Sampling Signals with Finite Rate of Innovation Project Report

Yang Zhao

I. Introduction

We investigate a new sampling approach for signals that are not bandlimited but still have a finite number of degrees of freedom per unit time. Such signals as streams of Diracs, nonuniform splines, and piecewise polynomials cannot be entirely recovered using sine sampling waves with finite frequency. Fortunately, it is demonstrated in [1] that with an appropriate choice of kernels, signals with finite rate of innovation (FRI) can be sampled with a rate below the Nyquist frequency and recovered without loss of information. [2] implements the sampling and reconstruction with physically realisable kernels, which enables applications as wideband communication and image super-resolution (SR).

In this project, we choose different polynomial reproducing kernels $\varphi(t)$ together with its shifted version to reproduce polynomials up to a certain degree, retrieve pulse locations and weights from the weighted sum of the observed samples by annihilating filter, examine the performance of total least squares (TLS) and Cadzow's algorithms against noise, and apply the new sampling scheme to merge multiple low-resolution (LR) images to form a super-resolved image with more pixels and details.

II. Theory and Methods

A. Sampling

Sampling indicates to represent a continuous-time signal by its value at discrete points. Denote the original signal as $x(t)$ and the sampling kernel $\varphi(t)$, the samples can be represented as:

$$y_n = \left\langle x(t), \varphi(\frac{t}{T} - n) \right\rangle = \int_{-\infty}^{+\infty} x(t)\varphi(\frac{t}{T} - n)dt \quad (1)$$

The sampling kernel is the time-reversed version of the filter input response. If we define a subspace spanned by shifted kernels $\{\varphi(t-n)\}$, sampling can be rendered as projecting the original signal onto the sampling subspace. The samples are of the compact size that only retains partial information of the original signal, but perfect reconstruction is possible if the kernel can suggest the rest. For instance, bandlimited signals can be projected onto the subspace spanned by sinc function and recovered by a linear combination of those kernels, as suggested by Shannon's sampling theorem. In this case, the information neglected during sampling can be adequately compensated by the properties of kernels. Therefore, it is significant to determine the basis according to the characteristics of the signal to be sampled.

B. Signals with Finite Rate of Innovation (FRI)

It has been demonstrated by [1] that sampling signals that are neither bandlimited nor in a fixed subspace is possible, as long as they have a parametric representation (i.e. polynomial, exponential, and rational) and finite number of degrees of freedom. Such signals are called those with FRI, and can be represented as:

$$x(t) = \sum_{k\in Z}\sum_{r=0}^{R} \lambda_{k,r} g_r(t - t_k) \quad (2)$$

where $\{g_r(t)\}_{r=0,1,...,R}$ is the chosen kernel, $\lambda_{k,r}$ is the coefficient, and t_k is the time shift. If $C_x(t_a, t_b)$ counts the number of free parameters of $x(t)$ during time interval $[t_a, t_b]$, the local rate of innovation at time t is defined as [1]:

$$\rho_\tau(t) = \frac{1}{\tau} C_x(t - \frac{\tau}{2}, t + \frac{\tau}{2}) \quad (3)$$

The sampling of bandlimited signals can be viewed as a particular case where $\rho = 2f_{\max}$.

C. Kernels and Coefficients

Kernels can be mainly categorised into three groups according to the signals to reconstruct: polynomial, exponential, and rational. In this project we only consider the polynomial reproducing kernels that can satisfy the Strang-Fix condition [3]:

$$\sum_{n\in Z} c_{m,n}\varphi(t - n) = t^m \quad (4)$$

Equation 4 suggests that such kernels and its shifted versions can replicate polynomials up to a certain order with an appropriate coefficient set. There is a restraint that the scaling function of wavelets with $N+1$ vanishing moments can only recreate polynomials of degree N [4]. If we choose our kernels as scaling functions of wavelets and assume $\{\varphi(t-n)\}_{n\in Z}$ spans a Riesz space, then the kernel and the dual basis $\left\{\widetilde{\varphi}(t-n)_{n\in Z}\right\}$ satisfy $\langle \varphi(t), \widetilde{\varphi}(t-n)\rangle = \delta_0$. Therefore, by multiplying both sides of Equation 4 with the dual basis, the coefficients can be expressed as:

$$c_{m,n} = \langle t^m, \widetilde{\varphi}(t-n)\rangle \quad (5)$$

Thus, the highest degree of the polynomial to be recovered should be known or at least estimated, so that we can decide the wavelets whose scaling function is to be the kernel. It is feasible to employ wavelets with large enough vanishing moments, but the computational complexity can be much higher. Once the basis is determined, the coefficients of all shifted kernels for all possible orders are fixed, which can be applied on samples to calculate the first $N+1$ moments of the signal. It is interesting to notice that B-spline is with the smallest possible support for the signal recreation, and any kernel can be decomposed as the convolution of B-splines and distribution $u(t)$ with $\int u(t)dt \neq 0$ [5].

D. Reconstruction of Diracs

Consider the case where the signal contains only K Diracs and the kernel can reproduce polynomials of maximum degree $N \geq 2K-1$. To estimate the locations and amplitudes, first, the first $N+1$ moments of the signal should be calculated with the samples and coefficients. The weighted sum of the observations τ is:

$$\tau_m = \sum\nolimits_n c_{m,n} y_n = \sum\nolimits_{k=0}^{K-1} a_k t_k^m, m = 0, 1, \ldots, N \quad (6)$$

These moments contain the information about the locations and amplitudes of all pulses. Then, we define the annihilating filter h_m such that its z-transform has roots correspond to the locations t_k. In the time domain, the relationship between filter coefficients and moments can be expressed as [4]:

$$h_m * \tau_m = \sum_{i=0}^{K} h_i \tau_{m-i} = \sum_{k=0}^{K-1} a_k t_k^m \underbrace{\sum_{i=0}^{K} h_i t_k^{-i}}_{0} = 0 \quad (7)$$

which can be written in matrix form:

$$\begin{pmatrix} \tau_{K-1} & \cdots & \tau_0 \\ \vdots & \ddots & \vdots \\ \tau_{N-1} & \cdots & \tau_{N-K} \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_K \end{pmatrix} = - \begin{pmatrix} \tau_K \\ \vdots \\ \tau_N \end{pmatrix} \quad (8)$$

Therefore, the annihilating filter can be uniquely constructed with signal moments. At least $2K$ moments are required to determine all filter coefficients to solve the equation. Thus, the kernel should be able to produce polynomials up to order $N \geq 2K-1$. As mentioned above, we can utilise the scaling function of wavelets that have high vanishing moments, but it results in an overdetermined system although the result can be the same. Once the annihilating filter is established, the locations can be retrieved by finding the roots of its z-transform.

Finally, the amplitudes of Diracs can be deduced from the Vandermonde system formed by the locations and moments:

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ t_0 & t_1 & \cdots & t_{K-1} \\ \vdots & \vdots & \ddots & \vdots \\ t_0^{K-1} & t_1^{K-1} & \cdots & t_{K-1}^{K-1} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{K-1} \end{pmatrix} = \begin{pmatrix} \tau_0 \\ \tau_1 \\ \vdots \\ \tau_{K-1} \end{pmatrix} \quad (9)$$

TABLE I
PARAMETERS IN SAMPLING AND RECONSTRUCTION

Number of Diracs	$K = 2$
Max degree of polynomials	$M = 3$
Vanishing moments	$N = 4$
Signal length	$L = 2048$
Sampling period	$T = 64$
Max amplitude	$A = 32$
Number of shifts	$H = \frac{L}{T} - 1 = 31$
Number of iterations	$I = \log_2(T) = 6$
Noise standard deviation	$\sigma = 0.1, 1, 10$

E. The Noisy Scenario

Assume the noise is addictive Additive white Gaussian noise (AWGN) and consider it directly at the moments due to the linearity. Since uncertainty is introduced into the Yule-Walker system modelled by Equation 8, more moments $N \geq 2K$ are needed to calibrate the annihilating filter. A larger moments matrix utilised in denoising approaches can be formed as:

$$S = \begin{pmatrix} \tau_K & \tau_{K-1} & \cdots & \tau_0 \\ \tau_{K+1} & \tau_K & \cdots & \tau_1 \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{N-1} & \tau_{N-2} & \cdots & \tau_{N-K-1} \end{pmatrix} \quad (10)$$

1) Total least-squares (TLS) algorithm: Equation 7 indicates that $Sh = 0$ in noiseless case. When there is noise, the annihilating filter is designed to minimise the norm $\|Sh\|$ with the constraint $\|h\| = 1$. This can be realised by singular value decomposition (SVD) $S = U\Lambda V^T$ and choose the last column of V as h. The following procedures are the same with noiseless cases.

2) Cadzow's Algorithm: Cadzow's algorithm is a denoising scheme before utilising TLS approach that is especially suitable for small signal-to-noise ratio (SNR) cases. The basic idea is that in noiseless case if we construct a more massive moment matrix S, it should be rank deficient and Toeplitz. Therefore, the noisy moment matrix can be manipulated to ensure both properties to reduce the influence of noise. First, SVD of S is performed, and only the most essential K entries of Λ are reserved while others are set to zero. Second, a new moments matrix is reconstructed by $S' = U\Lambda' V^T$. Third, we make the matrix Toeplitz by averaging along all diagonals. These steps are iterated until the new matrix is both rank deficient and Toeplitz. The result is considered denoised and then to be used in TLS algorithm.

III. Result and Analysis

A. Parameters

This simulation is based on the parameters presented by Table I unless otherwise mentioned.

B. Exercise 1

Scaling functions correspond to wavelets with $N+1$ vanishing moments together with its shifted version should be used to reproduce polynomials of degree N, which is 'dB4' for Daubechies wavelet. The length of the Daubechies filter is

$2N$. It is orthogonal since the scaling function $\varphi(t)$ satisfies $\langle \varphi(t), \varphi(t-n) \rangle = \delta_0$. Also, the length and smoothness of the scaling function $\varphi(t)$ and the wavelet function $\psi(t)$ can be larger when the vanishing moments increases.

The process of this part is as follows. First, a kernel of the same length L with the original signal is developed by padding zero at the end of the scaling function. Then, all possible shifted versions of the kernel are derived according to the number of shifts H available. Notice that instead of circular shift, the strategy removes the residue out of the signal range L to improve the stability and precision. Next, polynomials of order 0 to M are decomposed on the kernel set, and the coefficients of corresponding kernels $c_{m,n}$ are determined by Formula 5. These coefficients can be regarded as the projection length of the polynomials in the subspace spanned by kernels. Finally, the polynomials can be recovered by a linear combination of the kernel set with weights determined by coefficients $c_{m,n}$ as expressed by Formula 6. Figure 1 illustrates the reproduction of polynomials up to degree $M = 3$ by the scaling function of Daubechies 4 tap wavelet and its shifted version.

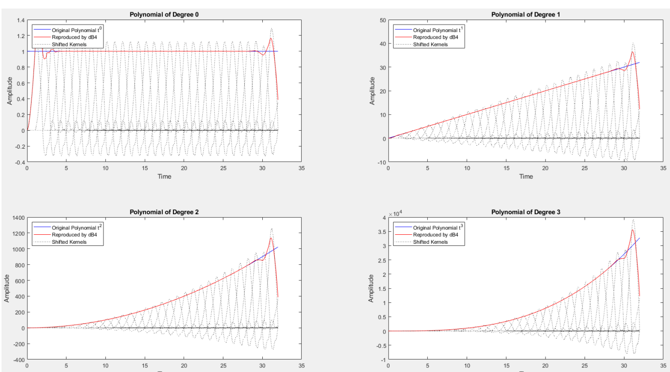

Fig. 1. Reproduction of polynomials of maximum degree $M = 3$ using scaling functions of 'dB4'

It can be observed that the dashed functions determined by the shifted version of the scaling function are applied with different scales correspond to coefficients, and their sums indicated by the red curves approximate the polynomials quite well apart from the endpoints. The reason is that the available kernels are insufficient when it comes to the edges, and the deviation can be higher than the central points. This problem can be mitigated by using a circular shift to generate kernels, but it can lead to a distortion of the weighted sum of the observed samples τ. Another solution is to increase the sampling rate, which can reduce the spacing between kernels and result in a more dense layout and narrow the distorted region.

C. Exercise 2

It is required to reconstruct the polynomials with the proper B-spline scaling function. Due to the biorthogonality, the decomposition should be based on the dual basis while the reconstruction requires the scaling function. [4] mentioned that B-spline of order N could reproduce polynomials of maximum degree N. Therefore cubic B-spline is utilised for polynomials of degree $0-3$.

B-spline of order N can be derived by [4]:

$$\beta_N(t) = \underbrace{\beta_0(t) * \beta_0(t) * \cdots * \beta_0(t)}_{N+1} \tag{11}$$

where $\beta_0(t)$ is the box function.

To obtain the cubic B-spline as the scaling function, the box function of length T is convolved four times in this case. Notice that in the simulation we add zeros to the beginning or the end of the result in each convolution to ensure it is a power of 2, and the peak happens at the centre. Then, the dual basis is derived by:

$$\widetilde{\Phi} = \Phi(\Phi * \Phi)^{-1} \tag{12}$$

where $G = \Phi * \Phi$ is the Gram matrix. In the implementation, the Gram matrix G is determined by:

$$G(v_1, \cdots, v_k) = \begin{pmatrix} \langle v_1, v_1 \rangle & \ldots & \langle v_1, v_k \rangle \\ \vdots & \ddots & \vdots \\ \langle v_k, v_1 \rangle & \cdots & \langle v_k, v_k \rangle \end{pmatrix} \tag{13}$$

where v_i is the value of the kernel. Afterwards, the coefficients $c_{m,n}$ is calculated by decomposition on the subspace spanned by shifted dual basis. Finally, the polynomials are reconstructed using coefficients $c_{m,n}$ and the shifted cubic B-spline set. The result is shown in Figure 2.

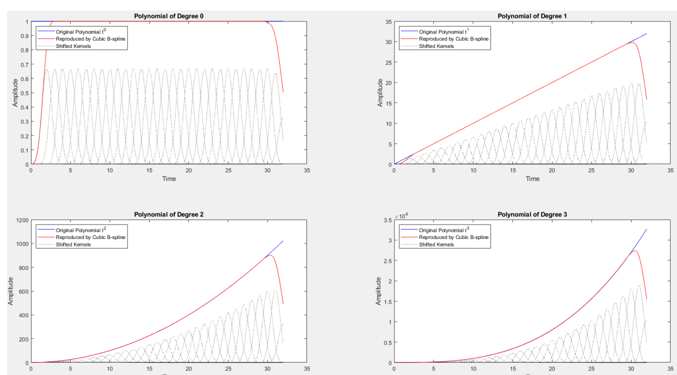

Fig. 2. Reproduction of polynomials of maximum degree $M = 3$ using cubic B-splines

The basic idea is similar as discussed in Exercise 1. Since the cubic B-spline is nonnegative, the amplitudes of components are smaller than the weighted sum. Consequently, the approximation is below the original signal for the deviations on the edges.

D. Exercise 3

The explanation of the annihilating filter method has been investigated in the theory section. It indicates that the location of Diracs can be estimated by the roots of the filter, and the relationship between coefficients and moments are represented by the Yule-Walker system as Equation 8. Once the locations are obtained, the corresponding amplitudes can be retrieved from the weighted samples τ by Equation 9. A plot of estimated Diracs is shown by Figure 3.

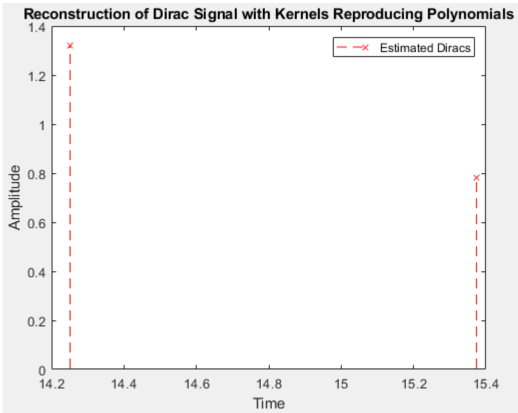

Fig. 3. Estimated Diracs for exercise 3

The locations are 14.25 and 15.375, while the amplitudes are 1.32 and 0.78 respectively.

E. Exercise 4

In this exercise, the locations and amplitudes are generated randomly. It is required that to recover K Diracs, the maximum degree of polynomials M should satisfy $M \geq 2K - 1$, which is determined to be 3 in this case to reduce the complexity. Therefore, the scaling function of the Daubechies filter 'dB4' is used to produce kernels. It is orthogonal so that the decomposition and reconstruction are based on the same kernel set. Then, the polynomials and coefficients are generated as described in the previous section. It is significant that once the filter is determined, the polynomials it can approximate and the corresponding coefficients $c_{m,n}$ are fixed, irrelevant to the Diracs to be sampled and reconstructed. Next, the signal $x[n]$ of length L that contains K Diracs is generated. The sampling process is as follows:

- Calculate the observations $y[n]$ by sampling the original signal $x[n]$ with the kernel set (the number of sampled points equals the number of kernels);
- Apply the coefficients $c_{m,n}$ correspond to the kernels on the samples $y[n]$ to obtain the weighted sum of the observed samples τ.

In this case, the degrees of possible polynomials are $0-3$, therefore there are four available τ. Equation 6 indicates that these τ include the information about locations and amplitudes of Diracs. Once all τ are obtained, the information of Diracs can be found as mentioned in Exercise 3. Figure 4 and Table II displays the simulation result.

It is demonstrated that this sampling and reconstruction approach works quite well. In this instance, the deviation is smaller than 0.2 percent for locations and no error for amplitudes. Nevertheless, in some case when the Diracs happen around zero, the estimation error can be significant. Figure 5 presents a typical situation.

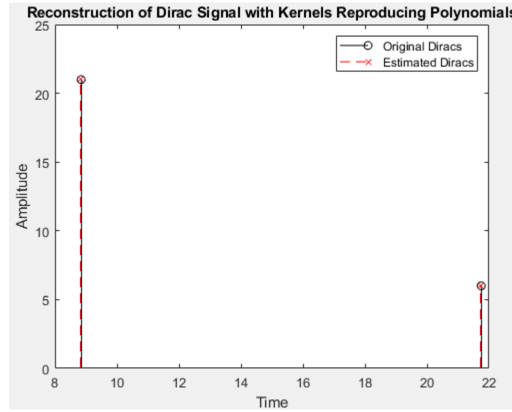

Fig. 4. Estimated Diracs for exercise 4 case a

TABLE II
EXACT AND ESTIMATED DIRACS

	Exact	Estimated
Location 1	8.84375	8.82812
Location 2	21.75	21.7344
Amplitude 1	21	21
Amplitude 2	6	6

Sampling is responsible for this problem. When the time is closed to zero, the corresponding part $a_k t_k^m$ is relatively small compared with the other components, and therefore the precision can be significantly affected. A possible solution is to start sampling before a pulse can occur, or leave a blank in the first section to ensure the accuracy.

F. Exercise 5

The sample is already given in this exercise. The idea of the reconstruction is the same as above. Figure 6 illustrate the result.

The estimated locations are 14.375 and 17.5 while the amplitudes are 2.63 and 1.48 respectively.

G. Exercise 6 and 7

In these exercises, we add noise to the system and vary the number of moments N from 5 to 10 to examine the performance of TLS and Cadzow algorithms. The maximum degree of polynomials $M = N - 1$ are $4 - 9$. Filters applied are still Daubechies with length $2N$ which produces orthogonal scaling functions, but the variation of the vanishing moments N leads to different kernels, coefficients, and hence the accuracy. It can be found that with the increasing of vanishing moments, the support of the scaling function and the smoothness of kernels are improved, but more complexity is required. Noise with a stand deviation $\sigma = 0.1, 1, 10$ is added to the weighted sum τ directly due to the linearity.

Subsequently, the result of TLS and Cadzow approaches are compared. The explanation of both algorithms has been covered in the theory section. The criterion of TLS is that the production of moment matrix S and annihilating filter h $Sh = 0$ in the noiseless case, which should be as small as possible for noisy cases. The demand for designing such an

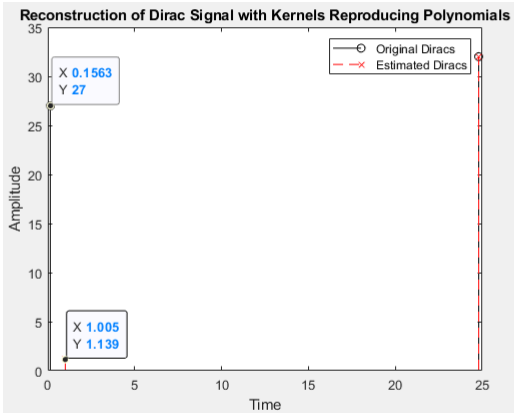

Fig. 5. Estimated Diracs for exercise 4 case b

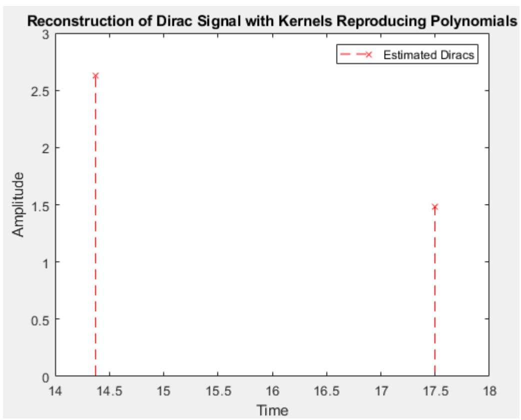

Fig. 6. Estimated Diracs for exercise 5

annihilating filter h to minimise the production can be realised by SVD of S and regard the last column as h. However, if the SNR is relatively low, the result can be improved by Cadzow denoising. It is mentioned in [4] that in noiseless case, S should be of rank K and Toeplitz, while it is full rank when noise exists. Nevertheless, our algorithm that generates Diracs randomly cannot verify the statement for minority cases. The proposed S is always Toeplitz, but the rank is not always the number of Diracs K. It means there may be some errors in the generation of τ, but unfortunately, the problem is not found until this report is written. Cadzow manipulates the moment matrix S to ensure it satisfies both Toeplitz and of rank K to minimise the influence of noise by the iteration discussed in the theory section. Figure 7 8 9 present the result when noise varies.

It can be observed that when the noise level is low, both algorithms can estimate the locations and amplitudes efficiently. For low SNR cases, the solutions proposed by the Cadzow strategy are closer to the actual Diracs than TLS. Therefore, Cadzow is proved as an effective extra denoising method. Moreover, the outputs indicate the accuracy can be improved significantly with the increase of moments N. The reason is that the overdetermined system provides more information

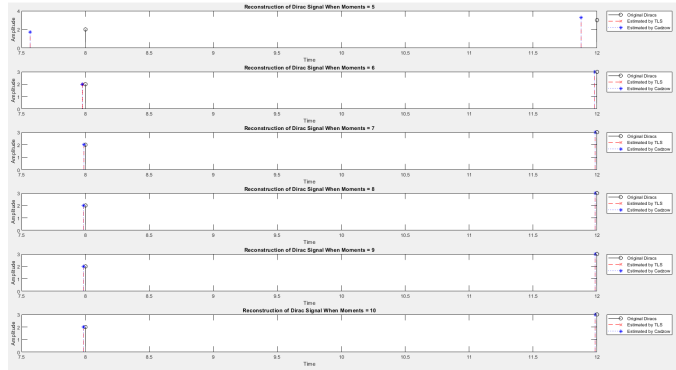

Fig. 7. Reconstruction comparison when $\sigma = 0.1$

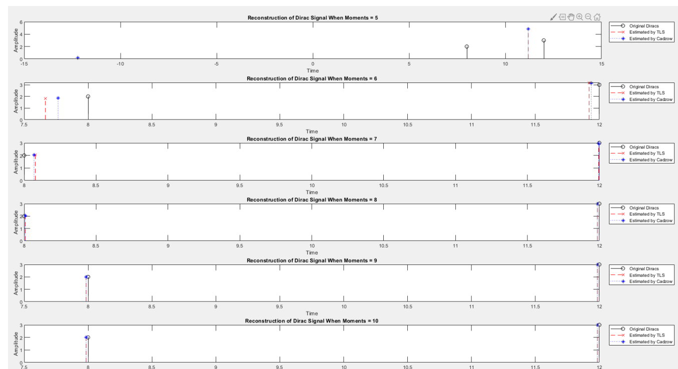

Fig. 8. Reconstruction comparison when $\sigma = 1$

when there are more equations (i.e. larger moments). Although those equations are inaccurate, the overall effect can be further reduced with more known conditions, and thus the result can be closer to the real value.

Nevertheless, due to the problem of rank mentioned above, there can be 'overfitting' problem when the rank of noiseless moment matrix S is not the number of Diracs K. Figure 10 shows a representative case when $\sigma = 10$.

Although TLS and Cadzow algorithms can solve the noise problem effectively, there appears to be a bug in the simulation which limits its application for all random Diracs.

H. Exercise 8

An application in image super-resolution is investigated in this exercise. An SR image is generated by fusing numerous LR images that observe the scene from different viewpoints. Our primary focus is image registration. First, the LR images are registered concerning the first image, in which the samples below the threshold are set to zero to reduce noise. Notice that to improve the peak signal to noise ratio (PSNR), the threshold of red, green, blue should be different. The optimal value (0.28 for red, 0.29 for green, 0.295 for blue) of the picture set utilised is determined by manual dichotomy, but it can be realised with code for more generalised cases. Then, moments and barycenters are computed for each colour layer. Finally, the first figure is selected as the reference, and the shift of barycenters are calculated. After registration, image

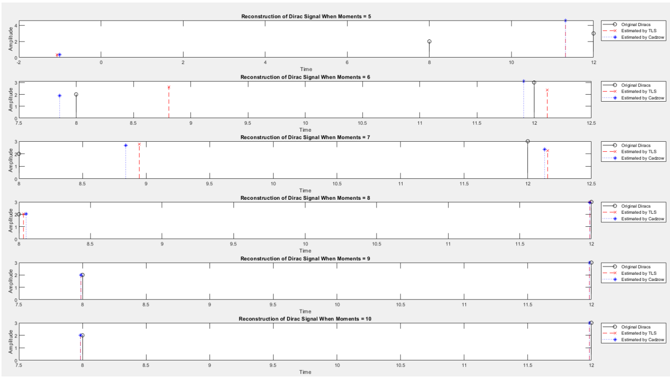

Fig. 9. Reconstruction comparison when $\sigma = 10$

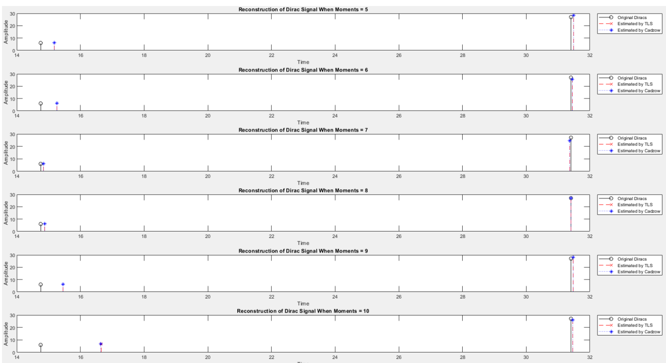

Fig. 10. Inaccurate estimation of Diracs occur near zero when $\sigma = 10$

fusion is realised by fusing all LR images on an HR grid and estimate the missing data, while restoration is achieved by removing blur and noise. In this case, 40 LR images of 64×64 pixels are fused to develop an HR image of 512×512 pixels. The result is shown in Figure 11.

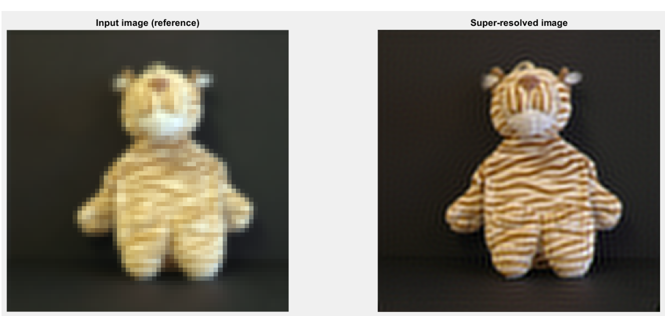

Fig. 11. Comparison of the low-resolution and super-resolution image

The PSNR of this SR image is 24.67 dB compared with the original high-resolution (HR) photo. The result demonstrates the feasibility of the examined sampling algorithm.

IV. Conclusion

In this experiment, we examined a new sampling and reconstruction algorithm for FRI signals by focusing on the observation and reproduction of Diracs using polynomial reproducing kernels. It is demonstrated that the polynomials can be approximated with shifted versions of scaling functions, and the corresponding coefficients can be utilised on the samples to develop the weighted sum from which the locations and amplitudes of Diracs can be retrieved through the annihilating filter method. Also, the effectiveness of TLS and Cadzow denoising algorithm have been proved in the noisy case. An application in image super-resolution has been investigated and the result provided a high PSNR as expected. Nevertheless, there appears to be an issue in random pulses sampling. In the noiseless case, the rank of the generated moment matrix is not always the number of Diracs. It leads to an 'overfitting' when the moments is high which reduces the accuracy of the result. Further optimisations can be developed with the code provided in the appendix.

V. Appendix: MATLAB code

The source code can be accessed via https://github.com/SnowzTail/.

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] 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.

[3] G. Strang and T. Nguyen, *Wavelets and filter banks*. SIAM, 1996.

[4] P. L. Dragotti, "Wavelets and applications," September 2018.

[5] T. Blu, P. Thcvenaz, and M. Unser, "Moms: Maximal-order interpolation of minimal support," *IEEE Transactions on Image Processing*, vol. 10, no. 7, pp. 1069–1080, 2001.