LEARNING TO SAMPLE FOR SPARSE SIGNALS

Satish Mulleti, Haiyang Zhang, and Yonina C. Eldar

Faculty of Math and Computer Science, Weizmann Institute of Science, Israel
Emails: mulleti.satish@gmail.com, haiyang.zhang@weizmann.ac.il, yonina.eldar@weizmann.ac.il

ABSTRACT

Finite-rate-of-innovation (FRI) signals are ubiquitous in radar, ultrasound, and time of flight imaging applications. In this paper, we propose a model-based deep learning approach to jointly design the subsampling and reconstruction of FRI signals. Specifically, our framework is a combination of a greedy subsampling algorithm and a learning-based sparse recovery method. Unlike existing learning-based techniques, the proposed algorithm can flexibly handle changes in the sampling rate and does not suffer from differentiability issues during training. Moreover, exact knowledge of the FRI pulse is not required. Numerical results show that the proposed joint design leads to lower reconstruction error for FRI signals compared with existing benchmark methods for a given number of samples. The method can easily adapt to other sparse recovery problems.

***Index Terms*—** Finite rate of innovation signal, sub-Nyquist sampling, greedy algorithm, learn to sample.

1. INTRODUCTION

Finite rate of innovation (FRI) signals are ubiquitous in applications such as radar signal processing, ultrasound imaging, and time of flight imaging [1, 2]. A typical FRI signal consists of a stream of known pulses, where the time-delays and amplitudes of these pulses are unknown parameters. To estimate these parameters, e.g., assume an FRI signal consisting of a stream of L pulses, the commonly used frequency domain estimation methods only require $2L$ Fourier samples in the absence of noise [3]. In the noisy scenario, however, more than $2L$ Fourier samples are necessary to achieve desired reconstruction accuracy.

In the framework of frequency domain based methods, the FRI signal is first filtered by a sampling kernel which determines the choice of the $K \geq 2L$ Fourier samples by design, and subsequently sampled at a rate proportional to K that is far below the Nyquist rate of the FRI signal. Then, the amplitude and time-delay parameters can be estimated from the corresponding Fourier-domain samples either by using high-resolution spectral estimation methods (HRSE) [4] or by applying compressive sensing (CS) techniques [5], such as iterative shrinkage/thresholding algorithms (ISTA) [6], or its fast-counterpart, fast-ISTA (FISTA) [7, Ch. 2].

For a fixed number of samples, the reconstruction accuracy in the presence of noise depends on the choice of Fourier samples and reconstruction algorithm. Therefore, it is desirable to optimize the subsampling of Fourier samples and parameters of the reconstruction algorithm to improve reconstruction accuracy. While prior knowledge of the FRI pulse shape, the sparsity structure, and noise levels are required for optimizing the selection of the Fourier samples and reconstruction method, they may not always be available in practice. For example, in radar imaging, the FRI pulse denotes the transmit signal that may be distorted during transmission and not known at the receiver. Under such circumstances, data-driven techniques have emerged as promising new approaches to learn hidden information (e.g., pulse shape, sparsity structure, and noise levels) by fine-tuning their parameters over examples, facilitating subsequent subsampling and reconstruction of FRI signals. To date, most of the works on data-driven subsampling and reconstruction can be broadly classified into two categories: (1) data-driven recovery methods with a fixed subsampling pattern [8–13]; (2) data-driven subsampling approaches with fixed recovery [14–20]. Learning-based ISTA or LISTA [10] is one of the most popular learning-based reconstruction techniques due to its computational efficiency and interpretable structure [11].

Unlike most of the existing works that considered independent subsampling and reconstruction, in this paper we propose a model-based deep learning approach to jointly design the subsampling and reconstruction of FRI signals. Specifically, our approach is a combination of a greedy algorithm that is used for subsampling the Fourier measurements and a LISTA network that is applied for recovery. We point out here that there are a few works that also considered data-driven joint subsampling and reconstruction in the context of MRI [21, 22], and ultrasound imaging [23]. However, they require exact knowledge of the measurement matrix, which, in the FRI context, is a function of the pulse shape that may be challenging to obtain in advance. Moreover, the above-mentioned joint subsampling and reconstruction approaches suffer from several drawbacks, such as the lack of flexibility to changes in the subsampling rate, e.g., in [22, 23], the networks require retraining from scratch when the sampling rate changes.

In our proposed joint subsampling and reconstruction (JSR) architecture, we only optimize over recovery parameters while the subsampling operation is aided by the greedy algorithm. Specifically, the greedy algorithm sequentially selects a new sample from the remaining ones such that it results in the lowest reconstruction error. This is in contrast to existing joint approaches where the networks are optimized over two sets of parameters, one for sampling and the other for recovery. The reconstruction error is computed from the ground truth and estimation of the sparse vector by a trained LISTA network. Thanks to the sequential greedy nature of the proposed algorithm, it can start from the current sampling pattern and does not require retraining when the sampling rate changes. Once the Fourier samples are selected, they are used to design a sampling kernel for FRI signals through a sum-of-sincs (SoS) sampling kernel where the location of sincs in the frequency-domain is the same as the locations of the selected Fourier samples [3, 24].

We compare the proposed algorithm with existing methods in which sampling and recovery are independent, and the knowledge of the pulse shape is required. Experimental results provided in Section

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme; Israel Science Foundation under Grant 0100101; and the QuantERA grant C'MON-QSENS!.

4 show that our approach results in lower error. In addition, we show that our joint learning-based approach can estimate the FRI signals up to -20 dB normalized mean-squared error (NMSE) from less than $2L$ measurements. In other words, one can consider sampling below the sub-Nyquist rate of FRI signals by using data-driven joint designs.

The rest of this paper is organized as follows: Section 2 formulates the problem of jointly designing the sampling kernel and reconstruction for FRI signals. Section 3 presents the proposed joint sampling and reconstruction algorithm, and Section 4 provides numerical results. Finally, Section 5 concludes the paper.

2. SIGNAL MODEL AND PROBLEM FORMULATION

Consider an FRI signal

$$f(t) = \sum_{\ell=1}^{L} a_\ell h(t - t_\ell), \tag{1}$$

where the pulse $h(t)$ is compactly supported and known. In this work, we assume that the time-delays t_ℓs lie on a grid of size Δ, that is, $t_\ell = n_\ell \Delta$ where $n_\ell \in \mathcal{N} = \{1, 2, \cdots, N\}$ for some integer $N \gg L$. FRI signal sampling and reconstruction deals with estimating the parameters $\{a_\ell, t_\ell\}_{\ell=1}^{L}$ from the fewest possible samples measured at a sub-Nyquist rate.

The recovery of FRI signals is typically carried out in the frequency domain. Consider uniform Fourier-domain samples of $f(t)$ as

$$F(k\omega_0) = H(k\omega_0) \sum_{\ell=1}^{L} a_\ell e^{\mathrm{j}k\omega_0 t_\ell}, \quad k \in \mathcal{N}, \tag{2}$$

where ω_0 is the sampling interval. By substituting the grid assumption in (2) together with $\omega_0 = \frac{2\pi}{t_{\max}}$, we have

$$F(k\omega_0) = H(k\omega_0) \sum_{n=1}^{N} x_n e^{\mathrm{j}2\pi kn/N}, \quad k \in \mathcal{N}, \tag{3}$$

where $\{x_{n_\ell} = a_\ell\}_{\ell=1}^{L}$ and $x_n = 0$ for $n \notin \{n_\ell\}_{\ell=1}^{L}$. Within the assumed settings, the Fourier samples in (3) can be written in matrix form as

$$\mathbf{f} = \mathrm{diag}(\mathbf{h})\mathbf{A}\mathbf{x}, \tag{4}$$

where $\mathbf{f} = [F(\omega_0), F(2\omega_0), \cdots, F(N\omega_0)]^{\mathrm{T}}$, $\mathbf{A}$ is an N-th order discrete Fourier transform (DFT) matrix, and $\mathbf{h} = [H(\omega_0), H(2\omega_0), \cdots, H(N\omega_0)]^{\mathrm{T}}$. The vector $\mathbf{x} = [x_1, \cdots, x_N]^{\mathrm{T}}$ is L-sparse where its support carries the information of time-delays and the values of its non-zero entries are equal to the amplitudes of the FRI signal. The sparse vector $\mathbf{x}$ can be estimated by applying CS methods on a minimum of $2L$ samples of $\mathbf{f}$ in the absence of noise and more than $2L$ in the presence of noise. Let $\mathcal{K} \subset \mathcal{N}$ denote indices of the desired Fourier samples where $|\mathcal{K}| \geq 2L$. Then there are $\binom{N}{|\mathcal{K}|}$ possible choices of Fourier samples. Our problem is to select the one that leads to lowest reconstruction accuracy.

Determination of the desired Fourier samples is achieved by carefully designing a sampling kernel and then sampling the output at a prescribed rate. Here we use a compactly supported SoS sampling kernel with the following impulse response

$$g_{\mathbf{c}}(t) = \sum_{k \in \mathcal{N}} c_k e^{\mathrm{j}k\omega_0 t}, \quad t \in [0, T_g], \tag{5}$$

where T_g is greater than the support of the FRI signal. The coefficients $\{c_k\}_{k\in\mathcal{N}}$ are parameters that can be designed to select specific Fourier coefficients as explained next. Specifically, we set $c_k = 1$ for $k \in \mathcal{K}$ and the remaining ones to zero.

It can be shown that for an appropriate choice of T_g the samples of the filtered signal $y(t) = f(t) * g_c(t)$ are given as (see [3, 24] for details)

$$y(nT_s) = \sum_{k\in\mathcal{N}} c_k F(k\omega_0) e^{\mathrm{j}k\omega_0 nT_s} = \sum_{k\in\mathcal{K}} F(k\omega_0) e^{\mathrm{j}k\omega_0 nT_s}, \tag{6}$$

for $n \in \mathcal{N}$. The Fourier samples $\{F(k\omega_0)\}_{k\in\mathcal{K}}$ can be uniquely determined from the time domain samples provided that $T_s = \frac{t_{\max}}{|\mathcal{K}|+\epsilon}$ where $1 > \epsilon > 0$.

The Fourier measurements computed through SoS filtering can be written compactly as

$$\bar{\mathbf{f}} = \mathrm{diag}(\mathbf{c})\mathbf{f} = \mathrm{diag}(\mathbf{c})\mathrm{diag}(\mathbf{h})\mathbf{A}\mathbf{x}, \tag{7}$$

where $\mathbf{c}$ consists of the coefficients of the SoS kernel. Note that $\mathbf{c} \in \{0, 1\}^N$ and the support of $\mathbf{c}$ is $\mathcal{K}$. From $\bar{\mathbf{f}} = \mathrm{diag}(\mathbf{c})\mathrm{diag}(\mathbf{h})\mathbf{A}\mathbf{x}$, the sparse signal $\mathbf{x}$ is estimated as $\hat{\mathbf{x}} = r_{\boldsymbol{\theta}}(\mathrm{diag}(\mathbf{c})\mathrm{diag}(\mathbf{h})\mathbf{A}\mathbf{x})$ where $r_{\boldsymbol{\theta}}$ is a parametric reconstruction method with parameter $\boldsymbol{\theta}$. The estimate $\hat{\mathbf{x}}$ is a function of $\boldsymbol{\theta}$ and $\mathbf{c}$ and it is desirable to choose them to minimize the error in estimation of $\mathbf{x}$.

For a given application, let a set of representative examples be given as

$$\mathcal{D} = \{\mathbf{f}_q = \mathrm{diag}(\mathbf{h})\mathbf{A}\mathbf{x}_q, \mathbf{x}_q\}_{q=1}^{Q} \tag{8}$$

for fixed $\mathbf{h}$. Our goal is to choose $\mathbf{c}$ and $\boldsymbol{\theta}$ to minimize the reconstruction error for sparse vectors from $\mathcal{D}$. Specifically, we consider the following optimization problem

$$\min_{\substack{\mathbf{c}\in\{0,1\}^N \\ \boldsymbol{\theta}\in\mathbb{C}^M}} \frac{1}{Q}\sum_{q=1}^{Q} \mathrm{C}\left(\mathbf{x}_q, r_{\boldsymbol{\theta}}\left(\mathrm{diag}(\mathbf{c})\mathbf{f}_q\right)\right) \quad \text{s. t.} \quad \|\mathbf{c}\|_0 = K, \tag{9}$$

where $K = |\mathcal{K}|$ is the number of samples to be selected for reconstruction and the cost function $\mathrm{C} : \mathbb{C}^N \times \mathbb{C}^N \to [0, \infty)$ is a measure of reconstruction error. From the solution of (9), a sampling kernel and recovery technique are designed that can be used to sample and reconstruct FRI signals similar to the representative examples.

3. JSR APPROACH

Our joint solution is based on a combination of a greedy algorithm and LISTA-based reconstruction. For ease of discussion, we first discuss the standard greedy method and describe the LISTA network. Then, we present the proposed joint algorithm.

3.1. Greedy Algorithm for Sampling

Let us reconsider the optimization problem (9) where the objective is to select only an optimal sampling pattern $\mathbf{c}$ for a fixed recovery. Specifically, consider the cost function for a fixed recovery as

$$\mathrm{S}(\mathcal{K}) = \frac{1}{Q}\sum_{q=1}^{Q} \mathrm{C}\left(\mathbf{x}_q, r_{\boldsymbol{\theta}}\left(\mathrm{diag}(\mathbf{c})\mathbf{f}_q\right)\right). \tag{10}$$

Algorithm 1 Greedy Algorithm

Initialize: $\mathcal{K} = \emptyset$
for $k = 1$ to K **do**
[S1] $i^* = \min_{i \in \mathcal{N} \setminus \mathcal{K}} \quad \mathcal{S}(\mathcal{K} \cup \{i\})$
[S2] $\mathcal{K} = \mathcal{K} \cup \{i^*\}$
end for

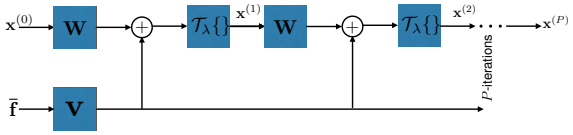

Fig. 1. Generalized network architecture of LISTA; The parameters $\{\mathbf{W}, \mathbf{V}, \lambda\}$ are optimized for each iteration of the greedy algorithm.

Then the resulting sampling problem,

$$\min_{\mathcal{K} \subset \mathcal{N}} \quad \mathrm{S}(\mathcal{K}) \quad \text{s. t.} \quad |\mathcal{K}| = K, \tag{11}$$

can be solved by a greedy approach in K-steps provided that the cost is submodular and monotone [25]. Typically, a greedy algorithm starts with zero samples. At each iteration, a new sample is added to the selected sample set to minimize the cost function. The steps of the greedy approach are summarized in Algorithm 1.

3.2. LISTA-Based Reconstruction

We use LISTA as our recovery strategy which is discussed next. To estimate a sparse $\mathbf{x}$ from its compressed measurements $\bar{\mathbf{f}} = \mathbf{B}\mathbf{x}$, where $\mathbf{B} = \mathrm{diag}(\mathbf{c})\mathrm{diag}(\mathbf{h})\mathbf{A}$ (cf. (7)), an l_1-relaxed optimization problem is considered

$$\min_{\mathbf{x} \in \mathbb{C}^N} \quad \frac{1}{2}\|\bar{\mathbf{f}} - \mathbf{B}\mathbf{x}\|_2^2 + \bar{\lambda}\|\mathbf{x}\|_1, \tag{12}$$

where $\bar{\lambda} > 0$ is a sparsity enforcing regularization parameter. One of the ways to solve (12) is to apply ISTA which starts from an initial solution $\mathbf{x}^{(0)}$ and iteratively updates it by using the following step:

$$\mathbf{x}^{(i+1)} = \mathcal{T}_{\frac{\bar{\lambda}}{\mu}} \left\{ \left(\mathbf{I}_N - \frac{1}{\mu}\mathbf{B}^{\mathrm{H}}\mathbf{B} \right) \mathbf{x}^{(i)} + \frac{1}{\mu}\mathbf{B}^{\mathrm{H}}\bar{\mathbf{f}} \right\}, \tag{13}$$

where μ is a constant parameter that controls the step size of each iteration and $\mathbf{I}_N$ denotes an $N \times N$ identity matrix. In (13), $\mathcal{T}_\alpha\{\cdot\}$ is an elementwise soft-thresholding operator defined as $\mathcal{T}_\alpha\{x\} = \mathrm{sign}(x)\ \max\{|x|, 0\}$.

By replacing $\lambda = \frac{\bar{\lambda}}{\mu}$, $\mathbf{W} = \left(\mathbf{I}_N - \frac{1}{\mu}\mathbf{B}^{\mathrm{H}}\mathbf{B}\right)$, and $\mathbf{V} = \frac{1}{\mu}\mathbf{B}^{\mathrm{H}}$, in (13), we have

$$\mathbf{x}^{i+1} = \mathcal{T}_\lambda \left\{ \mathbf{W}\mathbf{x}^i + \mathbf{V}\bar{\mathbf{f}} \right\}. \tag{14}$$

Note that $\boldsymbol{\theta} = \{\lambda, \mathbf{W}, \mathbf{V}\}$ are reconstruction parameters. While $\boldsymbol{\theta}$ is fixed in ISTA, it is learned from examples in LISTA using unrolled iterations as shown in Fig. 1. LISTA maps each iteration into a layer of the network. By stacking a finite number of layers (for example P layers in Fig 1), a network is constructed. Performing end to end training, $\boldsymbol{\theta}$ can be learned from a set of examples to minimize $\sum_{q=1}^{Q} \|\mathbf{x}_q - \hat{\mathbf{x}}_q\|_2^2$, where $\hat{\mathbf{x}}_q$ is the output of the P-th layer.

Algorithm 2 Joint Subsampling and Recovery Algorithm

Inputs: Data $\mathcal{D}$ and full sample indices $\mathcal{N}$
Initialize: $\mathcal{K}^{(0)} = \emptyset$
for $k = 1$ to K **do**
[S1] **for all** $i \in \mathcal{N} \setminus \mathcal{K}^{(k-1)}$ **do**
($\mathbf{a}$) For a binary-valued vector $\mathbf{c}_i \in \{0,1\}^{|\mathcal{N}|}$, set $\mathrm{supp}\{\mathbf{c}_i\} = \mathcal{K}^{(k-1)} \cup \{i\}$
($\mathbf{b}$) $\boldsymbol{\theta}_i^{(k)} = \arg\min_{\boldsymbol{\theta}} \frac{1}{Q}\sum_{q=1}^{Q} \|\mathbf{x}_q - r_{\boldsymbol{\theta}}\left(\mathrm{diag}(\mathbf{c}_i)\mathbf{f}_q\right)\|_2^2$ where $r_{\boldsymbol{\theta}}$ is a LISTA-based reconstruction
($\mathbf{c}$) $\mathbf{x}_{q,i} = r_{\boldsymbol{\theta}_i^{(k)}}\left(\mathrm{diag}(\mathbf{c}_i)\mathbf{f}_q\right)$ for $q = 1, \cdots, Q$
end for
[S2] $i_*^{(k)} = \arg\min_{i \in \mathcal{N} \setminus \mathcal{K}^{(k-1)}} \frac{1}{Q}\sum_{q=1}^{Q} \|\mathbf{x}_q - \mathbf{x}_{q,\mathcal{K}\setminus\{i\}}\|_2^2$
[S3] $\mathcal{K}^{(k)} = \mathcal{K}^{(k-1)} \cup \{i_*^{(k)}\}$
end for
Output: Optimal sampling set $\mathcal{K}^K \subseteq \mathcal{N}$ with $|\mathcal{K}^K| = K$ and corresponding reconstruction parameters $\boldsymbol{\theta}_{i_*^{(k)}}$

3.3. Proposed JSR Algorithm

Our proposed JSR algorithm is based on the combination of the greedy algorithm and LISTA-based reconstruction. Details of the proposed joint sub-sampling and reconstruction approaches are described in Algorithm 2. As in Algorithm 1, we start with zero samples and sequentially add a new sample at each iteration.

Comparing Algorithm 2 with Algorithm 1, we note that most of the steps remain the same, except the step [S1] where we optimize the reconstruction parameters for each sampling pattern in Algorithm 2 compared to a fixed reconstruction method in Algorithm 1. Since the sampling mechanism is non-parametric the proposed JSR algorithm is generic and can be optimized for any reconstruction method. The optimization problem in step [S1](b) is solved by a deep-learning approach. Specifically, each reconstruction step is replaced by a LISTA network, and the parameters $\boldsymbol{\theta}$ are learned through training. Note that since the reconstruction parameters are optimized for each iteration of the greedy algorithm, there is no backpropagation step from one iteration to another of [S1]. Consequently, there is no derivative step over the sampling patterns and the differentiability issue does not arise.

Suppose the network is trained for K samples. Then at the end of the K-th iteration of the greedy algorithm, we have a pair of optimal sampling patterns and corresponding reconstruction parameters $\{\mathcal{K}^{(k)}, \boldsymbol{\theta}_{i_*}^{(k)}\}$ for $k = 1, \cdots, K$. Hence, if the goal is to choose fewer samples than K, we already have a solution and no training is required. Whereas if it is required to increase the samples beyond K, then one can start from the selected K samples and follow the steps of Algorithm 2 to add additional samples without starting from scratch.

3.4. JSR Training

In Algorithm 2, only the sampling location set is passed between different sampling layers. Hence, we can use a simple sequential training approach to train each sampling layer (one iteration of the greedy algorithm). In particular, for each sampling layer, we trained a ded-

icated LISTA network for each candidate sampling pattern. Thus, we can train all the LISTA networks within the same sampling layer in a highly efficient parallel way. The initial parameters for each LISTA network within the same sampling layer are randomly generated with the same seed, which ensures that the best sampling location can be fairly selected. The loss function corresponding to the ith sampling pattern in the k-th sampling layer is given by

$$\boldsymbol{\theta}_i^{(k)} = \arg\min_{\boldsymbol{\theta}} \frac{1}{Q} \sum_{q=1}^{Q} \| \mathbf{x}_q - r_{\boldsymbol{\theta}} \left(\text{diag}(\mathbf{c}_i) \mathbf{f}_q \right) \|_2^2, \qquad (15)$$

which is used to adjust the trainable variables in LISTA. The Adam solver is used to stochastically optimize the network parameters, where the learning rate is set to be 0.001. For the LISTA module, we set the number of unfolding layers to 10.

4. NUMERICAL RESULTS

In this section, we provide numerical results to verify the performance of our proposed joint-sampling recovery (**JSR**) scheme presented in Algorithm 2. To this end, we compared the following methods: **Rand+FISTA:** The sub-sampling pattern is chosen randomly, and reconstruction is achieved by applying FISTA; **G-CRLB+FISTA:** The sub-sampling pattern is determined by using the greedy algorithm with a CRLB-based cost function [18] and reconstruction is achieved by applying FISTA. A closed form expression for CRLB-based time-delays and amplitudes estimation can be found in [26]. We note that except for our proposed **JSR** algorithm, the remaining two algorithms mentioned above require knowledge of the pulse shape $\mathbf{h}$.

In the simulations, performance is evaluated in terms of normalized MSE (NMSE) in the estimation of time-delays and amplitudes of the FRI signals. Since both these quantities are encoded in the support and amplitudes of the sparse vector $\mathbf{x}$, respectively (cf. (3)), we compute the NMSE for the test data $\{\mathbf{x}_i\}_{i=1}^{Q_{\text{Test}}}$ as follows:

$$\text{NMSE} = \frac{\sum_{i=1}^{Q_{\text{Test}}} \|\mathbf{x}_i - \hat{\mathbf{x}}_i\|_2^2}{\sum_{i=1}^{Q_{\text{Test}}} \|\mathbf{x}_i\|_2^2}, \qquad (16)$$

where $\hat{\mathbf{x}}_i$ is an estimate of $\mathbf{x}_i$. We consider the problem of sampling and reconstructing FRI signals with $L = 5$ and $t_{\max} = 1$. To generate the training and test examples as in (8), we set $N = 30$. The support of $\mathbf{x}_q$ is generated uniformly at random over the set $\{1, 2, \cdots, N\}$ and amplitudes of its non-zero entries are i.i.d. Gaussian random variables with mean 10 and variance 3. We consider an FRI pulse $h(t)$ such that its Fourier samples, $\mathbf{h}$, are non-vanishing an non constant with significant variation. Specifically, the n-th sample of $\mathbf{h} \in \mathbb{C}^N$ is given as $0.01 + e^{-0.04(n-3N/4)^2} + e^{-0.04(n-N/8)^2}$. We use $Q = 400,000$ examples for training, and $Q_{\text{test}} = 5,000$ samples for testing.

We first show results for the above-mentioned experimental setup when the sampling rate is less than the sub-Nyquist rate. In particular, we consider $K = 8$ samples for recovery, which is lower than the theoretical minimum of $2L = 10$ samples. One instance of recovery is shown in Fig. 2. We observe that, compared with **Rand+FISTA** and **G-CRLB+FISTA**, our proposed **JSR** is able to estimate the time-delays of the FRI signals or support of the sparse vector more accurately. The error is due to incorrect amplitude estimation. This leads to the conclusion that if the goal is to estimate only the time-delays as in radar imaging applications, one can go below the sub-Nyquist rates by carefully choosing the Fourier samples and reconstruction for a given data set.

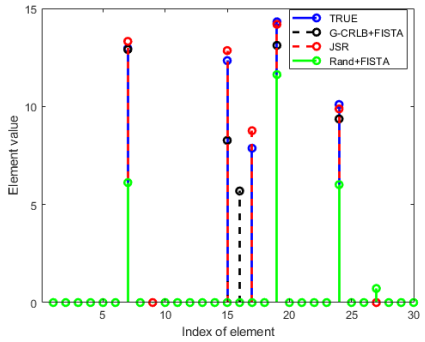

Fig. 2. Estimation of the time-delays and amplitudes from sub-sampled Fourier measurements for different sampling and reconstruction schemes for $L = 5$, $N = 30$, $K = 8$.

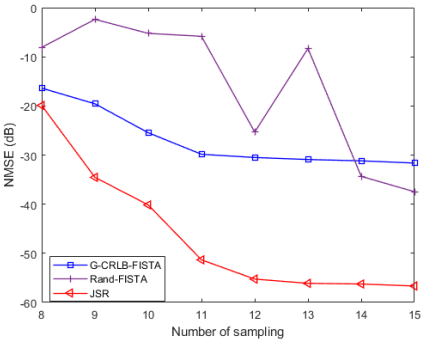

Fig. 3. A comparison of performances for different methods as a function of the number of samples. The proposed **JSR** algorithm results in lowest NMSE among different methods.

In Fig. 3 we compared the algorithms for different values of K. It is observed that **JSR** outperforms non-learning-based methods with more than 15 dB error margin, and it improves upon **G-CRLB+FISTA** by $5 - 25$ dB for different sampling rates. For $K = 8, 9$, except random sampling, all the methods result in an NMSE in the range of -15 to -20 dB. In this particular simulation, we note that while in theory, $K = 10$ is the minimum number of samples in the absence of noise, practical algorithms may not result in perfect recovery at this rate.

5. CONCLUSION

We proposed a model-based deep learning solution for joint subsampling and reconstruction of sparse signals. The proposed algorithm is a combination of the greedy subsampling algorithm and the LISTA-based sparse recovery method, which enjoys several advantages such as data-adaptive, interpretable, and flexible to changes in the sampling rate. We showed in simulations that our proposed algorithm is capable of achieving lower NMSE compared to independent design approaches. In addition, we also showed that the sampling rate can be reduced below the minimal sub-Nyquist rate known for FRI signals.

6. REFERENCES

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

[2] W. U. Bajwa, K. Gedalyahu, and Y. C. Eldar, "Identification of parametric underspread linear systems and super-resolution radar," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2548–2561, Jun. 2011.

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

[4] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*. Upper Saddle River, NJ: Prentice Hall, 1997.

[5] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012.

[6] I. Daubechies, M. Defrise, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Comm. on Pure and Applied Mathematics*, vol. 57, pp. 1413–1457, 2004.

[7] D. P. Palomar and Y. C. Eldar, *Convex Optimization in Signal Processing and Communications*. Cambridge University Press, 2010.

[8] H. Palangi, R. Ward, and L. Deng, "Distributed compressive sensing: A deep learning approach," *IEEE Trans. Signal Process.*, vol. 64, no. 17, pp. 4504–4518, 2016.

[9] S. Lohit, K. Kulkarni, R. Kerviche, P. Turaga, and A. Ashok, "Convolutional neural networks for noniterative reconstruction of compressively sensed images," *IEEE Trans. Comput. Imag.*, vol. 4, no. 3, pp. 326–340, 2018.

[10] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *proc. Int. Conf. Machine Learning (ICML)*. Madison, WI, USA: Omnipress, 2010, p. 399–406.

[11] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *IEEE Signal Processing Magazine*, vol. 38, no. 2, pp. 18–44, 2021.

[12] V. C. H. Leung, J.-J. Huang, and P. L. Dragotti, "Reconstruction of FRI signals using deep neural network approaches," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*, 2020, pp. 5430–5434.

[13] V. C. H. Leung, J.-J. Huang, Y. C. Eldar, and P. L. Dragotti, "Reconstruction of FRI signals using autoencoders with fixed decoders," in *Proc. European Signal Process. Conf. (EUSIPCO)*, 2021.

[14] B. Adcock, A. C. Hansen, and B. Roman, "The quest for optimal sampling: Computationally efficient, structure-exploiting measurements for compressed sensing," *in Compressed Sensing and its Applications. Springer*, pp. 143–167, 2015.

[15] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 58, no. 6, p. 1182–1195, 2007.

[16] S. Ravishankar and Y. Bresler, "Adaptive sampling design for compressed sensing MRI," in *Int. Conf. IEEE Eng. Medicine Biol. Soc.*, 2011, pp. 3751–3755.

[17] J. P. Haldar and D. Kim, "OEDIPUS: An experiment design framework for sparsity-constrained MRI," *IEEE Trans. Medical Imaging*, vol. 38, no. 7, pp. 1545–1558, 2019.

[18] S. Mulleti, C. Saha, H. S. Dhillon, and Y. C. Eldar, "A fast-learning sparse antenna array," in *Proc. IEEE Radar Conf. (RadarConf)*, 2020.

[19] L. Baldassarre, Y.-H. Li, J. Scarlett, B. Gözcü, I. Bogunovic, and V. Cevher, "Learning-based compressive subsampling," *IEEE J. Selected Topics in Signal Process.*, vol. 10, no. 4, pp. 809–822, 2016.

[20] B. Gözcü, R. K. Mahabadi, Y. Li, E. Ilıcak, T. Çukur, J. Scarlett, and V. Cevher, "Learning-based compressive MRI," *IEEE Trans. Medical Imag.*, vol. 37, no. 6, pp. 1394–1406, 2018.

[21] K. H. Jin, M. Unser, and K. M. Yi, "Self-supervised deep active accelerated MRI," *CoRR*, vol. abs/1901.04547, 2019. [Online]. Available: http://arxiv.org/abs/1901.04547

[22] H. K. Aggarwal and M. Jacob, "J-MoDL: Joint model-based deep learning for optimized sampling and reconstruction," *IEEE J. Selected Topics in Signal Process.*, vol. 14, no. 6, pp. 1151–1162, 2020.

[23] I. A. M. Huijben, B. S. Veeling, K. Janse, M. Mischi, and R. J. G. van Sloun, "Learning sub-sampling and signal recovery with applications in ultrasound imaging," *IEEE Trans Medical Imag.*, vol. 39, no. 12, pp. 3955–3966, 2020.

[24] S. Mulleti and C. S. Seelamantula, "Paley-Wiener characterization of kernels for finite-rate-of-innovation sampling," *IEEE Trans. Signal Process.*, vol. 65, no. 22, pp. 5860–5872, 2017.

[25] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions — I," *Math. Program.*, vol. 14, no. 1, pp. 265–294, 1978.

[26] C. S. Seelamantula and S. Mulleti, "Super-resolution reconstruction in frequency-domain optical-coherence tomography using the finite-rate-of-innovation principle," *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 5020–5029, Oct. 2014.