IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 19 May 2023, accepted 28 June 2023, date of publication 10 July 2023, date of current version 18 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3293637

RESEARCH ARTICLE

Learning to Sample: Data-Driven Sampling and Reconstruction of FRI Signals

SATISH MULLETI[1], (Member, IEEE), HAIYANG ZHANG[2], (Member, IEEE), AND YONINA C. ELDAR[3], (Fellow, IEEE)

[1]Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai 400076, India
[2]School of Communications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210049, China
[3]Faculty of Mathematics and Computer Science, The Weizmann Institute of Science, Rehovot 7610001, Israel

Corresponding author: Satish Mulleti (mulleti.satish@gmail.com)

This work was supported in part by the Israeli Council for Higher Education (CHE) via the Weizmann Data Science Research Center, in part by the European Union's Horizon 2020 Research and Innovation Program under Grant 646804-ERC-COG-BNYQ, and in part by the Israel Science Foundation under Grant 0100101.

ABSTRACT Finite-rate-of-innovation (FRI) signal model is well suited for time-of-flight imaging applications such as ultrasound, lidar, sonar, radar, and more. Due to their finite degrees of freedom, the sub-Nyquist sampling framework is used for FRI signals. In this framework, sampling is achieved by using appropriate sampling kernels. Reconstruction is performed by first computing Fourier samples of the FRI signal and then applying sparse-recovery algorithms. The choice of the Fourier samples and reconstruction method plays a crucial role in the reconstruction quality. In this paper, we consider jointly optimizing the choice of Fourier samples and reconstruction parameters. Our framework combines a greedy subsampling algorithm and a learning-based sparse recovery method. The combination has three distinct advantages. First, the network does not require knowledge of the FRI pulse shape, which is not the case with existing approaches. Second, the network does not suffer from differentiability issues during training which is common in sampling networks. Further, the proposed algorithm can flexibly handle changes in the sampling rate. Numerical results show that, for a given number of samples, the proposed joint design leads to lower reconstruction error for FRI signals than independent data-driven design methods for noisy and clean samples. We also propose a way to extend the approach for large-scale problems. Our learning-to-sample approach can be readily applied to other sampling setups, including compressed sensing problems.

INDEX TERMS FRI signal, sub-Nyquist sampling, sum-of-sincs filter, model-based deep learning, unrolling, greedy algorithm, LISTA, joint sampling and recovery, learn to sample.

I. INTRODUCTION

Sub-Nyquist approaches make use of signal structures beyond bandlimitedness [1]. A popular structure is that of finite rate of innovation (FRI), which has been extensively studied the time of flight imaging applications such as ultrasound imaging, radar imaging, and more [2], [3], [4], [5], [6]. A typical FRI signal is a linear combination of delayed copies of a known pulse, where the information of the signal lies in the amplitude and time-delays of the pulses. For an FRI signal consisting of a stream of L pulses there are $2L$ unknowns. It has been shown that a minimum of $2L$ Fourier samples of the FRI signal are sufficient to uniquely identify the unknowns in the absence of noise [6], [7], [8]. In the presence of noise, a larger number of Fourier samples are required.

The desired Fourier samples of the FRI signals can be computed by using a kernel-based sub-Nyquist sampling framework. In this framework, the FRI signal is passed through an appropriate sampling kernel prior to sampling. Then, the filtered signal is sampled at a rate proportional to the number of Fourier samples to be determined. The sampling rate is far below the effective bandwidth of the FRI signal, which results in sub-Nyquist sampling. The sampling

The associate editor coordinating the review of this manuscript and approving it for publication was Manuel Rosa-Zurera.

kernel, which is a function of the desired Fourier samples, plays a crucial role in determining Fourier measurements of the FRI signal from its sub-Nyquist samples. The amplitude and time-delays of the FRI signal are then determined from the Fourier samples. For off-grid time delays, high-resolution spectral estimation methods (HRSE) [9, Ch. 4] are applied. When the time delays are on a grid, compressive sensing (CS) methods [10] such as iterative shrinkage/thresholding algorithms (ISTA) [11], fast-ISTA (FISTA) [12, Ch. 2], [13] can be used.

In the aforementioned standard sampling and recovery framework, reconstruction accuracy depends on the parameters of the recovery algorithm and the choice of Fourier samples. Hence, for a given number of Fourier samples, the parameters and choices could be optimized to improve reconstruction accuracy. For example, the spectrum of the FRI pulse plays a crucial role in the reconstruction algorithm and choice of Fourier coefficients, especially in the presence of noise. Moreover, in many applications, the sparsity pattern of the FRI signals is further structured, such as tree sparsity or block sparsity. In this case, the number of measurements can be used to reduce below what is required by standard CS methods [14]. In general, the FRI pulse shape, the sparsity structure, and noise levels are required to optimize the Fourier sampling patterns and parameters of the reconstruction. However, prior knowledge of these quantities may not be available. For example, the pulse shape could be distorted during transmission in a radar application, and hence it is not known accurately at the receiver.

In the absence of such prior knowledge, data-driven techniques are well suited. In these approaches, a set of examples that are similar to the signal to be sampled are available. These examples implicitly carry information of the sparsity structure, FRI pulse shape, and noise levels. Even when the information is known explicitly, there does not exist an optimal data-independent technique to solve the problem of Fourier sample selection. Data-driven reconstruction and subsampling methods learn hidden information by optimizing their parameters by using examples.

The data-driven or learning approaches can be categorized as follows: (1) fixed subsampling with learning-based recovery methods; (2) learning-based subsampling with fixed reconstruction; and (3) joint subsampling and reconstruction methods. For example, deep-learning reconstruction techniques have been proposed to reduce the computational cost compared to standard CS algorithms and improve reconstruction accuracy in the presence of noise and sparsity structures [15], [16], [17]. Among these techniques, learned ISTA (LISTA) [17] is favored due to its interpretable structure and computational efficiency. In addition, due to its model-based structure, LISTA requires fewer examples to train compared to standard deep learning approaches [18]. Unlike HRSE or standard CS-based methods, LISTA does not require explicit knowledge of the FRI pulse shape. Recently, several alternative model-based networks have been proposed for sparse recovery [19], [20], [21].

There are several subsampling mechanisms considered in the literature [22], [23], [24], [25], [26], [27], [28] which result in lower reconstruction error compared to non-adaptive, random subsampling [29], [30]. Adcock et al. [22] considered a deterministic approach for subsampling Fourier measurements with an assumption that the signals are sparse in the wavelet domain. Lustig et al. [23] used a random variable density subsampling scheme for magnetic resonance imaging (MRI) where low frequencies are favored. In [25], [26], [27], and [28], greedy algorithms are used to design subsampling patterns. In [25] and [26] a Crámer-Rao lower bound (CRLB)-based cost function, which is independent of any recovery method, is used to determine the subsampling pattern. In [27] and [28], a means-squared error cost function is used, which is computed assuming a fixed reconstruction method. A direct application of these methods or their modifications to the Fourier subsampling problem for FRI signals requires knowledge of the FRI pulse shape.

Compared to independent subsampling and reconstruction optimization methods, joint design results in lower reconstruction error for a given number of measurements as shown for MRI [31], [32], [33], [34], [35], [36], [37], [38], and ultrasound [39]. These methods consider both subsampling and reconstruction as parametric functions and optimize the parameters over a given data set such that reconstruction error is minimized. These joint design methods can be applied to select Fourier samples and reconstruction parameters for FRI signals. However, these techniques have several drawbacks. First, the approaches require exact knowledge of the measurement matrix, which is a function of the pulse shape in the FRI context. Second, in the joint design works, the cost function is not differentiable with respect to the sampling pattern, so that gradient-based optimization methods are not applicable. In [32], the sampling patterns are constrained to be a continuous function of a known sampling template to address the differentiability issue. However, the selection of a sampling template is an additional challenge. In [31], [33], [37], [38], and [39], categorical random variables or probabilistic policies are used to reparameterize the discrete sampling locations to a continuous random variable. This relaxation results in additional parameters to be learned and increases the computational cost [33], [40]. In addition, it is difficult to understand the effect of relaxation on the performance of joint optimization. In [36], the authors used an alternate minimization approach where the sampling pattern is first optimized by fixing the reconstruction approach, and then the parameters of the reconstruction method are determined with the optimized sampling pattern in the first step. The sampling pattern optimization step uses a heuristic approach for subset selection. In the context of FRI sampling, this approach does not require knowledge of FRI pulse. However, the algorithm's convergence requires careful selection of the initialization

and learning rates. The flexibility of the networks with changes in the subsampling rate is another key issue in these approaches. The networks in [32] and [39] require retraining from scratch for changes in the sampling rate. Recently, several alternative learning-based joint sampling and reconstruction approaches have been proposed for images and videos [41], [42], [35], [43], [44].

In this paper, we propose a model-based, deep-learning framework for the sampling and reconstruction of FRI signals. Specifically, we design a data-driven approach by jointly optimizing the Fourier subsampling pattern and parameters of the reconstruction approach. Two main ingredients of our technique are a greedy algorithm that is used for subsampling and a LISTA network for FRI parameter estimation. At any given iteration of the greedy algorithm, a new sample is selected (or removed) from the remaining (or existing) samples such that it results in the lowest reconstruction error. A trained LISTA network is used to estimate the sparse vector, and then the reconstruction error between the estimate and the ground truth is computed. In our approach, at each iteration of the greedy step, reconstruction parameters are optimized for sampling patterns at that step. In this way, at the end of the final iteration, a choice of Fourier samples, along with a matched reconstruction method, is obtained. During reconstruction, the LISTA network is used, which does not require the knowledge of the FRI pulse shape. Similarly, the sampling pattern is optimized to reduce the reconstruction error, which is pulse-independent. Hence, the overall network does not require knowledge of the FRI pulse.

The sequential greedy selection approach used in the network is flexible to change in the sampling rate, and our network possesses the same flexibility. In particular, to optimize the network for a higher number of samples, one can start from the trained network instead of retraining. Further, the proposed network does not have differentiability issue that arises in most existing works. Due to the use of a greedy approach, our joint network does not have any parameters to optimize for sampling, and hence differentiability issue does not arise. Once the entire network is optimized, it results in optimal Fourier sampling locations and reconstruction method. The Fourier sampling locations are then used to design a sampling kernel. In this paper, a sum-of-sincs (SoS) kernel is designed by placing sinc functions at the selected sampling locations [6], [8]. Application of the proposed joint subsampling and reconstruction approach is not limited to FRI signals but can be extended to other subsampling problems such as MRI, compressive sensing, and more.

To assess of joint optimization algorithm, we compare it with the algorithms where sampling and reconstruction are independently optimized. Specifically, we compare our algorithm with greedy methods with fixed reconstruction [25], [26], [27], [28]. Compared to these methods, our joint approach results in a lower error for different noise levels. Further, we show that the joint-learning-based approach can sample below the sub-Nyquist rate of the FRI signal. Specifically, we show that an L-sparse FRI signal can be reconstructed up to -20 dB normalized mean-squared error (NMSE) from less than $2L$ Fourier samples. In addition, we show that the learning-based approach can learn sparsity structures in the signal by optimizing the sampling and reconstruction for a given set of examples.

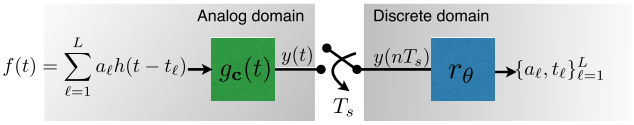


FIGURE 1. A schematic of FRI signal sampling and reconstruction: An FRI signal $f(t)$ is first passed through a filter $g_{\mathbf{c}}(t)$ where $\mathbf{c}$ is a parameter of the filter. The signal is uniformly sampled at a sub-Nyquist rate, and the signal parameters are estimated by applying a reconstruction method r_θ.

The paper is organized as follows. The proposed joint sampling and reconstruction problem for FRI signals is discussed in Section II. In Section III, we present the proposed joint-sampling and reconstruction algorithm. Numerical results are presented in Section IV followed conclusions in Section V.

II. PROBLEM FORMULATION

A. SIGNAL MODEL

Consider an FRI signal $f(t)$ that consists of a sum of L amplitude-scaled and time-shifted copies of a pulse $h(t)$:

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

In the standard FRI framework, $h(t)$ is assumed to be known, which results in a finite degrees of freedom of the FRI signal $f(t)$ in terms of the amplitudes $\{a_\ell\}_{\ell=1}^{L}$ and delays $\{t_\ell\}_{\ell=1}^{L}$. FRI signal sampling and reconstruction deal with estimating the parameters $\{a_\ell, t_\ell\}_{\ell=1}^{L}$ from the fewest possible samples measured at a sub-Nyquist rate.

In radar, ultrasound, and other time-of-flight-based applications, the pulse $h(t)$ denotes the transmit pulse. Assuming there are L point targets in the medium, the transmit pulse gets reflected from the targets and a linear combination of L reflected pulses are received. The received signal is modeled as the FRI signal in (1) where the amplitudes $\{a_\ell\}_{\ell=1}^{L}$ and $\{t_\ell\}_{\ell=1}^{L}$ denote area or size of the targets and location of the targets, respectively.

We make the following assumptions on the signal model. We assume that the number of targets, L, is known and, $t_\ell \in (0, t_{\max}]$, where $t_{\max}$ is the maximum time delay of the targets. In practice, the pulse $h(t)$ has compact support, and we assume that $h(t) = 0$ for $t \notin [0, T_h]$. The support assumption of $h(t)$ together with the assumption that $t_\ell \in (0, t_{\max}]$ results in compactly supported $f(t)$ over the interval $[0, T_h + t_{\max}]$.

A schematic of a typical FRI signal sampling and reconstruction framework is shown in Fig. 1. The sampling process consists of a sampling kernel $g_{\mathbf{c}}(t)$ followed by a sampler

operating at a rate $\frac{1}{T_s}$ samples per second. The output of the sampler is a sequence of samples $y(nT_s)$ from which the FRI parameters $\{a_\ell, t_\ell\}_{\ell=1}^L$ are computed by the reconstruction method $r_{\boldsymbol{\theta}}$. The sampling kernel and reconstruction method are parameterized by $\mathbf{c}$ and $\boldsymbol{\theta}$, respectively. Our goal is to jointly design a sampling kernel and reconstruction method such that the FRI signal parameters are determined uniquely from sub-Nyquist samples. Specifically, for a given set of FRI signals of the form (1), we seek to jointly optimize $\mathbf{c}$ and $\boldsymbol{\theta}$ to minimize the reconstruction error while operating at sub-Nyquist rates. In addition, we assume that the pulse shape $h(t)$ is not known a priori. Details of the problem formulation, together with the sampling and reconstruction methods, are discussed next.

B. FREQUENCY-DOMAIN RECONSTRUCTION

The reconstruction is typically carried out in the frequency domain as it facilitates application of either HRSE or CS-based methods. To elaborate, 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 $\omega_0 \in \mathbb{R}$ is the sampling interval and $\mathcal{N}$ is a set of integers. The sampling locations are chosen such that $\{H(k\omega_0)\}_{k\in\mathcal{N}}$ do not vanish. Since $\{H(k\omega_0)\}_{k\in\mathcal{N}}$ are computed a priori from the known pulse $h(t)$, from the Fourier samples the following sequence of samples are derived:

$$S(k\omega_0) = \frac{F(k\omega_0)}{H(k\omega_0)} = \sum_{\ell=1}^{L} a_\ell e^{\mathrm{j}k\omega_0 t_\ell}, \quad k \in \mathcal{N}. \tag{3}$$

The spectral samples $S(k\omega_0)$ are a function of the unknown FRI parameters $\{a_\ell, t_\ell\}_{\ell=1}^L$.

To estimate the FRI parameters through HRSE methods, such as annihilating filter [45], or matrix-pencil algorithms [46], the set $\mathcal{N}$ should consist of consecutive integers with cardinality $|\mathcal{N}| \geq 2L$. In other words, there should be $2L$ spectral samples to determine $2L$ unknowns. In addition, for uniquely identifying the time-delays, it is necessary that the elements in the $\{e^{\mathrm{j}\omega_0 t_\ell}\}_{\ell=1}^L$ are distinct. A straightforward way to ensure this requirement is to set $\omega_0 = \frac{2\pi}{t_{\max}}$.

Compared to HRSE methods, CS-based algorithms, such as orthogonal matching pursuit or iterative thresholding techniques, are shown to be robust in the presence of noise. This motivates us to consider CS-based algorithms for reconstruction in this paper. To apply these methods, the time-delays are assumed to be on a grid. To elaborate, let Δ be a grid size and $N = \frac{t_{\max}}{\Delta}$. Then for each t_ℓ there exists an integer $n_\ell \in \mathcal{N} = \{1, 2, \cdots, N\}$ such that $t_\ell = n_\ell\Delta$. By substituting the grid assumption in (2) together with $\omega_0 = \frac{2\pi}{t_{\max}}$, we have

$$F(k\omega_0) = H(k\omega_0)\sum_{n=1}^{N} x_n e^{\mathrm{j}2\pi kn/N}, \quad k \in \mathcal{N}, \tag{4}$$

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 (2) can be written in matrix notation as

$$\mathbf{f} = \mathrm{diag}(\mathbf{h})\mathbf{A}\mathbf{x}, \tag{5}$$

where $\mathbf{f} = [F(\omega_0), F(2\omega_0), \cdots, F(N\omega_0)]^{\mathrm{T}}$, $\mathbf{h} = [H(\omega_0), H(2\omega_0), \cdots, H(N\omega_0)]^{\mathrm{T}}$, and $\mathbf{A}$ is an $N \times N$ discrete Fourier transform (DFT) matrix. The vector $\mathbf{x} = [x_1, \cdots, x_N]^{\mathrm{T}}$ has L non-zero values. The support of $\mathbf{x}$ carries the information of time-delays, and the non-zero values are equal to the amplitudes. Theoretically, $2L$ Fourier samples are sufficient to recover $\mathbf{x}$ uniquely. However, more than $2L$ samples are required in the presence of noise.

C. KERNEL-BASED SAMPLING

For both on-grid or gridless premises, the frequency-domain-based reconstruction requires the computation of Fourier samples $\{F(k\omega_0)\}_{k\in\mathcal{N}}$. These samples can be determined by designing a suitable sampling kernel and then sampling the filtered signal at an appropriate rate. This is achieved through carefully designing the sampling kernel and the sampling rate. In the context of FRI sampling, different sampling kernels are proposed, such as SoS kernels, generalized kernels, Gaussian, ideal lowpass filters, exponential and polynomial generating kernels, and more. Among these, the SoS kernel, the generalized kernels, and the polynomial generating kernels are suitable for the present discussion. These kernels can be designed such that the filtered signal $y(t)$ is a linear combination of Fourier samples of $f(t)$ as required in the Fourier-domain recovery method. Moving forward, we consider an SoS kernel due to its simplicity in design and parametric nature.

An SoS kernel has compact support, and its impulse response is given by

$$g_{\mathbf{c}}(t) = \sum_{k\in\mathcal{N}} c_k e^{\mathrm{j}k\omega_0 t}, \quad t \in [0, T_g], \tag{6}$$

where $T_g > T_h + t_{\max}$. The coefficients $\{c_k\}_{k\in\mathcal{N}}$ act as design parameters of the filter and can be optimized to select the Fourier samples.

It can be shown that the filtered signal $y(t)$ by using the SoS filter is given as [6] and [8]

$$y(t) = \sum_{k\in\mathcal{N}} c_k F(k\omega_0) e^{\mathrm{j}k\omega_0 t}, \quad t \in [T_h + t_{\max}, T_g]. \tag{7}$$

Upon uniformly sampling $y(t)$ in the interval $[T_h + t_{\max}, T_g]$ we have the following samples

$$y(nT_s) = \sum_{k\in\mathcal{N}} c_k F(k\omega_0) e^{\mathrm{j}k\omega_0 nT_s}. \tag{8}$$

To uniquely determine scaled Fourier samples $\{F(k\omega_0)\}_{k\in\mathcal{N}}$ from the time-samples in (8), there should be at least $|\mathcal{N}|$ time-samples and the set of linear equations in (8) should be invertible. By choosing $T_g \geq T_h + t_{\max} + |\mathcal{N}|T_s$, it can be ensured that there are at least $|\mathcal{N}|$ time-samples. The linear equations are invertible provided that $T_s = \frac{t_{\max}}{|\mathcal{N}|}$. Note that

the sampling rate is proportional to the number of Fourier samples.

In practice, $N \gg L$ ensures that the on-grid assumption is close to any off-grid scenario. The condition $N \gg L$ results in a large number of time and Fourier samples beyond the minimal $2L$ samples. As the sampling rate is proportional to the number of Fourier samples, $N \gg L$ leads to oversampling.

To ensure that the desired number of Fourier samples are selected while sampling at the lowest possible rate, Fourier-domain subsampling is required. Specifically, let $\mathcal{K} \subset \mathcal{N}$ where $2L \leq |\mathcal{K}| < |\mathcal{N}|$ denotes the index of Fourier samples in $\bar{\mathbf{f}}$ that are considered for reconstruction. One of the simplest ways to consider these Fourier samples is to set $c_k = 1, k \in \mathcal{K}$ and the remaining coefficients to be zero. The non-zero values of c_k can be designed to take any value instead of $c_k = 1$. However, hardware implementation of the filter for arbitrary choice of c_ks may give rise to calibration issues. Hence for the remainder of the discussion, we consider binary valued c_ks.

With subsampling, (8) becomes

$$y(nT_s) = \sum_{k\in\mathcal{K}} c_k F(k\omega_0) e^{\mathrm{j}k\omega_0 nT_s}, \quad n \in \mathcal{N}, \tag{9}$$

where the summation is over $\mathcal{K}$ instead of $\mathcal{N}$. For reconstruction, $\{F(k\omega_0)\}_{k\in\mathcal{K}}$ needs to be determined from the time samples in (9). In the set of linear-equations in (9), the matrix relating the Fourier and time samples is a Vandermonde matrix with its (nk)-th component being $e^{\mathrm{j}k\omega_0 nT_s}$ where $n \in \mathcal{N}$ and $k \in \mathcal{K}$. The Vandermonde matrix is left-invertible provided that its seeds $\{e^{k\omega_0 T_s}\}_{k\in\mathcal{K}}$ are distinct. Since $\mathcal{K} \subset \mathcal{N}$, the choice $T_s = \frac{t_{\max}}{|\mathcal{N}|}$ ensures uniqueness of recovery. However, this results in a higher sampling rate.

To ensure that the sampling rate is proportional to the number of Fourier samples $|\mathcal{K}|$, we set the sampling interval as

$$T_s = \frac{t_{\max}}{|\mathcal{K}| + \epsilon}, \tag{10}$$

where $1 > \epsilon > 0$. The condition $\epsilon = 0$ should give a minimum desirable sampling rate. However, this may result in repetition in the set $\{e^{\mathrm{j}k\omega_0 T_s}\}_{k\in\mathcal{K}}$. For example, if $k_0, k_0 + |\mathcal{K}| \in \mathcal{K}$ for some integer k_0, then the seeds $e^{\mathrm{j}k_0\omega_0 T_s}$ and $e^{\mathrm{j}(k_0+|\mathcal{K}|)\omega_0 T_s}$ are indistinguishable for $\epsilon = 0$. The upper bound, $\epsilon < 1$, ensures that the sampling rate is close to the desired minimum rate while ensuring distinct seeds.

D. PROBLEM OF JOINT SAMPLING AND RECOVERY

The Fourier samples derived from the filtered signal are given as

$$\bar{\mathbf{f}} = \mathrm{diag}(\mathbf{c})\mathbf{f} = \mathrm{diag}(\mathbf{c})\mathrm{diag}(\mathbf{h})\mathbf{A}\mathbf{x}, \tag{11}$$

where elements of $\mathbf{c} \in \{0, 1\}^N$ are the coefficients of the SoS filter. The non-zero locations of $\mathbf{c}$ denote the selected Fourier samples. Specifically, we have that $\mathrm{supp}(\mathbf{c}) = \mathcal{K}$. Let $r_{\boldsymbol{\theta}}(\cdot)$ is reconstruction method with tunable parameters $\boldsymbol{\theta} \in \mathbb{C}^M$. Then the estimate of the sparse signal is given as $\hat{\mathbf{x}} = r_{\boldsymbol{\theta}}\,(\mathrm{diag}(\mathbf{c})\mathrm{diag}(\mathbf{h})\mathbf{A}\mathbf{x})$. Our objective is to optimize $\mathbf{c}$ and $\boldsymbol{\theta}$ to minimize error in estimation of $\mathbf{x}$. To this end, we follow a data-driven approach and optimize the parameters for a given set of examples.

For a given application, consider a data set of Q representative examples,

$$\mathcal{D} = \{\mathbf{f}_q = \mathrm{diag}(\mathbf{h})\mathbf{A}\mathbf{x}_q + \boldsymbol{\eta}_q, \mathbf{x}_q\}_{q=1}^{Q}, \tag{12}$$

for fixed $\mathbf{h}$. The vector $\boldsymbol{\eta}_q$ denotes measurement noise in the q-th example. To optimize $\mathbf{c}$ and $\boldsymbol{\theta}$ over these examples, we consider the following optimization problem

$$\min_{\substack{\mathbf{c}\in\{0,1\}^N \\ \boldsymbol{\theta}\in\mathbb{C}^M}} \quad \frac{1}{Q}\sum_{q=1}^{Q}\{\mathrm{C}(\mathbf{x}_q, r_{\boldsymbol{\theta}}\left(\mathrm{diag}(\mathbf{c})\mathbf{f}_q\right)\} \quad \text{s. t.} \quad \|\mathbf{c}\|_0 = K, \tag{13}$$

where K is the sampling budget and the cost function $\mathrm{C} : \mathbb{C}^N \times \mathbb{C}^N \to [0, \infty)$ is a error metric. In addition to being L-sparse, these representative examples could have structured sparsity. For example, the non-zero values could be bunched together or can occur in a specific pattern. These structures are not known a priori. Any solution to the optimization problem of the representative examples should be able to learn these structures and optimize $\mathbf{c}$ and $\boldsymbol{\theta}$ accordingly. A sampling kernel and recovery technique are designed from the solution of the optimization problem in (13). Then the sampling kernel and reconstruction method can be used to sample FRI signals whose sparsity structure and pulse shape are the same as the representative data set. Since the sampling kernel operates in an analog domain and the reconstruction is discrete, the above design principle is hybrid in nature. In the next section, we propose solutions to the problem discussed here.

III. DATA-DRIVEN JOINT SAMPLING AND RECONSTRUCTION ALGORITHM

In this section, we propose algorithms to solve the optimization problem in (13). The optimization problem concerning $\mathbf{c}$ is combinatorial. To address this issue, we follow a greedy approach. Specifically, the sampling problem is solved by a non-parametric, iterative, greedy method. For reconstruction, we apply the LISTA [17]. To solve the sampling and reconstruction problem jointly, we combine the greedy algorithm and LISTA. Before discussing the proposed joint learning algorithm, we first discuss the standard greedy algorithm and its variants suitable for our objective and the LISTA algorithm.

A. SUB-SAMPLING THROUGH GREEDY APPROACH

Consider the optimization problem (13). If we fix the recovery approach, the problem is reduced to designing an optimal sampling pattern $\mathbf{c}$. To address the combinatorial nature of the problem, convex relaxation approaches [47], [48], bit-wise mutation algorithms [49], [50], [51], and greedy

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

methods [26], [52], [53] have been suggested. Among these two methods, the greedy one is computationally efficient. In the greedy algorithm, the sampling pattern is computed in K-steps, and the performance of this method is close to optimal if the cost function is submodular [54].

In (13), for a fixed reconstruction method $r_{\boldsymbol{\theta}}$, the cost is a function of the set $\mathcal{K} = \text{supp}(\mathbf{c})$. Let

$$\text{S}(\mathcal{K}) = \frac{1}{Q} \sum_{q=1}^{Q} \text{C}(\mathbf{x}_q, r_{\boldsymbol{\theta}} \left(\text{diag}(\mathbf{c})\mathbf{f}_q\right). \tag{14}$$

With this cost, the optimization problem,

$$\min_{\mathcal{K} \subset \mathcal{N}} \quad \text{S}(\mathcal{K}) \quad \text{s. t.} \quad |\mathcal{K}| = K, \tag{15}$$

can be solved by a greedy algorithm if the cost is submodular and monotone. Mathematically, submodularity implies that $\forall \mathcal{K}_1 \subseteq \mathcal{K}_2 \subseteq \mathcal{N}$ we have that $\text{S}(\mathcal{K}_1 \cup \{i\}) \geq \text{S}(\mathcal{K}_2 \cup \{i\}), \forall i \in \mathcal{N} \setminus \mathcal{K}_2$. The cost is said to be monotonic if $\forall \mathcal{K}_1 \subseteq \mathcal{K}_2 \subseteq \mathcal{N}$ we have that $\text{S}(\mathcal{K}_1) \leq \text{S}(\mathcal{K}_2)$ [54]. For submodular and monotone cost functions, performance of the greedy approach reaches within $\left(1 - \frac{1}{e}\right)$ of the optimal solution [52].

Typically, greedy algorithms rely on a bottom-up approach where the algorithms start from zero samples and add a new sample at each iteration. The new sample selected from the remaining samples minimizes the cost at that iteration. The details are summarized in Algorithm 1.

Among several choices of submodular cost functions, cost as the log-determinant of the CRLB (log-det CRLB) is shown empirically to have a small error while estimating sparse vectors from subsampled Fourier measurements [26], [55]. In the present setup, assuming that $\{\mathbf{x}_q\}_{q=1}^{Q}$ are deterministic, the CRLB in estimation of $\mathbf{x}_q$ from $\mathbf{y}_q = \mathbf{B}\mathbf{x}_q + \boldsymbol{\eta}_q$ where $\mathbf{B} = \text{diag}(\mathbf{c})\text{diag}(\mathbf{h})\mathbf{A}$. Here CRLB or log-det CRLB is a function of the sampling pattern $\mathbf{c}$ or $\mathcal{K}$ as well as the pulse $\mathbf{h}$. Let $\text{CRLB}_q(\mathbf{x}_q, \mathbf{h}, \mathbf{c})$ denote the CRLB matrix in the estimation of $\mathbf{x}_q$. Specifically, $\text{CRLB}_q(\mathbf{x}_q, \mathbf{h}, \mathbf{c})$ denotes lower bound on the error covariance matrix in the estimation of $\mathbf{x}_q$ from $\mathbf{y}_q$, that is,

$$\text{E}_{\hat{\mathbf{x}}_q}\left((\mathbf{x}_q - \hat{\mathbf{x}}_q)(\mathbf{x}_q - \hat{\mathbf{x}}_q)^{\text{T}}\right) \geq \text{CRLB}(\mathbf{x}_q, \mathbf{h}, \mathcal{K}), \tag{16}$$

where $\hat{x}_q$ is an unbiased estimate of $\mathbf{x}_q$ and $\text{E}_{\hat{\mathbf{x}}_q}$ is expectation over $\hat{x}_q$. Then the cost function in (14) is given as

$$\mathcal{S}(\mathcal{K}) = \frac{1}{Q} \sum_{q=1}^{Q} \text{log-det}\, \text{CRLB}_q(\mathbf{x}_q, \mathbf{h}, \mathbf{c}). \tag{17}$$

A similar CRLB-based cost function has been considered in [25] for subsampling using a greedy technique.

While the CRLB-based greedy algorithms work, the selected sampling pattern is independent of the reconstruction $r_{\boldsymbol{\theta}}$. In [27] and [28], it is shown that the reconstruction error can be reduced by taking into account the reconstruction process while subsampling. For example, the CRLB cost can be replaced by mean-squared error or peak signal-to-noise ratio, which is computed after estimating the sparse vector by a fixed reconstruction $r_{\boldsymbol{\theta}}$. Details of this modified greedy algorithm that depends on a reconstruction method are summarized in Algorithm 2. In Algorithm 2, the sample locations selected at the end of the k-th iteration are denoted as $\mathcal{K}^{(k)}$. At the k-th iteration, to evaluate [S1], we need to apply a reconstruction algorithm $(N - k + 1)$-times. This results in an overall of $KN - \frac{K(K-1)}{2}$ recalls of the reconstruction algorithm.

Algorithm 2 Modified Greedy Algorithm

Inputs: Data $\mathcal{D}$, full-sample indices $\mathcal{N}$, and a reconstruction method $r_{\boldsymbol{\theta}}$
Output: Sampling set $\mathcal{K}$
Initialize: $\mathcal{K} = \emptyset$
for $k = 1$ to K **do**
[S1] For $i \in \mathcal{N} \setminus \mathcal{K}$, estimate $\mathbf{x}_{q,\mathcal{K}\cup\{i\}} = r_{\boldsymbol{\theta}}\left(\text{diag}(\mathbf{c})\mathbf{f}_q\right)$ where $c_k = 1$ for $k \in \mathcal{K} \cup \{i\}$ and the rest of the coefficients are zero
[S2] $i^* = \min_{i \in \mathcal{N} \setminus \mathcal{K}} \quad \frac{1}{Q} \sum_{q=1}^{Q} \text{C}\left(\mathbf{x}_q, \mathbf{x}_{q,\mathcal{K}\cup\{i\}}\right)$
[S3] $\mathcal{K} = \mathcal{K} \cup \{i^*\}$ and $\mathcal{K}^{(k)} = \mathcal{K}$
end for

In Algorithm 2, one starts with zero samples and sequentially adds newer samples. As the algorithm estimates the sparse signal with fewer measurements/samples, in the first few iterations, compared with the number of unknowns, the estimation error could be very high. Especially in the presence of noise, the method may not choose the right samples due to this large error and may be less efficient. An alternative approach is to start with full samples and sequentially remove one sample at a time until the desired number of samples is retained. This approach is also known as the backward greedy approach [56], whereas the standard approach (Algorithm 1 or 2) is referred to as the forward greedy approach. The sample removed at each iteration is the one for which a change in the cost function is the smallest, as described in Algorithm 3.

This algorithm requires $\frac{(N-K)(N+K+1)}{2}$ recalls of the reconstruction method. Algorithm 2 requires fewer computations compared to that of Algorithm 3 for $K < \frac{N}{2}$ and vice versa. The sampling pattern and hence the reconstruction accuracy of the two algorithms depend on the data used and noise levels. In addition, the cost functions are not necessarily sub-modular, and optimality of the algorithms is not guaranteed.

Algorithm 3 Modified Greedy Algorithm - 2

Inputs: Data $\mathcal{D}$, full sample indices $\mathcal{N}$, and a reconstruction method $r_{\boldsymbol{\theta}}$
Output: Sampling set $\mathcal{K} \subseteq \mathcal{N}$ with $|\mathcal{K}| = K$
Initialize: $\mathcal{K} = \mathcal{N}$
for $k = 1$ to $N - K$ **do**
[S1] For $i \in \mathcal{K}$, estimate $\mathbf{x}_{q,\mathcal{K}\backslash\{i\}} = r_{\boldsymbol{\theta}}\left(\text{diag}(\mathbf{c})\mathbf{f}_q\right)$ where $c_k = 1$ for $k \in \mathcal{K}\backslash\{i\}$ and the rest of the coefficients are zero

$$\text{[S2]}\ i^* = \min_{i\in\mathcal{K}} \frac{1}{Q}\sum_{q=1}^{Q} \text{C}\left(\mathbf{x}_q, \mathbf{x}_{q,\mathcal{K}\backslash\{i\}}\right)$$

[S3] $\mathcal{K} = \mathcal{K}\backslash\{i^*\}$ and $\mathcal{K}^{(k)} = \mathcal{K}$
end for

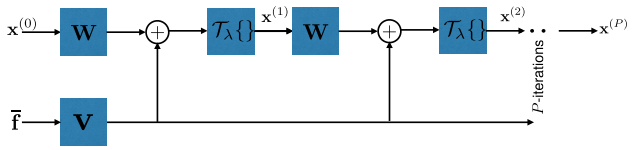


FIGURE 2. Flow diagram of ISTA and network architecture for LISTA: While the set of parameters $\theta = \{\lambda, \mathbf{W}, \mathbf{V}\}$ is fixed in ISTA whereas in LISTA θ is learned for a given set examples.

B. LISTA-BASED RECONSTRUCTION

In our approach, LISTA is applied for recovery, as discussed next. We consider the following l_1-relaxed optimization problem to estimate $\mathbf{x}$ from $\bar{\mathbf{f}} = \mathbf{B}\mathbf{x}$, where $\mathbf{B} = \text{diag}(\mathbf{c})\text{diag}(\mathbf{h})\mathbf{A}$

$$\min_{\mathbf{x}\in\mathbb{C}^N} \frac{1}{2}\|\bar{\mathbf{f}} - \mathbf{B}\mathbf{x}\|_2^2 + \bar{\lambda}\|\mathbf{x}\|_1, \tag{18}$$

where $\bar{\lambda} > 0$ is a regularization parameter. One can apply ISTA to solve (18) where starting from an initialization $\mathbf{x}^{(0)}$, a sparse solution is updated as

$$\mathbf{x}^{(i+1)} = \mathcal{T}_{\frac{\bar{\lambda}}{\mu}}\left\{\left(\mathbf{I}_N - \frac{1}{\mu}\mathbf{B}^{\text{H}}\mathbf{B}\right)\mathbf{x}^{(i)} + \frac{1}{\mu}\mathbf{B}^{\text{H}}\bar{\mathbf{f}}\right\},\ i = 1, 2, \cdots, \tag{19}$$

where μ is a constant parameter that controls the step size of each iteration and $\mathbf{I}_N$ is an $N \times N$ identity matrix. In (19), $\mathcal{T}_\alpha\{\cdot\}$ is an elementwise soft-thresholding operator defined as $\mathcal{T}_\alpha\{x\} = \text{sign}(x)\ \max\{|x|, 0\}$.

By replacing $\lambda = \frac{\bar{\lambda}}{\mu}$, $\mathbf{W} = \left(\mathbf{I}_N - \frac{1}{\mu}\mathbf{B}^{\text{H}}\mathbf{B}\right)$, and $\mathbf{V} = \frac{1}{\mu}\mathbf{B}^{\text{H}}$, in (19), we have

$$\mathbf{x}^{i+1} = \mathcal{T}_\lambda\left\{\mathbf{W}\mathbf{x}^i + \mathbf{V}\bar{\mathbf{f}}\right\}. \tag{20}$$

The iterations of ISTA are summarized in Fig. 2. In the above formulation, $\boldsymbol{\theta} = \{\lambda, \mathbf{W}, \mathbf{V}\}$ denote reconstruction parameters which are fixed in standard ISTA. Whereas, in LISTA, these parameters are learned by unrolling the ISTA algorithm, as shown in Fig. 2. Specifically, a network is constructed by stacking a finite number of layers where each layer is a map of one iteration. The parameters $\boldsymbol{\theta}$ are then learned to minimizing cost $\sum_{q=1}^{Q}\|\mathbf{x}_q - \hat{\mathbf{x}}_q\|_2^2$ over a set of example. Here $\hat{\mathbf{x}}_q$ is the output of the P-th layer.

The LISTA-based reconstruction algorithm is well suited for our goal; it leads to a parametric, data-driven reconstruction and does not require exact knowledge of the measurement matrix $\mathbf{B}$, which is a function of $\mathbf{h}$.

Algorithm 4 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** $i \in \mathcal{N}\backslash\mathcal{K}^{(k-1)}$ **do**
(**a**) For a binary-valued vector $\mathbf{c}_i \in \{0, 1\}^{|\mathcal{N}|}$, set $\text{supp}\{\mathbf{c}_i\} = \mathcal{K}^{(k-1)} \cup \{i\}$
(**b**) $\boldsymbol{\theta}_i^{(k)} = \arg\min_{\boldsymbol{\theta}} \frac{1}{Q}\sum_{q=1}^{Q}\|\mathbf{x}_q - r_{\boldsymbol{\theta}}\left(\text{diag}(\mathbf{c}_i)\mathbf{f}_q\right)\|_2^2$ where $r_{\boldsymbol{\theta}}$ is a LISTA-based reconstruction.
(**c**) $\mathbf{x}_{q,i} = r_{\boldsymbol{\theta}_i^{(k)}}\left(\text{diag}(\mathbf{c}_i)\mathbf{f}_q\right)$ for $q = 1, \cdots, Q$
end for

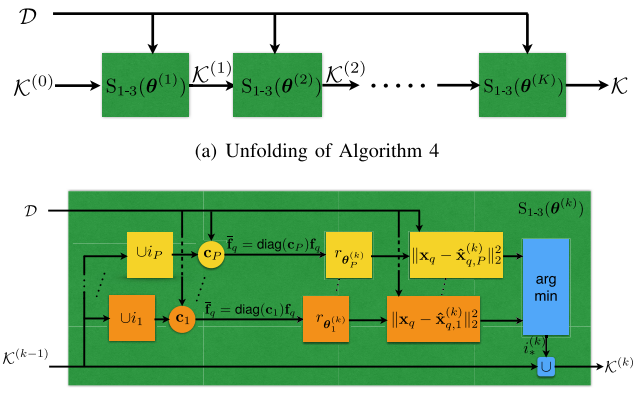
$$\text{[S2]}\ i_*^{(k)} = \arg\min_{i\in\mathcal{N}\backslash\mathcal{K}^{(k-1)}} \frac{1}{Q}\sum_{q=1}^{Q} \text{C}\left(\mathbf{x}_q, \mathbf{x}_{q,i}\right)$$

[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 $\boldsymbol{\theta}_{i_*^{(k)}}$

C. PROPOSED JOINT SAMPLING AND RECONSTRUCTION (JSR) ALGORITHM

Both Algorithm 2 and its variant Algorithm 3 are functions of the parameter of the reconstruction algorithm $\boldsymbol{\theta}$. While it is fixed in both of these algorithms, finding an optimal $\boldsymbol{\theta}$ for a given set of examples results in a solution to the joint optimization problem in (13). To this end, we use the LISTA-based reconstruction method while the greedy algorithm determines the sampling set. Details of the proposed joint sub-sampling and reconstruction approaches are described in Algorithm 4. As in the forward greedy approach (Algorithm 2), we start with zero samples and sequentially add a new sample. Alternatively, one can also start from full samples and remove one at a time, as in the backward greedy approach (Algorithm 3).

A flow diagram of the JSR algorithm is shown in Fig. 3 where for simplicity of the discussion the steps [S1]-[S3] are denoted by an operator $\text{S}_{1\text{-}3}(\boldsymbol{\theta})$. At the k-th iteration, the operator $\text{S}_{1\text{-}3}(\boldsymbol{\theta})$ acts on the previous sampling pattern $\mathcal{K}^{(k-1)}$ and outputs $\mathcal{K}^{(k)}$, that is, $\mathcal{K}^{(k)} = \text{S}_{1\text{-}3}(\boldsymbol{\theta})\{\mathcal{K}^{(k-1)}\}$. Note that the parameters of the reconstruction algorithm vary for each iteration as the number of samples varies with iteration. Moreover, within each iteration, for different sampling patterns, different reconstruction parameters are used, as shown in Fig. 3(b).



(a) Unfolding of Algorithm 4

(b) Flow diagram of $S_{1\text{-}3}(\boldsymbol{\theta})$ of k-th iteration

FIGURE 3. A flow diagram of Algorithm 4: (a) Steps [S1]-[S3] are denoted by $S_{1\text{-}3}(\theta)$, and the parameter θ of the reconstruction algorithm varies for each iteration. (b) For k-th iteration, $P = N - k$; For each r_θ, LISTA network shown in Fig. 2 is used, and the parameters are learned from the given data.

In Algorithm 4, LISTA is used to solve the optimization problem in [S1](b) (cf. Fig. 2). Note that the parameters of LISTA are learned sequentially for a given sampling pattern in each layer. Hence, backpropagation is not required from one iteration to another of [S1]. As a consequence, derivatives do not exist over sampling optimization. In addition, the network is flexible, as explained next.

Consider the problem of selecting K samples together with optimal reconstruction. Then, the proposed unfolding architecture results in a pair of optimal sampling patterns and corresponding reconstruction parameters $\{\mathcal{K}^{(k)}, \boldsymbol{\theta}_{i_*}^{(k)}\}$ for $k = 1, \cdots, K$. As a result, one need not retrain the network for the problem of selecting less than K samples. On the other hand, if the objective is to choose more than K samples together with optimal reconstruction parameters, then one can start from the sampling pattern $\mathcal{K}^{(K)}$ and add additional samples by using Algorithm 4. In this case, we need not train the network from scratch.

The formulation in (11) is also similar to the standard CS problem, where the objective is to recover sparse vectors from their compressed measurements. However, a major difference is that in a CS problem, the measurement matrix is assumed to be known, whereas it need not be the case here. Irrespective of this, the proposed joint sampling and reconstruction method can be applied to cases where CS can be applied.

D. TRAINING OF JSR NETWORK

As discussed in the previous section, in the proposed network, the parameters in each layer (correspond to one iteration of Algorithm 4) are learned sequentially. Hence, for selecting a new sampling pattern in each iteration, dedicated LISTA networks are trained. For example, at k-th iteration, to select the sampling pattern $\mathcal{K}^{(k)}$, individual LISTA networks are trained for each candidate sampling pattern of size k. These LISTA networks within a sampling layer are trained simultaneously with random initializations with the same seed. To optimize the LISTA parameter for ith sampling pattern in the k-th sampling layer we use the following loss function $\boldsymbol{\theta}_i^{(k)} = \arg\min_{\boldsymbol{\theta}} \frac{1}{Q} \sum_{q=1}^{Q} \|\mathbf{x}_q - r_{\boldsymbol{\theta}}\left(\text{diag}(\mathbf{c}_i)\mathbf{f}_q\right)\|_2^2$. Ten unfolding layers are used for each LISTA network. We use the Adam solver with a learning rate of 0.001 to optimize the network.

In terms of computational complexity, the algorithm requires multiple training of the LISTA network. However, within each iteration of the greedy algorithm, the LISTA networks can be trained in parallel to reduce the training time. In applications such as MRI, where Fourier samples are typically considered in a two-dimensional plane, N and K are large. To reduce the computations, instead of adding one sample at a time as the current approach, one can sequentially add one sampling pattern from a set of predefined sampling patterns as in [28]. For example, one can sequentially add rows of samples from a two-dimensional Fourier plane.

IV. NUMERICAL RESULTS

In this section, the performance of the proposed joint sampling and reconstruction network is assessed in comparison with the following methods:

Rand+FISTA: This is a non-learning-based approach where the sampling pattern is randomly chosen, and FISTA is used for reconstruction.

G-CRLB+FISTA: This method is non-learning and deterministic where Algorithm 3 with a CRLB-based cost function given in (17) is used for subsampling. Then FISTA is used for recovery. To evaluate cost, we use the expression for CRLB in the estimation of time-delays and amplitudes derived in [57].

G-FISTA+FISTA: We use Algorithm 3 with FISTA as the recovery method in step [S1] to generate the sampling pattern and used FISTA for recovery. In principle, this algorithm is similar to that in [27] and [28]. In [S2], we used squared-error as a cost function:

$$\text{C}\left(\mathbf{x}_q, \mathbf{x}_{q,\mathcal{K}\setminus\{i\}}\right) = \|\mathbf{x}_q - \mathbf{x}_{q,\mathcal{K}\setminus\{i\}}\|_2^2. \tag{21}$$

We use the same cost function in the following three algorithms.

G-FISTA+LISTA We use Algorithm 3 with FISTA as a recovery method in step [S1] to generate the sampling pattern. Then a LISTA network is optimized for recovery for the given sampling pattern.

JSR-1: We apply joint-sampling and recovery Algorithm 4 to determine the sampling pattern and recovery parameters.

JSR-2: Algorithm 4 is modified to start from full samples and sequentially remove one sample per iteration as in Algorithm 3.

Except **JSR-1** and **JSR-2** algorithms, all the algorithms mentioned above require the knowledge of the pulse shape $\mathbf{h}$. In addition, we pick the L strongest peaks of the estimated sparse vectors for comparison with the ground-truth sparse vector.

We consider two experiments. The first one is to compare different algorithms, as mentioned above, for a given data set. The second is to analyze the data-driven approach of the proposed algorithm. For both experiments, NMSE is used as the performance metric. For the test data $\{\mathbf{x}_i\}_{i=1}^{Q_{\text{Test}}}$, NMSE is computed as

$$\text{NMSE} = \frac{\sum_{i=1}^{Q_{\text{Test}}} \|\mathbf{x}_i - \hat{\mathbf{x}}_i\|_2^2}{\sum_{i=1}^{Q_{\text{Test}}} \|\mathbf{x}_i\|_2^2}, \tag{22}$$

where $\hat{\mathbf{x}}_i$ is an estimate of $\mathbf{x}_i$. To determine accuracy for support recovery, we compute the hit rate as

$$\text{Hit rate} = \frac{1}{LQ_{\text{Test}}} \sum_{i=1}^{Q_{\text{Test}}} \text{supp}(\mathbf{x}_i) \cap \text{supp}(\hat{\mathbf{x}}_i), \tag{23}$$

where the set $\text{supp}(\mathbf{x}_i)$ denotes L non-zero locations of $\mathbf{x}_i$. A hit rate equal to one implies that the support is estimated exactly over all the examples. To analyze the noisy scenario, we assume that the clean measurements $\bar{\mathbf{f}}$ are contaminated by circular, additive white Gaussian noise with zero mean and variance σ^2. For a given a clean signal $\bar{\mathbf{f}}$, the signal-to-noise ratio (SNR) is computed as

$$\text{SNR} = \frac{\|\bar{\mathbf{f}}\|_2^2}{N\sigma^2}. \tag{24}$$

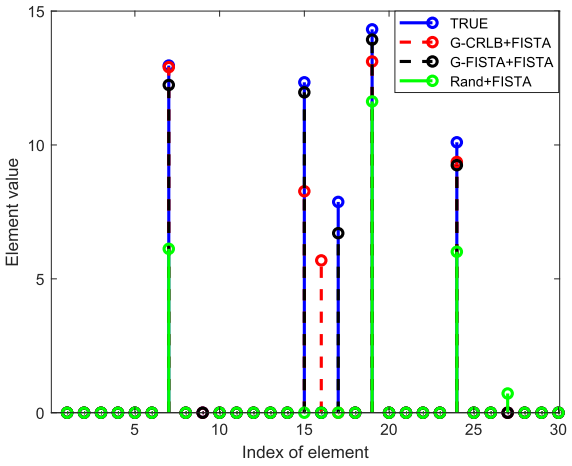
A. PERFORMANCE COMPARISON OF DIFFERENT METHODS

In this experiment, we consider the problem of sampling and reconstructing FRI signals with $L = 5$ and $t_{\max} = 1$. We generated $Q = 400,000$ examples for training and $Q_{\text{test}} = 5,000$ samples to test as in (12) for $N = 30$. For generating sparse vector $\mathbf{x}_q$, first, its support is chosen uniformly at random over the set $\{1, 2, \cdots, N\}$. Then amplitudes are generated as i.i.d. Gaussian random variables with a mean value of 10 and variance 3. A reason for choosing a large data set is that there should be sufficient examples for each sparsity pattern. For an N-length, L-sparse vector $\mathbf{x}$ there are $\binom{N}{L}$ possible sparsity patterns. For $N = 30$ and $L = 5$, there would be 142,506 possible sparsity patterns. Hence, we choose $Q = 400,000$ examples to train and expect that the networks all possible sparse vectors while training.

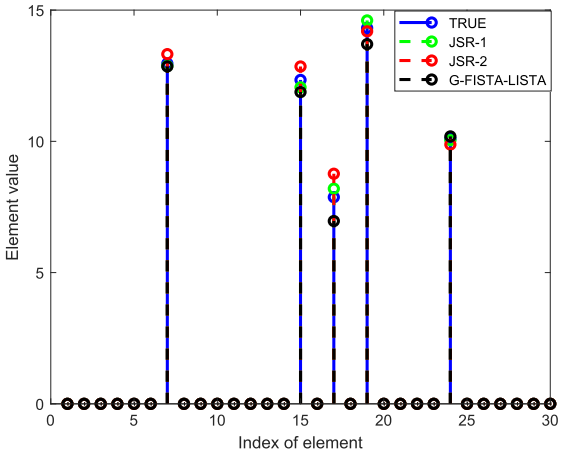
We consider an FRI pulse $h(t)$ such that its n-th sample of $\mathbf{h} \in \mathbb{C}^N$ is given as $0.01 + e^{-0.04(n-3N/4)^2} + e^{-0.04(n-N/8)^2}$. By generating the Fourier samples of $h(t)$ as a sum of Gaussian pulses, we ensure that they are non-zero as desired and have varying spectra. With varying spectral values, SNRs vary for each Fourier sample. Specifically, large values of $\mathbf{h}$ may result in higher SNR, which may play a role in the selection of the corresponding sample.

With the experimental setup, we learned sampling pattern and reconstruction parameters for $K = 8, \cdots, 15$. Note that for $L = 5$, theoretically, a minimum of $2L = 10$ samples are required for perfect reconstruction. In choosing $K < 2L$, our objective is to verify if the joint learning-based approaches break the theoretical barrier. In Fig. 4, we show one recovery of a sparse vector for $K = 8$. We observe that all the learning-based methods were able to determine the support of the sparse filter perfectly, as shown in Fig 4(b), whereas there is an error in the estimation of the amplitudes. Compared to the learning-based methods, the non-learning-based methods fail to estimate the sparse vectors perfectly (see Fig 4(a)). Since support recovery is sufficient in applications such as radar, we computed hit rates (averaged over all test examples) for different methods and plotted them in Fig. 5 as a function of the sampling rate. As in the previous observation, we note that the learning-based methods are able to estimate the support perfectly, even below the theoretical minimum sampling rate. Hence, one can go below the sub-Nyquist rate by using the proposed learning-based approaches.

In Fig. 6, we show an example of recovery of the sparse vector for SNR = 30 dB and $K = 2L = 10$. We observe that while all the methods were able to estimate the support



(a) Non-learning methods.



(b) Learning schemes

FIGURE 4. Comparison of different approaches in the absence of noise for $L = 5$, $N = 30$, $K = 8$.

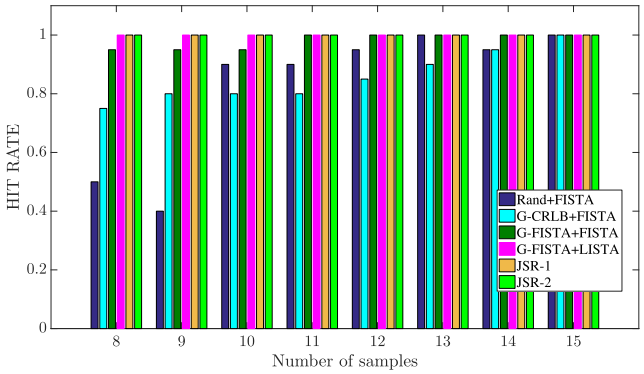
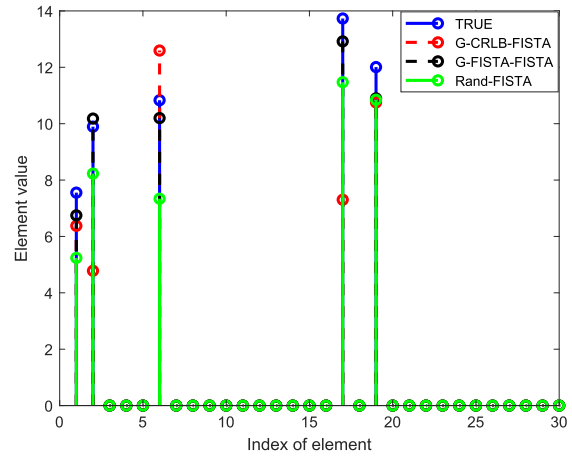
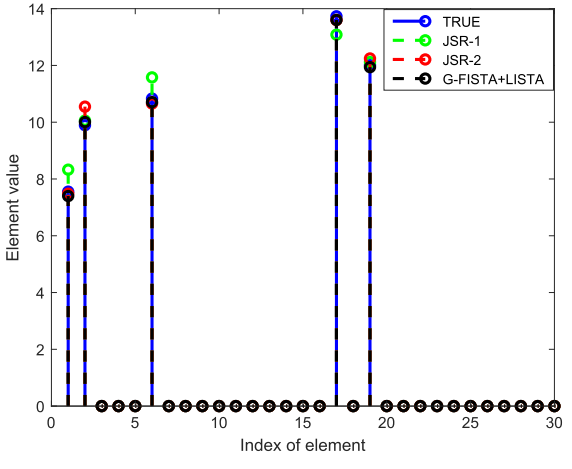


FIGURE 5. A comparison of hit rates for different methods as a function of number of samples in the absence of noise. The learning-based approaches are able to determine the support perfectly for the sampling rate below the sub-Nyquist rate ($K < 2L = 10$).



(a) Non-learning methods.



(b) Learning schemes

FIGURE 6. Comparison of different approaches in the presence of noise for $L = 5$, $N = 30$, $K = 8$, and SNR= 30 dB.

or the time-delays of the FRI signals perfectly, amplitude estimates are closer to the true values in the learning-based methods (Fig. 6(b)) compared to non-learning-based methods

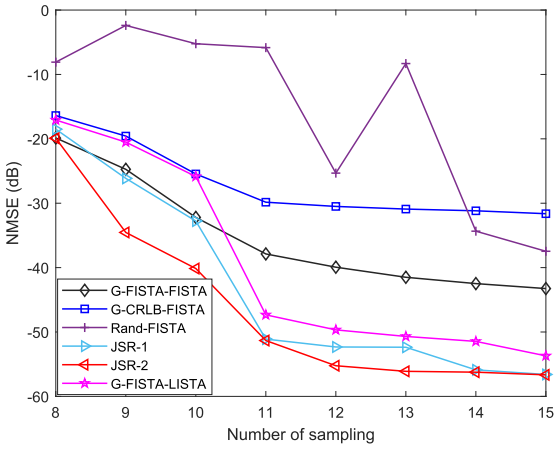


FIGURE 7. A comparison different methods as a function of number of samples in the absence of noise. The proposed JSR-2 algorithm results in lowest NMSE among different methods.

(Fig. 6(a)). For this particular example, among the learning-based methods, **G-FISTA+LISTA** is able to estimate the amplitudes more accurately compared to **JSR** methods. However, on average, the proposed **JSR-2** algorithm has the lowest NMSE for different values of K and SNRs, as discussed next.

In Fig. 7, we compared the algorithms for different values of K in the absence of noise. In comparison to **JSR-1**, **JSR-2** method has 8 dB lower NMSE for $K = 10$, and the gap between NMSEs reduces as K increases. **JSR-1** algorithm has $5 - 15$ lower NMSE compared with **G-FISTA+LISTA** and has a gain of more than 15 dB compared to non-learning-based methods. Below the theoretical barrier on the sampling rate, that is, for $K = 8, 9$ **JSR-2** has the lowest error compared with the other methods. From Fig. 7, we notice that for the Rand+FISTA scheme, the performance of 13 samples is worse than that of 12 samples. This is because, for the Rand+FISTA scheme, the sampling pattern is randomly generated, which further illustrates the importance of sampling strategy.

To compare the performance of the methods in the presence of noise, in Fig. 8, we show NMSE as a function of the number of samples for 30 dB SNR. The learning-based methods result in lower NMSE compared to non-learning-based algorithms. However, unlike the results in the absence of noise, the **JSR-1** method has higher NMSE compared to that of **G-FISTA+LISTA**. Moreover, the gap between NMSEs of **JSR-1** and **JSR-2** is larger. This behavior is a consequence of the fact that the latter approach starts with full measurements compared to zero measurements in the former approach. As a result, at any given iteration of the algorithm, the **JSR-2** method has a larger number of samples to estimate the sparse vectors compared to that in **JSR-1** and hence, it is able to perform better.

In Fig. 9, we compare NMSEs of different methods for SNRs varying from 20 to 40 dB with $K = 15$. We observe that the learning-based methods outperform non-learning-based

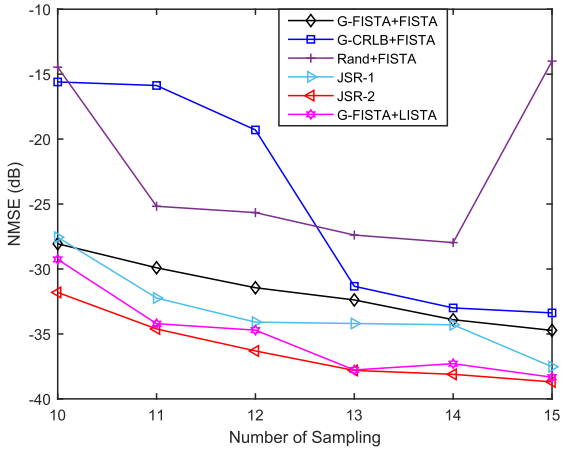


FIGURE 8. A comparison of performances different methods as a function of number of samples with 30 dB SNR. The proposed JSR-2 algorithm results in lowest NMSE among different methods.

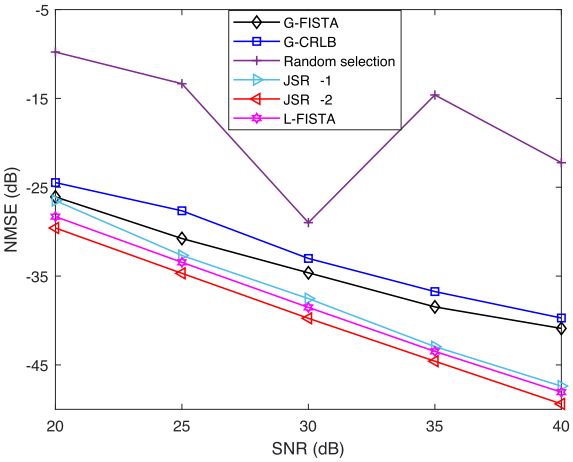
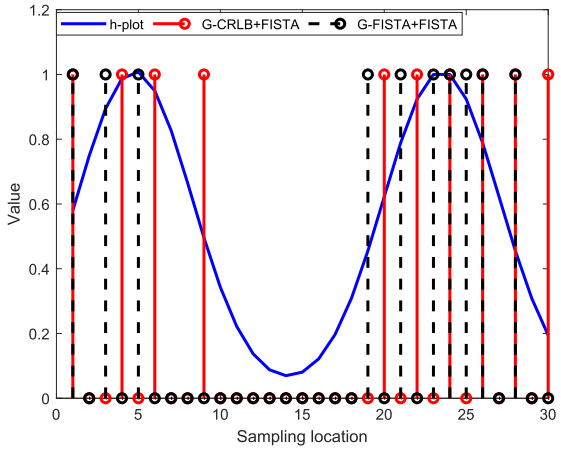
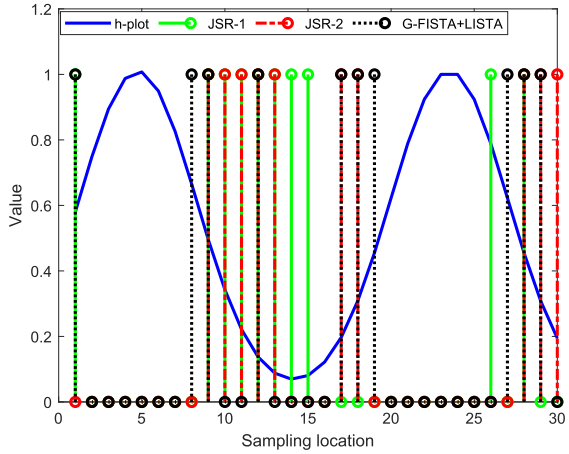


FIGURE 9. A Comparison of NMSE vs SNR for different methods for $K = 15$: learning-based methods have lower NMSE compared to non-learning-based algorithms JSR-2 algorithm results in lowest NMSE among different methods.



(a) Non-learning methods.



(b) Learning schemes

FIGURE 10. A comparison of sampling patterns estimated from various methods with a non-flat **h** (in blue). The sampling pattern of the learning-based methods are well spread compared to the non-learning algorithms.

methods with more than 5 dB improvement in NMSE, the gain in NMSE of **JSR-1** compared to NMSE of either **JSR-2** or **G-FISTA+LISTA** is only $1-3$ dB. In the noisy scenario, **G-FISTA+LISTA** algorithm is preferable over **JSR-2** as one LISTA network has to be trained.

Next, we study the sampling patterns resulted from each method with respect to the frequency response of the pulse as shown in Fig. 10 for $K = 10$ and SNR $= 30$ dB. We observe that the sampling pattern of **G-CRLB+FISTA** and **G-FISTA+FISTA** are localized around the peak of **h** compared to the learning-based methods. Both **JSR** and **G-FISTA+LISTA** depend on the recovery method, and we believe that a wider spread or less localization of the samples results in lower coherence in the resulting measurement matrix, which in turn leads to better estimation. Moreover, the sampling patterns show that it is not always necessary to consider samples from high-SNR regions, as shown in Fig. 10(b).

B. DATA-ADAPTABILITY OF JSR ALGORITHM

In this experiment, we study the data-adaptive nature of the proposed JSR algorithm to any additional structure within the data. We consider two different data sets for training and testing. For both the data sets, we have $N = 30$ and **h** remains the same as in the previous experiment. The goal is to determine the sampling patterns and LISTA parameters $\{\mathcal{K}_1, \boldsymbol{\theta}_1\}$ and $\{\mathcal{K}_2, \boldsymbol{\theta}_2\}$ and then study the effect of mismatch between sampling pattern and recovery. For both the data sets, we assume $L = 5$, but they differ in terms of their sparsity patterns. Specifically, for the first data set, the first two nonzero elements are chosen uniformly at random over the set $\{1, \cdots, 10\}$ and the remaining three from $\{21, \cdots, 30\}$. The nonzero values for the second data set are chosen uniformly at ransom over the set $\{11, \cdots, 20\}$. For both the data sets, the

TABLE 1. A comparison of NMSE while learning structured sparsity.

NMSE (dB)	$\mathcal{K}_1, \boldsymbol{\theta}_1$	$\mathcal{K}_2, \boldsymbol{\theta}_1$	$\mathcal{K}_1, \boldsymbol{\theta}_2$	$\mathcal{K}_2, \boldsymbol{\theta}_2$
Test Data 1	-52.25	-1.38	7.94	-0.96
Test Data 2	0.06	0.18	1.39	-47.09

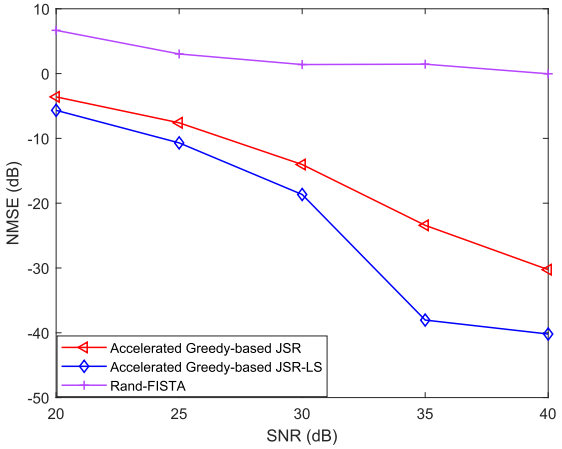


FIGURE 11. Large scale problem: A comparison of NMSE vs SNR for $N = 128$ and $K = 30$.

amplitude of the sparse vectors is chosen as in the previous experiment. We consider 40,000 samples for training and 2000 for testing with $|\mathcal{K}_1| = |\mathcal{K}_2| = K$. In Table 1, we show NMSE incurred by **JSR-2** algorithm for different combinations of sampling patterns and recovery parameters. We note that the error in the reconstruction is negligible when matched sampling and reconstructions are used, that is, for the combination $\{\mathcal{K}_1, \boldsymbol{\theta}_1\}$ and $\{\mathcal{K}_2, \boldsymbol{\theta}_2\}$. However, the NMSE is high when we used cross-combinations $\{\mathcal{K}_n, \boldsymbol{\theta}_m\}$, $n, m \in \{1, 2\}$ and $n \neq m$. We believe that the large NMSE for the cross-combinations is due to non-overlapping support of the sparse vectors. The observations lead to the conclusion that JSR approaches can better adapt to the sparse vectors' local structures. In addition, the learned sampling and reconstruction methods are not global and are highly data dependent.

C. EXTENSION TO A LARGE SCALE PROBLEM

For choosing K samples from total N measurements, the FISTA network to be trained $KN - \frac{K(K-1)}{2}$ and $\frac{(N-K)(N+K+1)}{2}$ times for the proposed **JSR-1** and **JSR-2** algorithms, respectively. Hence, for large N greedy algorithms may not be suitable even though one can apply parallel training to speed up the training or use a collection of samples as in MRI instead of individual ones. Several alternatives of the greedy algorithm have been suggested to improve the computational cost, such as distributed greedy approaches [58], [59] and accelerated greedy algorithm [60]. In the distributed greedy method, the data or the samples are distributed among several machines and solve smaller greedy problems in parallel. Then the solutions are placed on a single machine, and the standard greedy algorithm is used once more to select the final solution. The accelerated greedy approach [60] is very similar in principle to the standard greedy approach except for the fact that each sample is given a weight based on their contribution to the recovery process. In the conventional greedy approach, at each iteration, the cost functions for all the remaining samples are evaluated, and then the sample with minimum error is retained. The accelerated greedy algorithms start from the samples with higher weights and stop as soon as a sample with low error is chosen (cf. [60] more details). In this way, one need not perform too many evaluations of the cost function, which requires as many pieces of training of the FISTA network.

In this simulation, we modified **JSR-2** algorithm based on the accelerated greedy algorithm. We consider a problem of choosing $K = 30$ samples from $N = 128$ samples. All the remaining settings are the same as in the simulations of Section IV-A. In this case, **JSR-2** approach is required to be trained 7791 times, which is not feasible. In comparison, its accelerated version required only 264 instances of FISTA training. This shows that we need approximately 3.4% of the computational power. We compared the performance of the accelerated version with a random selection strategy for different SNRs, and the results are shown in Fig. 11. We observe that compared to random selection, the accelerated-greedy-based JSR approach has $10 - 30$ dB lower MSE in the estimation of sparse vectors.

In addition to these two approaches, we also compared results when the amplitudes of the sparse vectors are re-estimated by using the least-squares (LS). Specifically, we first estimate the support and then apply LS restricted to the estimated support. As we observed in the previous simulations, the learning-based methods were able to produce accurate estimates of the supports. Hence, one can use least-squares to improve the amplitude estimation by using the estimated supports. In Fig. 11, we show estimates for the standard approach and the LS approach. We observe a gain of $2 - 10$ dB by using least-squares. However, the least-squares approach requires an exact knowledge of the measurement matrix that is a function of the FRI pulse shape, limiting the least-square approach's applicability.

Training for all the networks was done for $Q = 400,000$ examples which could be very high for many applications. As this is the initial study of a combination of greedy approaches together with a model-based deep learning-based strategy for joint sampling and recovery, our focus is on the proof of concept and showing its potential performance gain. Therefore, we consider a relatively large data set. Since the accuracy of the trained model is highly related to the size of training sets, we consider a smaller, $Q = 40,000$, examples for training with the accelerated greedy approach for $N = 128$ and $K = 30$. The MSEs are shown in Fig. 12. As expected, we can see that the performance of the proposed algorithm with a small training set is a little bit worse than that with a large training set, but it is still better than the randomly selected approach. In addition, we would

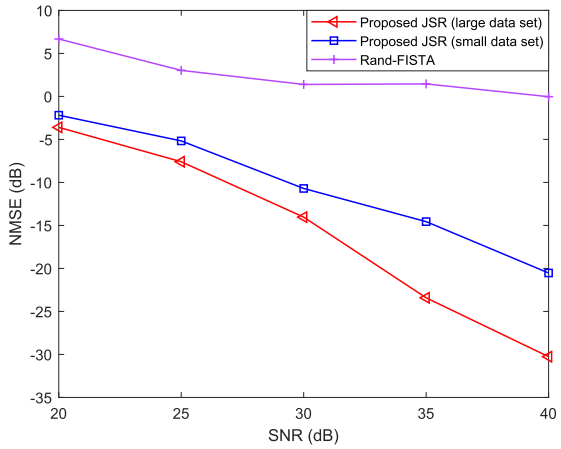


FIGURE 12. A comparison of NMSE vs SNR under different sizes of training data sets, with $N = 128$ and $K = 30$. The larger data set consists of 400,000 examples, and the smaller data set had 40,000 examples.

like to highlight that one of the advantages of our proposed joint sampling and recovery scheme is that our proposed scheme does not require the exact knowledge of FRI pulse shape, while this information is required by conventional approaches such as FISTA, making our proposed approach much more attractive in practical applications.

V. CONCLUSION

We propose a learning-based solution for a joint subsampling and reconstruction problem. The proposed algorithm is data-adaptive, circumvents the differentiability issue, and is flexible to change in the sampling rate. We show that the proposed algorithm can be utilized to design sampling kernels for FRI signals. Apart from sampling and reconstruction of FRI signals, our framework can be used for selecting k-space sampling in MRI, sinograms in computed tomography, and similar applications. Our methods have lower NMSE compared to independent design approaches. In addition, the sampling rate can be reduced below the sub-Nyquist rate which results in low-cost receivers.

REFERENCES

[1] Y. C. Eldar, *Sampling Theory: Beyond Bandlimited Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2015.

[2] 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.

[3] O. Bar-Ilan and Y. C. Eldar, "Sub-Nyquist radar via Doppler focusing," *IEEE Trans. Signal Process.*, vol. 62, no. 7, pp. 1796–1811, Apr. 2014.

[4] 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.

[5] S. Mulleti, S. Nagesh, R. Langoju, A. Patil, and C. S. Seelamantula, "Ultrasound image reconstruction using the finite-rate-of-innovation principle," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 1728–1732.

[6] 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.

[7] P. L. Dragotti, M. Vetterli, and T. Blu, "Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets Strang–Fix," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1741–1757, May 2007.

[8] 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, Nov. 2017.

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

[10] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[11] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, 2004.

[12] D. P. Palomar and Y. C. Eldar, *Convex Optimization in Signal Processing and Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[13] A. Beck and M. Teboulle, "A Fast Iterative Shrinkage-Thresholding Algorithm with application to wavelet-based image deblurring," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2009, pp. 693–696.

[14] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.

[15] 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, Sep. 2016.

[16] 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, Sep. 2018.

[17] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proc. Int. Conf. Mach. Learn. (ICML)*. Madison, WI, USA: Omnipress, Jun. 2010, pp. 399–406.

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

[19] M. Borgerding, P. Schniter, and S. Rangan, "AMP-inspired deep networks for sparse linear inverse problems," *IEEE Trans. Signal Process.*, vol. 65, no. 16, pp. 4293–4308, Aug. 2017.

[20] P. K. Pokala, A. G. Mahurkar, and C. S. Seelamantula, "FirmNet: A sparsity amplified deep network for solving linear inverse problems," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 2982–2986.

[21] D. Ito, S. Takabe, and T. Wadayama, "Trainable ISTA for sparse signal recovery," *IEEE Trans. Signal Process.*, vol. 67, no. 12, pp. 3113–3125, Jun. 2019.

[22] 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*. USA: Springer, 2015, pp. 143–167.

[23] 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, pp. 1182–1195, 2007.

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

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

[26] S. Mulleti, C. Saha, H. S. Dhillon, and Y. C. Eldar, "A fast-learning sparse antenna array," in *Proc. IEEE Radar Conf. (RadarConf)*, Sep. 2020, pp. 1–6.

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

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

[29] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[30] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[31] K. H. Jin, M. Unser, and K. M. Yi, "Self-supervised deep active accelerated MRI," 2019, *arXiv:1901.04547*.

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

[33] C. D. Bahadir, A. Q. Wang, A. V. Dalca, and M. R. Sabuncu, "Deep-learning-based optimization of the under-sampling pattern in MRI," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 1139–1152, 2020.

[34] T. Weiss, O. Senouf, S. Vedula, O. Michailovich, M. Zibulevsky, and A. Bronstein, "PILOT: Physics-informed learned optimized trajectories for accelerated MRI," 2019, *arXiv:1909.05773*.

[35] G. Wang, T. Luo, J. Nielsen, D. C. Noll, and J. A. Fessler, "B-spline parameterized joint optimization of reconstruction and K-space trajectories (BJORK) for accelerated 2D MRI," *IEEE Trans. Med. Imag.*, vol. 41, no. 9, pp. 2318–2330, Sep. 2022.

[36] M. V. W. Zibetti, F. Knoll, and R. R. Regatte, "Alternating learning approach for variational networks and undersampling pattern in parallel MRI applications," *IEEE Trans. Comput. Imag.*, vol. 8, pp. 449–461, 2022.

[37] Z. Huang and S. Ravishankar, "Single-pass object-adaptive data under-sampling and reconstruction for MRI," *IEEE Trans. Comput. Imag.*, vol. 8, pp. 333–345, 2022.

[38] T. Yin, Z. Wu, H. Sun, A. V. Dalca, Y. Yue, and K. L. Bouman, "End-to-end sequential sampling and reconstruction for MRI," 2021, *arXiv:2105.06460*.

[39] 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. Med. Imag.*, vol. 39, no. 12, pp. 3955–3966, Dec. 2020.

[40] F. Sherry, M. Benning, J. C. De los Reyes, M. J. Graves, G. Maierhofer, G. Williams, C. Schönlieb, and M. J. Ehrhardt, "Learning the sampling pattern for MRI," *IEEE Trans. Med. Imag.*, vol. 39, no. 12, pp. 4310–4321, Dec. 2020.

[41] Q. Yan, X. Xiong, K. Lei, Y. Zheng, and Y. Wang, "Sampling and reconstruction jointly optimized model unfolding network for single-pixel imaging," *Photonics*, vol. 10, no. 3, p. 232, Feb. 2023.

[42] J. Wang, Q. Yang, Q. Yang, L. Xu, C. Cai, and S. Cai, "Joint optimization of Cartesian sampling patterns and reconstruction for single-contrast and multi-contrast fast magnetic resonance imaging," *Comput. Methods Programs Biomed.*, vol. 226, Nov. 2022, Art. no. 107150.

[43] C. Chen, Y. Wu, C. Zhou, and D. Zhang, "JsrNet: A joint sampling–reconstruction framework for distributed compressive video sensing," *Sensors*, vol. 20, no. 1, p. 206, Dec. 2019.

[44] C. Zeng, J. Ye, Z. Wang, N. Zhao, and M. Wu, "JSRNN: Joint sampling and reconstruction neural networks for high quality image compressed sensing," 2022, *arXiv:2211.05963*.

[45] G. R. DeProny, "Essai experimental et analytique: Sur les lois de la dilatabilité de fluides élastiques et sur celles de la force expansive de la vapeur de l'eau et de la vapeur de l'alcool, à différentes températures," *J. de l'Ecole polytechnique*, vol. 1, no. 2, pp. 24–76, 1795.

[46] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 5, pp. 814–824, May 1990.

[47] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 451–462, Feb. 2009.

[48] S. P. Chepuri and G. Leus, "Sparsity-promoting sensor selection for non-linear measurement models," *IEEE Trans. Signal Process.*, vol. 63, no. 3, pp. 684–698, Feb. 2015.

[49] C. Qian, Y. Yu, and Z.-H. Zhou, "Subset selection by Pareto optimization," in *Proc. Adv. Neural Info. Process. Syst.*, vol. 28, 2015, pp. 1774–1782.

[50] C. Qian, "Distributed Pareto optimization for large-scale noisy subset selection," *IEEE Trans. Evol. Comput.*, vol. 24, no. 4, pp. 694–707, Aug. 2020.

[51] M. V. W. Zibetti, G. T. Herman, and R. R. Regatte, "Fast data-driven learning of MRI sampling pattern for large scale problems," 2020, *arXiv:2011.02322*.

[52] 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.

[53] H. Godrich, A. P. Petropulu, and H. V. Poor, "Sensor selection in distributed multiple-radar architectures for localization: A knapsack problem formulation," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 247–260, Jan. 2012.

[54] S. Fujishige, *Submodular Functions and Optimization* (ISSN). Amsterdam, The Netherlands: Elsevier, 2005.

[55] E. Tohidi, M. Coutino, S. P. Chepuri, H. Behroozi, M. M. Nayebi, and G. Leus, "Sparse antenna and pulse placement for colocated MIMO radar," *IEEE Trans. Signal Process.*, vol. 67, no. 3, pp. 579–593, Feb. 2019.

[56] C. Couvreur and Y. Bresler, "On the optimality of the backward greedy algorithm for the subset selection problem," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 3, pp. 797–808, Jan. 2000.

[57] 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.

[58] R. Barbosa, A. Ene, H. Nguyen, and J. Ward, "The power of randomization: Distributed submodular maximization on massive datasets," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2015, pp. 1236–1244.

[59] B. Mirzasoleiman, A. Karbasi, R. Sarkar, and A. Krause, "Distributed submodular maximization: Identifying representative elements in massive data," in *Proc. Neural Info. Process. Sys. (NIPS)*, 2013, pp. 2049–2057.

[60] M. Minoux, "Accelerated greedy algorithms for maximizing submodular set functions," in *Optimization Techniques* (Lecture Notes in Control and Information Sciences), J. Stoer Ed. Berlin, Germany: Springer, 1978, pp. 234–243.

SATISH MULLETI (Member, IEEE) received the Bachelor of Engineering degree from the Electronics and Communication Engineering Department, Jalpaiguri Government Engineering College, India, in 2005, and the Master of Engineering degree in electrical engineering from the Department of Electrical Engineering, Indian Institute of Technology Kanpur, India, in 2009. He was a Researcher with the Indian Space Research Organization (ISRO), India, and Tata Consultancy Services (TCS) Innovation Labs, Mumbai, India. In August 2011, he joined the Spectrum Laboratory, Department of Electrical Engineering, Indian Institute of Science, Bengaluru, for the Ph.D. degree. From April 2017 to August 2021, he was a Postdoctoral Researcher with the Department of Electrical Engineering, Technion—Israel Institute of Technology, and the Faculty of Mathematics and Computer Science, The Weizmann Institute of Science, Israel. He is currently an Assistant Professor with the Department of Electrical Engineering, Indian Institute of Technology (IIT) Bombay, India. His research interests include sampling theory, in particular, finite-rate-of-innovation signal sampling, compressive sensing, machine learning, blind deconvolution, sparse array signal processing, and spectral estimation.

HAIYANG ZHANG (Member, IEEE) received the B.S. degree in communication engineering from Lanzhou Jiaotong University, Lanzhou, China, in 2009, the M.S. degree in information and communication engineering from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2012, and the Ph.D. degree in information and communication engineering from Southeast University, Nanjing, in 2017. From 2017 to 2020, he was a Postdoctoral Research Fellow with the Singapore University of Technology and Design, Singapore. From 2020 to 2022, he was a Postdoctoral Research Fellow with The Weizmann Institute of Science, Israel, where he received the FGS Prize for outstanding research achievements. He is currently a Professor with the School of Communications and Information Engineering, Nanjing University of Posts and Telecommunications. His research interests include 6G near-field MIMO communications, learning and sampling theory, and physical-layer security. He serves as the Co-Chair for the IEEE ICC 2023 Workshop on near-field localization and communications for 6G.

YONINA C. ELDAR (Fellow, IEEE) received the B.Sc. degree in physics and the B.Sc. degree in electrical engineering from Tel Aviv University, Tel Aviv, Israel, in 1995 and 1996, respectively, and the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 2002. She was a Professor with the Department of Electrical Engineering, Technion, where she held the Edwards Chair in Engineering. She is currently a Professor with the Department of Mathematics and Computer Science, The Weizmann Institute of Science, Rehovot, Israel. She is also a Visiting Professor with MIT, a Visiting Scientist with the Broad Institute, and an Adjunct Professor with Duke University, and was a Visiting Professor with Stanford University. Her research interests include statistical signal processing, sampling theory and compressed sensing, learning and optimization methods, and their applications to biology and optics. She is a member of the Israel Academy of Sciences and Humanities (elected 2017) and a EURASIP Fellow. She was a Horev Fellow of the Leaders in Science and Technology Program at the Technion and an Alon Fellow. She was a member of the Young Israel Academy of Science and Humanities and the Israel Committee for Higher Education. She was a member of the IEEE Signal Processing Theory and Methods Technical Committee and the IEEE Bio-Imaging Signal Processing Technical Committee. She is a member of the IEEE Sensor Array and Multichannel Technical Committee. She serves on several other IEEE committees. She has received many awards for excellence in research and teaching, including the IEEE Signal Processing Society Technical Achievement Award, in 2013, the IEEE/AESS Fred Nathanson Memorial Radar Award, in 2014, and the IEEE Kiyo Tomiyasu Award, in 2016. She received the Michael Bruno Memorial Award from the Rothschild Foundation, the Weizmann Prize for Exact Sciences, the Wolf Foundation Krill Prize for Excellence in Scientific Research, the Henry Taub Prize for Excellence in Research (twice), the Hershel Rich Innovation Award (three times), the Award for Women with Distinguished Contributions, the Andre and Bella Meyer Lectureship, the Career Development Chair at the Technion, the Muriel & David Jacknow Award for Excellence in Teaching, and the Technion's Award for Excellence in Teaching (two times). She received several best paper awards and best demo awards together with her research students and colleagues, including the SIAM Outstanding Paper Prize, the UFFC Outstanding Paper Award, the Signal Processing Society Best Paper Award, and the IET Circuits, Devices and Systems Premium Award, and was selected as one of the 50 most influential women in Israel. She was the co-chair and the technical co-chair of several international conferences and workshops. She is the Editor-in-Chief of *Foundations and Trends in Signal Processing*. She served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the *EURASIP Journal of Signal Processing*, the *SIAM Journal on Matrix Analysis and Applications*, and the *SIAM Journal on Imaging Sciences*. She was a Signal Processing Society Distinguished Lecturer.