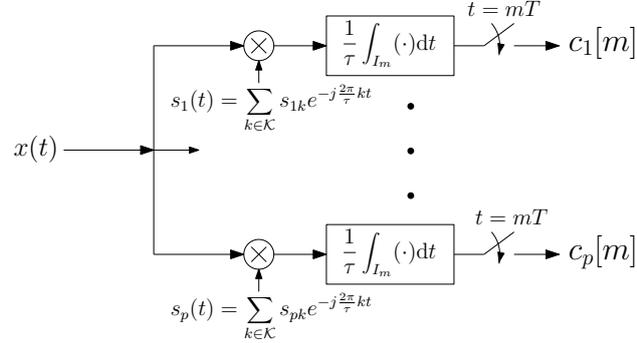


Figure 1.6 Multichannel sampling scheme for infinite FRI signals. Here $T = \tau$.

with the difference that now the samples from different periods can be jointly processed to improve performance.

Specifically, as before, we can recover $\mathbf{x}[m]$ from the output of the modulator bank. Since the delays are constant for each interval I_m , it can be shown (after normalizing the Fourier coefficients by the Fourier coefficients of the pulse if necessary) that in the frequency domain

$$\mathbf{x}[m] = \mathbf{N}(\mathbf{t})\mathbf{a}[m], \quad m \in \mathbb{Z}, \quad (1.49)$$

where $\mathbf{a}[m]$ is the vector of coefficients $a_k[m]$, and $\mathbf{N}(\mathbf{t})$ is the $P \times K$ Vandermonde matrix with $k\ell$ th element $e^{-j2\pi k \frac{t_\ell}{T}}$. When only one time instant m is available, we can solve (1.49) by using the annihilating filter method to recover the delays t_ℓ , and then the coefficients $a_k[m]$. However, now we have many vectors $\mathbf{x}[m]$ that share the same delays, namely, use the same matrix \mathbf{N} . This allows the use of robust methods that recover the delays more reliably, by jointly processing the samples for all m . Examples include the ESPRIT [46] or MUSIC [47] algorithms. These approaches, known as subspace methods, are far more robust than techniques based on a single set of samples. They proceed by computing the correlation matrix $\sum_{m \in \mathbb{Z}} \mathbf{x}[m]\mathbf{x}^T[m]$, and then separate the range of this matrix into two subspaces, the signal and noise subspaces. The delays associated with the matrix \mathbf{N} are then found by exploiting this decomposition.

Clearly, the condition for the general infinite model $P \geq 2K$ is a sufficient condition here as well in order to ensure recovery of $x(t)$. However the additional prior on the signal's structure can be used to reduce the number of sampling channels. In particular it is sufficient to use

$$P \geq 2K - \eta + 1 \quad (1.50)$$

channels, where η is the dimension of the minimal subspace containing the vector set $\{\mathbf{a}[m], m \in \mathbb{Z}\}$. This condition implies that in some cases the number of channels P can be reduced beyond the lower limit $2K$ for the general model.

An alternative scheme for the semi-periodic setting is the filterbank system. The advantage of this technique is that one need not assume the existence of

distinct pulse intervals, nor is it necessary for the pulse shape to have compact support [17]. Here as well we will exploit the periodicity to jointly process the samples by using subspace-methods.

When the pulse shape $p(t)$ is arbitrary, the derivation departs somewhat from the canonical technique presented in Section 1.3.1. This is a result of the fact that the signal is not periodic and cannot be divided into distinct intervals, so that one can no longer speak of its Fourier series. Instead, assume that the sampling interval T equals the signal period τ . The DTFT of the samples (1.46) is then

$$\hat{c}_\ell(e^{j\omega T}) = \frac{1}{T} \sum_{m \in \mathbb{Z}} \hat{s}_\ell^* \left(\omega - \frac{2\pi}{T} m \right) \hat{x} \left(\omega - \frac{2\pi}{T} m \right) \quad (1.51)$$

where the Fourier transform of a function $f(t)$ is denoted $\hat{f}(\omega)$. By computing the Fourier transform of the semi-periodic signal $x(t)$ of (1.20), we have

$$\hat{c}_\ell(e^{j\omega T}) = \sum_{k=0}^{K-1} \hat{a}_k(e^{j\omega T}) e^{-j\omega t_k} \frac{1}{T} \sum_{m \in \mathbb{Z}} \hat{s}_\ell^* \left(\omega - \frac{2\pi}{T} m \right) \hat{p} \left(\omega - \frac{2\pi}{T} m \right) e^{j \frac{2\pi}{T} m t_k}, \quad (1.52)$$

where we used the fact that $\hat{a}_k(e^{j\omega T})$ is $2\pi/T$ -periodic.

Let us restrict our attention to $\omega \in [0, \frac{2\pi}{T})$, which can be done without loss of information since expressions in the DTFT domain are $2\pi/T$ -periodic. Denote by $\hat{\mathbf{c}}(e^{j\omega T})$ the length- P column vector whose ℓ th element is $\hat{c}_\ell(e^{j\omega T})$, and by $\hat{\mathbf{a}}(e^{j\omega T})$ the length- K column vector whose k th element is $\hat{a}_k(e^{j\omega T})$. Also define the vector $\mathbf{t} = (t_0, \dots, t_{K-1})^T$. We can then write (1.52) in matrix form as

$$\hat{\mathbf{c}}(e^{j\omega T}) = \mathbf{M}(e^{j\omega T}, \mathbf{t}) \mathbf{D}(e^{j\omega T}, \mathbf{t}) \hat{\mathbf{a}}(e^{j\omega T}). \quad (1.53)$$

Here $\mathbf{M}(e^{j\omega T}, \mathbf{t})$ is a $P \times K$ matrix whose ℓk th element is

$$\mathbf{M}_{\ell k}(e^{j\omega T}, \mathbf{t}) = \frac{1}{T} \sum_{m \in \mathbb{Z}} \hat{s}_\ell^* \left(\omega - \frac{2\pi}{T} m \right) \hat{p} \left(\omega - \frac{2\pi}{T} m \right) e^{j \frac{2\pi}{T} m t_k}, \quad (1.54)$$

and $\mathbf{D}(e^{j\omega T}, \mathbf{t})$ is a diagonal matrix whose k th diagonal element equals $e^{-j\omega t_k}$.

Defining the vector $\mathbf{b}(e^{j\omega T})$ as

$$\mathbf{b}(e^{j\omega T}) = \mathbf{D}(e^{j\omega T}, \mathbf{t}) \hat{\mathbf{a}}(e^{j\omega T}), \quad (1.55)$$

we can rewrite (1.53) in the form

$$\hat{\mathbf{c}}(e^{j\omega T}) = \mathbf{M}(e^{j\omega T}, \mathbf{t}) \mathbf{b}(e^{j\omega T}). \quad (1.56)$$

Our problem can then be reformulated as that of recovering $\mathbf{b}(e^{j\omega T})$ and the unknown delay set \mathbf{t} from the vectors $\hat{\mathbf{c}}(e^{j\omega T})$, for all $\omega \in [0, \frac{2\pi}{T})$. Once these are known, the vectors $\hat{\mathbf{a}}(e^{j\omega T})$ can be recovered using the relation in (1.55).

To proceed, we focus our attention on sampling filters $\hat{s}_\ell(\omega)$ with finite support in the frequency domain, contained in the frequency range

$$\mathcal{F} = \left[\frac{2\pi}{T} \gamma, \frac{2\pi}{T} (P + \gamma) \right], \quad (1.57)$$

where $\gamma \in \mathbb{Z}$ is an index which determines the working frequency band \mathcal{F} . This choice should be such that it matches the frequency occupation of $p(t)$ (although $p(t)$ does not have to be bandlimited). This freedom allows our sampling scheme to support both complex and real valued signals. For simplicity, we assume here that $\gamma = 0$. Under this choice of filters, each element $\mathbf{M}_{\ell k}(e^{j\omega T}, \mathbf{t})$ of (1.54) can be expressed as

$$\mathbf{M}_{\ell k}(e^{j\omega T}, \mathbf{t}) = \sum_{m=1}^P \mathbf{W}_{\ell m}(e^{j\omega T}) \mathbf{N}_{mk}(\mathbf{t}), \quad (1.58)$$

where $\mathbf{W}(e^{j\omega T})$ is a $P \times P$ matrix whose ℓm th element is given by

$$\mathbf{W}_{\ell m}(e^{j\omega T}) = \frac{1}{T} \hat{s}_{\ell}^* \left(\omega + \frac{2\pi}{T} (m-1 + \gamma) \right) \hat{p} \left(\omega + \frac{2\pi}{T} (m-1 + \gamma) \right),$$

and $\mathbf{N}(\mathbf{t})$ is a $P \times K$ Vandermonde matrix. Substituting (1.58) into (1.56), we have

$$\hat{\mathbf{c}}(e^{j\omega T}) = \mathbf{W}(e^{j\omega T}) \mathbf{N}(\mathbf{t}) \mathbf{b}(e^{j\omega T}). \quad (1.59)$$

If $\mathbf{W}(e^{j\omega T})$ is stably invertible, then we can define the modified measurement vector $\mathbf{d}(e^{j\omega T})$ as $\mathbf{d}(e^{j\omega T}) = \mathbf{W}^{-1}(e^{j\omega T}) \hat{\mathbf{c}}(e^{j\omega T})$. This vector satisfies

$$\mathbf{d}(e^{j\omega T}) = \mathbf{N}(\mathbf{t}) \mathbf{b}(e^{j\omega T}). \quad (1.60)$$

Since $\mathbf{N}(\mathbf{t})$ is not a function of ω , from the linearity of the DTFT, we can express (1.60) in the time domain as

$$\mathbf{d}[n] = \mathbf{N}(\mathbf{t}) \mathbf{b}[n], \quad n \in \mathbb{Z}. \quad (1.61)$$

The elements of the vectors $\mathbf{d}[n]$ and $\mathbf{b}[n]$ are the discrete time sequences, obtained from the inverse DTFT of the elements of the vectors $\mathbf{b}(e^{j\omega T})$ and $\mathbf{d}(e^{j\omega T})$ respectively.

Equation (1.61) has the same structure as (1.49) and can therefore be treated in a similar fashion. Relying on methods such as ESPRIT and MUSIC one can first recover \mathbf{t} from the measurements [17]. After \mathbf{t} is known, the vectors $\mathbf{b}(e^{j\omega T})$ and $\hat{\mathbf{a}}(e^{j\omega T})$ can be found using linear filtering relations by

$$\mathbf{b}(e^{j\omega T}) = \mathbf{N}^\dagger(\mathbf{t}) \mathbf{d}(e^{j\omega T}). \quad (1.62)$$

Since $\mathbf{N}(\mathbf{t})$ is a Vandermonde matrix, its columns are linearly independent, and consequently $\mathbf{N}^\dagger \mathbf{N} = \mathbf{I}_K$. Using (1.55),

$$\hat{\mathbf{a}}(e^{j\omega T}) = \mathbf{D}^{-1}(e^{j\omega T}, \mathbf{t}) \mathbf{N}^\dagger(\mathbf{t}) \mathbf{d}(e^{j\omega T}). \quad (1.63)$$

The resulting sampling and reconstruction scheme is depicted in Fig. 1.7.

Our last step is to derive conditions on the filters $s_1^*(-t), \dots, s_P^*(-t)$ and the function $p(t)$ such that the matrix $\mathbf{W}(e^{j\omega T})$ will be stably invertible. To this end, we decompose the matrix $\mathbf{W}(e^{j\omega T})$ as

$$\mathbf{W}(e^{j\omega T}) = \mathbf{S}(e^{j\omega T}) \mathbf{P}(e^{j\omega T}) \quad (1.64)$$

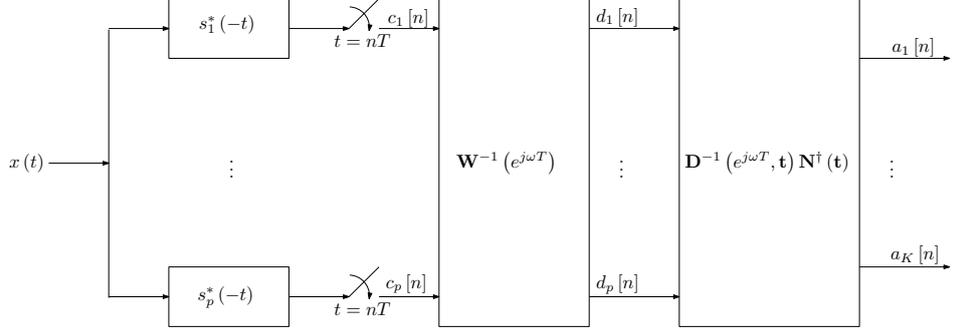


Figure 1.7 Sampling and reconstruction scheme for a semi-periodic signal.

where $\mathbf{S}(e^{j\omega T})$ is a $P \times P$ matrix whose ℓm th element is

$$\mathbf{S}_{\ell m}(e^{j\omega T}) = \frac{1}{T} \hat{s}_\ell^* \left(\omega + \frac{2\pi}{T} (m - 1 + \gamma) \right) \quad (1.65)$$

and $\mathbf{P}(e^{j\omega T})$ is a $P \times P$ diagonal matrix with m th diagonal element

$$\mathbf{P}_{mm}(e^{j\omega T}) = \hat{p} \left(\omega + \frac{2\pi}{T} (m - 1 + \gamma) \right). \quad (1.66)$$

We can guarantee stable invertibility of $\mathbf{W}(e^{j\omega T})$ by ensuring that both $\mathbf{S}(e^{j\omega T})$ and $\mathbf{P}(e^{j\omega T})$ are stably invertible. From (1.66), it is readily seen that the matrix $\mathbf{P}(e^{j\omega T})$ is stably invertible if there exists constants $a, b \in \mathbb{R}$ such that

$$0 < a \leq |\hat{p}(\omega)| \leq b < \infty \text{ almost everywhere } \omega \in \mathcal{F}. \quad (1.67)$$

In addition, the filters $s_\ell^*(-t)$ should be chosen in such a way that they form a stably invertible matrix $\mathbf{S}(e^{j\omega T})$. One example of a set of sampling kernels satisfying this requirement is the ideal bandpass filterbank given by

$$\hat{s}_\ell(\omega) = \begin{cases} T, & \omega \in [(\ell - 1)\frac{2\pi}{T}, \ell\frac{2\pi}{T}], \\ 0, & \text{otherwise.} \end{cases} \quad (1.68)$$

Another example is a LPF with cutoff $\frac{\pi P}{T}$ followed by a uniform sampler at a rate of $\frac{P}{T}$. The samples can then be converted into P parallel streams to mimic the output of P branches. Further discussion of sampling kernels satisfying these requirements can be found in [17].

To summarize, we derived a general technique for the recovery of pulse parameters from a semi-periodic pulse stream. The technique is outlined in Fig. 1.7. This method is guaranteed to perfectly recover the signal parameters from samples taken at a total rate of $\frac{2K}{T}$ or higher, provided that the pulse shape satisfies the stability condition (1.67) and the sampling kernels are chosen so as to yield a stable recovery matrix, for example by using the bandpass filterbank (1.68).

1.4 The effect of noise on FRI recovery

Real-world signals are often contaminated by noise and thus do not conform precisely to the FRI scheme. Furthermore, like any mathematical model, the FRI framework is an approximation which does not precisely hold in practical scenarios, an effect known as mismodeling error. It is therefore of interest to design noise-robust FRI recovery techniques.

Noise may arise both in the analog and digital domains, i.e., before and after sampling, as illustrated in Fig. 1.8. The resulting samples can then be written as

$$\tilde{y}_n = \langle x(t), h(t - nT) \rangle + \epsilon_n \quad (1.69)$$

with ϵ_n being the overall noise introduced in the process.

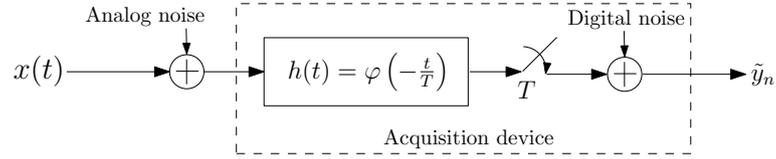


Figure 1.8 Noise perturbations in a “real-world” sampling set-up. The continuous signal $x(t)$ can be corrupted both in the analog and the digital paths.

When noise is present, it is no longer possible to perfectly recover the original signal from its samples. However, one can sometimes mitigate the effects of noise by oversampling, i.e., by increasing the sampling rate beyond the rate of innovation. In Section 1.4.3 we describe several modifications of the recovery techniques of Section 1.3 designed for situations in which a larger number of measurements is available. These are based on the noise model we introduce in Section 1.4.2.

Oversampling increases the number of measurements of the signal, and it is consequently not surprising that this technique can sometimes be used to improve performance under noise. However, the degree to which improvement is possible depends on the setting under consideration. Indeed, in some cases sampling at the rate of innovation is optimal even in the presence of noise, and cannot be improved by oversampling. Reaching such conclusions requires a theoretical analysis of the effects of noise on the ability to recover FRI signals. This issue will be discussed in Sections 1.4.1 and 1.4.2.

1.4.1 Performance bounds under continuous-time noise

In the next two sections, we analyze the effect of noise on the accuracy with which FRI signals can be recovered. A standard tool for accomplishing this is the Cramér–Rao bound (CRB), which is a lower bound on the MSE achievable by any unbiased estimator [48]. As such, it provides a measure of the difficulty of a given estimation problem, and can indicate whether or not existing techniques

come close to optimal. It can also be used to measure the relative merit of different types of measurements. Thus, we will see that the CRB can identify which of the sampling kernels proposed in Section 1.3 provides more robustness to noise, as well as quantify the benefit achievable by oversampling.

As we have already mentioned, in practical applications two types of noise may arise, namely, continuous-time noise which corrupts the signal prior to sampling, and discrete noise contributed by the sampling system (see Fig. 1.8). To simplify the discussion, we separately examine each of these models: we begin below with continuous-time noise and discuss sampling noise in Section 1.4.2. Further details concerning the combined effect of the two sources of noise can be found in [12].

For the purpose of the performance analysis, we focus on finite-duration FRI signals of the form (1.19). Thus, our signal $x(t)$ is determined by a finite number $2K$ of parameters $\{a_k, t_k\}_{k=0}^{K-1}$. For future use, we define the parameter vector

$$\boldsymbol{\theta} = (t_0, \dots, t_{K-1}, a_0, \dots, a_{K-1})^T. \quad (1.70)$$

An important aspect of continuous-time noise is that it is independent of the sampling process. This noise model can thus be used to identify ultimate limits on the achievable estimation accuracy of a given signal. To be specific, suppose we sample the signal

$$y(t) = x(t) + w(t) \quad (1.71)$$

where $x(t)$ is the finite-duration FRI signal (1.19) and $w(t)$ is continuous-time white Gaussian noise with variance σ^2 .

Sampling-indifferent bound

To bound the MSE that can be achieved by any sampling method, it is of interest to derive the CRB for estimating $x(t)$ directly from the continuous-time process $y(t)$. Clearly, no sampling mechanism can do better than exhausting all of the information contained in $y(t)$.

This bound turns out to have a particularly simple closed form expression which depends on the number of pulses in the signal (or, equivalently, on the rate of innovation) — but not on the class of FRI signals being estimated. Indeed, for a signal $x(t)$ of duration τ , it can be shown that the MSE of any unbiased, finite-variance estimator $\hat{x}(t)$ satisfies [12]

$$\frac{1}{\tau} \int \mathbb{E} (|x(t) - \hat{x}(t)|^2) dt \geq \rho_\tau \sigma^2 \quad (1.72)$$

where we recall that the τ -local rate of innovation satisfies $\rho_\tau = \frac{2K}{\tau}$ for finite FRI signals.

Thus, in the noisy setting, the rate of innovation can be given a new interpretation as the ratio between the best achievable MSE and the noise variance σ^2 . This is to be contrasted with the characterization of the rate of innovation in the noise-free case as the lowest sampling rate allowing for perfect recovery of the signal; indeed, when noise is present, perfect recovery is no longer possible.

Bound for sampled measurements

We next consider a lower bound for estimating $x(t)$ from *samples* of the signal $y(t)$ of (1.71). To keep the discussion general, we consider samples of the form

$$\tilde{y}_n = \langle y(t), \varphi_n(t) \rangle, \quad n = 0, \dots, N-1 \quad (1.73)$$

where $\{\varphi_n(t)\}$ is a set of sampling kernels. For example, pointwise sampling at the output of an anti-aliasing filter $\varphi(-t)$ corresponds to the sampling kernels $\varphi_n(t) = \varphi(t - nT)$. We denote by Φ the subspace spanned by the sampling kernels. In this setting, the samples inherit the noise $w(t)$ embedded in the signal $y(t)$. Note that unless the sampling kernels $\{\varphi_n(t)\}$ happen to be orthogonal, the resulting measurements will not be statistically independent. This is a crucial difference with respect to the sampling noise model of Section 1.4.2 below.

We assume that there exists a Fréchet derivative $\frac{\partial x}{\partial \boldsymbol{\theta}}$ which quantifies the sensitivity of $x(t)$ to changes in $\boldsymbol{\theta}$. Informally, $\frac{\partial x}{\partial \boldsymbol{\theta}}$ is an operator from \mathbb{R}^{2K} to the space of square-integrable functions L_2 such that

$$x(t)|_{\boldsymbol{\theta}+\boldsymbol{\delta}} \approx x(t)|_{\boldsymbol{\theta}} + \frac{\partial x}{\partial \boldsymbol{\theta}} \boldsymbol{\delta}. \quad (1.74)$$

Suppose for a moment that there exist elements in the range space of $\frac{\partial x}{\partial \boldsymbol{\theta}}$ which are orthogonal to Φ . This implies that one can perturb $x(t)$ without changing the distribution of the measurements $\tilde{y}_0, \dots, \tilde{y}_{N-1}$. This situation occurs, for example, when the number of measurements N is smaller than the number $2K$ of parameters defining $x(t)$. While it may still be possible to reconstruct some of the information concerning $x(t)$ from these measurements, this is an undesirable situation from an estimation point of view. Thus we will assume that

$$\frac{\partial x}{\partial \boldsymbol{\theta}} \cap \Phi^\perp = \{\mathbf{0}\}. \quad (1.75)$$

Under these assumptions, it can be shown that any unbiased, finite-variance estimator $\hat{x}(t)$ of $x(t)$ from the samples (1.73) satisfies [12]

$$\frac{1}{\tau} \int \mathbb{E} (|x(t) - \hat{x}(t)|^2) dt \geq \frac{\sigma^2}{\tau} \text{Tr} \left[\left(\frac{\partial x}{\partial \boldsymbol{\theta}} \right)^* \left(\frac{\partial x}{\partial \boldsymbol{\theta}} \right) \left(\left(\frac{\partial x}{\partial \boldsymbol{\theta}} \right)^* \mathbf{P}_\Phi \left(\frac{\partial x}{\partial \boldsymbol{\theta}} \right) \right)^{-1} \right] \quad (1.76)$$

where \mathbf{P}_Φ is the projection onto the subspace Φ .

Note that despite the involvement of continuous-time operators, the expression within the trace in (1.76) is a $2K \times 2K$ matrix and can therefore be computed numerically. Also observe that, in contrast to the continuous-time bound of (1.72), the sampled bound depends on the value of $\boldsymbol{\theta}$. Thus, for a specific sampling scheme, some signals can potentially be more difficult to estimate than others.

As expected, the sampled bound (1.76) is never lower than the ultimate (sample-indifferent) bound (1.72). However, the two bounds can sometimes coincide. If this occurs, then at least in terms of the performance bounds, estimators based on the samples (1.73) will suffer no degradation compared with the “ideal”

estimator based on the entire set of continuous-time measurements. Such a situation occurs if $x(t) \in \Phi$ for any feasible value of $x(t)$, a situation which we refer to as “Nyquist-equivalent” sampling. In this case, $\mathbf{P}_\Phi \frac{\partial x}{\partial \theta} = \frac{\partial x}{\partial \theta}$, so that (1.76) reduces to

$$\frac{1}{\tau} \int \mathbb{E} (|x(t) - \hat{x}(t)|^2) dt \geq \frac{\sigma^2}{\tau} \text{Tr}(\mathbf{I}_{2K \times 2K}) = \sigma^2 \rho_\tau \quad (1.77)$$

and the two bounds coincide.

Many practical FRI signal models are not contained in any finite-dimensional subspace, and in these cases, any increase in the sampling rate can improve estimation performance. Even if there exists a subspace containing the entire family of FRI signals, its dimension is often much larger than the number of parameters $2K$ defining the signal; consequently, fully exploiting the information in the signal requires sampling at the Nyquist-equivalent rate, which is potentially much higher than the rate of innovation. This fact provides an analytical explanation of the empirically observed phenomena that oversampling often provides improvement over sampling at the rate of innovation in the presence of noise. A practical example of the benefit of oversampling is described in Section 1.5.

It is interesting to examine this phenomenon from a union of subspaces viewpoint. Suppose that the set of feasible signals \mathfrak{X} can be described as a union of an infinite number of subspaces $\{\mathcal{U}_\alpha\}$ indexed by the continuous parameter α , so that

$$\mathfrak{X} = \bigcup_{\alpha} \mathcal{U}_\alpha. \quad (1.78)$$

In this case, a finite sampling rate captures all of the information present in the signal if and only if

$$\dim \left(\sum_{\alpha} \mathcal{U}_\alpha \right) < \infty \quad (1.79)$$

where $\dim(\mathcal{M})$ is the dimension of the subspace \mathcal{M} . By contrast, in the noise-free case, it has been previously shown [7] that the number of samples required to recover $x(t)$ is given by

$$\max_{\alpha_1, \alpha_2} \dim(\mathcal{U}_{\alpha_1} + \mathcal{U}_{\alpha_2}), \quad (1.80)$$

i.e., the largest dimension among sums of *two* subspaces belonging to the union. In general, the dimension of (1.79) will be much higher than (1.80), illustrating the qualitative difference between the noisy and noise-free settings. For example, if the subspaces \mathcal{U}_α are finite-dimensional, then (1.80) is also necessarily finite, whereas (1.79) need not be.

The bounds developed for analog noise can also be used to optimize the sampling kernels for a given fixed rate. Under certain assumptions, it can be shown that for the case of finite pulse streams, using exponential functions, or Fourier samples, as in the schemes of [10, 11], is optimal. However, in some cases of pulse

shapes the bounds demonstrate that there is room for substantial improvement in the reconstruction stage of these algorithms. Another insight gained from these bounds is that estimation in the semi-periodic setting is far more robust than in the infinite case. As we have discussed, this is because joint processing of the samples is possible. As a rule of thumb, it appears that for union of subspace signals, performance is improved at low rates if most of the parameters identify the position within the subspace, rather than the subspace itself.

1.4.2 Performance bounds under sampling noise

In this section, we derive the CRB for estimating the parameters of a finite-duration FRI signal (1.19) from samples, in the presence of discrete sampling noise. Specifically, we consider unbiased estimators of the parameters $\boldsymbol{\theta}$, as given in (1.70), from the noisy samples

$$\tilde{\mathbf{y}} = (\tilde{y}_0, \dots, \tilde{y}_{N-1})^T, \quad (1.81)$$

which are given by

$$\tilde{y}_n = \left\langle x(t), \varphi \left(\frac{t}{T} - n \right) \right\rangle + \epsilon_n. \quad (1.82)$$

We assume throughout that ϵ_n is white Gaussian noise with variance σ^2 .

This setting is distinct from the scenario discussed in Section 1.4.1 in two respects. First, we now consider noise introduced after the sampling process, rather than continuous-time noise. It is therefore possible to discuss performance bounds only in the context of a given sampling scheme, so that a sampling-indifferent lower bound such as (1.72) is not available in this case. Another implication is that since the noise originates from the sampling process, it is reasonable to assume that the noise in different samples is independent. Second, we consider in this section the problem of estimating the *parameters* $\boldsymbol{\theta}$ defining the signal $x(t)$, rather than the signal itself. Bounds on the recovery of $x(t)$ from samples corrupted by discrete noise can be found in [12].

For concreteness, we focus in this section on the problem of estimating a τ -periodic stream of Diracs, given by

$$x(t) = \sum_{m \in \mathbb{Z}} \sum_{k=0}^{K-1} a_k \delta(t - t_k - m\tau). \quad (1.83)$$

The samples (1.82) then become

$$\tilde{y}_n = \sum_{m \in \mathbb{Z}} \sum_{k=0}^{K-1} a_k \varphi(nT - t_k - m\tau) + \epsilon_n = f(\boldsymbol{\theta}, n) + \epsilon_n. \quad (1.84)$$

Thus, the measurement vector $\tilde{\mathbf{y}}$ has a Gaussian distribution with mean $(f(\boldsymbol{\theta}, 0), \dots, f(\boldsymbol{\theta}, N-1))^T$ and covariance $\sigma^2 \mathbf{I}_{N \times N}$. The CRB is given by [48]

$$\text{CRB}(\boldsymbol{\theta}) = (\mathbf{J}(\boldsymbol{\theta}))^{-1} \quad (1.85)$$

where $\mathbf{J}(\boldsymbol{\theta})$ is the Fisher information matrix

$$\mathbf{J}(\boldsymbol{\theta}) = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \nabla f(\boldsymbol{\theta}, n) \nabla f(\boldsymbol{\theta}, n)^T. \quad (1.86)$$

It follows that the MSE of any unbiased estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ satisfies

$$\mathbb{E} \left\{ \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 \right\} \geq \text{Tr} [(\mathbf{J}(\boldsymbol{\theta}))^{-1}]. \quad (1.87)$$

Note that a very similar technique can be used to obtain bounds on FRI signals composed of arbitrary pulse shapes, as well as periodic FRI signals. The only difference is that the expression for $f(\boldsymbol{\theta}, n)$ becomes more cumbersome.

Comparing sampling kernels in the presence of noise

As an example for which closed-form expressions of the CRB can be obtained, we now consider the special case of estimating the parameters of a periodic stream of Diracs in which each period contains a single pulse. We thus have $K = 1$, and the unknown parameters are $\boldsymbol{\theta} = (t_0, a_0)^T$. While this is a very simple case, the ability to derive a closed form will enable us to reach conclusions about the relative merit of various sampling schemes. In particular, we will compare the bounds obtained using the sinc, B-spline, E-spline, and SoS sampling kernels.

The CRBs for estimating this periodic FRI signal using various kernels are derived in the Appendix. The square root of the resulting MSE is then used to bound the uncertainties in the locations and amplitudes. The expressions obtained for B-splines and E-splines depend on t_0 . We remove this dependency by assuming that t_0 is uniformly distributed over τ and then compute the expected values of the uncertainties. We restrict our analysis to cardinal and trigonometric exponential splines [35]. For all the derivations and the summary given in this section we define the peak signal-to-noise ratio (SNR) as $\text{PSNR} = \left(\frac{a_0}{\sigma}\right)^2$. To obtain a fair comparison between the sampling kernels under consideration the kernels are normalized to have unit norm.

Table 1.1. Summary of the uncertainties on the locations and amplitudes for various sampling kernels. The uncertainties are obtained from the Cramér–Rao Bounds derived in the Appendix.

Kernel	$\frac{\Delta t_0}{\tau} \geq$	$\frac{\Delta a_0}{ a_0 } \geq$
sinc	$\frac{1}{\pi} \sqrt{\frac{\tau}{N} \frac{3}{(B^2 \tau^2 - 1)}} \text{PSNR}^{-\frac{1}{2}}$	$\sqrt{\frac{\tau}{N}} \text{PSNR}^{-\frac{1}{2}}$
B-spline	$\frac{2}{3} \frac{1}{N} \sqrt{\frac{\tau}{N}} \text{PSNR}^{-\frac{1}{2}}$	$\frac{2}{\sqrt{3}} \sqrt{\frac{\tau}{N}} \text{PSNR}^{-\frac{1}{2}}$
E-spline	$\frac{\omega_0 - \cos \omega_0 \sin \omega_0}{\omega_0 \sin \omega_0} \frac{1}{\omega_0 N} \sqrt{\frac{\tau}{N}} \text{PSNR}^{-\frac{1}{2}}$	$\frac{1}{\omega_0} \sqrt{\frac{\omega_0^2 - \cos^2 \omega_0 \sin^2 \omega_0}{\omega_0 \sin^2 \omega_0}} \sqrt{\frac{\tau}{N}} \text{PSNR}^{-\frac{1}{2}}$
SoS	$\frac{1}{2\pi} \sqrt{\frac{\tau}{N} \frac{\sum_{k \in \mathcal{K}} b_k ^2}{\sum_{k \in \mathcal{K}} k^2 b_k ^2}} \text{PSNR}^{-\frac{1}{2}}$	$\sqrt{\frac{\tau}{N}} \text{PSNR}^{-\frac{1}{2}}$

To compare the different kernels, assume that the sinc kernel is chosen with $B\tau = N$ for odd N and $B\tau = N - 1$ for even N , as is commonly accepted [19]. Also assume that the SoS kernel has $b_k = 1$ for all k , which yields optimal results for this kernel [10]. Under these assumptions, it can be seen from Table 1.1 that the uncertainties in the location for all the kernels follow the same trend, up to a constant factor: they are proportional to $\frac{1}{N}\sqrt{\frac{\tau}{N}}\text{PSNR}^{-\frac{1}{2}}$. Thus, performance improves considerably with an increase in the sampling rate (corresponding to larger values of N), and also improves as the square root of the SNR. Interestingly, it can easily be shown that the SoS kernel has precisely the same uncertainty as that of the sinc kernel. To see this, note that $|\mathcal{K}| = 2M + 1$ and that the number of samples has to satisfy $N \geq |\mathcal{K}| \geq 2K$.

Using typical values for the parameters of the results given in Table 1.1 we can compare the performance of the kernels. For instance, assume a fixed interval $\tau = 1$, and constant number of samples $N = 32$, with sampling period $T = \frac{\tau}{N}$ for all the kernels, $b_k = 1, \forall k$ for the SoS, $P = 1$ for the B-spline and $P = 1$ and $\omega_0 = \frac{2\pi}{N}$ for the E-spline, with only $K = 1$ Diracs. In this situation, the sinc and SoS kernels have the best behavior, both in terms of uncertainty in the location and amplitude. For the B-spline and E-spline kernels of lowest possible order ($P = 1$), the uncertainties are almost identical, and slightly worse than optimal. For any support larger than the minimum, the uncertainties achieved by these latter kernels increase.

1.4.3 FRI techniques improving robustness to sampling noise

A central step in each of the reconstruction algorithms examined in Section 1.3 was the search for an annihilating filter $\{h_m\}_{m=0}^K$ which satisfies a given system of linear equations (1.26). This annihilation equation was obtained by observing that $\hat{x} * h = 0$ for any filter $\{h_m\}$ whose z -transform has roots (zeroes) at the values $\{u_k = e^{-j2\pi\frac{tk}{\tau}}\}_{k=0}^{K-1}$. In the noise-free setting it was sensible to choose $\{h_m\}$ having degree K , the lowest possible degree for a filter with K zeroes. However, any filter of degree $L \geq K$ can also be chosen, as long as $\{u_k\}_{k=0}^{K-1}$ are among its L zeroes. Conversely, any filter which annihilates the coefficients $\{\hat{x}_m\}$ is also such that the values u_k are among its zeroes.

When noise is present in the system, we can no longer compute the sequence $\{\hat{x}_m\}$ precisely; instead, we have access only to a noisy version $\{\hat{\hat{x}}_m\}$. On the other hand, since the annihilating equation is satisfied for any contiguous sequence within $\{\hat{x}_m\}$, we can choose to increase the number of measurements, and consequently obtain the sequence $\{\hat{\hat{x}}_m\}$ in the range $-M \leq m \leq M$, for some $M > \frac{L}{2}$. The annihilating equation can then be written as

$$\begin{pmatrix} \hat{\hat{x}}_{-M+L} & \hat{\hat{x}}_{-M+L-1} & \cdots & \hat{\hat{x}}_{-M} \\ \hat{\hat{x}}_{-M+L+1} & \hat{\hat{x}}_{-M+L} & \cdots & \hat{\hat{x}}_{-M+1} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\hat{x}}_M & \hat{\hat{x}}_{M-1} & \cdots & \hat{\hat{x}}_{M-L} \end{pmatrix} \begin{pmatrix} h_0 \\ h_1 \\ \vdots \\ h_L \end{pmatrix} \approx \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (1.88)$$

which has $2M - L + 1$ equations and $L + 1$ unknowns. The equation is not satisfied exactly due to the presence of noise in the measurements $\{\hat{x}_m\}$. Equivalently, we can write the same equation more compactly as

$$\tilde{\mathbf{A}}\mathbf{h} \approx \mathbf{0} \quad (1.89)$$

where the tilde sign in $\tilde{\mathbf{A}}$ serves to remind us of the fact that this matrix contains noisy measurements. We will denote by \mathbf{A} the matrix obtained when we form the same system of equations with noiseless measurements.

Note that we do not require $h_0 = 1$. Indeed, there exist $L - K + 1$ linearly independent polynomials of degree L with zeros at u_k . Thus, there are $L - K + 1$ independent vectors \mathbf{h} that satisfy (1.88). In other words, the rank of $\tilde{\mathbf{A}}$ never exceeds K . This is a key point which forms the basis for many of the methods for signal reconstruction in the presence of noise. We now review two such techniques, namely the total least-squares approach and Cadzow iterative algorithm introduced in [19]. Note that for these techniques to work as explained next, the sampled noise ϵ_n has to be a set of additive, white and Gaussian measurements.

Total least-squares approach

In the presence of noise, the measurements $\{\hat{x}_m\}$ are not known precisely, and one therefore has access only to a noisy version $\tilde{\mathbf{A}}$ of matrix \mathbf{A} , so that the modified annihilating equation (1.89) is true. However, it is reasonable to seek an approximate solution to (1.89) by using the method of total least-squares (TLS) [19], which is defined as the solution to the minimization problem

$$\min_{\mathbf{h}} \|\tilde{\mathbf{A}}\mathbf{h}\|^2 \quad \text{subject to } \|\mathbf{h}\|^2 = 1. \quad (1.90)$$

It is not difficult to show that the filter \mathbf{h} solving (1.90) is given by the singular vector corresponding to the smallest singular value of $\tilde{\mathbf{A}}$. Once the filter \mathbf{h} is found, one can determine its roots and hence identify the time delays, as explained in Section 1.3.

Cadzow iterative denoising algorithm

When the level of noise increases, the TLS approach becomes unreliable. Therefore, it is necessary to use a technique that reduces the noise prior to applying TLS. The idea of the Cadzow technique is to exploit the fact that the noise-free matrix \mathbf{A} is Toeplitz with rank K . Our goal is therefore to find a rank- K Toeplitz matrix \mathbf{A}' which is closest to the noisy matrix $\tilde{\mathbf{A}}$, in the sense of a minimal Frobenius norm. Thus, we would like to solve the optimization problem

$$\min_{\mathbf{A}'} \|\tilde{\mathbf{A}} - \mathbf{A}'\|_F^2 \quad \text{such that } \text{rank}(\mathbf{A}') \leq K \text{ and } \mathbf{A}' \text{ is Toeplitz.} \quad (1.91)$$

To solve (1.91), we employ an algorithm that iteratively updates a target matrix \mathbf{B} until convergence. The iterations alternate between finding the best rank- K approximation and finding the best Toeplitz approximation to \mathbf{B} . Thus,

we must independently solve the two optimization problems

$$\min_{\mathbf{A}'} \|\mathbf{B} - \mathbf{A}'\|_F^2 \quad \text{such that } \text{rank}(\mathbf{A}') \leq K \quad (1.92)$$

and

$$\min_{\mathbf{A}'} \|\mathbf{B} - \mathbf{A}'\|_F^2 \quad \text{such that } \mathbf{A}' \text{ is Toeplitz.} \quad (1.93)$$

The solution to (1.93) is easily obtained by averaging the diagonals of \mathbf{B} . To solve (1.92), we compute the singular value decomposition (SVD) $\mathbf{B} = \mathbf{U}\mathbf{S}\mathbf{V}^*$ of \mathbf{B} , where \mathbf{U} and \mathbf{V} are unitary and \mathbf{S} is a diagonal matrix whose diagonal entries are the singular values of \mathbf{B} . We then discard all but the K largest singular values in \mathbf{S} . In other words, we construct a diagonal matrix \mathbf{S}' whose diagonal contains the K largest entries in \mathbf{S} , and zero elsewhere. The rank- K matrix closest to \mathbf{B} is then given by $\mathbf{U}\mathbf{S}'\mathbf{V}^*$.

The entire iterative algorithm for solving (1.91) can be summarized as follows:

- (1) Let \mathbf{B} equal the original (noisy) measurement matrix $\tilde{\mathbf{A}}$.
- (2) Compute the SVD decomposition of \mathbf{B} such that $\mathbf{B} = \mathbf{U}\mathbf{S}\mathbf{V}^*$, where \mathbf{U} and \mathbf{V} are unitary and \mathbf{S} is diagonal.
- (3) Build the diagonal matrix \mathbf{S}' consisting of the K largest elements in \mathbf{S} , and zero elsewhere.
- (4) Update \mathbf{B} to its best rank- K approximation $\mathbf{B} = \mathbf{U}\mathbf{S}'\mathbf{V}^*$.
- (5) Update \mathbf{B} to its best Toeplitz approximation by averaging over the diagonals of \mathbf{B} .
- (6) Repeat from step (2) until convergence or until a specified number of iterations has been performed.

Applying even a small number of iterations of Cadzow's algorithm will yield a matrix \mathbf{A}' whose error $\|\mathbf{A}' - \mathbf{A}\|_F^2$ is much lower than the error of the original measurement matrix $\tilde{\mathbf{A}}$. This procedure works best when $\tilde{\mathbf{A}}$ is as close as possible to a square matrix [19], and so a good choice would be to use $L = M = \lfloor \frac{B\tau}{2} \rfloor$. The denoised matrix \mathbf{A}' can then be used in conjunction with the TLS technique, as described previously.

1.5 Simulations

In this section we provide some results obtained from implementation of the FRI methods we described. We first show how perfect reconstruction is possible using the proposed kernels in the absence of noise. We then demonstrate the performance of the various kernels when samples are corrupted by additive iid Gaussian noise. In all simulations we consider only real-valued sampling kernels.

1.5.1 Sampling and reconstruction in the noiseless setting

Fig. 1.9 shows an example of the sampling and reconstruction process of Section 1.3.1 for periodic inputs consisting of Diracs. Note that in this setting, using a sinc sampling kernel, or an SoS filter with $b_k = 1$ is equivalent. In Fig. 1.9(a) we show the original and reconstructed signals plotted together, while in Fig. 1.9(b) we plot the filtered input and the samples taken at a uniform interval.

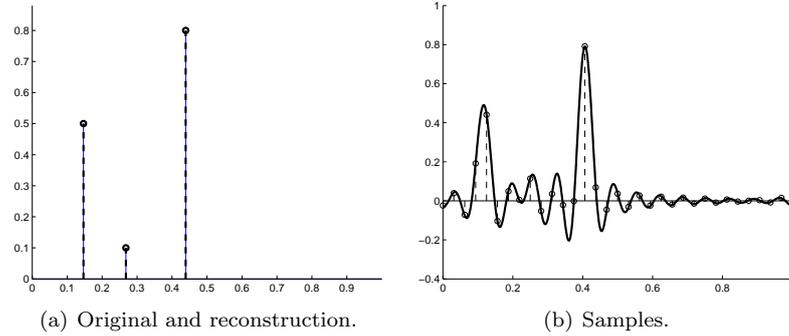


Figure 1.9 Example of sampling and reconstruction of a Stream of Diracs with a sinc kernel. (a) The original signal along with its reconstruction, exact to numerical precision. (b) Convolution of the sinc kernel with the input. The samples, taken at uniform intervals of T seconds, are also indicated.

Fig. 1.10 shows perfect reconstruction of $K = 4$ closely spaced Diracs using a real-valued E-spline. Here again, reconstruction is exact to numerical precision.

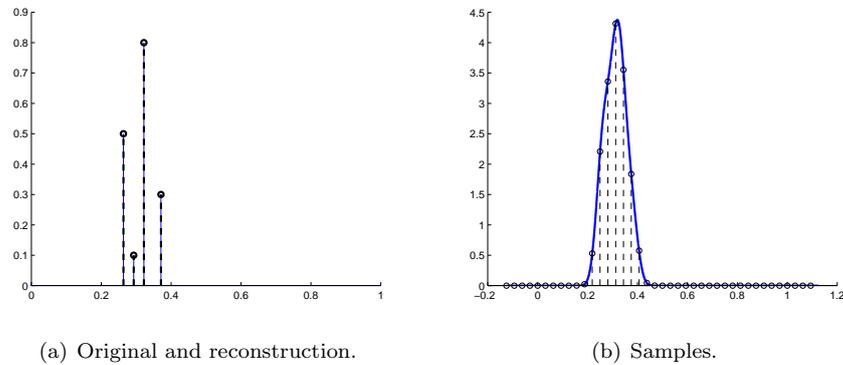


Figure 1.10 Sampling and reconstruction of $K = 4$ closely spaced Diracs with the E-spline kernel. (a) The original signal along with its reconstruction, exact to numerical precision. (b) Convolution of the E-spline kernel with the input. The samples, taken at uniform intervals of T seconds, are also indicated.

As a final example, consider a periodic input $x(t)$ in which each period consists of $K = 5$ delayed and weighted versions of a Gaussian pulse, with $\tau = 1$. We select the amplitudes and locations at random. Sampling is performed using an

SoS kernel with indices $\mathcal{K} = -K, \dots, K$ and cardinality $M = |\mathcal{K}| = 2K + 1 = 11$. We filter $x(t)$ with $g(t)$ defined in (1.34), and set the coefficients $b_k, k \in \mathcal{K}$ to be a length- M symmetric Hamming window. The output of the filter is sampled uniformly N times, with sampling period $T = \frac{\tau}{N}$, where $N = M = 11$. The sampling process is depicted in Fig. 1.11(b). The reconstructed and original signals are depicted in Fig. 1.11(a). Once again the estimation and reconstruction are exact to numerical precision.

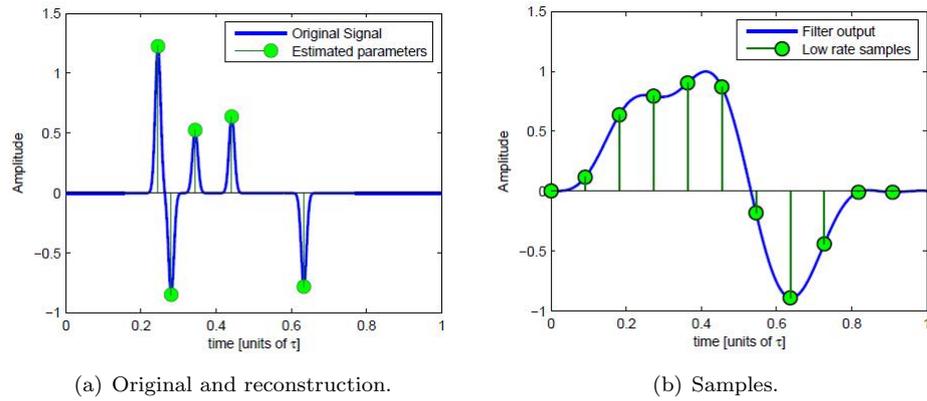


Figure 1.11 Example of sampling and reconstruction of a Stream of Pulses with an SoS kernel. (a) The train of pulses, with Gaussian shape, and the estimated parameters. (b) Convolution of the input with the SoS kernel, and the samples taken at uniform intervals.

1.5.2 Sampling and reconstruction in the presence of noise

In the presence of noise, exact retrieval of the input signal is no longer possible. In order to obtain reasonable recovery, it is necessary to employ some denoising strategies, such as those explained in Section 1.4.

Periodic pulse streams

We start by showing that the proposed robust reconstruction strategies can achieve the CRBs on digital noise given in Section 1.4.2 for a wide range of SNRs. We concentrate on the SoS kernel with coefficients $b_k = 1$. Notice that in this case, the SoS is the Dirichlet function and is therefore equivalent to the periodic sinc of [19]. Fig. 1.12 shows the results of the SoS kernel when the input is a periodic train of $K = 3$ Diracs, and the samples are corrupted by iid Gaussian noise with SNR = 10dB, where we define the SNR as $\text{SNR} = \frac{\|y\|_2^2}{N\sigma^2}$, for a single realization. We use $M = \lfloor \frac{B\tau}{2} \rfloor$, and 20 iterations of Cadzow. The result shows that there is a very small error in the estimated locations despite the fairly low SNR. There is, however, a bigger error when estimating the amplitudes. This happens because the kernel is optimized to minimize the error in the estimation of the location of the Diracs rather than in the amplitude.

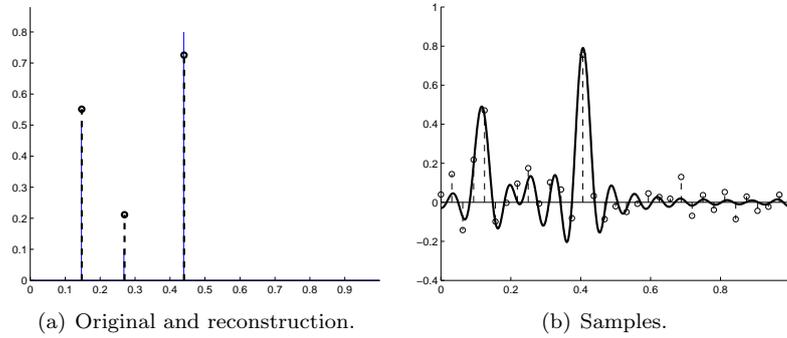


Figure 1.12 Example of sampling and reconstruction of a Stream of Diracs with an SoS kernel. (a) The original signal to be sampled and its reconstruction, overlapping the input. (b) Convolution of the kernel with the input. The noisy samples are also shown.

Now we consider a periodic stream with a single Dirac (e.g. $K = 1$). In the simulations, the amplitude of the Dirac is fixed. The samples are corrupted by iid Gaussian noise with variance σ^2 such that the SNR ranges from -10dB to 30dB. We define the error in time-delay estimation as the average over all experiments of $\|\mathbf{t} - \hat{\mathbf{t}}\|_2^2$, where \mathbf{t} and $\hat{\mathbf{t}}$ denote the true and estimated time-delays, respectively, sorted in increasing order. We then calculate the square root of the average to obtain the MSE, which equals the standard deviation for unbiased estimators. Fig. 1.13 shows the results obtained from averaging 10000 realizations and using 10 iterations of Cadzow's algorithm. More specifically, 1.13(a) shows the estimated positions with respect to the real location and 1.13(b) the estimation error compared to the deviation predicted by the CRB. The retrieval of the FRI signal made of one Dirac is almost optimal for SNR levels above 5dB since the uncertainty on these locations reaches the (unbiased) theoretical minimum given by Cramér–Rao bounds.

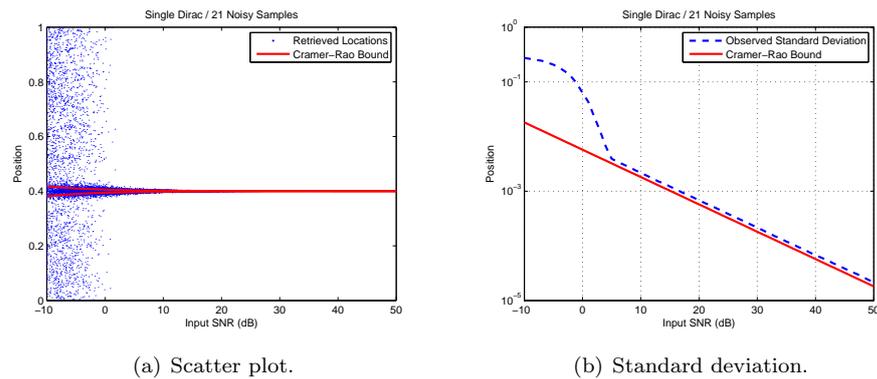


Figure 1.13 Retrieval of the locations of a FRI signal. (a) Scatterplot of the locations. (b) Standard deviation (averaged over 10,000 realizations) compared to the Cramér–Rao lower bound.

The reconstruction quality can be further improved at the expense of oversampling. This is illustrated in Fig. 1.14 where two Diracs are reconstructed. Here we show recovery performance for oversampling factors of 2, 4 and 8.

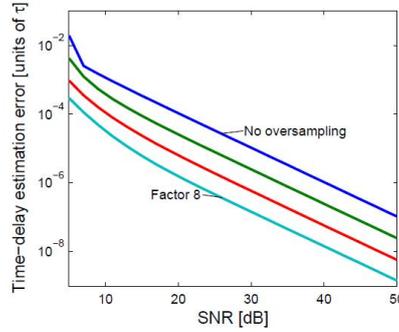


Figure 1.14 *Effect of oversampling.* The performance of the recovery improves for all SNR when more samples are available.

In the following simulation, we consider exponential reproducing kernels and analyze their performance in the presence of noise. Any exponential reproducing kernel is of the form $\varphi(t) = \psi(t) * \beta_{\vec{\alpha}}(t)$ where $\beta_{\vec{\alpha}}(t)$ is the E-spline with exponents $\vec{\alpha} = \{\alpha_0, \alpha_1, \dots, \alpha_P\}$ and $\psi(t)$ can essentially be any function, even a distribution. The aim here is to understand how to set both $\vec{\alpha}$ and $\psi(t)$ in order to have maximum resilience to noise when noise is additive iid Gaussian as assumed so far. It turns out that the best choice of the exponents is $\alpha_m = j2\pi \frac{m}{N}$ [36]. The choice of $\psi(t)$ is not unique and depends on the desired support of $\varphi(t)$. If $\varphi(t)$ has the same support as the SoS kernel, then the best choice of $\psi(t)$ leads to an exponential reproducing kernel with the property that its coefficients $c_{m,n}$ constitute a DFT. Moreover, when the order P of the resulting exponential reproducing kernel equals $P = N - 1$ then the kernel behaves like the Dirichlet function [36]. The simulation results of Fig. 1.15 confirm the above analysis. Here we retrieve two Diracs in the presence of noise using an E-spline with arbitrary exponents ($P = 9$, d), an E-spline with the correct exponents $\alpha_m = j2\pi \frac{m}{N}$ ($P = 9$, o), and finally using two of the most stable exponential reproducing kernels ($P = 15, 30$, n) (the best being the Dirichlet function). We use the notation “d” to indicate default kernel, “o” orthogonal rows of coefficients and “n” orthonormal rows.

Here, we have $N = 31$ samples and the input $x(t)$ is a τ -periodic stream of Diracs, where $\tau = 1$ second. We run 1000 experiments contaminating the samples with iid Gaussian noise of desired SNR by controlling its variance, and we denoise the calculated moments doing 30 iterations of Cadzow’s denoising algorithm. We can see the improvement in performance by going from the first to the last type of exponential reproducing kernel. In fact, as expected, proper choice of the exponents α_m improves the estimation of the locations, and the appropriate choice of $\psi(t)$ enhances the results further. Interestingly, if we use pure E-splines $\beta_{\vec{\alpha}}(t)$ then there is an order from which the performance declines. In

Fig. 1.15 we plot the optimum order ($P = 9$). In contrast, when we design the optimum exponential reproducing kernel the performance improves constantly until it matches that of the Dirichlet kernel.

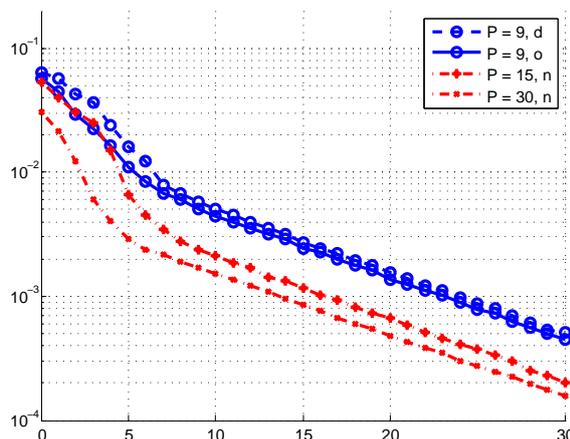


Figure 1.15 *Sampling with exponential reproducing kernels.* Results of the estimation of the location of $K = 2$ Diracs in the presence of noise. The performance of exponential reproducing kernels can be enhanced by proper selection of the parameter α_m (solid line with -o-) and depending on the choice of $\psi(t)$ (dashed-dotted lines).

Finite pulse streams

We now turn to demonstrate FRI recovery methods when using finite pulse streams. We examine 4 scenarios, in which the signal consists of $K = 2, 3, 5, 20$ Diracs¹. In our setup, the time-delays are equally distributed in the window $[0, \tau)$, with $\tau = 1$, and remain constant throughout the experiments. All amplitudes are set to one. The index set of the SoS filter is $\mathcal{K} = \{-K, \dots, K\}$. Both B-splines and E-splines are taken of order $2K - 1$, and for E-splines we use purely imaginary exponents, equally distributed around the complex unit circle. The sampling period for all methods is $T = \frac{\tau}{N}$, where the number of samples is $N = 2K + 1 = 5, 7, 11, 41$ for the SoS and $N = 2K + 1 + S = 9, 13, 21, 81$ for the spline-based methods, where S is the spline support. Hard thresholding was implemented in order to improve the spline methods. The threshold was chosen to be 3σ , where σ is the standard deviation of the noise. For the Gaussian sampling kernel the parameter σ_g was optimized and took on the value of $\sigma_g = 0.25, 0.28, 0.32, 0.9$, respectively.

The results are given in Fig. 1.16. For $K = 2$ all methods are stable, where E-splines exhibit better performance than B-splines, and Gaussian and SoS

¹ Due to computational complexity of calculating the time-domain expression for high order E-splines, the functions were simulated up to order 9, which allows for $K = 5$ pulses.

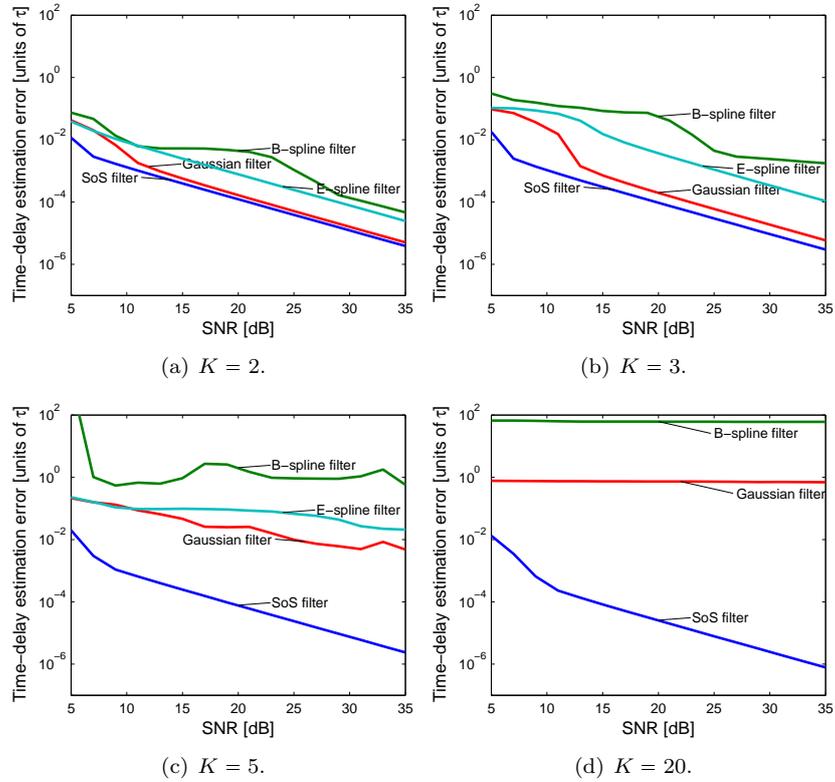


Figure 1.16 Performance in the presence of noise: finite stream case. SoS, B-spline, E-spline and Gaussian sampling kernels. (a) $K = 2$ Dirac pulses are present, (b) $K = 3$ pulses, (c) high value of $K = 5$ pulses, and (d) the performance for a very high value of $K = 20$ (without E-spline simulation, due to computational complexity of calculating the time-domain expression for high values of K).

approaches demonstrate the lowest errors. As the value of K grows, the advantage of the SoS filter becomes more prominent, where for $K \geq 5$, the performance of Gaussian and both spline methods deteriorate and have errors approaching the order of τ . In contrast, the SoS filter retains its performance nearly unchanged even up to $K = 20$, where the B-spline and Gaussian methods are unstable.

Infinite pulse streams

We now demonstrate the performance of FRI methods for infinite pulse streams in the presence of white Gaussian noise, when working at the rate of innovation. We compare three methods that can achieve the innovation rate in the infinite case: an integrator based approach detailed in [40], exponential filters [41], and the multichannel approach described in Section 1.3.4 based on modulators. For the modulators, we examine three waveforms: cosine and sine waveform (tones), filtered rectangular alternating pulses (rectangular) and waveforms obtained from delayed versions of the SoS filter (SoS). Following [41], the parameters

defining the impulse response of the exponential filters are chosen as $\alpha = 0.2T$ and $\beta = 0.8T$.

We focus on one period of the input signal, which consists of $K = 10$ Diracs with times chosen in the interval $[0, T)$ and amplitudes equal one, and $P = 21$ channels. The estimation error of the time-delays versus the SNR is depicted in Fig. 1.17, for the various approaches. The instability of the integrators and exponential filters based methods becomes apparent for these high orders. The SoS approach, in contrast, achieves good estimation results. There is a slight advantage for the schemes based on tones and SoS, over alternating pulses, where the first two configurations have similar performance.

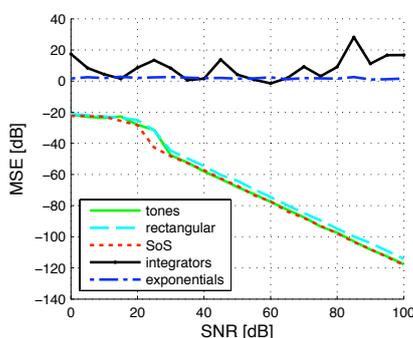


Figure 1.17 Performance in the presence of noise at the rate of innovation. The signal consists of $K = 10$ Diracs.

1.5.3 Periodic vs. semi-periodic FRI signals

As we have seen above, the reconstruction of signals of the form (1.18) in the presence of noise is often severely hampered when sampled at or slightly above the rate of innovation. Rather than indicating a lack of appropriate algorithms, in many cases this phenomenon results from fundamental limits on the ability to recover such signals from noisy measurements. A similar effect was demonstrated [10] in the finite pulse stream model (1.19). On the other hand, some types of FRI signals exhibit remarkable noise resilience, and do not appear to require substantial oversampling in the presence of noise [17]. As we now show, the CRB for analog noise can be used to verify that such phenomena arise from a fundamental difference between families of FRI signals.

As an example, we compare the CRB for reconstructing the periodic signal (1.18) with the semi-periodic signal (1.20). Recall that in the former case, each period consists of pulses having unknown amplitudes and time shifts. By contrast, in the latter signal, the time delays are identical throughout all periods, but the amplitudes can change from one period to the next.

While these are clearly different types of signals, an effort was made to form a fair comparison between the reconstruction capabilities in the two cases. To

this end, we chose an identical pulse $g(t)$ in both cases. We selected the signal segment $[0, \tau]$, where $\tau = 1$, and chose the signal parameters so as to guarantee an identical τ -local rate of innovation. We also used identical sampling kernels in both settings: specifically, we chose the kernels which measure the N lowest frequency components of the signal.

To simplify the analysis and focus on the fundamental differences between these settings, we will assume in this section that the pulses $p(t)$ are compactly supported, and that the time delays are chosen such that pulses from one period do not overlap with other periods. For the periodic signal, we chose $K = 10$ pulses with random delays and amplitudes. A period of $\tau = 1$ was selected. This implies that the signal of interest is determined by $2K = 20$ parameters (K amplitudes and K time delays). To construct a semi-periodic signal with the same number of parameters, we chose a period of $T = \frac{1}{9}$ containing $K = 2$ pulses. The segment $[0, \tau]$ then contains precisely $M = 9$ periods, for a total of 20 parameters. While it may seem plausible to require the same number of periods for both signals, this would actually disadvantage the periodic approach, as it would require the estimation of much more closely-spaced pulses.

Note that since the number of parameters to be estimated is identical in both signal models, the continuous-time CRB for the two settings coincides (see Section 1.4.1). Consequently, for a large number of measurements, the sampled bounds also converge to the same values. However, when the number of samples is closer to the rate of innovation, the bound on the reconstruction error for the semi-periodic signal is much lower than that of the periodic signal, as shown in Fig. 1.18. As mentioned above, this is in agreement with previously reported findings for the two types of signals [4, 11, 17].

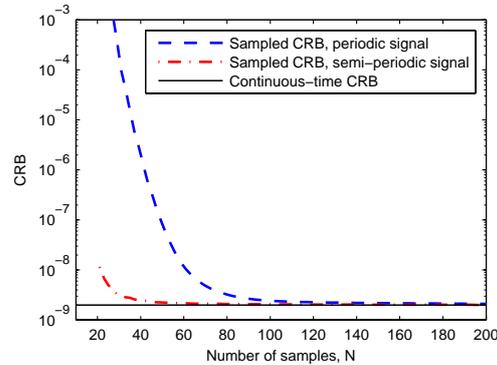


Figure 1.18 Comparison between the CRB for a periodic signal (1.18) and a semi-periodic signal (1.20).

To find an explanation for this difference, it is helpful to recall that both signals can be described using the union of subspaces viewpoint. Each of the signals in this experiment is defined by precisely 20 parameters, which determine the subspace to which the signal belongs and the position within this subspace. Specifi-

cally, the values of the time delays select the subspace, and the pulse amplitudes define a point within this subspace. Thus, in the above setting, the periodic signal contains 10 parameters for selecting the subspace and 10 additional parameters determining the position within it; whereas for the semi-periodic signal, only 2 parameters determine the subspace while the remaining 18 parameters set the location in the subspace. Evidently, identification of the subspace is challenging, especially in the presence of noise, but once the subspace is determined, the remaining parameters can be estimated using a simple linear operation (a projection onto the chosen subspace). Consequently, if many of the unknown parameters identify the position within a subspace, estimation can be performed more accurately. This may provide an explanation for the difference between the two examined signal models.

1.6 Extensions and applications

1.6.1 Sampling piecewise sinusoidal signals

While most of the previous sections have concentrated on sampling streams of pulses, the theory of FRI extends beyond this class of signals and can be applied, for instance, to sample piecewise polynomial signals or classes of 2-D signals. In this section we demonstrate that piecewise sinusoidal signals can also be sampled and perfectly reconstructed using FRI theory [16].

The signals we consider can be written as follows:

$$x(t) = \sum_{d=1}^D \sum_{n=1}^N A_{d,n} \cos(\omega_{d,n}t + \theta_{d,n}) \xi_d(t),$$

where $\xi_d(t) = u(t - t_d) - u(t - t_{d+1})$ and $-\infty < t_1 < \dots < t_d < \dots < t_{D+1} < \infty$. Namely, we consider piecewise sinusoidal signals with a maximum of D pieces and with a maximum of N sinusoids per piece. Piecewise sinusoidal signals are traditionally difficult to sample because they are not bandlimited, have information concentrated both in time and frequency (e.g., time location of the switching points, frequency of each sine wave) and finally cannot be sparsely described in a basis or a frame. However, they are completely specified by a finite number of parameters and are therefore FRI signals.

We assume, for simplicity, that the signal $x(t)$ is acquired using an exponential reproducing kernel, however, similar analysis applies to the sinc and SoS kernels. We have seen in (1.44) and (1.45) that given the samples y_n the new measurements $s_m = \sum_{n=0}^{N-1} c_{m,n} y_n$, $m = 0, 1, \dots, P$ correspond to the Laplace transform of $x(t)$ evaluated at $\alpha_m = \alpha_0 + m\lambda$. In the piecewise sinusoidal case the Laplace transform is given by:

$$s_m = \sum_{d=1}^D \sum_{n=1}^{2N} \bar{A}_{d,n} \frac{[e^{t_{d+1}(j\omega_{d,n} + \alpha_m)} - e^{t_d(j\omega_{d,n} + \alpha_m)}]}{(j\omega_{d,n} + \alpha_m)}, \quad (1.94)$$

where $\bar{A}_{d,n} = A_{d,n}e^{j\theta_{d,n}}$. We now define the polynomial $Q(\alpha_m)$ as follows:

$$Q(\alpha_m) = \prod_{d=1}^D \prod_{n=1}^{2N} (j\omega_{d,n} + \alpha_m) = \sum_{j=0}^J r_j \alpha_m^j. \quad (1.95)$$

Multiplying both sides of the equation by $Q(\alpha_m)$ we obtain:

$$Q(\alpha_m)s_m = \sum_{d=1}^D \sum_{n=1}^{2N} \bar{A}_{d,n} R(\alpha_m) [e^{t_{d+1}(j\omega_{d,n} + \alpha_m)} - e^{t_d(j\omega_{d,n} + \alpha_m)}], \quad (1.96)$$

where $R(\alpha_m)$ is a polynomial. Since $\alpha_m = \alpha_0 + \lambda m$ the right-hand side of (1.96) is a power-sum series and can be annihilated:

$$Q(\alpha_m)s_m * h_m = 0. \quad (1.97)$$

More precisely the right hand side of (1.96) is equivalent to $\sum_{d=1}^D \sum_{r=0}^{2DN-1} b_{r,d} m^r e^{\lambda t_d m}$ where $b_{r,d}$ are weights that depend on α_m but do not need to be computed here. Therefore a filter of the type:

$$\hat{h}(z) = \prod_{d=1}^D (1 - e^{\lambda t_d} z^{-1})^{2DN} = \sum_{k=0}^K h_k z^{-k}$$

will annihilate $Q(\alpha_m)s_m$. In matrix/vector form (1.97) can be written as

$$\begin{pmatrix} s_K & \alpha_K^J s_K & \cdots & s_0 & \cdots & \alpha_0^J s_0 \\ s_{K+1} & \alpha_{K+1}^J s_{K+1} & \cdots & s_1 & \cdots & \alpha_1^J s_1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ s_P & \alpha_P^J s_P & \cdots & s_0 & \cdots & \alpha_{P-K}^J s_{P-K} \end{pmatrix} \begin{pmatrix} h_0 r_0 \\ \vdots \\ h_0 r_J \\ \vdots \\ h_K r_0 \\ \vdots \\ h_K r_J \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Solving the system for $h_0 = 1$ enables finding the coefficients r_j , from which we can obtain the coefficients h_k . The roots of the filter $\hat{h}(z)$ and of the polynomial $Q(\alpha_m)$ give the locations of the switching points and the frequencies of the sine waves respectively. The number of values s_m required to build a system with enough equations to find the parameters of the signal is $P \geq 4D^3N^2 + 4D^2N^2 + 4D^2N + 6DN$.

An illustration of the sampling and reconstruction of a piecewise sinusoidal signal is shown in Fig. 1.19. For more details about the sampling of these signals we refer the reader to [16].

1.6.2 Signal compression

We have seen that specific classes of signals can be parsimoniously sampled using FRI sampling theory. Moreover, the sampling kernels involved include scaling

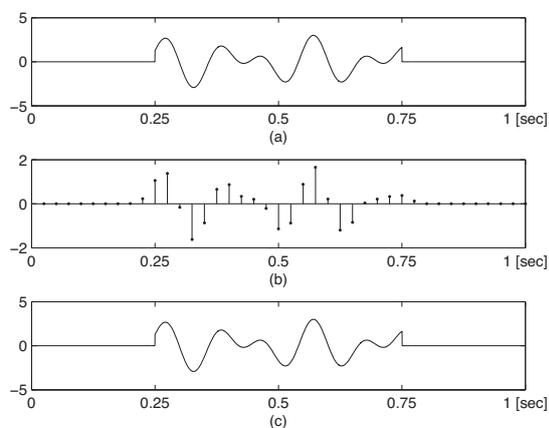


Figure 1.19 *Sampling a piecewise sinusoidal signal.* (a) The original continuous-time waveform, this is made of two truncated sinewaves. (b) The observed samples. (c) The reconstructed signal, where the retrieval of the two switch points and of the sine waves parameters is exact to machine precision.

functions used in the construction of wavelet bases such as, for example, B-splines or Daubechies scaling function.

We are now going to concentrate on this type of kernels and investigate the potential impact of such sampling schemes in compression where samples are also quantized and represented with a bit stream. In this context, the best way to analyze the compression algorithm is by using standard rate-distortion (R-D) theory since this gives the best achievable trade-off between the number of bits used and the reconstruction fidelity. It is often assumed that the error due to quantization can be modeled as additive noise. While this assumption is normally not accurate, it allows us to connect R-D theory with the CRB discussed in the previous section and therefore relate the theory of sampling FRI signals with compression.

The classes of signals we consider here are piecewise smooth functions, that is, functions which are made of regular pieces. The regularity of a function is normally measured using the Lipschitz coefficients [49]. We thus assume that the signals we consider are made of pieces with Lipschitz regularity α .

The FRI-based compression algorithm we propose is characterized by a simple linear encoding strategy and a more complex decoding. This is in contrast with standard wavelet based compression algorithms that involve a fairly sophisticated encoding strategy, but simple decoding. There might be situations, however, where it is important to have simple encoders. In our set-up, at the encoder the signal is decomposed using a standard wavelet transform and the resulting coefficients are quantized linearly. This means that the low-pass coefficients (equivalent to the samples in the FRI framework) are quantized first followed by the wavelet coefficients from the coarse to the finest scale.

At the decoder, the FRI reconstruction strategy is used to estimate the discontinuities in the signal using the scaling coefficients, while the other coefficients are used to reconstruct the smooth parts of the signals. By modeling the quantization error and any model mismatch as additive noise, one can use the CRB to estimate the performance of this compression strategy. The rate-distortion behavior of this FRI-based algorithm is [38, 39]:

$$D_{\text{FRI}}(R) \leq c_1 R^{-2\alpha} + c_2 \quad (1.98)$$

where c_2 is a systematic estimation error due to the model mismatch. Standard wavelet-based compression algorithms instead are characterized by a complex encoder and a simple decoder and can achieve the optimal rate distortion behavior [50]:

$$D_{\text{wave}}(R) \leq c_3 R^{-2\alpha}. \quad (1.99)$$

This indicates that if the systematic error in (1.98) is sufficiently small the FRI-based algorithm, which shifts the complexity from the encoder to the decoder, can achieve the same performance of the best wavelet-based compression algorithms for a wide range of bit rates.

1.6.3 Superresolution imaging

An image superresolution algorithm aims at creating a single detailed image, called a super-resolved image (SR) from a set of low-resolution input images of the same scene [51]. If different images from the same scene have been taken such that their relative shifts are not integer-multiple of the pixel size, then sub-pixel information exists among the set. This allows to obtain higher resolution accuracy of the scene once the images have been properly registered.

Image registration involves any group of transformations that removes the disparity between any two low resolution (LR) images. This is followed by image fusion, which blends the properly aligned LR images into a higher resolution output, possibly removing blur and noise introduced by the system [26].

The registration step is crucial in order to obtain a good quality SR image. The theory of FRI can be extended to provide superresolution imaging, combined with B-spline or E-spline processing. The key idea of this approach is that, using a proper model for the point-spread function of the scene acquisition system, it is possible to retrieve the underlying “continuous geometric moments” of the irradiance light-field. From this information, and assuming the disparity between any two images can be characterized by a global affine transformation, the set of images can be exactly registered.

Concretely, if the smoothing kernel that models the 2-D image acquisition is considered to be a B-spline or a more generic function such as a spline, then the continuous moments of the image can be found using a proper linear combination of the samples [25, 26]. From them, it is possible to find the central and complex moments of the signal, from which the disparities between any two LR images

can be estimated. Thus, this allows for proper registration of the set of input images, which can now be combined into a super-resolved output. Fig. 1.20 shows an example of the results obtained using the method presented in [26].

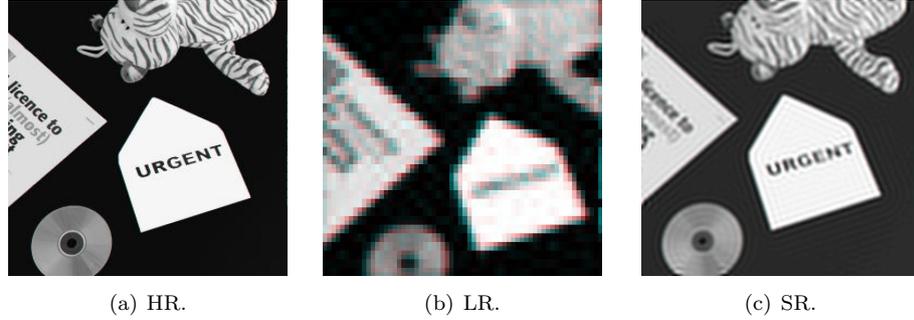


Figure 1.20 Image super-resolution from translated images with registration from the extracted edges and detected corners. (a) Original high resolution image (512x512pixels). (b) One of the 20 low-resolution images (64x64 pixels) used in the super-resolution simulation. (d) Super-resolved image with the proposed edge detector and Wiener Filter, 512x512 pixels, PSNR = 15.6 dB.

1.6.4 Ultrasound imaging

Another application of the stream of pulses FRI framework is ultrasound imaging [10]. In this application, an acoustic pulse is transmitted into the scanned tissue. Echoes of the pulse bounce off scatterers within the tissue, and create a signal consisting of a stream of pulses at the receiver. The time-delays and amplitudes of the echoes indicate the position and strength of the various scatterers, respectively. Therefore, determining these parameters from low rate samples of the received signal is an important problem. Reducing the rate allows more efficient processing which can translate to power and size reduction of the ultrasound imaging system.

The stream of pulses is finite since the pulse energy decays within the tissue. In order to demonstrate the viability of an FRI framework, we model the multiple echo signal recorded at the receiver as a finite stream of pulses, like (1.15). The unknown time-delays correspond to the locations of the various scatterers, whereas the amplitudes are the reflection coefficients. The pulse shape in this case is Gaussian, due the physical characteristics of the electro-acoustic transducer (mechanical damping).

As an example, we chose a phantom consisting of uniformly spaced pins, mimicking point scatterers, and scanned it by GE Healthcare's Vivid-i portable ultrasound imaging system, using a 3S-RS probe. We use the data recorded by a single element in the probe, which is modeled as a 1D stream of pulses. The center frequency of the probe is $f_c = 1.7021$ MHz, The width of the transmitted Gaussian pulse in this case is $\sigma = 3 \cdot 10^{-7}$ sec, and the depth of imaging is $R_{\max} = 0.16$ m

corresponding to a time window of² $\tau = 2.08 \cdot 10^{-4}$ sec. We carried out our sampling and reconstruction scheme on the data. We set $K = 4$, looking for the strongest 4 echoes. Since the data is corrupted by strong noise we oversampled the signal, obtaining twice the minimal number of samples. In addition, hard-thresholding of the samples was implemented, where we set the threshold to 10 percent of the maximal value. Fig. 1.21 depicts the reconstructed signal together with the full demodulated signal. Clearly, the time-delays were estimated with high precision. The amplitudes were estimated as well, however the amplitude of the second pulse has a large error. However, the exact locations of the scatterers is typically more important than the accurate reflection coefficients. This is because the time of arrival indicates the scatterer's location within the tissue. Accurate estimation of tissue boundaries and scatterer locations allows for reliable detection of certain illnesses, and is therefore of major clinical importance. The location of the boundaries is often more important than the power of the reflection which is incorporated in the received amplitudes.

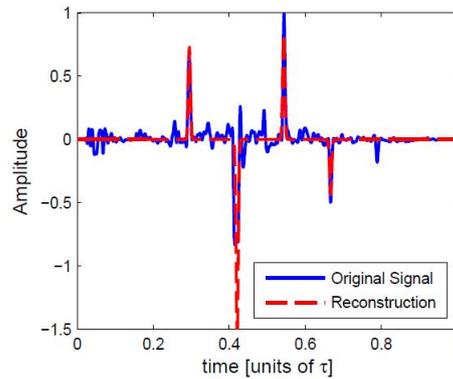


Figure 1.21 Example of sampling and reconstruction of real ultrasound imaging data. The input signal, in blue and continuous line, is sampled assuming there exist $K = 4$ pulses, and using an oversampling factor of 2. The output is a stream of Gaussian pulses, where the unknown locations and amplitudes have been estimated from the $N = 17$ samples obtained from the input, denoising with hard-thresholding.

Current ultrasound imaging technology operates at the high rate sampled data, e.g., $f_s = 20$ MHz in our setting. Since there are usually 100 different elements in a single ultrasonic probe each sampled at a very high rate, data throughput becomes very high, and imposes high computational complexity to the system, limiting its capabilities. Therefore, there is a demand for lowering the sampling rate, which in turn will reduce the complexity of reconstruction. Exploiting the parametric point of view, our sampling scheme reduces the sampling rate by over 2 orders of magnitude, while estimating the locations of the scatterers with high accuracy.

² The speed of sound within the tissue is 1550 m/sec.

1.6.5 Multipath medium identification

Another nice application of the FRI model is to the problem of time-varying channel estimation in wireless communication [17]. In such an application the aim of the receiver is to estimate the channel's parameters from the samples of the received signal [52].

We consider a baseband communication system operating in a multipath fading environment with pulse amplitude modulation (PAM). The data symbols are transmitted at a symbol rate of $\frac{1}{T}$, modulated by a known pulse $p(t)$. The transmitted signal $x_T(t)$ is given by

$$x_T(t) = \sum_{n=1}^{N_{sym}} d[n] p(t - nT) \quad (1.100)$$

where $d[n]$ are the data symbols taken from a finite alphabet, and N_{sym} is the total number of transmitted symbols.

The transmitted signal $x_T(t)$ passes through a baseband time-varying multipath channel whose impulse response is modeled as

$$h(\tau, t) = \sum_{k=1}^K \alpha_k(t) \delta(\tau - \tau_k) \quad (1.101)$$

where $\alpha_k(t)$ is the path time varying complex gain for the k th multipath propagation path and τ_k is the corresponding time delay. The total number of paths is denoted by K . We assume that the channel is slowly varying relative to the symbol rate, so that the path gains are considered to be constant over one symbol period:

$$\alpha_k(t) = \alpha_k[nT] \text{ for } t \in [nT, (n+1)T]. \quad (1.102)$$

In addition, we assume that the propagation delays are confined to one symbol, i.e. $\tau_k \in [0, T)$. Under these assumptions, the received signal at the receiver is given by

$$x_R(t) = \sum_{k=1}^K \sum_{n=1}^{N_{sym}} a_k[n] p(t - \tau_k - nT) + n(t) \quad (1.103)$$

where $a_k[n] = \alpha_k[nT] d[n]$ and $n(t)$ denotes the channel noise.

The received signal $x_R(t)$ fits the semi-periodic FRI signal model. Therefore, we can use the methods we described to recover the time delays of the propagation paths. In addition, if the transmitted symbols are known to the receiver, then the time varying path gains can be recovered from the sequences $a_k[n]$. As a result our sampling scheme can estimate the channel's parameters from samples of the output at a low rate, proportional to the number of paths.

As an example, we can look at the channel estimation problem in code division multiple access (CDMA) communication. This problem was handled using subspace techniques in [53, 54]. In these works the sampling is done at the chip

rate $\frac{1}{T_c}$ or above, where T_c is the chip duration given by $T_c = \frac{T}{N}$ and N is the spreading factor which is usually high (1023, for example, in GPS applications). In contrast, our sampling scheme can provide recovery of the channel's parameters at a sampling rate of $\frac{2K}{T}$. For a channel with a small number of paths, this sampling rate can be significantly lower than the chip rate.

1.6.6 Super-resolution radar

We end with an application of the semi-periodic model (1.20) to super-resolution radar [27].

In this context, we can translate the rate reduction to increased resolution, thus enabling super-resolution radar from low rate samples. Here the goal is to identify the range and velocity of a set of targets. The delay in this case captures the range while the time varying coefficients are a result of the Doppler delay related to the target velocity. More specifically, we assume that several targets can have the same delays but possibly different Doppler shifts so that $\{t_\ell\}_{\ell=1}^K$ denote the set of distinct delays. For each delay value t_ℓ there are K_ℓ values of associated Doppler shifts $\nu_{\ell k}$ and reflection coefficients $\alpha_{\ell k}$. It is further assumed that the system is highly underspread, namely $\nu_{\max}T \ll 1$, where ν_{\max} denotes the maximal Doppler shift, and T denotes the maximal delay. To identify the targets we transmit the signal

$$x_T = \sum_{n=0}^{N-1} x_n p(t - nT), \quad (1.104)$$

where x_n is a known N -length probing sequence, and $p(t)$ is a known pulse shape. The received signal can then be described in the form (1.20), where the sequences $a_\ell[n]$ satisfy

$$a_\ell[n] = x_n \sum_{k=1}^{K_\ell} \alpha_{\ell k} e^{j2\pi\nu_{\ell k}nT}. \quad (1.105)$$

The delays and the sequences $a_\ell[n]$ can be recovered using the general scheme for time delay recovery. The Doppler shifts and reflection coefficients are then determined from the sequences $a_\ell[n]$ using standard spectral estimation tools [18]. The targets can be exactly identified as long as the bandwidth of the transmitted pulse satisfies $\mathcal{W} \geq \frac{4\pi K}{T}$, and the length of the probing sequence satisfies $N \geq 2 \max K_\ell$ [27]. This leads to a minimal time-bandwidth product of the input signal of $\mathcal{WT} \geq 8\pi K \max K_\ell$, which is much lower than that obtained using standard radar processing techniques, such as matched-filtering (MF).

An example of the identification of nine close targets is illustrated in Fig. 1.22(a). The sampling filter used is a simple LPF. The original and recovered targets are shown on the Doppler-delay plane. Evidently all the targets were correctly identified using our FRI-based method. The result obtained by MF, with the same time-bandwidth product, is shown in Fig. 1.22(b). Clearly

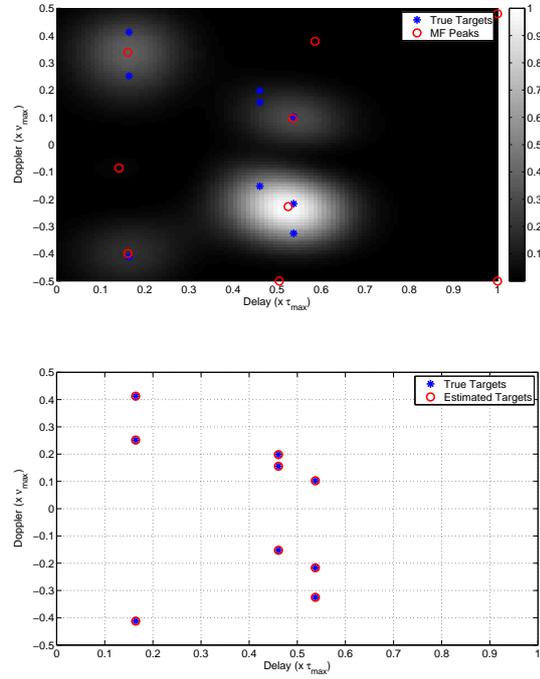


Figure 1.22 Comparison between the target-detection performance of matched-filtering and the procedure described in [27] for the case of nine targets (represented by $*$) in the delay–Doppler space with $\tau_{max} = 10\mu s$, $\nu_{max} = 10$ kHz, $\mathcal{W} = 1.2$ MHz, and $\mathcal{T} = 0.48$ ms. The probing sequence $\{x_n\}$ corresponds to a random binary (± 1) sequence with $N = 48$, the pulse $p(t)$ is designed to have a nearly-flat frequency response and the pulse repetition interval is $T = 10 \mu s$. (a) Target detection by matched-filtering. (b) Target detection using the proposed procedure with $P = 12$.

the FRI method has superior resolution than the standard MF. Thus, the FRI viewpoint not only offers a reduced-rate sampling method, but allows to increase the resolution in target identification.

1.7 Acknowledgement

Y. Eldar would like to thank Kfir Gedalyahu for many helpful comments.

Appendix to Chapter 1: Cramér–Rao bound derivations

Cramér–Rao bounds for the sinc kernel

Here we focus on the simplified case in which the input is a single τ -periodic Dirac, for which we can obtain a closed form expression for (1.85).

In the absence of noise, the samples taken at uniform intervals of time T with $K = 1$ Diracs, can be expressed as:

$$y_n = a_0 \psi(nT - t_0) = f(\boldsymbol{\theta}, n) \quad (.106)$$

where $\psi(t)$ is the Dirichlet kernel

$$\psi(t) = \sum_{m \in \mathbb{Z}} \phi_B(t - m\tau) = \frac{1}{B\tau} \frac{\sin\left(\frac{\pi(2M+1)t}{\tau}\right)}{\sin\left(\frac{\pi t}{\tau}\right)} = \frac{1}{B\tau} \frac{\sin(\pi Bt)}{\sin\left(\frac{\pi t}{\tau}\right)} \quad (.107)$$

and $\boldsymbol{\theta} = (t_0, a_0)^T$. The Fisher information matrix is the following square and size 2×2 matrix:

$$\mathbf{I}(\boldsymbol{\theta}) = \sigma^{-2} \begin{pmatrix} \sum_{n=0}^{N-1} (a_0 \psi'(nT - t_0))^2 & \sum_{n=0}^{N-1} a_0 \psi'(nT - t_0) \psi(nT - t_0) \\ \sum_{n=0}^{N-1} \psi(nT - t_0) a_0 \psi'(nT - t_0) & \sum_{n=0}^{N-1} (\psi(nT - t_0))^2 \end{pmatrix}. \quad (.108)$$

In order to evaluate the summations it is convenient to use the Fourier series representations of the signals $\psi(t)$ and $\psi'(t)$ because the following holds [55]:

$$\begin{aligned} \sum_{n=0}^{N-1} f(nT) g^*(nT) &\stackrel{(a)}{=} \sum_{n=0}^{N-1} \left(\sum_k \hat{f}_k e^{j2\pi k n \frac{T}{\tau}} \right) \left(\sum_{k'} \hat{g}_{k'}^* e^{-j2\pi k' n \frac{T}{\tau}} \right) \\ &= \sum_k \hat{f}_k \sum_{k'} \hat{g}_{k'}^* \frac{1 - e^{j2\pi(k-k')N\frac{T}{\tau}}}{e^{-j2\pi(k-k')\frac{T}{\tau}}} \\ &\stackrel{(b)}{=} \sum_k \hat{f}_k \sum_{k'} \hat{g}_{k'}^* N \delta_{k,k'} = N \sum_k \hat{f}_k \hat{g}_k^*, \end{aligned} \quad (.109)$$

where in (a) we have used the fact that $f(t)$ and $g(t)$ are assumed to be periodic, and in (b) the fact that for $\tau = NT$ the sum is only non-zero when $k = k'$.

Furthermore, if we call $\hat{\psi}_k$ the coefficients for $\psi(t)$, then $\hat{\psi}'_k = j2\pi \frac{k}{\tau} \hat{\psi}_k$ would be the coefficients for its derivative and $\hat{\psi}_k^{(t_0)} = e^{-j2\pi k \frac{t_0}{\tau}} \hat{\psi}_k$ the coefficients for its shifted version by t_0 .

These last equivalences and (.109) simplify the calculations of the sums in (.108), because the function $\psi(t)$ is characterized by the Fourier series coefficients $\hat{\psi}_k = \frac{1}{B\tau}|k| \leq M$ and $\hat{\psi}_k = 0$ otherwise. Element (1,1) in (.108) can therefore be calculated as

$$\sigma^{-2} \sum_{n=0}^{N-1} (a_0 \psi'(nT - t_0))^2 = \sigma^{-2} a_0^2 N \sum_{|k| \leq M} \psi'_k e^{-j2\pi k \frac{t_0}{\tau}} \psi_k^* e^{j2\pi k \frac{t_0}{\tau}} \quad (.110)$$

$$[\mathbf{I}(\boldsymbol{\theta})]_{11} = N \left(\frac{a_0}{\sigma} \frac{2\pi}{B\tau^2} \right)^2 \frac{M(M+1)(2M+1)}{3}.$$

The other elements in (.108) can be calculated likewise. Due to the fact that the elements in the anti-diagonal are zero, the inverse of the Fisher information matrix can be computed by just inverting the diagonal elements, yielding

$$\text{CRB}(\boldsymbol{\theta}) = \begin{pmatrix} \frac{1}{N} \left(\frac{\sigma}{a_0} \frac{B\tau^2}{2\pi} \right)^2 \frac{3}{M(M+1)(2M+1)} & 0 \\ 0 & \frac{1}{N} \sigma^2 B^2 \tau^2 \frac{1}{2M+1} \end{pmatrix}, \quad (.111)$$

where $M = \lfloor \frac{B\tau}{2} \rfloor$. Note that, for this M , it always holds that $2M+1 = B\tau$.

To end, we can determine the uncertainties in the location and the amplitude from the values derived in the CRB. We know that the diagonal values in (.111) are lower bounds for $\text{var}\{t_0\}$ and $\text{var}\{a_0\}$ respectively. And since we are interested in unbiased estimators, the variances equal the MSE for each unknown. Therefore, we can write the uncertainty in the location as follows:

$$\begin{aligned} \frac{\Delta t_0}{\tau} &\geq \sqrt{\frac{1}{N} \left(\frac{\sigma}{a_0} \frac{B\tau}{2\pi} \right)^2 \frac{3}{M(M+1)(2M+1)}} \quad (.112) \\ &\stackrel{(a)}{=} \frac{1}{\pi} \sqrt{\frac{3B\tau}{N(B^2\tau^2 - 1)}} \text{PSNR}^{-\frac{1}{2}}, \end{aligned}$$

where we have defined the peak signal-to-noise ratio as $\text{PSRN} = \left(\frac{a_0}{\sigma}\right)^2$, and in (a) we have used the fact that $4M(M+1) = (2M+1)^2 - 1 = B^2\tau^2 - 1$. We can also write the uncertainty in the amplitude as:

$$\frac{\Delta a_0}{|a_0|} \geq \sqrt{\frac{1}{N} \frac{\sigma^2}{a_0^2} B^2 \tau^2 \frac{1}{2M+1}} = \sqrt{\frac{B\tau}{N}} \text{PSNR}^{-\frac{1}{2}}. \quad (.113)$$

Cramér–Rao bounds for the SoS kernel

The derivation of the CRB for the SoS kernel follows exactly the same steps as in the previous section for the sinc. First, we express the samples as $y_n = a_0 \eta(t_0 - nT) = f(\boldsymbol{\theta}, n)$, where $\eta(t) = \sum_{m \in \mathbf{Z}} g(t - m\tau)$, and $g(t)$ is the filter defined in (1.34). Now, we can once more rely on the Fourier series representation of the

signals to calculate the summations. From (1.33), the Fourier series coefficients of the periodic expansion of the kernel $\eta(t)$ are $\eta_k = \frac{1}{\tau} \hat{g}\left(\frac{2\pi k}{\tau}\right) = b_k$, for $k \in \mathcal{K}$.

The elements of the Fisher information matrix are found as in the sinc case, using the equivalence shown in (.109) and the properties for the coefficients of the derivative of $\eta(t)$ and of its shifted version by t_0 . The only additional consideration is that, when computing the elements of the anti-diagonal, we encounter a term of the form $\sum_{k \in \mathcal{K}} k |b_k|^2$. This is always equal to zero as long as $|b_k| = |b_{-k}|$ which is true, for instance, if we want to design real filters, for which $b_k = b_{-k}^*$. Thus

$$\text{CRB}(\boldsymbol{\theta}) = \begin{pmatrix} \frac{1}{N} \left(\frac{\sigma}{a_0}\right)^2 \left(\frac{\tau}{2\pi}\right)^2 \frac{1}{\sum_{k \in \mathcal{K}} k^2 |b_k|^2} & 0 \\ 0 & \sigma^2 \frac{1}{N} \frac{1}{\sum_{k \in \mathcal{K}} |b_k|^2} \end{pmatrix}. \quad (.114)$$

The uncertainty in the location is

$$\frac{\Delta t_0}{\tau} \geq \frac{1}{2\pi} \sqrt{\frac{1}{N} \frac{1}{\sum_{k \in \mathcal{K}} k^2 |b_k|^2}} \text{PSNR}^{-\frac{1}{2}}, \quad (.115)$$

and the uncertainty in the amplitude

$$\frac{\Delta a_0}{|a_0|} \geq \sqrt{\frac{1}{N} \frac{1}{\sum_{k \in \mathcal{K}} |b_k|^2}} \text{PSNR}^{-\frac{1}{2}}. \quad (.116)$$

Cramér–Rao bounds for B-splines

We now derive the lower bounds on the variances when estimating the location t_0 and amplitude a_0 of a single τ -periodic Dirac when the sampling kernel is a B-spline. We restrict the analysis to the shortest possible B-spline capable of sampling one Dirac, i.e. the first order B-spline ($P = 1$) obtained as the convolution of two box functions. It has the following form:

$$\beta_1(t) = \begin{cases} t, & 0 \leq t < 1, \\ 2 - t, & 1 \leq t < 2. \end{cases} \quad (.117)$$

In the absence of noise, the samples taken at uniform intervals of time T can be expressed as $y_n = a_0 \sum_{m \in \mathbb{Z}} \beta_1\left(\frac{t_0}{T} - n - mN\right) = f(\boldsymbol{\theta}, n)$ where we have used $\tau = NT$.

If we want to sample an infinitely long signal with a finite support kernel, we need to have zero samples in between blocks of non-zero samples. For a kernel of size $L = P + 1$ we need at least 1 zero per period $\tau \geq (L + 1)T \leftrightarrow N \geq L + 1$. We assume the only non-zero samples are located in the positions $n = 0, \dots, L - 1$ (or we would do a circular shift otherwise).

We are working with a finite support kernel of length $L = P + 1 = 2$. This allows to remove the dependence of the Fisher information matrix on m , since fix-

ing $t_0 \in [PT, (P+1)T) = [T, 2T)$, which is equivalent to $n = 0, \dots, L-1$, makes the only possible m value to be equal to zero.

We can now evaluate the terms of the Fisher information matrix, which has a form identical to (.108). Contrary to the previous sections, now we have to work in the time domain, using the definition of the B-spline (.117) and of its derivative. We have finite length sums over n , so it is possible to derive closed form results. For example, the first element of the diagonal can be calculated as

$$\sigma^{-2} \sum_{n=0}^1 \left(\frac{a_0}{T} \beta'_1 \left(\frac{t_0}{T} - n \right) \right)^2 = \sigma^{-2} \left(\frac{a_0}{T} \right)^2 [1^2 + 1^2] \quad (.118)$$

$$[\mathbf{I}(\boldsymbol{\theta})]_{11} = 2\sigma^{-2} \left(\frac{a_0}{T} \right)^2.$$

Once we obtain all the terms, the CRB can be found by inverting the Fisher information matrix. In this scenario, the bounds depend on t_0 :

$$\text{CRB}(\boldsymbol{\theta}) = \begin{pmatrix} (2t_0^2 - 6Tt_0 + 5T^2) \left(\frac{\sigma}{a_0} \right)^2 & (3T - 2t_0) \frac{\sigma^2}{a_0} \\ (3T - 2t_0) \frac{\sigma^2}{a_0} & 2\sigma^2 \end{pmatrix}. \quad (.119)$$

In order to remove the dependence on t_0 , we can consider various options. For instance, we may calculate the expected value of $\text{CRB}(\boldsymbol{\theta})$ assuming that t_0 is uniformly distributed over τ . This leads to:

$$\frac{\Delta t_0}{\tau} \geq \frac{1}{N} \sqrt{\frac{2}{3}} \text{PSNR}^{-\frac{1}{2}}. \quad (.120)$$

$$\frac{\Delta a_0}{|a_0|} \geq \sqrt{2} \text{PSNR}^{-\frac{1}{2}}. \quad (.121)$$

Cramér–Rao bounds for E-splines

To conclude, we derive lower bounds on the variances of the estimated location and amplitude for a τ -periodic single Dirac when the sampling kernel is an E-spline. The method is the same as that explained for B-splines, but requires further assumptions.

We restrict the analysis to cardinal exponential splines, which we also assume to be trigonometric [35]. The first property means that the exponential splines are defined on a uniform grid, and the second property that the complex parameters are purely imaginary and equally spaced around the origin, which yields a real valued function.

We need to be careful when working with E-splines for two main reasons. The first is that the periodicity of the complex exponentials causes the moments to be periodic too. This imposes a limit in the locations that can be retrieved. The second is that E-splines are no longer a basis for certain combinations of the complex parameters [35]. These conditions, plus the fact that we want that

the exponential reproduction formula coefficients form an orthogonal basis with $\omega_0 = \frac{2\pi}{N}$, translate into a bound for t_0 which has to satisfy $t_0 < \frac{NT}{2} = \frac{\tau}{2}$, and for the number of samples, which requires $N > \max(P + 1, 2P)$. For a more detailed explanation of these conditions, we refer to [56].

If we focus on the first order real E-spline, then it is possible to derive a closed form expression for the CRB. Note that this function is obtained through convolution of the zero order components having complex parameters $\pm j\omega_0$. The obtained kernel has the following form:

$$e_1(t) = \begin{cases} \frac{\sin(\omega_0 t)}{\omega_0}, & 0 \leq t < 1, \\ -\frac{\sin(\omega_0(t-2))}{\omega_0}, & 1 \leq t < 2. \end{cases} \quad (.122)$$

The CRB can be obtained by inverting the Fisher information matrix, derived similarly to the B-spline case. In this situation, again the bounds depend on t_0 . Calculating the average values leads to:

$$\frac{\Delta t_0}{\tau} \geq \frac{1}{N} \sqrt{\frac{\omega_0 - \cos \omega_0 \sin \omega_0}{\omega_0 \sin^2 \omega_0}} \text{PSNR}^{-\frac{1}{2}}. \quad (.123)$$

$$\frac{\Delta a_0}{|a_0|} \geq \sqrt{\frac{\omega_0 + \cos \omega_0 \sin \omega_0}{\sin^2 \omega_0}} \text{PSNR}^{-\frac{1}{2}}. \quad (.124)$$

References

- [1] E. T. Whittaker, “On the Functions Which are Represented by the Expansions of the Interpolation Theory,” *Proc. Roy. Soc. Edinburgh A*, vol. 35, pp. 181–194, 1915.
- [2] H. Nyquist, “Certain topics in telegraph transmission theory,” *Trans. AIEE*, vol. 47, pp. 617–644, April 1928.
- [3] C. Shannon, “A mathematical theory of communication,” *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948.
- [4] M. Vetterli, P. Marziliano, and T. Blu, “Sampling signals with finite rate of innovation,” *IEEE Transactions on Signal Processing*, vol. 50, pp. 1417–1428, 2002.
- [5] Y. C. Eldar and T. Michaeli, “Beyond bandlimited sampling,” *IEEE Signal Processing Magazine*, vol. 26, pp. 48–68, May 2009.
- [6] M. Unser, “Sampling—50 years after Shannon,” in *Proceedings of the IEEE*, pp. 569–587, 2000.
- [7] Y. M. Lu and M. N. Do, “A Theory for Sampling Signals From a Union of Subspaces,” *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2334–2345, 2008.
- [8] M. Mishali and Y. C. Eldar, “Robust recovery of signals from a structured union of subspaces,” *IEEE Transactions on Information Theory*, vol. 55, pp. 5302–5316, November 2009.
- [9] 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, pp. 1741–1757, 2007.
- [10] R. Tur, Y. C. Eldar, and Z. Friedman, “Innovation Rate Sampling of Pulse Streams with Application to Ultrasound Imaging,” *To appear in IEEE Transactions on Signal Processing*.
- [11] K. Gedalyahu, R. Tur, and Y. C. Eldar, “Multichannel sampling of pulse streams at the rate of innovation,” *To appear in IEEE Transactions on Signal Processing*, 2010.
- [12] Z. Ben-Haim, T. Michaeli, and Y. C. Eldar, “Performance Bounds and Design Criteria for Estimating Finite Rate of Innovation Signals,” *To appear in IEEE Transactions on Information Theory*, 2010.
- [13] I. Maravic and M. Vetterli, “Sampling and reconstruction of signals with finite rate of innovation in the presence of noise,” *IEEE Transactions on Signal Processing*, vol. 53, pp. 2788–2805, 2005.
- [14] V. Y. F. Tan and V. K. Goyal, “Estimating signals with finite rate of innovation from noisy samples: A stochastic algorithm,” *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 5135–5146, 2008.
- [15] A. Erdozain and P. M. Crespo, “A new stochastic algorithm inspired on genetic algorithms to estimate signals with finite rate of innovation from noisy samples,” *Signal*

- Processing*, vol. 90, pp. 134–144, January 2010.
- [16] J. Berent, P. L. Dragotti, and T. Blu, “Sampling Piecewise Sinusoidal Signals With Finite Rate of Innovation Methods,” *IEEE Transactions on Signal Processing*, vol. 58, no. 2, pp. 613–625, 2010.
 - [17] K. Gedalyahu and Y. C. Eldar, “Time Delay Estimation from Low Rate Samples: A Union of Subspaces Approach,” *IEEE Transactions on Signal Processing*, vol. 58, pp. 3017–3031, June 2010.
 - [18] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 2000.
 - [19] 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.
 - [20] H. Akhondi Asl, P. L. Dragotti, and L. Baboulaz, “Multichannel Sampling of Signals with Finite Rate of Innovation,” *IEEE Signal Processing Letter*, vol. 17, pp. 762–765, August 2010.
 - [21] E. Matusiak and Y. C. Eldar, “Sub-Nyquist sampling of short pulses: Part I,” *To appear in IEEE J SP*, 2010. arXiv:1010.3132v1.
 - [22] A. Hormati, O. Roy, Y. M. Lu, and M. Vetterli, “Distributed sampling of signals linked by sparse filtering: Theory and applications,” *Trans. Signal Processing*, vol. 58, no. 3, pp. 1095–1109, 2010.
 - [23] I. Maravic and M. Vetterli, “Exact Sampling Results for Some Classes of Parametric Non-Bandlimited 2-D Signals,” *IEEE Transactions on Signal Processing*, vol. 52, no. 1, pp. 175–189, 2004.
 - [24] P. Shukla and P. L. Dragotti, “Sampling Schemes for Multidimensional Signals with Finite Rate of Innovation,” *IEEE trans. on Signal Processing*, 2006.
 - [25] L. Baboulaz and P. L. Dragotti, “Distributed acquisition and image super-resolution based on continuous moments from samples,” in *in Proceedings of IEEE International Conference on Image Processing (ICIP)*, pp. 3309–3312, 2006.
 - [26] L. Baboulaz and P. L. Dragotti, “Exact feature extraction using finite rate of innovation principles with an application to image super-resolution,” *IEEE Transaction on Image Processing*, vol. 18, no. 2, pp. 281–298, 2009.
 - [27] W. U. Bajwa, K. Gedalyahu, and Y. C. Eldar, “Identification of Parametric Under-spread Linear Systems and Super-Resolution Radar,” *To appear in IEEE Trans. Signal Processing*. arXiv:1008.0851v2.
 - [28] I. Maravic, J. Kusuma, and M. Vetterli, “Low-Sampling Rate UWB Channel Characterization and Synchronization,” *Journal of Communications and Networks KOR, special issue on ultra-wideband systems*, vol. 5, no. 4, pp. 319–327, 2003.
 - [29] I. Maravic, M. Vetterli, and K. Ramchandran, “Channel Estimation and Synchronization with Sub-Nyquist Sampling and Application to Ultra-Wideband Systems,” in *Proceedings IEEE International Symposium on Circuits and Systems*, vol. 5, pp. 381–384, 2004.
 - [30] M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*. Englewood Cliffs, NJ, USA: Prentice Hall, 1995.
 - [31] I. Gohberg and S. Goldberg, *Basic Operator Theory*. Boston, MA: Birkhäuser, 1981.
 - [32] Y. Hua and T. K. Sarkar, “Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 70, pp. 1272–1281, May 1990.

-
- [33] S. Y. Kung, K. S. Arun, and D. V. B. Rao, "State-space and singular-value decomposition-based approximation methods for the harmonic retrieval problem," *J. Opt. Soc. Am.*, vol. 73, pp. 1799–1811, December 1983.
- [34] B. D. Rao and K. S. Arun, "Model based processing of signals: a state space approach," *Proceedings of the IEEE*, vol. 80, no. 2, pp. 283–309, 1992.
- [35] M. Unser and T. Blu, "Cardinal Exponential Splines: Part I—Theory and Filtering Algorithms," *IEEE Transactions on Signal Processing*, vol. 53, pp. 1425–1438, April 2005.
- [36] J. A. Urigüen, P. L. Dragotti, and T. Blu, "On the Exponential Reproducing Kernels for Sampling Signals with Finite Rate of Innovation," *To appear in the Proceedings of Sampling Theory and Applications (SampTA)*, 2011.
- [37] G. Strang and G. Fix, "Fourier analysis of the finite element variational method," *Constructive Aspect of Functional Analysis*, pp. 796–830, 1971.
- [38] V. Chaisinthop and P. L. Dragotti, "Semi-parametric compression of piecewise smooth functions," *Proc. of European Signal Processing Conference (EUSIPCO)*, 2009.
- [39] V. Chaisinthop and P. L. Dragotti, "Centralized and Distributed Semi-Parametric Compression of Piecewise Smooth Functions," *To appear in IEEE Transaction on Signal Processing*, 2011.
- [40] J. Kusuma and V. Goyal, "Multichannel sampling of parametric signals with a successive approximation property," *IEEE Int. Conf. Image Process. (ICIP2006)*, pp. 1265–1268, Oct. 2006.
- [41] H. Olkkonen and J. Olkkonen, "Measurement and reconstruction of impulse train by parallel exponential filters," vol. 15, pp. 241–244, 2008.
- [42] M. Mishali, Y. C. Eldar, O. Dounaevsky, and E. Shoshan, "Xampling: Analog to digital at sub-nyquist rates," *IET Circuits, Devices and Systems*, vol. 5, pp. 8–20, Jan. 2011.
- [43] M. Mishali and Y. C. Eldar, "From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals," *IEEE J. Sel. Top. Sig. Proc.*, vol. 4, pp. 375–391, Apr. 2010.
- [44] Y. C. Eldar and V. Pohl, "Recovering signals from lowpass data," vol. 58, no. 5, pp. 2636–2646, 2010.
- [45] M. Mishali and Y. C. Eldar, "Blind Multiband Signal Reconstruction: Compressed Sensing for Analog Signals," *Signal Processing, IEEE Transactions on*, vol. 57, no. 3, pp. 993–1009, 2009.
- [46] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," vol. 37, pp. 984–995, Jul. 1989.
- [47] R. Schmidt, "Multiple emitter location and signal parameter estimation," vol. 34, pp. 276–280, Mar. 1986.
- [48] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice Hall, 1993.
- [49] S. Mallat, *A wavelet tour of signal processing*. Academic Press, 1998.
- [50] A. Cohen, I. Daubechies, O. G. Guleryuzz, and M. T. Orchard, "On the importance of combining wavelet-based nonlinear approximation with coding strategies," *IEEE Transactions on Information Theory*, vol. 48, pp. 1895–1921, 2002.
- [51] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Advances and Challenges in Super-Resolution," *International Journal of Imaging Systems and Technology*, vol. 13, no. 10, pp. 1327–1344, 2004.
- [52] H. Meyr, M. Moeneclaey, and S. A. Fechtel, *Digital communication receivers: synchronization, channel estimation, and signal processing*. New York, NY, USA: Wiley-

- Interscience, 1997.
- [53] S. E. Bensley and B. Aazhang, "Subspace-based channel estimation for code division multiple access communication systems," *IEEE Transactions on Communications*, vol. 44, no. 8, pp. 1009–1020, 1996.
 - [54] E. G. Strom, S. Parkvall, S. L. Miller, and B. E. Ottersten, "Dscdma synchronization in time-varying fading channels," *IEEE Journal on Selected Areas in Communications*, vol. 14, no. 8, pp. 1636–1642, 1996.
 - [55] L. Coulot, M. Vetterli, T. Blu, and P. L. Dragotti, "Sampling Signals with Finite Rate of Innovation in the Presence of Noise," tech. rep., Ecole Polytechnique Federale de Lausanne, Switzerland, 2007.
 - [56] F. J. Homann and P. L. Dragotti, "Robust Sampling of "Almost" Sparse Signals," tech. rep., Imperial College London, 2008.

Index

- Annihilating filter, 13
 - equation, 13
 - extended annihilating equation, 36
- Cadzow denoising, 37
- Cramér–Rao bound, *see* CRB
- CRB, 30
 - derivations, 56
 - kernel comparison, 35
 - periodic stream, 34
 - simulations, 41
- Exponential reproducing kernels, 19
 - CRB, 35
 - CRB derivation, 59
 - piecewise sinusoidal, 47
 - simulations, 42
- Finite rate of innovation, *see* FRI
- FRI, 2
 - applications, 47
 - definition, 8
 - examples, 8
 - extensions, 47
 - history, 4
 - sampling scheme, 2
 - simulations, 38
- Modulated wideband converter, 24
- Multichannel, 22
 - history, 5
 - infinite FRI, 25
 - periodic FRI, 23
 - semi-periodic FRI, 25
 - simulations, 44
- Noise, 30
 - annihilating filter, *see* Annihilating filter
 - development, 4
 - history, 5
 - matrix pencil, *see* matrix pencil
 - Oversampling
 - Cadzow, 37
 - TLS, 37
 - performance bounds
 - continuous-time noise, 30
 - sampling noise, 34
 - sampling scheme, 30
- Oversampling, *see* Noise
- Polynomial reproducing kernels, 21
 - CRB, 35
 - CRB derivation, 58
- Sampling, 2
 - equation, 3
 - noise, 30
 - noise-free, 12
- Sinc kernel, 12
 - CRB, 35
 - CRB derivation, 56
 - simulations, 39
- Singular value decomposition, *see* SVD
- SoS kernel, 15
 - CRB, 35
 - CRB derivation, 57
 - simulations, 39
- Sum of sines kernel, *see* SoS kernel
- SVD, 37
 - Cadzow, *see* Cadzow denoising
 - TLS, *see* TLS
- TLS, 37
- Total least squares denoising, *see* TLS