Reconstruction of Finite Rate of Innovation Signals with Model-Fitting Approach

Zafer Doğan, *Member, IEEE*, Christopher Gilliam, *Member, IEEE*, Thierry Blu, *Fellow, IEEE*, and
Dimitri Van De Ville, *Senior Member, IEEE*

Abstract—Finite rate of innovation (FRI) is a recent framework for sampling and reconstruction of a large class of parametric signals that are characterized by finite number of innovations (parameters) per unit interval. In the absence of noise, exact recovery of FRI signals has been demonstrated. In the noisy scenario, there exist techniques to deal with non-ideal measurements. Yet, the accuracy and resiliency to noise and model mismatch are still challenging problems for real-world applications. We address the reconstruction of FRI signals, specifically a stream of Diracs, from few signal samples degraded by noise and we propose a new FRI reconstruction method that is based on a model—fitting approach related to the structured—TLS problem. The model—fitting method is based on minimizing the training error, that is, the error between the computed and the recovered moments (i.e., the FRI-samples of the signal), subject to an annihilation system. We present our framework for three different constraints of the annihilation system. Moreover, we propose a model order selection framework to determine the innovation rate of the signal; i.e., the number of Diracs by estimating the noise level through the training error curve. We compare the performance of the model—fitting approach with known FRI reconstruction algorithms and Cramér–Rao's lower bound (CRLB) to validate these contributions.

Index Terms—Annihilating Filter, Cadzow, Cramér-Rao's lower bound (CRLB), finite-rate-of-innovation, iterative quadratic maximum likelihood (IQML), Kumaresan-Tufts, matrix pencil, model fitting, noise, reconstruction, sampling, structured total least squares (STLS), total least squares (TLS).

I. Introduction

$\mathbf{S}$AMPLING is essential in digital signal processing to convert a continuous—time signal into a discrete—time signal. Sampling theorems provide the fundamental bridge that allows the recovery of a continuous—time signal from a set of discrete samples. Over the past 60 years, the sampling process is predominantly based on the well-known Shannon sampling theorem that states uniform samples of a bandlimited (BL) signal at the Nyquist rate (i.e., at least twice of the bandwidth) are sufficient to perfectly reconstruct the signal [1], [2]. In this case, the reconstruction formula is given as a linear interpolation of the discrete samples with a sinc function. For non—BL signals, Shannon—type sampling can be interpreted as an approximation procedure where the original signal is projected onto the shift-invariant subspace of BL signals. This interpretation has been generalized for classes of non—BL signals that belong to shift invariant subspaces, such as uniform splines and scaling functions [2]–[4]. Such frameworks have proved to be useful from an implementation point of view by providing simpler interpolation models.

The recent theory of finite rate of innovation (FRI) has shown that it is possible to develop exact sampling and reconstruction schemes for specific classes of signals that are neither BL nor belong to a fixed subspace [5], [6]. Signals with FRI are characterized by a set of signal innovations. This representation includes the classes of signals that can be written as a weighted sum of (a set of known) functions with arbitrary shifts where the innovation weights and the innovation instants (i.e., shifts) are the only degrees of freedom of the signal. Examples of such FRI signals are streams of Diracs and stream of short pulses (with known pulse shape) [7], [8], uniform and non—uniform splines, piecewise polynomials and sinusoidal signals [9]–[11]. Moreover, further examples of multidimensional FRI signals have been studied in different settings [12]–[15].

In Fig. 1, the FRI framework is depicted with its three essential steps. For the acquisition, various sampling kernels have been proposed such as the infinite—support sinc and Gaussian kernels [5] as well as compact support sampling kernels that satisfy the generalized Strang-Fix condition for reproduction of polynomials and exponentials, i.e., the family of B-splines [9]. Recently, FRI sampling has been further extended to include arbitrary sampling kernels that allow only approximate reproduction of exponentials. As a consequence, the limitation on the choice of the sampling kernel has been successfully alleviated [16]. In the second step, namely the mapping, the ability of the sampling kernel to reproduce exponentials plays a central role. For this purpose, the mapping coefficients are first computed imposing the Strang-Fix condition to reproduce a set of exponentials. Then, these coefficients are used to combine linearly the signal samples to obtain the FRI samples which are actually the exponential moments of the FRI signal. Hence, the reconstruction is reduced to recovering the signal innovations from

Z. Doğan and D. Van De Ville are with the Medical Image Processing Lab (MIPLAB), Institute of Bioengineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne CH-1016, Switzerland, and with the Department of Radiology and Medical Informatics, University of Geneva, CH-1211 Geneva, Switzerland (e-mail: zafer.dogan@epfl.ch; dimitri.vandeville@epfl.ch).

C. Gilliam and T. Blu are with the Department of Electronics Engineering, The Chinese University of Hong Kong (CUHK), Hong Kong (e-mail: cgilliam@ee.cuhk.edu.hk; thierry.blu@m4x.org).
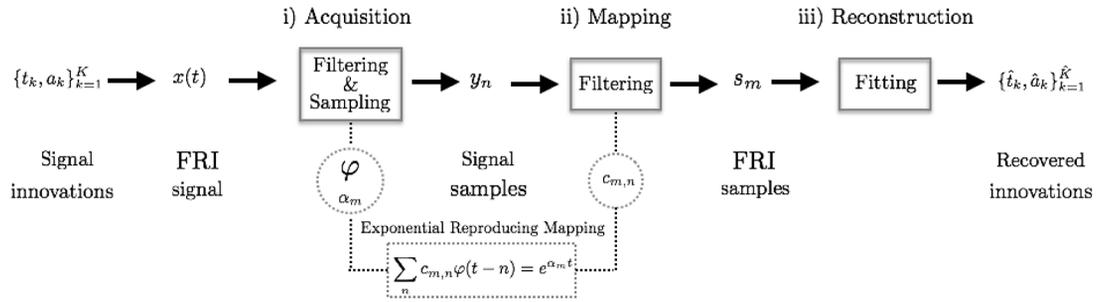
Fig. 1. FRI framework: The FRI signal is characterized by a finite number of signal innovations, namely innovation instants t_k and innovation weights a_k. (i) Acquisition: The FRI-signal $x(t)$ is filtered with a sampling kernel before being uniformly sampled. (ii) Mapping: The signal samples are mapped to FRI samples using the coefficients $c_{m,n}$ derived from the information of the sampling kernel. (iii) Reconstruction: The signal innovations is recovered from the FRI-samples using a spectral estimation method such as the annihilating filter method.

these FRI samples which typically reverts to a spectral estimation problem that involves a nonlinear estimation of the innovation instants and a least squares problem for the innovation weights.

In the absence of noise, perfect reconstruction is achieved using the annihilating filter (AF) method of [6], which is a variation of the well-known Prony's method [17]. However, this method becomes unstable and the accuracy of the reconstruction substantially degrades in the presence of noise. Several approaches to improve resiliency against measurement noise and model mismatch have been proposed [10], [18], [19]. In particular, there are two main approaches to remove the noise on the FRI samples that are derived from the several classes of high-resolution and subspace-based methods known from spectral estimation. The first approach is based on state space parametrization of the signal subspace that reduces the denoising problem into the estimation problem of the generalized eigenvalue of matrix pencil [18], [20]. A closely related algorithm, the ESPRIT algorithm, was developed as an extension of the state space method [21], [22]. The second approach, known as Cadzow denoising, is based on an iterative procedure that imposes the ideal properties of the FRI samples, namely the rank deficiency of the corresponding system in the AF method and the Toeplitz structure of this system at each iteration [6], [23]. Additionally, these FRI methods can achieve the optimal performance given by the CRLB only above a certain SNR and breaks down at lower SNR. Recently, in [24], this condition has been further analyzed and the condition of guaranteed performance in the FRI setting is given in terms of SNR, the sampling rate and the distance between adjacent Diracs.

Note that these methods can be classified as deterministic reconstruction algorithms except the random initialization for iterative methods. Alternatively, stochastic recovery algorithms have been proposed in FRI setting based on Gibbs sampling [25]. Briefly, the idea is to draw samples from a multivariate posterior probability distribution to infer the parameters of the FRI signal by calculating some statistic measures from the drawn samples. However, these methods are usually slower than the deterministic ones. An improvement of this approach is achieved in [26] inspired by a genetic algorithm but does not improve the computational efficiency. A deterministic approach has been proposed in [27] based on iterative maximum likelihood estimation to remedy the runtime of these methods.

Finally, another alternative method has been proposed based on fitting a signal model to the noisy data [25], [28]. Yet, robust estimation of the nonlinear signal innovations remains a challenging problem for practical applications.

Another challenge for the FRI reconstruction is to estimate the innovation rate of the FRI signal, i.e., the number of signal innovations per unit of time. In statistical estimation framework, this is known as the model order selection to estimate the dimension of the problem. In this case, the optimal model minimizes an information criterion with the principle of parsimony that favors simple models over complex ones for equal data fitting quality, e.g., the Bayesian information criterion (BIC), Akaike's information criterion (AIC), and network information criterion (NIC) [29]. For the FRI framework, this is translated in estimating the innovation rate from a set of FRI samples together with the signal innovations. Often, the innovation rate is assumed to be known by the conventional FRI recovery methods and, the standard way to compare the performance of the algorithms is to check the estimation accuracy of the signal innovations against the CRLB [5], [6], [16]. However, this may not be applicable if the true innovation rate or the true signal innovations are unknown. Hence, there are different approaches to estimate the innovation rate of the signal based on observation of the separation of the singular values of the system matrix in Cadzow denoising [6], [15] and in the subspace-based methods [10]. But, these methods can only provide a reliable estimation of the innovation rate if the amount of noise allows to differentiate a clear separation of the signal-related singular values from the remaining ones. Moreover, it is still necessary to provide adequate reconstruction that best explains the available data even if the noise level does not allow an accurate estimation of the true innovation rate. In [8], another practical approach is proposed to perform a consistency analysis of the retrieved innovation instants using a sliding window approach by building a histogram of retrieved locations among different windows.

In this paper, we propose a novel FRI reconstruction method based on a model—fitting approach. For this purpose, we formulate the nonlinear estimation of the innovation instants as a constrained optimization problem where we minimize the error between the measured FRI samples and the estimated FRI samples subject to the annihilation system. Our contributions are twofold: First, we propose a number of new model-fitting reconstruction methods exploring different constraints applicable

to the annihilation system for the estimation of the innovation instants. We show that the problem can be reduced to a simplified version of the structured total least squares (STLS) problem [30]. We then solve a complex least squares problem for the estimation of the innovation weights. Our second contribution is a model order selection framework to determine the innovation rate of the FRI signal that allows reliable estimation of the signal innovations even at low SNR. The model order selection is based on analyzing the training error curve that reflects the error between the FRI samples and the estimated FRI samples for different model order.

The outline of the paper is as follows. In Section II, we provide a brief overview of the theory of FRI sampling and reconstruction algorithms. Next in Section III, we formally define the FRI model-fitting problem and propose a number of FRI recovery algorithms together with a model order selection framework. In Section IV, we present simulation results to validate our contributions by comparing our results with respect to the common FRI recovery methods and the Cramér-Rao bound for our estimation problem. We also demonstrate the simulation results for the model order selection.

II. SAMPLING SIGNALS WITH FINITE RATE OF INNOVATION

In this section, we review the fundamentals of the FRI theory more formally. Consider a set of signal innovations $\{t_k, a_k\}_{k=0}^{K-1}$ that generates a signal characterized by a parametric representation of the form

$$x(t) = \sum_{k=0}^{K-1} a_k \delta(t - t_k). \tag{1}$$

The only degrees of freedom in the signal are the number of innovation K, the innovation instants, t_k and the innovation weights a_k. In this case, the signal $x(t)$ is defined as an FRI signal since the number of signal innovations is finite per unit time [5], [9].

A. Acquisition of FRI Signals

In a typical FRI framework, as shown in Fig. 1, the acquisition of FRI signals is handled by the standard filtering and sampling. This process can be written in terms of an inner product of the input signal with the sampling kernel. In this case, the signal samples are given by

$$y_n = \left\langle x(t), \varphi\left(\frac{t}{T} - n\right) \right\rangle, \tag{2}$$

where the sampling kernel, φ, is the scaled and time-reversed version of the filter's impulse response and T is the sampling interval. In fact, the impulse response of the filter accounts for the physical properties of the acquisition device and cannot be modified. Therefore, it is essential to develop sampling schemes that do not require the use of particular kernels. Initially, various sampling kernels have been proposed for exact sampling and reconstruction of FRI signals such as the infinite—support sinc and Gaussian functions [5], [6] and compact support polynomial and exponential reproducing kernels [9]. In order to guarantee exact reconstruction, the sampling kernels are required to satisfy the so-called generalized Strang-Fix conditions [31]. However, these limitations on the choice of the sampling kernels have

been successfully removed by relaxing the exact exponential reproduction requirement [8], [16].

Here, we review the condition for the exponential reproducing kernels that heavily relies on the theory of e-splines [32]. An exponential reproducing kernel is any function $\varphi(t)$ that, together with a linear combination of its shifted versions, can reproduce exponentials of the form $e^{\alpha t}$ with $\alpha \in \mathbb{C}$

$$\sum_{n \in \mathbb{Z}} c_n \varphi(t - n) = e^{\alpha t}, \tag{3}$$

which holds if and only if the sampling kernel satisfies the generalized Strang—Fix conditions,

$$\widehat{\varphi}(\alpha) \neq 0 \text{ and } \widehat{\varphi}(\alpha + 2i\pi l) = 0 \text{ for } l \in \mathbb{Z} \setminus \{0\}, \tag{4}$$

where $\widehat{\varphi}(\alpha)$ represents the bilateral Laplace transform of $\varphi(t)$; i.e., $\widehat{\varphi}(s) = \int_{-\infty}^{\infty} e^{-st} \varphi(t) dt$ at $s = \alpha$. In this case, the coefficients are given by

$$c_n = \left\langle e^{\alpha t}, \widetilde{\varphi}(t - n) \right\rangle = e^{\alpha n} \widehat{\widetilde{\varphi}}(-\alpha), \tag{5}$$

where $\widetilde{\varphi}(t)$ forms a biorthonormal set with $\varphi(t)$ [9]. Recently, it has been shown that these conditions can be relaxed to include arbitrary sampling kernels in which case (3) becomes an approximation problem [16].

B. Mapping Signal Samples to FRI Samples

Once the signal samples y_n, are available by a sampling kernel that satisfies (3), we first find the exponential reproducing mapping for a set of complex exponentials of $\alpha_m = \alpha_0 + m\alpha \in \mathbb{C}$ by imposing

$$\sum_n c_{m,n} \varphi(t - n) = e^{\alpha_m t}, \; m = 0, \ldots, M - 1 \tag{6}$$

to retrieve the coefficients $c_{m,n}$ that allow a mapping from signal samples to FRI samples. Next, we obtain a new sequence of samples that we name as *FRI samples* by linearly combining the signal samples y_n with the coefficients $c_{m,n}$:

$$s_m = \sum_n c_{m,n} y_n. \tag{7}$$

Then, using (2), we have

$$s_m = \sum_n c_{m,n} \left\langle x(t), \varphi\left(\frac{t}{T} - n\right) \right\rangle \tag{8}$$

$$= \left\langle x(t), \sum_n c_{m,n} \varphi\left(\frac{t}{T} - n\right) \right\rangle \tag{9}$$

$$= \left\langle x(t), e^{\alpha_m \frac{t}{T}} \right\rangle, \tag{10}$$

which states that the FRI samples are indeed the exponential moments of the FRI signal. Finally, using (1), the FRI samples will be

$$s_m = \sum_{k=0}^{K-1} a_k e^{\alpha_m \frac{t_k}{T}} \tag{11}$$

$$= \sum_{k=0}^{K-1} \tilde{a}_k u_k^m, \tag{12}$$

where $\tilde{a}_k = a_k e^{\alpha_0 \frac{t_k}{T}}$ and $u_k = e^{\alpha \frac{t_k}{T}}$. Note that the choice $\alpha_m = \alpha_0 + m\alpha \in \mathbb{C}$ and the form of the FRI signal allow us to write the FRI samples s_m in a power sum form.

C. Reconstruction of FRI Parameters

The mapping step transforms the reconstruction of FRI signal into a parameter estimation problem that is similar to spectral estimation. Now, the pairs of unknowns $\{a_k, t_k\}_{k=0}^{K-1}$ can be retrieved from the FRI samples s_m using the Prony's method [33]. This is done by defining a filter h with z-transform

$$H(z) = \sum_{m=0}^{K} h_m z^{-m} = h_0 \prod_{k=0}^{K-1} (1 - u_k z^{-1}), \quad (13)$$

where the roots corresponds to the values u_k in (12). Then, it follows that h_m annihilates the sequence s_m as:

$$(h * s)_m = \sum_{i=0}^{K} h_i s_{m-i} \quad (14)$$

$$= \sum_{k=0}^{K-1} \tilde{a}_k u_k^m \underbrace{\sum_{i=0}^{K} h_i u_k^{-i}}_{H(u_k)=0} = 0. \quad (15)$$

In matrix/vector representation, (14) can be written as

$$\mathbf{S}\mathbf{h} = \mathbf{0}, \quad (16)$$

where $\mathbf{S}$ is a rank-deficient Toeplitz matrix with $[\mathbf{S}]_{mi} = s_{m-i}$, $\mathbf{h}$ admits a solution in the nullspace of $\mathbf{S}$ if $\text{rank}(\mathbf{S}) = K$ for noiseless samples and we name (16) as the annihilation system. However, in general, the samples in (2) are not ideal and degraded by noise,

$$\tilde{y}_n = y_n + \epsilon_n, \quad (17)$$

where ϵ_n is assumed to be i.i.d. AWGN, with zero mean and variance, σ^2. Therefore, the system in (16) cannot be satisfied exactly since the FRI samples (7) will be also corrupted by noise,

$$\tilde{s}_m = s_m + \tilde{e}_m, \quad (18)$$

where $\tilde{e}_m$ is the filtered noise given by $\tilde{e}_m = \sum_n c_{m,n} \epsilon_n$. Hence, the matrix $\mathbf{S}$ is now perturbed by noise

$$\tilde{\mathbf{S}} = \mathbf{S} + \mathbf{E}, \quad (19)$$

where the $\mathbf{E}$ is the Toeplitz matrix corresponding to the filtered noise $\tilde{e}_m$ in (18). Hence, the Annihilating Filter (AF) of [6] looks for total least squares (TLS) solution that minimizes $\|\tilde{\mathbf{S}}\mathbf{h}\|$ under the constraint that $\|\mathbf{h}\|^2 = 1$. In this case, the solution to $\mathbf{h}$ is the eigenvector that corresponds to the smallest eigenvalue of $\tilde{\mathbf{S}}^H \tilde{\mathbf{S}}$.

One way to control the effects of noise in the measurements is by studying the rank deficiency property of the matrix $\mathbf{S}$, i.e., ideally $\text{rank}(\mathbf{S}) = K$. For this purpose, the classical AF method can be improved by denoising $\tilde{\mathbf{S}}$ using Cadzow algorithm before applying TLS [6]. The algorithm is based on the fact that

the matrix $\tilde{\mathbf{S}}$ becomes full rank in the presence of noise and it has a Toeplitz matrix structure by construction. Hence, matrix $\tilde{\mathbf{S}}$ can be denoised by iteratively imposing these two properties until convergence. Moreover, this method can be also used to estimate the model order K by observing the separation of singular values of the matrix $\tilde{\mathbf{S}}$ [6], [15].

III. FRI MODEL FITTING

FRI signals only carry a finite number of innovations. For example, a stream of K Diracs in (1) is completely determined from $2K$ parameters being the innovation instants t_k and the innovations weights a_k. Hence, the performance of FRI methods is measured based on the estimation quality of these parameters. However, this may not be reliable as it depends on the model order K that is an internal parameter of the estimation. Instead, we propose a novel FRI reconstruction using a model fitting approach based on minimizing the error between the measured FRI samples and the recovered FRI samples. Consequently, this approach allows to define a criterion to monitor the reconstruction quality and adopt the model order accordingly.

The goal of FRI reconstruction methods in Section II was to first denoise the matrix $\tilde{\mathbf{S}}$ before looking for a solution to the annihilation system in (16). As mentioned before, these methods are highly sensitive to noise. Moreover, at low SNR, the estimation of the model order K becomes unreliable. Here, we propose a novel FRI reconstruction to overcome these problems using a model fitting approach based on structured—TLS (STLS) problem for affinely structured matrices as proposed by De Moor [30]. For this purpose, we first revisit the fundamental concepts of the STLS problem by developing an equivalent formulation for the FRI framework, and then we propose several FRI reconstructions based on STLS.

Considering the FRI model fitting problem, assume that we have M FRI samples to fit an FRI model with K non-linear parameters. This problem can be reformulated as

$$\min_{\mathbf{s},\mathbf{h}} \|\tilde{\mathbf{s}} - \mathbf{s}\|^2, \text{ subject to } \mathbf{S}(\mathbf{s})\mathbf{h} = 0 \text{ and } \|\mathbf{h}\|^2 = 1, \quad (20)$$

where $\tilde{\mathbf{s}}, \mathbf{s} \in \mathbb{C}^{M \times 1}$ are the measured and recovered FRI sample vectors, $\mathbf{h} \in \mathbb{R}^{K+1}$ is the filter coefficients vector in (13) and $\mathbf{S}(\mathbf{s}) \doteq \sum_{i=0}^{K} s_i \mathbf{S}_i$ with $\mathbf{S}_i \in \mathbb{C}^{(M-K) \times (K+1)}$ being fixed indicator matrices to represent the Toeplitz structure of $\mathbf{S}$. Note that we write the matrix $\mathbf{S}$ intentionally as $\mathbf{S}(\mathbf{s})$ to emphasize the linearity of $\mathbf{S}(\mathbf{s})$ with regards to $\mathbf{s}$. Hence, it is particularly interesting to define the left and the right dual matrices $\mathbf{L}(\cdot)$ and $\mathbb{R}(\cdot)$ of $\mathbf{S}$ for further development.

Lemma 1: Given the linearity of $\mathbf{S}(\mathbf{s})$ with respect to $\mathbf{s}$, the left and the right dual matrices $\mathbf{L}(\cdot)$ and $\mathbb{R}(\cdot)$ are defined respectively by

$$\left. \begin{aligned} \mathbf{g}^H \mathbf{S}(\mathbf{s})\mathbf{h} &= \mathbf{g}^H (\mathbf{R}(\mathbf{h})\mathbf{s}) \\ &= (\mathbf{L}(\mathbf{g})\mathbf{s})^H \mathbf{h} \end{aligned} \right\} \quad \text{for all } \mathbf{g}, \mathbf{h} \text{ and } \mathbf{s},$$

which yields the following identities

$$\mathbf{S}(\mathbf{s})\mathbf{h} = \mathbf{R}(\mathbf{h})\mathbf{s}, \quad (21)$$

$$\mathbf{S}(\mathbf{s})^H \mathbf{g} = \mathbf{L}(\mathbf{g})\mathbf{s}, \quad (22)$$

$$\mathbf{L}(\mathbf{g})^H \mathbf{h} = \mathbf{R}(\mathbf{h})^H \mathbf{g}. \quad (23)$$

Proof: The first two identities are straightforward. The last one can be obtained from

$$\Re\left\{\mathbf{g}^H\mathbf{S}(\mathbf{s})\mathbf{h}\right\} = \Re\left\{\mathbf{s}^H\mathbf{L}(\mathbf{g})\mathbf{h}\right\}$$
$$= \Re\left\{\mathbf{g}^H\mathbf{R}(\mathbf{h})\mathbf{s}\right\}$$
$$= \Re\left\{\mathbf{s}^H\mathbf{R}(\mathbf{h})^H\mathbf{g}\right\},$$

which is satisfied for all $\mathbf{s}$ and where $\Re\{\cdot\}$ specifies the real part of its argument.

Theorem 1: The equations satisfied by the optimal solutions of the problem (20) are

$$\begin{cases} \mathbf{R}(\mathbf{h})^H\mathbf{u} = \tilde{\mathbf{s}} - \mathbf{s}, \\ \mathbf{L}(\mathbf{u})\mathbf{s} = 0, \\ \mathbf{R}(\mathbf{h})\mathbf{s} = 0, \\ \|\mathbf{h}\|^2 = 1, \end{cases} \quad (24)$$

where the unknowns are the $(N - K) \times 1$ vector $\mathbf{u}$, the $(K + 1) \times 1$ vector $\mathbf{h}$ and the $M \times 1$ vector $\mathbf{s}$.

Proof: We start by replacing the original problem (20) with the following unconstrained minimization problem

$$\min_{\mathbf{s},\mathbf{h}} \underbrace{\|\tilde{\mathbf{s}} - \mathbf{s}\|^2 + 2\Re\{\mathbf{u}^H\mathbf{S}\mathbf{h}\} + \lambda\left\{\|\mathbf{h}\|^2 - 1\right\}}_{\mathcal{L}\{\mathbf{s},\mathbf{h},\mathbf{u},\lambda\}}, \quad (25)$$

where $\mathcal{L}$ is the Lagrangian to be minimized, $\mathbf{u}$ is a vector Lagrange multiplier for the constraints $\mathbf{S}(\mathbf{s})\mathbf{h} = 0$ and λ is a scalar Lagrange multiplier for the constraint $\|\mathbf{h}\|^2 = 1$.

The solution is then obtained by setting all the partial derivatives of $\mathcal{L}\{\mathbf{s},\mathbf{h},\mathbf{u},\lambda\}$ with respect to $\mathbf{s},\mathbf{h},\mathbf{u},\lambda$ to 0

$$\frac{\partial\mathcal{L}\{\mathbf{s},\mathbf{h},\mathbf{u},\lambda\}}{\partial\mathbf{s}} = 0 \Leftrightarrow \mathbf{R}(\mathbf{h})^H\mathbf{u} + \mathbf{s} - \tilde{\mathbf{s}} = 0,$$

$$\frac{\partial\mathcal{L}\{\mathbf{s},\mathbf{h},\mathbf{u},\lambda\}}{\partial\mathbf{h}} = 0 \Leftrightarrow \mathbf{S}(\mathbf{s})^H\mathbf{u} + \lambda\mathbf{h} = 0,$$

$$\frac{\partial\mathcal{L}\{\mathbf{s},\mathbf{h},\mathbf{u},\lambda\}}{\partial\mathbf{u}} = 0 \Leftrightarrow \mathbf{S}(\mathbf{s})\mathbf{h} = 0,$$

$$\frac{\partial\mathcal{L}\{\mathbf{s},\mathbf{h},\mathbf{u},\lambda\}}{\partial\boldsymbol{\lambda}} = 0 \Leftrightarrow \|\mathbf{h}\|^2 = 1.$$

Here, the application of the constraints, i.e., the last two lines, shows that $\lambda = 0$. Finally, the use of the duality relations leads to the system of four (24). $\square$

The next step of the STLS framework is the elimination of the unknown $\mathbf{s}$ to further reduce the problem to two unknowns. For that, the unknown $\mathbf{s}$ can be readily obtained from the first equation of (24), and subsequently replaced in the other three equations. Using the identities from Lemma 1, the optimality equations restricted to the two unknowns $\mathbf{u}$ and $\mathbf{h}$ are

$$\begin{cases} \mathbf{L}(\mathbf{u})\mathbf{L}(\mathbf{u})^H\mathbf{h} - \mathbf{S}(\tilde{\mathbf{s}})^H\mathbf{u} = 0, \\ -\mathbf{S}(\tilde{\mathbf{s}})\mathbf{h} + \mathbf{R}(\mathbf{h})\mathbf{R}(\mathbf{h})^H\mathbf{u} = 0, \\ \|\mathbf{h}\|^2 = 1, \end{cases} \quad (26)$$

while the third unknown is given by $\mathbf{s} = \tilde{\mathbf{s}} - \mathbb{R}(\mathbf{h})^H\mathbf{u}$. Note that it is possible to check

$$\mathbf{u}^H\mathbf{D_h}\mathbf{u} = \left\|\mathbf{R}(\mathbf{h})^H\mathbf{u}\right\|^2 = \left\|\mathbf{L}(\mathbf{u})^H\mathbf{h}\right\|^2 = \mathbf{h}^H\mathbf{D_u}\mathbf{h},$$

given that

$$\mathbf{D_u} = \mathbf{L}(\mathbf{u})\mathbf{L}(\mathbf{u})^H \text{ and } \mathbf{D_h} = \mathbf{R}(\mathbf{h})\mathbf{R}(\mathbf{h})^H.$$

Hence, by replacing $\mathbf{h}$ with a non-unit vector $\mathbf{v}$ according to $\mathbf{h} = \tau\mathbf{v}$, the system (26) is essentially the same as the one in [30] except that, since $\mathbf{h}$ is not normalized anymore, one degree of freedom is left undetermined.

The nonlinear system of equations (26) cannot be solved directly for $\mathbf{u}$ and $\mathbf{h}$, and it is necessary to resort to an iterative procedure where both $\mathbf{u}$ and $\mathbf{h}$ are updated [30]. Here, we propose a simplified alternative being equipped with the analysis of the STLS. We first note that the $\mathbf{u}$ of the Theorem 1 can be readily obtained from the first and the third equations of (24)

$$\mathbf{u} = \mathbf{D_h}^{-1}\mathbf{R}(\mathbf{h})\tilde{\mathbf{s}}. \quad (27)$$

Then, using the Lemma 1 and Theorem 1, we can obtain

$$\begin{aligned} \|\tilde{\mathbf{s}} - \mathbf{s}\|^2 &= \left\|\mathbf{R}(\mathbf{h})^H\mathbf{u}\right\|^2 \\ &= \left\|\mathbf{R}(\mathbf{h})^H\mathbf{D_h}^{-1}\mathbf{R}(\mathbf{h})\tilde{\mathbf{s}}\right\|^2 \\ &= \tilde{\mathbf{s}}^H\mathbf{R}(\mathbf{h})^H\mathbf{D_h}^{-1}\mathbf{R}(\mathbf{h})\mathbf{R}(\mathbf{h})^H\mathbf{D_h}^{-1}\mathbf{R}(\mathbf{h})\tilde{\mathbf{s}} \\ &= \tilde{\mathbf{s}}^H\mathbf{R}(\mathbf{h})^H\mathbf{D_h}^{-1}\mathbf{R}(\mathbf{h})\tilde{\mathbf{s}} \\ &= \mathbf{h}^H\mathbf{S}(\tilde{\mathbf{s}})^H\mathbf{D_h}^{-1}\mathbf{S}(\tilde{\mathbf{s}})\mathbf{h}. \end{aligned}$$

Therefore, rather than minimizing (26) with respect to $\mathbf{u}$ and $\mathbf{h}$, we propose to minimize the following quadratic form with respect to $\mathbf{h}$ only,

$$\min_{\mathbf{h}} \mathbf{h}^H\mathbf{S}(\tilde{\mathbf{s}})^H\mathbf{D_h}^{-1}\mathbf{S}(\tilde{\mathbf{s}})\mathbf{h} \text{ subject to } \mathbf{h} \in \Theta, \quad (28)$$

where Θ is a constraint set necessary to avoid the trivial solution. However, since the FRI problem only requires the roots of the polynomial defined by the coefficients $\mathbf{h}$, neither the solution to this problem is unique nor there exists only one constraint that will lead to the same solution. We will further discuss about the possible conditions on $\mathbf{h}$ and their effect on the solution of the problem later on.

Next, we need to define the operator $\mathbb{R}(\mathbf{h})$ to proceed for the FRI model-fitting framework. Indeed, the dual of the annihilation (14) can be written using commutativity of the convolution as

$$(s * h)_m = \sum_{i=0}^M s_i h_{m-i} = 0, \quad (29)$$

which can be written in matrix vector form

$$\mathbf{R}(\mathbf{h})\mathbf{s} = \mathbf{0}, \quad (30)$$

where $\mathbb{R}$ is a rank-deficient Toeplitz matrix with $[\mathbb{R}]_{mi} = h_{m-i}$.

A. Constraints on the Filter Coefficients

In this section, we consider possible cases for the constraint set Θ defined in (28) to avoid the trivial zero solution for the filter coefficients $\mathbf{h}$.

1) Quadratic Norm Constraint (MF-1): We consider a quadratic norm constraint such as $\|\mathbf{h}\|^2 = 1$ for (28). We first

note that this is an appropriate condition since the coefficients of the filter in (13), i.e., $h_0, \ldots, h_K$, can always be scaled by a constant such as $1/\|\mathbf{h}\|$ without altering the roots of filter. With this condition, we want to minimize

$$\min_{\mathbf{h}} \mathbf{h}^H \mathbf{S}(\tilde{\mathbf{s}})^H \mathbf{D}_{\mathbf{h}}^{-1} \mathbf{S}(\tilde{\mathbf{s}})\mathbf{h} \text{ subject to } \|\mathbf{h}\|^2 = 1, \qquad (31)$$

where the term $\mathbf{D}_{\mathbf{h}}$ also depends on $\mathbf{h}$. Hence, the problem cannot be minimized directly and we propose an iterative scheme, also known as iterative quadratic maximum likelihood (IQML) algorithm [34]–[37], assuming the term $\mathbf{D}_{\mathbf{h}}$ remains constant at each iteration of the solution and is obtained using the previous estimate of the solution $\mathbf{h}$. Hence, the estimation of the filter coefficients is implemented by iteratively solving

$$\min_{\mathbf{h}^{(i)}} \mathbf{h}^{(i)} \mathbf{S}(\tilde{\mathbf{s}})^H \left(\mathbf{D}_{\mathbf{h}}^{-1}\right)^{(i-1)} \mathbf{S}(\tilde{\mathbf{s}})\mathbf{h}^{(i)} \text{ subject to } \left\|\mathbf{h}^{(i)}\right\|^2 = 1, \qquad (32)$$

where $(\mathbf{D}_{\mathbf{h}}^{-1})^{(i-1)} = (\mathbf{R}(\mathbf{h}^{(i-1)})\mathbf{R}(\mathbf{h}^{(i-1)})^H)^{-1}$ and $\mathbf{h}^{(0)}$ randomly initialized. Then, the solution $\mathbf{h}^{(i)}$ involves finding the eigenvector corresponding to the minimum eigenvalue of $\mathbf{X}_{\mathbf{h}}^{(i)} = \mathbf{S}(\tilde{\mathbf{s}})^H \mathbf{D}_{\mathbf{h}}^{(i-1)} \mathbf{S}(\tilde{\mathbf{s}})$ at each iteration. We denote this method as model fitting 1 (MF-1) and provide the details in Algorithm 1.

2) Linear Constraint (MF-2): We consider a linear constraint such that one element of the $\mathbf{h}$ to be 1. Specifically, we note that scaling the coefficients of the filter (13), i.e., $h_0, \ldots, h_K$ with respect to h_K will not alter the roots of the filter. With this condition, we want to minimize

$$\min_{\mathbf{h}} \mathbf{h}^H \mathbf{S}(\tilde{\mathbf{s}})^H \mathbf{D}_{\mathbf{h}}^{-1} \mathbf{S}(\tilde{\mathbf{s}})\mathbf{h} \text{ subject to } \mathbf{e}_{K+1}^T \mathbf{h} = 1, \qquad (33)$$

where $\mathbf{e}_{K+1}$ is a length-$K + 1$ discrete impulse vector with "1" in position $K + 1$ and "0" elsewhere. Here the problem cannot be minimized directly and the term $\mathbf{D}_{\mathbf{h}}$ depends on $\mathbf{h}$. Hence, we propose to use the IQML algorithm [34]–[37] assuming the term $\mathbf{D}_{\mathbf{h}}$ remains constant at each iteration. Hence, the estimation of the filter coefficients is implemented by iteratively solving

$$\min_{\mathbf{h}^{(i)}} \mathbf{h}^{(i)} \mathbf{S}(\tilde{s})^H \left(\mathbf{D}_{\mathbf{h}}^{-1}\right)^{(i-1)} \mathbf{S}(\tilde{s})\mathbf{h}^{(i)} \text{ subject to } \mathbf{e}_{K+1}^T \mathbf{h} = 1, \qquad (34)$$

where $(\mathbf{D}_{\mathbf{h}}^{-1})^{(i-1)} = (\mathbf{R}(\mathbf{h}^{(i-1)})\mathbf{R}(\mathbf{h}^{(i-1)})^H)^{-1}$ and $\mathbf{h}^{(0)}$ randomly initialized. Then, the solution update is given by

$$\mathbf{h}^{(i+1)} = \frac{\left(\mathbf{X}_{\mathbf{h}}^{(i)}\right)^{-1} \mathbf{e}_{K+1}}{\mathbf{e}_{K+1}^T \left(\mathbf{X}_{\mathbf{h}}^{(i)}\right)^{-1} \mathbf{e}_{K+1}}, \qquad (35)$$

where $\mathbf{X}_{\mathbf{h}}^{(i)} = \mathbf{S}(\tilde{\mathbf{s}})^H \mathbf{D}_{\mathbf{h}}^{(i-1)} \mathbf{S}(\tilde{\mathbf{s}})$ as before and $\mathbf{h}^{(0)}$ randomly initialized. We name this method as model fitting 2 (MF-2) and provide the details in Algorithm 1.

3) Randomized (Linear) Constraint (MF-3): We consider a randomized (linear) constraint based on the randomly initialized solution such that the inner product with $\mathbf{h}$ and the initial estimation remains 1; i.e., $(\mathbf{h}^{(0)})^H \mathbf{h} = 1$. Hence, we want to minimize the following problem

$$\min_{\mathbf{h}} \mathbf{h}^H \mathbf{S}(\tilde{\mathbf{s}})^H \mathbf{D}_{\mathbf{h}}^{-1} \mathbf{S}(\tilde{\mathbf{s}})\mathbf{h} \text{ subject to } \left(\mathbf{h}^{(0)}\right)^H \mathbf{h} = 1, \qquad (36)$$

which cannot be minimized directly and the term $\mathbf{D}_{\mathbf{h}}$ depends on $\mathbf{h}$. Similarly, we propose to use IQML algorithm [34]–[37] assuming the term $\mathbf{D}_{\mathbf{h}}$ remains constant at each iteration. Consequently, the estimation of the filter coefficients is implemented by iteratively solving

$$\min_{\mathbf{h}^{(i)}} \mathbf{h}^{(i)} \mathbf{S}(\tilde{\mathbf{s}})^H \left(\mathbf{D}_{\mathbf{h}}^{-1}\right)^{(i-1)} \mathbf{S}(\tilde{\mathbf{s}})\mathbf{h}^{(i)} \text{ subject to } \left(\mathbf{h}^{(0)}\right)^H \mathbf{h}^{(i)} = 1, \qquad (37)$$

where $(\mathbf{D}_{\mathbf{h}}^{-1})^{(i-1)} = (\mathbf{R}(\mathbf{h}^{(i-1)})\mathbf{R}(\mathbf{h}^{(i-1)})^H)^{-1}$ and $\mathbf{h}^{(0)}$ randomly initialized. Then, the solution update is

$$(\mathbf{h})^{(i+1)} = \frac{\left(\mathbf{X}_{\mathbf{h}}^{(i)}\right)^{-1} \mathbf{h}^{(0)}}{\left(\mathbf{h}^{(0)}\right)^H \left(\mathbf{X}_{\mathbf{h}}^{(i)}\right)^{-1} \mathbf{h}^{(0)}}, \qquad (38)$$

where we note that the effect of the random initialization will remain (only) in this case due to the imposed constraint. We express this method as model fitting 3 (MF-3) and provide the details in Algorithm 1.

Finally, we note that the methods MF-3 and MF-2 will be equivalent if the filter coefficients are initialized such that $\mathbf{h}^{(0)} = \mathbf{e}_{K+1}$. Hence, it is clear that MF-3 has a certain flexibility compared to MF-2 in updating the filter coefficients as seen in (35) and (38).

A Note on the Choice of the Constraint: The problem in (28) can be solved under different constraints. In order to better understand the underlying differences among the proposed constraints, we consider a geometric interpretation. As stated before, the FRI reconstruction requires the estimation of the K roots of the polynomial defined by the unknown $K + 1$ coefficients. The solution points lie on a line in this $K + 1$ dimensional space since any scaling of the filter coefficients forms an equally valid solution. In this representation, the constraint of MF-1 is a unit hypersphere where the MF-1 solution set is limited to the intersection between the hypersphere and the solution points line. Hence, there are two drawbacks of MF-1, due to the quadratic constraint: first, the solution set has two intersection points, i.e., no unique solution, and second the coefficient update is done on the unit hypersphere, which requires a solution of eigenvalue problem at each iteration. On the other hand, the linear constraints (MF-2 and MF3) represent a hyperplane, and the solution set is limited to the unique intersection point between this hyperplane and the solution points line. Moreover, the coefficient update is also done on this hyperplane, which allows a linear update term; i.e., the hyper plane yield simpler surface than the unit hypersphere. Specifically, the constraint of MF-2 is the hyperplane that is tangent to the unit hyperplane (i.e., the surface defined by the constraint of MF-1) whereas the constraint of MF-3 is the hyperplane that is tangent to a hypersphere of radius $\|\mathbf{h}^{(0)}\|$ at $\mathbf{h}^{(0)}$. Hence, where both constraint uses a simpler update term on a plane, the constraint of MF-3 has another advantage over the MF-2: the update plane can be randomly initialized rather than using a fixed plane defined by $\mathbf{e}_{K+1}^T \mathbf{h} = 1$.

B. FRI with Model Fitting Algorithm

We propose a full FRI reconstruction algorithm integrating the model fitting approach. Once we have the FRI samples $\tilde{\mathbf{s}}$, we first construct $\mathbf{S}(\tilde{\mathbf{s}})$ and $\mathbf{D_h}$ using randomly initialized $\mathbf{h}^{(0)}$. Then, choosing any of the FRI model fitting approaches (MF-1), (MF-2) or (MF-3), we iteratively obtain the filter coefficients $\mathbf{h}^{(i)}$ to retrieve $u_k^{(i)}$ by finding the roots of the filter in (13). We then obtain the locations $t_k^{(i)}$ using the solution $u_k^{(i)}$, since $u_k = e^{\alpha \frac{t_k}{T}}$ from (11). By construction, the zeros of the filter $\mathbf{h}$ uniquely define the values u_k provided that the instants t_k's are distinct.

In the next step, the amplitudes a_k can be determined either by solving first K consecutive equations of (2) or (10) with the estimated locations t_k. Specifically, given the locations $t_k^{(i)}$, we determine the amplitudes $a_k^{(i)}$ by solving K consecutive equations of (12). Indeed, we propose to minimize the following complex-valued least-squares problem constrained to the real solution

$$\min_{\mathbf{a}^{(i)}} \left\| \mathbf{B}\mathbf{a}^{(i)} - \mathbf{s} \right\|^2 \text{ subject to } \mathbf{a}^{(i)} \in \mathbb{R}^{K \times 1}, \quad (39)$$

where $\mathbf{B} \in \mathbb{C}^{M \times K}$ is a complex matrix with entries $[\mathbf{B}]_{mk} = e^{\alpha_m t_k^{(i)}/T}$ and $\mathbf{s} \in \mathbb{C}^{M \times 1}$ is the complex-moment vector. As a convergence test, we compute the recovered FRI samples as

$$\hat{s}_m^{(i)} = \sum_{k=0}^{K-1} a_k^{(i)} e^{\frac{\alpha_m t_k^{(i)}}{T}} \quad (40)$$

and check the numerical convergence by

$$\frac{\sum_{m=0}^{M-1} \hat{s}_m^{(i)} - \hat{s}_m^{(i-1)}}{\sum_{m=0}^{M-1} \hat{s}_m^{(i-1)}} \leq \varepsilon$$

where ε is a chosen threshold. Although we have no proof of convergence of this algorithm, numerical experiments demonstrate that convergence is typically reached after 10 iterations.

It is important to realize that we look for the convergence in terms of fitting quality in FRI samples, but not the filter coefficients. One could have chosen the convergence criterion based on the filter coefficients which would not require to extract the parameters of the signal in each iteration. However, we have observed experimentally the first condition converges earlier than the second one, especially with increasing noise levels.

We note that the filter coefficients $\mathbf{h}$ are initialized uniformly in a $K + 1$ dimensional unit hypercube. Hence, each element of $\mathbf{h}$ is uniformly distributed between 0 and 1 forming the unit hypercube. It is important to state that the solution might not converge for some realizations of the random initialization under the linear constraint. For example, consider the case for MF-3 when the initialization is realized orthogonal to the solution, i.e., $\mathbf{h}^{(0)}\mathbf{h}^* = 0$ where $\mathbf{h}^*$ is the solution to the problem. In this case, the solution update cannot converge to the solution due to the initialization. However, this event is very unlikely to take place especially in higher dimensions. Nevertheless, the convergence of the algorithm can be monitored and in case of a failure of convergence after certain number of iterations, the algorithm can be restarted with a different initialization.

Algorithm 1: FRI Model Fitting

Require: Inputs: $\tilde{\mathbf{s}}, K > 0$

1: Initialization: Choose $\mathbf{h}^{(0)}$ and construct $\mathbf{D_h}$ and $\mathbf{X_h}$

2: **for** $i = 0$ **till convergence do**

3:     $\mathbf{h}^{(i)} \leftarrow$ Solution to either (31) or (33) or (36)

4:     $u_k^{(i)} \leftarrow$ Find the roots of the filter (13)

5:     $t_k^{(i)} \leftarrow T/\alpha \ln u_k^{(i)}$

6:     $a_k^{(i)} \leftarrow$ Solution to (39)

7:     $\hat{s}_m^{(i)} \leftarrow$ Recovered FRI samples (40)

8:     If converged stop, else go to 2

9: **end for**

10: **return** $\{t_k, a_k\}$

C. Model Order Selection

The estimation of the model order K is a challenging part of all FRI frameworks when the samples are corrupted by noise. Although there exist some methods to estimate the model order, the performance significantly depends on the amount of the noise. In this case, current model order estimation methods for FRI fail in particular at low SNR. Here, we propose a novel model order selection procedure that allows us to choose the model number K based on the noise level of the samples.

We start by noting that the annihilation system (16) can be solved if we have $M \geq 2K$ samples of s_m to solve for K innovations; i.e., we need at least as many equations as unknowns. This implies that M samples can be used to reconstruct at most $\frac{M}{2}$ innovations. Now, we define the model order selection problem as estimating the optimal model order K, which can take the values from a discrete set of $\hat{k} = [1, \ldots, \frac{M}{2}]$, for a given M samples of $\tilde{s}$ degraded by noise. For this purpose, we define a training error of the estimator

$$E_T(\hat{k}) = \left\| \hat{\mathbf{s}}(\hat{k}) - \tilde{\mathbf{s}} \right\|_2^2, \quad (41)$$

where $\hat{\mathbf{s}}(\hat{k})$ is the recovered FRI sample for model order $\hat{k}$ in (40) with the estimated parameters $\{\hat{t}_k, \hat{a}_k\}_{k \in \hat{k}}$. Notice that the training error curve E_T reveals a U-curve pattern in which the initial drop and final rise regimes represent an under-fitting and over-fitting behavior, respectively, and the flat region in the middle shows the optimal fitting region. We then define the fitting level as the minimum of the training error

$$\hat{\sigma}^2 = \min_{\hat{k}} E_T(\hat{k}), \quad (42)$$

which is a good estimator of the input noise level σ^2 for small amount of noise. Finally, the plausible model order is chosen at the intersection of under-fitting and fitting regions; i.e., minimum $\hat{k}$ that yields the sharpest decrease in the training error and achieves the fitting level in the range $(1 \pm 0.5) \hat{\sigma}^2$. Hence, this framework allows to determine the model order based on SNR level. We refer the reader to Appendix B for a detailed discussion on the training error.

Algorithm 2: FRI model order selection

Require $\tilde{\mathbf{s}} \in \mathbb{C}^{M \times 1}$

1: **for** $\hat{k} = 1$ to $\frac{M}{2}$
2:    $\{\hat{t}_k, \hat{a}_k\} \leftarrow$ Apply Algorithm 1
3:    $E_T(\hat{k}) \leftarrow$ Find the training error (41)
4: **end for**
5: $\hat{\sigma}^2 \leftarrow$ Find the fitting level (42)
6: $\bar{k} \leftarrow$ Model order selection (see Section III-C)
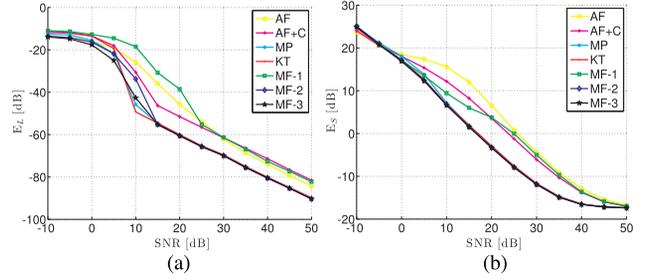7: **return** $\{\bar{k}, t_{\bar{k}}, a_{\bar{k}}\}$

Fig. 2. Comparison of FRI algorithms: Parameter estimation of a signal with $K = 5$ frequencies and $M = 4 \times K + 1$ for an average of $N = 1,000$ independent trials.

IV. NUMERICAL RESULTS AND EXPERIMENTS

We now present various simulation results to validate the proposed approach and demonstrate its practical feasibility. Specifically, we investigate the performance of the FRI model-fitting methods from Section III and compare them with the state-of-the-art FRI reconstruction algorithms introduced in Section II. Moreover, we show results for the model order selection framework developed in Section III-C.

We assume that an FRI signal (1) with K signal innovation, $\{t_k, a_k\}_{k=0}^{K-1}$, is sampled with a kernel that satisfies (3) as in Section II-A. Then, we assume that the signal samples are used to obtain the FRI samples using the FRI framework of Section II-B. Here, we specifically focus on the reconstruction of FRI signal innovations from FRI samples. Hence, we consider M FRI samples of

$$\tilde{s}_m = \underbrace{\sum_{k=1}^{K} a_k e^{i 2\pi t_k m}}_{s_m} + \tilde{e}_m, \quad m = 0, \ldots, M-1, \quad (43)$$

where $\tilde{e}_m = \tilde{e}_{\mathcal{R}m} + i \tilde{e}_{\mathcal{I}m}$ is complex additive white Gaussian noise with variance σ^2. The real and the imaginary parts of $\tilde{e}_m$ are uncorrelated and each has a variance of $\sigma^2/2$, i.e., $\tilde{e}_{\mathcal{R}m}, \tilde{e}_{\mathcal{I}m} \sim \mathcal{N}(0, \sigma^2/2)$, so that the covariance matrix of the noise is $\mathbf{R} = \mathrm{E}\{\tilde{\mathbf{e}}\tilde{\mathbf{e}}^H\} = \sigma^2 \mathbf{I}$. The signal-to-noise-ratio (SNR) is defined to be $\mathrm{SNR(dB)} = 10 \log_{10} \frac{\frac{1}{M} \sum_{m=0}^{M-1} (s_m)^2}{\sigma^2}$.

A. Comparison of FRI Algorithms

In the first part, we compare the performance of FRI model-fitting approach combined with three constraints (MF-1 to MF-3) developed in Section III with the most common methods in FRI reconstruction. Among these, the annihilating filter (AF) method and its extension with Cadzow (AF+C) enhancement were the first methods applied to FRI problem [6]. We also added Matrix Pencil (MP) method [20] and Kumaresan-Tufts (KT) algorithm [38] into our comparison [18], [19].

In the absence of noise, every method is able to recover the signal innovations exactly. However, as noise increases, the accuracy of each method significantly degrades. The FRI signal innovations are generated randomly such that we have K innovation instants between $t_k \in [0, 1]$ with innovation weights $a_k \sim \mathcal{N}(1, 0.1^2)$. Then, we take M samples of (43) degraded by noise. The variance is chosen according to the target SNR.

The signal is fully characterized by the innovation parameters: the instants $t_1, \ldots, t_k$ and the weights $a_1, \ldots, a_K$. For the numerical experiments, we define two metrics to facilitate comparisons: a normalized localization error in instant estimation, E_L, and a normalized error in FRI samples E_S

$$E_L = 10 \log_{10} \frac{1}{N} \sum_{n=1}^{N} \frac{\sum_k (\hat{t}_{n,k} - t_k)^2}{\sum_k (t_k)^2}, \quad (44)$$

$$E_S = 10 \log_{10} \frac{1}{N} \sum_{n=1}^{N} \frac{\sum_m |\hat{s}_{n,m} - s_m|^2}{\sum_m |s_m|^2}, \quad (45)$$

where $\hat{t}_{n,k}$ is the recovered instants, and $\hat{s}_{n,m}$ is the recovered FRI samples (40) in trial n, and N is the number of total independent trials. We further assume that the number of innovations is known and provided to each method so as to make a fair comparison.

We start with a particular realization where we have five innovations to recover with the parameters given as $\mathbf{t} = [0.30, 0.44, 0.56, 0.67, 0.88]^T$ and $\mathbf{a} = [0.914, 0.667, 1.025, 1.058, 0.770]^T$. In Fig. 2, we observe the plot of normalized localization and moments error with respect to the SNR level. Clearly, E_L, in Fig. 2(a), exhibits a thresholding effect with respect to SNR level due to the non-linear estimation step. Furthermore, we observe that three of the methods, namely, AF, AF with Cadzow, and the MF-1, differ in their asymptotic behavior for high SNR. They show significantly poor performance even for high SNR while Cadzow enhancement helps AF method to attain its asymptotic behavior at a lower SNR. In contrast, E_S, in Fig. 2(b), follows a rather smooth curve due to the compensation effect by the estimation of the weights as a solution to the complex least squares problem (39).

For the same setting, we now focus on the comparison of the performance of the methods on individual parameters. In Fig. 3, we observe the estimator's variance and the theoretical CRLB for each parameter. For the details of the derivation of the CRLB for complex AWGN, we refer the reader to Appendix A. Here, we show the observed standard deviation (over $N = 1,000$ independent realizations) compared to CRLB of each parameter. In Fig. 3, column (a), we observe the same thresholding effect for the innovation instants that reveals differences for different FRI algorithms. Note that MF-3, the model-fitting approach with randomized linear constraint, performs better for
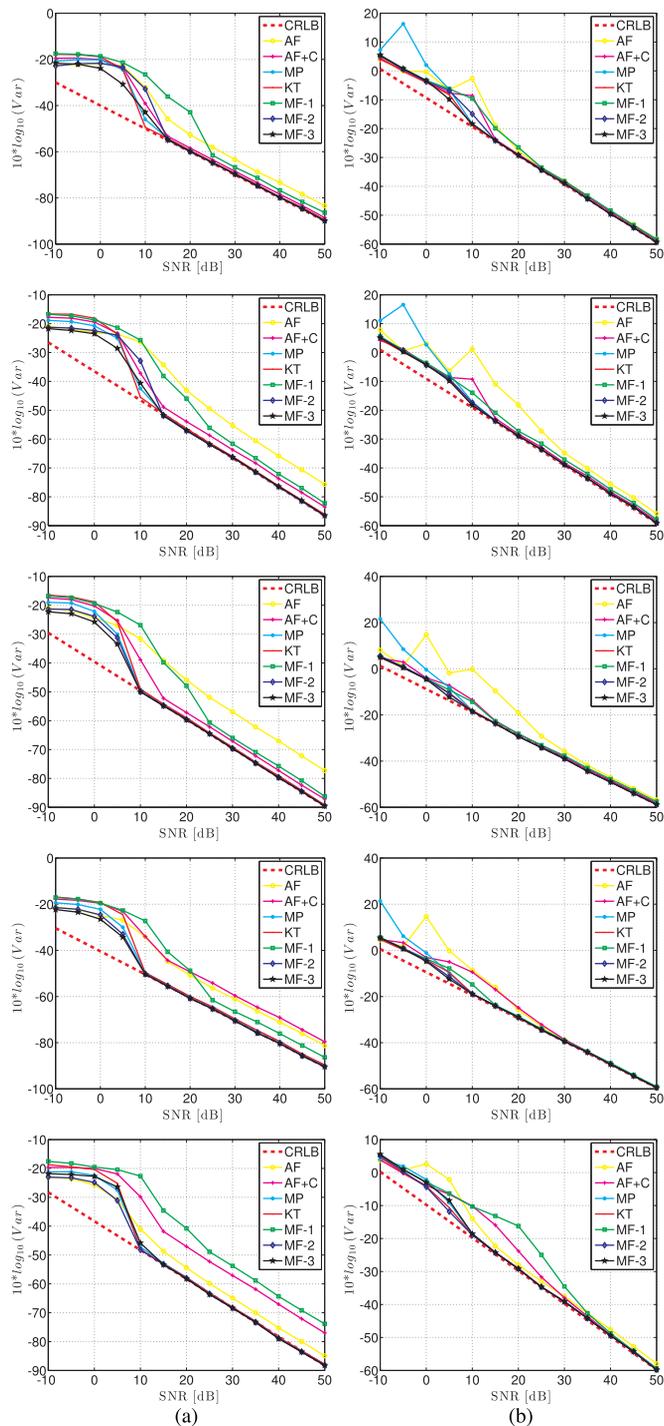
Fig. 3. Compare the performance of FRI algorithms based on estimator variance (Var) and CRLB with respect to each parameter for various SNR levels: Column (a) innovation instants t_1 to t_5 from top to bottom (b) innovation weights a_1 to a_5 from top to bottom. Parameter estimation of a signal with $K = 5$ frequencies and $M = 4 \times K + 1$ for an average of $N = 1,000$ independent trials.
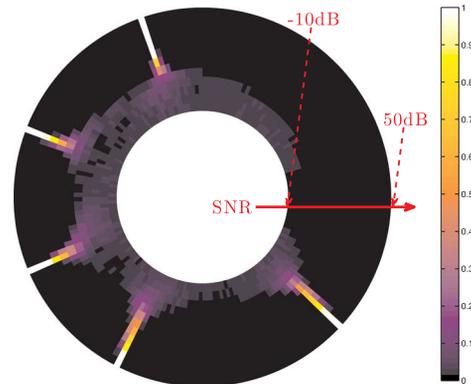

Fig. 4. SNR bagel: Visualization of frequency estimation for annihilating filter method, where each ring corresponds to a different SNR and the color represents probability of detection for each frequency averaged over $N = 1,000$ realization (see text for details).

SNR level, resolution and variance of the estimation. For this purpose, we design a visualization such that each trial of estimation for the innovation instant t_k can be mapped to a point on a ring with a phase given by $2\pi t_k$ and the radius of the ring determines the SNR level, hence termed an SNR bagel. The continuous instant t_k is mapped to $2\pi t_k$ with 1 degree radial resolution and the colorbar shows an average probability of detection for each innovation instant value. In Fig. 4, we observe the SNR bagel of the AF method where each ring corresponds to different SNR levels between -10 dB to 50 dB. For each SNR level, we compute the average over $N = 1,000$ independent realizations. Clearly, the uncertainty of the estimation drops as we move from inside to outside of the SNR bagel. Moreover, we also observe another interpretation of the CRLB, i.e., the varying thickness of the uncertainty cloud. We observe that in the low SNR region the estimations are spread reflecting the big variance on the estimation whereas the white narrow regions in high SNR show that the method achieved the minimum variance given by the CRLB. Moreover, we observe the thickness of the uncertainty region changes with respect to the spacing of the two instants which is also another interpretation of the non-diagonal entries of the inverse of the Fisher information matrix (49) in Appendix A.

With this new visualization, we can now observe the Fig. 5 and compare the performance of the known FRI reconstruction methods with our FRI model-fitting approach combined with three constraints. We observe that the MF-2, MF-3 and KT methods achieve better performance in the low SNR regime whereas all the methods performs similarly at high SNR. However, MF-1 clearly performs inferior with respect to the other two constraints of the model-fitting approach. Hence, we conclude linear constraints for our FRI model-fitting approach performs better than the quadratic constraint.

Next, we compare the FRI algorithms for varying model order K and SNR levels. For this experiment, we consider randomly generating $K = [3, \ldots, 10]$ innovations with instants in $t_k \in [0, 1]$ and random weights $a \sim \mathcal{N}(1, 0.1^2)$ using $M = 4 \times K + 1$ samples given by (43) degraded by complex AWGN as before. In Table I, we compare the performance of all FRI algorithms

the low SNR and together with MF-2, they achieve the theoretical limit given by the CRLM around 10 dB despite other methods attain this level around 20 dB. In contrast, the weights do not follow a general pattern expressing the compensation behavior the complex least squares step.

We next provide a different visualization of the innovation instants estimation to better illustrate the interaction between the
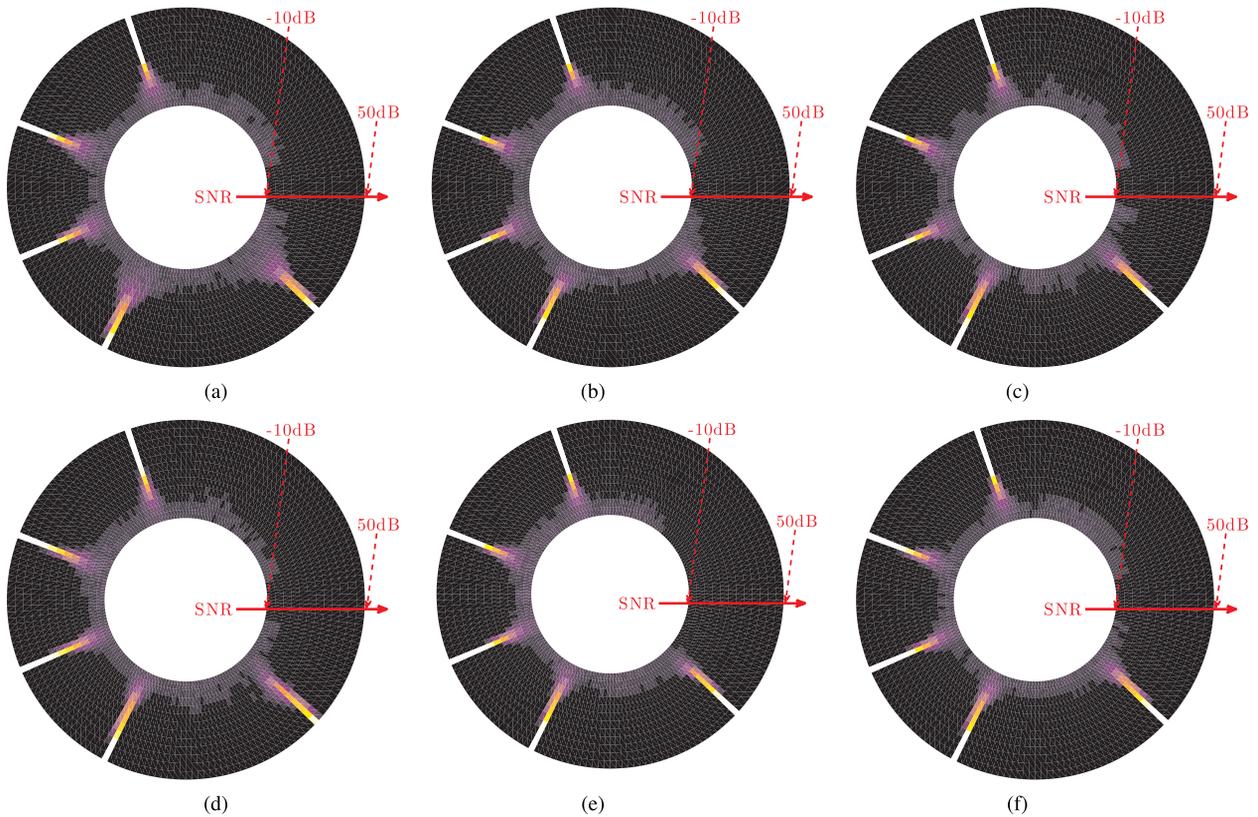
Fig. 5. Comparison of FRI algorithms on frequency estimation performance mapped to SNR Bagel: (a) Annihilating filter + Cadzow (b) Matrix Pencil (c) Kumaresan and Tufts (d) MF-1 (e) MF-2 (f) MF-3.

TABLE I
COMPARISON OF FRI ALGORITHMS FOR NORMALIZED LOCALIZATION ERROR E_L AVERAGED OVER $N = 1,000$ INDEPENDENT REALIZATIONS OF VARYING MODEL ORDER $K \in [3, \ldots, 10]$

	SNR(dB)				
	0	5	10	15	20
AF	-17.24	-21.99	-32.84	-41.69	-48.70
AF+C	-14.30	-18.38	-22.37	-26.13	-31.18
MP	-15.95	-22.47	-40.84	-48.13	-52.90
KT	-13.54	-19.19	-36.76	**-48.54**	-53.13
MF-1	-11.66	-12.99	-14.25	-14.10	-13.91
MF-2	-14.53	-16.20	-18.84	-25.02	-32.71
MF-3	**-17.29**	**-23.36**	**-42.44**	-48.49	**-53.83**

TABLE II
COMPARISON OF FRI ALGORITHMS FOR NORMALIZED FRI SAMPLE ERROR E_S AVERAGED OVER $N = 1,000$ INDEPENDENT REALIZATIONS OF VARYING MODEL ORDER $K \in [3, \ldots, 10]$

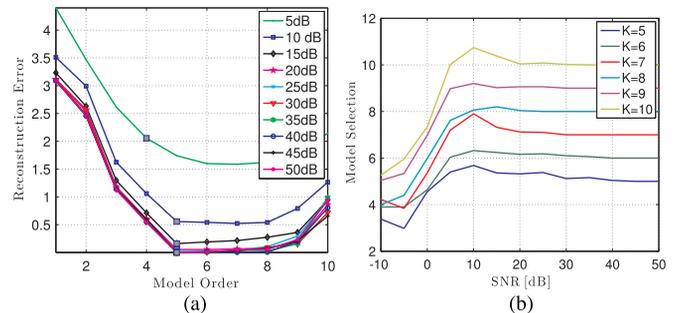|	SNR(dB)				
	0	5	10	15	20
AF	-0.41	-2.52	-5.77	-10.71	-16.53
AF+C	-0.84	-3.26	-5.70	-7.93	-10.40
MP	- -0.65	-4.40	-10.11	-15.91	-20.81
KT	**-1.75**	-6.17	-11.37	-16.20	-20.94
MF-1	-1.05	-3.96	-6.47	-7.61	-7.98
MF-2	-1.17	-4.42	-8.63	-14.24	-20.11
MF-3	-1.71	**-6.35**	**-11.51**	**-16.40**	**-21.36**

Fig. 6. FRI Model Selection: (a) Training error curve at different SNR for a true model order of $K = 5$, the selected order at each SNR level is given by the gray square (b) Model order selection for different true model order with respect to SNR levels an average of $N = 1,000$ independent trials.

B. Model Order Selection

Several applications of FRI require a robust model order selection as an integral part of the framework. In this section, we demonstrate the feasibility of the model order selection developed in Section III-C using the MF-3. We first observe the training error curve for the same setting designed in the previous part with $K = 5$ frequencies. In Fig. 6(a), we observe the corresponding U-curve, i.e., the training error curve (41), for each SNR level. Using Algorithm 2, the estimated model order is indicated as a gray square on the U-curves. Clearly, the model order selection algorithm successfully determines the true model order above 10 dB. Moreover, it allows to predict the best model selection for a low SNR-regime which results in reliable parameter estimation of the FRI signal.

with respect to normalized localization error in frequency estimation, E_L whereas in Table II we provide the comparison with respect to normalized FRI sample estimation error E_S. From both tables, we conclude that MF-3 outperforms the other algorithms in terms of innovation parameter estimation.

We further extend and demonstrate this SNR level dependent property of the model order selection framework in Fig. 6(b). Here, we consider randomly generating $K = [5, \ldots, 10]$ innovations in $t_k = [0, 1]$ with random weights $a \sim N(1, 0.1^2)$ using $M = 4 \times K + 1$ samples as before. We observe the average of the model selection over $N = 1'000$ independent trials. From each curve in Fig. 6(b), we clearly observe that the framework chooses a lower model order for low SNR (up to 10 dB) whereas it converges to the true model order for high SNR on the average. Hence, the model order selection framework displays robust performance with respect to model order K. We conclude that Algorithm 2 of Section III-C can be used for reliable estimation of the FRI signal innovations for a range of SNR.

V. CONCLUSION

We have considered the FRI reconstruction problem in presence of noise and proposed a novel reconstruction algorithm with model fitting approach and a model order selection that allows to estimate innovations even at very low SNR. Numerical results showed that the proposed algorithm with the linear constraint outperforms the other FRI reconstruction methods. This new method with model fitting approach would have a potential effect on applications with FRI signal degraded with realistic levels of measurements noise [8], [7], [39].

APPENDIX A
DERIVATION OF CRLB

The goal of the reconstruction problem given in (43) is to estimate the parameters $\Theta = [t_1, t_2, \ldots, t_K, a_1, a_2, \ldots, a_K]^T$ from a vector of M noisy samples $\tilde{s}_m = s_m + \tilde{e}_m$ where $\tilde{e}_m = \tilde{e}_{\mathcal{R}m} + i\tilde{e}_{\mathcal{I}m}$ is complex additive white Gaussian noise with variance σ^2. The real and the imaginary parts of $\tilde{e}_m$ are uncorrelated and each has a variance of $\sigma^2/2$, i.e., $\tilde{e}_{\mathcal{R}m}, \tilde{e}_{\mathcal{I}m} \sim \mathcal{N}(0, \sigma^2/2)$ so that the covariance matrix of the noise is $\mathbf{R} = \mathrm{E}\{\tilde{\mathbf{e}}\tilde{\mathbf{e}}^H\} = \sigma^2\mathbf{I}$. Now, assume that the PDF of $p(\tilde{\mathbf{s}}, \Theta)$ satisfies the regularity condition, i.e.,

$$\mathrm{E}\left[\frac{\partial \ln p(\tilde{\mathbf{s}}, \Theta)}{\partial \Theta}\right] = 0 \text{ for all } \Theta, \tag{46}$$

then, the variance of any unbiased estimator $\hat{\Theta}$ satisfies $\mathbf{C}_{\hat{\Theta}} - \mathbf{I_F}^{-1}(\Theta) \geq 0$, where $\mathbf{C}_{\hat{\Theta}}$ is the covariance matrix and $\mathbf{I_F}(\Theta)$ is the Fisher information matrix given by

$$\mathbf{I_F}_{i,j}(\Theta) = -\mathrm{E}\left[\frac{\partial^2 \ln p(\tilde{\mathbf{s}}, \Theta)}{\partial \Theta_i \partial \Theta_j}\right]. \tag{47}$$

Consider the PDF given by

$$p(\tilde{\mathbf{s}}, \Theta) = \frac{1}{(2\pi\sigma^2)^{\frac{M}{2}}} \exp\left(-\frac{1}{2\sigma^2} \sum_{m=1}^{M} (\tilde{s}_m - s_m(\Theta))^2\right), \tag{48}$$

in which case, the covariance matrix is lower bounded by

$$\mathbf{C}_{\hat{\Theta}} \geq \mathbf{I_F}^{-1}(\Theta), \tag{49}$$

where the Fisher information matrix is written explicitly by

$$\mathbf{I_F} = \mathbf{\Phi_S^H R^{-1} \Phi_S}, \tag{50}$$

where

$$\mathbf{\Phi_S} = \begin{bmatrix} \frac{\partial s_1}{\partial \Theta_1} & \frac{\partial s_1}{\partial \Theta_2} & \cdots & \frac{\partial s_1}{\partial \Theta_{2K}} \\ \frac{\partial s_2}{\partial \Theta_1} & \frac{\partial s_2}{\partial \Theta_2} & \cdots & \frac{\partial s_2}{\partial \Theta_{2K}} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial s_M}{\partial \Theta_1} & \frac{\partial s_M}{\partial \Theta_2} & \cdots & \frac{\partial s_M}{\partial \Theta_{2K}} \end{bmatrix}. \tag{51}$$

Hence, the variance of the estimator is bounded by

$$\mathrm{var}(\hat{\Theta}) \geq \mathrm{diag}\left(\mathbf{I_F}^{-1}\right). \tag{52}$$

APPENDIX B
TRAINING ERROR OF THE ESTIMATOR

We will further analyze the training error and the underlying model order selection framework developed in Section III-C. For a given parameter set $\Theta = [t_1, t_2, \ldots, t_K, a_1, a_2, \ldots, a_K]^T$, we assume that available samples are degraded versions of the true samples $\tilde{s}_m = s_m + \tilde{e}_m$ where $\tilde{e}_m = \tilde{e}_{\mathcal{R}m} + i\tilde{e}_{\mathcal{I}m}$ is complex additive white Gaussian noise with variance σ^2. Let us assume $Q_{\hat{k}}(\tilde{s})$ is a nonlinear denoising method that includes a nonlinear estimation method; e.g., the data fitting framework in our case. Then, the computation of the estimated complex moment is given by

$$Q_{\hat{k}}(\tilde{s}) = \sum_{k=0}^{\hat{K}-1} \hat{a}_k e^{\alpha_m \frac{\hat{t}_k}{T}}, \tag{53}$$

where $\{\hat{t}_k, \hat{a}_k\}_{k=1\ldots\hat{k}}$ are the estimated parameters for a model order $\hat{K}$. Finally, the training error curve is defined as

$$E_T(\hat{k}) = \left\|Q_{\hat{k}}(\tilde{s}) - \tilde{s}\right\|_2^2, \tag{54}$$

where $\hat{k} = [1, \ldots, \frac{M}{2}]$ and M is the number of samples of $\tilde{s}$. We can than write

$$\mathbb{E}\left[E_T(\hat{k})\right] = \mathbb{E}\left[\left\|Q_{\hat{k}}(\tilde{s}) - s\right\|_2^2\right] + \mathbb{E}\left[\|s - \tilde{s}\|_2^2\right] \\ - 2\mathbb{E}\left[\left\|\tilde{e}^H\left(Q_{\hat{k}}(\tilde{s})\right)\right\|_2^2\right] \tag{55}$$

where the three terms are known as the true MSE of the estimator, the noise power, i.e., $M\sigma^2$, and the divergence of the estimator for additive Gaussian noise case, respectively. Although this is well-known from Stein's unbiased risk estimate for real valued functions [40], [41], we develop the case for the complex valued functions for the sake of completeness. We work on the following expectation

$$\mathbb{E}\left[\left\|\tilde{e}^H(Q_{\hat{k}}(\tilde{s}))\right\|_2^2\right] = \mathbb{E}\left[\left\|\tilde{e}_I^H\left(Q_{\hat{k}}(\tilde{s})\right)\right\|_2^2\right] \\ - i\mathbb{E}\left[\left\|\tilde{e}_Q^H\left(Q_{\hat{k}}(\tilde{s})\right)\right\|_2^2\right],$$

where the first and the second terms are the real and the imaginary parts, respectively. We first assume the mapping $Q_{\hat{k}} : \mathbb{C}^\mathbb{M} \to \mathbb{C}^\mathbb{M}$ be weakly differentiable. Then,

$$\mathbb{E}\left[\left\|\tilde{e}_I^H\left(Q_{\hat{k}}(\tilde{s})\right)\right\|_2^2\right] \tag{56}$$

$$= \int p(\tilde{\mathbf{e}}_\mathbf{I}) \sum_{m=1}^{M} \tilde{e}_{I_m} Q_{\hat{k}}(s + \tilde{\mathbf{e}}_\mathbf{I} + i\tilde{\mathbf{e}}_\mathbf{Q}) d^M \tilde{\mathbf{e}}_\mathbf{I} \tag{57}$$

$$= \sum_{m=1}^{M} \int p(\tilde{\mathbf{e}}_\mathbf{I}) \tilde{e}_i Q_{\hat{k}}(s + \tilde{\mathbf{e}}_\mathbf{I} + i\tilde{\mathbf{e}}_\mathbf{Q}) d^M \tilde{\mathbf{e}}_\mathbf{I} \qquad (58)$$

$$= \sum_{m=1}^{M} \int \frac{-\sigma^2}{2} \frac{\partial}{\partial \tilde{e}_{I_m}} p(\tilde{\mathbf{e}}_\mathbf{I}) Q_{\hat{k}}(s + \tilde{\mathbf{e}}_\mathbf{I} + i\tilde{\mathbf{e}}_\mathbf{Q}) d^M \tilde{\mathbf{e}}_\mathbf{I} \quad (59)$$

$$= \mathbb{E}\left[\frac{\sigma^2}{2} \mathrm{div}_{\tilde{e}_I} Q_{\hat{k}}(\tilde{s}) \right], \qquad (60)$$

where $p(\tilde{e}_I)$ is the PDF of the real part of the noise and $\mathrm{div}_{\tilde{e}} = \sum_{m=1}^{M} \frac{\partial}{\partial \tilde{e}_m}$ is the divergence with respect to real part. Hence, the second term is written similarly

$$\mathbb{E}\left[\left\| \tilde{e}_Q^H \left(Q_{\hat{k}}(\tilde{s}) \right) \right\|_2^2 \right] = \mathbb{E}\left[\frac{\sigma^2}{2} \mathrm{div}_{\tilde{e}_Q} Q_{\hat{k}}(\tilde{s}) \right], \qquad (61)$$

which finally yields

$$\mathbb{E}\left[\left\| \tilde{e}^H \left(Q_{\hat{k}}(\tilde{s}) \right) \right\|_2^2 \right] = \frac{\sigma^2}{2} \mathbb{E}\left[\left(\mathrm{div}_{\tilde{e}_I} - i \mathrm{div}_{\tilde{e}_Q} \right) Q_{\hat{k}}(\tilde{s}) \right]. \qquad (62)$$

We note that this can be useful since one can define an unbiased MSE estimator using the training error, the noise variance σ^2 and the divergence of the estimator without knowing the true value of the signal [40], [41]. However, although an approximation of the divergence can be done using a Monte-Carlo SURE approach, computing the divergence of a nonlinear estimator is always a challenging problem itself. Hence, we leave this discussion at this point and focus only on the training error curve. The model selection requires to choose the number of degrees of freedom in our estimation problem. We rather proposed a model selection framework in Section III-C following our observation that the training curve exhibits a U-curve pattern. This can be easily seen from the first two terms of the analyze in (55) and ignoring the last term. Obviously, the second term does not change with respect to our mapping $Q_{\hat{k}}$. Hence, the first term, i.e., the MSE of the estimator, will dominate the training error for the under and the over fitting values of model order $\hat{k}$ as there will be a contribution coming from the bias of the estimator. However, for the true model order, the estimator MSE will be the variance for an unbiased estimator. Therefore, once the training error hits the fitting region defined by a Plato behavior, we choose the minimum of model order in this plato to favor a parsimonious model.

REFERENCES

[1] C. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, no. 1, pp. 10–21, Jan. 1949.

[2] M. Unser, "Sampling-50 years after shannon," *Proc. IEEE*, vol. 88, no. 4, pp. 569–587, Apr. 2000.

[3] M. Unser, "Splines: A perfect fit for signal and image processing," *IEEE Signal Process. Mag.*, vol. 16, no. 6, pp. 22–38, Nov. 1999.

[4] M. Unser and T. Blu, "Wavelet theory demystified," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 470–483, Feb. 2003.

[5] 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, 2002, 2006.

[6] 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, 2008.

[7] R. Tur, Y. 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.

[8] J. Oñativia, S. R. Schultz, and P. L. Dragotti, "A finite rate of innovation algorithm for fast and accurate spike detection from two-photon calcium imaging," *J. Neural Eng.* vol. 10, no. 4, p. 046017, 2013 [Online]. Available: http://stacks.iop.org/1741-2552/10/i=4/a=046017

[9] P. 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.

[10] A. Erdozain and P. Crespo, "Reconstruction of aperiodic fri signals and estimation of the rate of innovation based on the state space method," *Signal Process.*, vol. 91, no. 8, pp. 1709–1718, 2011.

[11] J. Berent, P. 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.

[12] I. Maravic 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.

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

[14] 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, January 2014.

[15] Z. Dogan, T. Blu, and D. Van De Ville, "Finite-rate-of-innovation for the inverse source problem of radiating fields," *Sampl. Theory Signal Image Process.*, vol. 13, no. 3, pp. 271–294, 2014.

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

[17] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1997.

[18] I. Maravic 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.

[19] K. Gedalyahu, R. Tur, and Y. Eldar, "Multichannel sampling of pulse streams at the rate of innovation," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1491–1504, Apr. 2011.

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

[21] B. Rao and K. Arun, "Model based processing of signals: A state space approach," *Proc. IEEE*, vol. 80, no. 2, pp. 283–309, Feb. 1992.

[22] R. Roy and T. Kailath, "Esprit-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Signal Process., Acoust., Speech*, vol. 37, no. 7, pp. 984–995, Jul. 1989.

[23] J. A. Cadzow, "Signal enhancement: A composite property mapping algorithm," *IEEE Trans. Signal Process., Acoust., Speech*, vol. 36, pp. 39–62, 1988.

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

[25] C. Gilliam and T. Blu, "Fitting instead of annihilation: Improved recovery of noisy fri signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2014, pp. 51–55.

[26] A. Erdozain and P. M. Crespo, "A new stochastic algorithm inspired on genetic algorithms to estimate signals with finite rate of innovation from noisy samples," *Signal Process.* vol. 90, no. 1, pp. 134–144, 2010 [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165168409002564

[27] A. Wein and L. Srinivasan, "Iterml: A fast, robust algorithm for estimating signals with finite rate of innovation," *IEEE Trans. Signal Process.*, vol. 61, no. 21, pp. 5324–5336, Nov. 2013.

[28] X. Wei and P. Dragotti, "Universal sampling of signals with finite rate of innovation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2014, pp. 1803–1807.

[29] A. Davison, *Statistical Models*, ser. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[30] 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.

[31] G. Strang and G. Fix, "A Fourier analysis of the finite element variational method," in *Constructive Aspects of Functional Analysis*, ser. C.I.M.E. Summer Schools, G. Geymonat, Ed. Berlin, Germany: Springer, 2011, vol. 57, pp. 793–840, doi = 10.1007/978-3-642-10984-3_7.

[32] M. Unser and T. Blu, "Cardinal exponential splines: Part itheory and filtering algorithms," *IEEE Trans. Signal Process.*, vol. 53, pp. 1425–1438, 2005.

[33] S. Kay, *Modern Spectral Estimation: Theory and Application*, ser. Prentice-Hall signal processing series. Englewood Cliffs, NJ, USA: Prentice-Hall, 1999.

[34] P. Lemmerling, L. Vanhamme, S. V. Huffel, and B. D. Moor, "Iqml-like algorithms for solving structured total least squares problems: A unified view," *Signal Process.* vol. 81, no. 9, pp. 1935–1945, 2001 [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165168401000937

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

[36] V. Nagesha and S. Kay, "On frequency estimation with the iqml algorithm," *IEEE Trans. Signal Process.* vol. 42, no. 9, pp. 2509–2513, Sep. 1994 [Online]. Available: http://dx.doi.org/10.1109/78.317875

[37] M. Clark and L. Scharf, "On the complexity of iqml algorithms," *IEEE Trans. Signal Process.* vol. 40, no. 7, pp. 1811–1813, Jul. 1992 [Online]. Available: http://dx.doi.org/10.1109/78.143451

[38] D. Tufts and R. Kumaresan, "Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood," *Proc. IEEE*, vol. 70, no. 9, pp. 975–989, Sept. 1982.

[39] Z. Dogan, T. Blu, and D. Van De Ville, "Detecting spontaneous brain activity in functional magnetic resonance imaging using finite rate of innovation," in *Proc. IEEE 11th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2014, pp. 1047–1050.

[40] F. Luisier, T. Blu, and M. Unser, "A new SURE approach to image denoising: Interscale orthonormal wavelet thresholding," *IEEE Trans. Image Process.*, vol. 16, no. 3, pp. 593–606, March 2007.

[41] T. Blu, "The SURE-LET methodology—A prior-free approach to signal and image denoising," presented at the 8th Int. Workshop on Sampling Theory Appl. (SampTA), May 18–22, 2009, plenary presentation.

Zafer Doğan (M'11) received his B.Sc. degree in electrical and electronics engineering from Middle East Technical University (METU), Turkey, his M.Sc. and Ph.D. degree in electrical engineering from Ecole Polytechique Federale de Lausanne (EPFL), Switzerland, in 2009, 2011 and 2015, respectively. Currently, he continues as a postdoctoral researcher at the Medical Image Processing Lab (MIPLAB) affiliated to the EPFL and the University of Geneva, Switzerland. His research interests include signal and image processing (e.g., sparse signal models, sampling, approximation and multi-resolution analysis) and inverse problems in (nonlinear) tomography.

Christopher Gilliam (M'10) is a Postdoctoral Fellow in the Image and Video Processing Group, Department of Electronic Engineering at The Chinese University of Hong Kong. He graduated from Imperial College London with a First Class Honors MEng in Electrical and Electronic Engineering in 2008 and with a Ph.D. in Electrical and Electronic Engineering in 2013. His research interests include sampling theory, approximation theory and its connection to the sampling of parametric non-bandlimited signals, and applications of sampling theory to image processing, in particular Image-Based Rendering (IBR).

Thierry Blu (M'96–SM'06–F'12) was born in Orléans, France, in 1964. He received the Diplôme d'ingénieur from École Polytechnique, France, in 1986 and from Télécom Paris (ENST), France, in 1988. In 1996, he obtained a Ph.D. 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 now a Professor in the Department of Electronic Engineering, the 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 was the recipient of 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 (20082013), 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).
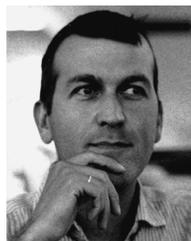
Dimitri Van De Ville (M'02–SM'12) received the M.S. degree in engineering and computer sciences and the Ph.D. degree from Ghent University, Belgium, in 1998, and 2002, respectively. After a post-doctoral stay (2002–2005) at the Biomedical Imaging Group of Prof. Michael Unser at the École Polytechnique Fédérale de Lausanne (EPFL), Switzerland, he became responsible for the Signal Processing Unit at the University Hospital of Geneva, Switzerland, as part of the Centre dImagerie Biomédicale (CIBM). In 2009, he has been the recipient of a Swiss National Science Foundation professorship and since 2015 became Professor of Bioengineering at the EPFL and the University of Geneva, Switzerland. His research interests include wavelets, sparsity, pattern recognition, and their applications in biomedical imaging, such as functional magnetic resonance imaging. He received the Pfizer Research Award 2012 in the category Neurosciences for his work on scale-free dynamics of EEG microstates.

Dr. Van De Ville served as an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING from 2006 to 2009 and the IEEE SIGNAL PROCESSING LETTERS from 2004 to 2006. He was the Chair of the Bio Imaging and Signal Processing (BISP) TC of the IEEE Signal Processing Society (2012–2013). He is Co-Chair of the biennial Wavelets & Sparsity series conferences, together with V. Goyal and M. Papadakis.