

# Sparse Sampling of Structured Information and its Application to Compression

Pier Luigi Dragotti

*Communications and Signal Processing Group  
Electrical and Electronic Engineering Department  
Imperial College London, United Kingdom  
p.dragotti@imperial.ac.uk*

**Abstract**—It has been shown recently that it is possible to sample classes of non-bandlimited signals which we call signals with Finite Rate of Innovation (FRI). Perfect reconstruction is possible based on a set of suitable measurements and this provides a sharp result on the sampling and reconstruction of sparse continuous-time signals.

In this paper, we first review the basic theory and results on sampling signals with finite rate of innovation. We then discuss variations of the above framework to handle noise and model mismatch. Finally, we present some results on compression of piecewise smooth signals based on the FRI framework.

## I. INTRODUCTION

The problem of reconstructing or estimating partially observed or sampled signals is an old and important one, and finds application in many areas of signal processing and communications. Traditional acquisition and reconstruction approaches are heavily influenced by the classical Shannon sampling theory which gives an exact sampling and interpolation formula for bandlimited signals. Recently, the classical Shannon sampling framework has been extended to classes of non-bandlimited structured signals. In these new sampling schemes, the prior that the signal is sparse in a basis or in a parametric space is put to contribution and perfect reconstruction is possible based on a set of suitable measurements.

Depending on the set-up and reconstruction method involved, the above sampling problem goes under different names like compressed sensing, compressive sampling [1], [2] or sampling signals with finite rate of innovation (FRI) [3], [4].

The set-up considered here is the one in [3], [4], where the acquisition process is modeled as in Fig. 1. Here the smoothing function  $\varphi(t)$  is called the sampling kernel and normally models the distortion due to the acquisition device. The sampling kernel used in [3] is the sinc function, while the work in [4] uses compactly supported functions like for example polynomial splines (B-splines) [5] or exponential splines (E-splines) [6]. In both works it is shown that perfect reconstruction of classes of FRI signals from the measurements  $y_n$  is achievable by using a variation of Prony's method also known as annihilating filter method [7]. Signals that

can be sampled with this method include streams of Diracs, piecewise polynomial and piecewise sinusoidal signals.

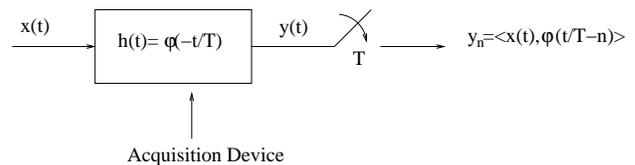


Fig. 1. Sampling setup. Here,  $x(t)$  is the continuous-time signal,  $h(t)$  the impulse response of the acquisition device and  $T$  the sampling period. The measured samples are  $y_n = \langle x(t), \varphi(t/T - n) \rangle$ .

In this paper, we first review the basic set-up and the fundamental results presented in [3], [4], we then discuss the problem of reconstructing signals when the samples have been corrupted by noise. In this context we present a variation of the robust reconstruction algorithm presented in [8] and we take full advantage of the fact that the kernels considered here have compact support.

We finally discuss the application of this new sampling framework to compression and present a new compression algorithm where complexity is shifted from the encoder to the decoder. We show that the algorithm performs as well as state-of-the-art compression algorithms based on the wavelet transform.

## II. SAMPLING SIGNALS WITH FINITE RATE OF INNOVATION

For the sake of clarity we restrict our analysis to the case where the observed signal  $x(t)$  is a stream of  $K$  Diracs with amplitudes  $a_k$  located at distinct instants  $t_k \in [0, \tau]$ :

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

Furthermore, we assume the sampling period is  $T = \tau/N$ . Consequently, the measurements are

$$\begin{aligned} y_n &= \langle x(t), \varphi(t/T - n) \rangle \\ &= \sum_{k=0}^{K-1} a_k \varphi(t_k/T - n), \quad n, = 0, 1, \dots, N-1. \end{aligned}$$

In [3], [4], [8], it was shown that with a proper choice of the acquisition kernel, it is possible to reconstruct  $x(t)$  from the samples  $y_n$  exactly. The kernels used in [3] are the sinc and the Gaussian functions. In this paper, we concentrate on the compact support kernels used in [4]. This includes:

- *Polynomial reproducing kernels:* Any kernel that satisfies

$$\sum_{n \in \mathbb{Z}} c_{m,n} \varphi(t-n) = t^m \quad m = 0, 1, \dots, P \quad (2)$$

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

- *Exponential reproducing kernels:* Any kernel that satisfies

$$\sum_{n \in \mathbb{Z}} c_{m,n} \varphi(t-n) = e^{\alpha_m t} \quad \text{with } \alpha_m = \alpha_0 + m\lambda \quad \text{and } m = 0, 1, \dots, P \quad (3)$$

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

The coefficients  $c_{m,n}$  in (2) are given by

$$c_{m,n} = \frac{1}{T} \int_{-\infty}^{\infty} t^m \tilde{\varphi}(t/T - n) dt,$$

where  $\tilde{\varphi}(t)$  is chosen to form with  $\varphi(t)$  a quasi-biorthonormal set [9]. This includes the particular case where  $\tilde{\varphi}(t)$  is the dual of  $\varphi(t)$ , that is,  $\langle \tilde{\varphi}(t-n), \varphi(t-k) \rangle = \delta_{n,k}$ . A similar expression applies to the coefficients  $c_{m,n}$  in (3).

The first family of kernels includes any function satisfying the so-called Strang-Fix conditions [10]. Namely,  $\varphi(t)$  satisfies Eq. (2) if and only if

$$\hat{\varphi}(0) \neq 0 \text{ and } \hat{\varphi}^{(m)}(2n\pi) = 0 \text{ for } n \neq 0 \text{ and } m = 0, 1, \dots, P,$$

where  $\hat{\varphi}(\omega)$  is the Fourier transform of  $\varphi(t)$  and the superscript  $(m)$  stands for the  $m$ -th derivative of  $\varphi(t)$ .

One important example of functions satisfying Strang-Fix conditions is given by the family of B-splines [5]. A B-spline of order  $P$  is a function of compact support  $L = P + 1$  and can reproduce polynomials up to degree  $P$ . It is obtained by the  $(P + 1)$ -fold convolution of the zero order B-spline and has the following Fourier transform

$$\hat{\beta}_P(\omega) = \left( \frac{1 - e^{j\omega}}{j\omega} \right)^{P+1}.$$

Strang-Fix conditions are used extensively in wavelet theory as well. In that context, the focus is on the design of wavelets with a certain number of vanishing moments [11]. The interesting point, here, is that a wavelet with  $P + 1$  vanishing moments is generated by a scaling function that can reproduce polynomials of degree  $P$ . This means that such a scaling function can be included in our family of sampling kernels. Given the central role played by wavelets in compression, we immediately realized the potential impact that these kernels may have not only for sampling but also for compression. This will become more evident in the following sections.

The family of E-splines represents an extension of the polynomial splines and the Fourier transform of the  $P$ -th order E-spline is:

$$\hat{\beta}_{\alpha}(\omega) = \prod_{m=0}^P \left( \frac{1 - e^{\alpha_m - j\omega}}{j\omega - \alpha_m} \right). \quad (4)$$

The above E-spline is able to reproduce the exponentials  $e^{\alpha_m t}$ ,  $m = 0, 1, \dots, P$ . Notice that the exponent  $\alpha_m$  in Eq. (4) can be complex which indicates that E-splines are usually complex functions. However, this can be avoided by choosing complex conjugate exponents.

The reconstruction scheme of [4] operates as follows: First the samples are linearly combined with the coefficients  $c_{m,n}$  of (2),(3) to obtain the new measurements

$$s_m = \sum_{n=0}^N c_{m,n} y_n \quad m = 0, 1, \dots, P. \quad (5)$$

Then, if the original signal is a stream of Diracs as the one in (1), one can show that

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

where  $u_k = t_k/T$  when polynomial splines are used and  $u_k = e^{\lambda t_k/T}$  when exponential splines are involved. In either cases, the pairs of unknowns  $\{a_k, u_k\}$  can be retrieved from the power series  $s_m = \sum_{k=0}^{K-1} a_k u_k^m$  using the classical Prony's method. The key ingredient of this method is the annihilating filter. Call  $h_m$ ,  $m = 0, 1, \dots, K$  the filter with  $z$ -transform

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

That is, the roots of  $H(z)$  correspond to the locations  $u_k$ . It clearly follows that

$$h_m * s_m = \sum_{i=0}^K h_i s_{m-i} = \sum_{k=0}^{K-1} a_k u_k^m \underbrace{\sum_{i=0}^K h_i u_k^{-i}}_{H(u_k)} = 0. \quad (6)$$

The filter  $h_m$  is thus called annihilating filter since it annihilates the observed series  $s_m$ . Moreover, the zeros of this filter uniquely define the set of locations  $u_k$  since the locations are distinct. The identity in (6) can be written in matrix/vector form as follows:

$$SH = 0 \quad (7)$$

which reveals that the Toeplitz matrix  $S$  is rank deficient. By solving the above system, we retrieve the  $u_k$ 's and therefore the locations  $t_k$ . Given the locations, the weights  $a_k$  are then obtained by solving a system of linear equations. Notice that the problem can be solved only when  $P \geq 2K - 1$ .

We thus conclude that perfect reconstruction of a stream of Diracs is possible with any kernel able to reproduce exponentials or polynomials. The reconstruction procedure is the same, the only difference is in the choice of the coefficients  $c_{n,m}$ , which depends on the properties of the chosen kernel.

Other FRI signals can be sampled and reconstructed using the same procedures. FRI signals that can be sampled and perfectly reconstructed include piecewise polynomial and piecewise sinusoidal signals [3], [4], [12], multidimensional signals [13], [14] and signals that have a sparse representation in a basis [15].

### III. THE NOISY SCENARIO

The signal and acquisition models discussed before are ideal and perturbations to this model need to be considered. For simplicity we assume the perturbation is introduced after sampling and is modeled as additive noise. Consequently, the new measurements are

$$\hat{y}_n = \langle x(t), \varphi(t/T - n) \rangle + \epsilon_n, \quad n = 0, 1, \dots, N - 1,$$

where  $\epsilon_n$  is assumed to be i.i.d. additive Gaussian noise with zero mean and variance  $\sigma^2$ .

In order to reduce the effect of noise, the reconstruction procedure discussed in the previous section need to be modified. The retrieval of the signal parameters in the FRI sampling framework is similar to a classical harmonic retrieval problem [7] and so standard techniques used in noisy harmonic retrieval can be used in this context. First of all because of noise Eq. (7) is not satisfied any more. We thus look for a solution that minimizes  $\|SH\|^2$  under the constrain that  $\|H\|^2 = 1$ . This is a classical total-least-square (TLS) problem that can be solved using Singular Value Decomposition (SVD).

The algorithm may be further improved by denoising  $S$  before applying TLS. This is done by using the Cadzow iterative algorithm [16].

Cadzow algorithm is based on the fact that, in the absence of any perturbation, the matrix  $S$  is Toeplitz and rank deficient (i.e., it has rank  $K$ , where  $K$  represents the number of Diracs in the signal). When noise is present  $S$  becomes full rank. So in the first step of the Cadzow iteration an SVD of  $S$  is performed leading to  $S = U\Lambda V$ , where  $\Lambda$  is a diagonal matrix. Then only the first  $K$  diagonal elements of  $\Lambda$  are kept and  $S$  is reconstructed. The new matrix  $S$  is now by construction rank deficient but is not Toeplitz anymore. This condition is then imposed by averaging the diagonal elements of  $S$ . The procedure is then iterated.

Finally, we further improve resilience to noise by exploiting the fact that the sampling kernels considered are of compact support. This means that in absence of noise many of the samples  $y_n$  are exactly zero. When noise is present this is not the case, we therefore set to zero the small observed samples which are probably carrying only noise and no signal information. The important point here is that after thresholding, groups of consecutive non-zero samples are separated by zero samples. We use this fact as an indication that the Diracs have generated samples that do not interfere with each other and can therefore be treated independently. We thus run the reconstruction algorithm on each group of non-zero samples independently.

The overall algorithm can summarized as follows:

- 1) Given the observed measurements  $\hat{y}_n$ , set to zero those whose amplitude is smaller than a predetermined threshold  $T_h$  (typically,  $T_h = 3\sigma$ ).
- 2) For each group of consecutive non-zero samples, **Do**
  - a) Construct the rectangular matrix  $S$ .
  - b) Estimate  $K$ .
  - c) Apply Cadzow iterative algorithm to  $S$ .

- d) Apply TLS method: Perform the singular value decomposition of  $S$  and choose the eigenvector  $[h_0, h_1, \dots, h_K]^T$  corresponding to the smallest eigenvalue.
- e) Compute the roots of  $H(z) = \sum_{k=0}^K h_k z^{-k}$  and retrieve the locations  $t_k$ ,  $k = 0, \dots, K - 1$ .

3) **End.**

An example of the behavior of the algorithm is shown in Fig. 2. In this example we have  $K = 6$  Diracs and we observe  $N = 128$  samples. The noiseless and the noisy samples are shown in Fig. 2(a), they are obtained using a real E-spline of order  $P = 13$ . In this example the SNR=5dB. The reconstructed Diracs are shown against the original signal in Fig. 2(b).

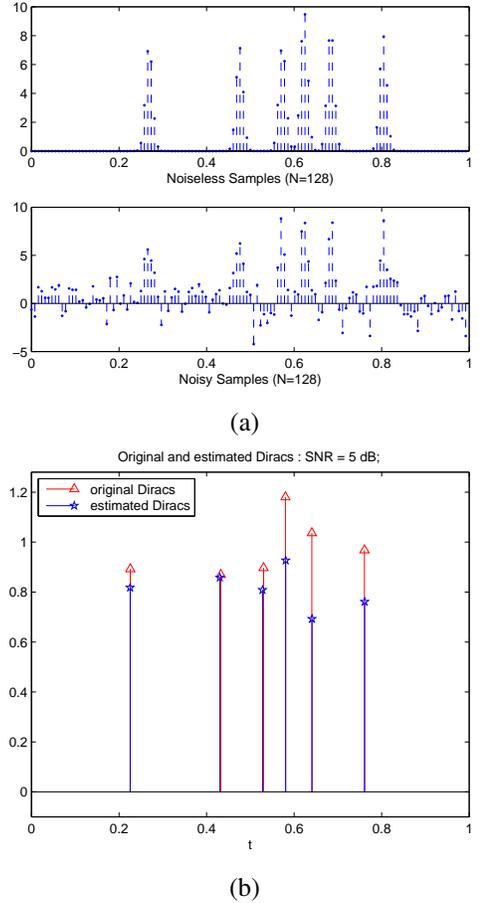


Fig. 2. Reconstruction of  $K = 6$  Diracs from  $N = 128$  noisy samples.

Notice that the sampling of FRI signals is equivalent to a parametric estimation problem. It is therefore possible to evaluate the performance of the algorithm by using the Cramer-Rao bounds (CRB). In [8], [17] such bounds were computed and it was shown that the proposed algorithm exhibit an almost optimal behavior since it can achieve the CRB up to noise levels of about 5dB.

#### IV. APPLICATIONS TO SIGNAL COMPRESSION

We have seen in the previous sections that specific classes of signals can be parsimoniously sampled using FRI sampling theory. Moreover, the sampling kernels involved include the scaling 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 we want to 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 connect 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 [11]. We thus assume that the signals we consider are made of pieces with Lipschitz regularity  $\alpha$ .

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 and it turns out that the rate-distortion behavior of this FRI-based algorithm is [18]:

$$D_{FRI}(R) \leq c_1 R^{-2\alpha} + c_2, \quad (8)$$

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 [19]:

$$D_{wave}(R) \leq c_3 R^{-2\alpha}. \quad (9)$$

This indicates that if the systematic error in (8) is sufficiently small the FRI-based algorithm which shift 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.

The R-D behavior derived in (8) has been validated through numerical simulations. In Figure 3, we show that the R-D decay of the proposed semi-parametric method is of the order  $R^{-2\alpha}$  as anticipated by the theory. The signal considered is depicted in Fig. 4 where the achieved reconstruction at 544 bits is shown and compared against an algorithm based only on the linear approximation of the wavelet coefficients.

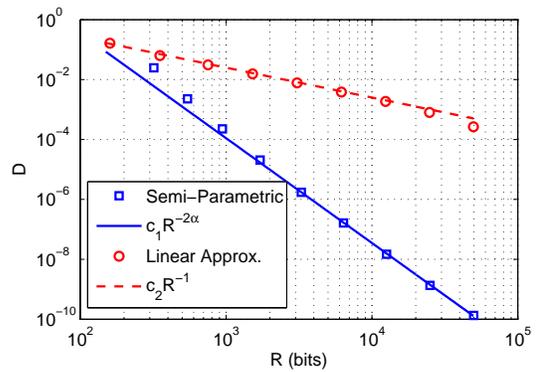


Fig. 3. R-D performance of the proposed semi-parametric reconstruction algorithm.

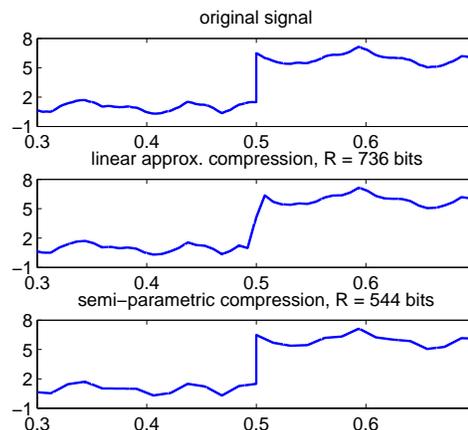


Fig. 4. Reconstruction of a piecewise smooth signal with the FRI-based reconstruction method discussed in Section IV.

#### V. CONCLUSIONS

Recent developments in sampling theory have shown that some classes of sparse signals can be sampled below the Nyquist rate.

In this paper, we have briefly reviewed the main aspects of the theory and then discussed variations of the method to handle noise and for compression. In both applications

numerical results have shown the potential impact of this new framework.

## REFERENCES

- [1] D. Donoho, "Compressed sensing," *IEEE Trans. Information Theory*, vol. 52(4), pp. 1289–1306, April 2006.
- [2] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principle: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Information Theory*, vol. 52(2), pp. 489–509, February 2006.
- [3] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. Signal Processing*, vol. 50(6), pp. 1417–1428, June 2002.
- [4] P. Dragotti, M. Vetterli, and T. Blu, "Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets Strang-Fix," *IEEE Trans. on Signal Processing*, vol. 55(5), pp. 1741–1757, May 2007.
- [5] M. Unser, "Splines: a perfect fit for signal and image processing," *IEEE Signal Processing Magazine*, vol. 16(6), pp. 22–38, November 1999.
- [6] M. Unser and T. Blu, "Cardinal Exponential Splines: Part I-theory and filtering algorithms," *IEEE Trans. on Signal Processing*, vol. 53(4), pp. 1425–1438, April 2005.
- [7] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Englewood Cliffs, NJ, Prentice-Hall, 2000.
- [8] T. Blu, P. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, "Sparse sampling of signal innovations: Theory, algorithms and performance bounds," *IEEE Signal Processing Magazine*, vol. 25(2), pp. 31–40, March 2008.
- [9] T. Blu and M. Unser, "Approximation errors for quasiinterpolators and (multi-) wavelet expansions," *Applied and Computational Harmonic Analysis*, vol. 6 (2), pp. 219–251, March 1999.
- [10] G. Strang and G. Fix, "A Fourier analysis of the finite element variational method," in *Constructive Aspect of Functional Analysis*, Rome, Italy, 1971, pp. 796–830.
- [11] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press, 1998.
- [12] J. Berent and P. Dragotti, "Perfect reconstruction schemes for sampling piecewise sinusoidal signals," in *Proc. of IEEE Int. Conf. on acoustic, speech and signal processing*, Toulouse (France), May 2006.
- [13] P. Shukla and P. Dragotti, "Sampling schemes for multidimensional signals with finite rate of innovation," *IEEE Trans. on Signal Processing*, vol. 55(7), pp. 3670–3686, July 2007.
- [14] I. Maravic and M. Vetterli, "Exact sampling results for some classes of parametric nonbandlimited 2-d signals," *IEEE Trans. on Signal Processing*, vol. 52(1), pp. 175–189, January 2004.
- [15] A. Hormati and M. Vetterli, "Annihilating Filter Based Decoding in the Compressed Sensing Framework," in *Wavelets XII, 2007*. [Online]. Available: [www.spie.org](http://www.spie.org)
- [16] J. Cadzow, "Signal enhancement a composite property mapping algorithm," *IEEE Trans. on Acoustic, Speech and Signal Processing*, vol. 36(1), pp. 149–62, January 1988.
- [17] P. Dragotti and F. Homann, "Sampling signals with finite rate of innovation in the presence of noise," in *Proc. of IEEE Int. Conf. on acoustic, speech and signal processing (ICASSP)*, Taipei (Taiwan), April 2009.
- [18] V. Chaisinthop and P. Dragotti, "Semi-parametric compression of piecewise smooth functions," in *Proc. of European Signal Processing Conference (EUSIPCO)*, Glasgow (UK), August 2009.
- [19] A. Cohen, I. Daubechies, O. Guleryuz, and M. Orchard, "On the importance of combining wavelet-based non-linear approximation with coding strategies," *IEEE Trans. on Information Theory*, vol. 48(7), pp. 1895–1921, July 2002.