Reconstruction of FRI Signals using Autoencoders with Fixed Decoders

Vincent C. H. Leung*, Jun-Jie Huang*, Yonina C. Eldar†, and Pier Luigi Dragotti*
*Department of Electrical and Electronic Engineering, Imperial College London, United Kingdom
†Faculty of Mathematics and Computer Science, Weizmann Institute of Science, Rehovot, Israel
*{chi.leung14, j.huang15, p.dragotti}@imperial.ac.uk, †yonina.eldar@weizmann.ac.il

***Abstract*—Finite Rate of Innovation (FRI) theory considers sampling and reconstruction of classes of non-bandlimited signals that have a small number of free parameters. The task of reconstructing continuous FRI signals from discrete samples is often transformed into a spectral estimation problem and solved using methods involving estimating signal subspaces. These techniques tend to break down at a certain peak signal-to-noise ratio (PSNR). To avoid this inherent breakdown, we consider an alternative learning-based approach that uses autoencoders with fixed decoders. We propose to determine the parameters of the decoders based on the information of the sampling kernel explicitly. The fixed decoders provide a regularizing effect on the output of the encoder and lead to a robust network. Simulations show significant improvements on the breakdown PSNR over both classical subspace-based methods and our previous work based on deep neural networks.**

***Index Terms*—Finite rate of innovation, neural network, autoencoders, signal reconstruction, deep learning.**

I. Introduction

Classical sampling theory has enabled us to perfectly reconstruct continuous bandlimited signals from their discrete samples. In recent years, the emergence of finite rate of innovation (FRI) sampling theory [1]–[6] has extended it to classes of non-bandlimited signals that have finite degrees of freedom per unit time. A common example is a stream of K Diracs, which has a $2K$ rate of innovations as the signal can be defined by the amplitudes and locations of the K Diracs.

Fig. 1 illustrates a typical acquisition process that involves filtering the input continuous signal $x(t)$ with $h(t) = \varphi(-t/T)$ and sampling at a regular interval $t = nT$. Perfect reconstruction of non-bandlimited FRI signals can be achieved by using specific classes of kernels $\varphi(t)$ (e.g. [1], [2], [4], [5]). Conventionally, the task is then transformed into a spectral estimation problem that can be solved by methods such as Prony's method with Cadzow denoising [7], [8] and matrix pencil [9]. These methods make use of the Singular Value Decomposition (SVD) to estimate signal subspaces. Under noisy conditions, the reconstruction performance follows the Cramér-Rao bound in the low noise regime [10], [11] until it breaks down when the peak signal-to-noise ratio (PSNR) drops below a certain threshold. The reason is conjectured to be the subspace swap event [12] which refers to the confusion of the orthogonal subspace with the signal subspace under noisy conditions [13].

Vincent C. H. Leung is funded by the President's PhD Scholarships of Imperial College London.

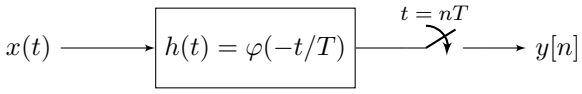


Fig. 1. Acquisition process that converts continuous time signal $x(t)$ into discrete time samples $y[n] = \langle x(t), \varphi(t/T - n)\rangle$.

Our aim is to avoid the inherent subspace swap events in subspace-based methods. In our previous work [14], we proposed solving the original FRI reconstruction problem instead of the transformed spectral estimation problem by directly inferring the locations of the Diracs from the noisy samples using deep neural networks. Simulation results have shown improvement over the low PSNR region with a slight compromise in the high PSNR region. Despite promising signs that solving the original reconstruction problem using neural networks allows us alleviate subspace swap events, the information of the sampling kernel and the knowledge of the true amplitudes of the Diracs in the training data are yet to be utilized in learning the network.

Here we propose to reconstruct FRI signals from noisy samples using autoencoders with fixed decoders to incorporate both pieces of information. The encoder $g_\phi(\cdot)$ follows the same structure as [14] and aims to estimate the locations of the Diracs from the noisy samples. The fixed decoder $f_\theta(\cdot)$ is then used to resynthesize the samples of the FRI signal based on the estimated locations given the true amplitudes of the Diracs in the training data. We propose to determine the parameters of the decoder θ based on the information of the sampling kernel while learning the parameters of the encoder ϕ through backpropagation. The loss function considers the error on both estimated locations and the corresponding discrete samples. This provides a regularizing effect on the output of the encoder network and improves its performance in retrieving the pulse locations of the FRI signals.

The rest of the paper is organized as follows: In Section II, we discuss the inherent subspace swap event and breakdown PSNR in classical FRI methods using the example of reconstructing a stream of K Diracs. We then present our proposed approach using autoencoder networks with fixed decoders in Section III. In Section IV, we compare simulation results of our technique with classical subspace-based methods and our previous work. We then conclude in Section V.

II. Breakdown in subspace-based methods

In this section, we explain the breakdown by going through the conventional reconstruction methods of the most basic FRI signal, a stream of K Diracs. To compare with the subspace-based techniques in their optimal settings, in this paper, we consider the reconstruction of a τ-periodic stream of K Diracs:

$$x(t) = \sum_{l\in\mathbb{Z}} \sum_{k=0}^{K-1} a_k \delta(t - t_k - l\tau), \tag{1}$$

where $\{a_k \in \mathbb{R}\}_{k=0}^{K-1}, \{t_k \in \mathbb{R}\}_{k=0}^{K-1}$ are the amplitudes and locations of the Diracs respectively. To sample the continuous signal $x(t)$, we use an exponential reproducing kernel $\varphi(t)$ that can reproduce complex exponentials:

$$\sum_{n\in\mathbb{Z}} c_{m,n}\varphi(t-n) = e^{j\omega_m t}, \tag{2}$$

with $\omega_m = \omega_0 + m\lambda$ for $m = 0, 1, ..., P$. Assuming sampling period $T = \tau/N$, it is possible to map the acquired samples

$$y[n] = \left\langle x(t), \varphi\left(\frac{t}{T} - n\right)\right\rangle = \sum_{k=0}^{K-1} a_k \varphi\left(\frac{t_k}{T} - n\right), \tag{3}$$

into a sum of exponentials:

$$\begin{aligned} s[m] &= \sum_{n=0}^{N-1} c_{m,n} y[n] = \sum_{k=0}^{K-1} a_k \sum_{n\in\mathbb{Z}} c_{m,n}\varphi\left(\frac{t_k}{T} - n\right) \\ &= \sum_{k=0}^{K-1} \underbrace{a_k e^{j\omega_0 t_k/T}}_{b_k} \left(\underbrace{e^{j\lambda t_k/T}}_{u_k}\right)^m = \sum_{k=0}^{K-1} b_k u_k^m. \end{aligned} \tag{4}$$

The amplitudes of the Diracs $\{a_k\}_{k=0}^{K-1}$ are mapped to the amplitudes of the exponentials $\{b_k\}_{k=0}^{K-1}$ while the locations of Diracs $\{t_k\}_{k=0}^{K-1}$ are transformed to $\{u_k\}_{k=0}^{K-1}$. This forms a spectral estimation problem that can be solved using subspace-based methods. Particularly, we are interested in retrieving the locations of the Diracs $\{t_k\}_{k=0}^{K-1}$ due to its non-linear nature in the problem seen in (4). On the contrary, the problem of retrieving the amplitudes of the Diracs is linear, which means that given the locations of the Diracs, we can directly estimate the amplitudes.

Previous works such as [3] have shown that subspace-based methods achieve optimal reconstruction performance defined by the Cramér-Rao bound yet break down at a certain PSNR threshold. It is conjectured [13] that the breakdown in subspace-based methods is due to the confusion between noise and signal subspaces in performing spectral estimation. In [12], a mathematical relationship was drawn between the breakdown PSNR and the relative distance between neighboring Diracs $\Delta t_k/T$ with $\Delta t_k = t_{k+1} - t_k$. For instance, when there are two Diracs of same amplitude ($K = 2, a_0 = a_1$), the necessary condition for the subspace swap event is when

$$\text{PSNR} < 10\log_{10} \frac{8\left(\frac{P}{2}+1\right)\ln\left(\frac{P}{2}+1\right)}{\left(\frac{P}{2}+1-\frac{\sin(\frac{\lambda}{2}(\frac{P}{2}+1)\Delta t_0/T)}{\sin(\frac{\lambda}{2}\Delta t_0/T)}\right)^2}. \tag{5}$$

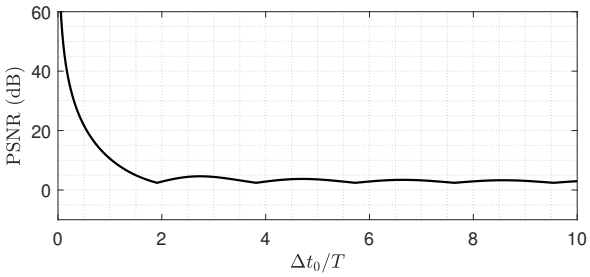


Fig. 2. Relationship between breakdown PSNR and the distance between Diracs in the case of $K = 2, N = 21 = P + 1$ and $\lambda = \frac{2\pi}{N}$ (after [12]).

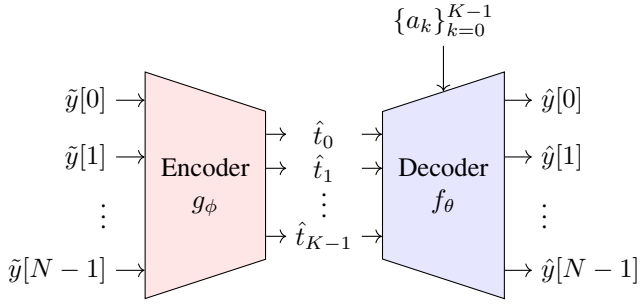


Fig. 3. The encoder maps the input noisy samples $\{\tilde{y}[n]\}_{n=0}^{N-1}$ to the estimated locations $\{\hat{t}_k\}_{k=0}^{K-1}$ of the Diracs. In the training phase, since the true amplitudes of training data are known, they are directly fed into the decoder. Given the estimated locations and true amplitudes, the decoder resynthesizes the noiseless samples $\{\hat{y}[n]\}_{n=0}^{N-1}$. Since the pulse shape is known, the parameters θ of the decoder are fixed and do not require learning. The parameters ϕ of the encoder are instead learned and the goal is to minimize the loss function in (10). In the evaluation phase, we only need the encoder to provide us the estimated locations. Given the locations, the amplitudes can be directly estimated using least squares method fitting $\left\{\varphi\left(\hat{t}_k/T - n\right)\right\}_{n=0}^{N-1}$ to $\{\tilde{y}[n]\}_{n=0}^{N-1}$ from the relationship described in (3).

This is visualized in Fig. 2, which shows that the smaller the distance between two nearby Diracs, the higher the breakdown PSNR will be. Thus, the subspace swap event occurring inherently in current FRI techniques precludes us from recovering FRI signals with a higher resolution under strong noise.

III. Proposed Method

In this section, we present an FRI reconstruction approach using autoencoders. We propose to construct the autoencoder using encoders learned through backpropagation and fixed decoders. We start by describing their functionality before delving into the architecture and loss function design in detail. A block diagram of the proposed system in shown in Fig. 3.

A. Encoder Design and Network Architecture

For the encoder $g_\phi(\cdot) : \mathbb{R}^N \to \mathbb{R}^K$, we use the same architecture as in our previous work in [14], which infers the locations of Diracs directly from the noisy samples, i.e. $\hat{t_k} = g_\phi(\tilde{y}[n])$. This is because the occurrence of the subspace swap event analyzed in [12] is inherent to subspace-based reconstruction methods. By directly inferring the locations, we hope to avoid the subspace swap events.

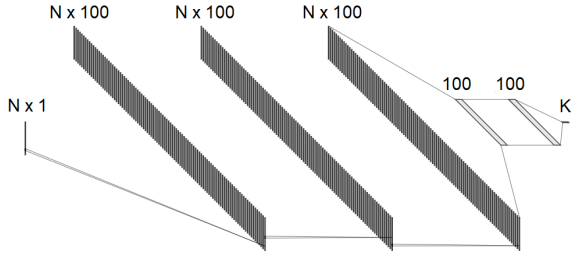


Fig. 4. Encoder network architecture to perform inference from the observed noisy samples $\{\tilde{y}[n]\}_{n=0}^{N-1}$ to the locations of Diracs $\{t_k\}_{k=0}^{K-1}$.

As shown in Fig. 4, the encoder consists of 3 convolutional layers followed by 3 fully connected (FC) layers of sizes $100, 100, N$ respectively. Each of the convolutional layers has 100 filters of size 3. Rectified linear unit (ReLU) is used as the activation function between each two layers.

B. Decoder Design and Network Architecture

The decoder $f_\theta(\cdot) : \mathbb{R}^K \to \mathbb{R}^N$ transforms the estimated locations back to the denoised samples using fully connected networks and ReLU, i.e. $\hat{y}[n] = f_\theta(\hat{t_k}) = f_\theta\left(g_\phi\left(\tilde{y}[n]\right)\right)$. While it is possible to learn a decoder through backpropagation, we propose to use an explicit construction approach to determine the parameters of the decoder network based on the sampling kernel and the true amplitudes of the pulses. During training, the fixed decoder provides an implicit and accurate regularization on the estimated pulse locations of the encoder network and will, therefore, improve the learning of the encoder network. During testing, only the encoder is used to estimate the pulse locations from the noisy samples.

As works such as [15]–[17] have suggested the capability of ReLU networks being turned into universal approximator of any arbitrary function of compact support, the decoder can follow the same framework and be used as an approximator of $\varphi(t)$. Theoretically, the approximation framework in [15], [16] allows us to approximate any kernel with an arbitrary and non-uniform resolution with ReLU networks. In this paper, we will focus on piecewise linear interpolation of $\varphi(t)$ with a step size of Δ. The process of approximating $\varphi(t)$ with a compact support L can be described as:

$$\hat{\varphi}_\Delta(t) = \sum_{i=0}^{I-1} d_i \text{ReLU}(t - i\Delta), \tag{6}$$

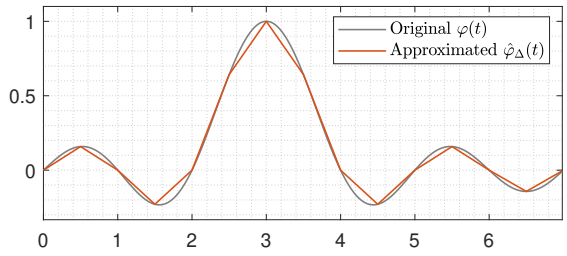


Fig. 5. A comparison of an arbitrary sampling kernel $\varphi(t)$ and its corresponding piecewise linear approximation $\hat{\varphi}_\Delta(t)$ using ReLU networks with a uniform step size of $\Delta = 1/2$.

TABLE I
DYNAMICS AND NUMBER OF OUTPUTS AT EACH LAYER OF THE DECODER

Layer	Output at Each Layer	# of Outputs
Input	$\{\hat{t}_k\}_{k=0}^{K-1}$	K
FC1	$\left\{\frac{\hat{t}_k}{T} - n\right\}_{k=0,n=0}^{K-1,N-1}$	KN
FC2+ReLU	$\left\{\text{ReLU}\left(\frac{\hat{t}_k}{T} - n - i\Delta\right)\right\}_{k=0,n=0,i=0}^{K-1,N-1,I-1}$	KNI
Output	$\left\{\hat{y}[n] = \sum_{k=0}^{K-1} a_k \hat{\varphi}_\Delta\left(\frac{\hat{t}_k}{T} - n\right)\right\}_{n=0}^{N-1}$	N

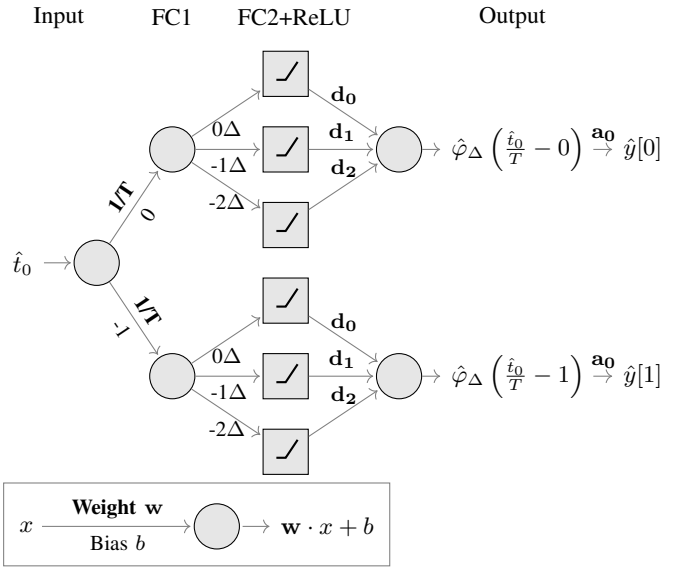


Fig. 6. An example of decoder architecture for acquiring $N = 2$ samples from sampling a stream of $K = 1$ Dirac using approximated sampling kernel $\hat{\varphi}_\Delta(t)$ with $I = 3$ linear segments.

with

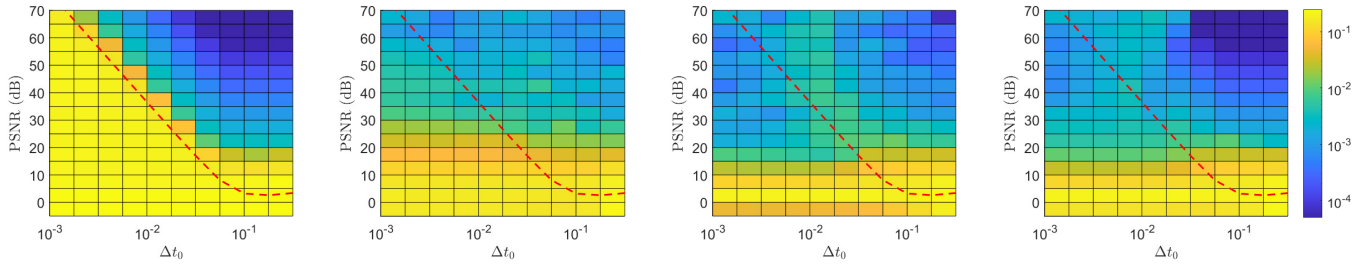
$$d_i = \frac{\varphi\left((i+1)\Delta\right) - \varphi(i\Delta)}{\Delta} - d_{i-1} \quad \text{and} \quad d_0 = 0. \tag{7}$$

As the amplitudes are known for the training phase, we can substitute (6) into (3) and express the estimated samples $\{\hat{y}[n]\}_{n=0}^{N-1}$ as

$$\hat{y}[n] = \sum_{k=0}^{K-1} a_k \hat{\varphi}_\Delta\left(\frac{\hat{t}_k}{T} - n\right) \tag{8}$$

$$= \sum_{k=0}^{K-1} a_k \sum_{i=0}^{I-1} d_i \text{ReLU}\left(\frac{\hat{t}_k}{T} - n - i\Delta\right). \tag{9}$$

By utilizing I ReLU units, we are effectively dividing the sampling kernel into I linear segments. Therefore, the total number of linear segments is given by $I = L/\Delta$. Fig. 5 displays an example of approximating an arbitrary kernel using ReLU networks. When step becomes infinitely small $\Delta \to 0$, $\hat{\varphi}_\Delta(t)$ will ultimately converge to the original sampling kernel $\varphi(t)$. A special note here is that both amplitudes and the sampling kernel information are only required during training.



(a) Prony's method with Cadzow denoising.

(b) Direct inference using DNN [14] (equivalent to training with only the encoder).

(c) Our proposed autoencoder with its encoder initialized from the trained model in (b).

(d) Our proposed method in (c) coupled with gradient descent using least-squares fitted amplitudes.

Fig. 7. Average standard deviation of the retrieved locations of a stream of Diracs ($N = 21, K = 2$) over 1000 realizations at each PSNR-Δt_0 pair using different methods. The red dotted line refers to the breakdown PSNR using subspace-based methods shown in Eq. (5) [12].

In evaluation stage, we would only need the encoder to infer the locations of Diracs from the noisy samples.

To implement this framework, the decoder consists of 3 fully connected hidden layers of sizes KN, KNI, N respectively. The detailed dynamics is listed in Table I, which performs the transformation from $\{\hat{t}_k\}_{k=0}^{K-1}$ to $\{\hat{y}[n]\}_{n=0}^{N-1}$. Here $\{\hat{t}_k\}_{k=0}^{K-1}$ are the estimated locations produced by the encoder. An example decoder for $N = 2, K = 1, I = 3$ is also shown in Fig. 6. As the decoder is modeled from approximating the true sampling kernel $\varphi(t)$, the weights and the biases of the decoder will be fixed and frozen during training.

C. Loss Function

Contrary to standard autoencoders, since we would like our recovered samples to be denoised, the loss function is the squared error between the output estimated samples $\{\hat{y}[n]\}_{n=0}^{N-1}$ and the ground truth noiseless samples $\{y[n]\}_{n=0}^{N-1}$.

Furthermore, we impose a constraint on the bottleneck by including the squared error between the estimated locations $\{\hat{t}_k\}_{k=0}^{K-1}$ and the ground truth locations $\{t_k\}_{k=0}^{K-1}$. Together, the resultant loss function can be written as

$$\mathcal{L}(\hat{\mathbf{t}}) = \sum_{n=0}^{N-1} \left(\hat{y}[n] - y[n]\right)^2 + \sum_{k=0}^{K-1} \left(\hat{t}_k - t_k\right)^2. \tag{10}$$

Backpropagation with Adam optimizer [18] is used for learning the encoder, while the decoder remains fixed to ensure the latent variable refers to the estimated location of the Diracs.

IV. Simulation Results

In this section, we compare the performance of our proposed autoencoders with the subspace-based Prony's method with Cadzow denoising [7], [8], as well as with our previous work that performs direct inference of FRI parameters from the noisy samples. The evaluation metric is the average standard deviation of the retrieved locations of Diracs, defined as:

$$SD_{avg} = \frac{1}{K} \sum_{k=0}^{K-1} \sqrt{\frac{\sum_{j=0}^{J-1} \left(\hat{t}_k^{(j)} - t_k\right)^2}{J}}, \tag{11}$$

where $\hat{t}_k^{(j)}$ and J are the j-th estimation and the number of realizations respectively.

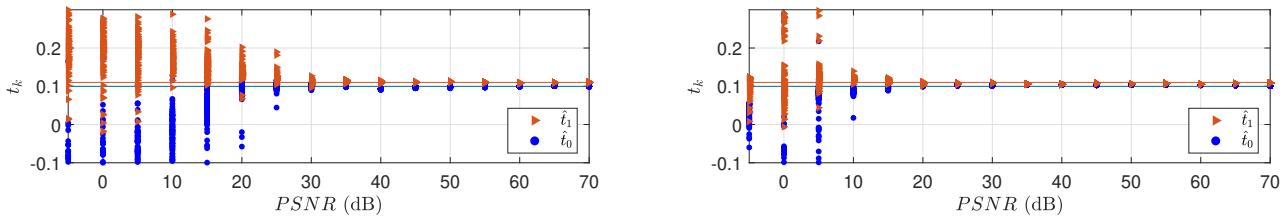
The simulation focuses on a basic setting of reconstructing a periodic stream of 2 Diracs with $t_k \in [-0.5, 0.5)$ and $a_k \in \mathbb{R}^+$. The number of samples and signal period are set to $N = 21$ and $\tau = 1$, respectively.

For a fair comparison with the classical subspace-based methods, we consider optimal settings for the subspace-based methods [12]. Thus, we choose the sampling kernel $\varphi(t)$ to be an exponential reproducing kernel of maximum order and minimum-support (eMOMS) [4] that can reproduce $P+1 = N$ exponentials with $\omega_0 = \frac{-P\pi}{P+1}$ and $\lambda = \frac{2\pi}{P+1}$. For the decoder, we opt for a high resolution of $\Delta = 1/64$, meaning that for every sampling period T, we approximate the sampling kernel by 64 linear pieces. As the sampling kernel is of support N, this results in a total of $I = N/\Delta = 1344$ linear segments.

To observe the performance of different reconstruction methods under noisy conditions, the samples $y[n]$ are corrupted with additive white Gaussian noise at different noise levels. An autoencoder is then trained for each PSNR $\in [-5, 70]$ dB with a step of 5 dB. The training set for each network consist of 10^6 training data with $t_k \in \mathcal{U}(-0.5, 0.5)$ and $a_k \in \mathcal{U}(0.5, 10)$ for $k = 0, 1$, where $\mathcal{U}(a, b)$ denotes uniform distribution between a and b.

In order to compare our results with the breakdown PSNR shown in Fig. 2, in the evaluation stage, we further assume constant amplitudes for two Diracs with $a_0 = a_1 = 2$. We then fix the first Dirac at $t_0 = 0.1$ and change $\Delta t_0 \in [10^{-0.5}, 10^{-3}]$ evenly on a logarithmic scale with a step of $10^{-0.25}$. For each pair of PSNR and Δt_0, Monte Carlo simulations with 1000 realizations have been performed.

Fig. 7 shows the respective reconstruction performance of subspace-based method, our previous work in direct inference network [14], which is equivalent to training the encoder only, and our newly proposed autoencoder. The encoder network of the autoencoder in Fig. 7c is initialized using the trained model of direct inference network not only for a fairer comparison but also to improve the stability of the network. The breakdown PSNR plotted in Fig. 2 is overlaid as the red dotted line to aid the visualization. As discussed in Section II, we can see that the performance of Prony's method with Cadzow denoising



(a) Direct inference using DNN [14].

(b) Autoencoder with encoder initialized from the trained model in Fig. 7b.

Fig. 8. Scatter plot of the retrieved locations over 100 realizations, where the horizontal lines indicate the true locations at $t_0 = 0.1$ and $t_1 = 0.1 + 10^{-2}$.

suffers from an abrupt deterioration at the red dotted line. This demonstrates the breakdown in performance due to the inherent subspace swap event in subspace-based methods. On the other hand, both DNN-based methods maintains a consistent performance across different Δt_0. For instance, when $\Delta t_0 = 10^{-2}$, the autoencoders break down at around PSNR $=$ 5 dB whereas Prony's method breaks down at PSNR $= 40$ dB. It shows that solving the original FRI reconstruction problem instead of the transformed spectral estimation problem enables us to recover FRI signals with a higher resolution under strong noise. Nonetheless, this comes with a slight compromise in the low noise region, which can be eased when our proposed method is coupled with gradient descent, as shown in Fig. 7d.

A key comparison is between our previous work in direct inference using DNN with our newly proposed autoencoder. As seen from Fig. 7b and Fig. 7c, the autoencoder outperforms the direct inference method globally. While the autoencoder provides a significant improvement over the breakdown PSNR in strong noise, it also provides a slightly more accurate reconstruction performance at low noise regimes above red line indicating the breakdown PSNR. For better visualization, we take a closer look at a representative case. Fig. 8 shows the scatter plot when the distance of the Diracs are $\Delta t_0 = 10^{-2}$, meaning that the Diracs are close together. We can observe that the direct inference method breaks down at around PSNR $= 20$ dB, while the newly proposed autoencoder only breaks down at PSNR $= 5$ dB. It shows that the addition of fixed decoders improves the performance of the learned encoder networks in reconstructing the locations of the Diracs.

V. Conclusion

This paper addresses the breakdown of performance in reconstruction of FRI signals caused by the so-called subspace swap event in traditional subspace-based methods under noisy conditions. We proposed a reconstruction method using autoencoders with fixed decoders to learn from training data pairs. Simulation results show that our proposed autoencoder can reconstruct FRI signals at a low PSNR region where the existing FRI methods would break down, while also present a significant general improvement over our previous work that performs direct inference of FRI parameters using DNN. The slight performance compromise in high PSNR region can also be alleviated when our proposed approach is coupled with gradient descent.

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, Jun. 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 I, pp. 1741–1757, 2007.

[3] T. Blu, P. L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, "Sparse sampling of signal innovations: Theory, algorithms, and performance bounds," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 31–40, 2008.

[4] 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.

[5] R. Tur, Y. C. Eldar, and Z. Friedman, "Innovation rate sampling of pulse streams with application to ultrasound imaging," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1827–1842, 2011.

[6] Y. C. Eldar, *Sampling theory: Beyond bandlimited systems*. Cambridge University Press, 2014.

[7] R. Prony, "Essai expérimental et analytique sur les lois de la dilatabilité des fluides élastiques, et sur celles de la force expansive de la vapeur de l'eau et de la vapeur de l'alkool, á différentes temperatures," *J. de l'Ecole Polytechnique*, vol. 1, pp. 24–76, 1795.

[8] 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.

[9] 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, May 1990.

[10] H. Cramér, *Mathematical methods of statistics*. Princeton university press, 1946.

[11] C. R. Rao, "Information and the Accuracy Attainable in the Estimation of Statistical Parameters," *Bulletin of Calcutta Mathematical Society*, vol. 37, pp. 81–89, 1945.

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

[13] J. K. Thomas, L. L. Scharf, and D. W. Tufts, "The Probability of a Subspace Swap in the SVD," *IEEE Transactions on Signal Processing*, vol. 43, no. 3, pp. 730–736, 1995.

[14] V. C. H. Leung, J.-J. Huang, and P. L. Dragotti, "Reconstruction of FRI Signals Using Deep Neural Network Approaches," in *2020 International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2020)*, 2020, pp. 5430–5434.

[15] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989.

[16] M. Leshno, V. Y. Lin, A. Pinkus, and S. Schocken, "Multilayer feedforward networks with a nonpolynomial activation function can approximate any function," *Neural Networks*, vol. 6, no. 6, pp. 861–867, 1993.

[17] U. Shaham, A. Cloninger, and R. R. Coifman, "Provable approximation properties for deep neural networks," *Applied and Computational Harmonic Analysis*, vol. 44, no. 3, pp. 759–773, Sep. 2018.

[18] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *CoRR*, vol. abs/1412.6, 2014.