Towards Generalized FRI Sampling With an Application to Source Resolution in Radioastronomy

Hanjie Pan, *Student Member, IEEE*, Thierry Blu, *Fellow, IEEE*, and Martin Vetterli, *Fellow, IEEE*

***Abstract*—It is a classic problem to estimate continuous-time sparse signals, like point sources in a direction-of-arrival problem, or pulses in a time-of-flight measurement. The earliest occurrence is the estimation of sinusoids in time series using Prony's method. This is at the root of a substantial line of work on high resolution spectral estimation. The estimation of continuous-time sparse signals from discrete-time samples is the goal of the sampling theory for finite rate of innovation (FRI) signals. Both spectral estimation and FRI sampling usually assume uniform sampling. But not all measurements are obtained uniformly, as exemplified by a concrete radioastronomy problem we set out to solve. Thus, we develop the theory and algorithm to reconstruct sparse signals, typically sum of sinusoids, from *nonuniform* samples. We achieve this by identifying a linear transformation that relates the unknown uniform samples of sinusoids to the given measurements. These uniform samples are known to satisfy the annihilation equations. A valid solution is then obtained by solving a constrained minimization such that the reconstructed signal is consistent with the given measurements and satisfies the annihilation constraint. Thanks to this new approach, we unify a variety of FRI-based methods. We demonstrate the versatility and robustness of the proposed approach with five FRI reconstruction problems, namely Dirac reconstructions with irregular time or Fourier domain samples, FRI curve reconstructions, Dirac reconstructions on the sphere, and point source reconstructions in radioastronomy. The proposed algorithm improves substantially over state-of-the-art methods and is able to reconstruct point sources accurately from irregularly sampled Fourier measurements under severe noise conditions.**

***Index Terms*—Finite rate of innovation (FRI), approximation, sparse reconstruction, irregular sampling, continuous-time sparsity, radio interferometry.**

I. Introduction

CONSIDER a classic array signal processing problem in radio interferometry. The electromagnetic (EM) waves emitted by celestial sources in the sky are collected by an array of antennas. The received signals at two antennas differ by a phase shift, which depends on the relative distance of the antennas and the point source locations in the sky (Fig. 1). It can be shown that the cross-correlation of the received EM waves is related to the Fourier transform of the underlying sky image (see Table I) sampled at *non-uniform* frequencies [4]. The goal is to reconstruct these point sources, which are modeled as a weighted sum of Dirac delta distributions, from the irregularly sampled Fourier measurements in continuous space.

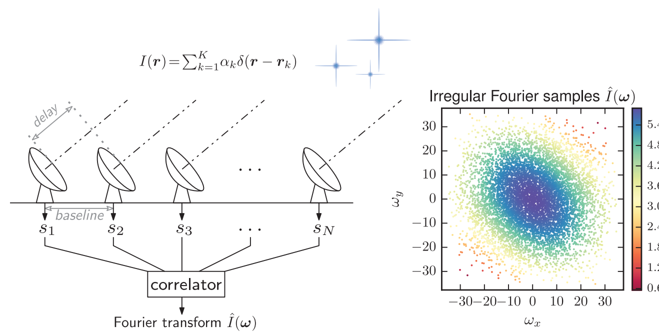


Fig. 1. Schematic diagram of a radio interferometer. The cross-correlations of the received signals at different antennas are related to the Fourier transform of the sky image (see Table I) at certain non-uniform frequencies.

TABLE I
SUMMARY OF RADIO ASTRONOMY TERMS

Term	*Meaning*
sky image	Brightness distribution of the sky.
point sources	Celestial sources that can be modeled as Dirac delta distributions.
visibility	Cross-correlations of the received signals at different antennas. It is related to the Fourier transform of the sky image.
uv-coverage	Fourier domain coverage. A radio interferometer can only cover part of the Fourier domain at some irregular frequencies.
dirty image	Inverse Fourier transform of the irregularly sampled Fourier transform of the sky image.

The classic approach in radioastronomy is to assume that the point sources are located on a discrete grid (i.e., griding). The associated *discretized* sky image is then reconstructed by taking the inverse discrete Fourier transform (see e.g. Fig. 2(a)) followed by an iterative deconvolution process [5]. Recently, it has been shown that the conventional discretized approach is related to compressed sensing, where the ℓ_1 norm of the *discretized* sky image is minimized subject to the data-fidelity constraint [6]–[8]. Note that the reconstruction accuracy of the sky image is inherently limited by the resolution of the grid: In order to obtain a more accurate reconstruction, a denser grid has to be used. Additionally, the measurement matrix, which relates the sky image to the Fourier measurements, is determined by the

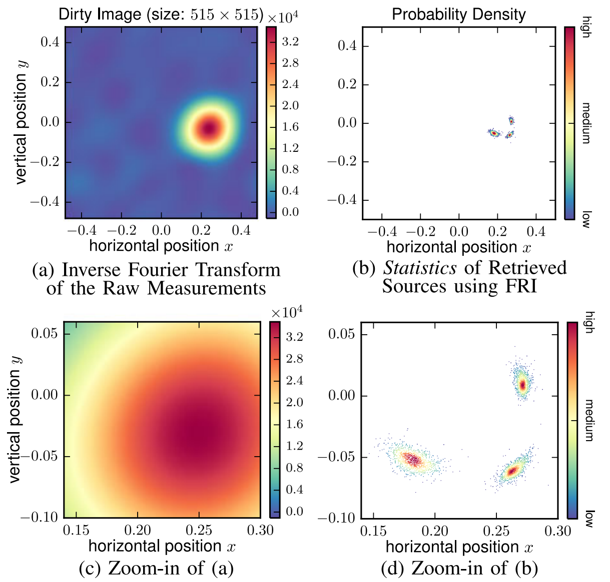
Manuscript received May 12, 2016; revised August 25, 2016 and October 8, 2016; accepted October 24, 2016. Date of publication November 4, 2016; date of current version December 5, 2016. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Wenwu Wang. This work was supported in part by Swiss National Science Foundation under Grant SNF–20FP-1_151073 and in part by the General Research Fund CUHK14600615 from the Hong Kong Research Grant Council.

H. Pan and M. Vetterli are with the School of Computer and Communication Sciences, Ecole polytechnique fédérale de Lausanne, Lausanne 1015, Switzerland (e-mail: hanjie.pan@epfl.ch; martin.vetterli@epfl.ch).

T. Blu is with the Department of Electronic Engineering, Chinese University of Hong Kong, Hong Kong (e-mail: thierry.blu@m4x.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

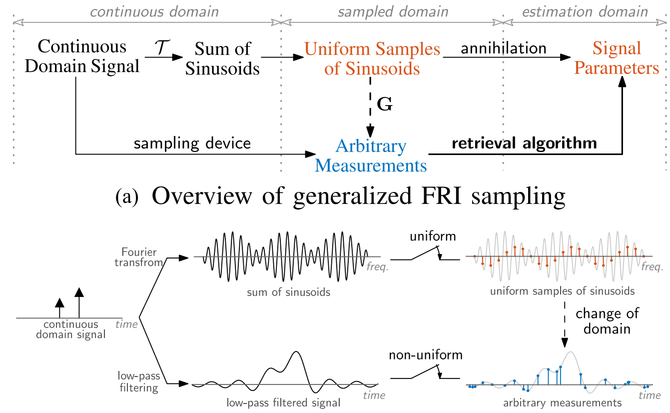
Digital Object Identifier 10.1109/TSP.2016.2625274



(a) Inverse Fourier Transform of the Raw Measurements (b) *Statistics* of Retrieved Sources using FRI

(c) Zoom-in of (a) (d) Zoom-in of (b)

Fig. 2. Accurate reconstruction of point sources' locations from partial Fourier domain measurements (number of irregular Fourier samples: 8000, SNR = 5 dB). (a) Spatial domain representation (a.k.a. "dirty image" in radioastronomy) associated with the given partial Fourier measurements. (b) Probability density of the reconstructed point source locations with the FRI approach (number of independent noise realizations:1000; average estimation error of Dirac locations: 8.07×10^{-3}). For comparison with other methods see Fig. 11.



(a) Overview of generalized FRI sampling

(b) An example of various signals in (a)

Fig. 3. Any continuous domain signal, which can be represented as a sum of sinusoids by applying a certain transformation $\mathcal{T}$, is an FRI signal. The classic FRI framework reconstructs the continuous domain signal from a set of *uniform* samples. Our focus in this paper is on cases where measurements are taken irregularly. We will identify a linear mapping $\mathbf{G}$ that relates uniform samples to these measurements with a good approximation.

layout of the antenna arrays. It does not necessarily satisfy the restricted isometry property required in standard compressed sensing theory.

Alternatively, we can address the point source reconstruction problem directly in *continuous* space by using algorithms developed for signals with finite rate of innovation [3] (FRI). A common feature of these signals is that they can either be represented as or transformed to a weighted sum of sinusoids, which is also the case for point source reconstruction (see (16) in Section V). The key to the reconstruction of FRI-type signals [3], [9]–[17] is the annihilating filter method, which is related to Prony's method [1] in spectrum estimation: We can build a finite length discrete filter (i.e., the annihilating filter) such that its convolution with *uniformly* sampled sinusoids is zero. The point source locations are then given by the roots of a polynomial, whose coefficients are specified by the annihilating filter. However, despite the ability to reconstruct the point sources directly in continuous-domain, the classic FRI approach cannot deal with irregularly sampled data.

In this paper, we propose a robust reconstruction algorithm that removes the uniform-sampling limitation from the FRI framework. Therefore, it allows to reconstruct the point sources in the continuous space with *irregularly* sampled Fourier measurements. We achieve this by establishing a linear relation between a set of unknown uniform Fourier samples, which can be annihilated by a discrete filter, and the given measurements (Fig. 3). We recast the point source reconstruction as an approximation problem, where we would like to find a sum of Dirac that is *consistent* with the measurements: The discrepancy between the measured and re-synthesized samples (based on the reconstruction) should stay within the (known or estimated) noise level. A valid solution of the signal approximation problem is obtained with a constrained optimization, where the approximation error is minimized subject to the annihilation constraint (see Section III for details). Thanks to the new approach, point sources are recovered accurately in continuous space even in severe noise conditions (see Fig. 2 and Fig. 11). The proposed approach shows a substantial improvement in both accuracy and resolvability of closely located sources over a state of the art method based on discrete ℓ_1 minimization (see Fig. 11).

It turns out that our contribution is much more general than the specific algorithm to solve the point source reconstruction in radioastronomy: In fact, all FRI reconstruction problems can be formulated concisely within the same algorithmic framework. In the proposed approach, we work directly with the given samples, which themselves may not be annihilated right away. With previous approaches [3], [18], [19], a linear transformation had to be applied to the samples first [20], [21]. We not only simplify the problem formulation but also can address cases that were overlooked and considered very challenging to solve with FRI (e.g., the Dirac reconstruction with non-uniform samples). We demonstrate the versatility and robustness with several examples, including the important Dirac reconstructions with *irregular* time/Fourier domain samples (Section IV-A and IV-B), FRI curve reconstructions (Section IV-C) and the recovery of Diracs on the sphere (Section IV-D).

Our goal is to provide a unified algorithmic framework, which is simple yet flexible so as to cope with various FRI sampling problems. To facilitate future research and the applications of the proposed algorithm, a Python implementation of all the examples is made available online.[1]

Before proceeding, we briefly point to literature related to sparse signal recovery. We shall not review various algorithms

[1] The codes are available at http://lcav.epfl.ch/people/hanjie.pan

for the classic spectrum estimation in detail but rather refer readers to standard textbooks (e.g., [2]) for comprehensive reviews. One notable algorithm in spectrum estimation is iterative quadratic maximum likelihood (IQML) [22] and was an inspiration for our approach. In a classical sparse recovery setting, the signal is recovered by minimizing the ℓ_1 norm of the target sparse signal [23]. Recently, it has been extended to continuous domain by minimizing the total variation [24] (or in general an atomic norm [25]) of the continuous sparse signal. It has been shown that the sparse signal can be reconstructed exactly provided that a minimum separation condition is met [24]–[28]. Alternatively, the optimal reconstruction can be formulated as a structured low-rank approximation, where the rank of a data matrix (typically of Toeplitz/Hankel structure) is minimized subject to a data fidelity constraint. This non-convex optimization is then solved either heuristically [29]–[31] or via convex relaxation [32]–[34].

The paper is organized as follows. We briefly review the classic sampling and reconstruction framework for signals with finite rate of innovation in Section II. It serves as the basis for Section III, where we propose a novel algorithmic framework. Both the problem formulation and the implementation are developed. Next we demonstrate the versatility and the robustness of the proposed approach by solving four different FRI problems in Section IV. Further, we illustrate in detail the application to radio interferometry in Section V before we conclude in Section VI.

II. The Classic FRI Reconstruction Framework

From an algorithmic point of view, if a continuous domain signal is or can be transformed into a finite sum of sinusoids, then it is a finite rate of innovation (FRI) signal [3], [18]. The FRI sampling problem then boils down to estimating frequencies of the sinusoids from the given measurements. From this perspective, it coincides with the classic harmonic retrieval problem [35], which is encountered in many applications [4], [36], [37].

One such example is point source reconstruction in radioastronomy as discussed briefly earlier. Similar principles are used for target localization in radar systems as well as in acoustic source localization with microphone arrays.

Another example is X-ray crystallography, where the goal is to determine the atom locations from measured diffraction patterns. The diffraction patterns are the Fourier transform of an autocorrelation function, which is a sum of sinusoids [36]. The distances between atoms are directly related with the frequencies of these sinusoids.

The classic FRI approach (Fig. 3) for solving the frequency retrieval problem consists in sampling the continuous domain signal uniformly. On the one hand, these uniform samples have a one-to-one correspondence with uniformly sampled sinusoids: by applying an inverse mapping (typically an inverse DFT transformation) to the uniform samples, we get an estimate of the sampled sinusoids. On the other hand, it is easy to show that these sampled sinusoids satisfy a set of annihilation equations [3], [18]: There exists a discrete filter (a.k.a. "annihilating" filter), which depends on the unknown frequencies of the sinusoids, such that its convolution with the sampled sinusoids is zero. The annihilating filter coefficients are reconstructed uniquely (up to a scaling factor) by solving a linear system of equations. The frequencies of the sinusoids are obtained from the roots of a polynomial whose coefficients are specified by the annihilating filter. Once we have recovered the frequencies of the sinusoids, the reconstruction of the sinusoids' amplitudes amounts to a simple least square minimization, which reverts to solving a linear system of equations.

Various reconstruction algorithms have been proposed to improve the robustness, notably total least square minimization [3], matrix pencil approach [21], [38] and the Cadzow denoising [18], [39]. In the total least square minimization, one obtains the filter that minimizes the ℓ_2 norm of the annihilation error; the matrix pencil approach takes advantage of the shift-invariant subspace property in the structured data matrix; the Cadzow denoising method exploits the fact that the convolution matrix associated with the annihilating filter has a Toeplitz structure and is rank deficient. The Cadzow method denoises the data by iterating between a thresholding step (to ensure that the matrix is singular) and a projection step (to make the matrix Toeplitz). Recent works [30], [31] generalize such a strategy by formulating the reconstruction problem explicitly as a structured low-rank approximation.

For the rest of the paper, we focus on more general cases where the continuous domain signal may not be sampled uniformly. In this case, we can no longer estimate the sampled sinusoids from the given measurements directly. We will develop a generic approach to solve the FRI reconstruction problem in this non-uniformly sampled case.

III. A Generic FRI Reconstruction Algorithm

In this section, we propose a robust reconstruction algorithm for signals with finite rate of innovation from arbitrary samples. The reconstruction problem is recast as a constrained optimization (Section III-A). We discuss the essential ingredients (Section III-B), the optimization strategy (Section III-C1) as well as the implementation details (Section III-C2) of the proposed algorithm.

A. FRI Reconstruction as a Constrained Optimization

We reformulate the generic FRI reconstruction question as an approximation problem, where we would like to fit an FRI model to the given measurements, or

Given a set of measurements, reconstruct an FRI signal that is consistent with the measurements.

The consistency constraint requires that if we re-synthesize the measurements based on the reconstructed FRI signal parameters, the difference with the given measurements should stay within the noise level (or in general within the allowed approximation error). But how can we ensure that the reconstructed signal satisfies our FRI signal model? One key feature of many FRI signals is that they can be transformed into a sum of sinusoids. The uniform samples of sinusoids are known to be annihilated by a filter with a specific structure that is related to the FRI signal parametrization [3], [18]. Therefore, a signal being

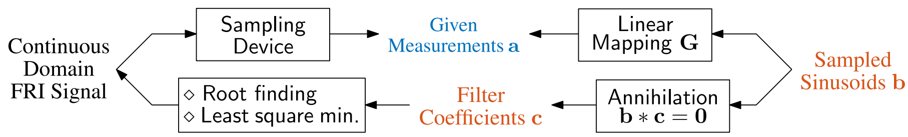


Fig. 4. Overview of the generic FRI reconstruction problem. The continuous domain FRI signal is reconstructed from the given measurements $\mathbf{a}$ by enforcing the annihilation constraint on the unknown sampled sinusoids $\mathbf{b}$, which are related with $\mathbf{a}$ via a linear mapping $\mathbf{G}$.

FRI is algorithmically equivalent to satisfying the annihilation constraint, after some linear transformation (see Fig. 3).

Problem 1 (Constrained Minimization):

$$\begin{aligned} \min_{\mathbf{b},\mathbf{c}\in\mathcal{C}} \quad & \|\mathbf{a}-\mathbf{G}\mathbf{b}\|_2^2 \\ \text{subject to} \quad & \mathbf{b}*\mathbf{c}=\mathbf{0}, \end{aligned} \tag{P1}$$

where (Fig. 4)

- $\mathbf{a}$ is the given set of measurements (sampled non-uniformly in general);
- $\mathbf{b}$ is the vector of uniform samples of the sinusoids to be annihilated. The convolution constraint guarantees that $\mathbf{b}$ is effectively a sum of sinusoids;
- $\mathbf{c}$ is the annihilating filter coefficients, which belongs to a certain feasible set $\mathcal{C}$ (see a precise discussion in the next subsection);

The Generic FRI Reconstruction Problem

Goal: Reconstruct an FRI signal $\mathbf{b}$ that is consistent (up to the noise level ε^2) with the given measurements $\mathbf{a}$:

$$\begin{aligned} \text{find} \quad & \mathbf{b},\mathbf{c}\in\mathcal{C} \\ \text{subject to} \quad & \mathbf{b}*\mathbf{c}=\mathbf{0}, && \text{// b is FRI} \\ & \|\mathbf{a}-\mathbf{G}\mathbf{b}\|_2^2 \le \varepsilon^2 && \text{// consistency} \end{aligned}$$

Key: (i) The linear mapping $\mathbf{G}$ between $\mathbf{a}$ and $\mathbf{b}$;
(ii) Constrained minimization (P1);
(iii) Stopping criterion: compare $\|\mathbf{a}-\mathbf{G}\mathbf{b}\|_2^2$ with the noise level ε^2 (Section III-B3);
(iv) Random initialization (Section III-B4).

- $\mathbf{G}$ models the linear mapping[2] between the measurements $\mathbf{a}$ and the uniform sinusoid samples $\mathbf{b}$. In general, in the presence of noise, we need to increase the number of measurements.

Superficially, the data term $\|\mathbf{a}-\mathbf{G}\mathbf{b}\|_2^2$ looks similar to the one used in compressed sensing (CS) based approaches. Yet, it arises from a completely different approach to resolution (see Fig. 5):

- On the one hand, in CS-based sparse recovery the final recoverable resolution is directly related to the step-size of the uniform grid that supports the samples $\mathbf{b}$ (see e.g., [25] Remark 1.2): Perfect reconstruction is obtained for Diracs that are separated by a minimum distance that is proportional to the grid step-size. The matrix $\mathbf{G}$ then encodes the "non-uniform" down-sampling that provides the known measurements $\mathbf{a}$; for that reason, $\mathbf{G}$ is necessarily a fat matrix.

[2]For simplicity, we will assume that the linear transformation $\mathbf{G}$ has full column rank throughout the paper.

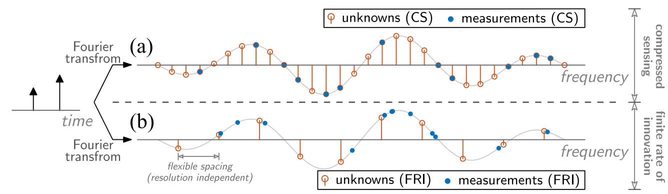


Fig. 5. Problem setup differences between compressed sensing and finite rate of innovation (FRI). (a) In compressed sensing based approaches, the recoverable resolution (commonly known as the minimum separation requirement in compressed sensing) is related to the step-size of the underlying grid. (b) In FRI, the reconstruction accuracy is independent of the spacing between the uniform samples (see texts after (P1) in Section III-A).

- On the other hand, in FRI the final resolution is only related to the noise (or model mismatch) level. This error is typically given by the Cramér-Rao lower bound, which is reached by FRI-based reconstruction algorithms experimentally [18], [31]. The samples $\mathbf{b}$ have to be taken on a uniform grid in order for the annihilation equation to be satisfied, but the step-size of the grid is flexible and is unrelated to the resolution of the method. In that context, the matrix $\mathbf{G}$ models a different type of linear relation[3] between arbitrary uniform samples $\mathbf{b}$ and arbitrary non-uniform known measurements $\mathbf{a}$ than in compressed sensing; for that reason, $\mathbf{G}$ can — and will, in this paper — be a tall matrix.

Note that (P1) is non-convex with respect to $(\mathbf{b},\mathbf{c})$ jointly. Despite various attempts to solve similar constrained optimizations [22], [40], a reliable algorithm for finding the optimall solution of (P1) has yet to be discovered. But do we actually need to obtain the optimal solution of (P1)? In many cases, we know (or can estimate accurately) the noise level present in the given measurements. Hence, we may use this additional information in validating whether a solution is feasible or not: we claim that any solution $\mathbf{b}$ is valid as long as it satisfies both the annihilation and consistency constraint for the given measurements (up to the noise level ε^2), which we formalize as Problem (P2).

Problem 2 (Constrained Approximation with Noise Level):

$$\begin{aligned} \text{find} \quad & \mathbf{b},\mathbf{c}\in\mathcal{C} \\ \text{subject to} \quad & \mathbf{b}*\mathbf{c}=\mathbf{0}, \\ & \|\mathbf{a}-\mathbf{G}\mathbf{b}\|_2^2 \le \varepsilon^2. \end{aligned} \tag{P2}$$

[3]One would expected that a $\mathbf{G}$ matrix that maps too few uniform samples to the non-uniformly sampled measurements, would lead to worse reconstruction. However, this imbalance seems not to be critical experimentally (see examples in Section IV-B and Section V-C).

One way to find a valid solution of (P2) is to resort to the constrained minimization (P1). However, we should keep in mind that it is not the optimal solution of (P1) that we seek but rather *a* valid solution that satisfies the constraints in (P2). Indeed, for any non-zero ε, (P2) has infinitely many solutions, among which one is the optimal solution of (P1). From an approximation point of view, all these solutions are valid because, for each of them, the reconstructed parametric signal explains the given measurements up to the noise level[4]. This subtle difference is important, since it allows to develop a reliable algorithm in the rest of this section.

B. Essential Ingredients

Before we present the algorithm that solves (P2), we want to highlight five key elements of the proposed constrained formulation.

1) "Bilinearity" of the Annihilation Constraint: The annihilation equation is nonlinear with respect to the $(\mathbf{b}, \mathbf{c})$-pair. However, if we fix one variable (e.g., $\mathbf{c}$), then the annihilation constraint reduces to a linear constraint with respect to the other variable (e.g., $\mathbf{b}$ here). Motivated by the bilinearity of the annihilation constraint, we define a right dual operator.

Definition 1: Denote the annihilation constraint in (P2) as $\mathbf{T}(\mathbf{b})\mathbf{c} = \mathbf{0}$, then the right dual of $\mathbf{T}(\cdot)$ is an operator $\mathbf{R}(\cdot)$ such that $\mathbf{R}(\mathbf{c})\mathbf{b} = \mathbf{T}(\mathbf{b})\mathbf{c}$ for all $\mathbf{b}, \mathbf{c}$.

In many FRI reconstruction problems, the annihilation equations are convolutions, which implies that $\mathbf{T}(\mathbf{b})$ and $\mathbf{R}(\mathbf{c})$ are Toeplitz-structured convolution matrices. We can justify the right dual definition from the commutativity of the convolution: $\mathbf{b} * \mathbf{c} = \mathbf{c} * \mathbf{b}$.

Thanks to the bilinearity of the annihilation constraint, it can be shown that the bivariate optimization (P1) is equivalent to a constrained optimization with respect to $\mathbf{c}$ alone. This equivalent formulation provides an iterative strategy for finding a valid solution of (P2) (see Section III-C).

2) Forward Mapping: Unlike most annihilating filter based reconstruction algorithms [3], [18], [19], we deal with the measurements directly *without* pre-processing the given measurements first (e.g., a truncated DFT transformation). The linear mapping $\mathbf{G}$, which links the measurements to a sequence that can be annihilated, is integrated in the reconstruction algorithm.

Thanks to the new approach, we are not only able to extend the FRI framework to cases with irregularly measured samples but also streamline otherwise rather complicated FRI reconstructions (see Section III-C2).

3) Stopping Criteria: Because of the non-convexity of (P1), we should not expect the algorithm to always find the global optimal solution in general. In fact, it is *not* the optimal solution of (P1) that we should seek but rather *a* solution that (i) satisfies the annihilation constraint and (ii) has a fitting error $\|\mathbf{a} - \mathbf{G}\mathbf{b}\|_2^2$ below the noise level [41]. After all, our goal is to use the constrained minimization *as a tool* to find a valid solution of (P2) — any solution that meets the two criteria is a valid one for the FRI reconstruction.

The criteria are constructive: If we can guarantee that the reconstructed signal always satisfies the annihilation constraint, which is the case with the proposed algorithm (see details in Section III-C1), then we only need to check the fitting error in order to decide whether to terminate the algorithm or not.

4) Random Initialization: Because of the non-convexity of (P1), a commonly used strategy is to initialize the algorithm with a "good" candidate solution, which is hopefully close to the ground truth, e.g., the total least square reconstruction. In our proposed algorithm, we choose to initialize the annihilating filter coefficients $\mathbf{c}$ with a random vector instead.

The randomness of the initialization actually gives the algorithm the flexibility to have fresh restarts to increase the likeliness of meeting the stopping criteria — if the algorithm fails to find a solution that meets the aforementioned stopping criteria (Section III-B3), we can always reinitialize the algorithm with a different annihilating filter. The random initialization strategy has been shown to result in a valid solution within a finite number of initializations (typically less than 15) in extensive tests [42].

5) Feasible Set $\mathcal{C}$ of the Annihilation Filter Coefficients: Observe that (P1) is scale invariant with respect to the annihilating filter coefficients $\mathbf{c}$. Without any normalization, we have a trivial solution $\mathbf{c} \equiv \mathbf{0}$. Experimentally, we have observed that the most robust performance is achieved by restricting[5] $\mathbf{c}_0^{\mathrm{H}}\mathbf{c} = 1$, where $\mathbf{c}_0$ is the random initialization for the algorithm (see [19] as well).

C. An Iterative Algorithm to Solve (P2)

1) Inspiration from (P1): For a given $\mathbf{c}$, (P1) is a constrained quadratic minimization with respect to $\mathbf{b}$. By substituting the solution $\mathbf{b}$ (in a function of $\mathbf{c}$) to (P1), we end up with an optimization over the annihilating filter coefficients $\mathbf{c}$ alone. Denote the annihilation constraint in (P1) as a matrix vector product: $\mathbf{T}(\mathbf{b})\mathbf{c} = \mathbf{0}$. It can be shown that (P1) is equivalent to (see Appendix A):

$$\begin{aligned} &\min_{\mathbf{c}} \ \mathbf{c}^{\mathrm{H}}\mathbf{T}^{\mathrm{H}}(\boldsymbol{\beta})(\mathbf{R}(\mathbf{c})(\mathbf{G}^{\mathrm{H}}\mathbf{G})^{-1}\mathbf{R}^{\mathrm{H}}(\mathbf{c}))^{-1}\mathbf{T}(\boldsymbol{\beta})\mathbf{c} \\ &\text{subject to } \ \mathbf{c}_0^{\mathrm{H}}\mathbf{c} = 1, \end{aligned} \tag{1}$$

where $\boldsymbol{\beta} = (\mathbf{G}^{\mathrm{H}}\mathbf{G})^{-1}\mathbf{G}^{\mathrm{H}}\mathbf{a}$, and $\mathbf{R}(\cdot)$ is the right dual of $\mathbf{T}(\cdot)$ defined in Definition 1.

Then, the reconstructed FRI signal can be expressed in a function of $\mathbf{c}$ as:

$$\mathbf{b} = \boldsymbol{\beta} - (\mathbf{G}^{\mathrm{H}}\mathbf{G})^{-1}\mathbf{R}^{\mathrm{H}}(\mathbf{c})(\mathbf{R}(\mathbf{c})(\mathbf{G}^{\mathrm{H}}\mathbf{G})^{-1}\mathbf{R}^{\mathrm{H}}(\mathbf{c}))^{-1}\mathbf{R}(\mathbf{c})\boldsymbol{\beta}. \tag{2}$$

[4]Notice that ε controls the approximation error with the ground truth: From the triangle inequality, the difference (2-norm) between the re-synthesized measurements with the ground truth signal and any one of the valid solution of (P2) is at most 2ε. This corresponds to a maximum $2\varepsilon/\sqrt{\lambda_{\min}(\mathbf{G}^{\mathrm{H}}\mathbf{G})}$ difference between the reconstructed and the ground truth $\mathbf{b}$, where $\lambda_{\min}(\cdot)$ is the smallest eigenvalue of a matrix.

[5]Other (natural, but less successful) normalization strategies [3], [18] include a quadratic constraint $\|\mathbf{c}\|_2^2 = 1$; or a linear constraint on one component of $\mathbf{c}$, e.g., $\mathbf{e}_0^{\mathrm{H}}\mathbf{c} = 1$, where $\mathbf{e}_0 = [1, 0, \ldots, 0]^{\mathrm{T}}$.

In general, it is very difficult to solve (1). But we can draw inspiration from this equivalent formulation and devise an iterative algorithm for finding a valid solution of (P2).

More specifically, our strategy amounts to minimizing the objective function in (1) iteratively: At each iteration, we build the matrix $(\mathbf{R}(\mathbf{c})(\mathbf{G}^{\mathrm{H}}\mathbf{G})^{-1}\mathbf{R}^{\mathrm{H}}(\mathbf{c}))^{-1}$ with $\mathbf{c} = \mathbf{c}_{n-1}$, the filter coefficients from the *previous* iteration. The updated $\mathbf{c}$ is then obtained by solving a quadratic minimization:

$$\begin{aligned} &\min_{\mathbf{c}} \ \mathbf{c}^{\mathrm{H}}\mathbf{T}^{\mathrm{H}}(\boldsymbol{\beta})(\mathbf{R}(\mathbf{c}_{n-1})(\mathbf{G}^{\mathrm{H}}\mathbf{G})^{-1}\mathbf{R}^{\mathrm{H}}(\mathbf{c}_{n-1}))^{-1}\mathbf{T}(\boldsymbol{\beta})\mathbf{c} \\ &\text{subject to } \ \mathbf{c}_0^{\mathrm{H}}\mathbf{c} = 1, \end{aligned} \tag{3}$$

which has a simple closed form solution. For the consideration of numerical stability, we are not going to implement the solution of (3) directly but revert to solving an equivalent linear system of equations instead (see Section III-C2). The uniform sinusoidal samples $\mathbf{b}_n$ is updated based on (2) with the reconstructed $\mathbf{c}_n$ at the *current* iteration.

Note that since (2) is obtained by solving (P1) for a given $\mathbf{c}$, by construction, $\mathbf{b}_n$ obtained this way will always satisfy the annihilation constraint: $\mathbf{R}(\mathbf{c}_n)\mathbf{b}_n = \mathbf{0}$. Hence, we only need to compute the approximation error $\|\mathbf{a} - \mathbf{G}\mathbf{b}_n\|_2^2$ and check whether it is below the noise level in order to terminate the iteration (see comments in Section III-B3).

The proposed approach may be judged similar to the *iterative quadratic maximum likelihood* [22] method in the spectrum estimation community, with the important difference: we only use the constrained optimization as a way to find a valid FRI reconstruction (P2). Hence, it is not the convergent solution of (P1) that matters (as in [22]) but rather any solution that meets the stopping criteria in Section III-B3. The randomness in the initialization and the linear constraint give the algorithm more flexibility and has been shown [42] to achieve more robust reconstruction results for the FRI problems.

2) Efficient Implementation: A direct implementation of (3) for the update of the annihilating filter coefficients involves several nested matrix inverses and would not only be inefficient (compared with solving linear system of equations) but also numerically unstable even with double-precision accuracy.

We can obtain the solution of (3) by solving a larger (compared with the dimension of $\mathbf{c}$) linear system of equations with a simple trick: We introduce an auxiliary variable as a substitute of a matrix inverse applied to the input vector. Extra equations, which only involve multiplication of the matrix (instead of its inverse), are subsequently added to ensure that the resultant problem is equivalent to the original one.

Proposition 1: The solution of (3) is given by solving a linear system of equations

$$\begin{bmatrix} \mathbf{0} & \mathbf{T}^{\mathrm{H}}(\boldsymbol{\beta}) & \mathbf{0} & \mathbf{c}_0 \\ \mathbf{T}(\boldsymbol{\beta}) & \mathbf{0} & -\mathbf{R}(\mathbf{c}_{n-1}) & \mathbf{0} \\ \mathbf{0} & -\mathbf{R}^{\mathrm{H}}(\mathbf{c}_{n-1}) & \mathbf{G}^{\mathrm{H}}\mathbf{G} & \mathbf{0} \\ \mathbf{c}_0^{\mathrm{H}} & \mathbf{0} & \mathbf{0} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{c}_n \\ \boldsymbol{\ell} \\ \mathbf{v} \\ \lambda \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ 1 \end{bmatrix}, \tag{4}$$

where $\boldsymbol{\ell}$, $\mathbf{v}$, and λ are newly introduced auxiliary variables. Similarly, the reconstructed FRI signal $\mathbf{b}_n$ is updated as the solution of

$$\begin{bmatrix} \mathbf{G}^{\mathrm{H}}\mathbf{G} & \mathbf{R}^{\mathrm{H}}(\mathbf{c}_n) \\ \mathbf{R}(\mathbf{c}_n) & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{b}_n \\ \boldsymbol{\ell} \end{bmatrix} = \begin{bmatrix} \mathbf{G}^{\mathrm{H}}\mathbf{a} \\ \mathbf{0} \end{bmatrix}, \tag{5}$$

where $\boldsymbol{\ell}$ is the auxiliary variable.

Proof: See Appendix B. ■

Algorithm 1: Robust FRI Signal Reconstruction.

Input : Measurements $\mathbf{a}$ of the FRI signal, transformation matrix $\mathbf{G}$, noise level ε^2
Output: Uniform sinusoid samples $\mathbf{b}$, annihilating filter coefficients $\mathbf{c}$

```
for loop ← 1 to max. initializations do
    Initialize c with a random vector c_0;
    for n ← 1 to max. iterations do
        Build the augmented matrix with c_{n-1} and update c_n by solving (4);
        Build the augmented matrix with c_n and update b_n by solving (5);
        if ||a − G b_n||_2^2 ≤ ε^2 then
            Terminate both loops;
        end
    end
end
b ← b_n, c ← c_n.
```

Instead of calculating several matrix inverses, we only need to solve a larger linear system of equations and extract the corresponding components of the solution in order to update $\mathbf{c}$ and $\mathbf{b}$ from iteration to iteration. For a vector $\mathbf{b}$ of size L, the computational complexity at each iteration is $\mathcal{O}(L^3)$ (see [43] Chapter 3).

In the actual implementation, we randomly initialize the algorithm with a maximum number of iterations, e.g., 50. At each iteration, we compute the approximation error $\|\mathbf{a} - \mathbf{G}\mathbf{b}_n\|_2^2$ with the current reconstruction $\mathbf{b}_n$ and compare it with the noise level. If the error is below the noise level, then the iteration is terminated. In the case where the algorithm fails to find such a solution after the maximum number of iterations is reached, we reinitialize the algorithm with a different random vector. We summarize the proposed algorithm in Algorithm 1. In [42], this strategy (50 inner iterations and 15 random initializations) is shown to succeed in 99.9% cases. Alternatively, we can always run the algorithm with a (fixed) maximum number of random initializations and return the reconstructed $(\mathbf{b}, \mathbf{c})$-pair that has the smallest fitting error. This strategy is useful for cases where we do not know (or do not have a good estimate of) the noise level a priori, albeit less efficient than the approach used in Algorithm 1: With a given (or estimated) noise level, Algorithm 1 usually terminates much earlier before reaching the maximum number of random initializations.

IV. Four Algorithmic Examples for Various Types of FRI Signals

In the previous section, we presented a generic reconstruction algorithm for signals with finite rate of innovation. To demonstrate the versatility of the proposed algorithm, we showcase several FRI reconstruction problems, including the Dirac reconstruction with *non-uniform* time/Fourier domain samples (Section IV-A and IV-B), FRI curve reconstruction (Section IV-C) and the reconstruction of Diracs on the sphere (Section IV-D).

A common misconception on annihilating-filter based FRI reconstruction algorithms is that they can only deal with *uniformly* taken samples, which have one to one correspondence with uniform samples of a sum of sinusoids, e.g., the DFT coefficients in the reconstruction of the Dirac stream (6). Such an artificial limitation is waived as soon as we are able to identify the forward mapping (or an approximation of it) that links the sequence to be annihilated and the given measurements, i.e., $\mathbf{G}$ in (P2). We show that we can either find the exact mapping (Section IV-A) or approximate it by interpolation (Section IV-B). The new formulation is flexible in the choice of the objective function or a proper constraint: instead of being restricted by the reconstruction algorithm, e.g., the Cadzow denoising method, we can use a different formulation, which is simpler and gives more robust reconstruction results (Section IV-C and Section IV-D).

A. Stream of Diracs With Irregular Time Domain Samples

Consider a τ-periodic stream of Diracs:

$$x(t) = \sum_{k' \in \mathbb{Z}} \sum_{k=1}^{K} \alpha_k \delta(t - t_k - k'\tau), \tag{6}$$

where α_k and t_k are unknown signal parameters. The goal is to reconstruct these parameters from a set of ideally low-pass filtered samples that are taken at *irregular* (but known) time instances $t'_1, \ldots, t'_L$:

$$y_\ell = \sum_{k=1}^{K} \alpha_k \varphi(t'_\ell - t_k) \quad \text{for } \ell = 1, \ldots, L. \tag{7}$$

Here φ is the Dirichlet kernel $\varphi(t) \stackrel{\text{def}}{=} \frac{\sin(\pi B t)}{B\tau \sin(\pi t/\tau)}$ and B is the bandwidth of the ideal lowpass filter.

- *Uniform Samples of Sinusoids*: Observe that $x(t)$ is a linear combination of the same function with different shifts — if we transform the signal to the Fourier domain, the spectrum is a sum of sinusoids:

$$\hat{x}_m = \frac{1}{\tau} \sum_{k=1}^{K} \alpha_k e^{-j\frac{2\pi m}{\tau} t_k},$$

  where $\hat{x}_m$ is the Fourier series coefficients of the *periodic* signal $x(t)$. Since $\hat{x}_m$ is a uniformly sampled sum of sinusoids, we know that it can be annihilated by a discrete filter.
- *Relation with the Given Measurements*: It is easy to show that the given measurements are linearly related with $\hat{x}_m$ via a truncated inverse DFT transformation:

$$y_\ell = \frac{1}{B} \sum_{|m| \leq \lfloor \frac{\tau B}{2} \rfloor} \hat{x}_m e^{j\frac{2\pi m}{\tau} t'_\ell}. \tag{8}$$

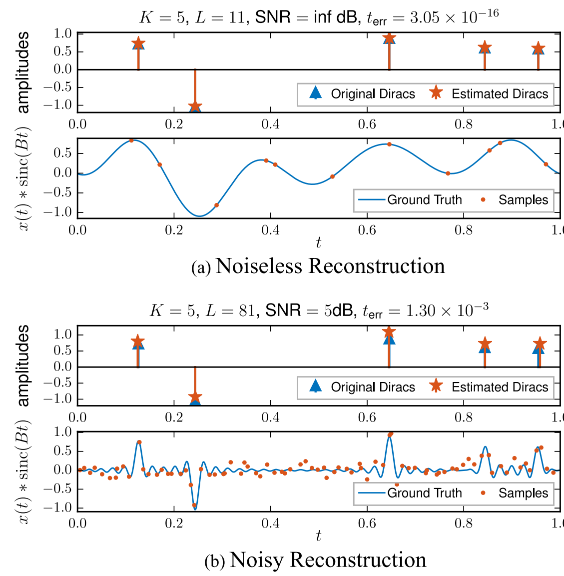


Fig. 6. Reconstruction of a stream of Diracs (6) from ideally low-pass filtered samples taken at irregular time instances (8). (a) Exact reconstruction in the noiseless case (filter bandwidth $B = 11$, number of samples $L = 11$). (b) Robust reconstruction in the noisy case (SNR $= 5$ dB, filter bandwidth $B = 81$, number of samples $L = 81$, average reconstruction error for t_k: 1.30×10^{-3}).

In terms of the reconstruction algorithm, we can rearrange the samples y_ℓ and $\hat{x}_m$ as column vectors $\mathbf{a}$ and $\mathbf{b}$, respectively; the linear mapping (8) between $\mathbf{b}$ and $\mathbf{a}$ is denoted by the matrix $\mathbf{G}$. Then the "denoised" FRI signal as well as the associated annihilating filter $\mathbf{c} = [c_0, \ldots, c_K]^{\mathrm{T}}$ are given by Algorithm 1. The Dirac locations are reconstructed by taking the roots of the polynomial whose coefficients are specified by $\mathbf{c}$; while the amplitudes are reconstructed with least square minimization [3], [18].

We summarize the reconstruction results in Fig. 6 for both the noiseless and noisy cases, where Gaussian white noise is added to the lowpass filtered samples. Note that the irregular sampling scheme does not change the minimum number of samples required in order to recover the original signal (Fig. 6(a)): With at least $2K + 1$ samples, the exact reconstruction (up to numerical accuracies) is obtained. In the presence of noise, we need to over-sample the signal. The proposed algorithm is robust enough to give a reliable reconstruction in the presence of severe noise (SNR $= 5$ dB).

B. Weighted Sum of Diracs With Irregular Fourier Domain Samples

In this example, we consider a slightly different Dirac reconstruction problem than that in the previous section. In particular, consider a sparse signal that consists of K weighted Diracs:

$$x(t) = \sum_{k=1}^{K} \alpha_k \delta(t - t_k), \tag{9}$$

with limited time support between $-\tau/2$ to $\tau/2$, i.e., $|t_k| \leq \tau/2$. Instead of taking the time domain samples (as in the previous example), the Fourier transform

$$X(\omega) = \sum_{k=1}^{K} \alpha_k e^{-j\omega t_k}, \tag{10}$$

is measured at some frequencies ω_ℓ for $\ell = 1, \ldots, L$.

The question at hand is: Can we recover the original signal (9) from non-uniform Fourier samples $X(\omega_\ell)$? In many applications, e.g., magnetic resonance imaging [44], holography [45], crystallography [36] and radio interferometry [46], direct Fourier domain measurements are available thus making the sparse reconstruction problem of particular interest.

- *Uniform Samples of Sinusoids*: Since the Fourier transform (10) is a weighted sum of sinusoids, the uniformly sampled Fourier transform on a grid: $X(2\pi m/\tau)$ for $m \in \mathbb{Z}$, can be annihilated.
- *Relation with the Given Measurements*: In general, the given Fourier measurements are taken non-uniformly. Hence, we cannot apply the annihilating filter method directly. However, not everything is lost: We may interpolate the Fourier transform over a finite interval, e.g., $\omega \in [-M\pi, M\pi]$:

$$X(\omega) \approx \sum_{|m| \leq \lfloor \frac{M\tau}{2} \rfloor} X\left(\frac{2\pi m}{\tau}\right) \psi\left(\frac{\omega}{2\pi/\tau} - m\right), \tag{11}$$

  where $\psi(\cdot)$ is a certain interpolation kernel, e.g., a spline function.

By evaluating (11) at ω_ℓ, we establish a linear mapping (i.e., $\mathbf{G}$ in the reconstruction algorithm) between the given Fourier measurements $\mathbf{a}: X(\omega_\ell)$ and the unknown sampled sinusoids $\mathbf{b}: X(2\pi m/\tau)$. Provided that we have sufficiently many measurements, i.e., $L \geq M\tau$, then we can reconstruct (9) with Algorithm 1 (see [30] for a similar strategy in spectral estimation).

We may justify such an approach by considering a specific case, where the Fourier transform $X(\omega)$ is periodic with period $2\pi M$ for some M such that $M\tau$ is an odd number. It is proved that we can represent $X(\omega)$ exactly by interpolating with the Dirichlet kernel $\psi(\omega) = \frac{\sin(\pi\omega)}{M\tau \sin(\frac{\pi\omega}{M\tau})}$ in this case (see Appendix C and Fig. 7).

Note that rather than enforcing the interpolation equation (11) as a hard constraint on the reconstructed signal, we only use it to derive a data-fidelity metric in (P2) that measures the approximation quality. The tradeoff is that we can no longer reconstruct the signal exactly in general — We do not have the actual mapping $\mathbf{G}_0$ (which depends on the unknown signal parameters α_k and t_k) but only its approximation from the interpolation (11): $\mathbf{G} = \mathbf{G}_0 + \mathbf{G}_\varepsilon$. Consequently, we will have a model mismatch even in the noiseless cases[6] (Fig. 8(a)). As we have mentioned in Section III-C2, one possible way to circumvent the difficulty in choosing ε in (P2) is to run the algorithm with fixed random

[6]Equivalently, we can view the noiseless measurements $\mathbf{a} = \mathbf{G}_0\mathbf{b}$ as being "noisy" with respect to $\mathbf{G}$, which is used in the data-fidelity constraint in (P2): $\mathbf{a} = \mathbf{G}\mathbf{b} + \text{noise}$, with noise $= -\mathbf{G}_\varepsilon\mathbf{b}$.

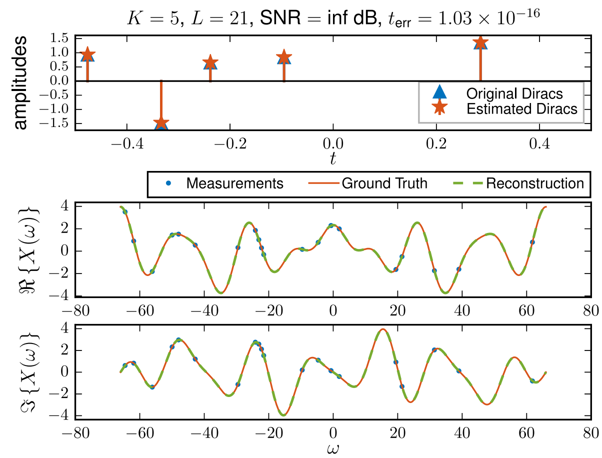
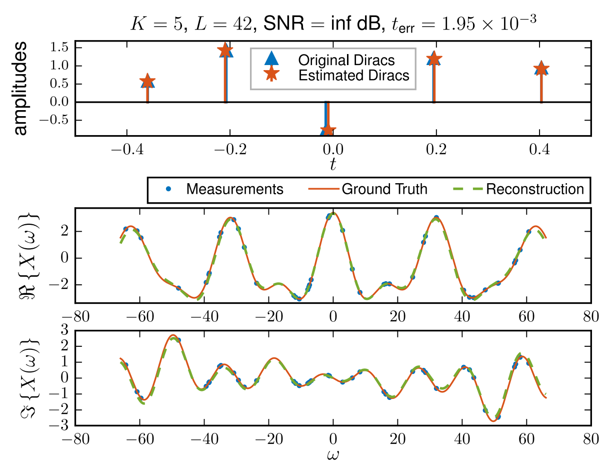
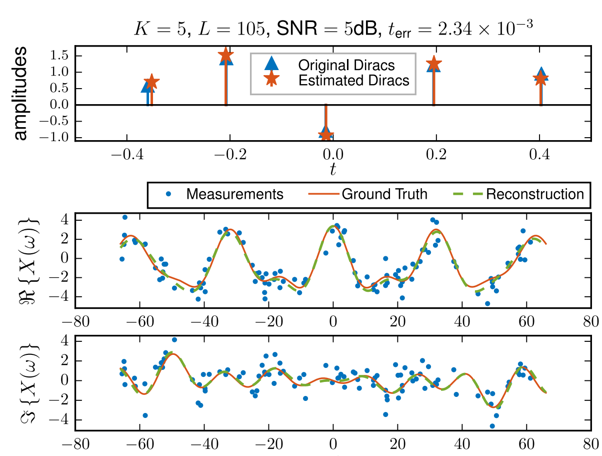


Fig. 7. Exact reconstruction of weighted Diracs (9) with periodic spectrum from irregularly sampled Fourier measurements. Dirichlet interpolation kernel is used to relate the uniform samples of sinusoids to the measurements (period of the spectrum: $2\pi \times 21$; number of samples $L = 21$).



(a) With Noiseless Samples



(b) With Noisy Samples

Fig. 8. Reconstruction of weighted Diracs (9) from non-uniform Fourier samples (10). The FRI framework makes use of the *piecewise linear interpolation* (11) with 21 uniform knots. (a) Reconstruction with noiseless Fourier domain samples (number of samples $L = 42$, average reconstruction error for t_k: 1.95×10^{-3}). (b) Robust reconstruction with noisy Fourier measurements (number of samples $L = 105$, SNR = 5 dB, average reconstruction error for t_k: 2.34×10^{-3}).

initializations. The solution that gives the minimum fitting error is taken as the reconstruction.

We demonstrate the effectiveness of the interpolation strategy (11) in Fig. 8, where the interpolation kernel is the first order B-spline:

$$\psi(\omega) = \begin{cases} 1-\omega & \text{if } \omega \in [0,1), \\ 1+\omega & \text{if } \omega \in [-1,0), \\ 0 & \text{otherwise.} \end{cases}$$

We have chosen 21 interpolation knots located uniformly on the interval $[-21\pi, 21\pi]$, where the Fourier transform $X(\omega)$ is approximated. Complex-valued Gaussian white noise is added to the Fourier samples in the noisy case (SNR = 5 dB). Even with such a coarse approximation, we still obtain robust and accurate reconstruction of Diracs in the presence of noise (Fig. 8).

C. FRI Curves

As we mentioned in the introduction, the Cadzow denoising algorithm [18] tries to find a structured matrix (typ. Toeplitz/block-Toeplitz) that satisfies the rank constraint while being as close as possible to the noisy data matrix. With the Cadzow denoising method, we are restricted to work directly with a sequence that can be annihilated (so that we can enforce the rank constraint on the matrix). In comparison, we have more freedom with the proposed algorithm in defining what is the unknown data $\mathbf{b}$ other than the obvious choice as the sampled sinusoids. We demonstrate this flexibility with an example of curves with finite rate of innovation [16].

Consider an interior indicator image associated with a curve:

$$I_C(x,y) = \begin{cases} 1 & \text{if } (x,y) \in \mathring{C} \\ 0 & \text{otherwise,} \end{cases}$$

where $\mathring{C}$ denotes the interior of the curve C. Our goal is to reconstruct the curve locations in the *continuous* domain from a set of ideally lowpass filtered samples of the binary image $I_C(x,y)$.

- *Uniform Samples of Sinusoids*: We may treat the derivative of the indicator image as an infinite sum (in fact a line integration) of Diracs along the curve. In the Fourier domain, the Wirtinger derivative (i.e., $\partial \overset{\text{def}}{=} \frac{\partial}{\partial x} + j\frac{\partial}{\partial y}$) is therefore a sum of sinusoids. Consequently, we know that the Fourier transform of the derivative image on a uniform grid
$$\widehat{\partial I}_{k,l} = \left(\frac{2\pi k}{\tau_1} + j\frac{2\pi l}{\tau_2}\right) \hat{I}_C\left(\frac{2\pi k}{\tau_1}, \frac{2\pi l}{\tau_2}\right) \qquad (12)$$
satisfies the annihilation equations: $c_{k,l} * \widehat{\partial I}_{k,l} = 0$. The curve locations are specified by the roots of a 2D polynomial with coefficients $c_{k,l}$.
- *Relation with the Given Measurements*: Similar to the 1D case in Section IV-A, the Fourier transform of the indicator image on a uniform grid is related with the ideally lowpass filtered samples[7] via a truncated inverse DFT. Combined with (12), we have a linear mapping from the unknown sampled sinusoids $\widehat{\partial I}_{k,l}$ to the measured spatial domain samples.

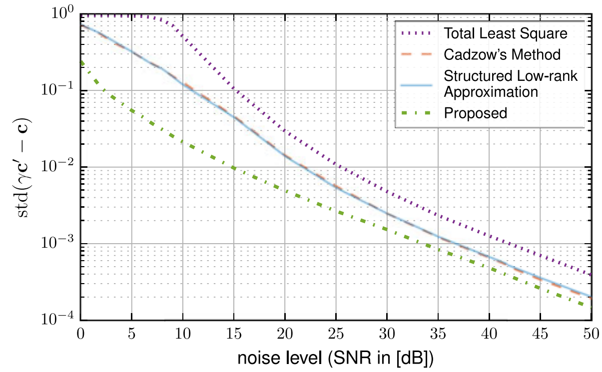


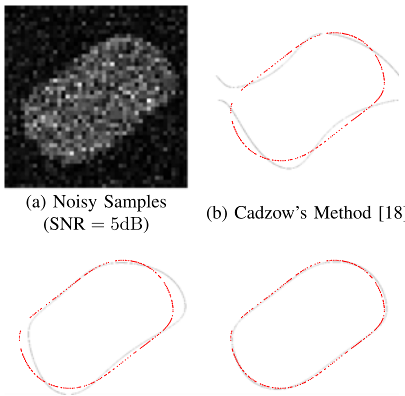
Fig. 9. FRI curve coefficient error against different noise levels (curve coefficients $c_{k,l}$ size: 3×3, sample size: 45×45, bandwidth $B_1 = B_2 = 25$, periods $\tau_1 = \tau_2 = 1$). The results are averaged over 500 independent noise realizations.

In our original approach [16], we first obtain $\hat{I}_C(\frac{2\pi k}{\tau_1}, \frac{2\pi l}{\tau_2})$ by applying a truncated DFT transformation to the given samples. Then we apply Cadzow's method to denoise $\widehat{\partial I}_{k,l}$, since it is $\widehat{\partial I}_{k,l}$ that satisfies the annihilation. Unfortunately, doing so inevitably amplifies the high frequency noise components, which explains the relatively limited performance of Cadzow's method for FRI curve reconstructions (Fig. 9).

In our new approach, we do not have such a restriction anymore: in (P2), we can choose $\mathbf{a}$ directly as the pixel values $I_{m,n}$, the unknown $\mathbf{b}$ as the Fourier transform $\hat{I}_C(\frac{2\pi k}{\tau_1}, \frac{2\pi l}{\tau_2})$, and the linear mapping $\mathbf{G}$ as the truncated inverse DFT transformation. The right dual matrix $\mathbf{R}(\cdot)$ in this case is no longer the convolution matrix associated with the filter $c_{k,l}$ alone — we should right multiply the convolution matrix by a diagonal matrix whose entries are specified by the corresponding frequencies $(\frac{2\pi k}{\tau_1}, \frac{2\pi l}{\tau_2})$ in (12).

We summarize the reconstruction results obtained with total least square minimization [3], [18], Cadzow's method [18], structured low-rank approximation [31] and the proposed method in Fig. 9, where different levels of Gaussian white noise is added to the ideally lowpass filtered samples. Since the annihilating filter coefficients are invariant with respect to any non-zero scaling, we measure the reconstruction error with a scale-invariant standard deviation of the error between the ground truth $\mathbf{c}$ and the reconstructed coefficients $\mathbf{c}'$: $\mathrm{std}(\gamma\mathbf{c}' - \mathbf{c})$. Here the scalar γ is chosen in such a way that $\|\gamma\mathbf{c}' - \mathbf{c}\|_2^2$ is minimized. An example at noise level SNR = 5 dB is also included for visual comparisons (Fig. 10). The proposed algorithm is more robust even in such a severe noise condition.

[7]Note that with the same argument as in Section IV-A, we can deal with ideally lowpass filtered samples that are taken non-uniformly.



(a) Noisy Samples (SNR = 5dB) (b) Cadzow's Method [18]

(c) Structured Low-rank Approximation [31] (d) Proposed Approach

Fig. 10. Visual comparisons of the reconstructed curves with Cadzow's method, structured low-rank approximation [31] and the proposed approach (noise level: 5 dB, curve coefficients $c_{k,l}$ size: 3×3, sample size: 45×45, periods $\tau_1 = \tau_2 = 1$). The solid black line is the reconstructed curve; while the dotted red line is the ground truth.

D. Diracs on the Sphere

In most cases, the given measurements of an FRI signal cannot be annihilated directly. With previous approaches, we had to apply an inverse transformation in order to obtain a sequence that can be annihilated. However, such an inverse transformation is not always easy to identify. Additionally, it makes the reconstruction problem unnecessarily complicated.

We demonstrate how the proposed algorithmic framework helps simplify the formulation of FRI problems. Consider an example of Diracs defined on the 2-sphere $\mathbb{S}^2 \stackrel{\text{def}}{=} \{\boldsymbol{r} \in \mathbb{R}^3 | \boldsymbol{r}^{\mathrm{T}}\boldsymbol{r} = 1\}$:

$$I(\theta, \varphi) = \sum_{k=1}^{K} \alpha_k \delta(\cos\theta - \cos\theta_k)\delta(\varphi - \varphi_k),$$

for $0 \le \theta, \theta_k < \pi$ and $0 \le \varphi, \varphi_k < 2\pi$. Here θ_k and φ_k are the unknown angles of the colatitude and azimuth of the k-th Dirac and α_k is its amplitude.

Similar to the planar case, if we apply a Fourier-like transformation to Diracs, we will get a sum of sinusoids. More specifically, the Fourier-like transformation on $\mathbb{S}^2$ involves spherical harmonics. It can be shown (see Appendix D) that the coefficients of the spherical harmonic decomposition is

$$\hat{I}_{l,m} = N_{l,m} \sum_{n=0}^{l-|m|} p_{n,|m|} \sum_{k=1}^{K} \alpha_k (\cos\theta_k)^n (\sin\theta_k)^{|m|} \mathrm{e}^{-jm\varphi_k}, \tag{13}$$

with $|m| \le l$ and $0 \le l \le L_0$, for some *fixed* coefficients $p_{n,|m|}$ that can be precomputed. Here $N_{l,m}$ is a normalization factor associated with the spherical harmonics of degree l and order m.

The FRI reconstruction problem that we want to solve is as follows: Reconstruct $(\alpha_k, \theta_k, \varphi_k)$ for $k = 1, \ldots, K$ from a given set of spherical harmonic coefficients (13). In an actual setup, the measurements are spatial domain samples (on $\mathbb{S}^2$), which have a linear relationship [15], [17] with the spherical harmonic coefficients (13). With the proposed algorithm, we can reconstruct the signal from the spatial domain samples directly. But the complexity is beyond the scope of this section and hence is omitted.

- *Uniform Samples of Sinusoids*: Complicated as it may appear, (13) is a linear combination of uniformly sampled sinusoids: $b_{n,m} \stackrel{\text{def}}{=} \sum_{k=1}^{K} \alpha_k (\cos\theta_k)^n (\sin\theta_k)^{|m|} \mathrm{e}^{-jm\varphi_k}$, then

$$b_{n,m} = \sum_{k=1}^{K} \tilde{\alpha}_{n,k}^{(1)} u_k^m = \sum_{k=1}^{K} \tilde{\alpha}_{m,k}^{(2)} v_k^n, \tag{14}$$

  where $\tilde{\alpha}_{n,k}^{(1)} = \alpha_k (\cos\theta_k)^n$, $u_k = \sin\theta_k\, \mathrm{e}^{-j\varphi_k}$ for $m \ge 0$ and $u_k = \frac{1}{\sin\theta_k} \mathrm{e}^{-j\varphi_k}$ for $m < 0$; $\tilde{\alpha}_{m,k}^{(2)} = \alpha_k (\sin\theta_k)^{|m|}$ and $v_k = \cos\theta_k$. Consequently, we know that there exist two discrete filters such that

$$b_{n,m} \underset{m}{*} c_m^{(1)} = 0 \;\; \forall n \quad \text{and} \quad b_{n,m} \underset{n}{*} c_n^{(2)} = 0 \;\; \forall m.$$

- *Relation with the Given Measurements*: The expression (13) is nothing but a linear mapping from the sampled sinusoids $b_{n,m}$ to the given measurements (i.e., the spherical harmonics $\hat{I}_{l,m}$).

Thanks to this analysis, we are now ready to formulate the Dirac reconstruction problem on the sphere as two (for u_k and v_k each) constrained approximations of the form (P2). Take the reconstruction of v_k as an example:

- $\mathbf{a}$ is the given spherical harmonics $\hat{I}_{l,m}$;
- $\mathbf{b}$ is $b_{n,m}$ as defined in (14);
- $\mathbf{c}$ is the annihilating filter coefficients $c_k^{(2)}$;
- $\mathbf{G}$ is the linear relation between the given measurements $\hat{I}_{l,m}$ and $b_{n,m}$ in (13).

By reconstructing u_k and v_k from the given spherical harmonics, the angles φ_k and θ_k are uniquely specified. The Dirac amplitudes α_k can be easily obtained using least square minimization once we have reconstructed the values of φ_k and θ_k.

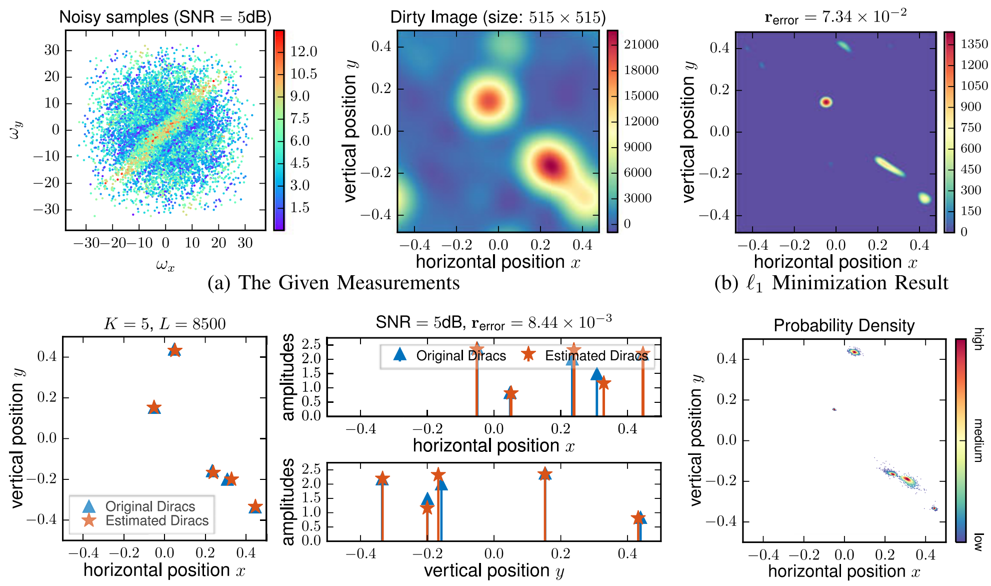
One major challenge in the earlier work [17] was to find the correct inverse transformation that should be applied to $\hat{I}_{l,m}$. In comparison, such an inverse mapping is no longer required with the proposed framework, which leads to a significantly more simplified formulation.

V. Application to Radio Interferometry

In this section, we apply the proposed reconstruction algorithm to a simplified radio interferometry problem. Cases with more realistic settings will be considered in a follow-up astronomy-oriented paper.

A. Data Acquisition and Signal Model

A radio interferometer consists of an array of antennas that collect the electromagnetic (EM) waves emitted by celestial sources in the sky. In a far field context, we can assume that these sources are located on a hypothetical celestial sphere and that the signals arriving at each antenna follow parallel lines (Fig. 1). Consequently, the received signals at two different antennas differ by a time delay, which is determined by the relative locations of the antennas with respect to the celestial



(a) The Given Measurements (b) ℓ_1 Minimization Result

(c) *Continuous* Domain Reconstruction (FRI) (d) *Statistics* of the Reconstructed Locations (FRI)

Fig. 11. Reconstruction of point sources from irregular Fourier measurements (SNR = 5 dB, number of Fourier measurements: $L = 8500$). (a) The given noisy Fourier samples and their spatial domain representation via inverse FFT (a.k.a., the the dirty image in radioastronomy). (b) The compressed sensing result by minimizing the ℓ_1 norm of the sky image (estimation error for point sources' locations: 7.34×10^{-2}). (c) The reconstructed point sources with FRI (estimation error for point sources' locations: 8.44×10^{-3}). (d) Probability density of the estimated point sources' locations with FRI approach (number of independent noise realizations: 1000; the average estimation error: 1.09×10^{-2}).

sources. It can be shown that under the assumption of a narrow field-of-view, the cross-correlation between the received EM waves at two different antennas (a.k.a. visibility in radioastronomy) is related to the Fourier transform of the underlying sky image $I(x, y)$ at a certain frequency (see [4] Chapter 3, equation (3.10)). Since there is a finite number of antennas with fixed locations, the radio interferometer will only have a *partial* Fourier domain coverage.

The conventional approach reconstructs the point sources in the discrete space by de-convolving the dirty image iteratively, which is the inverse discrete Fourier transform of the irregularly sampled Fourier measurements. Alternatively, as we demonstrate in the next section, we can directly address the reconstruction problem in the continuous-domain. In particular, our focus in this section is on the reconstruction of a sky image, which consists of *point sources* within the field of view:

$$I(x, y) = \sum_{k=1}^{K} \alpha_k \delta(x - x_k, y - y_k). \tag{15}$$

Here $\delta(\cdot, \cdot)$ is the Dirac delta distribution or generalized function, (x_k, y_k) is the location of the k-th point source, and $\alpha_k \geq 0$ is its intensity.

To summarize, the point source reconstruction problem in radio interferometry is as follows: How can we reconstruct the K Diracs on a 2D plane (15) from a given set of Fourier domain measurements at irregular frequencies:

$$\hat{I}(\omega_1^{(\ell)}, \omega_2^{(\ell)}) = \sum_{k=1}^{K} \alpha_k \mathrm{e}^{-j\omega_1^{(\ell)} x_k - j\omega_2^{(\ell)} y_k}, \tag{16}$$

for $\ell = 1, \ldots, L$? Note that in a realistic setting, these frequencies $(\omega_1^{(\ell)}, \omega_2^{(\ell)})$ should be based on the layout of the radio telescope. We have considered a simplified experimental setup here so as to be as close as possible to the algorithmic examples in the previous section. We leave the extra complications encountered in practice for an ongoing work on the processing of real data acquired with a radio telescope.

B. Reconstruction of Point Sources

Note that we have considered a similar 1D Dirac reconstruction problem in Section IV-B. Hence, we adopt the same strategy and approximate the Fourier transform over a finite area, e.g., $\omega_1 \times \omega_2 \in [-M\pi, M\pi] \times [-N\pi, N\pi]$. For practical considerations, we have chosen the Dirichlet interpolation kernel:

$$\psi(\omega_1, \omega_2) = \frac{\sin(\pi\omega_1)\sin(\pi\omega_2)}{MN\tau_1\tau_2 \sin(\frac{\pi\omega_1}{M\tau_1})\sin(\frac{\pi\omega_2}{N\tau_2})},$$

for some M and N such that $M\tau_1$ and $N\tau_2$ are odd numbers. The interpolation equation provides the link between the given Fourier measurements $\hat{I}(\omega_1^{(\ell)}, \omega_2^{(\ell)})$ and the sampled sinusoids $\hat{I}(2\pi m/\tau_1, 2\pi n/\tau_2)$ for $|m| \leq \lfloor \frac{M\tau_1}{2} \rfloor$ and $|n| \leq \lfloor \frac{N\tau_2}{2} \rfloor$, which can be annihilated by a discrete 2D filter.

In general, the solution that satisfies the 2D annihilation equations is a curve instead of a few isolated Diracs [16]. In fact, any Dirac that is located on the curve will satisfy the same set of 2D annihilation equations. One way to overcome such a difficulty is to reconstruct the Dirac's x and y locations *separately* by enforcing the annihilation constraint along each direction. Specifically, we would like to find two annihilating filters, whose

z-transforms are

$$C^{(1)}(z_1) = \sum_{k=0}^{K} c_k^{(1)} z_1^{-k} = c_0^{(1)} \prod_{k=1}^{K} (1 - u_k z_1^{-1})$$

and

$$C^{(2)}(z_2) = \sum_{k=0}^{K} c_k^{(2)} z_2^{-k} = c_0^{(2)} \prod_{k=1}^{K} (1 - v_k z_2^{-1})$$

with $u_k = \mathrm{e}^{-j\frac{2\pi}{\tau_1} x_k}$ and $v_k = \mathrm{e}^{-j\frac{2\pi}{\tau_2} y_k}$. The rows and columns of the Fourier transforms $\hat{I}(2\pi m/\tau_1, 2\pi n/\tau_2)$ are annihilated by the filter $[c_0^{(1)}, \cdots, c_K^{(1)}]$ and $[c_0^{(2)}, \cdots, c_K^{(2)}]$, respectively:

$$c_m^{(1)} \underset{m}{*} \hat{I}\left(\frac{2\pi m}{\tau_1}, \frac{2\pi n}{\tau_2}\right) = 0 \quad \text{and} \quad c_n^{(2)} \underset{n}{*} \hat{I}\left(\frac{2\pi m}{\tau_1}, \frac{2\pi n}{\tau_2}\right) = 0.$$

The Dirac locations are then reconstructed by solving two constrained approximation problems (P2). For the sake of brevity, we detail the exact formulation for the reconstruction of the Dirac vertical locations y_k only. The formulation for the reconstruction of x_k can be derived similarly.

Denote the annihilating filter coefficients c_k and the given Fourier measurements $\hat{I}(\omega_1^{(\ell)}, \omega_2^{(\ell)})$ as column vectors $\mathbf{c} = [c_0^{(2)}, \cdots, c_K^{(2)}]^{\mathrm{T}}$ and $\mathbf{a}$, respectively. We rearrange the unknown Fourier transform values $\hat{I}(2\pi m/\tau_1, 2\pi n/\tau_2)$ column by column as a vector $\mathbf{b}$. Then the constrained minimization that we would like to solve is:

$$\begin{aligned} \min_{\mathbf{b},\mathbf{c}} \quad & \left\|\mathbf{a} - \mathbf{G}\mathbf{b}\right\|_2^2 \\ \text{subject to} \quad & \mathbf{R}(\mathbf{c})\mathbf{b} = \mathbf{0}, \\ & \mathbf{c}_0^{\mathrm{H}}\mathbf{c} = 1, \end{aligned}$$

where $\mathbf{G} = [\widetilde{\mathbf{G}}^{(-\lfloor M\tau_1/2 \rfloor)} \cdots \widetilde{\mathbf{G}}^{(\lfloor M\tau_1/2 \rfloor)}]$ with $[\widetilde{\mathbf{G}}^{(m)}]_{\ell,n} = \psi(\frac{\omega_1^{(\ell)}}{2\pi/\tau_1} - m, \frac{\omega_2^{(\ell)}}{2\pi/\tau_2} - n)$; and[8]

$$\mathbf{R}(\mathbf{c}) = \overbrace{\begin{bmatrix} \widetilde{\mathbf{R}}(\mathbf{c}) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \widetilde{\mathbf{R}}(\mathbf{c}) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \widetilde{\mathbf{R}}(\mathbf{c}) \end{bmatrix}}^{2\lfloor M\tau_1/2 \rfloor \text{ blocks}},$$

with $\widetilde{\mathbf{R}}(\mathbf{c})$ the convolution matrix associated with the filter $\mathbf{c}$.

Once we have reconstructed the Dirac vertical and horizontal locations with the proposed algorithm in Section III, we still need to identify the correct associations — in principle, the 2D Diracs can be located on any one of the $K!$ possible combinations. The naive way would be the exhaustive search, where we try all the possible combinations and reconstruct the Dirac amplitudes α_k with the least square minimization. If the re-synthesized Fourier samples (16) based on the reconstructed parameters (α_k, x_k, y_k) are within the noise level, then we have correctly identified the Dirac locations.

However, such a straightforward approach is only computationally feasible for cases with few Diracs. Experimentally, we observed that we can find the correct associations of vertical and horizontal locations with a simple trick: We first reconstruct the amplitudes of the Diracs with the least square minimization by pretending there were K^2 Diracs (i.e., all the intersections of the reconstructed $\{x_k\}_{k=1}^K$ and $\{y_k\}_{k=1}^K$). Among these K^2 possible locations, we select K of them with the largest amplitudes. We should reconstruct α_k by solving the least square minimization once more with the correctly identified K Dirac locations on the 2D plane.

C. Simulation Results

We apply the algorithm to reconstruct 2D Diracs, which are located randomly in $[-0.5, 0.5] \times [-0.5, 0.5]$. The Dirac amplitudes follow a log-normal distribution: $\alpha_k \sim \log \mathcal{N}(\log 2, 0.5)$. The Fourier transform is sampled at $L = 8500$ irregular frequencies, which are randomly generated with higher concentrations around low frequencies. This is because low frequency Fourier measurements correspond to the cross-correlations between antennas that are close to each other, a case that is more convenient in practice. We have chosen 15×15 interpolation knots located uniformly on the area $[-12\pi, 12\pi] \times [-12\pi, 12\pi]$, where the Fourier transform is approximated. Complex-valued Gaussian white noise is added to the Fourier transforms so that the signal-to-noise ratio is 5 dB. The algorithm is able to reconstruct the Diracs correctly even in the presence of severe noise (Fig. 11). With our current Python implementation (which can be further optimized), it takes 42 seconds for the reconstruction on average with a Macbook Pro laptop. Following [47], we also include the classical sparse recovery result obtained when the ℓ_1-norm of the discretized sky image is minimized. As evidenced in Fig. 11, this approach is not only less accurate than the FRI method, but it also fails to resolve Diracs that are closely located.

VI. Conclusion

Motivated by the point source reconstruction problem in radio interferometry, we have developed a robust algorithmic framework for FRI reconstruction with arbitrary measurements, including the non-uniform sampling cases. We have unified all FRI-based methods concisely with a constrained formulation by establishing a linear relation between the given measurements and a set of unknown uniform samples of sinusoids. We have demonstrated the versatility of the proposed approach with various FRI signal recoveries in addition to an application to radio interferometry. The algorithm out-performs state of the art methods and is able to recover point sources accurately even in severe noise conditions. For future work, it would be interesting to consider an alternative (convex) formulation for FRI reconstructions, where a properly chosen atomic norm (see e.g., [25]) is minimized subject to a data-fitting constraint.

[8]As pointed out by a reviewer, it is possible to write $\mathbf{R}(\mathbf{c})$ compactly as the Kronecker product between $\widetilde{\mathbf{R}}(\mathbf{c})$ and an identity matrix of size $2\lfloor M\tau_1/2 \rfloor$: $\widetilde{\mathbf{R}}(\mathbf{c}) \otimes \mathbf{Id}_{2\lfloor M\tau_1/2 \rfloor}$.

Appendix A
Derivation of the Equivalent Formulation of (P1)

The constrained optimization for the reconstruction of FRI signals is

$$\begin{aligned} \min_{\mathbf{b},\mathbf{c}} \quad & \|\mathbf{a}-\mathbf{G}\mathbf{b}\|_2^2 \\ \text{subject to} \quad & \mathbf{R}(\mathbf{c})\mathbf{b}=\mathbf{0} \\ & \mathbf{c}_0^{\mathrm{H}}\mathbf{c}=1, \end{aligned} \tag{17}$$

where $\mathbf{c}_0$ is a random initialization for the annihilating filter coefficients.

For a fixed $\mathbf{c}$, (17) is a constrained quadratic minimization with respect to $\mathbf{b}$. The associated Lagrangian is:

$$\mathcal{L}(\mathbf{b},\boldsymbol{\ell}) = \frac{1}{2}\|\mathbf{a}-\mathbf{G}\mathbf{b}\|_2^2 + \boldsymbol{\ell}^{\mathrm{H}}\mathbf{R}(\mathbf{c})\mathbf{b},$$

where $\boldsymbol{\ell}$ is the Lagrange multiplier. From the optimality conditions, we have:

$$\begin{cases} \mathbf{G}^{\mathrm{H}}(\mathbf{G}\mathbf{b}-\mathbf{a}) + \mathbf{R}^{\mathrm{H}}(\mathbf{c})\boldsymbol{\ell} = \mathbf{0}, \\ \mathbf{R}(\mathbf{c})\mathbf{b} = \mathbf{0}. \end{cases} \tag{18}$$

Since $\mathbf{G}$ has full column rank (see footnote 2), from (18) we have

$$\mathbf{b} = \boldsymbol{\beta} - (\mathbf{G}^{\mathrm{H}}\mathbf{G})^{-1}\mathbf{R}^{\mathrm{H}}(\mathbf{c})(\mathbf{R}(\mathbf{c})(\mathbf{G}^{\mathrm{H}}\mathbf{G})^{-1}\mathbf{R}^{\mathrm{H}}(\mathbf{c}))^{-1}\mathbf{R}(\mathbf{c})\boldsymbol{\beta}, \tag{19}$$

where $\boldsymbol{\beta} = (\mathbf{G}^{\mathrm{H}}\mathbf{G})^{-1}\mathbf{G}^{\mathrm{H}}\mathbf{a}$.

We can substitute (19) into the objective function:

$$\begin{aligned} \|\mathbf{a}-\mathbf{G}\mathbf{b}\|_2^2 &= \boldsymbol{\beta}^{\mathrm{H}}\mathbf{R}^{\mathrm{H}}(\mathbf{c})(\mathbf{R}(\mathbf{c})(\mathbf{G}^{\mathrm{H}}\mathbf{G})^{-1}\mathbf{R}^{\mathrm{H}}(\mathbf{c}))^{-1}\mathbf{R}(\mathbf{c})\boldsymbol{\beta} \\ &\quad + \|\mathbf{a}-\mathbf{G}\boldsymbol{\beta}\|_2^2 \\ &\overset{(a)}{=} \mathbf{c}^{\mathrm{H}}\mathbf{T}^{\mathrm{H}}(\boldsymbol{\beta})(\mathbf{R}(\mathbf{c})(\mathbf{G}^{\mathrm{H}}\mathbf{G})^{-1}\mathbf{R}^{\mathrm{H}}(\mathbf{c}))^{-1}\mathbf{T}(\boldsymbol{\beta})\mathbf{c} \\ &\quad + \text{terms independent of } \mathbf{c}, \end{aligned}$$

where (a) results from the definition of the right dual matrix in Definition 1.

Appendix B
Equivalent Form for the Solution of (3)

The Lagrangian associated with the constrained optimization (3) is

$$\begin{aligned} \mathcal{L}(\mathbf{c},\lambda) &= \frac{1}{2}\mathbf{c}^{\mathrm{H}}\mathbf{T}^{\mathrm{H}}(\boldsymbol{\beta})(\mathbf{R}(\mathbf{c}_{n-1})(\mathbf{G}^{\mathrm{H}}\mathbf{G})^{-1}\mathbf{R}^{\mathrm{H}}(\mathbf{c}_{n-1}))^{-1}\mathbf{T}(\boldsymbol{\beta})\mathbf{c} \\ &\quad + \lambda(\mathbf{c}_0^{\mathrm{H}}\mathbf{c}-1). \end{aligned}$$

From the optimality conditions, we have

$$\begin{cases} \mathbf{T}^{\mathrm{H}}(\boldsymbol{\beta})(\mathbf{R}(\mathbf{c}_{n-1})(\mathbf{G}^{\mathrm{H}}\mathbf{G})^{-1}\mathbf{R}^{\mathrm{H}}(\mathbf{c}_{n-1}))^{-1}\mathbf{T}(\boldsymbol{\beta})\mathbf{c} + \lambda\mathbf{c}_0 = \mathbf{0}, \\ \mathbf{c}_0^{\mathrm{H}}\mathbf{c} = 1. \end{cases} \tag{20}$$

Denote an auxiliary variable

$$\boldsymbol{\ell} = (\mathbf{R}(\mathbf{c}_{n-1})(\mathbf{G}^{\mathrm{H}}\mathbf{G})^{-1}\mathbf{R}^{\mathrm{H}}(\mathbf{c}_{n-1}))^{-1}\mathbf{T}(\boldsymbol{\beta})\mathbf{c},$$

then (20) is equivalent to

$$\begin{cases} \mathbf{T}^{\mathrm{H}}(\boldsymbol{\beta})\boldsymbol{\ell} + \lambda\mathbf{c}_0 = \mathbf{0}, \\ \mathbf{R}(\mathbf{c}_{n-1})(\mathbf{G}^{\mathrm{H}}\mathbf{G})^{-1}\mathbf{R}^{\mathrm{H}}(\mathbf{c}_{n-1})\boldsymbol{\ell} = \mathbf{T}(\boldsymbol{\beta})\mathbf{c}, \\ \mathbf{c}_0^{\mathrm{H}}\mathbf{c} = 1. \end{cases} \tag{21}$$

We can apply the same manipulation again by introducing another auxiliary variable $\mathbf{v} = (\mathbf{G}^{\mathrm{H}}\mathbf{G})^{-1}\mathbf{R}^{\mathrm{H}}(\mathbf{c}_{n-1})\boldsymbol{\ell}$, then (21) is equivalent to:

$$\begin{cases} \mathbf{T}^{\mathrm{H}}(\boldsymbol{\beta})\boldsymbol{\ell} + \lambda\mathbf{c}_0 = \mathbf{0}, \\ \mathbf{R}(\mathbf{c}_{n-1})\mathbf{v} = \mathbf{T}(\boldsymbol{\beta})\mathbf{c}, \\ \mathbf{G}^{\mathrm{H}}\mathbf{G}\mathbf{v} = \mathbf{R}^{\mathrm{H}}(\mathbf{c}_{n-1})\boldsymbol{\ell}, \\ \mathbf{c}_0^{\mathrm{H}}\mathbf{c} = 1. \end{cases} \tag{22}$$

If we rearrange (22) in a matrix/vector form, we have (4).

Once we have the updated annihilating filter coefficients $\mathbf{c}_n$, (5) is obtained directly by rewriting the optimality conditions (18) with $\mathbf{c} = \mathbf{c}_n$ as a linear system.

Appendix C
Exact Interpolation With Dirichlet Kernel

Since $x(t)$ has finite time support between $-\frac{\tau}{2}$ and $\frac{\tau}{2}$, we can rewrite $x(t)$ with its periodized version multiplied by a rectangular window: $x(t) = \mathrm{rect}(t/\tau)\sum_{n\in\mathbb{Z}} x(t-n\tau)$. Hence, the Fourier transform of $x(t)$ is

$$\begin{aligned} X(\omega) &= \int_{-\infty}^{\infty} \mathrm{rect}(t/\tau)\sum_{n\in\mathbb{Z}} x(t-n\tau)\mathrm{e}^{-j\omega t}\mathrm{d}t \\ &\overset{(a)}{=} \int_{-\infty}^{\infty} \mathrm{rect}(t/\tau)\mathrm{e}^{-j\omega t}\cdot\frac{1}{\tau}\sum_{m\in\mathbb{Z}} X\left(\frac{2\pi m}{\tau}\right)\mathrm{e}^{j\frac{2\pi m}{\tau}t}\mathrm{d}t \\ &= \sum_{m\in\mathbb{Z}} X\left(\frac{2\pi m}{\tau}\right)\int_{-\infty}^{\infty}\frac{1}{\tau}\mathrm{rect}\left(\frac{t}{\tau}\right)\mathrm{e}^{-j(\omega-\frac{2\pi m}{\tau})t}\mathrm{d}t \\ &= \sum_{m\in\mathbb{Z}} X\left(\frac{2\pi m}{\tau}\right)\mathrm{sinc}\left(\frac{\tau}{2}\left(\omega-\frac{2\pi m}{\tau}\right)\right), \end{aligned} \tag{23}$$

where (a) is from the Poisson sum formula.

Further, from the periodicity of $X(\omega)$, we can rewrite the infinite summation in (23) as

$$\begin{aligned} X(\omega) &= \sum_{n\in\mathbb{Z}}\sum_{|m|\le\lfloor\frac{M\tau}{2}\rfloor} X\left(\frac{2\pi m}{\tau}\right)\mathrm{sinc}\left(\frac{\tau}{2}\left(\omega-\frac{2\pi(m+nM\tau)}{\tau}\right)\right) \\ &= \sum_{|m|\le\lfloor\frac{M\tau}{2}\rfloor} X\left(\frac{2\pi m}{\tau}\right)\sum_{n\in\mathbb{Z}}\mathrm{sinc}\left(\frac{\tau}{2}\left(\omega-\frac{2\pi(m+nM\tau)}{\tau}\right)\right) \\ &= \sum_{|m|\le\lfloor\frac{M\tau}{2}\rfloor} X\left(\frac{2\pi m}{\tau}\right)\frac{\sin(\frac{\tau\omega-2\pi m}{2})}{M\tau\sin\left(\frac{\tau\omega-2\pi m}{2M\tau}\right)}. \end{aligned}$$

Appendix D
Spherical Harmonics of Diracs on the Sphere

Conventionally, spherical harmonics of degree l and order m is defined as $Y_l^m(\theta,\varphi) = N_{l,m} P_l^{|m|}(\cos\theta) \mathrm{e}^{jm\varphi}$. Here the normalization factor $N_{l,m} = (-1)^{(m+|m|)/2}\sqrt{\frac{2l+1}{4\pi}\frac{(l-|m|)!}{(l+|m|)!}}$; and $P_l^{|m|}$ is the Legendre polynomial of degree l and order m ($|m| \leq l$):

$$P_l^{|m|}(t) = (-1)^{|m|}(1-t^2)^{|m|/2}\frac{\mathrm{d}^{|m|}}{\mathrm{d}t^{|m|}}P_l(t),$$

where $P_l(t) \stackrel{\text{def}}{=} \frac{1}{2^l l!}\frac{\mathrm{d}^l}{\mathrm{d}t^l}(t^2-1)^l$. The spherical harmonic coefficient is given by the inner product between the signal and the spherical harmonics basis on $\mathbb{S}^2$:

$$\begin{aligned}\hat{I}_{l,m} &= \langle I(\theta,\varphi), Y_l^m(\theta,\varphi)\rangle \\ &= \int_0^{2\pi}\int_0^{\pi} I(\theta,\varphi) N_{l,m} P_l^{|m|}(\cos\theta)\mathrm{e}^{-jm\varphi}\sin\theta \mathrm{d}\theta\mathrm{d}\varphi \\ &= N_{l,m}\sum_{k=1}^{K}\alpha_k P_l^{|m|}(\cos\theta_k)\mathrm{e}^{-jm\varphi_k}.\end{aligned}$$

Note that the $|m|$-th order derivative of the Legendre polynomial $P_l(t)$ is a polynomial of degree $l-|m|$. Hence, we may rewrite $P_l^{|m|}$ in terms of canonical polynomial bases as:

$$P_l^{|m|}(t) = (-1)^{|m|}(1-t^2)^{|m|/2}\sum_{n=0}^{l-|m|} p_{n,|m|}t^n,$$

for some coefficients $p_{n,|m|}$, which are independent of where the polynomial is evaluated, and can be precomputed. Consequently the spherical harmonic coefficient of the Diracs is:

$$\hat{I}_{l,m} = N_{l,m}\sum_{n=0}^{l-|m|} p_{n,|m|}\sum_{k=1}^{K}\alpha_k(\cos\theta_k)^n(\sin\theta_k)^{|m|}\mathrm{e}^{-jm\varphi_k}.$$

Acknowledgment

The authors would like to thank Matthieu Simeoni of EPFL and Paul Hurley of IBM Zürich for insightful discussions on radio interferometry. They thank the reviewers for their constructive comments which helped clarify the paper.

References

[1] R. Prony, "Essai expérimental et analytique sur les lois de la dilabilité des fluides elastique et sur celles de la force expansive de la vapeur de l'alkool, à différentes températures," *J. l'Ecole Polytechnique*, vol. 1, no. 2, pp. 24–76, 1795.

[2] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Upper Saddle River, NJ, USA: Pearson/Prentice-Hall, 2005.

[3] 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, Aug. 2002.

[4] A. R. Thompson, J. M. Moran, and G. W. Swenson, *Interferometry and Synthesis in Radio Astronomy*. New York, NY, USA: Wiley, 2001.

[5] J. Högbom, "Aperture synthesis with a non-regular distribution of interferometer baselines," *Astron. Astrophys. Suppl. Ser.*, vol. 15, pp. 417–426, 1974.

[6] Y. Wiaux, L. Jacques, G. Puy, A. Scaife, and P. Vandergheynst, "Compressed sensing imaging techniques for radio interferometry," *Monthly Notices Roy. Astron. Soc.*, vol. 395, no. 3, pp. 1733–1742, 2009.

[7] S. Wenger, M. Magnor, Y. Pihlström, S. Bhatnagar, and U. Rau, "SparseRI: A compressed sensing framework for aperture synthesis imaging in radio astronomy," *Publications Astron. Soc. Pacific*, vol. 122, no. 897, pp. 1367–1374, Oct. 2010.

[8] R. E. Carrillo, J. D. McEwen, and Y. Wiaux, "PURIFY: A new approach to radio-interferometric imaging," *Monthly Notices Roy. Astron. Soc.*, vol. 439, no. 4, pp. 3591–3604, 2014.

[9] P. Milanfar, G. C. Verghese, W. C. Karl, and A. S. Willsky, "Reconstructing polygons from moments with connections to array processing," *IEEE Trans. Signal Process.*, vol. 43, no. 2, pp. 432–443, Feb. 1995.

[10] I. Maravić and M. Vetterli, "Exact sampling results for some classes of parametric nonbandlimited 2-D signals," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 175–189, Jan. 2004.

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

[12] L. Baboulaz and P. L. Dragotti, "Exact feature extraction using finite rate of innovation principles with an application to image super-resolution," *IEEE Trans. Image Process.*, vol. 18, no. 2, pp. 281–298, Feb. 2009.

[13] J. Berent, P. L. Dragotti, and T. Blu, "Sampling piecewise sinusoidal signals with finite rate of innovation methods," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 613–625, Feb. 2010.

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

[15] S. Deslauriers-Gauthier and P. Marziliano, "Sampling signals with a finite rate of innovation on the sphere," *IEEE Trans. Signal Process.*, vol. 61, no. 18, pp. 4552–4561, Sep. 2013.

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

[17] I. Dokmanić and Y. M. Lu, "Sampling sparse signals on the sphere: Algorithms and applications," *IEEE Trans. Signal Process.*, vol. 64, no. 1, pp. 189–202, Jan. 2016.

[18] T. Blu, P. L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, "Sparse sampling of signal innovations," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 31–40, Mar. 2008.

[19] Z. Doğan, C. Gilliam, T. Blu, and D. Van De Ville, "Reconstruction of finite rate of innovation signals with model-fitting approach," *IEEE Trans. Signal Process.*, vol. 63, no. 22, pp. 6024–6036, Nov. 2015.

[20] 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.

[21] J. Urigüen, T. Blu, and P. L. Dragotti, "FRI sampling with arbitrary kernels," *IEEE Trans. Signal Process.*, vol. 61, no. 21, pp. 5310–5323, Nov. 2013.

[22] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP-34, no. 5, pp. 1081–1089, Oct. 1986.

[23] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[24] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Commun. Pure Appl. Math.*, vol. 67, no. 6, pp. 906–956, 2014.

[25] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7465–7490, Nov. 2013.

[26] C. Aubel, D. Stotz, and H. Bölcskei, "A theory of super-resolution from short-time Fourier transform measurements," *J. Fourier Anal. Appl.*, 2017, to appear.

[27] G. Tang, B. N. Bhaskar, and B. Recht, "Near minimax line spectral estimation," *IEEE Trans. Inf. Theory*, vol. 61, no. 1, pp. 499–512, Jan. 2015.

[28] Y. Li and Y. Chi, "Off-the-grid line spectrum denoising and estimation with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 64, no. 5, pp. 1257–1269, Mar. 2016.

[29] I. Markovsky, "How effective is the nuclear norm heuristic in solving data approximation problems?" in *Proc. 16th IFAC Symp. Syst. Identification*, 2012, pp. 316–321.

[30] F. Andersson, M. Carlsson, J.-Y. Tourneret, and H. Wendt, "A new frequency estimation method for equally and unequally spaced data," *IEEE Trans. Signal Process.*, vol. 62, no. 21, pp. 5761–5774, Nov. 1, 2014.

[31] L. Condat and A. Hirabayashi, "Cadzow denoising upgraded: A new projection method for the recovery of Dirac pulses from noisy linear measurements," *Sampl. Theory Signal Image Process.*, vol. 14, no. 1, pp. p–17, 2015.

[32] Y. Chen and Y. Chi, "Robust spectral compressed sensing via structured matrix completion," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6576–6601, Oct. 2014.

[33] J. C. Ye, J. M. Kim, K. H. Jin, and K. Lee, "Compressive sampling using annihilating filter-based low-rank interpolation," *IEEE Trans. Inf. Theory*, vol. PP, no. 99, 2016, doi: 10.1109/TIT.2016.2629078, preprint.

[34] G. Ongie, S. Biswas, and M. Jacob, "Structured low-rank recovery of piecewise constant signals with performance guarantees," in *Proc. IEEE Int. Conf. Image Process (ICIP)*, Sep. 2016, pp. 963–967.

[35] *Proc. IEEE, Special Issue on Spectral Estimation*, S. Haykin and J. A. Cadzow, Eds., vol. 70, no. 9, pp. 883–1125, Sep. 1982.

[36] R. P. Millane, "Phase retrieval in crystallography and optics," *J. Opt. Soc. Amer. A*, vol. 7, no. 3, pp. 394–411, 1990.

[37] T. Blu, H. Bay, and M. Unser, "A new high-resolution processing method for the deconvolution of optical coherence tomography signals," in *Proc. 1st IEEE Int. Symp. Biomed. Imaging: Macro Nano*, Jul. 7–10, 2002, vol. 3, pp. 777–780.

[38] I. Maravić and M. Vetterli, "Sampling and reconstruction of signals with finite rate of innovation in the presence of noise," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2788–2805, Aug. 2005.

[39] J. A. Cadzow, "Signal enhancement—A composite property mapping algorithm," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 1, pp. 49–62, Jan. 1988.

[40] B. De Moor, "Total least squares for affinely structured matrices and the noisy realization problem," *IEEE Trans. Signal Process.*, vol. 42, no. 11, pp. 3104–3113, Nov. 1994.

[41] C. Gilliam and T. Blu, "Fitting instead of annihilation: Improved recovery of noisy FRI signals," in *Proc. 39th IEEE Int. Conf. Acoust. Speech, Signal Process.*, May 4–9, 2014, pp. 51–55.

[42] C. Gilliam and T. Blu, "Finding the minimum rate of innovation in the presence of noise," in *Proc. 41st IEEE Int. Conf. Acoust. Speech Signal Process.*, Mar. 20–25, 2016, pp. 4019–4023.

[43] G. H. Golub and C. F. Van Loan, *Matrix Computations*, vol. 3. Baltimore, MD, USA: Johns Hopkins Univ. Press, 2012.

[44] P. C. Lauterbur, "Image formation by induced local interactions: Examples employing nuclear magnetic resonance," *Nature*, vol. 242, pp. 190–191, 1973.

[45] D. Gabor, "A new microscopic principle," *Nature*, vol. 161, no. 4098, pp. 777–778, 1948.

[46] R. Perley, F. Schwab, and A. Bridle, *Synthesis Imaging in Radio Astronomy*. San Francisco, CA, USA: Astronomical Soc. of the Pacific, Jan. 1989.

[47] H. Garsden *et al.*, "LOFAR sparse image reconstruction," *Astron. Astrophys.*, vol. 575, 2015, Art. no. A90.

Hanjie Pan (S'11) was born in Jiangsu, China, in 1988. In 2010, he received the B.Eng. degree with first honor in electronic engineering from the Chinese University of Hong Kong, Shatin, Hong Kong. He received the M.Phil. degree from the same institution in 2013. He is currently working toward the Ph.D. degree in the Laboratory of Audiovisual Communications, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. In the summer 2010, he was a Visiting Student at Imperial College, London, United Kingdom. His research interests include image restorations with sparsity constraints and sampling signals with finite rate of innovation.

Thierry Blu (M'96–SM'06–F'12) was born in Orléans, France, in 1964. He received the "Diplôme d'ingénieur" degree from École Polytechnique, Palaiseau, France, in 1986 and from Télécom Paris (ENST), Paris, France, in 1988. In 1996, he received the Ph.D. degree in electrical engineering from ENST for a study on iterated rational filterbanks, applied to wideband audio coding.

He was with the Biomedical Imaging Group, Swiss Federal Institute of Technology, Lausanne, Switzerland, from 1998 and 2007. He is currently a Professor in the Department of Electronic Engineering, Chinese University of Hong Kong. His research interests include wavelets, approximation and sampling theory, sparse representations, image denoising, biomedical imaging, optics, and wave propagation.

Dr. Blu received two best paper awards from the IEEE Signal Processing Society (2003 and 2006). He is also coauthor of a paper that received a Young Author best paper award (2009) from the same society.

He has been a member of the IEEE Signal Processing Theory and Methods Technical Committee (2008–2013), and an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING (2002–2006), the IEEE TRANSACTIONS ON SIGNAL PROCESSING (2006–2010), and *Elsevier Signal Processing* (2008–2011). He is currently on the board of *EURASIP Journal on Image and Video Processing* (since 2010).

Martin Vetterli (M'86–SM'90–F'95) received the Dipl. El.-Ing. degree from ETH Zurich (ETHZ), Zürich, Switzerland, in 1981, the M.S. degree from Stanford University, Stanford, CA, USA, in 1982, and the Doctoratés Sciences degree from EPF Lausanne (EPFL), Lausanne, Switzerland, in 1986.

He was a Research Assistant at Stanford and EPFL, and has worked for Siemens and AT&T Bell Laboratories. In 1986, he joined Columbia University in New York, where he was last an Associate Professor of electrical engineering and Co-Director of the Image and Advanced Television Laboratory. In 1993, he joined the University of California at Berkeley, where he was a Professor in the Department of Electrical Engineering and Computer Sciences until 1997, and has held an Adjunct Professor position until June 2010.

Since 1995, he is a Professor of communication systems at EPF Lausanne, Switzerland, where he chaired the Communications Systems Division (1996/1997), and Heads the Audiovisual Communications Laboratory. From 2001 to 2004, he directed the National Competence Center in research on mobile information and communication systems. He also was a Vice-President at EPFL from October 2004 to February 2011 in charge, among others, of international affairs and computing services. He has held visiting positions at ETHZ (1990) and Stanford (1998). From March 2011 onward, he is the Dean of the School of Computer and Communication Sciences, EPFL. Since January 2013, he leads the Swiss National Science Foundation. He has been nominated to the Presidency of EPFL as of January 2017. He has published about 170 journal papers on a variety of topics in signal/image processing and communications and holds about 50 patents and patent applications. His research interests include sampling, wavelets, multirate signal processing, computational complexity, signal processing for communications, digital image/video processing, joint source/channel coding, signal processing for sensor networks and inverse problems like acoustic tomography.

He is a Fellow of ACM, a Fellow of EURASIP, and a member of SIAM and NAE. He is on the editorial boards of *Applied and Computational Harmonic Analysis*, the *Journal of Fourier Analysis and Application*, and the IEEE JOURNAL ON SELECTED TOPICS IN SIGNAL PROCESSING and has been elected Foreign Member of the NAE in 2015.

He received the Best Paper Award of EURASIP in 1984, the Research Prize of the Brown Bovery Corporation (Switzerland) in 1986, the IEEE Signal Processing Society's Senior Paper Awards in 1991, in 1996 and in 2006 (for papers with D. LeGall, K. Ramchandran, and Marziliano and Blu, respectively). He won the Swiss National Latsis Prize in 1996, the SPIE Presidential award in 1999, the IEEE Signal Processing Technical Achievement Award in 2001, the IEEE Signal Processing Society Award in 2010 for fundamental contributions to signal processing theory, technology and education, and is an ISI highly cited researcher in engineering. He was a member of the Swiss Council on Science and Technology from 2000 to 2003.

He was a plenary speaker at various conferences (e.g., IEEE ICIP, ICASSP, ISIT) and is the co-author of three books with J. Kovacevic, *Wavelets and Subband Coding* (Prentice Hall, 1995), with P. Prandoni, *Signal Processing for Communications* (EPFL Press, 2008), and with J. Kovacevic and V. K. Goyal, *Foundations of Signal Processing* (Cambridge University Press, 2015).