MODEL ORDER SELECTION FOR SAMPLING FRI SIGNALS

Xiaoyao Wei and Pier Luigi Dragotti

Electrical and Electronic Engineering Department
Imperial College London
Email: ivy.wei@imperial.ac.uk, p.dragotti@imperial.ac.uk

ABSTRACT

Recently it has been shown that specific classes of non-bandlimited signals known as signals with finite rate of innovation (FRI) can be perfectly reconstructed by using appropriate sampling kernels and reconstruction schemes. The knowledge of the model order (i.e. the rate of innovation) is essential for correct reconstruction. In view of this, we devise an algorithm which can robustly identify the rate of innovation prior to the signal reconstruction in different noise levels and this extends the current scheme to a universal one that works with signals with unknown rate of innovation and using arbitrary kernels. We use the 'guaranteed performance' criterion to assess the performance and show a success rate close to 100% for SNR up to 10dB.

***Index Terms*—** finite rate of innovation, model order, sampling theory

1. INTRODUCTION

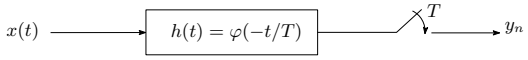


Fig. 1: A typical sampling set-up.

Finite Rate of Innovation (FRI) theory has demonstrated that it is possible to sample and perfectly reconstruct classes of continuous-time non-bandlimited signals [1, 2, 3]. Such signals are characterized by the fact that they are completely specified by a finite number of parameter per unit of time. For example a stream of pulses has a rate of innovation $2K$ if there are at most K pulses per unit of time. This is due to the fact that, given the knowledge of the pulse shape, only the K locations and K amplitudes of the pulses need to be retrieved in order to reconstruct the signal exactly.

FRI normally requires the acquisition setting to be as the one shown in Fig. 1, where the function $h(t)$ is the unit impulse response of the acquisition device. Under this model the samples are given by $y_n = \langle x(t), \varphi(t/T - n)\rangle$, where $\varphi(t) = h(-tT)$ is called the sampling kernel. The original FRI sampling theory has then been extended to multidimensional signals [4, 5, 6], alternative acquisition settings [7, 8, 9] and has also been applied in many domains (e.g. [10, 11]).

Typically, in noisy settings, estimation of FRI signals is performed using methods based around the SVD decomposition [12, 13, 14]. Examples include Cadzow iterative algorithm [15, 16] and matrix pencil methods [17]. Moreover, it has been observed [2, 18] that the performance of these algorithms for estimation of locations of Diracs achieves optimal results given by Cramér-Rao bound up to certain signal-to-noise ratios (SNR). The estimation departs from Cramér-Rao bound when SNR falls below a threshold due to the increasing probability of erroneous "outlier" estimates. In [18] we have explained the breakdown event by subspace swap (see also [19]) and our conjecture is that as long as there is no subspace swap event, the reconstruction algorithm is guaranteed to follow the performance given by the Cramér-Rao bound.

Traditional FRI schemes are not universal in that the sampling kernel needs to satisfy certain properties that depend on the rate of innovation of $x(t)$, and the rate of innovation needs to be known. For example, if the incoming signal is a stream of Diracs with at most K Diracs per unit of time, the rate of innovation is $\rho = 2K$ and the kernel is designed so that any stream of Diracs with $\rho \leq 2K$ can be reconstructed. However normally the same kernel cannot reconstruct signals with $\rho > 2K$ even if we increase the sampling rate. Moreover, estimation of ρ can be problematic in noisy settings, even though effective methods are now starting to appear [20, 21].

In this paper, we use the results in [22] to make the acquisition device independent of K, and we further extend the scheme by proposing a model selection algorithm, i.e. estimation of K, to achieve a *universal* scheme for sampling streams of Diracs with *unknown* rate of innovation and using *arbitrary* kernels. Our method does not require the knowledge of the statistic of the noise nor the variance of the noise. We use guaranteed no-breakdown performance to evaluate our proposed universal scheme. The core message is that if the minimum distance of the Diracs is above the value provided by the no-breakdown condition, then the correct K should be estimated. If it is below, the right selection $\tilde{K}$ should be below K. The algorithm is effective in noisy scenarios where we

This work is supported by the European Research Council (ERC) starting investigator award Nr. 277800 (RecoSamp).

show that it is more robust than the previous version [23] and achieve a success rate close to 100% for SNR up to 10dB.

The paper is organised as follows: in Section 2 we provide an overview of the theory of sampling FRI signals. In Section 3, we overview the guaranteed performance results of the FRI algorithm when using exponential reproducing kernel. In Section 4, we explain our proposed algorithm to enable universal sampling of signals with arbitrary FRI and explain how to use the guaranteed performance criterion to analyze its performance. Simulation results are then shown in Section 5. Finally we conclude in Section 6.

2. OVERVIEW OF FRI THEORY

2.1. Exact Framework

Consider the sampling set-up of Fig. 1. It has been shown in [3] that, when $\varphi(t)$ is an exponential reproducing function, perfect recovery of $x(t)$ from the samples $y_n = \langle x(t), \varphi(t/T - n)\rangle$ is possible. We say that $\varphi(t)$ is an exponential reproducing function of order P when, together with its integer shifts, it is able to reproduce $P+1$ complex exponentials:

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

for proper coefficients $c_{m,n}$.

Under this assumption, FRI algorithms compute the following weighted sums $s_m = \sum_n c_{m,n} y_n$, $m = 0, 1, ..., P$, where the weights $c_{m,n}$ are those in (1) that reproduce $\mathrm{e}^{\alpha_m t}$. When α_m is purely imaginary, it can be shown that s_m is exactly the Fourier transform of $x(t)$ at $\omega = \omega_m$ denoted by $\hat{x}(\alpha_m)$.

When $x(t)$ is a specific class of signals with FRI and α_m is chosen to be of the form $\alpha_m = \alpha_0 + m\lambda$, it is possible to establish a one-to-one mapping between $\hat{x}(\alpha_m)$ and $x(t)$. For example, if $x(t)$ is a stream of K Diracs located at t_k, that is

$$x(t) = \sum_{k=1}^{K} a_k \delta(t - t_k), \quad (2)$$

then the locations $\{t_k\}_{k=1}^K$ of the Diracs can be found from $\{\hat{x}(\alpha_m)\}_{m=0}^M$ by Prony's method (annihilating filter method [1, 3]), and then the amplitudes $\{a_k\}_{k=1}^K$ are retrieved by solving a linear system of equations. For retrieval of FRI signals in the presence of noise, where we assume the samples y_n are corrupted by white Gaussian noise with variance σ_y^2, Cadzow method [15] and matrix pencil [17] are proven to be effective. They are SVD-based methods and operate by splitting the measurement space into an estimated signal-subspace and an orthogonal-subspace and they work well when the noise has no significant effect on the signal subspace.

This formulation is not universal in that that it requires the acquisition device to behave like an exponential reproducing function and its order must be equal to or larger than the rate of innovation of the signal with FRI, specifically, for this example $P \geq 2K - 1$. This means that if the incoming signal has more than K Diracs, e.g. $K' > K$, it cannot be reconstructed with this kernel and this even when $N \geq 2K'$. Moreover, without knowing the order K in the parametric model (2) the reconstruction is not possible.

2.2. Approximate Framework

Recently, the FRI sampling theory has been extended so that any acquisition device can be used [22].

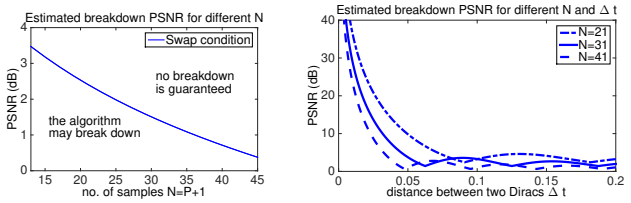
Consider an arbitrary kernel $\varphi(t)$. We want to find a linear combination of $\varphi(t)$ with its shifted versions that provides the best approximation to a specific exponential, more specifically, find coefficients c_n such that $\sum_{n\in\mathbb{Z}} c_n\varphi(t-n) \approx \mathrm{e}^{\alpha t}$. This approximation is exact only when the kernel $\varphi(t)$ satisfies the generalized Strang-Fix condition. For any other function, we use the coefficients $c_n = \hat{\varphi}(\alpha)^{-1}\,\mathrm{e}^{\alpha n}$ so that the approximation error is minimised.

We note that in the approximate framework N samples can give us N approximate exponentials and this directly relates the highest rate of innovation it can recover to the sampling rate rather than the order of the kernel. Hence any acquisition device is always usable for signals with arbitrary rate of innovation below the sampling rate. Although theoretically N samples are enough for recovering $N/2$ Diracs, in reality we require a slightly higher number of samples per unit of time since the Fourier coefficients are all approximated. Later on in Sec. 4 we propose an algorithm for identifying the rate of innovation and this will extend the approximate framework to a universal one that works with any arbitrary acquisition device and arbitrarily unknown number of Diracs.

3. GUARANTEED PERFORMANCE IN FRI

In both the exact or the arbitrary sampling framework, the performance of SVD-based estimation algorithms for robust FRI reconstruction achieves optimal results given by Cramér-Rao bound up to certain signal-to-noise ratio (SNR) and breaks down for smaller SNRs.

In [18], we have explained the breakdown event by subspace swap, which refers to the situation where due to noise the orthogonal subspace mixes with the signal subspace making the retrieval of the signal unreliable and this has been broadly recognised as the reason of performance breakdown in SVD-based parameter estimation algorithms [19]. In our work we have described a worst subspace swap scenario, and from there we have worked out the necessary condition for subspace swap to happen. In this way, we derived, for exponential reproducing kernels at which noise level the absence of subspace swap is guaranteed, i.e. no breakdown is guaranteed.



(a) one Dirac (b) two Diracs

Fig. 2: Necessary condition for subspace swap in different sampling settings

Specifically, when there is one Dirac the breakdown PSNR=$10\log_{10}\frac{a^2}{\sigma_y^2}$ is related to the order P of the kernel and the necessary condition for breakdown is shown in Fig. 2(a) by solid curve. When there are two Diracs with same amplitudes, the breakdown PSNR is related to the sampling period T, the distance Δt between the two Diracs, the frequency interval λ of α_m and the order of the kernel P. The necessary condition is shown in Fig. 2(b). Breakdown may happen for settings in the area below the curve and it is *guaranteed* no-breakdown will happen for the area above the curve. In other words, FRI achieves the Cramér-Rao bounds in settings above the curve. Note that the results in Fig. 2 is optimistic when it comes to an arbitrary sampling kernel setting, since only approximate reproduction of exponentials can be achieved in this case. In this setting, we estimate the breakdown condition empirically. The following explains how to measure the breakdown points.

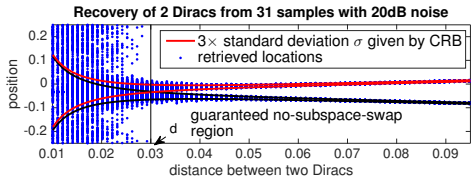


Fig. 3: Scatterplot of locations retrieved in 20dB noise using FRI reconstruction algorithm (Cadzow) compared to 3 times the standard deviation given by Cramér-Rao bounds.

In Fig. 3 we plot the estimated locations and the Cramér-Rao bounds for the situation where there are two Diracs with same amplitude sampled with the 5th order B-spline at the rate $1/T = 31$ in SNR=20dB noise, where SNR(dB) is defined as $10\log\frac{\|\mathbf{y}\|^2}{N\sigma_y^2}$. It shows that the location reconstruction algorithm in general achieves the Cramér-Rao bounds for distance d beyond some specific critical value. So we can say Dirac distance$> d$ is the no-breakdown condition when using the 5th order B-spline as the sampling kernel at sampling rate of 31Hz in noise level of SNR=20dB. When the no-breakdown condition is not satisfied, subspace swap may happen and these two Diracs becomes indistinguishable. The FRI reconstruction algorithm reconstructs them as one tall Dirac situated in between the true Diracs and one Dirac far away from the true Diracs with negligible amplitude. In this case, the more reasonable reconstruction should be neglecting the Dirac with negligible amplitude.

We can use no-breakdown condition to evaluate our proposed algorithm for model order selection (finding K in (2)). We aim for a model order selection algorithm that can correctly identify the number of Diracs when guaranteed no-breakdown condition is satisfied and that can automatically make the right selection of number of Diracs which should be smaller than K when subspace swap happens and adjacent Diracs cannot be resolved.

4. RATE OF INNOVATION IDENTIFICATION ALGORITHM

In this section we use the approximate framework and propose a novel method to identify the rate of innovation and this results in a universal scheme for sampling signals with unknown rate of innovation and using arbitrary kernels.

The idea of the algorithm is as follows. Given cN ($c > 1$) samples y_n of the input stream of Diracs $x(t)$ taken by an arbitrary kernel $\varphi(t)$, we are able to obtain cN approximated Fourier coefficients $\hat{x}(\alpha_m), m = 1, \ldots, cN$ using the method described in Sec. 2.2. From these Fourier coefficients we estimate at most $N/2$ Diracs.

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

We expect that the error on the samples will first decrease gradually when the number of Diracs p we assumed approaches the true number K. When we further increase p, the errors will either rise slightly or further decrease but at a much slower rate. The turning point will be our estimated number of Diracs. Note that to recognize the turning point we do not need to have p going all the way to $N/2$. In fact, the iteration can be stopped when ϵ_p increases and we denote this point by K_{stop}. We recognize the turning point by a scoring system which gives scores to all the numbers $p = 1, \ldots, K_{\text{stop}}$ based on resynthesis error ϵ_p on samples, first derivative of ϵ_p and second derivative of ϵ_p. More specifically, the score S_p for $p = 1, \ldots, K_{\text{stop}}$ is as follows:

$$S_p = \epsilon_p'' - 2|\epsilon_p'| - 2\epsilon_p - (\epsilon_1 - \epsilon_2)p. \tag{3}$$

The term ϵ'' in S_p rewards p which is a sudden turning point in ϵ_p, and $-2|\epsilon_p'|$ penalizes p where the speed of decreasing in resynthesis error is comparatively high. $-\epsilon_p$ penalizes p with large resynthesis error. At last, $-(\epsilon_p + (\epsilon_1 - \epsilon_2)p)$ penalizes large p that cannot reach the expected decreasing rate of resynthesis error. The estimated number of Diracs with

highest score will be our chosen number of Diracs. Note that this scoring system decides the number of Diracs by choosing the p with highest possibility among all instead of an absolute numerical criterion, it can operate regardless of the number of Diracs, signal to noise ratio, amplitudes of Diracs and does not require knowledge of the statistics of the noise.

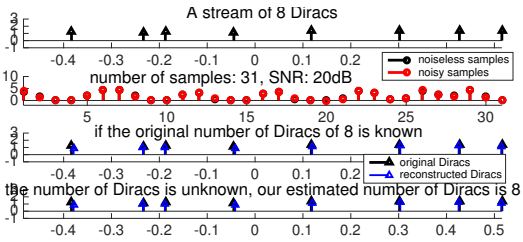
Once the number of Diracs K is determined, the input signal $x(t)$ can be recovered using the parametric model with correct order. We summarize the algorithm as follows:

Algorithm 1: Reconstruction of a stream of unknown number of Diracs $x(t)$ from its cN samples $\{y_n\}_{n=1}^{cN}$.

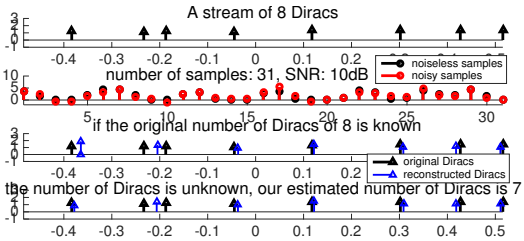
Obtain cN Fourier coefficients $\hat{x}(\alpha_m)$ from $\{y_n\}_{n=1}^{cN}$;
$p = 0$;
while *assumed number of Diracs* $p < N/2$ **and** $errDecreaseDetector$ **do**
  $p = p + 1$;
  Estimate location(s) $\hat{t}_k$ and amplitude(s) $\hat{a}_k$ of p Diracs from y_n (with Cadzow or matrix pencil);
 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

5. SIMULATIONS

In this part, we want to test the performance of the proposed algorithm. We want to show that our algorithm can identify the number of Diracs when the sampling setting is in the no-breakdown region, and when it is not the algorithm is able to choose the most reasonable number of Diracs. In the example of Fig. 4, there are 8 Diracs sampled at $1/T = 31$, in which two close-by Diracs have a distance satisfying the 20dB no-breakdown condition but not the 10dB no-breakdown condition. We notice that in 20dB noise the number of Diracs is identified correctly and all the 8 Diracs are accurately reconstructed. In 10dB noise subspace swap happens, FRI reconstruction methods are not able to resolve the two Diracs no matter what the number of Diracs we assumes, but $K = 7$ results in a better reconstruction. Our proposed algorithm makes the right selection of the number of Diracs when there is subspace swap event.



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



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

Fig. 4: Universal sampling of a stream of unknown number of Diracs using 5th order B-spline in the presence of noise.

In the following we do a more systematic test where we generate a stream of 4 Diracs with random locations and random amplitudes between 1 to 1.5, satisfying the minimum distance required by the no-breakdown condition. We reconstruct the unknown number of Diracs from 51 samples taken with the kernel of 5th order B-spline. We have 10000 realisations for each noise level.

The percentage of correct estimations for different level of noise is shown in Table 1. The result is compared with our previous version of identification algorithm which is slower and only based on the second derivative of resynthesis errors [23]. It shows that our proposed algorithm is much more robust than the previous version and can identify the number of Diracs with high success rate even in high level of noise.

Table 1: The percentage of correct identification of the number (4) of Diracs from 51 samples in different noise level. 10000 realisations for each noise level.

SNR	Inf.	20dB	15dB	10dB	5dB
alg. in [23]	100%	99.92%	97.38%	89.32%	63.55%
proposed alg.	100%	100%	100%	99.99%	95.72%

6. CONCLUSIONS

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

7. REFERENCES

[1] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Transactions on Signal Processing*, vol. 50, no. 6, pp. 1417–1428, 2002.

[2] 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.

[3] P. L. Dragotti, M. Vetterli, and T. Blu, "Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets Strang-Fix," *IEEE Transactions on Signal Processing*, vol. 55, no. 5, Part 1, pp. 1741–1757, 2007.

[4] P. Shukla and P. L. Dragotti, "Sampling schemes for multidimensional signals with finite rate of innovation," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3670–3686, 2007.

[5] H. Pan, T. Blu, and P. L. Dragotti, "Sampling curves with finite rate of innovation," *IEEE Transactions on Signal Processing*, vol. 62, no. 2, pp. 458–471, 2014.

[6] C. Chen, P. Marziliano, and A. C. Kot, "2D finite rate of innovation reconstruction method for step edge and polygon signals in the presence of noise," *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 2851–2859, 2012.

[7] H. A. Asl, P. L. Dragotti, and L. Baboulaz, "Multichannel sampling of signals with finite rate of innovation," *IEEE Signal Processing Letters*, vol. 17, no. 8, pp. 762–765, 2010.

[8] K. Gedalyahu, R. Tur, and Y. C. Eldar, "Multichannel sampling of pulse streams at the rate of innovation," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1491–1504, 2011.

[9] C. S. Seelamantula and M. Unser, "A generalized sampling method for finite-rate-of-innovation-signal reconstruction," *IEEE Signal Processing Letters*, vol. 15, pp. 813–816, 2008.

[10] X. Wei and P. L. Dragotti, "FRESH - FRI-based single-image super-resolution algorithm," *IEEE Transactions on Image Processing*, vol. 25, no. 8, pp. 3723–3735, 2016.

[11] J. Onativia, S. R. Schultz, and P. L. Dragotti, "A finite rate of innovation algorithm for fast and accurate spike detection from two-photon calcium imaging," *Journal of Neural Engineering*, vol. 10, no. 4, 2013.

[12] 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, no. 8, pp. 2788–2805, 2005.

[13] C. Aubel and H. Bölcskei, "Deterministic performance analysis of subspace methods for cisoid parameter estimation," in *2016 IEEE International Symposium on Information Theory*, July 2016, pp. 1551–1555.

[14] A. Jayawant and A. Kumar, "High rate quantization analysis for a class of finite rate of innovation signals," in *2016 24th European Signal Processing Conference*, Aug 2016, pp. 423–427.

[15] J. A. Cadzow, "Signal enhancement-a composite property mapping algorithm," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, no. 1, pp. 49–62, 1988.

[16] L. Condat and A. Hirabayashi, "Cadzow denoising upgraded: A new projection method for the recovery of dirac pulses from noisy linear measurements," *Sampling Theory in Signal and Image Processing*, vol. 14, no. 1, pp. 17–47, 2015.

[17] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, no. 5, pp. 814–824, 1990.

[18] X. Wei and P. L. Dragotti, "Guaranteed performance in the FRI setting," *IEEE Signal Processing Letters*, vol. 22, no. 10, pp. 1661–1665, 2015.

[19] D. W. Tufts, A. C. Kot, and R. J. Vaccaro, "The threshold analysis of svd-based algorithms," in *1988 International Conference on Acoustics, Speech, and Signal Processing*. IEEE, 1988, pp. 2416–2419.

[20] C. Gilliam and T. Blu, "Fitting instead of annihilation: Improved recovery of noisy FRI signals," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2014, pp. 51–55.

[21] C. Gilliam and T. Blu, "Finding the minimum rate of innovation in the presence of noise," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2016, pp. 4019–4023.

[22] J. A. Urigüen, T. Blu, and P. L. Dragotti, "FRI sampling with arbitrary kernels," *IEEE Transactions on Signal Processing*, vol. 61, no. 21, pp. 5310–5323, 2013.

[23] X. Wei and P. L. Dragotti, "Universal sampling of signals with finite rate of innovation," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2014, pp. 1803–1807.